College of Polytechnics Jihlava

31st International Conference

Mathematical Methods in Economics 2013

Proceedings – Part I.

11 – 13 September 2013

Jihlava, Czech Republic

Proceedings of the 31st International Conference
Mathematical Methods in Economics 2013
Publisher: College of Polytechnics Jihlava, Tolstého 16, Jihlava, Czech Republic
Print: Polygrafický servis Brno, František Šabart, Srbská 28
Editor: Hana Vojáčková

This publication is not a subject of language check.
Papers are sorted by authors names in alphabetical order.
All papers passed a blind review process.

ISBN: 978-80-87035-76-4

Programme Committee

Martina Kuncová
Department of Economic Studies, College of Polytechnics Jihlava
Petr Fiala
Faculty of Informatics and Statistics, University of Economics, Prague
Jana Hančlová
Faculty of Economics, Technical University of Ostrava
Marie Hojdarová
Department of Mathematics, College of Polytechnics Jihlava
Josef Jablonský
Faculty of Informatics and Statistics, University of Economics, Prague
Ladislav Lukáš
Faculty of Economics, University of West Bohemia, Pilsen
Jan Pelikán
Faculty of Informatics and Statistics, University of Economics, Prague
Karel Sladký
Institut of Informatics Theory and Automation of the Academy of Science sof CR, Prague
Tomáš Šubrt
Faculty of Economics and Managment, University of Life Science, Prague
Karel Zimmermann
Faculty of Mathematics and Physics Charles University, Prague

Organizing Committee

Zbyněk Bureš
Kristine Karakhanyan
Andrea Kubišová
Martina Kuncová
Lenka Lízalová
Tomáš Souček
Hana Vojáčková

Conference website: https://mme2013.vspj.cz/

List of Referees:

Bartoška Jan
Bisová Sára
Borovička Adam
Brožová Helena
Cahlík Tomáš
Černý Michal
Čížek Ondřej
Dlouhá Zuzana
Dlouhý Martin
Domeová Ludmila
Fábry Jan
Fiala Petr
Formánek Tomáš
Friebel Ludvík
Fribelová Jana
Hančlová Jana
Havlíček Jaroslav
Hedija Veronika
Hojdarová Marie
Horský Richard
Houšek Roman
Houška Milan
Houšková – Beránková Martina
Jablonský Josef
Jiříček Petr
Kaňková Vlasta
Klicnar Martin
Klicnarová Jana
Klůfa Jindřich
Kobzareva Maria
Kodera Jan
Kopa Miloš
Krejčí Igor
Krištůfek Ladislav
Kučera Petr
Kuncová Martina
Lukáš Ladislav
Mazouch Petr
Melecký Lukáš
Pánková Václava
Pelikán Jan
Plevný Miroslav
Princ Peter
Ramík Jaroslav
Říčař Michal
Sladký Karel
Sekničková Jana
Staníčková Michaela
Šimek Milan
Šimpach Ondřej
Školuda Václav
Šmíd Martin
Šubrt Tomáš
Tichý Tomáš
Tvrz Stanislav
Vašíček Osvald
Vlach Milan
Vltavská Kristýna
Vošvrda Miloslav
Zámková Martina
Zavacká Jana
Zimmermann Karel
Zmeškal Zdeněk
Zouhar Jan
Zouharová Martina

Contents

Part I.

Part II.

Decision support system for plant variety testing in the Czech Republic

Václav Adamec[1], David Hampel[2], Jitka Janová[3]

Abstract. Plant variety testing through field experiments and subsequent statistical evaluations provide necessary information for variety entry in national and international registers. A newly registered variety must possess unique properties making it distinct from other already registered varieties. It must be uniform, stable and provide original characteristics, which set it above the current level of production potential. Qualified decision-making about variety registration is conditioned upon availability of relevant information. The current paper describes the process of creating a support system for decision-making (DSS) to assist planning and evaluations of field experiments according to guidelines set by the National Plant Variety Office of Central Institute for Supervising and Testing in Agriculture. The decision support system shall represent a powerful instrument to provide assistance in the area of quantitative analysis, a part of the variety testing and evaluations. DSS generates optimum design of field experiments through suitable placement of varieties in field trial plans and secures the follow-up statistical analysis of the experimental results. The paper describes statistical and mathematical methods implemented in original or modified form and focuses on modules of the decision support system as a whole. Issues of validating the implemented algorithms are discussed. The DSS system was implemented in the MATLAB language and supplied with suitable user interface.

Keywords: decision support system, field experiments, NNA, plant variety testing, REML.

JEL classification: C13
AMS classification: 65C20

1 Introduction

During the next 50 years, new challenges are expected to appear in the sector of agriculture: increase food production, reduce soil and ecosystem degradation, efficiently use and protect water resources and cut emissions of greenhouse gases. Developing new plant varieties can help tackle these tasks. It can be accomplished by intensive breeding with the objective to bring new plant varieties with better resistance, higher yield and lower need of fertilizers, water and energy.

Listing in National Register is a legal requirement for new varieties of main plant species, which seeks to ensure that no variety can be marketed unless it is genuinely new, and it constitutes an improvement to the varieties already being marketed. For farmers, the National Listing guarantees value for cultivation, quality of the planting material, and also safety to human health, animals, plants and the environment. Legal framework for variety listing in National Register is provided by the International Treaty on Legal Protection of New Plant Varieties of 1991. In the EU it is determined by EC directive No. 2100/94 and in the Czech Republic by Act No. 408/2000 Coll., Act No. 147/2002 Coll. and Act No. 219/2003 Coll.

Authorities responsible for protecting plant variety rights are the Central Institute for Supervising and Testing in Agriculture (ÚKZÚZ) and the Ministry of Agriculture (MA). Testing and certification of

[1]Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operational Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, vadamec@mendelu.cz
[2]Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operational Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, qqhampel@mendelu.cz
[3]Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operational Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, janova@mendelu.cz

newly developed or imported plant varieties in the Czech Republic is carried out by the National Plant Variety Office of the Central Institute for Supervising and Testing in Agriculture (NOÚ ÚKZÚZ), an organization of the MA. Objective of testing and certification is to verify that a new variety is distinct from others, uniform in its characteristics and stable in the long run. Impartial variety testing provides qualitative and quantitative information essential for entry in the National Register.

Administration of modern certification system, independent testing, advice on the varieties and identification of the most profitable planting material for farmers creates on the side of the ÚKZÚZ the need for efficient system capable of handling experimental design, processing primary data coming from the testing stations and performing statistical analyses and evaluations of test data. The objective of this project is development of computerized decision support system (DSS) generating design of variety trials, speedy statistical analyses of experimental data and presentation of the results in user-friendly format (Excel). DSS must be able to provide quality analyses of large datasets via graphical user interface and offer reliable transport of results to the end user. Open code must permit updates of the statistical procedures, if required. The DSS modules are programmed in modern MATLAB language and computing environment, thus allowing full utilization of computing power in current computer hardware.

The DSS system is expected to assist the ÚKZÚZ in the process of awarding registrations and breeder's rights of the varieties, provide valuable information about specific varieties to the farmers, the public and the companies processing and marketing agricultural commodities and recommend novel domestic and foreign crop varieties to be cultivated and marketed. Inevitably, DSS will promote economic viability of Czech farmers in the EU agricultural market and increase reputation of the Czech testing and certification system. DSS may also assist the MA in implementation of the country's and EU policies in agriculture.

DSS can be accessed via graphical user interface. The modules provide planning of field experiments, evaluation of a single trial or entire experimental series. As a result, the DSS modules can be used on data from several years or applied to isolated experiments. Main parts of the DSS are shown in Fig. 1.

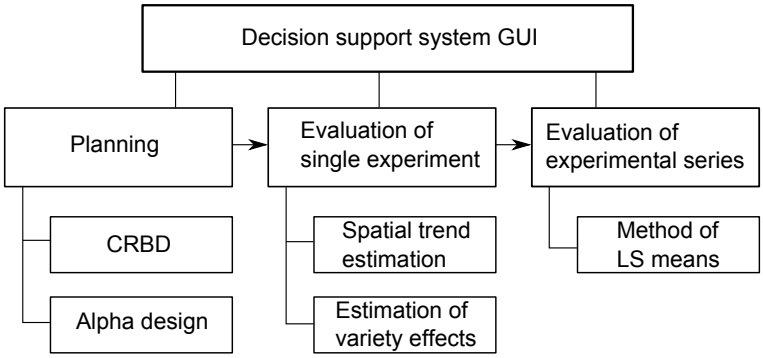


Figure 1 Modules of the DSS system

A typical application of DSS begins with construction of appropriate experimental design for several testing stations and a calendar year by a cultivar analyst. The experimental design is either the Completely Randomized Block Design (CRBD) or the Alpha Design [3], depending on the number of varieties in test. Following the field trial completion, the analyst enters the data and analyzes individual experiments. Approximate impact of soil heterogeneity on the crop yields is revealed by the Nearest Neighbor Analysis (NNA) of spatial trend [12]; final estimates of variety effects are obtained by AI-REML [2]. The analyst then combines estimates from the current year with historical results and evaluates the whole experimental series by the method of LS means. The output of the Fisher's Least Significant Difference (LSD) then allows to make inferences about means of specific varieties, testing stations, calendar years and testing systems. The following text focuses on estimation of spatial trend by the NNA and mixed model analysis of trial data by AI-REML with subsequent pair-wise comparisons.

2 Nearest Neighbor Adjustment for spatial trend

The NNA method of adjusting the crop variety yields for spatial heterogeneity of soil and water environment, i.e. spatial trend, was implemented in the second module. Computing algorithm of the Wilkinson's NNA method was described by Stroup *et al.* [11], who also provided own implementation in SAS code.

The Stroup's realization of NNA mimics functionality of the ANOFT software programmed in Visual Basic 3.0 by Erik Schwarzbach [8] in early 1980s.

In theory, identical performance of the named NNA variants is anticipated in crop yield adjustment for environmental heterogeneity and estimation of local environmental trend used for construction of the trend map. The trend map of the field experiment displays the estimated effects of the local soil and environmental trends on the observed crop yield. It is prepared in tabular and graphical form.

In the field experiment, the plots of rectangular shape are arranged linearly in block design. The varieties are assigned to the plots according to prepared design [3], a single variety per row. A variety must not be placed in adjacent plots nor in marginal plots only, since this arrangement would prevent separation of the variety effects on the observed yield from the environmental trend by statistical methods and void the whole experiment. The scheme in Fig. 2 illustrates the customary arrangement of the testing plots in the field experiment with three rows; y_{ijk} denotes observed yield of the k-th variety in i-th row and j-th column; e_{ijk} denotes deviation of the observed k-th variety yield in i-th row and j-th column from the mean yield of the k-th variety.

		$e_{i-1,j}$		
	$e_{i,j-1}$	y_{ijk}	$e_{i,j+1}$	
		$e_{i+1,j}$		

Figure 2 Arrangement of plots in the testing experiment

The NNA analysis by Wilkinson can be described in the following steps:

1. Firstly, arithmetic means of crop yield $\bar{y}_{..k}$ are calculated for every variety included in the testing experiment. Then, deviations of observed yield y_{ijk} from the mean yield of the corresponding k-th variety are received for all plots in the testing trial by the formula $e_{ijk} = y_{ijk} - \bar{y}_{..k}$.
2. Two auxiliary spatial trend variables are formed from the deviated observations. Within the i-th row, averaging two neighboring deviations produces variable κ for the observed yield y_{ijk}. For the marginal observations, means are obtained from two closest deviations within the same row. Variable κ is a reflection of the spatial trend within a row. It is expected to exert significant influence on the observed yield. The second variable ω is derived for observations of the middle rows by averaging two neighboring deviations from the rows on the right and left side of the observation y_{ijk}. This ω variable reflects the spatial trend of observations in the same column. Column trend is assumed to have marginal effect on the results of crop testing.
3. An auxiliary model of ANCOVA with no intercept $y_{ijk} = \alpha_k + \beta_1\kappa_{ij} + \beta_2\omega_{ij} + \varepsilon_{ijk}$ is constructed in the next step. It describes variability of observed yield y_{ijk} as a function of variety α_k, spatial trend κ_{ij} within a row, local trend ω_{ij} pertaining to column and random disturbance ε_{ijk}.
4. Estimated variety yields $\hat{\alpha}_k$ partially corrected for local trends and estimated errors $\hat{\varepsilon}_{ijk}$ are extracted there from. Error sum of squares (ESS) is calculated.
5. In the iterative variant of NNA, deviations of observed yield y_{ijk} from estimated expected yield of the varieties from the preceding round of iteration are calculated. Points 2 through 4 are then repeated until change in ESS from two subsequent iterations becomes negligible.

After convergence, estimated coefficients of variety yields from the Wilkinson method corrected for local trend can be compared to variety estimates from ANOFT software by paired t-test. In addition, F-test of concurrent restrictions for parameters of linear calibration model $Y_{NNA} = \beta_0 + \beta_1 Y_{ANOFT} + \varepsilon$ with $H_0 : \beta_0 = 0$ and simultaneously $\beta_1 = 1$, described in [1], can verify similarity of variety yield estimates received from the two methods.

Upon convergence, estimated deviations of spatial trend for rows and columns associated with the experimental plots are expressed on the relative scale. The diagram of the trend map then depicts the local trend deviations exceeding 5 % of the overall mean in absolute value. The authors' implementation of NNA in MATLAB was validated on datasets provided by the recipient institution. Testing data available from ÚKZÚZ were affected by severe, medium-size and near absent environmental trends. Estimates of trend-adjusted yield and trend map from the authors' NNA software and the ANOFT program by Erik Schwarzbach were available to evaluate performance of NNA applied in the current software.

3 Mixed model estimation by Average Information REML

The estimation procedure in linear models with fixed and random effects includes interconnected processes of estimating the fixed effects, predicting the random effects and estimation of variance components. Estimation of the fixed effects is often realized by OLS or GLS; prediction of random effects is based on theory developed early by C. R. Henderson and it is referred to as the Best Linear Unbiased Prediction (BLUP). Estimation of variance components can be realized by the method of Maximum Likelihood (ML), see [4], which in 1960s replaced Henderson's methods I, II and III. A few years later, the Residual Maximum Likelihood (REML) substituted ML, and currently, it is the superior method.

Theoretical foundations of REML can be traced back to Patterson and Thompson's paper [6]. REML maximizes the residual (restricted) likelihood derived from the error contrasts after filtering out the fixed effects by OLS or GLS. REML method corrects for the bias in ML estimates of variance components by accounting for the degrees of freedom attributed to the fixed effects. Residual Maximum Likelihood is a nonlinear function of the variance components; therefore, iterative algorithms must be applied to estimate the variances.

Computational algorithms of REML can be either derivative (based on first or second order derivatives of the natural logarithm of likelihood) or derivative-free (DF-REML), which use other principles. Average Information REML (AI-REML) belongs to the second-order derivative methods along with method of Newton-Raphson (NR) and Fisher-Scoring (FS). In estimation of variance components, however, AI-REML uses the average information matrix. Among the second-order derivative methods, AI-REML is the most computationally effective, especially when applied to large datasets. Obtaining estimated variance components in AI-REML requires less iteration in comparison with other derivative methods. For its favorable properties, AI-REML algorithm has been implemented in many software products, e.g. AS-REML or GenStat [2], [10].

Symbol n designates sample size; p indicates number of unique levels of varieties; k denotes number of random effects excluding error, and q_m denotes number of levels of the m-th random effect. We start with model equation with fixed effects of varieties and random effects of blocks and replications

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y}$ is column vector storing observations of the crop yield in experimental plots $n \times 1$, $\boldsymbol{X}$ is full column rank design matrix for the fixed effects of varieties without intercept $n \times p$, $\boldsymbol{\beta}$ is vector of fixed effects parameters for varieties $p \times 1$, $\boldsymbol{Z}$ is design matrix for random effects of blocks and replications without constant $n \times (q_1 + q_2)$, $\boldsymbol{u}$ is vector of random parameters of blocks and replications $(q_1 + q_2) \times 1$, $\boldsymbol{\varepsilon}$ is column vector of random disturbances $n \times 1$.

Expectation of the dependent variable is $E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta}$. Covariance matrix of the dependent variable is $D(\boldsymbol{y}) = \boldsymbol{H} = \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^T + \boldsymbol{R}$, where $\boldsymbol{G} = \bigoplus_{m=1}^{k} \sigma_m^2 \boldsymbol{I}_{q_m}$ is a block-diagonal covariance matrix[1] of other than error random effects and $\boldsymbol{R}$ is covariance matrix of the error terms. The design matrix $\boldsymbol{Z}$ for non-error random effects can be partitioned to matrices of the random effects, i.e. $\boldsymbol{Z} = [\boldsymbol{Z}_1 \boldsymbol{Z}_2]$. Multivariate distribution of the random effects is assumed

$$\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{\varepsilon} \end{bmatrix} \sim N\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{G} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R} \end{bmatrix}\right).$$

Absorption matrix $\boldsymbol{P}$ is defined in the form

$$\boldsymbol{P} = \boldsymbol{H}^{-1} - \boldsymbol{H}^{-1}\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1}.$$

Natural logarithm of the restricted likelihood for linear model with fixed and random effects (mixed model) can be written

$$\ell = -\frac{1}{2}\left(\ln|\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X}| + \ln|\boldsymbol{H}| + (n-p)\ln\sigma_0^2 + \boldsymbol{y}^T\boldsymbol{P}\boldsymbol{y}/\sigma_0^2\right) + c.$$

Assuming vector of variances of random effects to be estimated is $\boldsymbol{\kappa} = (\sigma_0^2, \sigma_1^2, \sigma_2^2)$. Symbol σ_0^2 indicates error variance, σ_1^2 and σ_2^2 denote variances of blocks and replications, respectively and c is a constant.

[1] Note, that $\boldsymbol{A} \oplus \boldsymbol{B} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B} \end{bmatrix}$ is the direct sum of $\boldsymbol{A}$ and $\boldsymbol{B}$.

For mixed model with three random effects of error, block and replication, the first partial derivatives of log-likelihood with respect to variance components can be written [2]

$$\frac{\partial \ell}{\partial \sigma_0^2} = -\frac{1}{2}\left((n-p)/\sigma_0^2 - \boldsymbol{y}^T\boldsymbol{P}\boldsymbol{y}/\sigma_0^4\right),$$
$$\frac{\partial \ell}{\partial \sigma_1^2} = -\frac{1}{2}\left(tr(\boldsymbol{P}\boldsymbol{Z}_1\boldsymbol{Z}_1^T) - \boldsymbol{y}^T\boldsymbol{P}\boldsymbol{Z}_1\boldsymbol{Z}_1^T\boldsymbol{P}\boldsymbol{y}/\sigma_0^2\right),$$
$$\frac{\partial \ell}{\partial \sigma_2^2} = -\frac{1}{2}\left(tr(\boldsymbol{P}\boldsymbol{Z}_2\boldsymbol{Z}_2^T) - \boldsymbol{y}^T\boldsymbol{P}\boldsymbol{Z}_2\boldsymbol{Z}_2^T\boldsymbol{P}\boldsymbol{y}/\sigma_0^2\right).$$

$\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are the partial design matrices for random blocks and replications, respectively. In the next step, elements of the observed information matrix $-\frac{\partial^2 \ell}{\partial \kappa_i \partial \kappa_j}$ in Newton-Raphson or the expected information matrix $-E\left(\frac{\partial^2 \ell}{\partial \kappa_i \partial \kappa_j}\right)$ in Fisher-Scoring must be initiated. For estimating the mixed model by AI-REML, "average" information matrix $\boldsymbol{A}$ is assembled from elements

$$\boldsymbol{A}(\sigma_0^2,\sigma_0^2) = \frac{1}{2}\boldsymbol{y}^T\boldsymbol{P}\boldsymbol{y}/\sigma_0^6, \qquad \boldsymbol{A}(\sigma_0^2,\sigma_1^2) = \frac{1}{2}\boldsymbol{y}^T\boldsymbol{P}\boldsymbol{Z}_1\boldsymbol{Z}_1^T\boldsymbol{P}\boldsymbol{y}/\sigma_0^4,$$
$$\boldsymbol{A}(\sigma_0^2,\sigma_2^2) = \frac{1}{2}\boldsymbol{y}^T\boldsymbol{P}\boldsymbol{Z}_2\boldsymbol{Z}_2^T\boldsymbol{P}\boldsymbol{y}/\sigma_0^4, \qquad \boldsymbol{A}(\sigma_1^2,\sigma_1^2) = \frac{1}{2}\boldsymbol{y}^T\boldsymbol{P}\boldsymbol{Z}_1\boldsymbol{Z}_1^T\boldsymbol{P}\boldsymbol{Z}_1\boldsymbol{Z}_1^T\boldsymbol{P}\boldsymbol{y}/\sigma_0^2,$$
$$\boldsymbol{A}(\sigma_1^2,\sigma_2^2) = \frac{1}{2}\boldsymbol{y}^T\boldsymbol{P}\boldsymbol{Z}_1\boldsymbol{Z}_1^T\boldsymbol{P}\boldsymbol{Z}_2\boldsymbol{Z}_2^T\boldsymbol{P}\boldsymbol{y}/\sigma_0^2, \qquad \boldsymbol{A}(\sigma_2^2,\sigma_2^2) = \frac{1}{2}\boldsymbol{y}^T\boldsymbol{P}\boldsymbol{Z}_2\boldsymbol{Z}_2^T\boldsymbol{P}\boldsymbol{Z}_2\boldsymbol{Z}_2^T\boldsymbol{P}\boldsymbol{y}/\sigma_0^2.$$

Estimated components of variance are obtained by iteration using the scoring method [9]

$$\boldsymbol{\kappa}^{(j+1)} = \boldsymbol{\kappa}^{(j)} + \left(\boldsymbol{A}^{(j)}\right)^{-1}\frac{\partial \ell}{\partial \boldsymbol{\kappa}}(\boldsymbol{\kappa} = \boldsymbol{\kappa}^{(j)})$$

until convergence is reached. Symbol $(\boldsymbol{A}^{(j)})^{-1}$ indicates inverted average information matrix and $\frac{\partial \ell}{\partial \boldsymbol{\kappa}}(\boldsymbol{\kappa} = \boldsymbol{\kappa}^{(j)})$ a vector of first derivatives of the log-likelihood ℓ with respect to variance components $\boldsymbol{\kappa}$, evaluated at estimates received in j-th round. Suitable non-negative starting values of variance components $\boldsymbol{\kappa}^{(0)}$ are found by applying several iterations of first-order derivative EM-REML [5], known to be less sensitive to poor starting values, before switching to AI-REML. It is recommended, that AI-REML iteration to convergence be traced.

Upon convergence, estimates from two successive rounds satisfy the criterion $|\kappa_m^{(j)} - \kappa_m^{(j-1)}| < 10^{-6}$. Fixed effects for varieties are then estimated by OLS or GLS using

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{y}.$$

In the calculation, covariance matrix $\boldsymbol{H}$ originates from the most recent round of iteration. Covariance matrix of the fixed effects of varieties $D(\hat{\boldsymbol{\beta}})$ is estimated by

$$D(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\hat{\sigma}_0^2,$$

where $\hat{\sigma}_0^2$ denotes REML estimate of error variance from the last iteration. Levels of random effects associated with blocks and replications are predicted from

$$\hat{\boldsymbol{u}} = \boldsymbol{G}\boldsymbol{Z}^T\boldsymbol{H}^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right).$$

Fisher's Least Significant Difference $LSD_\alpha(r,s)$ for pairwise differences between mean yields of varieties r and s can be calculated

$$LSD_\alpha(r,s) = t_{1-\frac{\alpha}{2}}(\nu_\varepsilon)\cdot SE\left(\hat{\beta}_r - \hat{\beta}_s\right).$$

Standard errors of the differences between variety means are obtained from

$$SE\left(\hat{\beta}_r - \hat{\beta}_s\right) = \sqrt{D(\hat{\beta}_r) + D(\hat{\beta}_s) - 2Cov(\hat{\beta}_r,\hat{\beta}_s)}$$

using matching elements of $D(\hat{\boldsymbol{\beta}})$ matrix. Symbol ν_ε denotes residual degrees of freedom.

Value of log restricted likelihood, estimated effects of varieties, $LSD_\alpha(r,s)$ for pair-wise comparisons, variance components, predicted random effects and residuals from GenStat and R-software [7] were available to us to verify implementation of the AI-REML algorithm in MATLAB. Statistical methods and models, required output, testing data and results of the original procedures were provided by the recipient institution.

4 Conclusions

At present, plant variety yield data from testing experiments is statistically analyzed by existing software and purchased third-party software working mostly under MS-DOS operating system. It offers limited user comfort and prevents necessary updates to meet current demands for quality and fast data processing, since only precompiled code is now available. Decision support system (DSS) developed by the authors provides transfer of statistical procedures to modern programming platform represented by MATLAB, supports critical updates of the statistical methods, secures speedy analysis of large datasets and provides user-friendly environment and prompt transfer of results to the authorities and the farmers. After completion, DSS system will assist with generating quality decisions about variety registrations, protect breeder's rights, increase amount of information about variety production potential available to the government and the public and thus increase economic endurance of the Czech agriculture.

Acknowledgements

This project was supported by the Agency of Technology of the Czech Republic, grant No. TD010128.

References

[1] Fox, J., and Weisberg, S.: *An R Companion to Applied Regression*, Second Edition, Sage Publications, Inc., 2011.

[2] Gilmour, A. R., Thompson, R., and Cullis, B. R.: Average Information REML: An efficient algorithm for variance parameter estimation in linear mixed models, *Biometrics*, **51**, (1995), 1440–1450.

[3] Hampel, D., Hartmann, J., and Tláskal, M.: Konstrukce alfa designu, In: Biometrické metódy a modely v pódohospodárskej vede, výskume a výučbe. XVIII. Letná škola Biometriky, Račkova dolina, Vydavatelstvo SPU v Nitre: SAPV, 2008, 37–44.

[4] Hartley, H., and Rao, J.: Maximum likelihood estimation for the mixed analysis of variance, *Biometrika*, **54**, (1967), 93–108.

[5] Hoeschele, I.: STAT 5554: *Variance components course pack*, Virginia Polytechnic Institute & State University, Blacksburg, VA, USA, 1998.

[6] Patterson, H. D., and Thompson, R.: Recovery of interblock information when block sizes are unequal, *Biometrika,* **58**, (1971), 545–554.

[7] R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012. http://www.R-project.org/.

[8] Schwarzbach, E.: A new approach in evaluation of field trials, *Vorträge für Pflanzenzüchtung*, **6**, (1984), 249–259.

[9] Searle, S. R., Cassella, G., and McCulloch, C. E.: *Variance Components*. John Wiley & Sons Inc., New York, 2006.

[10] Smith, A. B., Cullis, B. R., and Thompson, R.: The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches, *Journal of Agricultural Science*, **143**, (2005), 449–462.

[11] Stroup, W. W., Baenziger, P. S., and Mulitze, D. K.: A comparison of methods to account for spatial variation in wheat yield trials, In: *Proceedings of the 19-th. hard red winter wheat workers conference*, Lincoln, NE, 1992, 113–117.

[12] Wilkinson, G. N., Eckert, S. R., Hancock, T. W., and Mayo, O.: Nearest Neighbour (NN) Analysis of Field Experiments, *Journal of the Royal Statistical Society*, **45**, (1983), 151–211.

International portfolio selection with Markov processes and liquidity constraints

Enrico Angelelli[1], Sergio Ortobelli[2], Tomáš Tichý[3], Daniele Toninelli[4]

Abstract. This paper proposes an ex-post comparison of portfolio selection strategies applied to 64 stock exchange markets during the period 2008-2011. Assets are considered according to three different classes of liquidity constraints which differ for their average traded daily value. Moreover, in order to simplify the computational complexity of the choice we preselected assets among the global market considering different return characteristics: the joint Markovian behavior of the returns; their association with market stochastic bounds and their ex-ante reward-risk performance. For each liquidity class of assets we compare the ex-post performance of three portfolio selection strategies: one of them is based on the maximization of the Sharpe ratio; the other two are based on the maximization of a reward-risk performance measure that considers a Markovian evolution of the portfolio returns. The ex-post analysis shows that the reward-risk functionals allow better performance in terms of final wealth and demonstrates, with respect to liquidity classes, different features of portfolio composition in terms of assets nationality.

Keywords: portfolio selection strategies, Markov chains, liquidity constraints.

JEL Classification: G11, G01, G17
AMS Classification: 60J22, 91B28, 91G10

1 Introduction

The last financial crises highlight how world financial markets are exposed to risks that are magnified from the close connections between the markets of different countries. Thus, it is clear that it becomes necessary to find instruments that can help investors in their portfolio choices, helping them to maximize their return and minimize the investments' risk. In this paper we carry on an analysis through a study based on data covering the last financial crises (from 2008 to 2011). In particular, the purpose of this paper is to compare, ex-post, three different portfolio selection strategies on three groups of stocks preselected within the global market, for a total of 64 stock exchange markets.

In order to simplify the computational complexity of the task, we preselected assets among the global stock market considering three different classes of liquidity constraints that differ for their average traded daily value. Moreover, we preselected several stocks for each portfolio selection problem, taking into account different return characteristics: the joint Markovian behavior of the portfolio returns (Angelelli and Ortobelli 2009a, 2009b); their association with market stochastic bounds (see also Ortobelli *et al.*, 2011); their ex-ante reward-risk performance. Among Markovian models we essentially distinguish two categories: (i) parametric models (see: Cont and Tankov, 2003; Cox et al., 1979), and (ii) non parametric models (e.g., Angelelli and Ortobelli 2009a; D'Amico, 2006). For the first category the Markovian hypothesis is used for diffusive models of the log returns; in the parametric models the time series are used to estimate the transition among the states. In our research we assume a non-parametric Markovian model for portfolio selection, describing the returns time evolution by means of a homogeneous Markov chain (for other nonparametric models, see also Ait-Sahalia 1996; Stutzer, 1996). With this methodology, the estimation of the future wealth distribution can be a heavy task; nevertheless, the computational complexity can be controlled by means of a Markov chain. Moreover, the states of Markov chain are chosen to induce a recombining effect on the future wealth (Angelelli and Ortobelli 2009a).

[1] University of Brescia, Department Economics and Management, C.da S. Chiara n. 50, angele@eco.unibs.it
[2] University of Bergamo, Department of Management, Economics and Quantitative Methods, via dei Caniana 2, 24127, Bergamo, Italy / VŠB-Technical University of Ostrava, Faculty of Economics, Department of Finance, Sokolská 33, 70121 Ostrava, Czech Republic, sergio.ortobelli@unibg.it.
[3] VŠB-Technical University of Ostrava, Faculty of Economics, Department of Finance, Sokolská 33, 70121 Ostrava, Czech Republic, tomas.tichy@vsb.cz
[4] University of Bergamo, Department of Management, Economics and Quantitative Methods, via dei Caniana 2, 24127, Bergamo, Italy / VŠB-Technical University of Ostrava, Faculty of Economics, Department of Finance, Sokolská 33, 70121 Ostrava, Czech Republic, daniele.toninelli@unibg.it.

For each liquidity class of assets we compare the ex-post performance of three portfolio selection strategies: one of them is based on the maximization of the Sharpe ratio; the other two are based on the maximization of a reward-risk performance that considers a Markovian evolution of the portfolio returns. The comparison of the ex-post final wealth obtained by the optimization of the reward-risk functionals remarks very different characterristics of international portfolio composition among the three liquidity classes of assets. Thus the results emphasize the importance of liquidity constraints in selecting international portfolios.

In the following section of the paper we introduce the reduction of the dimensionality of the portfolio problem within Markov processes and we provide more details about the Markovian behavior of wealth. The results of the empirical analysis are discussed in Section 3.

2 Criteria to reduce the dimensionality of the portfolio problem

Our analysis is applied to *n* risky assets, and in our simulation we assume that investors want to maximize the performance of their choices at a given future time *T* and that the portfolio returns can be described by a homogeneous Markov chain with *N* states defined on a filtered probability space $\left(\Omega,\Im,\left(\Im_t\right)_{0\le t\le\infty},P\right)$. This Markovian hypothesis is useful to forecast future behavior of the wealth. The vector of the positions taken by an investor in the *n* preselected risky assets is denoted by $x=[x_1,\ldots,x_n]'$ and the portfolio return during the period [*t*, *t*+1] is given by $z_{(x),t+1}=x'z_{t+1}=\sum_{i=1}^{n}x_i z_{i,t+1}$, where $z_{t+1}=[z_{1,t+1},\ldots,z_{n,t+1}]'$ is the vector of the gross returns. The objective is to maximize a given performance measure applied to the forecasted future wealth produced by an investment strategy.

In general the approach to future wealth can be considered static or dynamic. In the latter the gross returns are considered following a Markov chain. Therefore, the investor maximizes a functional of future wealth of the type $\max_{x\in S} f\left(W_T\left(z_{(x)}\right)\right)$, where $W_T\left(z_{(x)}\right)$ is the predicted wealth after *T* trading days obtained investing in the portfolio $z_{(x)}=x'z$. On the other side, with the static approach, the investor maximizes a functional of future wealth that is assumed independent from his temporal horizon [0, *T*], i.e. he solves problems of the type $\max_{x\in S} f\left(z_{(x),t+1}\right)$.

In our work we consider assets selected by both the approaches (static and dynamic). More into details, the selection is based on three steps: 1) consider a set of desirable ordering criteria; 2) assets are sorted by each criterion; 3) intersect top ranked assets so that a fixed number of assets are selected (about 100 assets satisfying the dynamic approach criteria and about 50 assets satisfying the static approach criteria). In particular, in our analysis, the selected assets satisfy some properties of wealth behavior and optimality criteria (Ortobelli et al. 2011), such as: 1) Timing (we want both to maximize the time taken for wealth to reach a given lower bound and to minimize the time to reach a given upper bound; 2) An asymptotic and a Markovian behavior of wealth; 3) Consistency with investors' preferences; 4) Association with market stochastic bounds.

A common problem with other portfolio selection strategies is the reduction of the dimensionality of the problem. If this is the target to be generally reached, other studies (for example, Papp *et al*., 2005, and Kondor *et al.*, 2007) have shown that the number of observations should increase proportionally with the number of assets, if we want to obtain a good approximation of the portfolio risk-reward measures. Thus in our research we have to find a compromise between a good statistical approximation of the historical series, and a reduced number of parameters. Usually, the dimensionality of a group of preselected assets can be reduced identifying a few common factors to approximate the asset returns. In particular, the principal component analysis (PCA) can be applied to the Pearson correlation matrix (that can be computed on the Markovian forecast of future returns). Therefore, the returns can be approximated applying a regression analysis on the first few components obtained by means of the PCA. Following the suggestions of Ortobelli and Tichy (2010, 2011), we can "reduce the dimensionality of the problem approximating the return series with a *k*-fund separation model (or other regression-type models) that depends on an adequate number (not too large) of parameters" (Ortobelli and Tichy, 2011). The first step of this analysis is the application of the Principal Component Analysis (PCA) to identify portfolios with the highest return variability (se also Biglova *et al.*, 2009). For this purpose, we work on the gross returns, replacing the 150 preselected time series $\left(z_i; i=1,...,n=150\right)$ with 150 uncorrelated ones $\left(R_i; i=1,...,n=150\right)$. Each z_i can be obtained as a linear combination of R_i. The reduction of dimensionality is obtained selecting, within the R_i group, only *s* time series with a significant dispersion measure; they are called f_i factors (*i* = 1, …, *s*). We know that these factors can be seen as a linear combination of z_i, as follows:

$f_i = \sum_{i=1}^{n} x_i z_i$, where $\sum_{k=1}^{n} x_k^2 = 1$. We can consider the remaining *n*-*s* time series as errors $\left(\varepsilon_j; j = s+1,...,n\right)$. Seen this all, the *n* selected time series can be viewed as a linear combination of factors f_i and a linear combination of factors with small dispersion measure (uncorrelated noise ε_i); the relation is sinthesized in the following formula:

$$z_i = \sum_{k=1}^{s} a_{ik} f_k + \sum_{j=s+1}^{n} a_{ij} R_j = \sum_{k=1}^{s} a_{ik} f_k + \varepsilon_i, \tag{1}$$

In our analysis we apply the PCA both to the Pearson correlation matrix and to the linear correlation matrix $Q = [\rho_{i,j}]$, where $\rho_{i,j} = O_{2,\Im_1}(z_i, z_j)$ for the same sigma algebra $\Im_1$ considered in Ortobelli and Tichy, 2011. Recall that the matrix $O_{p,\Im_1}(X,Y) = \text{cor}((X^{<\frac{p}{2}>} - \text{E}(X^{<\frac{p}{2}>} \mid \Im_1)), (Y^{<\frac{p}{2}>} - \text{E}(Y^{<\frac{p}{2}>} \mid \Im_1)))^{<\min(\frac{p}{2},2)>}$, where $\Im_1$ is a sub-sigma algebra of $\Im$ (i.e. $\Im_1 \subset \Im$) and X and Y are not $\Im_1$ measurable, refers to a logical extension of the Pearson correlation measure, since we obtain the Pearson correlation measure if $p = 2$ and $\Im_1 = \{\varnothing; \Omega\}$. Then we determine the 12 factors with the highest variability of the preselected gross returns for each linear correlation matrix (Pearson and Q). In particular, we assume that the two correlation matrixes are computed on the forecasted wealth $W_{20}(z_i)$ obtained investing in each asset z_i after 20 trading days and assuming a Markovian beahovior of the wealth. Finally we approximate the returns regressing the series on the 24 f_j factors (12 determined with the PCA on the Pearson correlation matrix and the other 12 determined with the PCA on the matrix $Q = [\rho_{i,j}]$), as follows:

$$z_i = b_{i,0} + \sum_{j=1}^{24} b_{i,j} f_j + \varepsilon_i. \tag{2}$$

2.1 The Markovian behavior of wealth

In our research we assume that each portfolio follows a Markov chain. If we define with $z_{(x),k}$ the *k*-th observed gross return of the portfolio $z_{(x)}$, and if we assume that the gross returns are ordered ($z_{(x)}^{(i)} > z_{(x)}^{(i+1)}$ for $i = 1,..., N-1$), the range of the portfolio gross return is: $(\min_k z_{(x),k}; \max_k z_{(x),k})$. This range can be divided into *N* intervals $(a_{(x),i}; a_{(x),i-1})$, such as: $a_{(x),i} = \left(\frac{\min_k z_{(x),k}}{\max_k z_{(x),k}}\right)^{i/N} \cdot \max_k z_{(x),k}$, $i = 0, 1, \ldots, N$. The return of each state, $z_{(x),k}$, can be computed as the geometric average of the extremes of the interval: $z_{(x)}^{(i)} = \sqrt{a_{(x),i} a_{(x),i-1}} = \max_k z_{(x),k} \left(\frac{\max z_{(x),k}}{\min z_{(x),k}}\right)^{\left(\frac{1-2i}{2N}\right)}$, where: $i = 1, 2, \ldots, N$. Consequently: $z_{(x)}^{(i)} = z_{(x)}^{(1)} u^{1-i}$, where: $u = \left(\frac{\max z_{(x),k}}{\min z_{(x),k}}\right)^{1/N} > 1$. Let now assume that the initial wealth W_0 at time 0 is equal to 1, and that, for each possible wealth W_t at time *t*, we have *N* possible different values $W_{t+1} = W_t z_{(x)}^{(i)}$ at time *t*+1 (*i* = 1, 2, …, *N*). Taking into consideration the recombining effect of the Markov chain, after *k* steps we would obtain $1 + k(N-1)$ values of wealth $W_k(z_{(x)})$. If we assume a homogeneous Markov chain, the transition matrix $P = [p_{i,j}]$ does not depend on time and the entries $p_{i,j}$ are estimated using the maximum likelihood estimates $\hat{p}_{i,j} = \frac{\pi_{ij}(K)}{\pi_i(K)}$, where $\pi_{ij}(K)$ is the number of observations (out of *K* observations) that transit from the *i*-th state to the *j*-th state and $\pi_i(K)$ is the number of observations (out of *K* observations) in the *i*-th state. Similarly to what made in Iaquinta and Ortobelli (2006), the distribution function of future gross returns can now be computed. In particular, the $(N-1)k+1$ dimensional vector $p^{(k)}$ represents the unconditional distribution of wealth $W_k(z_{(x)})$ at a given time *k* (*k* = 0, 1, 2, ..., *T*). This distribution can be computed by means of a sequence of matrixes $\left\{Q^{(k)}\right\}_{k=0,1,\ldots,T}$, where $Q^{(k)} = [q_{i,j}^{(k)}]_{\substack{1 \le i \le (N-1)k+1 \\ 1 \le j \le N}}$ and $q_{i,j}^{(k)}$ is the unconditional probability of obtaining a wealth $w_{(x)}^{(i,k)}$ and of being in the

$z_{(x)}^{(j)}$ state at time *k*. The $Q^{(k)}$ matrix is called *unconditional evolution matrix*. In the following analysis we assume to forecast wealth for 20 days assuming 9 states in the Markov chain.

3 An empirical ex-post comparison among international portfolio strategies

In this section we propose an ex-post comparison among portfolio models based on three different portfolio strategies during a period of 41 months (where one month contains 20 trading days) starting from December 2007 to February 2011. In particular, we compare the ex-post performance of three portfolio selection strategies applied to three different clusters of stocks preselected from various financial markets in the world[5] that contain more than 90000 stocks during the analysed period. We distinguish the traded assets for their level of liquidity. We consider the stocks whose average of daily value (expressed in USD) of contracts belongs to a given interval. The value of daily contracts is given by the formula:

$$\text{value of daily contracts} = (\text{closure price}) * (\text{daily volume}),$$

and the average of the value of daily contracts is computed on the last two years (500 trading days). The first group of stocks (with the biggest liquidity) consider the assets with average daily values bigger than 90 million USD. The second group, (defined "Setmid") is composed by stocks with average daily values included between 250,000 and 90 million USD. The third group includes stocks with average daily values between 10,000 and 250,000 USD.

Moreover, since we value the ex-post value of wealth of three models during a period with high volatility of the markets and presence of systemic risk, we introduce an "alarm" that tells us about the presence of systemic risk. The "alarm" is a simple rule that counts the assets whose average returns on the last 20 days is lower than the mean on the last 100 working days. If the number of these assets is higher than 65% of the assets we deduce that probably there is systemic risk on the market and 65% of the assets are losing value. Therefore, in presence of systemic risk we do not invest in the market for any strategy. When we do not observe the presence of systemic risk, we suggest to invest on the 150 preselected stock returns. In particular, when the proposed alarm does not find out the presence of systemic risk, we preselect the 150 stocks using the criteria suggested by Ortobelli *et al* 2011. Then we perform the approximation in formula (2) of the preselected returns. Finally, on the preselected approximated gross returns we maximize a reward-risk performance ratio assuming that no short sales are allowed and we cannot invest more than 10% in a single asset. Doing so, starting by an initial wealth $W_{t_0} = 1$, we evaluate the ex-post wealth sample paths for three different reward-risk portfolio strategies. The three reward risk performance measures we maximize are:

1) The Sharpe ratio (see, Sharpe 1994) $SR(x'z) = \dfrac{E\left(x'z - z_f\right)}{St.dev\left(x'z\right)}$, where the mean and the standard deviation of the portfolio *x'z* are approximated by the empirical mean and standard deviation of the last 125 trading days (we assume that the riskless is null, that is $z_f = 1$).

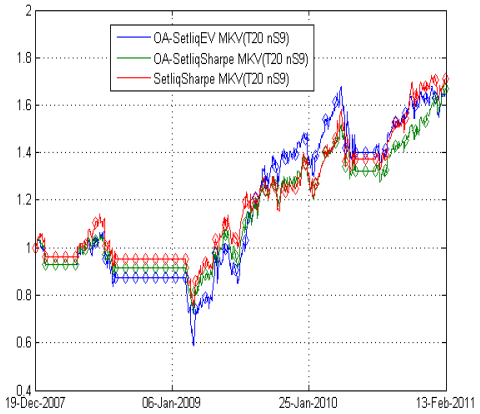
2) The OA-Sharpe ratio: $\dfrac{E\left(W_{20}\left(x'z\right)\right)-1}{St.dev.\left(W_{20}\left(x'z\right)\right)}$, where $W_T\left(x'z\right)$ is the predicted wealth obtained after $T = 20$ working days investing in the *i*-th asset (for any $i = 1, 2, \ldots, n$).

3) The OA-Expected value: $OA-EV\left(W_{20}\left(x'z\right)\right) = \dfrac{\sum_{t=1}^{20} E\left(W_t\left(z_{(x)}\right)/W_t\left(z_{(x)}\right) > 1.2\right)}{\sum_{t=1}^{20} E\left(W_t\left(z_{(x)}\right)/W_t\left(z_{(x)}\right) < 0.97\right)}$ that assumes as: reward measure the expected value of the future wealth conditional that this wealth is greater than a benchmark 1.2; as risk measure the expected value of the future wealth conditional that this wealth is smaller than a benchmark 0.97.

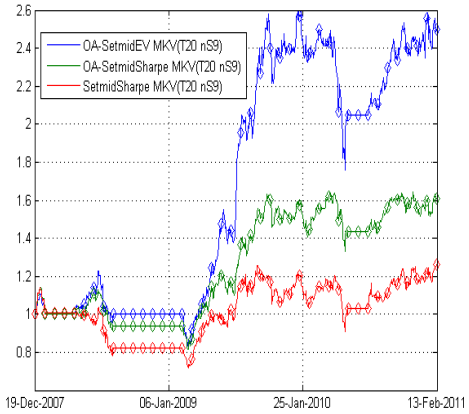
In Figure 1 we report the international diversification and the ex-post comparison of the sample paths of wealth obtained with the different portfolio strategies and for different liquidity constraints that we denote as: "Setliq", "Setmid" and "Setsmall" the strategies respectively with biggest, medium, smallest liquidity.

[5] Considered markets: Australia, New Zealand; Sao Paulo, Mexico, Buenos Aires, Bogotà, Caracas, Lima, Santiago; NYSE, AMEX, NASDAQ, Toronto; Ghana, Johannesburg, Cairo, Casablanca, Bombay, National India, Tel Aviv, Istanbul, Shanghai, Shenzen, Hong Kong, Corea, Taiwan, Bangkok, Indonesia, Kuala Lumpur, Philippine Stock Exchange, Singapore, Thailand, Kuwait City, Tokyo, London, SIX Swiss, Euronext (Amsterdam, Brussels, Lisbon, Paris), Berlin, Madrid, Milan, Helsinki, Athens, Cyprus, Dublin, Iceland, Vienna S. E., Copenhagen S. E., Stockholm, Oslo S. E., Bratislava, Kiev, Riga, Tallinn, Lithuania, Ljubljana, Prague, Belgrade, Zagreb, Bucharest, Budapest S. E., Sofia, Warsaw, Russian Trading System, Moskow.

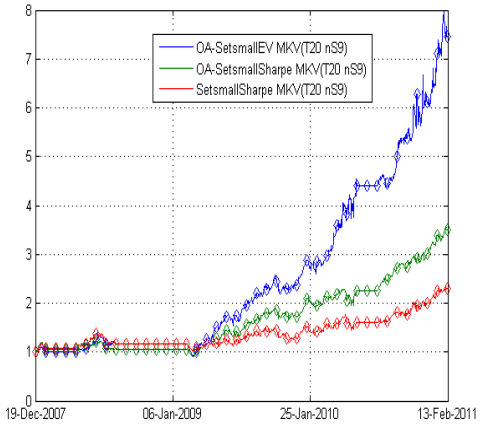
a) Ex post wealth with biggest liquidity



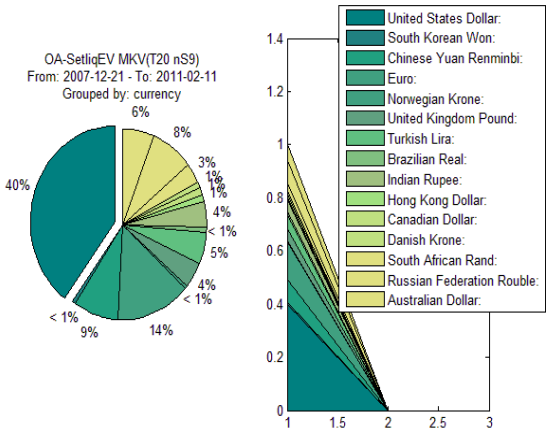
b) Ex post wealth with medium liquidity



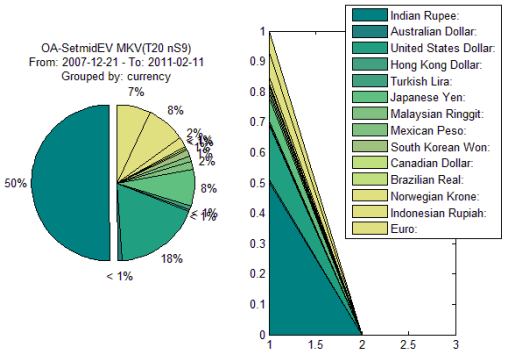
c) Ex post wealth with smallest liquidity



d) International portfolio diversification with OA-SetliqEV



e) International portfolio diversification with OA-SetmidEV



f) International portfolio diversification with OA-SetsmallEV

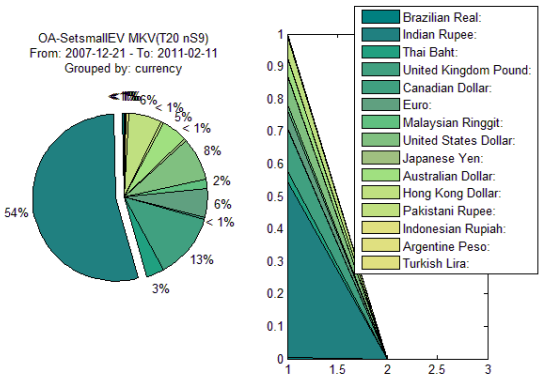


Figure 1 Ex-post wealth and portfolio diversification with different liquidity constraints

4 Conclusion

In the paper, we focused in more details on the problem of portfolio selection assuming internationally diversified portfolio and considering Markovian processes and liquidity constraints.

We have shown a better ex-post performance of the strategy based on the maximization of the OA-Expected value ratio. It clearly appears (in Figure 1 a) b) and c)) that both strategies based on the Markovian evolution of the returns present higher final wealth than the classic one based on the maximization of the Sharpe ratio.

Moreover, the proposed empirical analysis demonstrates that the alarm inserted to detect the presence of systemic risk works well, since it is able to stop, to identify and, to forecast the largest period of systemic risk of the recent crises (sub-prime crisis 2008-2009 and the country credit risk crisis 2010-2011).

Finally, we also discover that the turnover behavior does not change too much among the three strategies. In addition, the three strategies diversify among different assets but not among different countries for each class of liquidity. However, while the most liquid strategies invest most of the wealth in the most industrialized countries (US, Europe); the less liquid strategies invest most of the wealth in the BRIC countries (in particular India, China and Brazil).

Acknowledgements

The research was supported through the Czech Science Foundation (GAČR) under project 13-13142S and SP2013/3, an SGS research project of VSB-TU Ostrava, and furthermore by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070), including the access to the supercomputing capacity, and the European Social Fund in the framework of CZ.1.07/2.3.00/20.0296 (second and third author) and CZ.1.07/2.3.00/30.0016 (fourth author) and the state budget of the Czech Republic. The paper has been also supported by the Italian funds ex MURST 60% 2011 and MIUR PRIN MISURA Project, 2013–2015.

References

[1] Ait-Sahalia, Y.: Nonparametric Pricing of Interest Rate Derivative Securities. *Econometrica* **64** (1996), 527–560.

[2] Angelelli, E., and Ortobelli, S.: American and European Portfolio Selection Strategies: the Markovian Approach. In: *Financial Hedging* (Catlere P.N., eds.). Nova Science Publishers, 2009(a),119–152.

[3] Angelelli, E., and Ortobelli, S.: Maximum Expected Utility of Markovian Predicted Wealth. In: *Computer Science 5545* (Lecture Notes, G. Allen *et al.* eds., ICCS 2009, Part II, LNCS). 2009(b), 588–597.

[4] Biglova, A., Ortobelli, S., Rachev; S., and Fabozzi, F.: Modeling, Estimation, and Optimization of Equity Portfolios with Heavy-tailed Distributions. In: *Optimizing Optimization: the Next Generation of Optimization Applications and Theory* (Satchell S., eds.). 2009, 117–141.

[5] Cont, R., and Tankov, P.: *Financial Modeling with Jump Processes*. Chapman & Hall/CRC Press, Boca Raton, 2003.

[6] Cox, J.C., Stephen, A.R., and Rubinstein, M.: Option Pricing: A Simplified Approach. *Journal of Financial Economics* **7** (1979), 229–263.

[7] D'Amico, G.: Statistical Inference for Markov Chain European Option: Estimating the Price, the Bare Risk and the Theta by Historical Distributions of Markov Chain. *Journal of Statistics & Management Systems* **9/3** (2006), 737–754.

[8] Iaquinta, G., and Ortobelli S.: Distributional Approximation of Asset Returns with non Parametric Markovian Trees. *International Journal of Computer Science & Network Security* **6/11** (2006), 69–74.

[9] Kondor, I., Pafka, S., and Nagy, G. Noise sensitivity of portfolio selection under various risk measures, *Journal of Banking and Finance* **31** (2007) 1545–1573.

[10] Ortobelli S. and Tichý, T. On the impact of association measures in the portfolio theory. *In Proceeding of 47th EWGFM meeting* (Tichý, T., Kopa, M. eds), 2010, pp.119–130

[11] Ortobelli, S, and Tichý, T.: The impact of association measures within the portfolio dimensionality reduction problem. *Ekonomická revue – Central European Review of Economic Issues* **14** (2011), 257–264.

[12] Ortobelli, S., Angelelli, E., and Toninelli, D.: Set-Portfolio Selection with the Use of Market Stochastic Bounds. *Emerging Markets Finance & Trade* **47/5** (2011), 5–24.

[13] Papp, G., Pafka, S., Nowak, M.A., Kondor, I. Random Matrix Filtering in Portfolio Optimization, *ACTA Physica Polonica B* **36** (2005), 2757–2765.

[14] Sharpe, W.F. The Sharpe ratio, *Journal of Portfolio Management*, (1994), 45–58.

[15] Stutzer, M.J.: A Simple Nonparametric Approach to Derivative Security Valuation. *Journal of Finance* **51** (1996), 1633–1652.

Inverse fuzzy eigenproblem in databases

Martin Bacovský[1], Martin Gavalec[2], Hana Tomášková[3]

Abstract. Fuzzy algebra is defined as the real unit interval equipped by two binary operations: maximum and minimum, which are used analogously as addition and multiplication in the classical linear algebra. Fuzzy algebra is used in applications such as fuzzy control systems, or stability of discrete event systems. Eigenvectors of a fuzzy matrix correspond to steady states of a given discrete-events system.

The inverse eigenproblem is motivated by the task of finding all possible systems possessing a given steady state. In the contribution, the inverse problem is studied as a system of fuzzy equations, possibly with upper end lower bounds. Application of the inverse approach to efficient search in databases is presented.

Keywords: fuzzy algebra, max-min algebra, eigenproblem, database search

JEL classification: C44
AMS classification: 90C15

1 Introduction

By max-min algebra (sometimes called: fuzzy algebra) we understand a linear structure on a linearly ordered set with two binary operations maximum and minimum, used similarly as addition and multiplication in the classical linear algebra. Fuzzy algebra is used in many applications such as fuzzy control systems or stability of discrete event systems. Eigenvectors of a fuzzy matrix correspond to steady states of a given discrete-events system. Investigation of the max-min eigenvectors of a given matrix is therefore of great practical importance. The eigenproblem in max-min algebra has been studied by many authors. Interesting results were found in describing the structure of the eigenspace, and algorithms for computing the maximal eigenvector of a given matrix were suggested, see e.g. [3], [4], [5], [6], [7].

While the standard eigenproblem looks for the steady states of a system characterized by given transition matrix, the inverse eigenproblem is motivated by the task of finding all possible systems with given steady state vector. In this paper the inverse problem is studied for fuzzy linear systems. The problem is a special case of solving a system of max-min fuzzy equations, however, the direct way to inverse solution allows description of solutions with more details and with smaller computational complexity.

Application of the inverse fuzzy eigenproblem to databases helps to find suitable intervals for database search systems fulfilling some given conditions. This will narrow down the scanned field, and the problem may be solved for greater data sets. The model represents the database searching by the preferences of the user.

2 Notions and notation

Many applications can naturally be solved by using operations maximum and minimum on a linearly ordered set. The so-called max-min algebras were frequently studied by number of authors and various algorithms have been developed for solving problems analogous to those in classical linear algebra.

[1]University of Hradec Králové, Faculty of Informatics and Management, Rokitanského 62, 50003 Hradec Králové, Czech Republic, e-mail: martin.bacovsky@uhk.cz
[2]University of Hradec Králové, Faculty of Informatics and Management, Rokitanského 62, 50003 Hradec Králové, Czech Republic, e-mail: martin.gavalec@uhk.cz
[3]University of Hradec Králové, Faculty of Informatics and Management, Rokitanského 62, 50003 Hradec Králové, Czech Republic, e-mail: hana.tomaskova@uhk.cz

Max-min algebra is defined as a triplet $(\mathcal{B}, \oplus, \otimes)$, where $\mathcal{B}$ is a linearly ordered set, and $\oplus = \max$, $\otimes = \min$ are binary operations on $\mathcal{B}$. The notation $\mathcal{B}(n,n)$ ($\mathcal{B}(n)$) denotes the set of all square matrices (all vectors) of given dimension n over $\mathcal{B}$. Operations $\oplus$, $\otimes$ are extended to matrices and vectors in a formal way.

The eigenproblem for a given matrix $A \in \mathcal{B}(n,n)$ in max-min algebra consists of finding a vector $x \in \mathcal{B}(n)$ (eigenvector) such that the equation

$$A \otimes x = x \tag{1}$$

holds true. By the eigenspace of a given matrix we mean the set of all its eigenvectors.

In the inverse eigenproblem the eigenvector x is considered as fixed, and the matrix A is unknown. It is clear that the equation (1) with fixed x and variable A is a special case of the equation (2) below, with fixed coefficients x, b. In the next two sections, the equation (2) with upper and lower bounds for the solution matrix A will be studied. The maximal (greatest) solution and the set of all minimal solutions will be described. The applications of the results are briefly presented in the last section.

3 The greatest solution

Definition 1. Let $x, b \in \mathcal{B}(n)$ and $\mathbf{A} = [\underline{A}, \overline{A}] \subseteq \mathcal{B}(n,n)$ be given. Then by *inverse interval problem* we understand linear equation

$$A \otimes x = b \tag{2}$$

w.r.t. the unknown matrix $A \in \mathbf{A}$.

Following matrix will play a crucial role in solving problem (2). For given $\mathbf{A} = [\underline{A}, \overline{A}] \subseteq \mathcal{B}(n,n)$, x and $b \in \mathcal{B}(n)$, the matrix $\hat{A}(x,b)$ is defined by its entries:

$$\hat{a}_{ij}(x,b) = \begin{cases} b_i & \text{when } \overline{a}_{ij} \otimes x_j > b_i, \\ \overline{a}_{ij} & \text{otherwise.} \end{cases}$$

Proposition 1. $\hat{A}(x,b) \leq \overline{A}$.

Proof. By contradiction: let exist $i, j \in N$ such that $\hat{a}_{ij}(x,b) > \overline{a}_{ij}$, then $\hat{a}_{ij}(x,b) = b_i < \overline{a}_{ij}$, contradiction. □

Proposition 2. $\hat{A}(x,b) \otimes x \leq b$.

Proof. By contradiction: let exist $i, j \in N$ such that $\hat{a}_{ij}(x,b) \otimes x_j > b_i$, but then $\hat{a}_{ij}(x,b) = b_i$, contradiction. □

Proposition 3. *If* $A \otimes x = b$, $A \leq \overline{A}$, *then* $A \leq \hat{A}(x,b)$.

Proof. By contradiction: let exist $i, j \in N$ such that $a_{ij} > \hat{a}_{ij}(x,b)$, then $\hat{a}_{ij}(x,b) = b_i$ and $x_j > b_i$, which yields $a_{ij} \otimes x_j > b_i$, contradiction. □

Proposition 4. *If* $b \leq \overline{A} \otimes x$, *then* $\hat{A}(x,b) \otimes x = b$.

Proof. By contradiction: let $i_* \in N$ be such that $\bigoplus_{j=1}^{n} \hat{a}_{i_* j}(x,b) \otimes x_j < b_{i_*}$, then it follows that $\hat{a}_{i_* j}(x,b) = \overline{a}_{i_* j}$ for all $j \in N$, thence $\bigoplus_{j=1}^{n} \overline{a}_{i_* j} \otimes x_j < b_{i_*}$, which is in contradiction with $b \leq \overline{A} \otimes x$. □

Proposition 5. *If* $\underline{A} \otimes x \leq b$, *then* $\underline{A} \leq \hat{A}(x,b)$.

Proof. By contradiction: let element $\hat{a}_{i_*j_*}(x,b)$ be such that $\hat{a}_{i_*j_*}(x,b) < \underline{a}_{i_*j_*}$, then $\hat{a}_{i_*j_*}(x,b) = b_{i_*}$ and $x_{j_*} > b_{i_*}$, from $\underline{a}_{i_*j_*} \otimes x_{j_*} > b_{i_*}$ it follows that $\bigoplus_{j=1}^{n} \underline{a}_{i_*j} \otimes x_j > b_{i_*}$, which is in contradiction with $\underline{A} \otimes x \leq b$. □

Remark 1. Preceding propositions 3 till 5 show that if there exists some solution of (2) in $\mathbf{A}$, then $\hat{A}(x,b)$ is also a solution and moreover it is the greatest one in $\mathbf{A}$. This finding justifies next definition.

Definition 2. For a given problem (2), a matrix $\hat{A}(x,b)$ introduced before will be called *the greatest solution.*

Remark 2. For $\mathbf{A} = [\underline{A}, \overline{A}] \subseteq \mathcal{B}(n,n)$ and $x \in \mathcal{B}(n)$ we will denote $\mathbf{A} \otimes x := [\underline{A} \otimes x, \overline{A} \otimes x]$.

Theorem 6. *There is at least one solution of (2) in* $\mathbf{A} = [\underline{A}, \overline{A}]$ *iff* $b \in \mathbf{A} \otimes x$.

Proof. Implication $\Rightarrow$ is clear, converse implication follows from preceding propositions. □

Theorem 7. *Let* $x, y \in \mathcal{B}(n)$ *and* $\mathbf{A} = [\underline{A}, \overline{A}] \subseteq \mathcal{B}(n,n)$ *be given. Then the following statements are equivalent:*

(i) There exists A *in* $\mathbf{A}$ *such that* $A \otimes x = A \otimes y$,

(ii) $\mathbf{A} \otimes x \cap \mathbf{A} \otimes y \neq \emptyset$,

(iii) $(\underline{A} \otimes x) \oplus (\underline{A} \otimes y) \leq (\overline{A} \otimes x) \oplus' (\overline{A} \otimes y)$.

Proof. $(i) \Rightarrow (ii)$: for such A clearly $A \otimes x \in \mathbf{A} \otimes x \cap \mathbf{A} \otimes y$, hence the intersection is nonempty.
$(ii) \Rightarrow (i)$: let us choose arbitrary but fixed $b \in \mathbf{A} \otimes x \cap \mathbf{A} \otimes y$ and construct matrix A with entries:

$$a_{ij} = \begin{cases} b_i & \text{when } \overline{a}_{ij} \otimes \max(x_j, y_j) > b_i, \\ \overline{a}_{ij} & \text{otherwise.} \end{cases}$$

We will show that A is a solution. Since $b \in \mathbf{A} \otimes x$, there exists for every $i \in N$ such index $j_* \in N$ that $\hat{a}_{ij_*}(x,b) \otimes x_{j_*} = b_i$ holds. Therefore either $x_{j*} = b_i \ \wedge \ \hat{a}_{ij_*}(x,b) \geq b_i$ or $x_{j_*} > b_i \ \wedge \ \hat{a}_{ij_*}(x,b) = b_i$. In both cases $a_{ij_*} \otimes x_{j_*} = b_i$. From the latter and the fact that $A \leq \hat{A}(x,b)$ it follows that

$$\bigoplus_{j \in N} a_{ij} \otimes x_j = b_i,$$

which holds for every $i \in N$, i.e. $A \otimes x = b$. By the same considerations we derive $A \otimes y = b$.
$(ii) \Leftrightarrow (iii)$: clearly $\mathbf{A} \otimes x \cap \mathbf{A} \otimes y = [(\underline{A} \otimes x) \oplus (\underline{A} \otimes y), (\overline{A} \otimes x) \oplus' (\overline{A} \otimes y)]$. □

4 Minimal solutions

Let us introduce some notation: Let $\mathbf{A} = [\underline{A}, \overline{A}]$, where $\underline{A}$ and $\overline{A} \in \mathcal{B}(n,n)$, and $x, b \in \mathcal{B}(n)$ for some $n \in \mathbb{N}$. Then for every $i \in N$ we define:

- $e(i) = \{j \in N |\ \underline{a}_{ij} \otimes x_j = b_i\}$,
- $k(i) = \begin{cases} \emptyset & \text{when } e(i) \neq \emptyset, \\ \{j \in N | x_j \geq b_i \wedge \underline{a}_{ij} < b_i \leq \overline{a}_{ij}\} & \text{otherwise,} \end{cases}$
- $l_i = \begin{cases} 1 & \text{when } e(i) \neq \emptyset, \\ |k(i)| & \text{otherwise.} \end{cases}$
- $l = \prod_{j \in N} l_j$, $L = \{1, 2, \ldots, l\}$.

In addition, let ϕ be n-tuple with elements satisfying

$$\phi(i) = \begin{cases} 0 & \text{for } k(i) = \emptyset, \\ j \in k(i) & \text{for } k(i) \neq \emptyset. \end{cases}$$

Then we define matrix $\check{A}(x, b, \phi)$ with entries $\check{a}_{ij}(x, b, \phi) = \begin{cases} b_i & \text{when } j = \phi(i), k(i) \neq \emptyset, \\ \underline{a}_{ij} & \text{otherwise.} \end{cases}$

Theorem 8. *Let* $\mathbf{A}$, x, b *a* l *be from the latter notation and moreover* $\underline{A} \otimes x \leq b$. *Then the following holds: equation (2) has in the interval* $\mathbf{A}$ *exactly* l *mutually incomparable minimal solutions. These solutions are in the form* $\check{A}(x, b, \phi)$ *for any* ϕ *from the latter notation.*

Proof. First, we will prove that $\check{A}(x, b, \phi) \otimes x = b$ for some ϕ. For ith entry of this product two cases may arise: Either $e(i) = \emptyset$ and $k(i) \neq \emptyset$, then for index $j_0 = \phi(i)$ must hold

$$\check{a}_{ij_0}(x, b, \phi) \otimes x_{j_0} = b_i \otimes x_{j_0} = b_i$$

thence

$$\bigoplus_{j \in N} \check{a}_{ij}(x, b, \phi) \otimes x_j \geq b_i.$$

If the last inequality was strictly greater, there would exist index j_1 such that $x_{j_1} > b_i$ and also $\check{a}_{ij_1}(x, b, \phi) = \underline{a}_{ij_1} > b_i$, which is in contradiction with the assumption $\underline{A} \otimes x \leq b$.

When $k(i) = \emptyset$, $e(i)$ is either nonempty, from which it follows that the product will give b_i from the similar considerations as in the last case, or it is empty too, i.e. $l = 0$. In this case we have

$$x_j > b_i \Rightarrow b_i < \underline{a}_{ij} \vee b_i > \overline{a}_{ij}, \quad x_j = b_i \Rightarrow b_i > \overline{a}_{ij} \geq \underline{a}_{ij}.$$

Since $\underline{A} \otimes x \leq b$ could not hold for $x_j > b_i \wedge b_i < \underline{a}_{ij}$, b_i has to be $b_i > \overline{a}_{ij}$ for $x_j > b_i$. In sum, from $x_j \geq b_i$ it follows that $b_i > \overline{a}_{ij}$. But then

$$(\overline{A} \otimes x)_i = \bigoplus_{j \in N} \overline{a}_{ij} \otimes x_j < b_i$$

and there exists no solution in the interval $\mathbf{A}$. This result corresponds with $l = 0$. That proves the equality $\check{A}(x, b, \phi) \otimes x = b$.

Now we prove $\check{A}(x, b, \phi) \in \mathbf{A}$:

$$\underline{A} \leq \check{A}(x, b, \phi) \Leftrightarrow \underline{a}_{ij} \leq \check{a}_{ij}(x, b, \phi) = \begin{cases} \underline{a}_{ij} & \text{inequality holds,} \\ b_i & \text{from } k(i) \text{ definition also holds,} \end{cases}$$

$$\overline{A} \geq \check{A}(x, b, \phi) \Leftrightarrow \overline{a}_{ij} \geq \check{a}_{ij}(x, b, \phi) = \begin{cases} \underline{a}_{ij} & \text{inequality holds,} \\ b_i & \text{from } k(i) \text{ definition also holds.} \end{cases}$$

Therefore we have l matrices $\check{A}(x, b, \phi)$ from the interval $\mathbf{A}$ that satisfy $\check{A}(x, b, \phi) \otimes x = b$. We will now show their incomparability: let us have $\underline{A}(x, b, \phi_1)$ and $\underline{A}(x, b, \phi_2)$, where $\phi_1 \neq \phi_2$. Then must exist such $i \in N$ that $\phi_1(i) \neq \phi_2(i)$. By mutual matrices's entries comparison

$$\begin{array}{rcccccl} \check{a}_{i,\phi_1(i)}(x, b, \phi_1) & = & b_i & > & \underline{a}_{i,\phi_1(i)} & = & \check{a}_{i,\phi_1(i)}(x, b, \phi_2), \\ \check{a}_{i,\phi_2(i)}(x, b, \phi_1) & = & \underline{a}_{i,\phi_2(i)} & < & b_i & = & \check{a}_{i,\phi_2(i)}(x, b, \phi_2), \end{array}$$

which yields their incomparability.

Finally, we will prove their minimality: Let $A \in \mathbf{A}$, $A \otimes x = b$ and A be at least at one element a_{ij} lesser than $\check{A}(x, b, \phi)$ for some ϕ. Then exists $(i_0, j_0) \in N \times N$ such that $a_{i_0,j_0} < \check{a}_{i_0,j_0}(x, b, \phi) = b_{i_0}$ (second case $\underline{a}_{i_0,j_0}$ can not arise, because $A \in \mathbf{A}$), thus $x_{j_0} \geq b_{i_0}$ together with $a_{i_0,j_0} \otimes x_{j_0} < b_{i_0}$ and $j_0 = \phi(i_0)$. From $\check{a}_{i_0,j_0}(x, b, \phi) \neq \underline{a}_{i_0,j_0}$ follows that $e(i_0) = \emptyset$ and $k(i_0) \neq \emptyset$. Then there exists $u \in k(i_0)$, i.e. $x_u \geq b_{i_0}$ and $b_{i_0} > \underline{a}_{i_0,u}$ such that $a_{i_0,u} = b_{i_0} > \underline{a}_{i_0,u} = \check{a}_{i_0,u}(x, b, \phi)$. Other i_0th row entries can not be less than $\underline{a}_{i_0,j}$, otherwise A would not be from $\mathbf{A}$, and even not greater, since then A could not be minimal.

Let us define $\phi'(i_0) = u$. This is possible, because from preceding considerations u is from $k(i_0)$. Since the same deduction can be applied to all rows of matrix A that contain elements less than corresponding elements of $\check{A}(x, b, \phi)$ and that otherwise we can define ϕ' in the same way as ϕ, for such ϕ' it holds that $A = \check{A}(x, b, \phi')$ and moreover $\phi' = \phi_\lambda$ for some $\lambda \in L$.

Any minimal matrix A is thus the same as some $\check{A}(x, b, \phi_\lambda)$ for appropriate $\lambda \in L$. These matrices are therefore minimal and the only minimal in the interval $\mathbf{A}$. □

Definition 3. For a given problem (2), a matrix $\check{A}(x, b, \phi)$ for some ϕ as introduced before will be called *a minimal solution.*

Remark 3. Number of minimal solutions is at most n^n.

Example 1. Let us define for all $i, j \in N$: $(\underline{A})_{ij} = 1$, $(\overline{A})_{ij} = 3$, $x_i = b_i = 2$. Then for each $i \in N$ is $e(i) = \emptyset, k(i) = N$ and $\phi : N \to N$. There exist n^n such mappings ϕ.

Remark 4. A solution set of (2) can be described as the union of intervals with a lower bound equal to $\check{A}(x, b, \phi)$ and an upper bound to $\hat{A}(x, b)$. These intervals are not disjoint, e.g. $\hat{A}(x, b)$ lies always in their intersection. Special cases of exactly one solution (i.e. $\check{A}(x, b, \phi) = \hat{A}(x, b)$) or of no solution are also possible.

By *inverse eigenproblem* we understand problem (2) with the right-hand side b equal to x. As it is easily seen, this constraint does not affect introduced algorithms for computing the greatest and minimal elements, therefore they can be used for inverse eigenproblems without modifications.

5 Applications in databases

Motivation of the inverse eigenproblem is to find the set of all possible systems corresponding to given conditions such as user preferences. The user preferences are written into vector, whereas the set of possible systems is represented by an interval matrix. In this case we have given the vector of preferences and we are looking for the matrix, which fulfil the conditions given by vector. The first step is to find the maximal matrix which accomplishes the conditions. The maximal matrix is uniquely determined. Next we decide the form of the lower bound - minimal matrix. Unlike the maximal matrix, the minimal matrix is not unique and the set of all minimal matrices (solutions of the inverse eigenproblem) may be very large. In spite of this fact, the solution set of the inverse eigenproblem consists of all matrices between the maximal and one of the minimal solutions.

Applications of the inverse eigenvector problem below is derived from the next two articles. In paper [1] a solution of recommendation system based on the user profile and heterogeneous components connected via web services was presented. The prototype developed under the code name Personal Recommendation (PRECO) is located on the following website http://preco.uhk.cz . The architecture of the system is guided by the need for centralized store of user preferences on one hand and on the other hand by the necessity to track the users' behaviour.

In response to this article the max-min inverse approach can be used as follows. Each database user is able to set the degree of importance of any discipline that is listed in the database. This importance depends on the focus orientation of the user. Inputs of interval matrix is given as the preference matrix of relevant disciplines. For any value of input a_{ij} the next notation is used: a_{ij} = rate of priorization of discipline i in comparison with discipline j. By using the inverse approach we are able to find any set of systems according to the requirements specified by the user. This will help us to narrow down the scanned field in database.

The next paper [2], talks about the idea to adapt principles and technologies from the BI concept to public health management. In this article the authors explore the decision support systems of support for Public Health in the Management of Biological Incidents using simulations, judgemental, and other methods.

In this application the input vector x can be represented by degree of intensity of manifested symptoms (conditions). A survey conducted by the authors, was arranged the necessary data, from which we can build two types of approaches. The first type is the largest accumulation, i.e. it assembles the largest possible array of possible measures of overlapping symptoms, conditions or circumstances. Second type is the smallest overlapping, i.e. it prepares a matrix of the lowest possible rates possible concurrence

of symptoms, conditions or circumstances. By using these types of the inverse approach a narrower interval overlapping symptoms can be found. The motivation is an acceleration of the search for potential biological incidents.

Acknowledgement

The support of INKOV project # CZ.1.07/2.3.00/20.0001 financed from EU and Czech Republic funds, and Czech Science Foundation project # 402/09/0405 is kindly acknowledged.

(Received July 10, 2013.)

References

[1] V. Bureš, P. Čech, 2007. Recommendation of web resources for academics - architecture and components. 3rd International Conference on Web Information Systems and Technologies, INSICC, 437–440, ISBN : 978-972-8865-78-8.

[2] V. Bureš, P. Čech, T. Otčenášková, 2012. Proposal of simulation-based management of biological or chemical incidents as a smart solution. International Review on Computers and Software 7(5), 2173–2178.

[3] K. Cechlárová: Eigenvectors in bottleneck algebra. Lin. Algebra Appl. **175** (1992), 63–73.

[4] R.A. Cuninghame-Green: Minimax Algebra and Application. In: Advances in Imaging and Electron Physics 90, (P. W. Hawkes, Ed.). Academic Press, New York, 1995.

[5] M. Gavalec: Solvability and unique solvability of max-min fuzzy equations. Fuzzy Sets and Systems **124** (2001), 385–393.

[6] J. Plavka: l-parametric eigenproblem in max algebra. Discrete Applied Mathematics **150** (2005), 16–28.

[7] K. Zimmermann: Extremal Algebra (in Czech). Ekon. ústav ČSAV Praha, 1976.

A note on the methods of constructing weekly stock market returns

Eduard Baumöhl[1], Štefan Lyócsa[2], Tomáš Výrost[3]

Abstract. To avoid the non-synchronicity of daily stock market data, the weekly returns are commonly used by the empirical researchers. We focused three methods on construction of weekly returns: Wednesday-to-Wednesday, Friday-to-Friday, and returns calculated from weekly price averages. Using data from the CEE-3 (the Czech PX, the Hungarian BUX, and the Polish WIG) and three developed stock markets (the MSCI Germany, the Dow Jones Euro Stoxx 50, and the MSCI World) we show, that the statistical properties of weekly returns depend on the method used to construct the series. Simulations are also performed to observe the differences between variance and autocorrelations of the weekly returns.

Keywords: stock markets, weekly returns, statistical properties.

JEL Classification: C10, C80, G10
AMS Classification: 62-07, 62P20, 62M10, 91B84

1 Introduction

Because of non-synchronicity, researchers using and analyzing data from different stock markets often try to avoid the usage of daily closing prices. We distinguish between two types of non-synchronicity in daily stock market data (Baumöhl and Lyócsa [1]):

- The "non-synchronous trading effect I" – different number of observation between daily samples arises from the fact, that the stock markets are subject to different national, religious, and other holidays, and other unexpected events;
- The "non-synchronous trading effect II" – the fact that the national stock exchanges operate in different time zones induces cross-autocorrelations in closing prices.[4] Similarly, even within the stock markets with same trading hours, non-synchronicity occurs as the last trades of individual stocks may occur at different times.

According to Lo and MacKinlay [9], ignoring the non-synchronous trading effects can result in biased inference regarding the short-run behavior of prices. They proposed a data generating process (DGP) which models returns and allows for various forms of stochastic non-trading. Among others, using the proposed DGP they showed how non-trading can effect autocorrelation of the observed time-series. To cope with non-synchronous trading effects, Hamao et al. [7] disaggregated returns into open-to-close and close-to-open returns, while Forbes and Rigobon [6] used overlapping multiday returns.[5] Burns et al. [2] proposed to forecast the missing observations from values of known (presumably correlated) returns, therefore achieving synchronized returns.[6] The idea is that such a prediction, although not necessarily correct, will be an unbiased estimator of the true returns. They also show, that by ignoring non-synchronous trading effects, the correlations are understated, which has important implications in portfolio management (see also Martens and Poon [10] or Scherer [11]).

Probably the most popular approach to deal with non-synchronous trading effects involves using lower frequency data, i.e. weekly or monthly returns. The idea is, that time differences between closing prices are much smaller relative to 168 hours than to 24 hours (Martens and Poon [10]). Many authors who prefer using weekly returns claim that they mitigate or control the effects of non-synchronous trading, infrequent trading, bid-ask spreads, non-overlapping trading hours, spurious spill-over effects, or other microstructure frictions which can bias subsequent estimation.

[1] University of Economics, Faculty of Business Economics, Department of Economics, Tajovského 13, Košice, Slovakia, eduard.baumohl@euke.sk
[2] University of Economics, Faculty of Business Economics, Department of Business Informatics and Mathematics, Tajovského 13, Košice, Slovakia, stefan.lyocsa@euke.sk
[3] University of Economics, Faculty of Business Economics, Department of Finance and Accounting, Tajovského 13, Košice, Slovakia, tomas.vyrost@euke.sk
[4] The information set included in the closing prices of equities in Tokyo is obviously distinct from the information set at the end of the same trading day in New York.
[5] Thus introducing autocorrelation in the resulting series.
[6] The forecasts are based on a vector first order moving average model of returns with a GARCH covariance matrix.

There are several popular approaches used to construct weekly returns. While it is clear that the resulting series will differ with regard to their statistical properties, it is the goal of this paper to show, that these differences might be quite substantial. Unfortunately, in empirical studies, the method for data construction is rarely[7] specified.[8]

2 Data and methodology

Our dataset for empirical part of the paper covers six stock market indices, namely Polish WIG 20 (WIG), Czech PX, Hungarian BUX, MSCI Germany (GER), Dow Jones Euro Stoxx50 (STX) and MSCI World (MSCI) and a period from January 4, 1999 to September 28, 2012. Daily closing prices were obtained from the Thomson Reuters Datastream. Weekly returns were estimated using three methods, which can be grouped into conceptually two different approaches.

The *first approach* is a systematic sampling from the initial series of daily closing prices. Let P_n be the closing price of a stock market index in given day $n = 1, 2, ..., N$. Then assuming a week of five trading days, new series is created by sampling every fifth observation from the initial series of daily closing prices. If days from this subsample correspond to Fridays, the resulting series is denoted as S^F_t. Similarly, if these days correspond to Wednesday, the resulting series is denoted as S^W_t. The weekly (continuous) returns are then calculated as $ln(S_t/S_{t-1})$, where $t = 1, 2, \ldots, T$ corresponds to a given week and are denoted as *F-to-F* (Friday-to-Friday) and *W-to-W* (Wednesday-to-Wednesday). In case of *W-to-W* returns, if Wednesday prices were not available, they were substituted by prices in the following order: Thursday, Tuesday, Friday, and Monday. In case of *F-to-F* returns, if Friday prices were not available, they were substituted by prices in the following order: Thursday, Wednesday, Tuesday, and Monday. This approach corresponds to an intuitively appealing definition of a weekly return, where weekly stock market returns should correspond to a nominal continuous return acquired over five trading days.

The *second approach* takes averages over P_n of a given week t, i.e., $S^A_t = n(t)^{-1}\sum_{i=1}^{n(t)} P_{t,i}$, where $n(t)$ is the number of trading days (indexed as $i = 1, 2, \ldots n(t)$) in a given week t. The weekly returns are then calculated as before and denoted as *AV*. Using *AV* returns, one could argue, that average price is a representative value of prices in a given week. Therefore weekly returns from such series may also be regarded as representative returns. Note, that if one assumes, that stock prices can be regarded as a continuous-time stochastic process, then the resulting return is a return realized between prices from both weeks, but the time difference between them is stochastic, in contrast to *F-to-F* and *W-to-W* methods, where this difference is fixed and set to five trading days.

3 Results

3.1 Simulations

Knowing the DGP of the daily stock prices, one could derive the variance and autocorrelation of the resulting weekly returns. For example, assuming that daily closing prices follow $P_n = P_{n-1} + \delta_n$, with $\delta_n = \rho\delta_{n-1} + v_n$, where $v_n \sim N(0,\sigma^2)$ are *iid* and $|\rho| < 1$, we could derive the variance, autocovariance and autocorrelation of weekly re-

[7] For this version of the paper, we have performed a preliminary screening of 147 papers published via a selected publisher. In 79 cases (53.7%) the method used to construct weekly returns was not explicitly specified, although in 21 from these cases researchers used data from Datastream, where weekly returns correspond to the end of the week closing prices. Apart from that, end of the week prices were explicitly specified in 15 (10.2%) cases. Apart from few studies (e.g. Arab stock markets), end of the week prices refer to Friday closing prices. Corresponding returns are denoted as *F-to-F* (Friday-to-Friday). In 39 cases (26.5%) researchers used returns calculated from *W-to-W* (Wednesday-to-Wednesday) closing prices. Other days of the week were used in as much as 14 cases (9.5%). Note, that we are not concern by studies, where one is interested in in- or out-of sample forecasting of specific prices. Our sample includes studies, where weekly prices are used mostly due to non-synchronous trading effects. The use of *F-to-F* closing prices was rarely justified. The use of *W-to-W* prices was justified for various reasons (which depend on the markets): fewer holidays on Wednesdays compared to Fridays, higher trading volume on Wednesdays than in other trading days, elimination of the beginning-of-the-week and end-of-the-week effects. If Friday or Wednesday closing prices were unavailable, the last known closing prices are used, or a specific order of prices is specified (based on researchers choice and data availability, e.g. "*if Wednesday closing prices are unavailable we used Thursday, Tuesday, Friday or Monday closing prices*").

[8] If possible, the recommended approach is to use "synchronous closing prices", also called a "common window" (e.g. Schotman and Zalewska [12]; Égert and Kočenda [5]). This however requires high frequency data, which are not always available, particularly for analyzed periods longer than a decade. Even if daily data are available, in many instances, they are not required as for the problem under investigation, they can be too noisy.

turns. However, the assumptions put on v_n are rather simplistic. We have therefore decided to perform a simple simulation study, where the goal was to observe the differences between variance and autocorrelations of the simulated weekly returns. As defined in previous section, two approaches were compared.

The simulation was conducted as follows:

1. Using daily (continuously compounded) returns, we estimated ARMA(p, q)-GARCH(r, s) models. In this step, we have considered several specifications of the variance equation. For each of the following specifications; GARCH, CSGARCH, TGARCH, AVGARCH, NGARCH, FGARCH, APARCH, GJR-GARCH, NAGARCH we choose p, q (up to 5) and r, s (up to 2) such that no autocorrelation and conditional heteroskedasticity was indicated (using the Ljung-Box test) in the resulting standardized residuals. From these specifications, we chose the one, with the lowest BIC.[9]
2. Estimated coefficients were used to simulate returns.
3. Simulated returns were used to reconstruct the series of daily closing prices.
4. The simulated daily closing prices were used to draw weekly prices and calculate weekly returns using the two approaches as described in Section 2.
5. The variance and autocorrelation at lag 1 were computed.
6. Steps 2 – 5 were repeated 10000 times. The results may be found in Table 1 and Figure 1.

This simulation study revealed, that autocorrelation at first lag is higher for *AV* returns. Note, that when computing *AV* returns, the averaging is performed over non-overlapping trading days. One of the possible explanations of this autocorrelation is that averaging acts as a filtering process, thus reducing the noise of the series (note a lower variance in all series). This allows us to measure the underlying dependence among returns, which is otherwise masked by larger noise.

Panel A: CEE markets												
	WIG				PX				BUX			
	I. approach		*II. approach*		*I. approach*		*II. approach*		*I. approach*		*II. approach*	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
ACF(1)	0.0410	0.0500	0.2650	0.0440	0.0774	0.0563	0.3146	0.0501	-0.0125	0.0457	0.2224	0.0431
VAR	0.0011	0.0002	0.0007	0.0002	0.0011	0.0002	0.0008	0.0002	0.0012	0.0002	0.0008	0.0001
Panel B: Developed markets												
	GER				STX				MSCI			
	I. approach		*II. approach*		*I. approach*		*II. approach*		*I. approach*		*II. approach*	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
ACF(1)	-0.0024	0.0569	0.2323	0.0523	-0.0336	0.0672	0.2023	0.0615	0.0193	0.0548	0.2405	0.0500
VAR	0.0013	0.0005	0.0009	0.0004	0.0014	0.0018	0.0009	0.0011	0.0007	0.0004	0.0005	0.0003

Table 1 Mean and standard deviation of the autocorrelation coefficient and variance of simulated weekly returns

Notes: Std. – standard deviation, VAR – variance, ACF(1) – autocorrelation at lag 1.

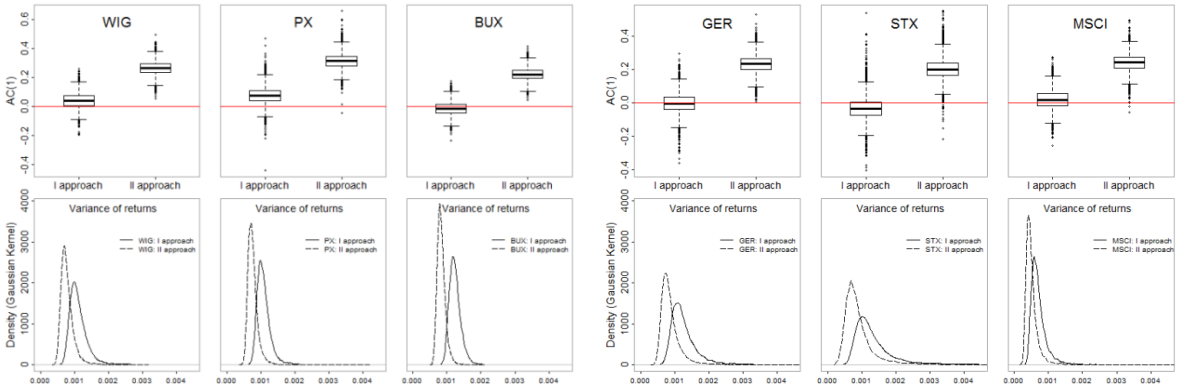


Figure 1 Simulated autocorrelation coefficients and variances of weekly returns

[9] For the sake of brevity, results from these models are not reported.

3.2 Empirical comparisons

Table 2 presents some descriptive statistics of the resulting weekly returns. These are in line with previous simulations. The standard deviation and extremes of returns for *AV* were lowest for each of the series. Similarly as in simulations, the systematically positive autocorrelation of *AV* returns was found. Small and positive first order autocorrelations suggest that *AV* returns have some short memory. This might have some implications for the random walk hypothesis (see Ding et al. [4]; Campbell et al. [3]). Therefore one may readily see that in this framework of weekly returns, the choice of the method of constructing weekly returns can already have some implications in empirical studies. Positive autocorrelation is not a rarity in financial time series. On the other hand, *W-to-W* or *F-to-F* returns have small autocorrelation, which might be more in line with the efficient market hypothesis and such series requires less filtering in a subsequent analysis.

Panel A: CEE markets									
	WIG			**PX**			**BUX**		
	AV	*W-to-W*	*F-to-F*	*AV*	*W-to-W*	*F-to-F*	*AV*	*W-to-W*	*F-to-F*
Mean	0.002	0.002	0.002	0.001	0.001	0.001	0.002	0.001	0.002
Std.	0.027	0.033	0.032	0.028	0.034	0.033	0.030	0.037	0.038
Min.	-0.128	-0.175	-0.171	-0.162	-0.170	-0.305	-0.176	-0.201	-0.269
Max.	0.093	0.125	0.120	0.135	0.133	0.156	0.095	0.151	0.152
ACF(1)	0.226	-0.037	0.042	0.240	-0.013	0.076	0.233	0.001	-0.003
JB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Panel B: Developed markets									
	GER			**STX**			**MSCI**		
	AV	*W-to-W*	*F-to-F*	*AV*	*W-to-W*	*F-to-F*	*AV*	*W-to-W*	*F-to-F*
Mean	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Std.	0.028	0.035	0.035	0.027	0.034	0.033	0.022	0.025	0.026
Min.	-0.140	-0.165	-0.239	-0.121	-0.148	-0.251	-0.153	-0.165	-0.224
Max.	0.116	0.165	0.156	0.109	0.168	0.136	0.075	0.093	0.116
ACF(1)	0.143	-0.113	-0.039	0.101	-0.175	-0.083	0.154	-0.061	-0.022
JB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 2 Descriptive statistics of weekly returns

Notes: Std. – standard deviation, Min. – minimum, Max. – maximum, ACF(1) – autocorrelation at lag 1, JB – the p-value of Jarque – Bera normality test.

Panel A: CEE markets									
	WIG			**PX**			**BUX**		
[100%]	$\frac{(Wr-Fr)}{Fr}$	$\frac{(Wr-AV)}{AV}$	$\frac{(Fr-AV)}{AV}$	$\frac{(Wr-Fr)}{Fr}$	$\frac{(Wr-AV)}{AV}$	$\frac{(Fr-AV)}{AV}$	$\frac{(Wr-Fr)}{Fr}$	$\frac{(Wr-AV)}{AV}$	$\frac{(Fr-AV)}{AV}$
1st qrt	-114.599	-28.918	-99.668	-113.676	-30.325	-87.651	-123.860	-36.498	-104.290
median	-33.706	11.816	-23.031	-33.925	12.537	-13.689	-44.580	10.004	-10.367
3rd qrt	43.893	60.315	49.302	45.237	62.934	63.028	27.720	53.410	64.231
IQR	158.491	89.233	148.970	158.913	93.259	150.678	151.580	89.907	168.521
Panel B: Developed markets									
	GER			**STX**			**MSCI**		
[100%]	$\frac{(Wr-Fr)}{Fr}$	$\frac{(Wr-AV)}{AV}$	$\frac{(Fr-AV)}{AV}$	$\frac{(Wr-Fr)}{Fr}$	$\frac{(Wr-AV)}{AV}$	$\frac{(Fr-AV)}{AV}$	$\frac{(Wr-Fr)}{Fr}$	$\frac{(Wr-AV)}{AV}$	$\frac{(Fr-AV)}{AV}$
1st qrt	-122.837	-34.460	-84.648	-130.875	-34.019	-116.234	-114.768	-28.049	-97.059
median	-39.449	11.388	-11.437	-41.251	13.303	-16.264	-39.442	14.831	-16.201
3rd qrt	37.013	53.036	72.706	37.548	64.393	58.321	36.199	51.214	63.890
IQR	159.850	87.496	157.354	168.423	98.412	174.555	150.967	79.263	160.949

Table 3 Descriptive statistics: Percentage of differences between weekly returns

Notes: Wr – W-to-W returns, Fr – F-to-F return, AV – weekly returns calculated from averages. 1st qrt – first quartile, 3rd qrt – third quartile, IQR – interquartile range.

Calculating percentage differences between weekly returns offers a more straightforward approach to comparison. As our dataset had many outliers, Table 3 presents quartiles.[10] *W-to-W* returns appear to be systematically smaller (from 33.706% to 44.580%) than *F-to-F* returns. *F-to-F* returns are also systematically larger than *AV*

[10] The presence of outliers is of no surprise, as in denominator, often very small (near zero) returns are present. This largely invalidates the use of moment based statistics like: mean, standard deviation, skewness, and kurtosis.

returns (from 10.367% to 23.031%). Finally, *W-to-W* returns appear to be systematically larger than *AV* returns (from 10.004% to 14.831%). Note, that this anomaly is different from day-of-the-week effects, although clearly it has implications on the series of weekly returns.

Perhaps even more important is the comparison of the variability of the percentage differences between weekly returns. Because of outliers, we have used interquartile range (IQR). Whenever *F-to-F* returns were used, the variability was large, at least compared to the variability of percentage differences between *W-to-W* and *AV* returns.

As previous analysis suggested, the differences between *W-to-W* and *AV* returns were smaller. We have performed a series of simple regressions to observe the dependence between different types of weekly returns. The regressions were in the following form: $Ar_t = \beta_0 + \beta_1 Fr_t + u_t$. Based on the estimates of the β_1 coefficients, we have tested whether the coefficient might be regarded as statistically different from 0, or in other instance different from 1. Inability to reject the second hypothesis would suggest that the differences between returns are not statistically significant. As errors from these regressions are probably autocorrelated and heteroskedastic, the standard errors of regression coefficients were calculated using HAC estimates of standard errors with kernel weights based on quadratic spectral and automatic bandwidth selection as proposed in Kiefer and Vogelsang [8]. Table 4 reports the results and the corresponding critical values, which are different from those derived from standard normal probability distribution.

As observed in Table 4, the dependence was highest among *W-to-W* and *AV* returns. Again, where *F-to-F* returns were involved, the dependence was smaller. All test rejected the null hypothesis, therefore all coefficients might be regarded as significantly different from 0, but at the same time, significantly different from 1, i.e., the differences between series are systematic.

Panel A: CEE markets									
	WIG			**PX**			**BUX**		
	Ar ~ Fr	*Ar ~ Wr*	*Wr ~ Fr*	*Ar ~ Fr*	*Ar ~ Wr*	*Wr ~ Fr*	*Ar ~ Fr*	*Ar ~ Wr*	*Wr ~ Fr*
Estimate of β_1	0.633	0.767	0.636	0.612	0.774	0.629	0.592	0.757	0.617
Std. error	0.049	0.019	0.062	0.033	0.020	0.053	0.025	0.013	0.026
H_0: $\beta_1 \leq 0$ [*t*-stat]	12.918	40.368	10.291	18.545	38.700	11.868	23.680	58.231	23.731
H_0: $\beta_1 \geq 1$ [|*t*-stat|]	7.490	12.263	5.890	11.758	11.300	7.000	16.320	18.692	14.731
b	0.279	0.191	0.434	0.257	0.160	0.375	0.216	0.180	0.372
5% critical values	2.732	2.341	3.670	2.627	2.173	3.297	2.442	2.261	3.297
Panel B: Developed markets									
	GER			**STX**			**MSCI**		
	Ar ~ Fr	*Ar ~ Wr*	*Wr ~ Fr*	*Ar ~ Fr*	*Ar ~ Wr*	*Wr ~ Fr*	*Ar ~ Fr*	*Ar ~ Wr*	*Wr ~ Fr*
Estimate of β_1	0.600	0.754	0.590	0.556	0.722	0.523	0.600	0.794	0.585
Std. error	0.016	0.012	0.026	0.014	0.017	0.016	0.008	0.014	0.012
H_0: $\beta_1 \leq 0$ [*t*-stat]	36.585	62.833	22.692	39.714	42.471	32.688	74.074	55.139	48.750
H_0: $\beta_1 \geq 1$ [|*t*-stat|]	24.390	20.500	15.769	31.714	16.353	29.813	49.383	14.306	34.583
b	0.322	0.160	0.453	0.329	0.159	0.470	0.326	0.194	0.441
5% critical values	2.966	2.173	3.816	2.966	2.173	3.947	2.966	2.341	3.670

Table 4 Regression coefficients: dependence between different types of weekly returns

Notes: Ar ~ Fr – this notation means that: "weekly returns calculated from averaged prices are explained by weekly returns calculated from Friday prices" (and analogously in other cases). Std. error – standard errors, b – corresponds to the proportion of the total sample size used as a bandwidth. 5% critical values are asymptotic critical values from Kiefer – Vogelsang [8].

4 Conclusion

Based on a simulation study and empirical data, we found evidence of qualitatively distinctive statistical properties of the weekly returns of selected stock market indices. Our main conclusion is that using *AV* returns introduces small and positive autocorrelation into the series, while *W-to-W* and *F-to-F* returns do not display this autocorrelation, which is usually found in daily stock returns. This is probably due to larger variance of the *W-to-W* and *F-to-F* weekly returns.

It is still premature to recommend one method over other. Clearly, if one is interested in in- or out-of-sample forecasting, *AV* returns might not be suitable at all. On the other hand, in most of the empirical studies, when one is interested in verifying certain relationships, *AV* returns have following advantages over *W-to-W* or *F-to-F* returns: (1) *AV* returns smear out the volatility of daily returns, which is lower compared to *W-to-W* or *F-to-F*

returns as well, (2) *AV* returns are less prone to day-of-the-week effects, (3) *AV* returns are less prone to non-synchronous trading effect II, which is still present for both *W-to-W* or *F-to-F* weekly returns.

Acknowledgements

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0666-11. The authors also acknowledge the funding support of the Slovak Grant Agency for Science (VEGA Project No. 1/0393/12).

References

[1] Baumöhl, E., and Lyócsa, Š.: Constructing weekly returns based on daily stock market data: A puzzle for empirical research? MPRA Working Paper 43431, 2012, University Library of Munich, Germany.
[2] Burns, P., Engle, R., and Mezrich, J.: Correlations and volatilities of asynchronous data. *The Journal of Derivatives* **5** (1998), 7–18.
[3] Campbell, J., Lo, A., and MacKinlay, A. C.: *The Econometrics of Financial Markets.* Princeton University Press, New Jersey, 1997.
[4] Ding, Z., Granger, C. W. J., and Engle, R. F.: A long memory property of stock market returns and a new model. *Journal of Empirical Finance* **1** (1993), 83–106.
[5] Égert, B., and Kočenda, E.: Time-Varying Synchronization of the European Stock Markets. *Empirical Economics* **40** (2011), 393–407.
[6] Forbes, K., and Rigobon, R.: No contagion, only interdependence: Measuring stock market comovements. *Journal of Finance* **57** (2002), 2223–61.
[7] Hamao, Y., Masulis, R. W., and Ng, V.: Correlations in Price Changes and Across International Stock Markets. *Review of Financial Studies* **3** (1990), 281–307.
[8] Kiefer, N. M., and Vogelsang, T. J.: Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation. *Econometrica* **70** (2002), 2093–95.
[9] Lo, A. W., and MacKinlay, A. C.: An econometric analysis of nonsynchronous trading. *Journal of Econometrics* **45** (1990), 181–211.
[10] Martens, M., and Poon, S. H.: Returns Synchronization and Daily Correlation Dynamics between International Stock Markets. *Journal of Banking & Finance* **25** (2001), 1805–27.
[11] Scherer, B.: Synchronize Your Data or Get Out of Step with Your Risks. *Journal of Derivatives* **20** (2013), 75–84.
[12] Schotman, P. C., and Zalewska, A.: Non – synchronous Trading and Testing for Market Integration in Central European Emerging Markets. *Journal of Empirical Finance* **13** (2006), 462–94.

Generalization of geometric median

Přemysl Bejda[1]

Abstract. In the paper are suggested new robust estimators of location and variance. It is proved that these estimators have a breakdown point one half. The used method comes from a geometric median. In the first step it is shown that we can employ one half of observations and the estimate stays robust in the sense of the breakdown point. In the second step we show that we can add even more observations which are in some sense close to the geometric median and still get robust results. The robustness is proved in both steps for a multidimensional case. Since we can employ more observations and stay robust in the sense of the breakdown point, we enlarge the used information in comparison to other robust estimators like median and therefore get better results. We combine the advantage of the robust estimator and the classical mean. Our estimators are compared by simulation study with classical estimators like mean, median or alpha windsorised estimator. The comparison is done for different distributions like normal, Cauchy or exponential. We also consider the case, when the normal distribution is with some probability contaminated by another distribution. In the last section are shown some illustrations of the approach.

Keywords: breakdown point, contamination, geometric median, robustness

JEL classification: C14
AMS classification: 91G70

1 Generalized geometric median

We will consider any normed vector space.

Further we consider the concept of a breakdown point which was introduced in [1]. We employ the definition which can be found in [3].

Consider $S_{\boldsymbol{x}_i} = \sum_{j=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$.

Definition 1. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are any observations from some normed vector space. By a geometric median we mean the number

$$\hat{\boldsymbol{x}} = \underset{i=1,\ldots,n}{\operatorname{argmin}} \sum_{j=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|. \tag{1}$$

If there are more observations which satisfies (1), we take an average of only two of them, which are different i.e. $\boldsymbol{x}_i \neq \boldsymbol{x}_j$.

Remark 1. It is not necessary to take the average of only two observations to gain geometric median. It is done only for purpose of the following proposition.

The standard definition of geometric median is slightly different from the ours.

Remark 2. By $\|\cdot\|_q$, $q \in \mathbb{N}$ we understand a norm in $\mathbb{R}^p$ such that any $\boldsymbol{x} \in R^p$ satisfies

$$\|\boldsymbol{x}\|_q = \sqrt[q]{|x_1|^q + \cdots + |x_p|^q}.$$

[1]Mathematical-Physical faculty, Charles University, Dept. of statistics, Sokolovská 83, Prague 8, Czech Republic, premyslbejda@gmail.com

And we denote $\|\boldsymbol{x}\|_\infty = \max_{i=1,\dots,p} |x_i|$.

In special one dimensional case are all norms $\|\cdot\|_q$ equal. The distance between x and y is for us $\|x-y\|$.

Proposition 1. *For any norm $\|\cdot\|_q$ and one dimensional case is the geometric median $\hat{x}$ equal to a median x^{m}.*

Proof. We know that the median can be find by $\operatorname{argmin}_{a\in\mathbb{R}} \sum_{i=1}^{n} |x_i - a|$. We order the observations. In case of n odd we get the observation in the middle. So this observation has to be as well the geometric median.

If n is even then the median is the average of two observations in the middle of our ordered observations (for simplicity we neglect some degenerated cases). Denote these two observations $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$ and let $\boldsymbol{x}^1 \leq \boldsymbol{x}^2$. Now we will compute $S_{\boldsymbol{x}^1}$ and $S_{\boldsymbol{x}^2}$. Denote D^1 the sum of distances between $\boldsymbol{x}^1$ and all observations less than $\boldsymbol{x}^1$. Similarly D^2 the sum of distances between $\boldsymbol{x}^2$ and all observations greater than $\boldsymbol{x}^2$. $D^3 = \|\boldsymbol{x}^2 - \boldsymbol{x}^1\| = |\boldsymbol{x}^2 - \boldsymbol{x}^1|$. Then $S_{\boldsymbol{x}^1} = D_1 + D_3 + \frac{n-2}{2} D_3 + D_2 = S_{\boldsymbol{x}^2}$. So the distances are same.

We know that if we remove $\boldsymbol{x}_{(n)}$ then $S_{\boldsymbol{x}^1}$ is the smallest. If we return $x_{(n)}$ back then for all $\boldsymbol{x}_i \leq \boldsymbol{x}^1$ is $S_{\boldsymbol{x}^1} \leq S_{\boldsymbol{x}_i}$, because the distance between $\boldsymbol{x}^1$ and $\boldsymbol{x}_{(n)}$ is smaller than between other these observations and $\boldsymbol{x}_{(n)}$. The same is true for $\boldsymbol{x}^2$ and observations greater than this one. So we get that the value $S_{\boldsymbol{x}^1} = S_{\boldsymbol{x}^2}$ is the smallest.

We should note that in this proof is necessary to use the last sentence from the definition of our case of the geometric median. □

Now we will deal with the breakdown point of $\hat{\boldsymbol{x}}$. According to the definition of the breakdown point we try to find when $\|\hat{\boldsymbol{x}}\| = \infty$ if enough observations satisfy $\|\boldsymbol{x}\| = \infty$.

Proposition 2. *The break down point of geometric median $\hat{\boldsymbol{x}}$ is one half.*

Proof. Let $k < \lceil n/2 \rceil$ denote the number of observations, which can be contaminated (i.e. $\|\boldsymbol{x}\| = \infty$). We propose that k points can be arbitrary far from the rest of points and our estimator stay among the original set. Denote by K_1 the set of points which were changed and by K_0 the set of points which stayed unviolated. We will proceed by a contradiction. Let us assume that $\hat{\boldsymbol{x}} \in K_1$. We denote $\boldsymbol{x}^l = \operatorname{argmin}_{i\in K_0} \|\hat{\boldsymbol{x}} - \boldsymbol{x}_i\|$, $A = \min_{i\in K_0} \|\hat{\boldsymbol{x}} - \boldsymbol{x}_i\|$, $B = \sum_{x_i \in K_1} \|\hat{\boldsymbol{x}} - \boldsymbol{x}_i\|$ and $a = \max_{i,j\in K_0} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$. Observations in K_1 are arbitrary far from the original observations. Let then $A \geq \min_{\boldsymbol{x}_i\in K_0, \boldsymbol{x}_j \in K_1} \|\boldsymbol{x}_i - \boldsymbol{x}_j\| > (n-1)a$. Now compute

$$S_{\hat{\boldsymbol{x}}} = \sum_{\boldsymbol{x}_i \in K_0} \|\hat{\boldsymbol{x}} - \boldsymbol{x}_i\| + \sum_{\boldsymbol{x}_i \in K_1} \|\hat{\boldsymbol{x}} - \boldsymbol{x}_i\| \geq (n-k)A + B.$$

For any $\boldsymbol{x} \in K_0$

$$\begin{aligned} S_{\boldsymbol{x}} = \sum_{\boldsymbol{x}_i \in K_0} \|\boldsymbol{x} - \boldsymbol{x}_i\| + \sum_{\boldsymbol{x}_i \in K_1} \|\boldsymbol{x} - \boldsymbol{x}_i\| \leq \\ (n-k-1)a + \sum_{\boldsymbol{x}_i \in K_1} \|\boldsymbol{x} - \hat{\boldsymbol{x}} + \hat{\boldsymbol{x}} - \boldsymbol{x}_i\| \leq \\ (n-k-1)a + k\|\boldsymbol{x} - \hat{\boldsymbol{x}}\| + B \leq \\ (n-k-1)a + k\|\boldsymbol{x} - \boldsymbol{x}^l\| + k\|\boldsymbol{x}^l - \hat{\boldsymbol{x}}\| + B \leq \\ (n-k-1)a + ka + kA + B = (n-1)a + kA + B. \end{aligned}$$

We get the contradiction if $S_{\boldsymbol{x}} < S_{\hat{\boldsymbol{x}}}$. This holds when $(n-1)a < (n-2k)A$. But $n-2k$ is according to the definition of k at least 1. So it is enough when $(n-1)a < A$. But this is the contradiction. □

Let us now derive another kind of estimators from the geometric median.

Definition 2. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are any observations from some normed vector space. Let in a set L is one half of observations which distances from geometric median are less than a distance of any other observation which does not lie in the L. By a G2 estimator we mean the number which is computed as an average of $\lfloor n/2 \rfloor$ observations lying in L. Denote this by $\tilde{\boldsymbol{x}}$.

Proposition 3. *The break down point of $\tilde{\boldsymbol{x}}$ is one half.*

Proof. Once more it can be contaminated at most $k < \lceil \frac{n}{2} \rceil$. L is the same set as in the definition. For such k we know that $\hat{\boldsymbol{x}} \in L$ and it is not an outlier. For a contradiction we suppose that exist $\boldsymbol{x}_u \in L$, which is an outlier. Let K_0 is a set of not outlying observations and K_1 is a set of outlying observations and $A = \max_{\boldsymbol{x}_i \in K_0} \|\boldsymbol{x}_i - \hat{\boldsymbol{x}}\|$. Let $\boldsymbol{x}$ be the estimate which was gained from the sample, where were not any outlying observations. Put $B = \min_{\boldsymbol{x}_i \in K_1} \|\boldsymbol{x}_i - \boldsymbol{x}\|$, $b = \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|$. Now $\|\boldsymbol{x}_u - \boldsymbol{x}\| \leq \|\boldsymbol{x}_u - \hat{\boldsymbol{x}}\| + \|\hat{\boldsymbol{x}} - \boldsymbol{x}\|$ from that follows $\|\boldsymbol{x}_u - \hat{\boldsymbol{x}}\| \geq B - b$. Since the outlier can be arbitrarily far from $\boldsymbol{x}$, we can choose B large enough i. e. $B - b > A$ to get the contradiction. □

The set L can be employed for constructing different kinds of estimators. We can take for instance once more the median from the set or $r = \max_{\boldsymbol{x}_i, \boldsymbol{x}_j \in L} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$ can serve as an estimator of inter quartile range (from the similar estimator we can, under assumption of normality, easily derive an estimator of variance) etc.

Definition 3. Let $r = \max_{\boldsymbol{x}_i, \boldsymbol{x}_j \in L} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$. $b > 0$. Let the set E contains all observations $\boldsymbol{x}$ such that $\min_{\boldsymbol{x}_i \in L} \|\boldsymbol{x}_i - \boldsymbol{x}\| \leq br = a$.

Proposition 4. *If we make any estimator $\tilde{\boldsymbol{x}}$ from observations in E then the estimator has the break down point one half. It means that the estimator $\tilde{\boldsymbol{x}}$ is a function of the observations from E. The function has to be bounded on any finite interval.*

Proof. Because of the property of the function, from which is the estimator constructed, we have to investigate only the situation when observation from E tends to infinity. The rest of the proof is based on the same idea as the previous proof, so we skip it. □

Remark 3. The last proposition instruct us, how to include more observations and so increase efficiency. But a should not be too large. We can for instance employ theoretical quantiles of a normal distribution to find an appropriate constant a. Let us denote the constant a as a constant of widening.

Remark 4. Thanks to the properties of a norm we can easily show that our estimators (we employ an average on the set L and E respectively) are scale and location equivariant.

2 One dimensional case and simulation study

Let us now consider only one dimensional case. We want to compute a mean from observations in E, but we have to gain the set.

Let X is a random variable with a mean μ, a standard deviation σ and let $q_{\mu,\sigma}(\alpha)$ denotes its quantile function. Then for any continuous distribution G (in case that the second moment is finite) the following ratio is constant for every μ and σ

$$K_{G,\alpha} = \frac{q_{\mu,\sigma}(1 - \alpha/2) - q_{\mu,\sigma}(\alpha/2)}{q_{\mu,\sigma}(0.75) - q_{\mu,\sigma}(0.25)}$$

This comes from the fact $\sigma q_{0,1}(\alpha) + \mu = q_{\mu,\sigma}(\alpha)$.

Let us further suppose that $X \sim \mathrm{N}(\mu, \sigma)$. We construct estimate of $q_{\mu,\sigma}(0.75) - q_{\mu,\sigma}(0.25)$ such that we put $\mathrm{RIQR} = \mathrm{argmax}_L(x) - \mathrm{argmin}_L(x)$. In this manner we construct a robust estimator

of inter quartile range RIQR and it has according to the proposition 3 the breakdown point one half. We put the constant of widening as $a = K_{\mathrm{N}(0,1),\alpha}\mathrm{RIQR}$.

Otherwise than in the definition 3 we take all observations which distance from their median is less than a. Let us denote the set GM3S. The only difference is that we do not take into account all observations from the set L, but only one from these observations. So the proposition 4 stays valid. This approach gives also better results in practise. We can employ other symmetric distributions than normal.

All employed methods and a generation of random samples were implemented in R. For each situation we have 1000 samples and in each sample 100 observations. The estimators used in simulation study are: mean, median, $\alpha-$winsorised mean, truncated mean, GM2 and GM3, where GM2 and GM3 are means from the sets L and GM3S respectively.

During the simulation study we truncate 20 percent of observations from each side for the $\alpha-$winsorised mean. We do the same for the the $\alpha-$truncated mean. And we employ $\alpha = 0.01$ for the GM3.

In the following tables is employed a function $\sum_{i=1}^{N} |\widehat{\boldsymbol{x}}_i - \mu|$, where N is a number of samples (1000), $\widehat{\boldsymbol{x}}_i$ is some estimate (mean, median, etc.) from sample i, and μ is a parameter of location, which is known (it is usually a median of the not contaminated distribution).

In the first table is a normal distribution $\mathrm{N}(0,1)$ contaminated by some other distribution with probability p.

An expression $\mathrm{U}(a,b)$ denotes a uniform distribution on the interval (a,b).

Distribution	Mean	Med	Trim	Winsor	GM2	GM3
$p = 5\%$						
$\mathrm{N}(0,100)$	187	106	90	89	123	85
Cauchy	247	99	86	86	115	86
$\mathrm{U}(-10,10)$	131	108	92	90	123	89
$p = 10\%$						
$\mathrm{N}(0,100)$	270	104	93	94	122	87
Cauchy	481	101	89	88	116	88
$\mathrm{U}(-10,10)$	170	115	102	103	130	95
$p = 40\%$						
$\mathrm{U}(-20,10)$	1960	277	546	1062	184	155

Table 1: A normal distribution contaminated by a distribution from the first column with a probability p.

Distribution	Mean	Med	Trim	Winsor	GM2	GM3
$\mathrm{N}(0,1)$	78	100	85	83	117	84
$\mathrm{N}(0,100)$	774	974	830	811	1132	824
Cauchy	5101	121	138	168	121	135
$\mathrm{U}(-10,10)$	476	803	638	561	1034	476
t_5	104	105	94	95	119	98
t_{10}	91	104	91	91	119	92
Laplace	118	87	95	105	94	99

Table 2: Different symmetric distributions.

In the table 2 are employed some symmetric distributions always with parameter of location equal to 0 (for Laplace distribution is scale equal to 1).

2.1 Conclusion

From the simulation study is visible that GM3 is robust and employ enough observations. Under assumption of normality gives GM3, $\alpha-$trimmed and $\alpha-$winsorised means almost the same results, but for more contaminated cases is our estimator much better.

A disadvantage of GM3 can appear when we set wrongly the widening constant (for different distribution), but the table 2 shows that this does not violate the results too much (especially when our choice was rather conservative for normal distribution). In case of asymmetric distribution we could spread the set L asymmetrically.

GM2 is good only in cases of really bad contamination or for some heavy tailed distributions. But it serves as the first step for GM3.

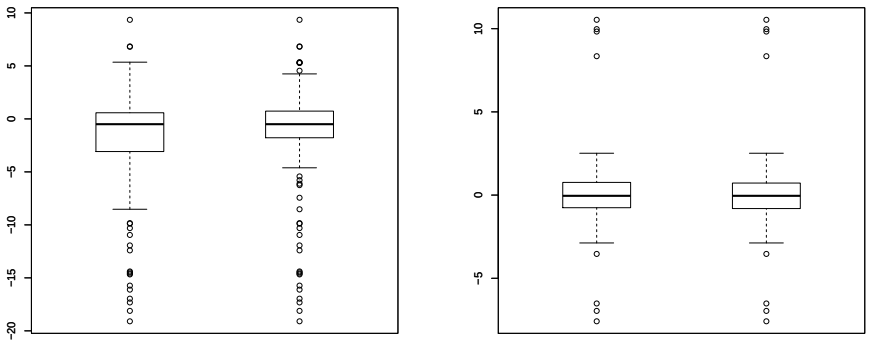
Computational complexity for both GM2 and GM3 is $n \ln n$ as for other robust estimators. But in practise it is more time consuming than computation of median.

The theoretical equivalent of GM2 and GM3 is not easy to grasp.

3 Illustrations

3.1 Boxplot

In boxplots we can employ instead of IQR the RIQR estimator. As whiskers can be taken borders, which are given by constant of widening and are used for construction of GM3 (the last and the first observation from GM3S). To get proper results it is necessary to have symmetric density of a not contaminated distribution. In a case of really bad contamination we are able to classify outliers better.



(a) $p = 0.4$, n=100, U$(-20, 10)$ (b) $p = 0.1$, n=100, N$(0, 100)$

Figure 1: On the left side is the classical boxplot and on the right side is our modification. In both cases we contaminate N$(0, 1)$ with a probability p by the observations with a distribution which is stated under the figures.

In the figure 1 is visible that the standard boxplot and our method give very similar results in

case of mild contamination. When the contamination is higher than 25 % then our boxplot is still able to find outliers in comparison to the standard boxplot.

3.2 Outlying shortfall

We can construct a similar statistic to an expected shortfall. Very brief overview about a topic of VaR and expected shortfall can be found e.g. in [2].

Let r_t is a logarithmic return in time t. We suppose that r_t are independent and has a normal distribution contaminated by some other distribution $r_t \sim p \cdot G + (1-p) \cdot \mathrm{N}(\mu, \sigma^2)$. We are interested in how much we can loss because of the contamination. We take a negative value of a mean from all observations which are less than argmin(GM3S).

As a small illustration of this approach we employ EUR/USD hourly rates from 1. 8. 2012 to 3. 11. 2012.

Results:

- A 95 % VaR is 0.00140406. As the VaR we take a negative value of a 5 % quantile from observed returns.
- A 95 % expected shortfall is 0.00219346. As the expected shortfall we take a negative value of an average of all returns lower than the 5 % quantile.
- An outlying shortfall (with a constant of widening with $\alpha = 0.01$) is 0.002376221.
- A 99 % VaR is 0.002517572.
- A 99 % expected shortfall is 0.003608807.

A disadvantage of this approach is its worse interpretation. We can also determine how often we suffer a loss because of outlying observations. In our case it is 3.91 %.

Acknowledgement: The work was supported by the grant SVV-2013 - 267 315.

References

[1] Donoho, D. L., and Huber, P. J.: The notion of breakdown point. In: *A Festschrift for Erich Lehman* (P. J. Bickel, K. A. Doksum and J. L. Hodges, eds.), Wadsworth, California, 1983, 157–184.

[2] Jorion, P.: *Financial riskmanager handbook*. JOHN WILEY & SONS, New Jersey, 2003.

[3] Jurečková, J., and Picek, J.: *Robust statistical methods with R*. Chapman & Hall/CRC, Boca Raton, Florida, 2005.

The Use of Belief Functions for the Detection of Internet Auction Fraud

Ladislav Beránek[1], Václav Nýdl[2]

Abstract. Currently, Internet auction portals are a standard part of business activities on the Internet. Anyone can easily participate in online auctions, either as a seller or a buyer (bidder), and total turnover on Internet auction portals attains billions of dollars. However, the amount of fraud in these Internet auctions is also related to their popularity. To prevent discovery, fraudsters perform normal trading behaviors and disguise themselves as honest members. It is therefore not easy to detect fraud in online auctions. Users must rely on information that is on the web auction systems available, for example user's ratings (reputation), descriptions of the offered items, time frames of different activities and records of transactions. This paper presents a set of characteristics that characterize fraudulent behavior on the internet auction – the sale of counterfeit or stolen goods. To evaluation, whether or not this fraudulent behavior occurs, an approach based on belief functions is applied. Experimental results show that this approach gives good results in detection of this type of fraud on online auctions.

Keywords: e-commerce, fraud, belief function, online auction

JEL classification: L81
AMS classification: 68T37

1 Introduction

Currently, many users participate in online auctions organized by many different Internet online auction systems operating on Internet infrastructure. Aukro [1] (Czech auction company with a turnover of USD 250 million and 2.5 million users in January 2012) is an example of such Internet auction system. Internet auctions allow their users to buy or sell a great amount of products and services. A large number of users use Internet auctions even as their main means of trading. For example Aukro states [1] that from 2.5 million users 9,300 users are professional dealers.

On the other hand, the amount of frauds on online auction is also increasing. The most common frauds are incorrect (purposely) description of goods, undelivered goods, irredeemable payments, sale of stolen goods, and more. Fraudsters are attracted by low admission costs and high profit potential.

This paper examines the characteristics of Internet auction fraud on Czech Aukro online auction systm [1] with a focus on specific fraudulent behaviour - the sale of stolen goods. The rest of this paper is organized as follows: section 2 presents the basic principles of the theory of belief functions. Section 3 summarizes some related work concerning the use of this theory in the field of Internet auctions. Section 4 presents our approach and describes key definitions of belief functions to represent fraudulent behavior. Section 5 presents our experimental results and model verification. Section 6 describes some interesting conclusions and directions for further research.

[1]University of South Bohemia, Department of Applied Mathematics and Informatics, Studentska 13, Ceske Budejovice, beranek@ef.jcu.cz
[2]University of South Bohemia, Department of Applied Mathematics and Informatics, Studentska 13, Ceske Budejovice, nydl@ef.jcu.cz

2 Background and Related Work

The online auction sellers and bidders are not in physical contact and bidders can not even physically see the auctioned items. This situation provides opportunities for cheating [6, 7].

From the perspective of the seller, online auctions bring along following risks [9, 11], particularly:

- Bidder will not to pay for the goods supplied.
- Bidder wrongly claims that the goods were not delivered.

From the perspective of the bidder:

- The seller refuses to send the goods.
- The description of the auctioned object is false.
- Seller sends a different goods or goods that does not match by its quality [8].
- Goods intended for auction is a fake or stolen.

Fraudulent behavior is on online auctions relatively widespread. The main reasons is that it is relatively easy to be carried out:

- Online auction participants are largely anonymous – they act under a pseudonyms. Internet auction systems use different methods for verifying the identity of users. These methods, however, may not be sufficiently reliable.
- Operators of online auctions insufficiently monitor course of auctions.
- The laws are often unclear with regard to specific situations or to the diversities of legal systems in different countries.

3 The Suggestion of Our Model

We used the approach similar to [4, 5, 12] and we performed statistical analysis of 28 real cases of online auctions on Aukro.cz auction system [1] in order to find out the characteristics of online auction fraud related to the sale of counterfeit or stolen goods. These cases were chosen on the basis of the complaints referred to in various internet forums, for example, [2, 3]. Users on these forums complained that somebody had stolen some their thing (radio from their car) and this thing appeared shortly after the theft on some Internet auction. We evaluated 8 auctions as auctions in which the stolen goods are sold. Following characteristics of online auction offering stolen goods observed:

1. Stolen goods are sold at inadequate low prices (at least about 20% less than the price of legitimate goods).
2. Fraudsters prefer to sold for fixed price.
3. A variety of goods are sold via fraudulent account (such as car accessories, footwear, sporting goods etc.).
4. Activities of respective fraudulent account is very short (often less than ten days).
5. In most of the cases, the goods are sold within several days of creating the account.
6. Fraudsters have accounts on multiple auction systems, the value of their reputational score is not high.

We will discuss these characteristics in the following paragraphs. We will define the corresponding belief functions. We will perform combination of respective belief functions and we will formulize conclusion whether the respective seller sells stolen goods on respective Internet auction account or not.

We have chosen the following attributes of sale of stolen goods on the basis of our analysis. (the other ones are too difficult to verify or to express them mathematically):

1. Inadequate low price;
2. Goods are sold mostly at fixed price;
3. A variety of goods being sold.

In the following text, we will denote $\Theta_i = \{stolen_i, \neg stolen_i\}$ as a frame of discernment [10] concerning the our belief that the seler i sells stolen goods.

Inadequate low price

This attribute shows that the seller i sells stolen good for lower price than it is the average price of legitimate good. The belief functions have the following form:

$$\begin{aligned} m_L(\{stolen_i\}) &= \begin{cases} v_{LA}\frac{\overline{P}-P_i}{\overline{P}} & \text{for } P_i \leq \overline{P} \\ 0 & \text{for } P_i > \overline{P} \end{cases} \\ m_L(\{\neg stolen_i\}) &= 0 \\ m_L(\Theta_i) &= \begin{cases} 1 - v_{LA}\frac{\overline{P}-P_i}{\overline{P}} & \text{for } P_i \leq \overline{P} \\ 1 & \text{for } P_i > \overline{P} \end{cases} \end{aligned} \tag{1}$$

where v_L is the weight of this evidence. We can intuitively read this weight as a reliability of this evidence, P_i – is the average price at which the seller sells certain goods. $\overline{P}$ is the average price of this goods offered through online auction system.

With this equation, we have expressed the inadequate low price of goods offered by the seller i. Usually, the lower the price of goods offered by a seller i, compared to the average price of respective goods j, the higher the suspicion that the seller offers stolen goods. Therefore, we assume that the equation reflecting the offering of legitimate goods, does not show that the seller does not offer "*stolen*", i.e. $m_L(\{stolen_i\}) = 0$.

Goods are sold mostly at fixed price

The sellers (fraudsters) want to sell his goods as quickly as possible. They want to get rid of it as quickly as possible. Therefore they prefer to sell goods at a fixed price (on Internet auction systems it is the option "buy now"). When a seller sells goods at fixed price, the auction ends and the seller does not have to wait to the end of the auction. Belief functions will have the following forms:

$$\begin{aligned} m_F(\{stolen_i\}) &= v_F\frac{N_{Fi}}{N_i} \\ m_F(\{\neg stolen_i\}) &= 0 \\ m_F(\Theta_i) &= 1 - v_F\frac{N_{Fi}}{N_i} \end{aligned} \tag{2}$$

where N_{Fi} is the number of goods of the seller i for the fixed price, N_i is the total number of goods sold by this seller.

It is valid that the higher the number of goods sold at fixed price, compared to the total number of goods sold by this seller, the higher the suspicion that this seller sells "*stolen*" goods. Therefore, we also assume that the given equation does not indicate that the seller does not sells "*stolen*" goods, i.e. $m_F(\{stolen_i\}) = 0$. The parameter v_F is in these equations the weight of evidence. We can intuitively interpret this weight as the reliability of the given evidence.

A variety of goods being sold

Let's suppose that the seller sells "*stolen*" goods. This fact has an effect that he sells the goods that he "gets". The variety of goods being sold is then higher than at the average proper seller. The belief functions of this attribute have the following form:

$$\begin{aligned} m_V(\{stolen_i\}) &= \begin{cases} v_V \frac{V_i - \overline{V}}{V_i} & \text{for } \mathrm{V}_i \geq \overline{V} \\ 0 & \text{for } \mathrm{V}_i < \overline{V} \end{cases} \\ m_V(\{\neg stolen_i\}) &= 0 \\ m_V(\{\Theta_i\}) &= \begin{cases} 1 - v_V \frac{V_i - \overline{V}}{V_i} & \text{for } \mathrm{V}_i \geq \overline{V} \\ 1 & \text{for } \mathrm{V}_i < \overline{V} \end{cases} \end{aligned} \tag{3}$$

where V_i is the amount of different types of goods sold by seller i, $\overline{V}$is the amount of different types of goods sold by proper sellers in a respective category. The v_V parameter is the weight of evidence. We can intuitively interpret this weight as the reliability of the given evidence.

It is valid that the more amount of different goods the seller sells, compared to the average types of goods sold by proper seller, the higher the suspicion that the seller sells a "fake" goods. Therefore, we also assume that the given equation does not indicate that the seller does sell "*stolen*" goods, i.e. $m_V(\{stolen_i\}) = 0$.

Combination of characteristic signs (proofs) of shill behavior

One characteristic alone is not enough to identify fraudulent behavior. This, once we have obtained the belief functions expressing our belief regarding fraudulent behavior, we combine them in a consistent manner to get a more complete assessment of what the whole group of signs indicates. The combination of belief functions is done with the help of the Dempster's combination rule [10]. We express the assumption that a given seller i sells "*stolen*" goods with the help of belief function $m(\{stolen_i\})$. We calculate the value $m(\{stolen_i\})$ using the combination of single belief functions expressing appropriate evidence:

$$m(\{stolen_i\}) = (m_L \oplus m_F \oplus m_V)(\{stolen_i\}) \tag{4}$$

The operator $\oplus$ is the Dempster's rule of belief function combination [10].

We perform the combination of multiple proofs according to the Dempster's rule – first we combine two belief functions, then we combine the result with the third belief function, fourth belief function and so forth.

Categorization of users according to the resulting belief functions representing the behavior - selling of stolen goods.

When constructing belief functions, we assume that they do not reflect that a given seller does not sell stolen goods. It is valid that $m(\{stolen_i\}) = 0$. After calculating the belief value that the seller i shows the character of illegal behavior (selling stolen goods), the value $m(\{stolen_i\})$ is assigned to the seller i as a measure which indicates the strength of the conviction that the user i sells stolen goods [4]. Now, we will categorize users into categories according to the view that a certain user i sells stolen goods. We will classify them into two categories: "*Seller sells stolen goods*" and "*Proper user*".

We have to define threshold η. We will categorize according to these threshold as follows: if $m(\{stolen_i\}) \geq \eta$ then the user i sells stolen goods. If $m(\{stolen_i\}) < \eta$ then the user i is a proper user. The threshold η will be qualified on the basis of statistical evaluations of analyzed auctions.

4 Case Study and Analysis of Results

To demonstrate the feasibility of the suggested reputation mechanism, we tested our methodology using real auction data from Aukro.cz [1]. We explored the bidding history of 28 auctions and various Internet discussions dedicated to selling stolen goods on online auctions. We investigated past auctions hosted by a particular seller. We evaluated 8 auctions as auctions in which the stolen goods are sold.

We explored mainly prices, number of auction carried out for fixed prices, number of various goods sold by sellers. We had to investigate all information manually because Aukro does not have (in contrast

to eBay any API interface enabling automatic gathering of information. We calculated the degree of belief that the seller sells "stolen" goods. We expressed belief functions defined on the section 4 on the basis of observed values. Then, we calculated the total degree that the seller sells "stolen" goods.

Some typical results of our exploration of auctions on Aukro are presented in Tables I and II. The basic masses assigned for evidence specified in our model and the resulting values are shown in Table II.

Table I. Chosen auction data collected from Czech online auction Aukro [1]

Seller i	Average price at which the seller i sells certain goods [in CZK]	Average price of this goods through on-line auction system [in CZK]	Number of goods which the seller i sells at fixed price	Total number of goods sold by seller i	Amount of different type of goods offered by seller i	Average amount of different type sold goods sold by sellers in respective category
D***r	1500	2525	2	2	2	2
O***2	700	1850	1	1	1	2
m***k	1250	1600	1	2	2	2
d***l	650	750	2	6	6	2
2***j	1420	1540	1	7	5	2
b***s	1200	1450	0	8	5	2

The values of belief of selling stolen goods are calculated using equations (1), (2), (3) and (4). Calculations are presented in Table II. The weights of evidence v_L, v_F and v_V were set in agreement with our experiments at the level 0.9, 0.7 and 0.8. We consider the characters "Inadequate low price", as the most predicative. The character "Goods sold at fixed prices" is the less reliable in determining a selling of fake goods. These values correspond to our survey of the relevance of single described characteristics from examined auctions. The value of threshold η was set on the basis of our experiments at the levels 0.80. This value correspond to results of our exploration of the relevance of single characteristics from examined auctions.

Table II. The basic masses assigned to single "selling stolen goods" characteristics ((1), (2), (3) and (4).

Seller i	$m_L(\{stolen\})$	$m_F(\{stolen\})$	$m_V(\{stolen\})$	$m(\{stolen\})$	$m(\Theta)$	Result
D***r	0.405941	0.7	0	0.821782	0.178218	Seller sells stolen goods
O***2	0.621622	0.7	0	0.886486	0.113514	Seller sells stolen goods
m***k	0.21875	0.35	0	0.492188	0.507813	Proper seller
d***l	0.133333	0.233333	0.533333	0.689926	0.310074	Proper seller
2***j	0.077922	0.1	0.48	0.568468	0.431532	Proper seller
b***s	0.172414	0	0.48	0.569655	0.430345	Proper seller

Values of $m(\{stolen\})$ express our belief that the seller sells stolen goods. On the other side, values $m(\Theta)$ represent our uncertainty or rather ignorance concerning the classification if the seller sells legal goods or if the seller sells stolen goods. The values of $m(\{stolen\})$ are high at the sellers D***r and O***2. They are greater than the threshold η.

5 Conclusion and Future Work

In our work, we presented a computational model of trust to seller on online auctions. It is based on the seller's rating obtained after performed transactions and his possible fraudulent behavior. We verified our model on Czech online auction Aukro. We performed a number of experiments on this auction. We make certain that we can increase the prediction of trust to seller by using reputation (rating information) as a basis which is completed with evaluation of possible fraudulent behavior (shilling) as additional information. Nevertheless we are also aware that the mathematical formalization of parameters used in our model (especially the parameters v_L, v_F, v_V) is necessary to increase the practical usefulness of our model.

In our future work, we want to define these parameters by the help of mathematical formulas. We will perform further statistical analyses of online auctions to verify these formulas and values of the parameters of our model.

References

[1] *Aukro* [online]. 2012 [ref. 2012-10-08]. Available from: www.aukro.cz.

[2] http://ekonomika.eurozpravy.cz/ceska-republika/6145-cesi-stale-kupuji-kradene-veci/

[3] http://forum.skodahome.cz/topic/114051-odcizeni-columbuse/

[4] Beranek, L. and Knizek, J. : The Usage of Contextual Discounting and Opposition in Determining the Trustfulness of Users in Online Auctions *Journal of Theoretical and Applied Electronic Commerce Research* **7** (2012). 34–50.

[5] Beranek, L., Tlusty, P. and Remes, R. The Usage of Belief Functions for an Online Auction Reputation Model. In: *Proceedings of the 28th Conference on Mathematical Method in Economy 2010.* Economical faculty. University of South Bohemia, Ceske Budejovice, 2010, 49–54.

[6] Chae, M., Shim, S., Cho, H. and LeeTrevathan, B. and Read, W : Empirical analysis of online auction fraud: Credit card phantom transactions *Expert Systems with Applications* **37** (2010). 2991–2999.

[7] Dong, F., Shatz, S. M. and Xu, H. : Combating online in-auction fraud: Clues, techniques and challenges *Computer Science Review* **3** (2009). 245–258.

[8] Gregg, S.D. and Scott, J : A typology of complaints about eBay sellers *Communications of the ACM* **51** (2008). 69–74.

[9] Sebesta, M. : Risks and problems of contemporary electronic C2C auction systems, *Systemova integrace* **16** (2009). 7–19.

[10] Shafer, G.: *A mathematical theory of evidence.* Princeton University Press, Princeton, 1976.

[11] Trevathan, J. and Read, W : Secure online English auctions, *E-Business and Telecommunication Networks* **21** (2008). 119–133.

[12] Turcinek, P., Stastny, J. and Motycka, A. : Usage of Cluster Analysis in Consumer Behavior Research. In *Proceedings of the 12th WSEAS International Conference on Applied Informatics and Communications (AIC '12).* Istanbul, Turkey, WSEAS Press, 2012, 172-177.

Average rate of return of pension or investment funds based on original, stochastic and continuous price index

Jacek Białek[1]

Abstract. In this paper we consider the problem of the proper construction of the average rate of return of pension (or investment) funds. The economic postulates for this kind of measure were formulated by Gajek and Kałuszka [6]. It is easy to show (see Białek [5]) that the definition of the average return given by these authors (and an original definition) can be expressed by using a chain price index. In our paper we propose and discuss an analogical continuous-time formula. We consider the situation when the prices and the number of the participating units are stochastic processes. We also show in a simulation study that all presented formulas and the continuous formula approximate each other.

Keywords: average rate of return of funds, price index theory

JEL Classification: C43, G12, G23
AMS Classification: 62P20

1 Introduction

There is a number of measures of the efficiency of open pension (or investment) funds (see Białek [3] or Białek [4]). The measures should be properly defined – it means that all changes of fund's assets, connected with any investment, should have impact on the given measure. The information about the average return of the group of funds is very important both for fund clients and fund managers. Firstly, it allows to compare the financial outcome of the given fund to the rest of funds. It may be helpful to customers in making a decision about money allocation. Secondly, having the knowledge about the average returns of investment funds from different sectors (manufacturing, agricultural, service etc.) provides important information about the financial situation within these sectors. And finally, in case of pension funds we can find legal regulations defining the *minimal rate of return* of funds based on the average rate of return. For example, in the Polish law regulations (The Law on Organization and Operation of Pension Funds, Art. 173, Dziennik Ustaw Nr 139 poz. 934, Art. 173; for the English translation see *Polish Pension…*, 1997, see [21]) half of the average return of a group of funds or the averege return minus four percentage points (depending on which of these values is higher) determines a minimal rate for any pension fund. In case of a deficit the *weak* fund has to cover it. It is always a very dangerous situation for the fund[2]. Under the Polish law the average return of a group of n pension funds is defined as:

$$\bar{r}_0(T_1,T_2)=\sum_{i=1}^{n}\frac{1}{2}r_i(T_1,T_2)\cdot\left(\frac{A_i(T_1)}{\sum_{i=1}^{n}A_i(T_1)}+\frac{A_i(T_2)}{\sum_{i=1}^{n}A_i(T_2)}\right) \tag{1}$$

where $r_i(T_1,T_2)$ denotes the rate of return of the $i-$th fund during a given time period $[T_1,T_2]$ and $A_i(t)$ denotes the value of $i-$th fund's assets at time t. Since 2004 the results of funds for the last 36 months have been verified twice a year. Unfortunately, the measure defined in (1) does not satisfy some economic postulates given by Gajek and Kałuszka [6]. Moreover, considering an even number of funds, where half of them have the return rates equal to 50% and the rest of funds have the return rates equal to (-50%), we should get the real average return rate on the level 0%. But using formula (1) we get 12.5 %. In our opinion, this is an argument for searching new definitions of the average rate of return of a group of funds. Our propositions for a discrete time can be found in Białek's paper [4]. In this work we propose and discuss an analogical formula for the continuous time. We consider the situation when the prices and the number of the participating units are some stochastic processes. We also show in a simulation study that presented discrete formulae and the continuous formula approximate each other.

[1] University of Łódź, Institute of Statistics and Demography, Department of Statistical Methods, address: 3/5 Polskiej Organizacji Wojskowej St., 90-255 Łódź, Poland, e-mail: jbialek@uni.lodz.pl
[2] In Poland, in 2001 and 2002 Bankowy Fund did not reach the minimal rate of return.

2 Economic postulates and discrete formulas of average return

At first sight the problem of constructing the average rate of return of funds seems to be straightforward. But if we look at postulates of Gajek and Kałuszka [6], which are quite natural and economically legitimate, we have to verify this opinion. The above-mentioned authors propose seven economic postulates (see also Białek [2]) and they prove that the Polish measure described by (1) violates four of them. Moreover, Gajek and Kałuszka [7] propose their own definition of the average rate of return of funds:

$$\bar{r}_{GK}(T_1,T_2)=\prod_{t=T_1}^{T_2-1}(1+\sum_{i=1}^{n}A_i^*(t)r_i(t,t+1))-1, \tag{2}$$

where

$$A_i^*(t)=\frac{A_i(t)}{\sum_{i=1}^{n}A_i(t)}=\frac{p_i(t)q_i(t)}{\sum_{i=1}^{n}p_i(t)q_i(t)}, \tag{3}$$

and $p_i(t)$ denotes the value of the participation unit of the $i-$th fund at time t, $q_i(t)$ denotes the number of units of the $i-$th fund at time t. In this part of the paper we treat the group of funds as an aggregate that contains n commodities (funds) with prices $p_i(t)$ and quantities $q_i(t)$, where $t\in[T_1,T_2]$. Let us denote by $P^L(t,t+1)$ the Laspeyres price index and by $P^{LL}(t,t+1)$ the logarithmic Laspeyres price index as follows (see von der Lippe [24])

$$P^L(t,t+1)=\frac{\sum_{i=1}^{n}q_i(t)p_i(t+1)}{\sum_{i=1}^{n}q_i(t)p_i(t)}, \tag{4}$$

$$P^{LL}(t,t+1)=\prod_{i=1}^{n}\left(\frac{p_i(t+1)}{p_i(t)}\right)^{A_i^*(t)}. \tag{5}$$

The definition (2) can be written with the use of the Laspeyres chain index. In fact we have (see Białek [5])

$$\prod_{t=T_1}^{T_2-1}P^L(t,t+1)-1=\prod_{t=T_1}^{T_2-1}\left(1+\sum_{i=1}^{n}\frac{q_i(t)p_i(t)}{\sum_{i=1}^{n}q_i(t)p_i(t)}\cdot\frac{p_i(t+1)-p_i(t)}{p_i(t)}\right)-1=\bar{r}_{GK}(T_1,T_2)\cdot \tag{6}$$

Białek [3] proposes another definition of the average return. His formula $\bar{r}_B$ can be written with the use of the logarithmic Laspeyres chain index (see Białek [5])

$$\bar{r}_B(T_1,T_2)=\prod_{t=T_1}^{T_2-1}P^{LL}(t,t+1)-1=\prod_{t=T_1}^{T_2-1}\prod_{i=1}^{n}\left(\frac{p_i(t+1)}{p_i(t)}\right)^{A_i^*(t)}-1=\prod_{t=T_1}^{T_2-1}\exp\left(\sum_{i=1}^{n}A_i^*(t)\ln\frac{p_i(t+1)}{p_i(t)}\right)-1\cdot \tag{7}$$

In the papers by Gajek and Kałuszka [7] or Białek [2], [3] we can find proofs, that measures $\bar{r}_{GK}$ and $\bar{r}_B$ satisfy all the above-mentioned postulates from Gajek and Kałuszka. It can be shown that $\bar{r}_B(T_1,T_2)\le\bar{r}_{GK}(T_1,T_2)$. Moreover, if $p_i(t+1)\approx p_i(t)$ for each i and $t\in[T_1,T_2]$, then $\bar{r}_B(T_1,T_2)\approx\bar{r}_{GK}(T_1,T_2)$. Gajek and Kałuszka claim that the Polish measure defined in (1) overestimates the real value of the average rate of return of funds. These authors consider not only the discrete stochastic model but they also propose continuous (deterministic and stochastic) measures (see Gajek and Kałuszka [8]). In the next part of the paper we present an original, stochastic and continuous measure of the average return. It seems to be a natural next step in using the chain index theory for constructing the average rate of return of funds.

3 Continuous time stochastic model

Let $\{p_i(t): t \geq 0\}$ denote the stochastic process of the price of unit of $i-$th fund ($i = 1,2,...,n$) defined on a probability space $(\Omega, \Im, P)$ and let $\{q_i(t): t \geq 0\}$ denote the stochastic process of the number of units of $i-$th fund defined on the same probability space. Let $F = \{\Im_t : t = 0,1,2,...\}$ be a filtration, i.e. each $\Im_t$ is an $\sigma-$algebra of Ω with $\Im_0 \subseteq \Im_s \subseteq \Im_t \subseteq \Im$ for any $s < t$. Without loss of generality, we assume $\Im_0 = \{\emptyset, \Omega\}$. The filtration F describes how the information about the market is revealed to the observer. We assume additionally that processes $p_i(t)$ and $q_i(t)$ are progressively measurable with respect to the family $\{\Im_t : t \geq 0\}$. In practice, the price and quantity processes have positive values. Thus in finance, the processes of share prices are often described by the geometric Brownian (Wiener) motion[3] (also known as exponential Brownian motion) as follows (see Koo [15])

$$dp_i(t) = \alpha_i p_i(t)dt + \beta_i p_i(t)dW_i(t), i \in \{1,2,...,n\}, \tag{8}$$

where the percentage drift α_i and the percentage volatility β_i are constants, $W_i(t)$ denotes the standard Wiener process. For an arbitrary initial real value $p_i(0)$ the stochastic differential equation (8) has the analytic solution (under Ito's interpretation, see Ito [13])

$$p_i(t) = p_i(0)\exp((\alpha_i - \frac{\beta_i^2}{2})t + \beta_i W_i(t)), i \in \{1,2,...,n\}. \tag{9}$$

Thus the price processes described in (9) have always positive values and additionally $p_i(t)$ is log-normally distributed (see Oksendal [18]). Let us assume that not only prices of units are described by the geometric Wiener process but also processes of number of units of funds are described as follows

$$dq_i(t) = \gamma_i q_i(t)dt + \theta_i q_i(t)dW_i(t), i \in \{1,2,...,n\}, \tag{10}$$

and thus

$$q_i(t) = q_i(0)\exp((\gamma_i - \frac{\theta_i^2}{2})t + \theta_i W_i(t)), i \in \{1,2,...,n\}, \tag{11}$$

Under above assumptions and significations we propose the following definition of the average rate of return of a group of funds on a time interval $[T_1, T_2]$

$$R_P(T_1,T_2) = \exp[\int_{T_1}^{T_2}(\sum_{i=1}^{n} A_i^*(t)\alpha_i + \frac{1}{2}\sum_{i=1}^{n} A_i^*(t)\beta_i\theta_i - \frac{1}{2}\sum_{i=1}^{n}(A_i^*(t))^2\beta_i^2 - \frac{1}{2}\sum_{i=1}^{n}(A_i^*(t))^2\beta_i\theta_i)dt + \sum_{i=1}^{n}\int_{T_1}^{T_2} A_i^*(t)\beta_i dW_i(t)] - 1 \tag{12}$$

where the integral on the right side of a formula (12) is the Ito integral (see Karatzas and Shreve [14]). Let us notice that if we reduce in (12) the random factor connected with the Wiener process taking $\beta_i(t) = 0$, and thus $\alpha_i(t) = dp_i(t)/dt$, then we obtain

$$R_P(T_1,T_2) = \exp(\int_{T_1}^{T_2}\sum_{i=1}^{n} A_i^*(t)\alpha_i dt) - 1 = \exp(\int_{T_1}^{T_2}\sum_{i=1}^{n} A_i^*(t)dp_i(t)) - 1 = P^{Div}(T_1,T_2) - 1, \tag{13}$$

where P^{Div} denotes the continuous Divisia price index (see Banerjee [1], Hulten [12]).

[3] Geometric Brownian Motion is used to model stock prices in the Black–Scholes model and is the most widely used model of stock price behavior (see Hull [11]).

If $\beta_i(t) = 0$ we also have from (9)

$$\frac{p_i(T_2)}{p_i(T_1)} = \exp[\alpha_i(T_2 - T_1)]. \tag{14}$$

It can be shown that from (13) and (14) we obtain

$$R_p(T_1, T_2) = \prod_{i=1}^{n} \left(\frac{p_i(T_2)}{p_i(T_1)}\right)^{w_i} - 1 = P^{CD} - 1, \tag{15}$$

where

$$w_i = \frac{\int_{T_1}^{T_2} A_i^*(t)dt}{T_2 - T_1} \quad , \quad \sum_{i=1}^{n} w_i = \frac{1}{T_2 - T_1} \sum_{i=1}^{n} \int_{T_1}^{T_2} A_i^*(t)dt = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} \sum_{i=1}^{n} A_i^*(t)dt = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} dt = 1 , \tag{16}$$

and P^{CD} is the well known Cobb-Douglas price index (see von der Lippe [24]). From (6), (7), (13) and (15) we conclude that the stochastic proposition of the average return rate (12) seems to be well-constructed. It can be also shown (the proof is omitted) that in general, stochastic case one of the most important postulates (from Gajek and Kałuszka) holds, namely

$$R_A(T_1, T_2) + 1 = (R_P(T_1, T_2) + 1)(R_Q(T_1, T_2) + 1), \tag{17}$$

where $R_A(T_1, T_2)$ and $R_Q(T_1, T_2)$ denote respectively the relative change in net assets and the number of clients of funds.

4 Simulation study

Let us take into consideration a group of $n = 4$ funds, the time horizon of observations $T = 1$ and the following parameters of prices of units and numbers of units processes

(a) the case of prices of units (see (8)):

$$\alpha_1 = 0{,}3,\ \beta_1 = 0{,}25,\ \alpha_2 = -0{,}12,\ \beta_2 = 0{,}05,\ \alpha_3 = 0{,}55,\ \beta_3 = 0{,}3,\ \alpha_4 = -0{,}45,\ \beta_4 = 0{,}07,$$

(b) the case of numbers of units (see (10)):

$$\gamma_1 = 0{,}25,\ \theta_1 = 0{,}05,\ \gamma_2 = -0{,}45,\ \theta_2 = 0{,}1,\ \gamma_3 = 0{,}7,\ \theta_3 = 0{,}25,\ \gamma_4 = 0{,}3,\ \theta_4 = 0{,}03.$$

Without loss of generality we assume that $p_i(0) = q_i(0) = 1$ for each $i \in \{1,2,...,4\}$. Some realization of the average return rate $R_P(0,t)$ for $t \in [0,1]$ is presented in Fig.1. The generated values of index $R_P(0,1)$ for each of $i-th$ realization of price and quantity processes ($R_{Pi}(0,1) : i = 1,2,...,100$) are presented in Fig. 2.

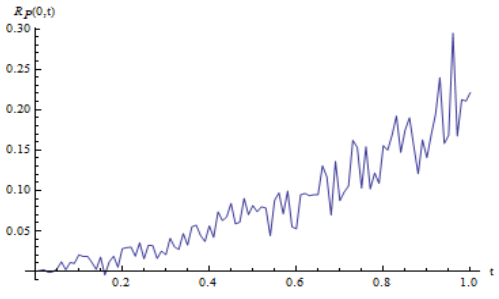


Figure 1 Realization of $R_P(0,t)$ process

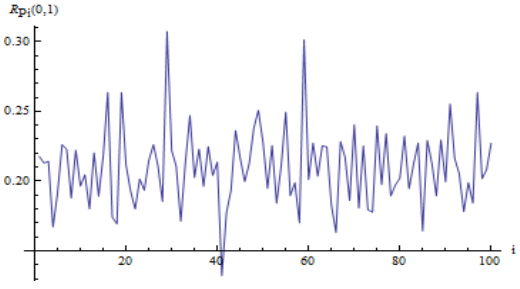


Figure 2 Generated values of index $R_P(0,1)$

Next we compare the stochastic version of the average rate of return of funds $R_P(0,1)$ with average rates of return $\bar{r}_{GK}(0,1)$ and $\bar{r}_B(0,1)$ for which we divide the time interval $[0,1]$ into ten subintervals of the same length. The results of our comparisons for $n = 10000$ generated realizations[4] of prices of units and numbers of units processes are presented in Tab.1 (to read more about estimation of mean value and variance and the bias of this estimation see Żądło [23], Małecka [17] or Papież, Śmiech [19]).

Parameter	$R_P(0,1)$	$\bar{r}_B(0,1)$	$\bar{r}_{GK}(0,1)$
Mean	0,233	0,432	0,479
Standard deviation	0,020	0,216	0,238
Median	**0,232**	**0,241**	**0,273**
Median deviation	0,014	0,141	0,130

Table 1 Basic parameters of average return rates

5 Conclusions

The form of the R_p measure seems to be proper – in the deterministic case, where $\beta_i(t) = 0$, it can be expressed by using some known chain indices (see (13) or (15)). The well-constructed $\bar{r}_{GK}$ and $\bar{r}_B$ have the same property in a discrete version (see (6) and (7)). Moreover, the stochastic definition R_p and the mentioned measures seems to approximate each other (see our simulation study). But let us notice that only $R_P(0,1)$ rate has a small volatility in our research (see Tab. 1). We suppose that some extreme realizations of price or (and) quantity processes lead to extreme values of the considered chain indices. Using medians for our comparison and thus ruling out these extreme realizations we obtain quite good approximations (for example: 0,232 in case of $R_P(0,1)$ and 0,241 in case of $\bar{r}_B(0,1)$).

Apart from measuring fund return rates, the proposed formula might find use in measuring mean efficiency of any class of investable assets, and in particular mutual funds. If the class is too broad to be completely enumerated, a representative sample might be chosen according to various criteria of representativeness discussed by Kruskal [16] and Gamrot [10]. The measure could be also used to calculate a premium rate in motor insurance (see Szymańska [22]).

References

[1] Banarjee, K.S.: A note on the Divisia indices. *Statistische Hefte* 20 (1979), 172-175.

[2] Białek, J.: Jak mierzyć rentowność grupy funduszy emerytalnych? Model stochastyczny. In: *Modelowanie Preferencji a Ryzyko'05, Praca zbiorowa pod redakcją naukową Tadeusza Trzaskalika*, Wydawnictwo Akademii Ekonomicznej w Katowicach, Katowice, 2005, 329-343.

[3] Białek, J.: New definition of the average rate of return of a group of pension funds. In: *Financial Markets: Principles of Modelling, Forecasting and Decision-Making*, Łódź, 2008, 126-135.

[4] Białek, J.: The Proposition of Maesure of Pension Fund's Effectiveness. *Acta Universitatis Lodziensis, Folia Oeconomica* 228 (2009), 331-339.

[5] Białek J.: Zastosowanie statystycznych indeksów łańcuchowych do oceny przeciętnego zwrotu grupy OFE. In press: *Inwestycje finansowe i ubezpieczenia – tendencje światowe a polski rynek*, Wrocław, 2012.

[6] Gajek, L., Kałuszka, M.: On the average return rate for a group of investment funds. Acta Universitas Lodziensis, Folia Oeconomica 152 (2000), 161-171.

[7] Gajek, L., Kałuszka, M.: On some properties of the average rate of return – a discrete time stochastic model. *Working paper*, 2001.

[8] Gajek, L., Kałuszka, M.: On some properties of the average rate of return – a continuous time stochastic model. *Working paper*, 2002.

[9] Gamrot, W.: Mean Value Estimation Using Two-Phase Samples with Missing Data in Both Phases. *Acta Appl Math* 96 (2007), 215–220.

[4] Many authors claim that $n = 10000$ is a sufficient number of replications (see for example Pekasiewicz [20]).

[10] Gamrot, W.: Representative Sample Selection via Random Search with Application to Surveying Communication Lines, In: *Proceedings of 26th International Conference, Mathematical Methods in Economics 2008*, (Eds: P. Rehorova, K. Marsikova, Z. Hubinka). Technical University of Liberec, Czech Republic, 2008, 125-131.
[11] Hull, J.: *Options, Futures, and other Derivatives* (7 ed.), Prentice Hall, New Jersey, 2009.
[12] Hulten, C.R.: Divisia Index Numbers, *Econometrica* 41 (1973), 1017-1025.
[13] Ito, K.: On stochastic differential equations. Memoirs, *American Mathematical Society* 4 (1951), 1–51.
[14] Karatzas, I., Shreve, S.: *Brownian Motion and Stochastic Calculus* (2 ed.), Springer, 1991.
[15] Koo, H.: Consumption and portfolio selection with labor income: a continuous time approach, *Mathematical Finance* 8 (1998), 49-65.
[16] Kruskal W., Mosteller F.: Representative Sampling III: the Statistical Literature. *International Statistical Review* 47 (1979), 245-265.
[17] Małecka, M.: Prognozowanie zmienności indeksów giełdowych przy wykorzystaniu modelu klasy GARCH. *Ekonomista* 6 (2011), 843-860.
[18] Oksendal, B. K.: *Stochastic Differential Equations: An Introduction with Applications*, Springer, 2002.
[19] Papież, M., Śmiech, S.: *Causality in mean and variance within the international steam coal market*. *Energy Economics* 36 (2013), 594–604.
[20] Pekasiewicz, D.:, *Bayesian statistical tests for parameters of structural changes model with Bernoulli variable,* In: *Proceedings of 7th Professor Aleksander Zeliaś International Conference Modelling and Forecasting of Socio-Economic Phenomena (Eds :M. Papież., S. Śmiech).* Cracow University of Economics, Poland, 2013, 115-123.
[21] *Polish Pension Reform Package*, Office of Government Plenipotentiary for Social Security Reform, 1997.
[22] Szymańska, A.*: The use of asymmetric loss function for estimating premium rates in motor insurance*, In: *Proceedings of 7th Professor Aleksander Zeliaś International Conference Modelling and Forecasting of Socio-Economic Phenomena (Eds :M. Papież., S. Śmiech).* Cracow University of Economics, Poland, 2013, 181-189.
[23] Żądło, T.: On prediction of total value in incompletely specified domains. *Australian & New Zealand Journal of Statistics* 48 (2006), 269–283.
[24] von der Lippe, P.: *Index Theory and Price Statistics*, Peter Lang, Frankfurt, Germany, 2007.

Measuring European railway efficiency using DEA approach.

Jaroslav Bil[1]

Abstract. There has been a great deal of interest in analysing deregulating and restructuring policies in the European railway sector. A large amount of studies is based on data envelopment analysis (DEA) for evaluating such policies. One of the main reasons why DEA gain so much popularity is, that it allows to measure efficiency, when production process presents the structure of multiple inputs and outputs, without knowing information about prices. On the other hand the complete freedom in choosing weights for inputs and outputs can cause overestimation of efficiency levels for some DMUs. The objective of this paper is to evaluate relevance of this fact in railway sector. For this purpose it's performed common CCR, BCC and SBM model and after a deeper insight into the results assurance region method is drawn up. Finally efficiency levels of selected models and resulting rankings of European countries are compared. Analysis is performed on the national rail system data for the year 2008 for 23 European countries using passenger and freight services as outputs and length of lines,staff, freight and passenger transport stock as inputs.

Keywords: DEA, railway, efficiency, CCR, BCC, SBM, assurance region method

JEL classification: C670,L920
AMS classification: 90B06,90B10,90B90

1 Introduction

During the 90's the European rail industry began to be reformed and deregulated in the hope of boast efficiency by promoting competition. Generally we can say that reforms go through two levels: by vertical dimension, when infrastructure separated from operation and horizontal dimension when service on rails has been enable to more operators. But different counties gradually introduced reforms differently with respect to the extent and time frame. These structural changes in the big industry as the railway sector is, created room for empirical investigation of the impact of reforms. Each such analysis starts by measuring efficiency of countries. Then in the analysis of the second stage obtained efficiency levels are regressed on dummy variables identifying reforms and other control variables. There are two common methodology for measuring efficiency levels. The first one is deterministic non parametric approach known under the name DEA (Data Envelopment Analysis) and the second one is based on the estimation of stochastic parametric distance function, called SFA (Stochastic Frontier Analysis). In this paper I will pay attention to the former.

The main reasons for which DEA has gained so much popularity is that it allows to handle multiple outputs and inputs without knowing information about prices and about functional form for the production function. On the other hand, these undemanding model assumptions bear its drawbacks. One of the most known disadvantage is connected with the fact, that DEA method is deterministic and so very sensitive to outliers. Another problem, which however empirical papers doesn't address, is related to the selection of the weights for evaluation of the efficiency level. DEA allows each item to select weights that maximise its own efficiency score. But sometimes we can find large differences in weights from item to item and even DEA can admit to set up values of certain inputs or outputs to zeros. Then such a selection doesn't have to appropriately reflect reality and by freedom in choosing weights efficiency levels of some countries may be overestimated. The aim of the paper is to assess the relevance of consequences of this fact.

[1]Masaryk University, Faculty of Economics and Administration, Department of Economics, Brno, jbil@mail.muni.cz

For this purpose, at first common DEA CCR, BCC and not so known SBM model are performed. Then assurance-region model is drawn up and obtained efficiency scores are compared.

2 Data

The production process in railway sector is very complex making measurement of performance complicated. Since railways transport both passengers and freight, two standard outputs are used: passenger-km transported (pkm) for passenger transport and tonnes-km transported (tkm) for freight transport [1]. In terms of input variables, the following are considered: mean annual staff strength (l) recalculated to full time equivalents, length of lines (d) measured as number of km of railway infrastructure and finally a proxy for capital stock (kf) indicating the number of wagons in freight transport and (kp) embodying the number of railcars and coaches in passenger transport. The analysis is performed for 23 European countries on the sample for the year 2008. Data information was taken from the reports published by International Union of Railways. Since in some countries there are more than one operator in freight or passenger transport, data of all companies were summed up for each country.

	Mean	**St, dev.**	**Min**	**Max**
pkm	14788	23754	78	86664
tkm	17335	21575	743	91178
d	8605	9039	699	33862
l	50717	59502	2938	240008
kp	3987	4978	123	18671
kf	23294	27303	1498	119916

Table 1 Summary statistics

3 Methodology and results

Data envelopment analysis was at first introduced in [2]. It's linear programming method which try to identify piecewise linear production possibility frontier. On its basis efficiency score of decision making units, here represented by countries, is determined. Efficiency score denoted θ is then defined as ratio of weighted outputs to weighted inputs, where the weights are selected in a manner to maximize resulting score. θ can also be interpreted as a ratio to which a given country could proportionally reduce its inputs without losing any outputs[2]. Though this method enable to work with the production process of multiple inputs and outputs and doesn't require any information about prices neither functional form of production, it's implicitly assumed, that a constructed production function is achievable for all countries.

3.1 CCR,BCC and SBM model

CCR is the most standard model, which production possibility set is based on the constant return to scale assumption. The linear programming model for each country is defined as follows:

$$\begin{aligned} \max uy_0 \quad & \text{s.t.} & & & \min \theta \quad & \text{s.t.} & (1)\\ vx_0 &= 1 & \iff & & Y\lambda &\geq y_0 & (2)\\ uY - vX &\leq 0 & & & \theta x_0 - X\lambda &\geq 0 & (3)\\ u,\, v &\geq 0 & & & \lambda &\geq 0, & (4) \end{aligned}$$

[1] A possible alternative in place of passenger-km and tonnes-km is to use train-km. But the former is preferred by the vast majority of studies, because it takes into consideration also allocation efficiency and not only the technical one.

[2] This type of model is called input-orientation and is more convenient for railway industry. Also there exist a output-oriented form of the model, when the efficiency score is interpreted as how much we can raise outputs without altering inputs.

where y_0 is the column vector of outputs for country 0, $y_0 = (pkm_0, tkm_0)$, x_0 is the column vector of inputs for country 0, $x_0 = (l_0, d_0, kf_0, kp_0)$, Y is the output matrix of dimension 2×23 one column for each country, X is the input matrix of dimension 4×23, one column for each country and u, v are the row vectors of output/input weights, which are unknown. Finally θ represents an efficiency score and λ's are the Lagrange multipliers of linear problem (and also searched coefficients of dual problem)

On the left hand side the model is in the multiplier form, which could be be informally interpreted as the effort to maximize weighted output (efficiency score) under the limitation, that at given weights any country cannot achieve greater efficiency score than 1, eq.(2). Its dual problem is on the right hand side, which can be seen as maximal possibility to radially lower inputs guaranteeing the same output (eq. 2) assuming that the combination $(\theta x_0, y_0)$ remain insight production possibility set (eq. 3). In other words it can be represented as a non-negative combination of inputs and outputs of DMUs (here captured by vector λ). Banker, Charnes and Cooper in [1] enhanced the model to implementation variable return to scale from decreasing to increasing. Formally it's simple done by adding one restriction to dual problem in the following form:

$$\sum_{i=1}^{23} \lambda_i = 1. \tag{5}$$

This enables to take only convex combinations of DMUs for construction of production possibility frontier. Since these two models in dual form differ only in that one constraint, production possibility set of the BCC model is the subset of production possibility set of the CCR model and so resulting efficiency score of the BCC model cannot be smaller than that of the CCR model. This fact can also be supported by the results contained in the first two columns of the table 2. Empirical studies in railway sector are based on the one of these two models [3]. Eventually they compare both approaches as e.g.[3]. From their results *Cantos at al* claim that the differences between two methods are especially notable for small countries. Authors explain the fact in the way, that there are economies of scale in small railway systems which fail in exploiting them. But as is stated in [5], a DMU that has a minimum input value for any input item, or a maximum output value for any output item, is BBC-efficient. Thus countries whit extreme values of any input factor are considered as efficient without knowing anything about what reality about efficiency score in fact is. In that study, it concerns Germany and Macedonia. Especially in the case of Macedonia, on the basis of CCR model it's hard to believe, that this country would lie on production possibility frontier. In the deeper look into result, specifically into reference sets[4], it could be possible to notice, that a vast majority of small countries (such as Bosnia and Herzegovina, Greece, Moldova and Slovenia) have Macedonia in their reference set in BCC model but not in the CCR model. This could be a more important reason than variable returns to scale, why small countries have bigger efficiency score in BCC model.

Now let's turn look on the SBM (slack based measure) model, which was introduced by Tone in [8]. Beside the CCR model [5] it take into computation of efficiency scores moreover values of slacks. Slacks are output shortfalls or input excesses which after radial projection onto production possibility set remain. To ensure CCR or SBM efficiency these slacks have to be zero. SBM efficiency score is defined by following product formula:

$$\rho = \left(\frac{1}{4} \sum_{i=1}^{4} \frac{x_{i0} - s_i^-}{x_{i0}} \right) \cdot \left(\frac{1}{2} \sum_{j=1}^{2} \frac{y_{j0} + s_j^+}{y_{j0}} \right)^{-1}, \tag{6}$$

where the first term evaluates the mean proportional reduction rate of inputs and the second evaluates the mean proportional expansion rate of outputs. After some transformations, SBM model can be expressed in linear form as follows:

[3] Either they prefer the CCR model e.g.[4] or to BCC model e.g.[6].

[4] Reference set is a set of countries with which a particular country is evaluated, in other words a set of those countries which are connected with non-zero λ^*.

[5] It can be also implemented in BCC model, but due to greater credibility of CCR results, the SBM model was implemented in this form.

$$\min t - \frac{1}{4}\sum_{i=1}^{4}\frac{S_i^-}{x_{i0}} \quad \text{s.t.} \quad 1 = t + \frac{1}{2}\sum_{j=1}^{2}\frac{S_i^+}{y_{j0}}$$
$$tx_0 = X\Lambda + S^-$$
$$ty_0 = Y\Lambda - S^+$$
$$\Lambda \geq 0,\ S^- \geq 0,\ S^+ \geq 0,\ t > 0,$$

where S^-, S^+ and Λ are defined as $S^- = ts^-$, $S^+ = ts^+$ and $\Lambda = t\lambda$ and where s^- denotes input slack, s^+ output slack and t convenient positive constant. All other letters have the same meaning as that in problem describe by (1-4).

The results of the efficiency score can be found in fourth column in the table 2. Incorporating slacks results in lower efficiency scores of the SBM model in comparison with the CCR model. But for some countries as Croatia and Turkey the drop is more significant and on the other hand in the Czech Republic, Slovakia and Slovenia the drop was gentle and accordingly they slightly jumped upwards in the ranking, see 7th column of the table 2.

DMU	θ_{CRR}	θ_{BCC}	θ_{SBM}	r_C	r_B	r_S	$\frac{u_{pkm}}{u_{tkm}}$	$\frac{v_{kp}}{v_{kf}}$	$\frac{v_l}{v_{kp}+v_{kf}}$	$\frac{v_d}{v_{kp}+v_{kf}}$	θ_{AR_C}	θ_{AR_B}	r_{AR_C}	r_{AR_B}	$\frac{v_d}{v_{kp}+v_{kf}}$	$\frac{v_l}{v_{kp}+v_{kf}}$	$\frac{v_{kp}}{v_{kf}}$	$\frac{v_{pkm}}{v_{tkm}}$	AR_{C2}	r_2
AUS	0.79	0.79	0.67	11	15	10	4.76	1.2E10	0.00	0.35	0.62	0.67	10	12	0.30	0.20	20.00	2.08	0.61	10
BAH	0.22	0.73	0.17	23	17	23	9.01	3.5E09	0.00	0.00	0.17	0.48	23	16	0.30	0.20	0.20	0.43	0.16	23
BEL	1.00	1.00	1.000	1	1	1	0.6	1.10	0.01	1.18	0.86	1.00	7	1	0.30	0.20	0.20	0.43	0.84	6
BLG	0.32	0.37	0.25	22	22	22	9.01	2.3E10	0.00	0.00	0.22	0.27	22	23	0.30	0.20	0.20	0.43	0.21	22
CRO	0.71	0.84	0.45	12	14	13	9.01	2.1E11	0.00	0.00	0.38	0.49	15	15	0.3	0.20	20.00	2.08	0.34	15
CZE	0.36	0.36	0.33	19	23	17	4.76	1.1E09	0.00	0.35	0.33	0.33	17	21	0.3	0.20	20.00	2.08	0.32	17
ESP	1.00	1.00	1.00	1	1	1	37.75	258.13	0.12	0.05	0.94	0.97	5	8	0.21	0.29	1.82	5.00	0.88	5
FIN	1.00	1.00	1.00	1	1	1	4.07	79.94	0.09	0.02	0.91	0.92	6	9	0.20	0.30	20	1.92	0.84	7
FRA	1.00	1.00	1.00	1	1	1	9.9	0.64	0.02	0.00	1.00	1.00	1	1	0.30	0.20	0.55	3.06	1.00	1
GER	0.88	1.00	0.72	9	1	9	4.76	1.5E09	0.00	0.35	0.69	1	9	1	0.30	0.20	20	2.08	0.65	9
GRE	0.42	0.59	0.30	17	18	18	5.86	1.1E11	0.08	0.00	0.34	0.52	16	14	0.24	0.26	20.00	5.00	0.32	16
HUN	0.49	0.55	0.40	15	19	15	0.35	0.46	0.06	0.00	0.43	0.46	13	17	0.30	0.20	0.2	0.43	0.39	14
ITA	0.93	0.94	0.80	8	11	8	3.6E07	6.8E07	0.20	0.25	0.83	0.89	8	10	0.30	0.20	20.00	5.00	0.78	8
LIT	1.00	1.00	1.00	1	1	1	0.66	236.84	0.11	0.27	1.00	1.00	1	1	0.27	0.23	12.53	0.64	1.00	1
MKD	0.33	1.00	0.27	21	1	21	9.01	1.0E11	0.00	0.00	0.28	1	19	1	0.30	0.20	0.20	0.43	0.26	19
MOL	0.34	0.77	0.28	20	16	20	4.76	4.9E09	0.00	0.35	0.23	0.39	21	19	0.30	0.20	0.20	0.43	0.24	21
POL	0.55	1.00	0.44	14	1	14	9.01	3.0E06	0.00	0.00	0.42	0.6	14	13	0.30	0.20	20.00	2.08	0.39	13
POR	1.00	1.00	1.00	1	1	1	2.61	64.47	5.12	0.38	1.00	1.00	1	1	0.20	0.30	10.73	4.00	1.00	1
RO	0.45	0.45	0.30	16	20	18	9.01	2.3E10	0.00	0.00	0.28	0.28	20	22	0.30	0.20	20.00	2.08	0.26	20
SLO	0.65	0.86	0.57	13	13	11	0.71	2.57	0.00	0.00	0.58	0.79	11	11	0.30	0.20	0.20	0.43	0.55	11
SUI	1.00	1.00	1.00	1	1	1	818.63	0.28	0.39	3.36	1.00	1.00	1	1	0.30	0.20	1.32	1.37	1.00	1
SVK	0.39	0.43	0.34	18	21	16	4.76	1.5E13	0.00	0.35	0.32	0.34	18	20	0.30	0.20	0.20	0.43	0.32	18
TUR	0.86	0.88	0.52	10	12	12	9.01	3.5E10	0.00	0.00	0.45	0.45	12	18	0.21	0.29	20.00	1.97	0.40	12

Table 2 Efficiency levels, ranking and weights

3.2 Assurance-region model

Above models allow for computation efficiency scores select both for inputs and outputs arbitrary non-negative weights. So this great flexibility is associate with the weakness, when can be huge differences in weights assigned to individual countries which may distort relative comparisons and also bias efficiency levels upward. Even for inefficient countries DEA enable for some inputs or outputs assign weight equal to zero. One way how to neutralise this problem is to add constraints for weight control. This type of model is called assurance-region method and first it was introduced in 1986 in [7].

But the task to select admissible bounds for the ratios of weights is not so simple. As is recommended in [5], it's possible to use of the knowledge of experts or of the evaluated weights of CCR model for preferable efficient countries. In the look on selected ratios of weights contained in columns 8–11 in table 2, it can be seen that both great relative magnitude in weights and many zeros are present in performed CCR model. Among efficient countries is really not easy to find some representative ones. With the help of the mean values of selected variables (table 1), given weights from the CCR model and some theoretical background, following bounds were set up. Bounds were selected in a very relaxed way, rather from the point of view to prevent unrealistic weighting. It was imposed so that ratio of weights of passenger and tonne kilometre would be between 0.2 and 5 ($0.2 \leq \frac{u_{pkm}}{u_{tkm}} \leq 5$); ratio of weights of passenger and freight capital stock would be between 0.2 and 20 ($0.2 \leq \frac{v_{kp}}{v_{kf}} \leq 20$); ratio of weights of staff and distance of traffic

lines to total capital stock would be between 0.2 and 1 ($0.2 \leq \frac{v_l}{v_{kp}+v_{kf}} \leq 1$, $0.2 \leq \frac{v_d}{v_{kp}+v_{kf}} \leq 1$) and finally the sum of staff and lines to total capital to be at least 0.5 ($0.5 \leq \frac{v_l+v_d}{v_{kp}+v_{kf}}$). The linear programming problem for each country is formulated as follows:

$$\begin{aligned} \min \theta \quad \text{s.t.} \quad & Y\lambda + Q\tau \geq y_0 \\ & \theta x_0 - X\lambda + P\pi \geq 0 \\ & \lambda \geq 0,\ \pi \geq 0,\ \tau \geq 0, \end{aligned}$$

where added restrictions in linear representation are contained in matrices P and Q, so that

$$P = \begin{pmatrix} 0 & 0 & 0 & 0 & -1 & 1 & -1 \\ 0 & 0 & -1 & 1 & 0 & 0 & -1 \\ -1 & 1 & 0.2 & -1 & 0.2 & -1 & 0.5 \\ 0.2 & -20 & 0.2 & -1 & 0.2 & -1 & 0.5 \end{pmatrix}, \quad Q = \begin{pmatrix} -1 & 1 \\ 0.2 & -5 \end{pmatrix}.$$

In comparison of the efficiency scores of the AR-CCR with CCR model, following facts can be noticed. First efficiency scores of AR-CCR model are always lower. But this consequences is imposed by adding restriction, which can limit achieving highest possible score if the corresponding bounds don't lie within bounds. In average efficiency level drop about 18 %, with a standard deviation of 14 which is not negligible and also not uniform change. On the other hand the influence on the ranking of countries is not so great. It can be found only small changes (for example Czech Republic and Slovenia jump by two places up otherwise Croatia and Romania droped by three and four places down). Also Belarus, Spain and Finland lost the position of efficient countries, so that production possibility function has changed. Finally in 4 columns to the end of the table 2, ratios of weights of AR-CCR model are referred, where it's confirmed, that in a majority of cases added constraints for ratios of weights are binding.

In the 13th column of the table 2 results of AR-BCC model are shown for completeness and the corresponding ranking is listed two columns further. It is worth noting here, that in the case of Poland, dramatic drop of efficiency level between AR-BCC and BCC model is present. Though Poland has efficiency score equal to 1, it isn't BCC efficient, since slacks aren't zero and in fact very high. Thus some linear combination of countries of reference set of Poland (BEL, LIT, FRA) caused by chance, that Poland lies on the edge of the production possibility frontier though with a high output shortfall and input excesses. But bounds for ratios of weights don't enable such as fitting and consequently more realistic scores are returned.

Further another AR-CCR model was performed with more restrictive bounds to evaluate sensitivity of efficiency scores on given bounds. Following restriction has changed to the these ones: $2 \leq \frac{v_{kp}}{v_{kf}} \leq 10$, $0.2 \leq \frac{v_d}{v_{kp}+v_{kf}} \leq 1$, $1 \leq \frac{v_l+v_d}{v_{kp}+v_{kf}}$. Results, efficiency scores and ranking, can be found in last two columns of table 2. By a minor modification of restriction, the results didn't change significantly and also the rankings remain the same to the swapping of Belorus to Finland. So efficiency scores are rather sensitive to the inclusion of bounds but their minor modification doesn't have practically any impact.

	CCR	**BCC**	**SBM**	**AR-CCR**	**AR-BCC**	**AR-CCR(2)**
CCR	1.000	0.748	0.987	0.967	0.722	0.968
BCC	0.748	1.000	0.732	0.740	0.906	0.745
SBM	0.987	0.732	1.000	0.975	0.730	0.980
AR-CCR	0.967	0.740	0.975	1.000	0.776	0.999
AR-BCC	0.723	0.906	0.730	0.776	1.000	0.775
AR-CCR(2)	0.968	0.745	0.980	0.999	0.775	1.000

Table 3 Spearman rank correlation coefficients

For final summarization of results Spearman rank correlation coefficient were calculated and are contained in the table above (table 3). All the coefficients lie above 0.7, thus the choice of DEA method is not so crucial, nevertheless has non negligible impact on efficiency scores and for precise analysis should

definitely taken into account. Also this empirical railway study shows, that it does matter in the choice between CCR and BCC model.

4 Conclusion

Nowadays due to many reforms and structural changes in European railway industry, measuring efficiency levels becomes actual and starting point for further analysis. This paper discussed and compared standard techniques based on the DEA method. The most common in this industry is CCR and BCC model. Some authors claim that BCC model is better because it enable to handle variable returns to scale. But as this study shown this advantage is questionable and one have to be careful if some inefficient small or big country don't create production possibility frontier and consequently don't raise artificially resulting efficiency scores for some countries. Also SBM model, which incorporates input excesses and output shortfalls was performed to give more precise results. Resulting efficiency scores didn't change significantly but some corrections can be found.

The main objective of this paper was to check the impact of freedom in choosing weight for inputs and outputs by adding lower and upper bounds for selected ratios of weights. At first empirical results show, that for a majority of countries selected weights, which maximize efficiency scores, are far from reality. By imposition of some though relaxed bounds, the efficiency score decrease in average nearly by 20%. Even though efficiency scores changed significantly, the ranking of European countries don't change dramatically. However to a gentle modification of ratios of weights the results are robust. So it depends mainly on introducing of bounds than on their concrete values. If we want to measure railway efficiency, it would be reasonable to pay attention to good model specification, since results are relatively sensitive to it. But if the efficiency scores serves only as one step of complete analysis in railway industry, choice of model specification could probably not have a significant impact on final conclusions. This fact, however, should be further scrutinized.

Acknowledgements

This work was supported by Student Project Grant MUNI/A/0797/2012.

References

[1] Banker, R. D., Charnes, A and Cooper, W. W.: Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, *Management Science* **30** (1984), 1078–1092.

[2] Charnes, A., Cooper, W. W. and Rhodes, E.: Measuring the efficiency of decision making units, *European Journal of Operational Research* **2** (1978), 429–444.

[3] Cantos, P., Pastor, M. J., and Serrano, L.: Evaluating European railway deregulation using different approaches, *Transport Policy* **24** (2012), 67–72.

[4] Cantos, P., Pastor, M. J., and Serrano, L.: Vertical and Horizontal Separation in the European Railway Sector and its Effects on Productivity, *Journal of Transport Economics and Policy* **44** (2010), 139–160.

[5] Cooper, W. W.: *Data envelopment analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software, Second Edition*. Springer, New York, 2007.

[6] Driessen, G., Lijesen, M.G. and Mulder, M.: *The Impact of Competition on Productive Efficiency in European Railways*. CPB Netherlands Bureau for Economic Policy Analysis, The Hague, 2006.

[7] Thompson, R.G.,Singleton, F.D.,Thrall, R.M. and Smith, A.: Comparative site evaluations for locating a high-energy physics lab, *Texas Interface* **16** (1986), 35–49.

[8] Tone, K.: A slacks-based measure of efficiency in data envelopment analysis, *European Journal of Operational Research* **130** (2001), 498–509.

Responses to Monetary Policy Shock in the Czech Republic

Sára Bisová[1]

Abstract. The concept of VAR models is widely used approach for the econometric analyses of the monetary policy transmission mechanism. Dynamic models are crucial in the monetary policy analyses due to significant lags between an action on one hand and the appropriate effects in the economy on the other hand. This paper estimates the responses of the key macroeconomic indicators to an exogenous monetary policy shock in the Czech economy using the VAR models. We estimate an identified VAR using recursive and non-recursive identification schemes (assuming the small open economy). To avoid the possibility of a price puzzle effect, due to mixing policy regimes in the sample period, we try to use data from single monetary policy regime starting 1998, when the Czech central bank switched to the inflation targeting regime. The model is based on the set of variables concerning output, consumer price index, short term interest rate and exchange rate, all measured at quarterly frequency. The importance is stressed on inclusion of the interest rate and the exchange rate channels as they are fundamental in the inflation targeting transmission mechanism. In the empirical analysis, the obtained results indicate a relatively high sensitivity to the identification scheme in SVAR.

Keywords: Choleski decomposition, identification, impulse response functions, monetary policy, SVAR.

JEL Classification: C32, C51, E52, E58
AMS Classification: 62M10, 91B84

1 Introduction

The concept of VAR models is frequently used for the econometric modeling of the macroeconomic stabilization policies. We estimated three identified VAR models for the analysis of the monetary transmission mechanism using the recursive and non-recursive identification techniques. The advantage of the non-recursive assumption lies in the fact that we can employ presumed structure of the economy and its monetary transmission mechanism. Most studies stress the importance of the interest rate channel and the exchange rate channel, in some papers also the money channel, during the transmission mechanism of monetary policy. We include two crucial transmission channels in the model – the interest rate channel and the exchange rate channel.

Using impulse response functions (IRF) for the real Czech data we quantified the effects of an unexpected monetary policy shock on the analyzed variables, particularly the price level (target variable). The sample period covers the single monetary policy regime starting 1998[2], when the Czech central bank switched to the inflation targeting regime, although earlier data are available. The reason is excluding policy regime switching during the sample period to avoid possible puzzle effects in estimated outputs, due to mixing policy regimes in the sample period. Regime switching VAR models were employed for example in [3][3]. We analyze the sensitivity to the identification scheme in SVAR using three different structures of identification matrix in the same model.

The paper is organized as follows: section 2 states theoretical background of VARs, data are presented in section 3 and empirical analysis in section 4. Conclusions are summarized in section 5.

2 VAR models

The unrestricted VAR approach in the field of macroeconomic analyses was introduced in [11]. Opposed to structural VARs, this concept avoid any a priori assumptions based on short-run or long-run restriction derived from the economic theory.

The structural VAR (or identified VAR) approach allows us to identify the IRF by imposing a priori restrictions on the covariance matrix of the structural errors (a priori short-run restrictions on contemporaneous

[1] University of Economics Prague, Dept. Of Econometrics, W. Churchill Sq. 4, Prague 3, xbiss00@vse.cz.
[2] In 1998 the Czech central bank accepted inflation targeting regime.
[3] The authors specified a regime switching VAR model for the U.S. fiscal policy analysis to allow differences among historical episodes with different economic environments.

effects of shocks and long-run restrictions to employ long-run impact of shocks to identify the IRF). The structural form (SVAR) without intercepts can be defined as follows [6]

$$\mathbf{A}\mathbf{y}_t = \mathbf{\Pi}(L)\mathbf{y}_{t-1} + \mathbf{B}\mathbf{u}_t, \tag{1}$$

where $\mathbf{u}_t$ is a structural disturbance vector generated by vector white noise process, with identity covariance matrix $E(\mathbf{u}_t\mathbf{u}_t^{\mathrm{T}}) = \mathbf{\Lambda}$, where diagonal elements are variances of structural shocks, $\mathbf{\Pi}(L)$ is a polynomial matrix in the lag operator and $\mathbf{y}_t$ is a vector of m endogenous variables. The elements outside the diagonal in matrix $\mathbf{B}$ may be non-zero, therefore, some of the shocks can influence more endogenous variables of the system [6].

We can estimate the **reduced form**

$$\mathbf{y}_t = \mathbf{A}^{-1}\mathbf{\Pi}(L)\mathbf{y}_{t-1} + \mathbf{v}_t, \tag{2}$$

where $\mathbf{v}_t$ is a disturbance vector of reduced form with a covariance matrix $E(\mathbf{v}_t\mathbf{v}_t^{\mathrm{T}}) = \mathbf{\Sigma}$. For the disturbance vectors $\mathbf{u}_t$ and $\mathbf{v}_t$ holds the following formulae [6]

$$\mathbf{v}_t = \mathbf{A}^{-1}\mathbf{B}\mathbf{u}_t \quad \text{or} \quad \mathbf{A}\mathbf{v}_t = \mathbf{B}\mathbf{u}_t \tag{3}$$

and

$$E(\mathbf{v}_t\mathbf{v}_t^{\mathrm{T}}) = \mathbf{A}^{-1}\mathbf{B}^{\mathrm{T}}E(\mathbf{u}_t\mathbf{u}_t^{\mathrm{T}})\mathbf{B}^{\mathrm{T}}\mathbf{A}^{-1} \quad \text{or} \quad \mathbf{\Sigma} = \mathbf{A}^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{\Lambda}\mathbf{B}^{\mathrm{T}}\mathbf{A}^{-1}. \tag{4}$$

The matrix $\mathbf{A}$ identifies the relation between the structural disturbances $\mathbf{u}_t$ and the reduced form shocks $\mathbf{v}_t$. There are many identification techniques based on imposing restriction on the parameters of the system mentioned above.

For structural VAR we employ AB-model, see (3), where $\mathrm{var}(\mathbf{u}_t) = \mathbf{\Lambda}$ is unit matrix. In order to get a just identified system we need $(m^2 - m)/2$ restrictions. Usually the **recursive identification** scheme, so-called **Choleski decomposition**, is used to obtain a just identified VAR, where the matrix $\mathbf{A}$ is lower triangular and the matrix $\mathbf{B}$ is diagonal, see [11]

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{21} & 1 & 0 & \cdots & 0 \\ a_{31} & a_{32} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ a_{m1} & a_{m2} & a_{m3} & \cdots & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_{11} & 0 & 0 & \cdots & 0 \\ 0 & b_{22} & 0 & \cdots & 0 \\ 0 & 0 & b_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & b_{mm} \end{bmatrix}. \tag{5}$$

We can also use a **non-recursive scheme**, where the matrix $\mathbf{A}$ takes different form from (5), see empirical part of this paper in section 3 or e.g. [1], [2], [5], [7], [8], [10]. This approach allows us to employ more assumptions originated in the economic theory about the relations between analyzed variables.

The estimation of VAR models is usually followed by construction of impulse response functions (IRF). IRF are substantial in the macroeconomic policy analyses, especially in case of analyzing and anticipating the fiscal and monetary policy effects. The order of endogenous variables in vector $\mathbf{y}_t$ influences the IRF when using (5). For deriving and interpretation of IRF see [9].

3 Data

For the analysis, four macroeconomic indicators were chosen. The set of endogenous variables includes output (gross domestic product in constant prices of 2005), consumer price index (2005=100), short term interest rate (3M PRIBOR) and exchange rate (Euro/ECU exchange rate in national currency), all measured at quarterly frequency. Similar to [2], [5], [10] we include one exogenous variable – world commodity price index. The reason of considering the world or foreign variables lies in the need to avoid a misinterpretation of domestic monetary shock in case of reaction to a foreign shock[4]. To reduce the possibility of puzzle effects, in response to mixing policy regimes in the sample period[5], we use the data from single monetary policy regime starting 1998, as men-

[4] For example oil shocks, foreign policy shocks etc. [5]
[5] Observed for example in [2].

tioned above. The sample period considered for the analysis consists of Q1 1998 – Q4 2012 (60 observations). Data series were obtained from Eurostat[6] (interest rate, exchange rate), CSO[7] (CPI and GDP) and The World Bank[8] (commodity price index). The abbreviations for variables used in following tables and graphs are *GDP* for gross domestic product, *CPI* for consumer price index, *IR* for interest rate, *ER* for exchange rate and *COM* for commodity price index.

GDP and CPI were seasonally adjusted by X12 ARIMA method in order to reduce the number of parameters to be estimated in the model. For the estimation EViews 7 and GRETL 1.9.5 software were used.

Table 1 shows the ADF tests of all the above mentioned variables in levels. All the variables are non-stationary and therefore should be transformed to reach the stationarity for the application in VAR. In order to eliminate the non-stationarity, differences between logarithms[9] were computed (for their useful interpretation as growth rates)

$$\Delta \ln(y_t) = \ln y_t - \ln y_{t-1} = \ln \frac{y_t}{y_{t-1}}, \tag{6}$$

where y_t represents each of the analysed variables in period *t*. All the time series are stationary in first differences – see table 1.

LEVEL	t-statistic	Prob.*	D-LOG	t-statistic	Prob.*
GDP	-1.083167	0.7170	***GDP***	-5.192390	0.0001
CPI	-0.485191	0.8864	***CPI***	-4.590175	0.0004
IR	-2.154578	0.2248	***IR***	-4.068376	0.0022
ER	-1.381421	0.5855	***ER***	-6.496229	0.0000
COM	-0.754072	0.8241	***COM***	-6.255058	0.0000

*MacKinnon (1996) one-sided p-values
Notes: Test critical values -3.54 (1% level), -2.91 (5% level), -2.59 (10% level), IR not in logs

Table 1 ADF tests for variables in levels

Figure 1 shows graphs of the time series in levels (*GDP* and *CPI* are seasonally adjusted as mentioned above). We can clearly see the non-stationarity of all variables in their level form.

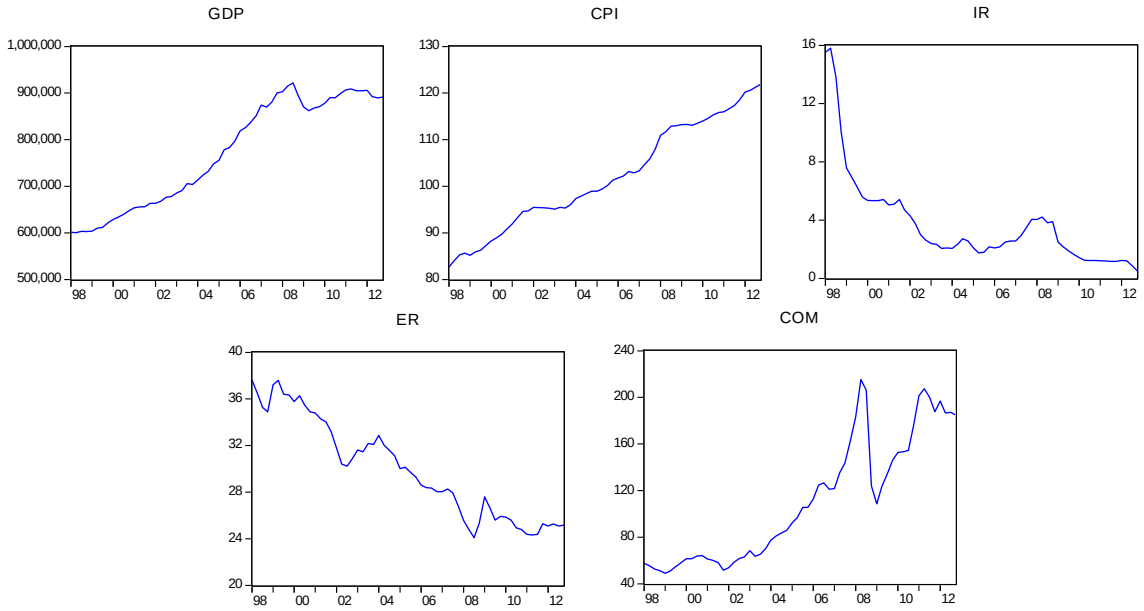


Figure 1 Time series plots

[6] http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database.
[7] Czech statistical office - www.czso.cz.
[8] http://econ.worldbank.org.
[9] Except for interest rate which is in percentage form and therefore transformed directly into first differences.

4 Application

Many studies using VAR methodology for analyzing monetary policies do not include money stock in the models, see [5], [7] and [10][10]. Oppositely, a number of studies use money supply in the vector of endogenous variables, see [1], [2], [8]. In this study we construct the four-variable VAR excluding the money stock. The extended five-variable model is the subject of upcoming studies.

The four-variable SVAR was estimated applying the recursive and non-recursive identification schemes. The non-recursive scheme takes the following form

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ a_{21} & 1 & 0 & 0 \\ 0 & a_{32} & 1 & a_{34} \\ a_{41} & a_{42} & a_{43} & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_{11} & 0 & 0 & 0 \\ 0 & b_{22} & 0 & 0 \\ 0 & 0 & b_{33} & 0 \\ 0 & 0 & 0 & b_{44} \end{bmatrix}, \tag{7}$$

where parameters a_{ij} are unconstrained and the vector of endogenous variables is transformed, as mentioned in section 3

$$\mathbf{y}_t = (gdp_t, cpi_t, ir_t, er_t).^{11} \tag{8}$$

We alternatively use the form (7) with the modification used in [5], where $a_{32} = 0$ instead of $a_{31} = 0$.

The non-recursive identification scheme (7) assumes simultaneity between the interest rate and the exchange rate[12] and that the monetary authority considers contemporaneous prices rather than output, when making monetary policy decisions concerning stimulation of interest rates in the inflation targeting regime. In [5] authors hold the assumption of the forward-looking monetary authority, which considers product rather than prices ($a_{32} = 0$ instead of $a_{31} = 0$). In this study, we compare the three models described above. We use world commodity price index as an exogenous variable and intercepts in all models. The recursive identification scheme uses the same vector of endogenous variables (8) and the structure (5).

Firstly we analyze the lag length in the VAR model. The calculated lag-length criteria for the four-variable VAR are listed in table 2.

LAG	LogL	LR	FPE	AIC	SC	HQ
0	493.6188	NA	2.52e-13	-17.65887	-17.36689	-17.54596
1	526.1082	57.89029	1.39e-13	-18.25848	-17.38255*	-17.91975*
2	545.6839	32.03295*	1.23e-13	-18.38851	-16.92863	-17.82396
3	562.8328	25.56738	1.22e-13*	-18.43028*	-16.38645	-17.63992
4	573.4610	14.29979	1.56e-13	-18.23495	-15.60716	-17.21876

*indicates lag order selected by the criterion
Notes: LR: sequential modified LR test statistic (each test at 5% level), FPE: Final prediction error, AIC: Akaike information criterion, SC: Schwarz information criterion, HQ: Hannan-Quinn information criterion.

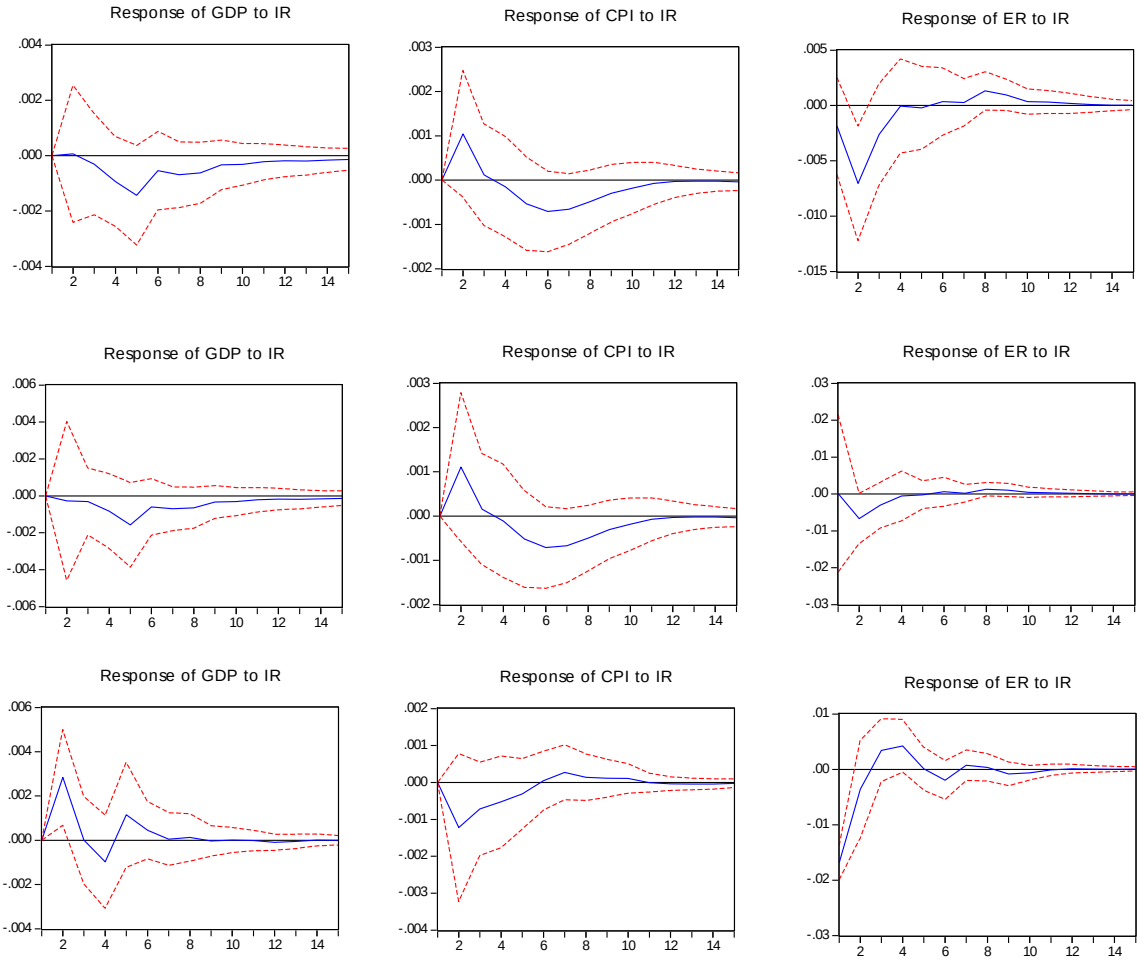
Table 2 VAR lag order selection criteria

The VAR(3) model was estimated as the AIC recommends (the inclusion of dynamics is desirable - there is a significant delay between the initial stimulus of the central bank and the change in the target variables in the monetary transmission mechanism – between 12 and 18 months. For the purpose of estimation, the reduced form and the OLS estimator were applied.[13]

Figure 2 shows selected responses to positive exogenous shock in the interest rate (monetary restriction) using the recursive identification scheme (the first row), the non-recursive identification scheme described in (7) (the second row) and the modification of (7) replacing $a_{32} = 0$ for $a_{31} = 0$ (the third row). In the columns are given responses of product, prices and exchange rate, respectively, to one standard deviation monetary policy

[10] The reason is usually a weakening effect of the money supply in the monetary transmission.
[11] Output, price level, interest rate and exchange rate, respectively.
[12] See for example [2], [5].
[13] The diagnostic control of the estimated model suggests the presence of heteroskedasticity of the error terms (joint test). Therefore, the robust standard errors were computed.

shock (domestic). Our assumptions about responses to an exogenous monetary contraction are following: the price level declines, the output does not increase and the exchange rate appreciates under the flexible exchange rate regime[14].



Notes: the recursive identification scheme in the first row, the non-recursive schemes in the rows 2 and 3.

Figure 2 IRF, four-variable VAR(3)

Using the first and second benchmark models, the results suggest similar responses to an exogenous monetary policy shock. On the other hand, the third version of the identification provides different behavior.

In the first two models, the contractionary monetary policy shock is followed by an immediate exchange rate appreciation with a slight depreciation after around two years in the first model – one possible explanation can lead to delayed overshooting puzzle[15], on the other hand, the confidence intervals are relative wide and therefore bring some uncertainty to our interpretation. The prices start to fall after about 3 quarters of persistence in response to the monetary policy shock with a bottom after approximately 6 quarters that is consistent with the assumed targeting horizon between 12 and 18 months. The output falls in response to monetary policy tightening with the bottom after about one year - it is in line with our assumptions.

In contrary to the first and second model, the third benchmark model shows different results. The interpretations of output and exchange rate are somewhat difficult while changing sings of responses over the horizon. Nevertheless, in the case of exchange rate there is an immediate appreciation after the exogenous positive shock in interest rate. The counterintuitive results in responses of output can be observed over the longer term horizon. The prices are falling, except for a slight increase in period between 8 to 10 quarters. The minimum is reached in about 2 quarters.

[14] While almost no theory predicts an increase in output and prices in response to an exogenous contractionary monetary policy shock, there is a number of theories concerning the exchange rate behavior, see [8].
[15] See for example [8].

The first two models exhibit similar results corresponding with our assumptions and with most studies providing similar analyses (e.g. [2], [5]). We can conclude that the responses are close to expected movements of analyzed variables. The third model provides different results, less consistent with our assumptions and difficult to interpret in case of some responses. We do not get similar results to [5] - one possible reason consists in using different definition of product – the authors use output gap instead of GDP growth (ex-post data), see [5]. We can conclude that the IRF are very sensitive to the structure of matrix **A** which identifies the relation between the structural disturbances and the reduced form residuals.

5 Conclusion

Using the structural VAR model, we analyzed the transmission mechanism of monetary policy in the Czech economy. We estimated three alternative four-variable VAR(3) models, imposing zero restrictions on some contemporaneous structural parameters. In the first model we employed just-identified VAR using the usual recursive identification scheme. In the second model we assumed the authority which considers contemporaneous prices when making monetary policy decisions about interest rates and in the third model we considered output instead of prices. The second and third models also assume simultaneous relation between the interest rate and the exchange rate. Compared to our previous analyses (e.g. [4]), and provided results, we used data period from Q1 1998 to Q2 2013 to eliminate puzzle effects due to mixing policy regimes, although earlier data are available. We also included one exogenous variable in the model not to misinterpret domestic monetary shock with the reaction to foreign shocks.

Considering the analyzed models, the second benchmark model with the assumed structure of the economy is preferred. In this model, an unexpected rise in interest rate (a contractionary monetary policy shock) implies a temporary fall in output that peaks around 5 quarters after the shock, a decrease in prices, although prices tend to be more persistent, and exchange rate appreciation. The third model does not give us persuasive results - the wave-shaped responses are difficult to interpret. In all the models, confidence intervals are relatively wide. There is important to note, that our model concentrates only on a basic part of the real system. The extended models - e.g. a five-variable SVAR and a regime switching VAR - are subjects of upcoming analyses.

Acknowledgements

The research project was supported by Grant No. IGA F4/1/2012, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Anzuini, A., Levy., A.: Monetary policy shocks in the new EU members: a VAR approach. *Applied Economics* **39** (2007), 1147-1161.
[2] Arnoštová, K., Hurník, J.: The Monetary Transmission Mechanism in the Czech Republic (Evidence from VAR Analysis). *CNB Working Paper* **4** (2005).
[3] Auerbach, A. J. and Gorodnichenko, Y.: Measuring the Output Responses to Fiscal Policy. *American Economic Journal: Economic Policy*, 2010, Vol. 4, No. 2, 1-27.
[4] Bisová, S.: Analýza monetární politiky pomocí modelů VAR – aplikace na českou ekonomiku. *Vědecká konference doktorského studia FIS 2013.* Oeconomica, Praha, 2013, 106–117.
[5] Borys, M. M., Horváth, R.: The effects of monetary policy in the Crech republic: An empirical study. *William Davidson Institute Working Paper* **922** (2008).
[6] Hušek, R.: *Ekonometrická analýza.* Oeconomica, Praha, 2007.
[7] Jarociński, M.: Responses to Monetary Policy Shocks in the East and the West of Europe: A Comparison. *Barcelona and CASE (Center for Social and Economic Research)*. Universitat Pompeu Fabra, Warsaw, 2004.
[8] Kim, S.: Monetary policy, foreign exchange policy and delayed overshooting. *Journal of money, credit and banking*, 2005, Vol. 37, No. 4, 775-782.
[9] Lütkepohl, H.: *New introduction to multiple time series analysis*. Springer, 2005.
[10] Mojon, B., Peersman, G.: A VAR description of the effects of monetary policy in the individual countries of the EURO area. *ECB Working Paper* **92** (2001).
[11] Sims, Ch. A.: Macroeconomics and reality. *Econometrica*, 1980, Vol. 48, No. 1, 1-48.

Input-output interactions in a DSGE framework

Sára Bisová[1], Eva Javorská[2], Kristýna Vltavská[3], Jan Zouhar[4]

Abstract. Despite their (heavily advertised) microeconomic foundations, DSGE models are usually implemented for economic analyses using country-level aggregate data. However, national statistical offices of the EU members create national accounts that contain detailed information about individual industries, such as the Input-Output tables. In this paper, we present a preliminary study of a multi-sector extension of a canonical RBC model that allows to use data on the input-output structure with multiple sectors. We formulate a simple baseline model that allows for an arbitrary number of sectors with an arbitrary input-output structure. The practical obstacle to using the model in practice lies in the need to find approximate steady-state values of the variables. In the general case, finding a solution to the steady-state problem is diffucult; however, we do provide an analytic solution to a special symmetric case, which can sometimes provide a close enough approximation for non-symmetric problems as well.

Keywords: DSGE models, Input-Output tables, National Accounts

JEL classification: E16, E17
AMS classification: 91B51

1 Introduction

In today's macroeconomic analyses, Vector Autoregressions (VAR), Vector Error Correction models (VEC) or DSGE models are typically used, with DSGE models being the newest of the three approaches, and one that is becoming increasingly popular with researchers and regulators alike. For instance, the Czech National Bank has developed a DSGE model "g3" [1], which is used for both the interpretation of current economic phenomena and for economic forecasting. A key feature that makes DSGE models so favourable is the fact that they are built on microeconomic foundations, and thus should be resistant to the Lucas critique. In our opinion, however, the idea of DSGE models' microfoundations is at odds with the aggregate nature of data that are typically used for model calibration; country-level data are almost invariably used, even though industry level-data are available in all EU member states' national accounts and presented e.g. in the Input-Output tables [4].

It has to be noted that some of the existing models do in fact divide the *firms* sector into a small number of "industries", typically distinguishing either between *manufacturers* and *retailers*, or, as in case of the Hubert model that has been developed and used by the Ministry of Finance of the Czech Republic (see [5], [6]), between *manufacturers*, *retailers*, and *importers*. Nevertheless, in such settings, (i) the model comprises a very small number of industries, (ii) the distinction between the individual industries does not follow the structure of national accounts, and (iii) the input-output structure is limited to a one-sided relationship between retailers and manufacturers. The aim of this paper is to formulate a simple baseline model that allows for an arbitrary number of sectors with an arbitrary input-output structure.

To the best of our knowledge, the only existing studies that develop a similar type of a model are the works of Bouakez et al [3], [2]. Even though we were undoubtedly inspired by these works, our model differs from that of Boukaez et al in several aspects. Firstly, their model is built to utilize the structure of the U.S. national accounts – which differs from that in both the Czech Republic and the EU as a whole; most importantly, the U.S. national accounts collect data on the capital flows among

[1]Univ. of Economics, Prague, Dept. of Econometrics, nám. W. Churchilla 4, 130 67 Praha 3, xbiss00@vse.cz
[2]Univ. of Economics, Prague, Dept. of Economic Statistics, nám. W. Churchilla 4, 130 67 Praha 3, xjave00@vse.cz
[3]Univ. of Economics, Prague, Dept. of Economic Statistics, nám. W. Churchilla 4, 130 67 Praha 3, kristyna.vltavska@vse.cz
[4]Univ. of Economics, Prague, Dept. of Econometrics, nám. W. Churchilla 4, 130 67 Praha 3, honza.zouhar@gmail.com

the individual sectors (the Capital Flow Tables), which allow for a simple incorporation of the capital formation proces into the DSGE model. As we aim to apply our model to Czech data in future, we had to resort to a different, less straightforward method of combining the input-output processes and capital stock creation, which proved to complicate the equilibrium derivation. Secondly, while the models of Bouakez et al take on the New Keynesian perspective, ours is mostly in line with the real business cycle (RBC) literature.

2 The model

The model economy consists of a single infinitely-lived household and J firms, representing one sector each. The household is the sole labour supplier to all firms, and the firm's production can be either consumed by the household, used as a material input in other sectors, or used in the form of a capital investment. In our treatment, we use the social-planner formulation of the model, which should give identical equilibrium conditions as the competitive equilibrium formulation, but with a lower notational burden, see [7]. The social planner's aim is to maximize the expected sum of households (discounted) utilities, i.e. to maximize

$$\mathbb{E}\sum_{t=0}^{\infty}\beta^t U(C_t, H_t) = \mathbb{E}\sum_{t=0}^{\infty}\beta^t\big[\log C_t - \phi\log H_t\big],$$

where U is the (instantaneous) utility function, C_t and H_t are the consumption and labour-time bundle in period t and ϕ is a non-negative parameter (labour disutility). The consumption bundle is defined as a weighted Dixit-Stiglitz aggregate of sector-specific consumptions:

$$C_t \equiv \Big[\textstyle\sum_j \gamma_j C_{jt}^{\theta}\Big]^{1/\theta},$$

where C_{jt} is the quantity consumed of goods from sector j in period t, γ_j is the relative weight of the goods from sector j in the consumption bundle (γ_j are nonnegative and sum to one), and θ is a parameter greater than 1, such that $\theta/(\theta-1)$ is the elasticity of substition among the sectors' production.

Similarly, the labour time bundle is defined as

$$H_t \equiv \Big[\textstyle\sum_j H_{jt}^{\upsilon}\Big]^{1/\upsilon},$$

where H_{jt} are hours worked in sector j and period t and υ is a parameter less than 1 representing the substition among hours spent working in individual sectors. Note that in this setting the household prefers diversity in labour time.

The production function in sector j has the form of a Cobb-Douglas function with constant returns to scale:

$$Y_{jt} = (e^{Z_t}H_{jt})^{\alpha_{jH}}K_{j,t-1}^{\alpha_{jK}}M_{jt}^{\alpha_{jM}},$$

where Z_t is an exogenous aggregate labour productivity shock, $K_{j,t-1}$ is the capital stock in sector j at the beginning of period t (i.e. capital stock determined by the end of period $t-1$), M_{jt} is a bundle of all material flows to sector j from other sectors, and the α's are sector-specific output elasticities that sum to one. The material bundle is defined as

$$M_{jt} \equiv \Big[\textstyle\sum_i \mu_{ij}M_{ijt}^{\theta}\Big]^{1/\theta},$$

where M_{ijt} is the quantity of sector i's output being supplied (as an input) to sector j, and μ_{ij} is the weight of sector i's output in sector j's material (input) bundle. The labor productivity shock is assumed to follow the standard AR(1) process,

$$Z_t = \rho Z_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0,\sigma^2),\ \text{i.i.d.} \tag{1}$$

The budget constraint is given as

$$Y_{jt} = C_{jt} + I_{jt} + \textstyle\sum_k M_{jkt},$$

where I_{jt} denotes capital investment in sector j and period t defined as

$$I_{jt} \equiv K_{jt} - (1-\delta)K_{j,t-1},$$

where δ is the depreciation rate.

Comments on alternative model specifications and calibration. The original sector-specific data (and, most notably, Input-Output tables) presented by the Czech Statistical Office are divided into 81 industries. Technically, such a number of sectors would probably prove too high for the resulting model to be computationally tractable. Therefore, it seems necessary to aggregate these industries in order to reduce the number of sectors to an acceptable level. Following the CZ-NACE classification, the 81 industries can be aggregated into as few as 19 sectors. As for calibration, the μ_{ij} parameters are supposed to be taken directly from Input-Output tables, and the γ_j parameters will be calibrated from the consumer price index. Moreover, we need the parameters of the Cobb-Douglas production function; we intend to use estimates based on the index number approach. The θ parameter can be estimated as the ratio of gross value added in current prices and compensation of employees. These data are published in the national accounts.

3 Solving the model

3.1 Equilibrium conditions.

First, we form the Lagrangian as[2]

$$\begin{aligned} L &= \mathbb{E}\sum_{t=0}^{\infty}\beta^t\Big[\log C_t - \phi\log H_t - \textstyle\sum_j \lambda_{jt}\Big(-Y_{jt} + C_{jt} + I_{jt} + \sum_k M_{jkt}\Big)\Big] \\ &= \mathbb{E}\sum_{t=0}^{\infty}\beta^t\Big[\log C_t - \phi\log H_t + \textstyle\sum_j \lambda_{jt}\Big(-Y_{jt} + C_{jt} + K_{jt} - (1-\delta)K_{j,t-1}\Big) - \sum_i\sum_j \lambda_{it}M_{ijt}\Big]. \end{aligned}$$

The first-order conditions are obtained by setting the partial derivatives of the Lagrangian (w.r.t. the decision variables and Lagrange multipliers) to zero. After the elimination of Lagrange multipliers and some obvious algebraic manipulation, we can list the equilibrium conditions as follows:

$$\begin{aligned} \alpha_{jH}\,\gamma_j\,Y_{jt}\,C_{jt}^{\theta-2}\,C_t^{-\theta} &= \phi\,(H_{jt}/H_t)^{\upsilon} && \text{for all } j, \\ \gamma_i\,C_{it}^{\theta-2}\,M_{ijt}^{1-\theta} &= \gamma_j\,\alpha_{jM}\,\mu_{ij}\,Y_{jt}\,C_{jt}^{\theta-2}\,M_{jt}^{-\theta} && \text{for all } i,j, \\ C_{jt}^{\theta-2}C_t^{-\theta} &= \beta\,\mathbb{E}_t\Big[C_{j,t+1}^{\theta-2}\,C_{t+1}^{-\theta}\Big(\alpha_{jK}(Y_{j,t+1}/K_{jt}) + 1 - \delta\Big)\Big] && \text{for all } j, \\ Y_{jt} &= (e^{Z_t}\,H_{jt})^{\alpha_{jH}}\,K_{j,t-1}^{\alpha_{jK}}\,M_{jt}^{\alpha_{jM}} && \text{for all } j, \\ Y_{jt} &= C_{jt} + I_{jt} + \textstyle\sum_k M_{jkt} && \text{for all } j, \\ Z_t &= \rho Z_{t-1} + \varepsilon_t, \\ \varepsilon_t &\sim \mathcal{N}(0,\sigma^2),\ \text{i.i.d.} \end{aligned}$$

Altogether, this gives $(J^2 + 4J + 1)$ equations. Note that if one wants to use a software package like Dynare, symbolic expressions containing loops and sums are not supported in the code. Even for J as low as 3, writing down these equations in hand is a very tedious task, and typing errors are likely to creep into the code. For these reasons, we put up a Matlab script that creates the appropriate Dynare code for an arbitrary number of sectors; we will gladly provide the Matlab script to anyone interested at an email request.

3.2 Steady state conditions

In the analysis of equilibrium dynamics, one has to find the steady state first. The steady state conditions are easily obtained from the equilibrium conditions by dropping the time subscript (t) and the error term

[2] We deliberately omitted the constraint (1) and its Lagrange multiplier in order to save space; note that Z_t is an exogenous random process, so it does not enter any first-order conditions w.r.t. the decision variables. The equilibrium conditions given below, however, do include the constraint on Z_t.

(ε_t). After a slight algebraic manipulation[3] we obtain a system of non-linear equations

$$\alpha_{jH}\,\gamma_j\,Y_j\,C_j^{\theta-2}\,C^{-\theta} = \phi\,(H_j/H_t)^{\upsilon} \qquad \text{for all } j, \tag{2}$$

$$\gamma_i\,C_i^{\theta-2}\,M_{ij}^{1-\theta} = \gamma_j\,\alpha_{jM}\,\mu_{ij}\,Y_j\,C_j^{\theta-2}\,M_j^{-\theta} \qquad \text{for all } i,j, \tag{3}$$

$$1 = \beta\Big(\alpha_{jK}(Y_j/K_j) + 1 - \delta\Big) \qquad \text{for all } j, \tag{4}$$

$$Y_j = H_j^{\alpha_{jH}}\,K_j^{\alpha_{jK}}\,M_j^{\alpha_{jM}} \qquad \text{for all } j, \tag{5}$$

$$Y_j = C_j + I_j + \textstyle\sum_k M_{jk} \qquad \text{for all } j. \tag{6}$$

3.3 Steady state – symmetric case

Consider now a special symmetric case where for all j, $\alpha_jK = \alpha_K$, $\alpha_jH = \alpha_H$, $\alpha_jM = \alpha_M$, $\gamma_j = \gamma$, and $\mu_{ij} = \mu$ for all $i \neq j$, $\mu_{ij} = 0$ otherwise. In this case, the analytic steady-state solution has the form given in the proposition below.

Proposition 1. *Let*

$$R = \frac{\alpha_K}{1/\beta - 1 + \delta}, \quad S = \mu^{1/\theta}\alpha_M(J-1)^{(1-\theta)/\theta}, \quad T = R^{-\frac{\alpha_K}{\alpha_H}}S^{-\frac{\alpha_M}{\alpha_H}}.$$

Then, in the steady state of the symmetric problem, for all j,

$$Y_j = \frac{\phi}{\alpha_H}C_j^2\,, \tag{7}$$

$$M_{ij} = \frac{\alpha_M}{J-1}Y_j\,,$$

$$K_j = R\,Y_j\,, \tag{8}$$

$$C_j = \frac{\alpha_H}{\phi(1-\delta R-\alpha_M)}\,, \tag{9}$$

$$H_j = S\,Y_j. \tag{10}$$

Proof. It is useful to derive the following identities first:

$$H^{\upsilon} = \textstyle\sum_j H_j^{\upsilon} = JH_j^{\upsilon}\,, \quad \text{and so} \qquad (H_j/H)^{\upsilon} = 1/J\,, \tag{11}$$

$$C^{\theta} = \textstyle\sum_j \gamma C_j^{\theta} = \gamma J C_j^{\theta}\,, \quad \text{and so} \qquad (C_j/C)^{\theta} = 1/(\gamma J)\,, \tag{12}$$

$$M_j^{\theta} = \textstyle\sum_{i\neq j}\mu M_{ij}^{\theta} = \mu(J-1)M_{ij}^{\theta}\,, \quad \text{and so} \qquad (M_{ij}/M_j)^{\theta} = 1/[\mu(J-1)]\,, \tag{13}$$

the last equation holds for all $i \neq j$. Now, plugging (11) and (12) into (2) immediately yields $\alpha_H Y_j = \phi C_j^2$, a simple rearrangement of (7). In (3), the terms γ_i, γ_j and C_i, C_j can be cancelled due to symmetry, and plugging in (13) gives $M_{ij} = \frac{\alpha_M}{J-1}Y_j$. Note that summing this formula across the second subscript yields

$$\textstyle\sum_k M_{jk} = \sum_{k\neq j}\frac{\alpha_M}{J-1}Y_k = \alpha_M Y_j\,. \tag{14}$$

The equation $K_j = RY_j$ follows immediately from (4). Next, note that

$$M_j = \Big[\sum_i \mu M_{ij}^{\theta}\Big]^{1/\theta} = \Big[\mu(J-1)\left(\frac{\alpha_M}{J-1}Y_j\right)^{\theta}\Big]^{1/\theta} = S\,Y_j.$$

The equation (9) can be obtained by plugging (7), (8) and (14) into (6). From (5) and the previous results, we have

$$Y_j = H_j^{\alpha_H}(R\,Y_j)^{\alpha_K}(S\,Y_j)^{\alpha_M},$$

which gives (10), and completes the proof. □

It is easily seen that the equations in Proposition 1 follow a recursive scheme in the decision variables – C_j is expressed in terms of the model parameters only, Y_j is in terms of C_j, and the rest is given in terms of Y_j.

[3] To be specific, we cancelled equal terms on both sides of the third equation, and noted that in the equilibrium, $Z = \rho Z$, and hence $Z = 0$.

4 Conclusions

As we pointed out in the introduction, this paper was primarily meant to show that DSGE models can be formulated to work with a disaggregated structure of the economy, and start the academic discussion that will hopefully lead to creating DSGE models that explain some of the real-life phenomena that were not captured by the extant models. We carried out first steps in this direction by devising a simple RBC-based DSGE model that can accomodate an arbitrary number of sectors, the interactions of which are described by the input-output tables produced by the national accounts along the accepted EU accounting standards. In chapter 3, we derived the equilibrium conditions that can be e.g. entered into the widely-used Dynare software package.

However, even in this simple setting, there are some practical obstacles that need to be tackled before the model can be used with real-life data. Concretely, efficient numeric solutions of the steady state are required. As of yet, we have only obtained some preliminary results, namely the characterization of the steady state in the special case of a symmetric problem, described in Proposition 1. In future, we aim to test to what extent can these results be used to find initial guesses for the numeric algorithms that solve the problem in the general case.

Other goals for future reseach include calibrating the model on Czech data and extending the baseline models to a more elaborate setting; these extensions might be as follows:

- more sector-specific parameters (such as the depreciation rate, technology shocks etc.),
- the introduction of the government and foreign sectors,
- explicit treatment of the labour market and its imperfections,
- incorporating possible price, labour, and capital adjustment costs.

Acknowledgements

This research was supported by the Internal Grant Agency of the University of Economics, Prague, project No. 18/2013, "The usage of DSGE models in national accounts".

References

[1] Ambriško, R., Babecký, J., Ryšánek, J., Valenta, V.: *Assesing the Impact of Fiscal Measures on the Czech Economy.* CNB Working paper series, 15/2012.

[2] Bouakez, H., Cardia, E., Ruge-Murcia, F. J..: *Sectoral Price Rigidity and Aggregate Dynamics.* CIRPEE Working Paper 09-06, 2009.

[3] Bouakez, H., Cardia, E., Ruge-Murcia, F. J..: *The Transmission of Monetary Policy in a Multi-Sector Economy.* CIREQ Working Paper 20-2005, 2005.

[4] EUROSTAT: *European System of Accounts (ESA 1995).* Eurostat, Luxemburg, 1996.

[5] Štork, Z., Závacká, J., Vávra, M.: *Hubert: DSGE Model of the Czech Economy. Working Paper 2/2009, Ministry of Finance of the Czech Republic.*

[6] Štork, Z.: *A DSGE Model of the Czech Economy: a Ministry of Finance Approach.* Ministry of Finance of the Czech Republic 4/2011.

[7] Uhlig, H.: *A Toolkit for Analyzing Nonlinear Dynamic Stochastic Model Easily.* CentER, 1995.

Fuzzy multiple criteria evaluation method

Adam Borovička[1]

Abstract. The article deals with new a proposed fuzzy multiple criteria evaluation method. This approach enables to include some uncertainties in a decision making process. Criteria values and weights of criteria may be expressed in a vague, uncertain form. A decision maker sets an importance of criteria linguistically. Some criteria values can be stochastic, random variables with some probability distribution, respectively. The criteria values and weights are expressed as triangular fuzzy numbers. The method requires the ranking of fuzzy numbers which is made via modified McCohone's approach. The algorithm is computationally namely based on preference relations among alternatives and also differences in (non)fuzzy evaluations of alternatives. The proposed multiple attribute decision making method is able to choose primarily one effective alternative. Finally, the approach is clearly applied in the investment decision making process in the capital market with open shares funds.

Keywords: fuzzy number, membership function, uncertainty.

JEL Classification: C44, G11
AMS Classification: 90-08, 91B28

1 Introduction

Many uncertainties can occur in a decision making process, even stochastic character of criteria values, vague preferences of a decision maker etc. To make the decision making procedure more real these factors should be included in used methodical approaches.

When we want to choose some alternative from the set some multiple criteria evaluation method can be applied. There are many methods which offer full ranking of alternatives and mostly are not able to involve some uncertain, vague input information. But we are trying to choose only one effective alternative and mainly take into account uncertainties in the decision making process.

The proposed fuzzy method uses the basic concepts of fuzzy set theory in order to express the uncertainties quantitatively. It evaluates the alternatives on the basis of preference relations between alternatives constructed via some operations with fuzzy weights of criteria and also makes provision for the differences in criterial values, the distance from the ideal alternative is computed.

The reason of the proposal of the fuzzy multi-criteria evaluation method is the particular decision making situation. It is a choice of one open shares funds from each set in the capital market. For this purpose, we projected the method which could include the stochastic elements and vague information from the decision maker. Further, we focused on its computation complexity to be able to apply it in practice without bigger obstructions.

Of course, some illustrative investment practical application of this method is included at the end of the article.

2 Fuzzy multiple attribute decision making method

The proposed method uses fuzzy numbers in order to express the uncertainties. The approach evaluates the alternatives by means of preference relations and also takes into account differences in the criteria values. Modified McCahone's approach is employed for ranking of fuzzy valuations and the modified concept of Hamming distance is applied to compute distance between fuzzy numbers. The introduced method does not offer a full ranking of all alternatives, but chooses only one effective.

Fuzzy multiple criteria evaluation methods will be described in terms of several following steps.

Step 1: Given the matrix $Y = (y_{ij})$, where y_{ij} $(i = 1,2,...,p;\ j = 1,2,...,k)$ represents valuation of *i-th* alternative by *j-th* criterion. Some of them can be stochastic, the criteria values are random variables with some probability distribution, respectively. It will be approximated by normal probability distribution $N(\mu, \sigma^2)$.

[1] University of Economics, Department of Econometrics, W. Churchill Sq. 4, 130 67 Prague 3, adam.borovicka@vse.cz.

Then the vague valuation is expressed as triangular fuzzy number[2], or fuzzy number with triangular membership function as follows

$$T_X = (\mu - 3\sigma, \mu, \mu + 3\sigma),$$

where μ is a mean and σ denotes a standard deviation of values of random variables *X*. This function covers 99,7 % of the set. To make the problem easier the nonlinear probability density function is replaced by linear form in parts as a triangular probability density function. This simplification is clearly displayed in the following figure (Figure 1).

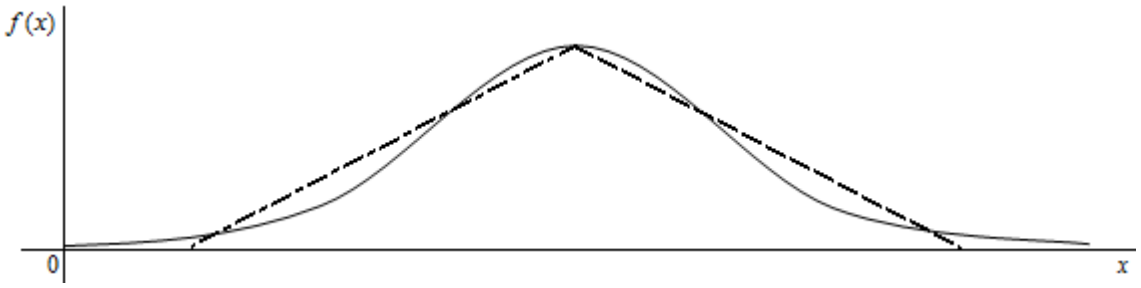


Figure 1 Comparison of two probability density functions

If the criteria values are not vague, uncertain, they can also be expressed as triangular fuzzy numbers with the same value of all three elements as follows

$$T_A = (a, a, a),$$

where *a* is the certain criteria value. Or as mentioned below, this transformation to fuzzy number is not necessary.

An importance of particular criteria is stated via weights formulated as triangular fuzzy numbers. The fuzzy numbers can be set on the basis of linguistic references about criteria importance from a decision maker. In other words, concrete value from interval $\langle 0,1 \rangle$ with some tolerance is assigned to each criterion (see more [2]).

Step 2: In the second step, we must find out definite ranking for all alternatives according to each criterion. This is a little problem in the case of fuzzy numbers. For this procedure, the McCahone's approach is applied with some smaller modifications and improvements.

Modified McCahone's approach

This approach compares the fuzzy numbers via the fuzzy maximum and fuzzy minimum [8]. Given *n* fuzzy numbers $F_1, F_2, ..., F_n$. The fuzzy max, or the membership function of fuzzy max can be formulated as

$$\mu_{\max}(x) = \sup_{x = x_1 \vee x_2 \vee ... \vee x_n} [\mu_{F_1}(x_1) \wedge \mu_{F_2}(x_2) \wedge ... \wedge \mu_{F_n}(x_n)] \quad \forall x, x_1, x_2, ..., x_n,$$

where $\mu_{F_1}(x_1), \mu_{F_2}(x_2), ..., \mu_{F_n}(x_n)$ are the membership functions of fuzzy numbers $F_1, F_2, ..., F_n$. We can analogously specify fuzzy min as

$$\mu_{\min}(x) = \sup_{x = x_1 \wedge x_2 \wedge ... \wedge x_n} [\mu_{F_1}(x_1) \wedge \mu_{F_2}(x_2) \wedge ... \wedge \mu_{F_n}(x_n)] \quad \forall x, x_1, x_2, ..., x_n.$$

The fuzzy max and fuzzy min are depicted in the following two graphs (Figure 2).

[2] *Triangular fuzzy number* is a convex fuzzy set with triangular shape of membership function (see more [4] or [9]). *Fuzzy set* is such a set whose elements are included with certain grade of membership (see more [3] or [10]). *Membership function* measures the grade of set membership. It takes the value from interval $\langle 0,1 \rangle$. The higher value denotes the higher degree of set membership [5].

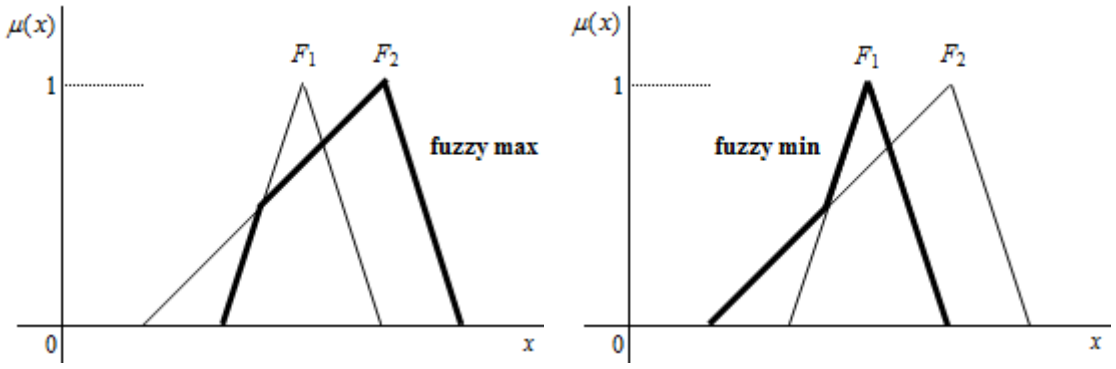


Figure 2 The fuzzy max a fuzzy min

Now we compute the contribution of fuzzy number F_i $(i = 1, 2, ..., n)$ toward the fuzzy max by the following formula

$$P(F_i) = \frac{\int_{S(F_i)} [\mu_{\max}(x) \wedge \mu_{F_i}(x)]\, dx}{\int_{S(F_i)} \mu_{F_i}(x)\, dx},$$

where $S(F_i)$ is the definition scope of *i-th* fuzzy number F_i.

Similarly, the contribution of fuzzy number F_i $(i = 1, 2, ..., n)$ toward the fuzzy min is defined by

$$N(F_i) = \frac{\int_{S(F_i)} [\mu_{\min}(x) \wedge \mu_{F_i}(x)]\, dx}{\int_{S(F_i)} \mu_{F_i}(x)\, dx}.$$

The contribution of fuzzy number F_i toward to fuzzy max, or fuzzy min is displayed by the following two graphs (Figure 3).

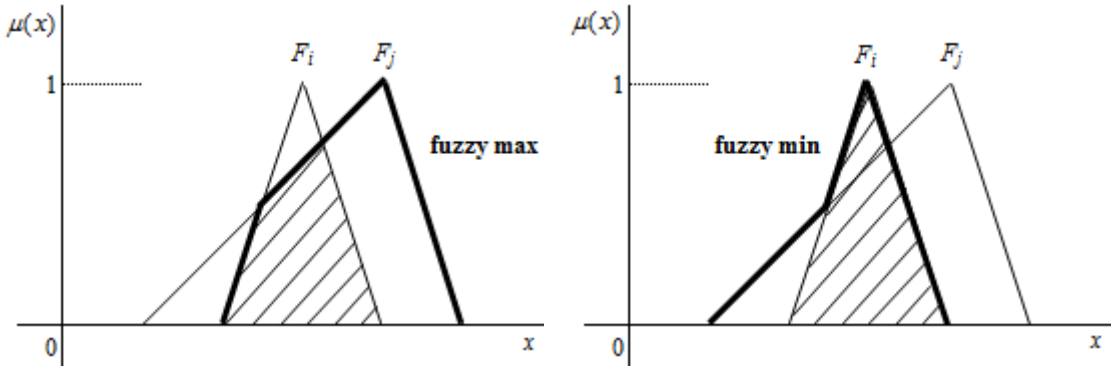


Figure 3 The contribution of F_i toward the fuzzy max and fuzzy min

In the next step, we rank the fuzzy numbers according to $P(F_i)$ descending and $N(F_i)$ upwardly. Two rankings are compared. If both raking orders are identical, the algorithm stops. If not, we will pick the fuzzy numbers sharing the same positions and perform pairwise comparison via the rules described in the following section.

We calculate the composite index

$$CP(F_i) = \frac{P(F_i)}{P(F_i) + N(F_i)}.$$

And now all fuzzy numbers sharing the same position are ranked descending according to this indicator. If the index is identical for more fuzzy numbers, it is not possible to distinguish them. Then we use the second rule, thus we compare the absolute sum of $P(F_i)$ and $N(F_i)$. One of the following relations must hold

$$\text{if } P(F_k)+N(F_k) > P(F_l)+N(F_l), \text{ then } F_k > F_l,$$
$$\text{if } P(F_k)+N(F_k) < P(F_l)+N(F_l), \text{ then } F_k < F_l,$$
$$\text{if } P(F_k)+N(F_k) = P(F_l)+N(F_l), \text{ then } F_k = F_l.$$

The problem of described method is that the locations of fuzzy numbers are not always considered. The indiscrimination case can be as follows. Generally, if we have a set of fuzzy numbers $\{F_1, F_2, ..., F_n\}$, where F_1 is the fuzzy min and F_n is the fuzzy max. Both of them do not overlap with other fuzzy numbers. It is obvious, that $F_2, F_3, ..., F_{n-1}$ cannot be distinguished by McCahone's approach. The indiscrimination case can be illustrated as follows (Figure 4).

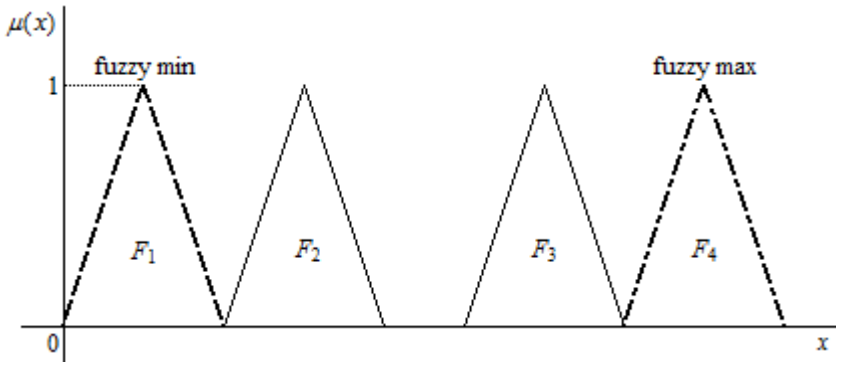


Figure 4 The indiscrimination case of McCahone's method

In this ominous situation, it holds $P(F_2) = ... = P(F_{n-1}) = N(F_2) = ... = N(F_{n-1}) = 0$. It is no doubt, that these fuzzy numbers will share the same position according to this algorithm and it is not possible to differentiate them even if they can unambiguously different. We suggest to apply the described McCahone's approach once again only for these fuzzy numbers to rank them in terms of common positions and then make a full ranking of all fuzzy numbers.

Step 3: And now we specify the set $I_{iRj} = \{r \mid y_{ir} \geq y_{jr}; i, j = 1, 2, ..., p; i \neq j\}$ containing the indices of criteria *r* according to which the alternative *i* is evaluated better or equally than the alternative *j*. Then we can formulate the matrix $\mathbf{S} = (s_{ij})$, where

$$s_{ij} = \sum_{q \in I_{iRj}} v_q \qquad i, j = 1, 2, ..., p, i \neq j,$$
$$s_{ij} = - \qquad i, j = 1, 2, ..., p, i \neq j.$$

The element of the matrix s_{ij} may be interpreted as the volume of the preference of the *i-th* alternative against the *j-th* alternative expressed as fuzzy number[3]. For each *i-th* alternative the average preference relation is calculated as follows

$$s_i^{average} = \frac{\sum_{\substack{j=1 \\ i \neq j}}^{p} s_{ij}}{p-1}.$$

Finally, the threshold value is computed as the following fuzzy number

$$s = \frac{\sum_{\substack{j=1 \\ i \neq j}}^{p} s_{ij}}{p(p-1)} = \frac{\sum_{i=1}^{p} s_i^{average}}{p}.$$

[3] The basic computational operations with fuzzy numbers are described in [6]. If some fuzzy numbers contain negative values, then positive constant is added to transform all values to nonnegative values in order to make the basic operations with fuzzy numbers easier.

Step 4: Firstly, we choose the alternatives which satisfy the following formula

$$s_i^{average} \geq s \,.$$

Because both sides of the constraint are represented by fuzzy number, the fuzzy ranking method (McCahone's approach) is applied once again. Secondly, the distance from ideal alternative is computed for choice alternatives. In the case of nonfuzzy valuation y_{ij} $(i = 1, 2, ..., p; j = 1, 2, ..., k)$, the distance from the best value by *j-th* criterion is as follows

$$d_{ij}^{nonfuzzy} = H_j^{nonfuzzy} - y_{ij}^{nonfuzzy} \,,$$

where $H_j^{nonfuzzy} = \max_i(y_{ij}^{nonfuzzy})$ is the best valuation according to *j-th* criterion.

If the criteria values are fuzzy numbers, then we use the modified concept of Hamming distance in order to find out the distance from the best valuation. The Hamming distance between two fuzzy number F_i and F_j is defined as [7]

$$d(F_i, F_j) = \int_{-\infty}^{+\infty} | \mu_{F_i}(x) - \mu_{F_j}(x) | \, dx \;\; .$$

To calculate the representative distance from the best value we will distinguish the parts, where $\mu_{F_i}(x) > \mu_{F_j}(x)$ and $\mu_{F_i}(x) < \mu_{F_j}(x)$. Let's denote fuzzy number F_j as the best valuation and F_i as a particular valuation whose distance from the best one is calculated. Specify

$$X^1 = \left\{ x; \mu_{F_i}(x) \geq \mu_{F_j}(x) \right\}, \quad X^2 = \left\{ x; \mu_{F_i}(x) < \mu_{F_j}(x) \right\}.$$

Then the distance of valuation from the best one will be defined as follows

$$d(F_i, F_j) = \int_{X^2} | \mu_{F_i}(x) - \mu_{F_j}(x) | \, dx - \int_{X^1} | \mu_{F_i}(x) - \mu_{F_j}(x) | \, dx \,.$$

In the case of fuzzy valuation, we should mark d_{ij}^{fuzzy} as the distance of *i-th* alternative from the best valuation by *j-th* criterion that is computed via the previous formula.

Finally, the general distance from ideal alternative for *i-th* chosen alternative is computed as

$$d_i = \sum_{j=1}^{k} \left[\frac{d_{ij}^{nonfuzzy}}{\max_i(d_{ij}^{nonfuzzy})} + \frac{d_{ij}^{fuzzy}}{\max_i(d_{ij}^{fuzzy})} \right].$$

The partial distance by each criterion *j* is standardized with the view of their comparability. The alternative with least distance from the ideal one is chosen as effective. This alternative is always nondominated. It is also possible to choose more alternatives which then could be dominated by effective one.

3 Practical application in the capital market with open shares funds

A potential investor decided to invest some money in open shares funds offered and managed by Investment company Česká spořitelna. There are three groups of shares funds – mixed, bond and stock funds. For illustrative instance, the group of bond funds is analysed. Thus we have five bond open shares funds – Sporoinvest, Sporobond, Trendbond, Korporátní dluhopisový and High Yield dluhopisový. The investor evaluates the investment alternative according to three main criteria – return, risk and costs.

We have 6-year time series of monthly returns of each open shares fund. According to statistic test, the distribution of returns can be described by Student's probability distribution because of its heavy tails and greater kurtosis. The distribution of returns is also usually negatively skewed. For simplification, we expect that the probability distribution is normal. As mentioned above, the probability density function is approximated by triangular shape. Then the return is expressed as a triangular fuzzy number. The risk is set as a standard deviation of monthly returns in last six years. The costs are represented by entry fee and TER (total expense ratio) containing (e. g.) management or license fee.

The preferences about criteria importance is described linguistically by the investor. The risk is very important, the return is important and the costs are little important. Each linguistic expression is transformed to a quantitative formula by means of fuzzy numbers. All necessary data is in the following table (Table 1).

Shares fund	Return	Risk	Costs
Sporoinvest	(-1.17, 0.06, 1.29)	0.41	1.07
Sporobond	(-3.7, 0.3, 4.30)	1.33	2.10
Trendbond	(-6.02, 0.18, 6.38)	2.07	2.74
Korporátní dluhopisový	(-9.63, 0.38, 10.38)	3.34	2.74
High Yield dluhopisový	(-13.15, 0.33, 13.81)	4.49	2.56
Weights	*(0.25, 0.3, 0.35)*	*(0.45, 0.55, 0.65)*	*(0.12, 0.15, 0.18)*

Table 1 The data about bond open shares funds

Shares fund Sporoinvest is the best in criteria risk and costs. According to modified McCahone's approach, Sporobond is the best in return. To choose one effective alternative the proposed fuzzy method is employed. Before the measuring distance from ideal alternative, two funds are chosen – Sporoinvest, Sporobond. Sporoinvest has the shortest distance from the ideal alternative, so it is the effective alternative according to applied introduced fuzzy method. Anyway it was quite expectable, because it dominates in the most important criterion.

This methodical approach can be applied in each group of shares funds and then the final investment portfolio can be created by appropriate methods (see more [1] or [2]).

4 Conclusion

The main contribution of this paper is a new fuzzy multiple criteria evaluation method which is able to contain some uncertain, vague information in the decision making process. The proposed approach makes use of the fuzzy set theory to quantify the uncertain elements, concretely triangular fuzzy numbers. Then the stochastic criteria values or vague information about importance of criteria may be included. This approach takes into account differences in criterial values and also computes preference relations between alternatives. In terms of the methodical procedure, the ranking of fuzzy number is done by McCahone's approach in the modified form and distance between fuzzy numbers is computed by means of the modified concept of Hamming distance.

In the case of fuzzy methods, the problem can be found in the fuzzy ranking procedure. It is obvious, that making ranking of fuzzy numbers may not be easy. So the result is also based on used fuzzy ranking approach. The disadvantage of the proposed method (in some practical decision making situation) can be the fact that it does not offer the full ranking of alternatives. But we are using this method in the concrete situation in the capital market when we want to choose only one alternative from several groups of open shares funds to make a portfolio.

Acknowledgements

The research project was supported by Grant No. IGA F4/19/2013 of the Internal Grant Agency, Faculty of Informatics and Statistics, University of Economics, Prague and Grant No. 13-07350S of the Grant Agency of the Czech Republic.

References

[1] Borovička, A. Metodický aparát k vytváření portfolia otevřených podílových fondů. In: *Vědecká konference doktorského studia FIS 2013*, Oeconomica, Praha, 2013, 118–133.
[2] Borovička, A. The investment decision making under uncertainty. In: *Mathematical methods in economics 2012*, University of Opava, Opava, 2012, 37 – 42.
[3] Dubois, D. and Prade, H.: *Fuzzy Sets and Systems: Theory and Applications*. Academic Press, Inc., New York, 1980.
[4] Gupta, M. and Bhattacharjee, D.: Multi Objective Problem in Fuzzy Environment where Resources are Triangular Fuzzy Number. *European Journal of Scientific Research* **46** (2010), 99-106.
[5] Klir, G. J. and Yuan, B.: *Fuzzy Sets and Fuzzy Logic – Theory and Application*. Prentice Hall PTR, New Jersey , 1995.
[6] Lai Y. J. and Hwang, C. L.: *Fuzzy Mathematical Programming: Methods and Applications*. Springer-Verlag, Berlin, 1992.
[7] Lai Y. J. and Hwang, C. L.: *Fuzzy Multiple Objective Decision Making: Methods and Applications*. Springer-Verlag, Berlin, 1994.
[8] McCahone, C.: *Fuzzy Set Theory Applied to Production and Inventory Control*. Ph.D. Thesis, Department of Industrial Engineering, Kansas State University, 1987.
[9] Novák, V.: *Základy fuzzy modelování*. BEN – technická literatura, Praha, 2000.
[10] Zadeh, L. A.: Fuzzy Sets. *Inf. & Control* **8** (1965), 338-353.

Consumption tax as a tool for greening transport

Botlíková Milena[1], Botlík Josef[2], Václavínková Klára[3]

Abstract. Development of transport is one of the region's competitiveness assumptions and its development is bound with European conventions. Strategic documents of the development of transport networks are based on the objectives of the White Paper, which is promoting environmentally friendly transport system, where is an appeal to create preconditions for an attractiveness of rail transport and the promotion of public transport. One of the steps to transfer traffic from road to rail and public transport was an increase of consumption taxes on fuels and lubricants, which led to an increase of fuel prices. Whether there is any relationship between the change of excise duty and change of preferred transport habits based on the performance of transport, registration of vehicles and questioning will be determined on the basis of statistical analysis such as correlation and regression[4], etc.

Keywords: taxes, consumer, automotive, consumer.

JEL Classification: C44, C01, C51
AMS Classification: 62P20, 62J12, 91B70

Introduction

In the Czech Republic are present various taxes which are state revenue and some tax acts as a regulator. Among the taxes with regulatory character belongs excise duty on fuels and lubricants. The excise tax on fuels and lubricants is an indirect tax, which is reflected in the price of fuel (1). This tax makes up 1.6% of the state budget and 9% of the total revenue of this tax is part of the State Fund for Transport Infrastructure (SFTI).

$$Fuel\ price = price + dealer\ excise\ duty + VAT \tag{1}$$

apply: Price …… dealer cost of buying = selling price of fuel, the cost of transportation. etc.
Excise……duty A fixed rate by the Ministry of Finance.

VAT …… value added tax-fixed by the Ministry of Finance

Efforts to increase the revenues of the state budget and the lack of financial resources SFDI, the development and renewal of road space, led government to increase the excise tax on fuels and lubricants in 2010. The structure of increased excise taxes is:

- Increase of excise tax on gasoline from11.84 CZK to 12.84 CZK,
- Increase of excise duty on diesel from 9.95 CZK to 10.95 CZK.

Increase of excise taxes on fuels and lubricants, together with the increase of VAT in the same year 2010 from 19% to 20% led to a significant increase of fuel prices.

The impact of increased excise taxes on fuels and lubricants

Increase of excise tax should be a tool for the creation of financial resources for rehabilitation and modernization of the transport area in the Czech Republic. After development of transport infrastructure in 2007, considerable problems occurred in financing of transport infrastructure. European funds were replaced by national; many buildings were slowed or stopped and conserved after 2009. After an increase of excise taxes SFDI items does not increase (see Figure 1). In the case of excise taxes on fuels and lubricants clearly acted Lafer curve and revenues from excise tax decreases. According to a study of Economics, which dealt with an issue of excise duty, it was revealed a need to reduce the excise duty. The reason for this reduction is clearly supportive and will not only help the state budget, but also domestic carriers, retailers and households.

[1]Silesian university, School of Business Administration University, Department of Logistisc, Univerzitní nám. 1934, 733 40, Karviná, botlikova@opf.slu.cz
[2]Silesian university, School of Business Administration University, Department of Informatics, Univerzitní nám. 1934, 733 40, Karviná, botlik@opf.slu.cz
[3]Silesian university, School of Business AdministrationUniversity, Department of Tourism, Univerzitní nám. 1934, 733 40, Karviná, vaclavinkova@opf.slu.cz
[4] Paper is based on research by SGS 5/2013, Silesian University in Opava, School of Business Administration in Karvina

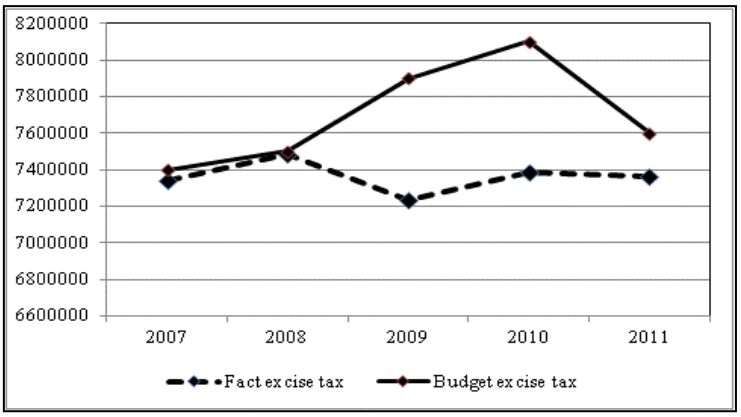


Figure 1 Excise tax SFTI 2007 - 2011

2.1 Excise tax as a means of greening transport

Another argument of the government was that the increase in excise duty PHM would not be the only tool to increase financial resources, but also a tool for increasing the use of environmentally-friendly transport and routing of the ecological behavior in transport, efforts to increase the use of public and rail transport, increase of availability of vehicles etc .. Indirect form should be to reduce congestion on the roads, reducing the environmental impacts resulting from transportation - emissions of CO2 and other greenhouse effects, etc.

Due to the increase of the real price of fuel it is expected that there will be a decrease in consumption as a result of improving the technical parameters of vehicles, restricting the use of vehicles with high capacity to change to another transport. If you want to increase the use of environmentally friendlier transport, then it should be a subsequent increase, for example [date 7]:

- Improving the performance of public transport, the number of passengers.
- Improving the Performance of railway passenger traffic, number of passengers.
- Improving the registration number of vehicles with alternative power.
- Reducing the registration of passenger vehicles.
- Reducing the number of trips by sharing transport.
- Reducing emissions REZZO 4[5]

2.2 Analysis of the impact of excise taxes on transportation habits

To assess the effect of increased excise tax and the increase in fuel prices on traffic patterns relationship between fuel price and individual characteristics was investigated. In an analysis, which is based on research SGS 5/2013, was determined using the relationship between each modality and the change in fuel prices due to excise duty. The analysis was made on the basis of the evaluation closeness of relationships between the characteristics and the price of fuel for various government interventions. Dependence was investigated in the case where was no increase of excise taxes in 2010, only VAT (Petrol_1; Diesel_1) was increased and then the relationship between characteristics and fuel price without state intervention was examined(Petrol_2; Diesel_2)

Correlation was used to determine the direct relationship between the amount of excise tax, which is reflected in the price of fuel, and individual characteristics. When data do not show a normal distribution and is not expected linear dependence, it is appropriate to use Spearman (ordinal) correlation coefficient. For qualified dependency evaluation and assessment which depends on the data with normal distribution Pearson's coefficient (1) can be used, which is more concise. Given that the data on the basis of testing SPSS a normal distribution can be the coefficient of leakage evaluation depending used.

$$r = \frac{n * \left(\sum_{i=1}^{n} x_i y_i \right) - \left(\sum_{i=1}^{n} x_i \right) \left(\sum_{i=1}^{n} y_i \right)}{\sqrt{\left(n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2 \right) \left(n \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i \right)^2 \right)}} \quad (1)$$

[5] REZZO4 - Register of Emissions and Air Pollution Sources - Number 4 - mobile device or other internal combustion engines that pollute the air, especially road and motor vehicles, rail vehicles, vessels and aircraft

Based on the analysis, it was tested whether individual variables are correlated variables, if the correlation coefficient is equal to zero, ie, if true than: H_0: $r = 0$ the independence of two variables against the alternative hypothesis H_1: $r \neq 0$, the dependent variables and whether the individual coefficients correlation considered statistically significant, and thus applies the value of the test criterion T (2) is greater than the critical value of the coefficient $t_{(n-2)(\alpha)}$ (3).

$$T = \left| \frac{r}{\sqrt{(1-r^2)}} \right| * \sqrt{(n-2)} \tag{2}$$

$$|T| > \text{critical value } t_{(\alpha, n-2)} = 2{,}06 \tag{3}$$

apply: *n –Number of observations*

The values of the correlation coefficient in the range -1 < y < 1 can be interpreted on the basis of the claim as follows: [4,5] (Figure 2). The negative value of r indicates the relationship contradictory, if the there is growth of one character (X, respectively. Y) there will be a decrease of the second (X, respectively. Y). A positive correlation coefficient represents an increase of values for both variables, ie., With the growth of one variable (X respectively. Y) the second will fluctuate (X, respectively. Y).

Correlation coefficient	Interpretation
$\|r\| \geq 1$	absolute dependence (functional)
$1 > \|r\| \geq 0,9$	very high
$0,9 > \|r\| \geq 0,7$	high
$0,7 > \|r\| \geq 0,4$	medium
$0,4 > \|r\| \geq 0,2$	low dependence
$0,2 > \|r\| \geq 0,00$	low useless
$\|r\| = 0$	complete independence

Figure 2 The interpretation of correlation coefficient

Based on the analysis and interpretation of SPSS (see Figure 3) by Chráska was found functional dependence between the 95 octane petrol and bus, city bus, tram and underground transport. It was found that, at the current a tax burden exhibit different magnitudes greater dependence than if the tax burden was maintained at the original level. If the price of petrol 95 octane and diesel increases, the number passengers in bus and tram transport will decrease. If the fuel prices increase, there will be increase of passenger stations. The analysis confirmed the correlation between the price of fuel performance shipments (million passenger-kilometers) of railway and metro.

		BUS	RAILWAY	BUS MHD	TROLEJBUS	TRAMVAJ	METRO
PETROL 95	Pearson Correlation	-0.414	-0.179	**-0.754**	-0.148	**-0.648**	**0.673**
	Sig (2-tailed)	0.029	0.381	0.000	0.458	0.000	0.000
	N	28	28	28	28	28	28
DIESEL	Pearson Correlation	-0.307	-0.081	**-0.661**	-0.029	**-0.599**	**0.632**
	Sig (2-tailed)	0.112	0.681	0.000	0.882	0.001	0.000
	N	28	28	28	28	28	28
PETROL_1	Pearson Correlation	-0.442	-0.117	**-0.681**	-0.134	**-0.568**	**0.580**
	Sig (2-tailed)	0.018	0.555	0.000	0.496	0.002	0.001
	N	28	28	28	28	28	28
DIESEL_1	Pearson Correlation	-0.314	-0.008	**-0.569**	-0.001	**-0.509**	**0.530**
	Sig (2-tailed)	0.104	0.967	0.002	0.996	0.006	0.004
	N	28	28	28	28	28	28
PETROL_2	Pearson Correlation	**-0.448**	-0.100	**-0.660**	-0.129	**-0.546**	**0.555**
	Sig (2-tailed)	0.017	0.612	0.000	0.512	0.003	0.002
	N	28	28	28	28	28	28
DIESEL_2	Pearson Correlation	-0.314	0.010	**-0.545**	0.007	**-0.485**	**0.503**
	Sig (2-tailed)	0.104	0.961	0.003	0.972	0.009	0.006
	N	28	28	28	28	28	28

Figure 3 The relationship between the price of fuel and the number of persons on selected modalities

* Bus - bus lines and irregular (national and international), ** Bus_1 - bus service within IDOS (urban transport)

With the increase of price of fuel there will be an increase in traffic of railway and metro. Higher correlation dependencies were waylaid between the performance of rail transport in case of maintaining excise duty on the border of CZK 11.84 for petrol and 9.95 for diesel and maintenance of VAT at 19%. In the case of bulk power stations were established interdependence between the price of fuel (petrol and diesel) with excise duties and VAT, whose development was accompanied by changes.

On the contrary, the case of rising fuel prices will lead to a reduction in performance of trolleybus transport. Higher dependence between gasoline price and performance of trolleybus transport was demonstrated in the case of oil prices from the treated excise taxes in the case of gasoline excise original and modified, but the original VAT (see Figure 4).

		PER BUS	PER RAILWAY	PER BUS MHD	PER TROLEJBUS	PER TRAMVAJ	PER METRO
PETROL 95	Pearson Correlation	-0.062	0.379	-0.185	**-0.550**	-0.125	**0.442**
	Sig (2-tailed)	0.755	0.047	0.376	0.002	0.528	0.019
	N	28	28	28	28	28	28
DIESEL	Pearson Correlation	-0.100	0.392	-0.107	**-0.455**	-0.082	**0.485**
	Sig (2-tailed)	0.613	0.039	0.589	0.015	0.677	0.009
	N	28	28	28	28	28	28
PETROL_1	Pearson Correlation	-0.088	**0.416**	-0.213	**-0.568**	-0.139	0.392
	Sig (2-tailed)	0.657	0.028	0.277	0.002	0.480	0.039
	N	28	28	28	28	28	28
DIESEL_1	Pearson Correlation	-0.128	**0.423**	-0.120	**-0.447**	0.089	**0.437**
	Sig (2-tailed)	0.516	0.025	0.542	0.017	0.653	0.020
	N	28	28	28	28	28	28
PETROL_2	Pearson Correlation	-0.093	**0.424**	-0.219	**-0.564**	-0.142	**0.377**
	Sig (2-tailed)	0.636	0.025	0.262	0.002	0.470	0.048
	N	28	28	28	28	28	28
DIESEL_2	Pearson Correlation	-0.133	**0.429**	-0.124	**-0.442**	-0.090	**0.423**
	Sig (2-tailed)	0.499	0.023	0.530	0.018	0.649	0.025
	N	28	28	28	28	28	28

Figure 4 The relationship between fuel price and performance of selected modalities

In the event of an increase of fuel prices was seen positive correlation between the price of fuel and registrations of new cars with diesel engines and drive vehicles LPG, CNG (see Figure 5).

		PETRO CAR	DIESELCAR	ECOCAR	OTHERCAR	OLD CAR
PETROL 95	Pearson Correlation	0.094	**0.658**	**0.779**	-0.272	-0,312
	Sig (2-tailed)	0.633	0.000	0.000	0.161	0,105
	N	28	28	28	28	28
DIESEL	Pearson Correlation	0.094	**0.539**	**0.756**	-0.139	-0,211
	Sig (2-tailed)	0.634	0.003	0.000	0.480	0,282
	N	28	28	28	28	28
PETROL_1	Pearson Correlation	0.092	**0.583**	**0.760**	-0.244	-0,215
	Sig (2-tailed)	0.641	0.001	0.000	0.211	0,272
	N	28	28	28	28	28
DIESEL_1	Pearson Correlation	0.091	**0.445**	**0.724**	-0.092	-0,102
	Sig (2-tailed)	0.646	0.018	0.000	0.642	0,605
	N	28	28	28	28	28
PETROL_2	Pearson Correlation	0.091	**0.652**	**0.752**	-0.235	-0,188
	Sig (2-tailed)	0.644	0.002	0.000	0.228	0,337
	N	28	28	28	28	28
DIESEL_2	Pearson Correlation	0.090	**0.420**	**0.712**	-0.078	-0,074
	Sig (2-tailed)	0.650	0.026	0.000	0.692	0,709
	N	28	28	28	28	28

Figure 5 The relationship between fuel price and new car registrations by type of fuel

This clarified the annual growth of each new car registrations (see Figure 6). Although, there was a presumption that the increased price of fuel will be restricted in the use of vehicles with higher consumption, where you can include the vehicle's age, dependency between registration of old vehicles and fuel price has not been demonstrated.

in %	Petrol	Diesel	CNG	LPG	E85	Biodiesel	Elektric	Hybrid	Other
2006	69,25	28,13	0	0	0	0	0	0	2,62
2007	69,52	27,73	0	0	0	0	0	0	2,73
2008	70,97	23,91	0,05	0	0,01	0	0	0	5,07
2009	64,88	31,77	0,06	0,01	0,02	0	0	0	3,25
2010	57,28	40,00	0,09	0,04	0,02	0	0	0	2,57
2011	57,80	39,54	0,11	0,07	0,20	0	0,03	0	2,25
2012	55,79	41,38	0,27	0,30	0,34	0	0,05	0,21	1,67

Figure 6 Development of new car registrations by fuel used

Quality of correlation dependence was analyzed by comparing the test criteria T and critical values *t* of for statistically significant we can consider the interdependence, $| T | \geq t\ (n-2)\ (\alpha)$ ie. $| T | \geq 2.06$, correlation or rejected. Based on the calculations, we can confirm the correlation to be statistically significant at the significance level $\alpha = 0.05$ for all selected variables (see Figure 7).

	Petrol 95_2 x Bus	Petrol 95x Bus_1	Petrol 95 x Tramvaj	Petrol 95 x Metro	Diesel x Bus_1	Diesel x Tramvaj	Diesel x Metro	Petrol 95_2 x perf railway	Petrol x perf trolejbus	Petrol x perf metro	Diesel_2 x perf railway	Diesel x perf trolejbus	Diesel x perf metro	Petrol 95 x Diesel Car	Petrol 95 x Eco Car	Diesel x Diesel Car	Diesel x Eco Car
Test crieteria *T*	2.56	5.58	4.34	4.63	4.49	3.81	4.16	2.39	3.48	2.51	2.42	2.61	2.83	4.46	6.33	3.26	5.84
Critical values *t*	> 2.06																

Figure 7 Test the statistical significance of the correlation between the different variables

Confirmation of the existence of relationships between the price of fuel and transport behavior of individual consumers was based on a test of independence of two characters performed using a pilot survey.

Fuel price/How high fuel costs, by forcing you to changing traffic patterns	SHARING	ALTERNATIVE TRANSPORTATION	FUEL ALTERNATIVE	OTHER
Non specified	1	2	2	3
35 - 39	1	1	1	1
40 - 43	1	1	1	1
44 - 47	3	4	1	5
> 47	2	4	1	12

Figure 8 Pilot survey

Reaching respondents were not forced to deal with changes in transportation habits at the current price level, however, if there was an increase in fuel prices over 47 CZK they would consider their current modes of transportation (see Figure 8).

Confirmation or rejection of the relationship between fuel price and changes in travel behavior was based on the determination of hypotheses $H_0 = p_i = p_j$ independence of characters, comparing the test criterion and a critical value. If the value of *G* belonged to the interval (4), then the alternative H_1 reject the values of independence and pays between variables on doing. Since G = 6.93, falls within the field admission means. that lies in the interval (5), the hypothesis was the independence of two characters taken. It was therefore confirmed the relationship between the price of fuel and possible changes in transport practices. [2]

$$G = \sum_{i=1}^{J} \frac{(n_i - \psi_i)^2}{\psi_i} \tag{4}$$

Critical value

$$\chi^2_{(r-1)(s-1)}(\alpha) = 21.02$$

when: *r = 4; s = 5; α = 0.05*

Industry adoption $H_{o:}$ $\left\langle 0; \chi^2_{(r-1)(s-1)}(\alpha)\right\rangle = \left\langle 0; 21.02\right\rangle$ (5)

Field rejection of $H_{o:}$ $\left\langle \chi^2_{(r\ \ 1)(s\ \ 1)}(\alpha); +\infty \right\rangle = \left\langle 21.02; +\infty\right)$

Conclusion

This analysis did not find a causal relationship. There were found only association relationships. It has been shown that if the X (or Y) changed also Y (or X) would change, but it was not confirmed whether this change was caused by X or Y. There is no confirmation that it affects fuel prices preference for modality or modalities demand affecting prices.

In general it can be said that the majority of selected variables has demonstrated correlation dependence. Positive correlation dependence was observed in the number of shipments and transport operations underground, even if state intervenes in fuel prices. A positive correlation was also demonstrated in rail transport performance, higher dependence was demonstrated in the case with no changes in excise duty and VAT in 2010. For other quantities, if there was proven correlation dependence, there was dependent negative character, with rising fuel prices will be reduced the usage of bus, city bus, trolleybus and tram system, which can cause increasing congestion, reducing security and other problems associated with transport within cities. A positive aspect is that the increasing fuel prices lead to greening the fleet.

References

[1] Brůhová – Foltýnová, H. *Application of microsimulation models in passenger transport: experiences from the Czech Republic and abroad.* Center for the Environment University in Prague. GA CR "Theoretical analysis of traffic regulation No. 402/07/P484

[2] Čemerková, Š., Mielcová, E.: *Statistical calculations in Excel. Textbook* SU OPF. Karviná (2007), s.90

[3] Czech Statistical Office. [On-line 2012]. 2012. Availabe from: www.cszo.cz

[4] Chráska, M.: *Introduction to Research in Education* (Basic quantitatively oriented research). Publishing Palacky University, Olomouc in (2003)

[5] Chráska, M.: *Bases of the research in pedagogic*. Publishing Palacky University, Olomouc (2000)

[6] *Ministry of Transport. Transport statistics.* [On-line 2012]. , 2012. Available from: http:www.sydos.

The Empirical Analysis of the Taylor Rule

Milan Bouda[1]

Abstract. This paper deals with the empirical analysis of the Taylor rule. This rule is used in all New Keynesian DSGE models. The Taylor rule represents the behavior of the central bank. Currently, there are many versions of Taylor rules. This paper deals with the Taylor rule which is used by the Ministry of Finance of the Czech Republic in the model which is called HUBERT. It is a DSGE model that quantifies impacts of fiscal policy. The Taylor rule consists of gap and steady state variables. The empirical time series must be transformed into the steady state variables using the Hodrick-Prescott filter. The parameter estimates are performed using two different methods. The first approach uses the instrumental variables technique. It is necessary to estimate the compound expression which is present in the Taylor rule and in the second is the whole Taylor rule estimated. For both steps is used the Least-Squares method. The second approach estimates the whole Taylor rule at once using the Marquardt algorithm. It is the Least-Squares method which allows estimating the non-linear parameters. These two approaches are compared and the Marquardt algorithm is chosen as the best one. The paper verifies the quality of the Taylor rule and the results of the parameter estimates might be used for the calibration of the New Keynesian DSGE model of the Czech Republic.

Keywords: Taylor rule, H-P filter, HUBERT, DSGE.

JEL Classification: C01
AMS Classification: 91B84

1 Introduction

Monetary policy is the process by which the monetary authority of the Czech Republic controls the supply of money, often targeting a rate of interest for the purpose of promoting economic growth and stability. In economics, a Taylor rule is a monetary policy rule that stipulates how much the central bank should change the nominal interest rate in response to changes in inflation, output, or other economic conditions. In particular, the rule stipulates that for each one-percent increase of inflation, the central bank should raise the nominal interest rate by more than one percentage point. This aspect of the rule is often called the Taylor principle. For further details see Taylor [15].

The paper deals with the empirical analysis of the Taylor rule used in the model which is used by the Ministry of Finance (MF) of the Czech Republic. The MF uses a simple DSGE model of the Czech economy called HUBERT, see Štork, Závacká and Vávra [14]. The model describes the behavior of four basic agents in the economy: households, firms, government and world. Although Hubert is rather a simple version of standard DSGE models it incorporates standard features of New Keynesian economics such as imperfect competition, habit formation of households, nominal and real rigidities.

The main goal of this paper is to analyze the time series which are present in the Taylor rule. This rule is used in each New Keynesian model to predict the behavior of the central bank. It is necessary to analyze the Taylor rule and estimate its parameters. These estimated values can be used in future as a calibrated values of the New Keynesian model of the Czech Republic.

2 Taylor rule

The Taylor rule was developed by Taylor [15] and discussed by Svensson [13]. The rule takes the following form

$$i_t = (1-\phi_i)\left[\bar{i} + \lambda_\pi \hat{\pi}_t + \lambda_y \hat{y}_t\right] + \phi_i i_{t-1}, \tag{1}$$

[1] University of Economics, Prague, Department of Econometrics, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic, bouda.mil@seznam.cz

where i_t is short-term nominal interest rate, $\bar{i}$ steady state value of short-term interest rate, $\hat{\pi}_t$ deviation of inflation rate from its steady state (target) value, $\hat{y}_t$ output gap, λ_y output gap weight, λ_π inflation weight, ϕ_i interest rate smoothing parameter.

According to Srour [12], there are many reasons for interest rate smoothing. First, the behaving of the central bank is important for investors and smoothing of interest rate ϕ_i can reduce volatility of a term premium and therefore volatility of long-term interest rates and other financial market instruments. Second, the central bank has usually limited information about the shocks hitting the economy. Third, many shocks are serially correlated. According to Levin et al. [8], simple monetary policy rules with a high degree of interest rate smoothing $(\phi_i \to 1)$ are also surprisingly robust against model uncertainty and misspecification. Unfortunately, this is probably a characteristic feature for large closed economies only. Coté et al. [4] find that the most robust rule is the original $(\phi_i \to 1)$ for small open economy. Much worse, Natvik [10] shows that extending a DSGE model for a fiscal block can lead to a serious determinacy problem. From this point of view, a cautionary note should be made for straightforward application of Taylor rules.

3 Hodrick-Prescott filter

The Hodrick–Prescott (HP) filter is a mathematical tool used in macroeconomics, especially in a real business cycle theory to separate the cyclical component of a time series from raw data. It is used to obtain a smoothed-curve representation of a time series, one that is more sensitive to long-term than to short-term fluctuations. The adjustment of sensitivity of the trend to short-term fluctuations is achieved by modifying a multiplier λ. The filter was popularized in the field of economics in the 1990s by economists Robert J. Hodrick and Nobel Memorial Prize winner Edward C. Prescott, see Hodrick and Prescott [6].

The H-P filter allows to separate trend and business cycle from data using the following optimization problem

$$\min_{\left\{y_t, y_t^{trend}\right\}} \sum_{t=1}^{T} \left(y_t\right)^2 + \lambda \sum_{t=1}^{T} \left[\left(y_{t+1}^{trend} - y_t^{trend}\right) - \left(y_t^{trend} - y_{t-1}^{trend}\right)\right]^2 \tag{2}$$

subject to

$$y_t + y_t^{trend} = \log\left(Y_t\right), \tag{3}$$

where T is the last period. We require smooth trend which corresponds to data. This compromise is given by the parameter λ which weight the following expression

$$\left[\left(y_{t+1}^{trend} - y_t^{trend}\right) - \left(y_t^{trend} - y_{t-1}^{trend}\right)\right]^2. \tag{4}$$

The most important is a choice of λ parameter. Hodrick and Prescott [6] recommend for quarterly data $\lambda = 1600$. The smoothed GDP by H-P filter is very often called potential product. The difference between the original and smoothed data is interpreted as output gap. The H-P filter enables to quantify the steady state variables and the output gap. This approach was also performed by Bjørnland [3], Proietti [11] and Billmeier [2].

4 Data

Data are obtained from ARAD (database of the Czech National Bank) and Czech Statistical Office. In Table 1 is a list of all variables which are present in the Taylor rule. In Table 1 is also a short description of these variables and their source. There is a time period from Q1 2000 to Q3 2012 (51 observations). Following three variables are used as a source: PRIBOR 3M (in %), Consumer Price Index ($t-12=100$) and Gross Domestic Product (in millions) is in previous year average prices and it is seasonally adjusted. The target value of inflation is taken from the Czech National Bank. These time series are used in the following data transformation which transforms this time series to steady state or gap time series.

variable	description	source/transformation
i_t	short-term nominal interest rate	PRIBOR 3M
$\bar{i}$	steady state value of short-term interest rate	H-P filtered PRIBOR 3M
$\hat{\pi}_t$	deviation of inflation rate from its steady state (target) value	(Inflation) – (Inflation target set by the CNB)
$\hat{y}_t$	output gap in %	GDP minus H-P filtered GDP divided by H-P filtered GDP

Table 1 Description and source of all variables

5 Empirical analysis

This chapter is focused on empirical analysis of the Taylor rule. First, it is necessary to quantify the steady state and gap variables. This quantification is performed by the Hodrick-Prescott filter. The transformed variables are displayed in Figure 1, Figure 2 and Figure 3. Next, these variables are used in the empirical analysis.

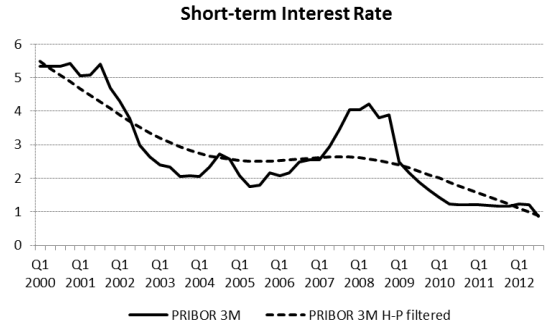


Figure 1 H-P filtered PRIBOR 3M

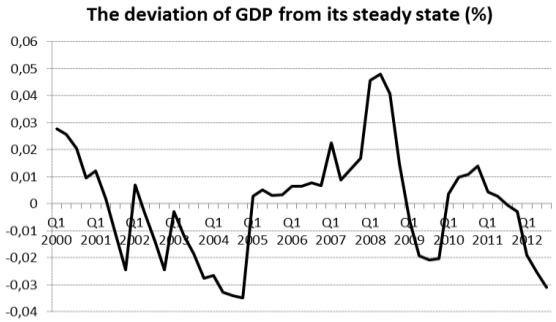


Figure 2 Output gap

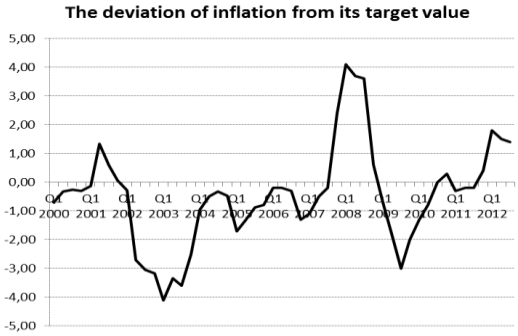


Figure 3 Deviation of inflation rate from its steady state

The empirical analysis verifies the quality of the Taylor rule. The verification is performed using two methods. First, model is estimated by the two steps instrumental method and the second approach is the estimates Taylor rule by the Marquardt method which allows estimating the nonlinear parameters.

Before the model is formulated, the stationarity of all variables is tested. The stationarity may be verified just by looking at the data or with the exact statistical tests, see Greene [5] or Arlt and Arltová [1]. The method used in this paper is KPSS test, for details see Kwiatkowski et al. [7]. The null hypothesis says that the time series is stationary. The Table 2 contains results of the KPSS test. In column KPSS are results of used time series and in column KPSS there are results, there are first differences of used time series. Level of significance is set to 5%. This test proves that variables i_t and $\bar{i}$ are not stationary. Yet it was decided to use original data and do not use first differences. Both presented estimation methods do not assume stationarity of time series and if the time series were modified the economic interpretation would be lost.

variable	KPSS	KPSS for differenced data
i_t	non-stationary	stationary
$\bar{i}$	non-stationary	stationary
$\hat{\pi}_t$	stationary	stationary
$\hat{y}_t$	stationary	stationary

Table 2 KPSS test for all variables

The (1) is estimated in two steps. The goal of the first step is to estimate the composite expression of (1). The following expression is estimated using the Ordinary Least Squares (OLS)

$$i_i = const + \alpha \bar{i} + \lambda_\pi \hat{\pi}_t + \lambda_y \hat{y}_t + \varepsilon_t, \tag{5}$$

$$i_i = -0.52 + 1.25\bar{i} + 0.30\hat{\pi}_t + 1.83\hat{y}_t. \tag{6}$$

The second step is the estimation of the Taylor rule

$$i_t = const + (1-\phi_i)\tilde{i}_t + \phi_i i_{t-1} + \varepsilon_t, \tag{7}$$

$$i_t = -0.06 + 0.50\tilde{i}_t + 0.50 i_{t-1}, \tag{8}$$

where the $\tilde{i}_t$ is represented by the fitted values from the (6). The estimated Taylor rule can be written in this form

$$i_t = (1-0.5)\left[\bar{i} + 0.30\hat{\pi}_t + 1.83\hat{y}_t\right] + 0.5 i_{t-1}. \tag{9}$$

Both equations (5) and (7) are estimated using the OLS in SAS.

The second method of the estimation of the Taylor rule is one step method. The estimation of all parameters in the (1) is performed using the SAS software and the Marquardt algorithm. This is an algorithm for least-squares estimation of nonlinear parameters. For details see Marquardt [9]. The results of the estimation procedure are in the Table 3.

parameter	description	estimate	std. error	lower	upper	pr > |t|
ϕ_i	interest rate smoothing parameter	0.6213	0.0675	0.4855	0.7571	<0.0001
λ_π	inflation weight	0.3462	0.0702	0.2049	0.4874	<0.0001
λ_y	output gap weight	6.1907	5.7605	-5.3979	17.7793	0.2880

Table 3 Parameter estimates

Figure 4 and Figure 5 show the short-term interest rate and fitted values. It is obvious that the Taylor rule works well in the Czech economy. The bad fact is the poor estimate of λ_y. Root Mean Square Error is in case of two step (instrumental) method equal to 0.071 and in case of Marquardt estimation method is equal to 0.065.

The Ministry of Finance of the Czech Republic calibrates parameters as follows (Table 4)

parameter	description	estimate	HUBERT
ϕ_i	interest rate smoothing parameter	0.6213	0.70
λ_π	inflation weight	0.3462	1.50
λ_y	output gap weight	6.1907	0.25

Table 4 Parameter estimates vs. Calibrated values in DSGE model HUBERT

As one can see, except of interest rate smoothing parameter ϕ_i the values are completely different.

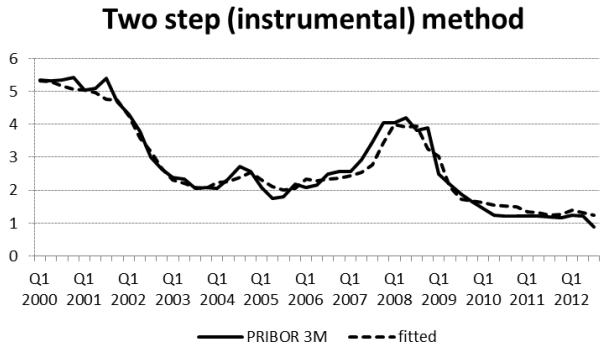


Figure 4 Instrumental method and its fit to data

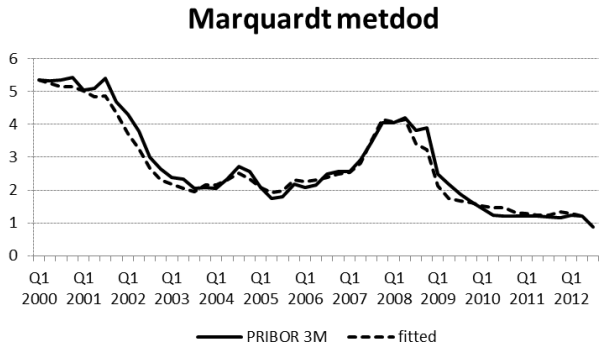


Figure 5 Marquardt method and its fit to data

6 Conclusion

The paper deals with the empirical analysis of the Taylor rule. This rule is very common in the New Keynesian models. The Taylor rule exists now in a lot of forms, but in the paper is analyzed the Taylor rule discussed by Svensson [13]. The rule is implemented in the DSGE model HUBERT which is used by the Ministry of Finance of the Czech Republic and its purpose is to analyze the impacts of fiscal policy. In the estimation of the DSGE model one has to calibrate values of the parameters which are in the Taylor rule. These parameters of the model have to be calibrated and this paper provides the deeper analysis of the rule and its fit to data.

First chapter describes the Taylor rule and its origin and purpose. Next chapter discusses the data sources and necessary data transformation. This chapter contains the description of all variables and parameters. Third chapter briefly describes Hodrick-Prescott filter and this filter is then applied on time series and one get as a result the gap or steady state variables. The last chapter deals with the empirical study. The estimation of parameters is performed using two different methods. The first is the two step instrumental variable method. The purpose of the method is to gain the estimate of the parameters inside the composite member of the Taylor rule. Next, in the second step the complete equation is estimated. The latter method estimates the equation at once. This method uses the Marquardt algorithm which allows estimation of non-linear equations. The quality of the final fit to the short-term interest rate is almost the same for both methods. According to Root Mean Square Error is the Marquardt algorithm better than two step instrumental variable method. Finally, values used by the Ministry of Finance in calibration of their DSGE model differ from parameter estimates which are obtained by Marquardt algorithm in this paper. According to this finding it is recommended to consider values used by the Ministry of Finance.

Acknowledgements

This paper is supported by funding of IGA VŠE, project IG403052.

References

[1] Arlt, J., Arltová, M.: *Ekonomické časové řady*. Professional Publishing, 2009.
[2] Billmeier, A., Brubakk, L., Jore, A. S.: Ghostbusting: which output gap really matters?. *International Economics and Economic Policy* 6/4 (2009), 391-419.
[3] Bjørnland, C., Brubakk, L., Jore, A. S.: Forecasting inflation with an uncertain output gap. *Empirical Economics* 35/3 (2008), 413-436.
[4] Coté, D., Kuszczak, J., Lam, J., Ying, L., St-Amant, P.: The Performance and Robustness of Simple Monetary Policy Rules in Models of the Canadian Economy. *Bank of Canada Technical Reports* 96 (2002).
[5] Greene, W.: *Econometric analysis*. 7th ed. Boston: Prentice Hall, 2012.
[6] Hodrick, R., Prescott, E. C.: Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit, and Banking* 29/1 (1997), 1-16.
[7] Kwiatkowski, D., Phillips, P., Schmidt, P., Shin, Y.: Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root. *Journal of Econometrics* 54 (1992), 159-178.
[8] Levin, A., Wieland, V., Williams, J.: Robustness of Simple Monetary Policy Rules under Model Uncertainty. *NBER Working Papers* 6570 (1998).
[9] Marquardt, D., W.: An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics* 11/2 (1963), 431-441.
[10] Natvik, G.: Government Spending and the Taylor Principle. *Norges Bank, Working Paper* 2006/11 (2006).
[11] Proietti, T., Musso, A., Westerman, T.: Estimating potential output and the output gap for the euro area: a model-based production function approach. *Empirical Economics* 33/1 (2007), 85-113.
[12] Srour, G.: Why Do Central Banks Smooth Interest rates? *Bank of Canada Working Papers* 17 (2001).
[13] Svensson, L.: Open –Economy Inflation Targeting. *The American Economic Review* 80 (1998).
[14] Štork, Z., Závacká, J., Vávra, M.: HUBERT: a DSGE mode of the Czech economy. *Working Paper of Ministry of Finance of the Czech Republic* 2/2009 (2009), 1-39.
[15] Taylor, J.: Discretion versus policy rules in practice. *Journal of business and Economic Statistics* 19 (1993).

Diversification-consistent DEA-risk tests – solution techniques and an empirical comparison

Martin Branda[1]

Abstract.

In this paper, we will propose numerically tractable formulations of the diversification-consistent DEA tests, which generalize traditional DEA tests as well as mean-risk models. We employ general deviation measures to measure risk of the investment opportunities. We will compare strength of the tests and give characterizations of efficient and inefficient investment opportunities. US industry representative portfolios will be ranked using the proposed DEA tests.

Keywords: Data envelopment analysis, diversification, deviation measures, efficiency, US industry representative portfolios

JEL classification: C44
AMS classification: 90C15

1 Introduction

Data Envelopment Analysis (DEA) was introduced by Charnes, Cooper and Rhodes [11] as a tool for testing efficiency of decision making units with the same structure of consumed inputs and produced outputs. Many generalizations have been proposed since than taking into account various aspects of production theory and practical applications, see, e.g., Cooper et al. [13]. A special attention has been given to applications of DEA models in finance. Murthy et al. [19] accessed in their seminal work the efficiency of mutual funds based on their mean-risk profile and transaction costs. More precisely, the indicators, which are preferably minimized by investors (e.g. risk, transaction costs), served as the inputs, and those, which are maximized (e.g. expected return), were used as the outputs. DEA model with variable return to scale introduced by Banker et al. [1] was employed. Similar efficiency tests based on DEA were proposed also by Basso and Funari [2], Branda and Kopa [8], Chen and Lin [12], where also other inputs and outputs were considered.

Several authors realized that the standard DEA tests employed in finance do not take into account dependencies between the assets. Namely, risk measures used as the inputs were combined linearly in the dual formulations, which does not correspond to proper diversification. It was even shown by Branda [6] that ignoring diversification leads to weaker tests which identify significantly higher number of efficient investment opportunities. DEA tests with diversification were introduced by Briec et al. [10], Joro and Na [14], Lozano and Gutiérrez [16], Branda [3]. However, these models were limited to particular inputs and outputs. Recently, Lamb and Tee [15] employed positive parts of coherent risk measures as the inputs and return measures as the outputs and introduced a general class of diversification-consistent tests. Several extension were proposed by Branda [6] who introduced input and input-output oriented tests focusing on the strength of the tests. Moreover, he avoided cutting the negative part of risk measures and suggested general deviation measures (Rockafellar et al. [20]) to quantify riskiness as the inputs. The resulting models can be seen as generalizations of Markowitz mean-risk models [17, 18], where variance and semivariance were used to quantify risk and which both belong to the class of general deviation measures. Branda and Kopa [9] even showed equivalence between particular new DEA tests and stochastic dominance efficiency tests.

[1]Charles University in Prague
Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics
Sokolovská 83, Prague 186 75, Czech Republic
tel.: +420 221 913 404, fax: +420 222 323 316, branda@karlin.mff.cuni.cz

This paper is organized as follows. In Section 2, we will propose basic notation and axiomatic definitions of deviation and return measures. Standard DEA tests and diversification-consistent extensions are proposed in Section 3. In Section 4, we employ the DEA tests to access efficiency of 46 US industry representative portfolios.

2 Preliminaries and notation

In this section, we introduce basic notion and review the axioms of deviation and return measures, which will serve as the inputs and outputs in the following DEA tests. Let $\mathcal{X}$ be a set of random returns of available investment opportunities corresponding to one single asset or to a portfolio consisting of many assets. We consider n assets and denote R_i the return of i-th asset which is a real random variable defined on the probability space $(\Omega, \mathcal{A}, P)$. We will use the set with no short sales enabling full diversification: $\mathcal{X} = \{\sum_{i=1}^{n} R_i x_i : \sum_{i=1}^{n} x_i = 1, x_i \geq 0\}$. Other choices of the set are also possible, e.g. with limited number of assets, allowing short sales, borrowing or including proportional and fixed transaction costs, cf. Branda [4, 6].

Functionals $\mathcal{D} : \mathcal{L}_2(\Omega) \to [0, \infty]$ are called general deviation measures if they satisfy:
(D1) translation invariance: $\mathcal{D}(X + C) = \mathcal{D}(X)$ for all X and constants C,
(D2) positive homogenity: $\mathcal{D}(0) = 0$, and $\mathcal{D}(\lambda X) = \lambda \mathcal{D}(X)$ for all X and all $\lambda > 0$,
(D3) subadditivity: $\mathcal{D}(X + Y) \leq \mathcal{D}(X) + \mathcal{D}(Y)$ for all X and Y,
(D4) nonnegativity: $\mathcal{D}(X) \geq 0$ for all X, with $\mathcal{D}(X) > 0$ for nonconstant X.
Note that the axioms (D2) and (D3) imply convexity. The main examples are standard deviation and semideviations.

Functionals $\mathcal{E} : \mathcal{L}_p(\Omega) \to (-\infty, \infty]$ are called return measures if they satisfy:
(E1) translation equivariance: $\mathcal{E}(X + C) = \mathcal{E}(X) + C$ for all X and constants C,
(E2) positive homogenity: $\mathcal{E}(0) = 0$, and $\mathcal{E}(\lambda X) = \lambda \mathcal{E}(X)$ for all X and all $\lambda > 0$,
(E3) superadditivity: $\mathcal{E}(X + Y) \geq \mathcal{E}(X) + \mathcal{E}(Y)$ for all X and Y,
(E4) monotonicity: $\mathcal{E}(X) \geq \mathcal{E}(Y)$ when $X \geq Y$.
The space $\mathcal{L}_p(\Omega)$ is selected so as the measures are finite, usually $p = 1$. The axioms (E2) and (E3) imply concavity of the functional. It is obvious that expectation fulfills the axioms. Moreover, coherent risk measures multiplied by a negative constant can be used as return functionals, see Lamb and Tee [15].

3 DEA-risk tests

In this section, we will formulate several DEA tests, which will be employed in the numerical study to access efficiency of US market portfolios. We start with the standard model with variable return to scale. We can simply demonstrate its drawbacks and motivate an introduction of the DEA tests with diversification. The efficiency is expressed by the optimal value of the corresponding DEA tests – we say that an investment opportunity is efficient if the optimal value is equal to 1, otherwise we say that it is inefficient.

Traditional mean-risk (mean-deviation) efficiency can be defined as follows: an investment opportunity is efficient if there is no other investment opportunity with higher or equal expected return and lower of equal risk (deviation) with at least one inequality strict. This approach to efficiency is extended by the following DEA tests where K deviation measures and J return measures are employed at the same time.

3.1 Traditional input oriented tests

We are going to access efficiency of an investment opportunity R_0. We assume that the benchmark R_0 is not constant, which implies that the employed deviation measures are positive. The standard DEA test with variable return to scale introduced by Banker et al. [1] can be formulated in the dual form as

a linear program as follows:

$$\begin{aligned}
\theta_{VRS}(R_0) &= \min_{\theta, x_i} \theta \\
\text{s.t. } \sum_{i=1}^{n} x_i \mathcal{E}_j(R_i) &\geq \mathcal{E}_j(R_0),\ j = 1, \ldots, J, \\
\sum_{i=1}^{n} x_i \mathcal{D}_k(R_i) &\leq \theta \cdot \mathcal{D}_k(R_0),\ k = 1, \ldots, K, \\
\sum_{i=1}^{n} x_i &= 1,\ x_i \geq 0,\ 1 \geq \theta \geq 0.
\end{aligned}$$

This dual formulation shows the main problem of the standard DEA tests in finance. It compares benchmark deviation with a linear combination of assets deviations, which has no financial or economical meaning. This drawback is removed by the diversification consistent (DC) tests proposed in the following parts.

3.2 Input oriented diversification-consistent tests

The input oriented test can be extended to take into account diversification. This can be simply done by interchanging the deviation and sum operators resulting into

$$\begin{aligned}
\theta_I(R_0) &= \min_{\theta, x_i} \theta \\
\text{s.t. } \mathcal{E}_j\left(\sum_{i=1}^{n} R_i x_i\right) &\geq \mathcal{E}_j(R_0),\ j = 1, \ldots, J, \\
\mathcal{D}_k\left(\sum_{i=1}^{n} R_i x_i\right) &\leq \theta \cdot \mathcal{D}_k(R_0),\ k = 1, \ldots, K, \\
\sum_{i=1}^{n} x_i &= 1,\ x_i \geq 0,\ 1 \geq \theta \geq 0.
\end{aligned}$$

It can be shown that for convex deviation measures and concave return measures this test is stronger than the standard one, i.e. for any benchmark R_0 it holds $\theta_{VRS}(R_0) \geq \theta_I(R_0)$ for arbitrary $R_0 \in \mathcal{X}$, see Branda [6]. We can obtain a mean-deviation efficiency test by restricting the number of considered inputs and outputs to one per class.

3.3 Input-output oriented diversification-consistent tests

We assume that $\mathcal{E}_j(R_0)$ is positive for at least one j. An input-output oriented test where inefficiency is measured with respect to the inputs and outputs separately can be formulated as follows

$$\begin{aligned}
\theta_{I-O}(R_0) &= \min_{\theta, \varphi, x_i} \frac{\theta}{\varphi} \\
\text{s.t. } \mathcal{E}_j\left(\sum_{i=1}^{n} R_i x_i\right) &\geq \varphi \cdot \mathcal{E}_j(R_0),\ j = 1, \ldots, J, \\
\mathcal{D}_k\left(\sum_{i=1}^{n} R_i x_i\right) &\leq \theta \cdot \mathcal{D}_k(R_0),\ k = 1, \ldots, K, \\
\sum_{i=1}^{n} x_i &= 1,\ x_i \geq 0,\ \varphi \geq 1,\ 1 \geq \theta \geq 0.
\end{aligned}$$

The benchmark investment opportunity R_0 is efficient if and only if there is no other investment opportunity with lower or equal deviations and higher or equal returns with strict inequalities for at least one group. The optimal value can be seen as a ratio of minimal necessary improvement (=decrease) of inputs and minimal necessary improvement (=increase) of outputs to reach the efficiency.

The following reformulation was derived by Branda [7]. If we set $1/t = \varphi$ and substitute $\tilde{x}_i = tx_i$, $\tilde{\theta} = t\theta$, and $\tilde{\varphi} = t\varphi$, the decision variables $\tilde{\varphi}$ and t can be omitted resulting into an input oriented DEA test with nonincreasing return to scale (NIRS):

$$\begin{aligned}
\theta_{I-O}(R_0) = \min_{\tilde{\theta},\tilde{x}_i} & \tilde{\theta} \\
\text{s.t. } \mathcal{E}_j\left(\sum_{i=1}^n R_i\tilde{x}_i\right) &\geq \mathcal{E}_j(R_0),\ j=1,\dots,J, \\
\mathcal{D}_k\left(\sum_{i=1}^n R_i\tilde{x}_i\right) &\leq \tilde{\theta}\cdot\mathcal{D}_k(R_0),\ k=1,\dots,K, \\
\sum_{i=1}^n \tilde{x}_i &\leq 1,\ \tilde{x}_i\geq 0,\ 1\geq\tilde{\theta}\geq 0.
\end{aligned}$$

Note that it is important for the reformulation that all inputs $\mathcal{D}_k$ and all outputs $\mathcal{E}_j$ are positively homogeneous. Since the general deviation measures are convex and the return measures are assumed to be concave, we obtained a convex programming problem. Similar class of input oriented tests was proposed by Lamb and Tee [15], where the assumption of NIRS was suggested for financial applications. Note that the input-output oriented test is stronger than the input oriented test, i.e. we obtain the relation between the optimal values $\theta_I(R_0) \geq \theta_{I-O}(R_0)$ for arbitrary $R_0 \in \mathcal{X}$.

4 Numerical results

In this section, we employ the DEA tests proposed above to access efficiency of 46 US industry representative portfolios observed monthly from January 2002 to December 2011, see Branda and Kopa [9] for details. We consider discretely distributed returns r_{is}, $s = 1,\dots,S$ with equal probabilities $1/S$, in our case $S = 120$. For general continuous distributions we can obtain similar problems using sample approximation technique, cf. Branda [5]. CVaR deviation on level α can be then formulated as

$$\mathcal{D}_\alpha^S\left(\sum_{i=1}^n R_i x_i\right) = \min_{\xi\in\mathbb{R}} \frac{1}{S}\sum_{s=1}^S \max\left\{(\sum_{i=1}^n x_i r_{is} - \xi), \frac{\alpha}{1-\alpha}(\xi - \sum_{i=1}^n x_i r_{is})\right\}$$

We consider $\mathcal{D}_\alpha^S$ for various levels $\alpha_k \in (0,1), k = 1,\dots,K$, in particular $\alpha_k \in \{0.75, 0.9, 0.95, 0.99\}$ and $K = 4$, as the inputs and the expectation as an output, i.e. $J = 1$ and $\mathcal{E}_1(X) = \mathbb{E}X$. This leads to the following linear programming problem for the input-output oriented test:

$$\begin{aligned}
\theta_{I-O}(R_0) = \min_{\theta,x_i,u_{sk},\xi_k} & \theta \\
\text{s.t. } \sum_{i=1}^n \mathbb{E}[R_i]x_i &\geq \mathbb{E}[R_0], \\
\frac{1}{S}\sum_{s=1}^S u_{sk} &\leq \theta\cdot\mathcal{D}_{\alpha_k}^S(R_0),\ k=1,\dots,K, \\
u_{sk} &\geq \left(\sum_{i=1}^n x_i r_{is} - \xi_k\right),\ s=1,\dots,S, k=1,\dots,K, \\
u_{sk} &\geq \frac{\alpha_k}{1-\alpha_k}\left(\xi_k - \sum_{i=1}^n x_i r_{is}\right),\ s=1,\dots,S, k=1,\dots,K, \\
\sum_{i=1}^n x_i &\leq 1,\ x_i\geq 0,\ i=1,\dots,n.
\end{aligned}$$

Similar reformulations were obtained by Branda [6, 7] for the input oriented and input-output oriented DEA tests with general probabilities. However, no numerical comparison of these approaches was reported in the previous papers.

We solved the DEA problems by the CPLEX 12.1 solver using the modeling system GAMS 23.2. Selected efficiency scores can be found in Table 1. We selected only the seven portfolios which are

efficient according to the VRS DEA test which is the weakest. The stronger DC tests identified only the portfolio Coal as efficient. Most of the VRS efficient portfolios are highly ranked also by the DC tests. The only exception is Drugs, which is ranked by the input-output oriented DC test very badly. We also computed correlations between the optimal values. Perhaps surprisingly we observed higher correlation between the efficiency scores of the VRS and input oriented test (0.97) than between the DC tests (0.70). The simplest explanation is the common orientation on inputs. Ranking of all representative portfolios can be found in Table 2.

	Food	Smoke	Hshld[1]	Drugs	Mines	Coal	Meals
VRS	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DC Inp	0.93	0.87	0.87	0.91	0.83	1.00	0.86
DC I-O	0.65	0.87	0.55	0.27	0.83	1.00	0.84

Table 1 Efficient industry representative portfolios and scores

5 Conclusions

In this paper, we have compared several approaches to efficiency of investment opportunities based on DEA. The traditional DEA test with variable return to scale has been shown much weaker that the extensions with diversification. These diversification consistent tests identified only one portfolio as efficient. However, the resulting rankings are significantly different. The diversification-consistent tests can be also seen as superefficiency models for the standard VRS DEA tests saying how much the benchmark inputs and outputs need to be improved to reach the efficiency frontier. Future research will be devoted to multiperiod–dynamic extensions.

	Agric	Food	Soda	Beer	Smoke	Toys	Fun	Hshld	Clths	Hlth
VRS	18	1	17	8	1	30	42	1	13	26
DC Inp	19	2	21	8	4	32	42	4	13	25
DC I-O	13	8	11	14	2	37	27	15	7	42
	MedEq	Drugs	Chems	Rubbr	Txtls	BldMt	Cnstr	Steel	FabPr	Mach
VRS	21	1	22	36	46	39	38	45	44	26
DC Inp	15	3	27	34	45	38	39	46	44	31
DC I-O	29	35	18	25	38	26	30	35	34	19
	ElcEq	Autos	Aero	Ships	Guns	Gold	Mines	Coal	Oil	Util
VRS	31	40	20	10	28	14	1	1	11	9
DC Inp	33	41	23	18	30	11	7	1	12	9
DC I-O	21	41	15	9	22	5	4	1	6	12
	Telcm	PerSv	BusSv	Comps	Chips	LabEq	Paper	Boxes	Trans	Whlsl
VRS	24	29	25	35	41	33	23	15	16	19
DC Inp	19	29	24	36	40	35	22	16	14	17
DC I-O	40	32	39	23	45	33	28	10	17	20
	Rtail	Meals	Insur	RlEst	Fin	Other				
VRS	12	1	34	43	37	32				
DC Inp	10	6	28	43	37	25				
DC I-O	24	3	43	31	44	46				

Table 2 Ranking of the industry representative portfolios

Acknowledgements

This work was supported by the Czech Science Foundation under the Grants P402/12/0558, P402/13/03749P.

References

[1] Banker, R.D., Charnes, A., Cooper, W.: Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis. *Management Science* **30** (9) (1984), 1078–1092.

[2] Basso, A., Funari, S.: A data envelopment analysis approach to measure the mutual fund performance. *European Journal of Operational Research* **135** (3) (2001), 477–492.

[3] Branda, M.: Third-degree stochastic dominance and DEA efficiency – relations and numerical comparison. *Proceedings of the 29th International Conference on Mathematical Methods in Economics 2011*, M. Dlouhý, V. Skočdopolová eds., University of Economics in Prague, Jánská Dolina, Slovakia (2011), 64–69.

[4] Branda, M.: Chance constrained problems: penalty reformulation and performance of sample approximation technique. *Kybernetika* **48** (1) (2012), 105–122.

[5] Branda, M.: Stochastic programming problems with generalized integrated chance constraints. *Optimization* **61** (8) (2012), 949–968.

[6] Branda, M.: Diversification-consistent data envelopment analysis with general deviation measures. *European Journal of Operational Research* **226** (3) (2013), 626–635.

[7] Branda, M.: Reformulations of input-output oriented DEA tests with diversification. *Operations Research Letters* **41** (5) (2013), 516–520.

[8] Branda, M., Kopa, M.: DEA-risk efficiency and stochastic dominance efficiency of stock indices. *Czech Journal of Economics and Finance* **62** (2) (2012), 106–124.

[9] Branda, M., Kopa, M.: On relations between DEA-risk models and stochastic dominance efficiency tests. *Central European Journal of Operations Research*, available online (2013). DOI: 10.1007/s10100-012-0283-2

[10] Briec, W., Kerstens, K., Lesourd, J.-B.: Single period Markowitz portfolio selection, performance gauging and duality: a variation on the Luenberger shortage function. *Journal of Optimization Theory and Applications* **120** (1) (2004), 1–27.

[11] Charnes, A., Cooper, W., Rhodes, E.: Measuring the efficiency of decision-making units. *European Journal of Operational Research* **2** (1978), 429–444.

[12] Chen, Z., Lin, R.: Mutual fund performance evaluation using data envelopment analysis with new risk measures. *OR Spectrum* **28** (2006), 375–398.

[13] Cooper, W.W., Seiford, L.M., Zhu, J.: *Handbook on data envelopment analysis*, Springer, New York, 2011.

[14] Joro, T., Na, P.: Portfolio performance evaluation in a mean–variance–skewness framework. *European Journal of Operational Research* **175** (2006), 446–461.

[15] Lamb, J.D., Tee, K-H.: Data envelopment analysis models of investment funds. *European Journal of Operational Research* **216** (3) (2012), 687–696.

[16] Lozano, S., Gutiérrez, E.: Data envelopment analysis of mutual funds based on second-order stochastic dominance. *European Journal of Operational Research* **189** (2008), 230–244.

[17] Markowitz, H.M.: Portfolio selection. *The Journal of Finance* **7** (1) (1952), 77–91.

[18] Markowitz, H.M.: *Portfolio selection: efficient diversification in investments.* John Wiley & Sons, Inc., New York, 1959.

[19] Murthi, B.P.S., Choi, Y.K., Desai, P.: Efficiency of mutual funds and portfolio performance measurement: a non-parametric approach. *European Journal of Operational Research* **98** (2) (1997), 408–418.

[20] Rockafellar, R.T., Uryasev, S., Zabarankin, M.: Generalized deviations in risk analysis. *Finance and Stochastics* **10** (2006), 51–74.

Multiple criteria evaluation of the activities criticalness in the Project management

Brožová Helena, Bartoška Jan, Šubrt Tomáš, Rydval Jan[1]

Abstract. The primary aim of project management is to ensure completion of the project under many constraints as scope, time, quality and budget and optimal allocation of necessary recourses. The crucial question is to find out what can cause delaying of the project due date or failure of project objectives. Some authors try to analyse activity parameters called criticalness.
Criticalness usually expressed by Criticality Index reflects the risk of influencing project due date by this activities. Another view via project simulation is how often a particular activity should be element of the critical path. All of these methods are primary derived from activity duration, type of links between activities and type of constraints. In the reality many different criteria (soft or hard) should influence criticalness of the activity. Criticalness in our approach is defined in a slightly different way.
We derive this parameter from the character of activity duration (known or random), resource assignments (in right time and right place), activity costs and number of parallel activities. In this work we suppose different multiple criteria approaches to criticalness calculations and we compare multiplicative and additive models and DEA.

Keywords: Project management, activity, criticalness, multiple criteria, evaluation.

JEL Classification: O22, C44
AMS Classification: 90B50, 90B99

1 Introduction

The principal purpose of project management is a successful implementation of a project: finishing the project within planned time, with planned costs and with a fulfilled objective. The success of the project is conditioned by the level of recognition of weak spots in the project. The criticalness of project activities, such as the level of amount and strength of weak spots in the project, is not only given by the surroundings and environment of the project but also by internal arrangement and structure of the project.

The project risk deals with risk management, however, without a quantitative view of the sequence and arrangement of project activities and its other quantitative qualities. In the area of risk management a number of different techniques and approaches have been derived in order to reduce the risk of a project and its partial activities [10]. The criticality of project activities is often defined from a time perspective only, using stochastic approaches ([1]; [4]), fuzzy sets methods ([3]; [12]) or using the findings of a network analysis ([7]; [2]). Authors [7] mention that ignoring the impact of noncritical activities which may easily become critical, is the most frequent criticism of project time analysis methods. Another point of view of activities criticalness is given by the structure of relations in the project. In [1] or also [11] deal with a stochastic analysis of a project network where the criticality of activities in the project is derived from the relation between activity duration and the whole project, and on the basis of a number of resources used for an activity and the whole project.

A deeper insight into the issue is presented by [1], who defines the criticality of the activity as uncertainty given by activity duration and a number of resources. A different approach to a project network has been adopted by [9], who analyse the growth or decline in the criticality of activities in the project during network structure changes. A number of preceding and following activities always has a direct impact on the criticality of the activity. Authors [9] claim that if a project network is composed of several parallel on-going chains, a lower variability and criticality of activities in the project can be expected. The criticalness of activities in the project and search for its weak spots needs to be accepted from various perspectives. It is impossible to expect a generally valid principle to determine the criticality of activities. Each project and each situation in the project is unique.

The main aim of the paper is to introduce heuristic based approach finding weak points in project completion process, identifying key activities for meeting project objectives. In this work we deal activities criticalness iden-

[1] Czech University of Life Sciences Prague, Faculty of Economics and Management,
Department of Systems Engineering, Kamýcká 129, 165 21 Praha 6 – Suchdol, brozova@pef.czu.cz

tification, we suppose different multiple criteria approaches to criticalness calculations and we compare multiplicative and additive models and DEA.

2 Indicators of activities criticalness and multiple criteria decision making methods

Estimation of overall criticalness of the project activities are based on the multiple criteria decision making methods using five indicators of the criticalness.

- **Time criticalness**

As a measure of the time criticalness of the activities can be used or time duration or indicators related to the project duration.

$$t_I \quad \text{or} \quad ct_I = \frac{t_I}{T} \tag{3}$$

where
t_I, t_K are the duration of the activities I, K,
T is the project duration,
ct_I is the time criticalness of the activity I,

- **Probability of Critical path**

This indicator evaluates the probability the activity will lie on critical pass related to the project structure.

$$p_{1j} = \frac{1}{h_1} \text{ and } p_{ij} = \frac{\sum_k p_{ki}}{h_i}, k \neq 1 \tag{1}$$

where
p_{1j}, p_{ki}, p_{ij} are the probability that the activities $1j$, ki, ij will be on the critical path,
h_i is the number of activities starting in the node i,

- **Slack criticalness**

The measure of the slack criticalness can be slacks or indicator related to maximal slack.

$$s_I \quad \text{or} \quad cs_I = \frac{\max_K s_K - s_I}{\max_K s_K} \tag{4}$$

where
s_I, s_K are the slack of the activities I, K,
cs_I is the slack criticalness of the activity I,

- **Cost criticalness**

The cost criticalness of the activities is based on the cost or indicators related to the project cost.

$$c_I \quad \text{or} \quad cc_I = \frac{c_I}{C} \tag{5}$$

where
C is the total project cost,
c_I, c_K are the cost of the activities I, K,
cc_I is the cost criticalness of the activity I,

- **Work criticalness**

As the indicator of the work criticalness of the activities can be used the work amount or indicators related to the project work.

$$v_I \quad \text{or} \quad cv_I = \frac{v_I}{V} \tag{6}$$

where
v_I, v_K are the total amount of work of the activities I, K,
cv_I is the work criticalness of the activity I,
V is the total amount of work of the project.

These indicators are used as input for the multiple criteria decision making approaches. We create and use multiplicative and additive multiple criteria models [5] and DEA model without outputs for evaluation of the project activities criticalness.

- **Multiplicative model**

In this approach all criteria evaluations are multiplied using formula $w_I = \prod_p r_{Ip}$, where w_I is global evaluation of the activities criticalness, r_{Ip} is evaluation of each components of activities criticalness and p is number of used components of criticalness.

- **Additive model**

In this approach all criteria evaluations are multiplied using formula $w_I = \sum_p r_{Ip}$, where w_I is global evaluation of the activities criticalness, r_{Ip} is evaluation of each components of activities criticalness and p is number of used components of criticalness.

- **Additive model with weights**

In this approach all criteria evaluations are multiplied using formula $w_I = \sum_p u_p r_{Ip}$, where w_I is global evaluation of the activities criticalness, r_{Ip} is evaluation of each components of activities criticalness, p is number of used components of criticalness and u_p is the weight of the p component of criticalness.

- **Data Envelopment Analysis Method**

Data Envelopment Analysis (DEA) serves for evaluation of units against the best unit. The DEA is a non-linear programming model for the estimation of an efficiency of units, based on the relationship between multiple outputs and multiple inputs. The DEA measure of the efficiency of any DMU is obtained as the maximum of a ratio of weighted outputs to weighted inputs, subject to the condition that the similar ratio for every units is less than or equal to 1. The simplest DEA model assumes constant returns to scale, this model is called the CCR model, according to its authors, Charnes, Cooper, and Rhodes [6]. Primal model is:

$$\Phi_H = \sum_{j=1}^{n} u_{jH} y_{jH} \rightarrow MAX$$

subject to

$$\begin{aligned} &\sum_{i=1}^{m} v_{iH} x_{iH} = 1 \\ &-\sum_{i=1}^{m} v_{iH} x_{ik} + \sum_{j=1}^{n} u_{jH} y_{jk} \leq 0,\ k = 1, 2, \ldots, p, \\ &u_{jH} \geq 0,\ j = 1, 2, \ldots, n, \\ &v_{iH} \geq 0,\ i = 1, 2, \ldots, m. \end{aligned} \tag{7}$$

Supposed DEA model has five inputs t_I, p_I, s_I, c_I, v_I and one output y_I meaning an expected criticalness of the activity I. At the beginning of the calculation all expected activities criticalness are set as $y_I = 1$. (The output of the DEA model without outputs is always unitary [8]) It means, that we suppose the equal criticalness of all activities. Using the CCR model, the real criticalness of a particular activity I is calculated using the primal output oriented model:

The relative efficiencies of the output orientated DEA model generally show how much an output must be increased. The more an output should increase, the higher criticalness this activity has, i. e. higher inefficiency means higher criticalness of particular activity. Therefore, these activities represent critical parts of a project, which require more attentions of a project manager. Those activities that lie on the efficient frontier do not represent critical parts of a project.

Authors used Efficiency Measurement System (EMS) SW for the calculation of the DEA model.

3 Evaluation of the project activities criticalness

To evaluate a project as a whole, from the point of view of its activity criticalness, is not easy. An unambiguous and fully sufficient approach is still non-existent. Activities on a critical path are considered the most threatening activities. The finding of a critical path and a follow-up analysis of reserves does not always suffice. In practice it can often be seen that even activities off a critical path have an extreme impact on the course and success of a project. To demonstrate the contribution of derived indicators of project activity criticalness, the following small-scale project is used as an illustrative example.

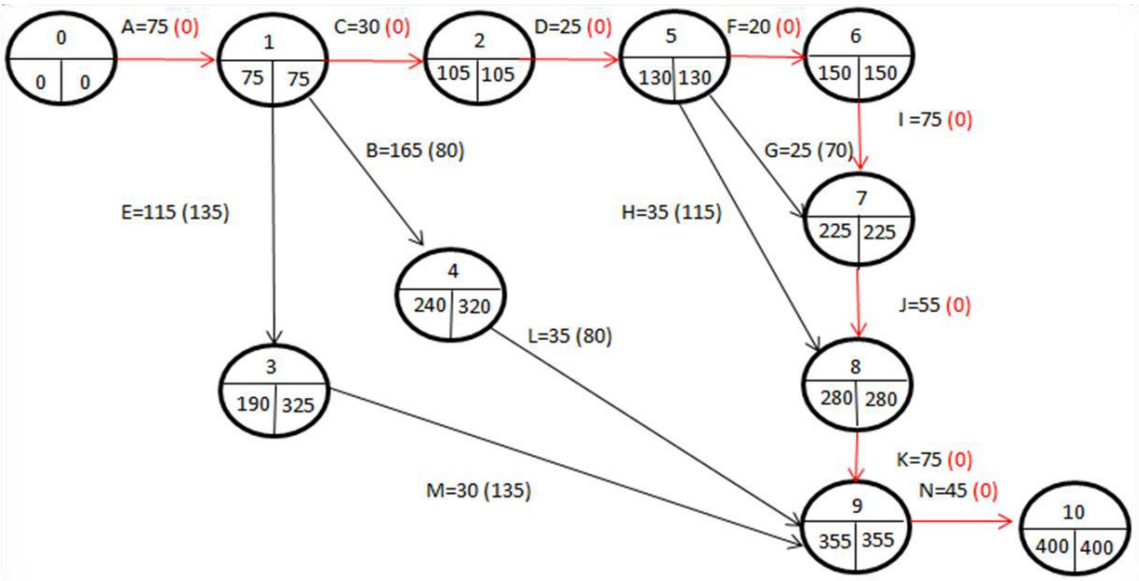


Figure 1 Small-scale project with the indication of a critical path.

A critical path of a project (**Figure 1**) is composed of activities A, C, D, F, I, J, K and N. If we derive proposed criticalness indicators for all activities (**Table 1**), some activities, which do not lie on a critical path, can be regarded, as concerns their values, as a greater threat to the project than activities which lie on a critical path. Several activities have much higher criticalness towards the project than a critical path method reveals.

Task	Days	Time criticalness	Probability of critical path	Slack	Slack criticalness	Cost	Cost criticalness	Work	Work criticalness
A	75	0.188	1	140	1	1350	0.121	450	0.049
B	165	0.413	0.33	60	0.407	990	0.088	495	0.053
C	30	0.075	0.33	140	1	630	0.056	630	0.068
D	25	0.063	0.33	140	1	175	0.016	175	0.019
E	115	0.288	0.33	5	0	690	0.062	345	0.037
F	20	0.050	0.11	140	1	60	0.005	60	0.006
G	25	0.063	0.11	70	0.481	150	0.013	150	0.016
H	35	0.088	0.11	25	0.148	210	0.019	210	0.023
I	75	0.188	0.11	140	1	1575	0.141	1575	0.170
J	55	0.138	0.22	140	1	1155	0.103	1155	0.125
K	75	0.188	0.33	140	1	2025	0.181	2025	0.218
L	35	0.088	0.33	60	0.407	875	0.078	875	0.094
M	30	0.075	0.33	5	0	360	0.032	180	0.019
N	45	0.113	1	140	1	945	0.084	945	0.102
MADM 1		**MAX**	**MAX**		**MAX**		**MAX**		**MAX**
DEA	**INPUT**		**INPUT**	**INPUT**		**INPUT**		**INPUT**	

Table 1 Activity criticalness indicators, input values for MADM.

Activities B and L which are not on a critical path and have a time reserve have, in regard to MADM (**Table 2**), higher criticalness values than activities that lie on a critical path. The order of activities is determined according to a total rank. In particular, activity B surpassed in its values other three activities on a critical path. The significance and impact of activity B towards the whole project is much higher than it can be assumed from the results of a critical path method.

Task	Multiplicative model	Rank	Aditive model	Rank	Additive model with weights	Rank	CCR-O super	Rank	Suma of ranks	Total rank
A	0.00110	2	2.357	1	0.395	1	282%	1	5	1
K	0.00245	1	1.917	3	0.325	3	266%	2	9	2
N	0.00097	3	2.299	2	0.384	2	183%	5	12	3
J	0.00039	5	1.585	5	0.252	5	200%	4	19	4
I	0.00049	4	1.608	4	0.260	4	100%	9	21	5
B	0.00026	6	1.292	8	0.220	7	243%	3	24	6
C	0.00009	7	1.529	6	0.236	6	133%	6	25	7
D	0.00001	9	1.427	7	0.211	8	117%	8	32	8
L	0.00009	8	0.997	10	0.174	9	130%	7	34	9
F	0.0000002	12	1.172	9	0.161	10	40%	13	44	10
E	0	13	0.716	11	0.136	11	100%	9	44	10
G	0.0000007	10	0.684	12	0.101	12	90%	11	45	12
H	0.0000006	11	0.387	14	0.065	14	56%	12	51	13
M	0	13	0.457	13	0.088	13	37%	14	53	14

Table 2 Results of MADM models and the rank of activities based on criticalness.

Deeper consideration should be given to the model DEA (CCR-O super) which introduces certain sturdiness and better results interpretation to activity criticalness evaluation. The appearing value *100%* for some activities can be interpreted as the activity which is at its inputs (Time criticalness, Probability of Critical path, Slack criticalness, Cost criticalness, Work criticalness) adequately critical toward a project and which is at its limit of criticalness towards the project and can be regarded as critical. Therefore, a significant finding could again be value *243%* for activity B which does not lie on a critical path, has a time reserve and is the third most critical activity towards the project based on the model CCR-O super.

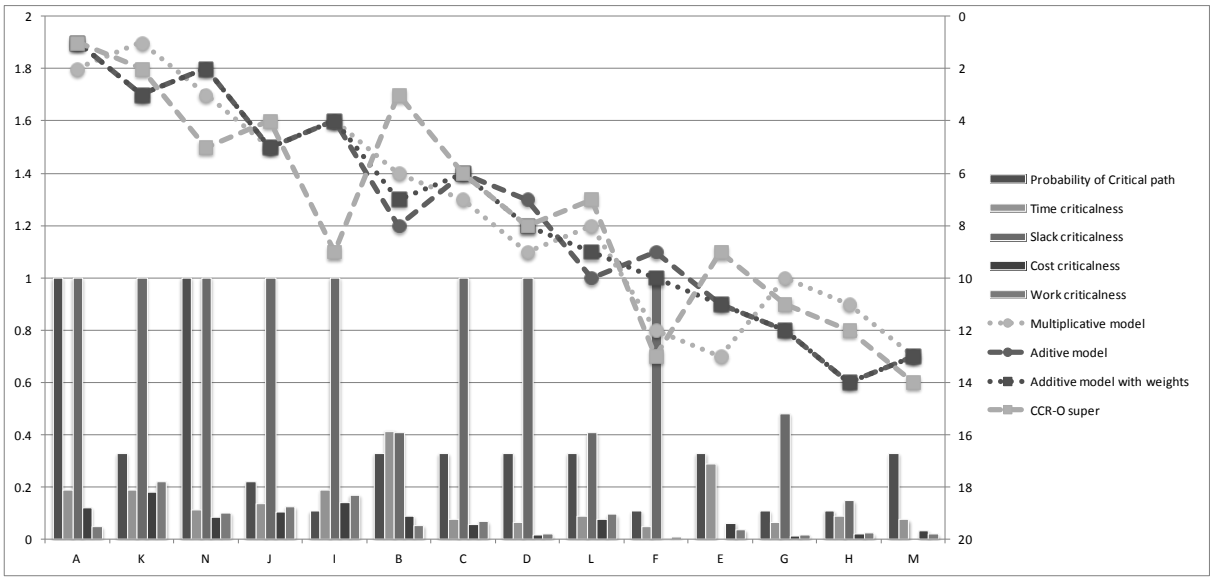


Figure 2 Comparison of activity criticalness indicators and MADM model results.

The final values of indicators and the rank obtained using MADM can be further illustrated in a graph (**Figure 2**), which shows that activity criticalness in the project is versatile and across activities descending, and it does not really correspond to the results of a critical path method.

4 Conclusion

The paper proposes indicators of project activity criticalness for the evaluation of activities from the perspective of their significance towards a project as a whole. Moreover, the use of the indicators as inputs for multiplicative, additive and DEA model was carried out. The calculated values of the indicators and the results of models for a small-scale project, used as an illustrative example support the fact that some activities which do not lie on a critical path, have a higher criticalness towards a project than activities that lie on a critical path. The significance and impact of activities in a project cannot be derived solely from the incidence of activities on a critical path.

Acknowledgements

The research is supported by the Internal Grant Agency of the University of Life Sciences Prague – project IGA PEF 20131028 - Multicriterial Analysis of Criticality of Project Activities.

References

[1] Bowers, J.: Identifying Critical Activities in Stochastic Resource Constrained Networks. *Omega - International Journal of Management Science* **24** (1996), 37-46.
[2] Chanas, S., Zielinski, P.: On the hardness of evaluating criticality of activities in a planar network with duration intervals. *Operation Research Letters* **31** (2003), 53–59.
[3] Chen, C. T., Huang, S. F.: Applying fuzzy method for measuring criticality in project network. *Information Sciences* **177** (2007), 2448–2458.
[4] Cruz, S., García, J., Herrerías, R.: Stochastic models alternative to the classical PERT for the treatment of the risk: mesokurtic and of constant variance. *Central European Journal of Operations Research* **7** (1999), 159-175.
[5] Hwang, Ch-L., Yoon, K.: *Multiple Attribute Decision Making*. Springer Verlang Berlin Heidelberg, New York, 1981.
[6] Charnes, A., Cooper, W.W., Rhodes, E.: Measuring the efficiency of decision-making units, *European Journal of Operational Research* **2** (1978), 429–444.
[7] Gong, D., Rowings, J. E.: Calculation of safe float use in risk-analysis-oriented network scheduling. *International Journal of Project Management* **13** (1995), 187-194.
[8] Lovell, C. A. K., Pastor, J. T.: Radial DEA models without inputs or without outputs. *European Journal of Operational Research* **1** (1999), 46-51.
[9] Madadi, M., Iranmanesh, H.: A management oriented approach to reduce a project duration and its risk (variability). *European Journal of Operational Research* **219** (2012), 751-761.
[10] Wideman, R.M.: *Project and Program Risk Management*. Project Management Institute, 1992.
[11] Williams, T. M.: Criticality in stochastic networks. *Journal of Operational Research Society* **43** (1992), 353-357.
[12] Yakhchali, S. H.: A path enumeration approach for the analysis of critical activities in fuzzy networks. *Information Sciences* **204** (2012), 23-35.

On Graphical Optimization of Linear Programming Models in the Column Space

Helena Brožová[1], Milan Houška[2]

Abstract. The necessary conditions for representing and solving a linear programming model graphically are well known. The model should contain up to two decision variables (a number of constraints could be unlimited, but finite) or up to two constraints (a number of decision variables could be unlimited, but finite). In the first case, we solve the model graphically in a so called "Row Space", where the axes of the graph represent decision variables. In the other case, we solve the model in a so called "Column Space", where the axes represent individual constraints.
In this paper, we focus on the optimization of linear programming models in the Column Space. There is a standard procedure to solve it, but it can be used if and only if all cost coefficients in the objective function are positive or zero. We make the procedure more general and show how to carry out the graphical optimization in the Column Space correctly, even if at least one cost coefficient is negative. We also demonstrate the proposed algorithm on a numerical example.

Keywords: linear programming, graphical optimization, column space.

JEL Classification: C61
AMS Classification: 90C05

1 Introduction

Linear programming (LP) is one of the most important areas in Operational Research. Many authors use the LP models to solve practical problems in various application domains, e.g. in economics [1], industry [2, 3], agriculture [4] or logistics [5, 6]. Linear programming models are also used as auxiliary models in other branches of Operational Research, e.g. in Project Management [7], Data Envelopment Analysis [8, 9] or Game Theory [10, 11].

Linear programming models of a small size (up to two decision variables or two constraints) could be solved using a graphical optimization technique. The graphical optimization procedure for the LP models containing two variables in the Row Space is commonly used and always described in textbooks dealing with Operational Research. Surprisingly, the procedure for the LP models containing two constraints and more than two decision variables (Column Space) is usually omitted there. The procedure is described in Simmonard [12] partially, but we have not found any other sources describing the procedure systematically. Withal, the graphical optimization techniques play important role at least when one starts dealing with the LP models [13].

We noticed that the procedure for the graphical optimization of the LP models in the Column Space is incomplete. It does not cover all cases, which may occur when a practical problem is solved. According to our best knowledge, the problem of negative cost coefficients in the objective function has not been mentioned and solved in literature. In this paper we enhance the standard procedure of the Column Space to be correct also for such coefficients, and provide necessary proofs. Finally, we demonstrate our approach on an illustrative numerical example.

2 Optimization in the Column Space

Let's have a general linear programming model containing two constraints

[1] Czech University of Life Sciences Prague, Faculty of Economics and Management, Department of Systems Engineering, Kamýcká 129, 165 21, Prague 6, Czech Republic, brozova@pef.czu.cz.
[2] Czech University of Life Sciences Prague, Faculty of Economics and Management, Department of Systems Engineering, Kamýcká 129, 165 21, Prague 6, Czech Republic, houska@pef.czu.cz.

$$\begin{aligned} &a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ &a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ &z = c_1x_1 + c_2x_2 + \ldots + c_nx_n \rightarrow MAX \\ &x_j \geq 0,\ j = 1,2,\ldots,n \end{aligned} \tag{1}$$

If the constraints are formulated as inequations, they have to be transformed to equations through slack variables added into the constraints. Then, the model is solvable in the Column Space using the following procedure [12].

Step 1: Elimination of different cost coefficients in the objective function

We transform the column vectors from the **A** matrix to ones, which reflect different cost coefficients of their variables. For *n* variables, we receive *n* transformed vectors denoted as $\boldsymbol{\alpha}_j$ as

$$\boldsymbol{\alpha}_j = \left(\frac{a_{1j}}{c_j}; \frac{a_{2j}}{c_j} \right)^T,\ j = 1,2,\ldots,n \tag{2}$$

After this operation we change the sense of the **A** matrix coefficients from "how many units of constraints are satisfied with one unit of the variable" to "how many units of constraints are satisfied with an amount of units of the variable obtained for one unit of the objective function".

Step 2: Identification of feasible solutions

Lemma 1. A feasible solution exists, if the cone of the column vectors exists that contains the right hand side vector.

Proof. If the column vectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ create feasible solution, the real numbers $x_1, x_2 \geq 0$ exist that $\mathbf{b} = \boldsymbol{\alpha}_1 x_1 + \boldsymbol{\alpha}_2 x_2$. Graphically, the cone of the vectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ has to contain the vector **b**. Any feasible solution does not exist, when no cone of the column vectors contains the right hand side vector.

Step 3: Optimization

Suppose that all coefficients of the objective function are positive and at least one feasible solution exists. If all column vectors lay in one half-plane and the feasible solution exists, the optimal solution exists. The column vectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ create feasible solution $\mathbf{b} = \boldsymbol{\alpha}_1 x_1 + \boldsymbol{\alpha}_2 x_2$ for the real numbers $x_1, x_2 \geq 0$. The equations $\mathbf{b} = \boldsymbol{\alpha}_1 x_1 + \boldsymbol{\alpha}_2 x_2$ for all feasible pairs of vectors have finite solutions and the problem is to select such a pair of vectors, for which the objective function $x_1 + x_2$ has the optimal value.

Lemma 2. Let the vectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ create a feasible solution, the coefficients of the objective function **c** = **1** and $\mathbf{b} = \boldsymbol{\alpha}_1 x_1 + \boldsymbol{\alpha}_2 x_2$ for $x_1, x_2 \geq 0$, then the objective function value is $z(\mathbf{x}) = \frac{\|\mathbf{b}\|}{\|\mathbf{p}\|}$, where the vector **p** is vector with the ending point P which is the intersection point of vector **b** and the abscissa between the ending points of the vectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$.

Proof. Suppose now that the vector **b** lies on the axis *X*, so **b** = (b_1, 0).

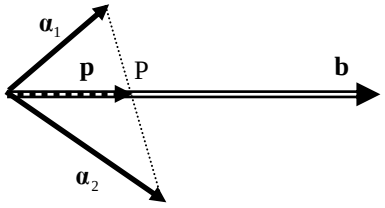


Figure 1 Combining the vectors

For the point *P* it holds $x_1 + x_2 = \frac{b_1}{p_1}$.

General case is transformed using rotation. As the rotation is isometry, for the general vectors **b** and **p** the following formula holds

$$z(\mathbf{x}) = x_1 + x_2 = \frac{\|\mathbf{b}\|}{\|\mathbf{p}\|} \tag{3}$$

This formula is also fulfilled regardless to the direction of vectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ and **b**.

The Lemma 2 shows that the optimal solution has to be selected according to the point P. If the point P is moving far from origin of coordinates system in the direction of the vector **b**, the value $z(\mathbf{x})$ is decreasing. The value $z(\mathbf{x})$ is increasing if the point P is moving to the origin.

Note: For all negative coefficients in the objective function this lemma is formulated similarly. The value of the objective function is $z(\mathbf{x}) = -\frac{\|\mathbf{b}\|}{\|\mathbf{p}\|}$ *If the point P is moving far from origin of coordinates system in the direction of the vector -**b**, the value z(**x**) is decreasing. The value z(**x**) is increasing if the point P is moving to the origin.*

3 Mixed negative and positive cost coefficients

A problem appears, when some of the cost coefficients in the objective function are positive and some of them are negative. In this case we cannot use the above-given procedure correctly. In Step 1, we would change the direction of the vector $\boldsymbol{\alpha}_j$, when the cost coefficient $\mathbf{c}_j$ is negative. It has negative impact on the Step 2, because we could omit some feasible combinations of variables, which apparently violate the Lemma 1, but in fact they are feasible. That is why we propose the following modification of the standard algorithm.

Step 1: Elimination of different cost coefficients in the objective function

Regarding to the Step 2, we should avoid redirecting the $\boldsymbol{\alpha}_j$ vectors. Thus we change the Eq. 2 to

$$\boldsymbol{\alpha}_j = \left(\frac{a_{1j}}{|c_j|}; \frac{a_{2j}}{|c_j|} \right)^{\mathrm{T}}, \; j = 1, 2, \ldots, n \tag{4}$$

and change the graphical notation of the vectors for the variables of the negative cost coefficients. Suppose that the cost coefficient of the variable x_1 is negative and the cost coefficient of the variable x_2 is positive. The vector $\boldsymbol{\alpha}_1'$ describes that the cost coefficient of variable x_1 is negative; the vector $\boldsymbol{\alpha}_1$ is shifted, so that the ending point of this vector is in the origin of the coordinates.

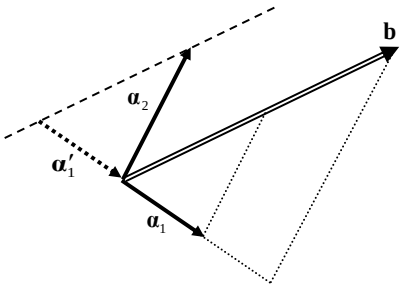


Figure 2 Graphical representation of the vectors of negative cost coefficients

After this modification, we can apply the Step 2 of the original algorithm to find all feasible solutions. But the Step 3 requires revisions as follows.

Step 3: Optimization

First we have to redirect the $\boldsymbol{\alpha}_j$ vectors for the variables of the negative cost coefficients. As the feasible solutions are determined with vectors of the equal orientation like in the **A** matrix, now we have to distinguish vectors increasing the objective function value (of the positive cost coefficient) or decreasing it (of the negative cost coefficient). Thus we recalculate back

$$\boldsymbol{\alpha}_j = \left(\frac{a_{1j}}{c_j}; \frac{a_{2j}}{c_j} \right)^{\mathrm{T}}, j = 1, 2, \ldots, n \tag{5}$$

Lemma 3. Let the vectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ create a feasible solution, the coefficients of the objective function $c_1 = 1$ and $c_2 = -1$ and $\mathbf{b} = \boldsymbol{\alpha}_1 x_1 + \boldsymbol{\alpha}_2 x_2$ for $x_1, x_2 \geq 0$, then the objective function value is $z(\mathbf{x}) = -\frac{\|\mathbf{b}\|}{\|\mathbf{p}\|}$, where the vector **p** is vector with the ending point P which is the intersection point of the vector **b** and the straight line defined by the ending points of the vectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$.

Proof. The vector **b** lies on the axis *X*, so **b**=(b_1, 0).

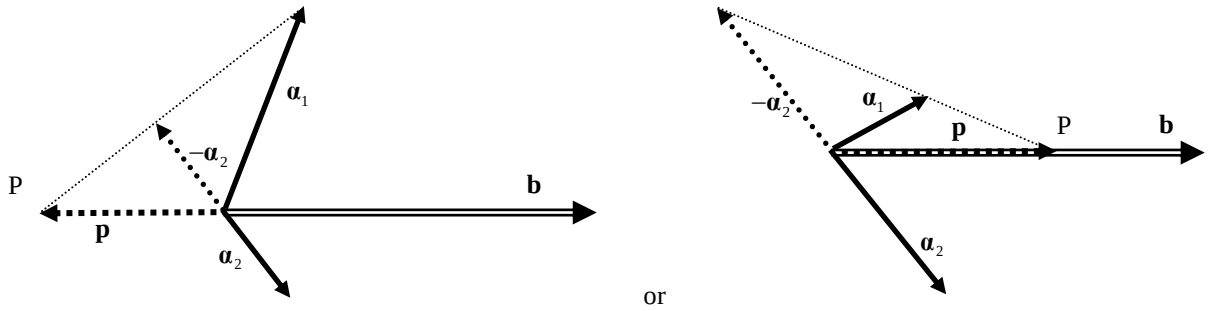


Figure 3 Combining the vectors for optimization

For the point *P* holds $x_1 - x_2 = \frac{b_1}{p_1}$. This value is positive or negative according to the value p_1.

If the point P does not exist, $x_1 - x_2 = 0$, the value p_1 is increasing without limits.

General case is received after isometry transformation - rotation. For the general vectors **b** and **p** the following formula depends on the direction of these vectors. If **b** and **p** have the same direction, the objective function value is

$$z(\mathbf{x}) = x_1 - x_2 = \frac{\|\mathbf{b}\|}{\|\mathbf{p}\|} \tag{6}$$

If **b** and **p** have the opposite direction, the objective function value is

$$z(\mathbf{x}) = x_1 - x_2 = -\frac{\|\mathbf{b}\|}{\|\mathbf{p}\|} \tag{7}$$

If the point P does not exist, $z(\mathbf{x}) = x_1 - x_2 = 0$.

Based on the Lemma 3 we can formulate a simple rule to order individual feasible solutions according to the values of the objective function. In descending order

- in the Quadrant I: from the nearest point P to the origin of coordinates system to the most distant one;
- then in the Quadrant III: from the most distant point P to the origin of coordinates system to the nearest one.

4 Numerical example

Using the above-described algorithm, we solve the following model

$$\begin{aligned} &4x_1 + 2x_2 - 0.5x_3 - x_4 \leq 16 \\ &2x_1 - x_2 + 2x_3 - x_4 \geq 6 \\ &z = 8x_1 + 2x_2 - 10x_3 - x_4 \rightarrow MAX \\ &x_{1,2,3,4} \geq 0 \end{aligned}$$

with added slack variables

$$\begin{aligned} &4x_1 + 2x_2 - 0.5x_3 - x_4 + s_1 = 16 \\ &2x_1 - x_2 + 2x_3 - x_4 - s_2 = 6 \\ &z = 8x_1 + 2x_2 - 10x_3 - x_4 + 0s_1 + 0s_2 \rightarrow MAX \\ &x_{1,2,3,4} \geq 0, s_{1,2} \geq 0 \end{aligned}$$

Step 1: Elimination of different cost coefficients in the objective function

Using the Eq. 4 we obtain

$$\boldsymbol{\alpha}_1 = \begin{pmatrix} 0.5 \\ 0.25 \end{pmatrix}; \boldsymbol{\alpha}_2 = \begin{pmatrix} 1 \\ -0.5 \end{pmatrix}; \boldsymbol{\alpha}_3 = \begin{pmatrix} -0.05 \\ 0.2 \end{pmatrix}; \boldsymbol{\alpha}_4 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}; \mathbf{s}_1' = \begin{pmatrix} \infty \\ 0 \end{pmatrix}; \mathbf{s}_2' = \begin{pmatrix} 0 \\ -\infty \end{pmatrix}$$

Step 2: Feasible combinations of variables in basis

The feasible combinations of the variables are highlighted in Figure 4.

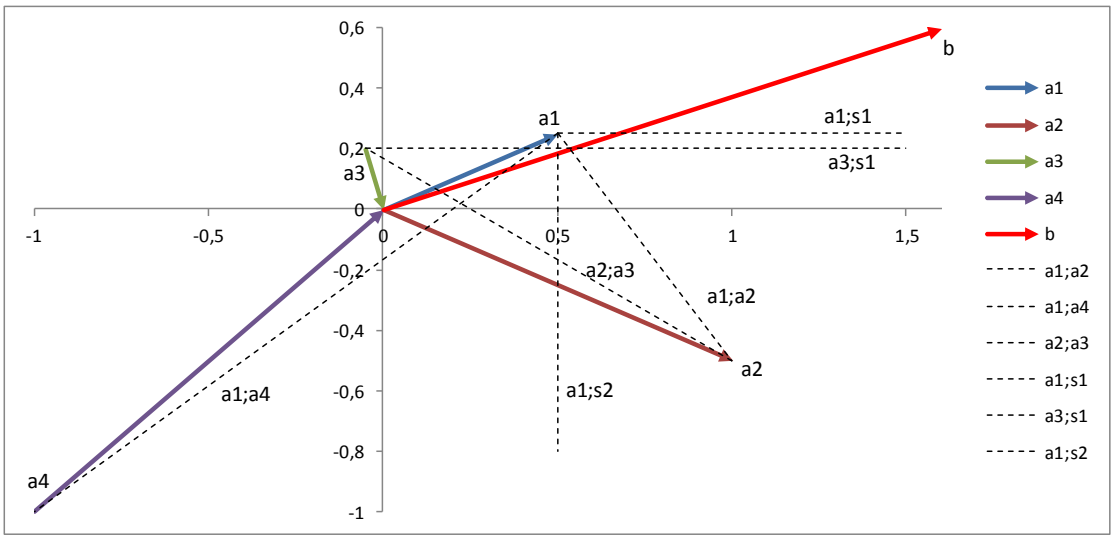


Figure 4 Feasible solutions

Step 3: Optimization

As the cost coefficients c_3 and c_4 are negative, the vectors $\mathbf{a_3}$ and $\mathbf{a_4}$ have to be redirected before optimization. Multiplying these vectors by (-1) we obtain

$$a_1 = \begin{pmatrix} 0.5 \\ 0.25 \end{pmatrix}; a_2 = \begin{pmatrix} 1 \\ -0.5 \end{pmatrix}; a_3 = \begin{pmatrix} 0.05 \\ -0.2 \end{pmatrix}; a_4 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}; s_1' = \begin{pmatrix} \infty \\ 0 \end{pmatrix}; s_2' = \begin{pmatrix} 0 \\ -\infty \end{pmatrix}$$

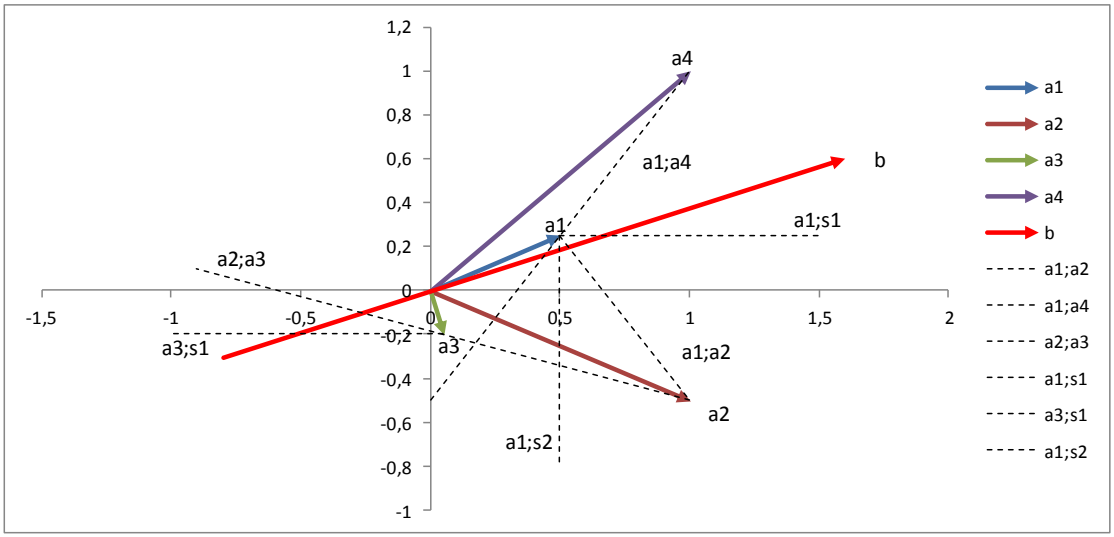


Figure 5 Optimization

Based on Figure 5, we can simply order the feasible basic solutions subject to descending values of the objective function

- in Quadrant I: $\mathbf{a_1}$; $\mathbf{a_4} - \mathbf{a_1}$; $\mathbf{s_2} - \mathbf{a_1}$; $\mathbf{a_2} - \mathbf{a_1}$; $\mathbf{s_1}$ (the closer to [0;0] the higher value of the objective function);
- in Quadrant III: $\mathbf{a_3}$; $\mathbf{s_1} - \mathbf{a_2}$; $\mathbf{a_3}$ (the closer to [0;0] the lower value of the objective function).

The vector of basic solution maximizing the objective function value is

$$\mathbf{x_B} = (x_1; 0; 0; x_4; 0; 0)^T$$

where the values of the variables x_1 and x_4 are calculated from the linear equations system

$$4x_1 - x_4 = 16$$
$$2x_1 - x_4 = 6$$

so $x_1 = 5$; $x_4 = 4$ and objective function value $z^{opt.} = 36$.

5 Conclusion

In this paper we provide the extension of the algorithm for graphical optimization of linear programming models in the Column Space. We showed how to solve the models, when both positive and negative values of the cost coefficients in the objective function occur. Of course, another way to overcome the problem is to construct the dual model, solve it in the Row Space, and finally re-interpret it back in terms of the primal model. We hardly recommend this way. We understand the graphical optimization in the Column Space as a regular algorithm, so we prefer to improve its imperfections rather than stop using it.

The authors know that there is another problem to solve: how to identify and recognize cases, when the objective function of the model is unbounded. The solution would be similar, but there are some difficulties, which could not be described and overcome in this paper. This is the topic for future work.

Acknowledgements

The authors thank to Professor Petr Němec, Head of the Department of Mathematics, Faculty of Engineering, Czech University of Life Sciences Prague, for his valuable comments to their work.

References

[1] Simic, V., and Dimitrijevic, B.: Risk explicit interval linear programming model for long-term planning of vehicle recycling in the EU legislative context under uncertainty. *Resources Conservation and Recycling* **73** (2013), 197-210.

[2] Catala, L. P., Durand, G. A., Blanco, A. M., and Bandoni, J. A.: Mathematical model for strategic planning optimization in the pome fruit industry. *Agricultural Systems*, **115** (2013), 63-71.

[3] Liu, S., and Papageorgiou, L. G.: Multiobjective optimisation of production, distribution and capacity planning of global supply chains in the process industry. *Omega-International Journal of Management Science*, **41** (2013), 369-382.

[4] Dejtanon, W., et al.: Extended deming's model and linear programming approach for agriculture land use for energy crops management: A case study of nakhon ratchasima, Thailand. In: *Proceedings of the International Conference on Advances Science and Contemporary Engineering*, 2012, 95-100.

[5] Cardoso, S. R., Barbosa-Povoa, A. P. F. D., and Relvas, S.: Design and planning of supply chains with integration of reverse logistics activities under demand uncertainty. *European Journal of Operational Research*, **226** (2013), 436-451.

[6] Garcia, J., et al.: Combining linear programming and automated planning to solve intermodal transportation problems. *European Journal of Operational Research*, **227** (2013), 216-226.

[7] Kone, O., Artigues, C., Lopez, P., and Mongeau, M.: Comparison of mixed integer linear programming models for the resource-constrained project scheduling problem with consumption and production of resources. *Flexible Services and Manufacturing Journal*, **25** (2013), 25-47.

[8] Tohidi, G., and Razavyan, S.: A circular global profit malmquist productivity index in data envelopment analysis. *Applied Mathematical Modelling*, **37** (2013), 216-227.

[9] Wang, X., Hockmann, H., and Bai, J. Technical efficiency and producers' individual technology: Accounting for within and between regional farm heterogeneity. *Canadian Journal of Agricultural Economics-Revue Canadienne D Agroeconomie*, **60** (2012), 561-576.

[10] Cabrera Garcia, S., Imbert Tamayo, J. E., Carbonell-Olivares, J., and Pacheco Cabrera, Y.: Application of the game theory with perfect information to an agricultural company. *Agricultural Economics-Zemedelska Ekonomika*, **59** (2013), 1-7.

[11] Nan, J., and Li, D.: Linear programming approach to matrix games with intuitionistic fuzzy goals. *International Journal of Computational Intelligence Systems*, **6** (2013), 186-197.

[12] Simmonard, M. *Programmation lineaire*. Dunod, Paris, 1962.

[13] Bartoska, J., Fejfar, J., and Jarkovska, M. (2011). The efficiency of linear programming teaching by means of system diagrams. In: *Proceedings of the 8th International Conference on Efficiency and Responsibility in Education* (Kvasnicka R. ed.). 2011, 15-22.

Empirical analysis of labour markets: An International comparison

Jan Brůha[1], Jiří Polanský[2]

Abstract. The goal of this paper is to document and summarize main stylized facts about labour market data in transition and developed countries. First, we report various correlations among labour market variables at different leads and lags and at various frequencies (growth rates and cycles), and compare them across countries. Second, we investigate the time-varying nature of these features paying attention to times of financial distress. We conclude that there are robust relations among the variables across countries and times at business cycle frequency, that the cyclical comovement between real output and selected labor market indicators is strong. On the other hand, correlations in growth rates are not similar across countries and times. The paper concludes with our assessment of what these findings imply for dynamic general-equilibrium macroeconomic models with labour market.

Keywords: Labour market modeling, stylised facts, time-varying correlations

JEL classification: E24, J21, J30
AMS classification: 62P20

1 Introduction

The goal of this paper is to document and summarize main stylized facts about labour market data in several countries. This may be useful for various reasons. First, one may be interested to know which data features are robust across countries and which depends on countries' characteristics, such as labour market regulation. Second, the results can be used to constructing a set of empirical checks when constructing structural models with explicit labour market blocks. Prominent examples of such structural models are Dynamic Stochastic General Equilibrium (DSGE) models with labour market frictions. In this paper, we are interested mainly in the second question and seek to find facts that are robust across countries and times and which a successful structural model should replicate.

An alternative check of a structural model is a comparison of its impulse responses with those based on theoretical models, mostly often structural vector autoregression (SVAR) models. We do not opt for this alternative. Firstly, this has been done countlessly before, and secondly, realistic and credible identification of SVARs is notoriously difficult if possible at all.[1] Hence we instead focus on correlation analysis to characterise stylized patterns at various frequencies (mainly at cycles), and their stability. Thus, we report various correlations among labour market variables at different leads and lags and compare them across countries. Then, we investigate the time-varying nature of these features paying a special attention to times of financial distress.

We find that there are robust relations among some variables across countries and times at business cycle frequency, that the cyclical comovement between real output and selected labor market indicators (mostly unemployment and hours) is strong and that these cyclical features hold not only in normal times but also during the times of financial crises. On the other hand, we find little evidence for such a robust relation between wages and output or wages and unemployment or hours. This implies that the structural models should pay more attention to fit the first type of the feature, as its more quantitative important and more stable in time and in space. This paper is a part of our ongoing research focus and in

[1]Czech National Bank, Macroeconomic Forecasting Division, Na Příkopě 28, Praha 1, Jan.Bruha@cnb.cz
[2]Czech National Bank, Macroeconomic Forecasting Division, Na Příkopě 28, Praha 1 and Faculty of Economics and Administration, Masaryk University, Jiri.Polansky@cnb.cz

[1]See [4] or [1] who provide examples of impossibility of identification of SVAR models in economies with realistic structures.

future we will extend presented results by a more detailed analysis of data and more extensive discussion of results.

2 Data

Up to now, we were able to obtain data from 33 countries for the analysis. When selecting countries, we want to cover a majority of European economies (transition as well as developed) and some other developed (OECD) countries. We have also data on Turkish economy, as it is now a market economy with its rapid economic development. Such a variety allows us to provide as robust analysis as possible due to different structural characteristics of labour market in different countries and also to deal with short time series in some countries. Some countries have relatively short times series (less than 10 years). However, if we find that some feature is present in most countries regardless their degree of development and regardless the time span of available data, such a feature should be judged as important, even if its identification in a single given country would be difficult.

All data come from statistical offices and national banks' databases. For each country, we collect real GDP, real consumption, real investment, various employment time series, total hours worked, unemployment rate and its level, nominal wages, compensation of employees, labour force and participation rates. However, we were unable to collect all data for all countries. Most data have quarterly frequency, some data were available at monthly frequency only. In that case, we have transformed these monthly data to quarterly data. Our all analyses are thus done on the quarterly data.

All in all, we were able to collect data on these countries: Austria, Australia, Belgium, Bulgaria, Canada, Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Iceland, Ireland, Italy, Japan, Latvia, Lithuania, Malta, Netherlands, Norway, New Zealand, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, the United Kingdom, the United States, Turkey.[2]

3 Correlation analysis

As a first step in our analysis, we report various sample correlations among labour market variables at different leads and lags and at various frequencies (trends and cycles), and compare them across countries. Gap time series have been acquired by the Hodrick-Prescott filter [5] filter with the usual value of the smoothing parameter for quarterly frequency $\lambda = 1600$.[3] We carry out the filtration for all time series, including unemployment, in order to filter out some structural changes in the labour market (which can be recently relevant for example for Germany).

Figure 3 presents correlation analysis among selected variables at the -6 to 6 leads-lags interval. The figure contains the statistics for all countries in the sample. In each panel, the bold line depicts the median of the sample correlations at the selected lag, the dark shadow area is the interquartile range, and the light shadow area range based on all available countries.[4] If the dark (or even light shadow) area is thin, it means that the underlying correlation is very similar for all countries in the sample. The first row in this figure shows correlations between GDP at and hours worked, employment level (persons) and unemployment rate. In the second row, the first subplot presents correlation between wage bill and the GDP. Next two panels show correlations between nominal wages and employment level and unemployment rate, respectively. The third row in the figure presents correlations between real consumption and GDP, unemployment rate and the wage bill. The next three rows are analogous for growth rates (trends).

There are two main interesting findings derived from Figure 3. First, there are robust relations among the variables across countries at business cycle frequency and that the cyclical comovement between real output and some of labor market indicators is strong. This is particularly true for hours worked, employment and unemployment. The output-wage comovement is weak (if it is present at all) and the degree and the sign of this comovement vary across countries. On the other hand, at lower frequencies, all the relations are relatively weak. This is corroborated by the fact that the comovements is apparently much weaker for growth rates (which implicitly contain the movements across all frequencies), while they

[2] The data come from Eurostat, Australian Bureau of Statistics, Statistics Canada, OECD National Accounts Statistics, OECD Main Economic Indicators, Statistics New Zealand, Statistics Korea, Turkish Statistical Institute, and FRED.

[3] We check the results also using the band pass Christiano-Fitzgerald filter [3] and the results are essentially the same.

[4] Since not all data are available for all countries, the sample of countries can be different for each panel. E.g. since the hours worked are not available for some countries, the first panel is based on less countries than the second etc.

are strong for gap variables. The implication for developing of a DSGE model with an elaborate labour market block, it is much more important to capture these cyclical patterns than correlations in growth rates.

An interesting question is the typical lag where the correlation between any two gaps peaks (i.e., it is the largest in the absolute value). Figure 3 presents a histogram with these correlation peaks for all countries. E.g. the correlation between the real GDP gap and the gap in hours worked is maximal for 10 countries with no lag, while for another 10 countries it is maximal when the gap in hours lags the output gap by one quarter. From the first row of Figure 3, it is apparent that for most countries, employment and unemployment lagged the output gap by one or two quarters. For a smaller subset of countries, the correlation is largest for contemporaneous variables[5]. Again, for the correlation of wages with the rest of variable, we do not find any interesting pattern (see the second row in Figure 3).

Figure 3 plots the boxplots of the (base 10) logarithm of the standard deviations of gaps in variables relative to the standard deviations of the output $\varsigma_x^{rel} = \log_{10}\left(\frac{\sigma_x^{gap}}{\sigma_{GDP}^{gap}}\right)$, where σ_x^{gap} is the sample standard deviations of the cyclical part of the time series x. The boxplots are organized as follows: on each box, the central red mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme datapoints considered to be not outliers[6], and the outliers are plotted individually by red crosses. The logarithmic transformation was chosen for better readibility of the picture. Obviously $\varsigma_x^{rel} < 0$ means that the gap in the respective variable is less volatile than the output gap, while $\varsigma_x^{rel} > 0$ means the opposite. If $\varsigma_x^{rel} \cong 1$, the gap in the variable x is about 10times volatile than the output gap.

Again, some robust facts emerge. The investment and unemployment gaps are much more volatile than the output gaps in all countries. The consumption gap has on average the same volatility as the output gap, while the gaps in employment, hours worked and labour force is typically less volatile than the output gap, even if there are few outliers here. Nevertheless, for a typical economy in the sample, the volatility of these three variables is lower.

The results for the gaps in wages and wage bills are diverse and as in the case of correlation we cannot characterize a typical country in the sample: there is no systematic pattern of the relative volatilities for these countries. We also looked whether there is a relation between relative volatilities of wage gaps and employment or hour gaps, but we have not found any systematical pattern[7].

4 The stability of correlation in time

Recently with the advances in computational power, researchers started using the sophisticated models with time-varying parameters. Many times, the researchers conclude that the parameters are not stable, especially with the financial and macroeconomic crises after 2009. In lights of these findings, the natural question therefore is whether the facts reported in the previous part of the paper are stable in times.

To answer this question, we employ the approach suggested by [2]. This approach is based on the computation of the recursive correlations:

$$\rho_{it}^{s_1 s_2} = \frac{\sum_{\tau=t-s_1}^{t+s_2}(x_\tau - \bar{x}_t^{s_1 s_2})(y_\tau - \bar{y}_t^{s_1 s_2})}{\sqrt{\sum_{\tau=t-s_1}^{t+s_2}(x_\tau - \bar{x}_t^{s_1 s_2})^2}\sqrt{\sum_{\tau=t-s_1}^{t+s_2}(y_\tau - \bar{y}_t^{s_1 s_2})^2}},$$

where $\rho_{it}^{s_1 s_2}$ is the correlation between variables x and y in country i, centered at time t, $\bar{x}_t^{s_1 s_2}$ is the recursive mean of the variable x ($\bar{x}_t^{s_1 s_2} = \frac{1}{s_1+s_2+1}\sum_{\tau=t-2_1}^{t+s_2} x_\tau$) and analogously for $\bar{y}_t^{s_1 s_2}$.

Given a binary indicator z_t, one can test whether the recursive correlations $\rho_{it}^{s_1 s_2}$ are systematically different in times when $z_t = 1$ comparing to times when $z_t = 0$. The details of the test are described in [2]. Alternatively, one could ran a regression of $\rho_{it}^{s_1 s_2}$ on z_t. We apply this procedure for the case of recent crisis and hence it is a binary variable equal to one to the post Lehman times and zero otherwise.

[5] Note that for some countries, the correlations seem to be strongest for 5 or 6 quarter lags. This is a spurious result caused by previous business cycle.

[6] Outliers are defined as observations larger than $P_{75} + 1.5(P_{75} - P_{25})$ or smaller than $P_{25} - 1.5(P_{75} - P_{25})$, where P_{25} and P_{75} are the 25th and 75th percentiles, respectively.

[7] The idea was that countries with high wage volatility could have lower unemployment or employment gap volatility as the shocks to the economy would be absorbed by prices (i.e. wages) rather than quantities (employment). This conjecture does not prove.

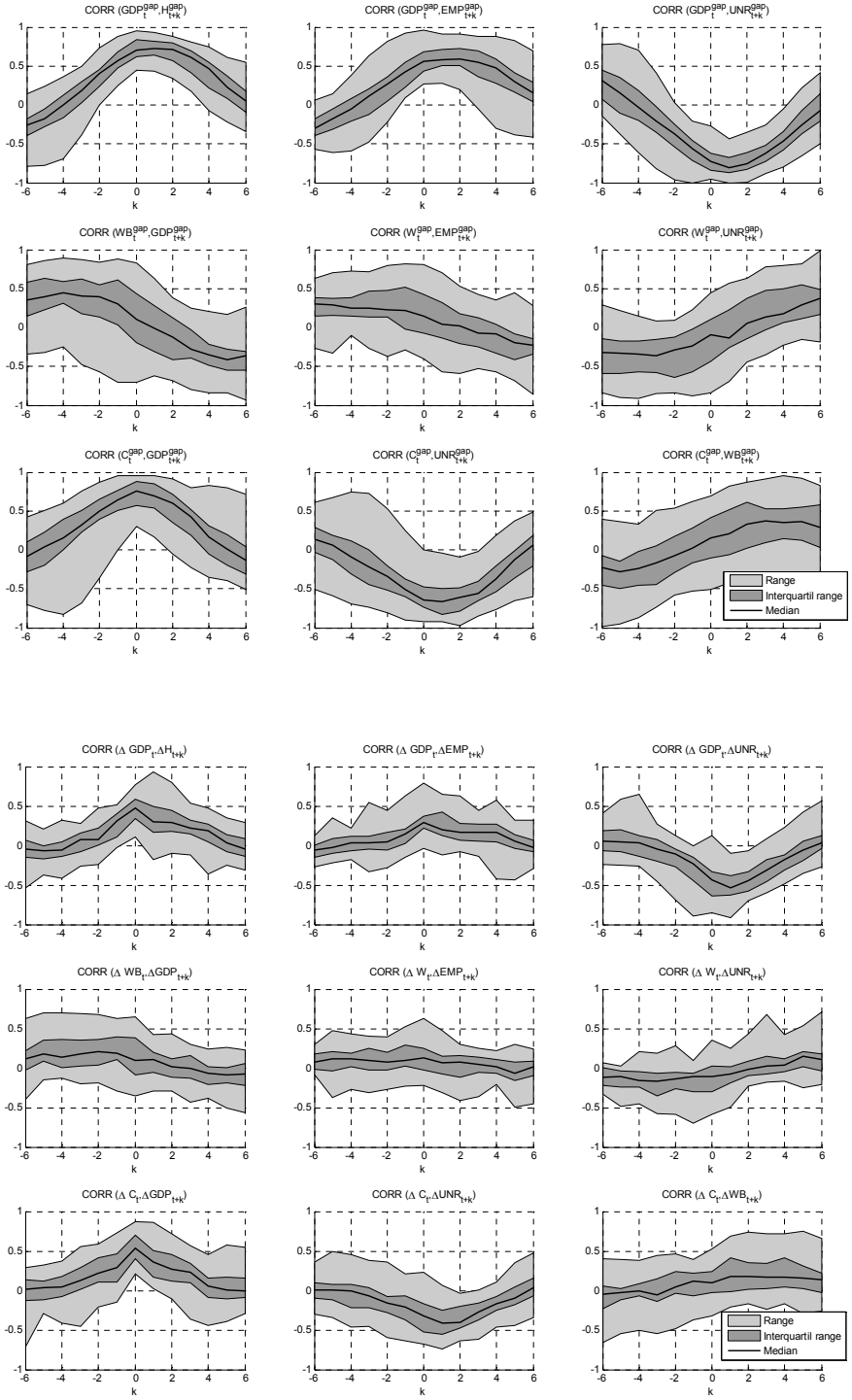


Figure 1 Correlation among labour market data

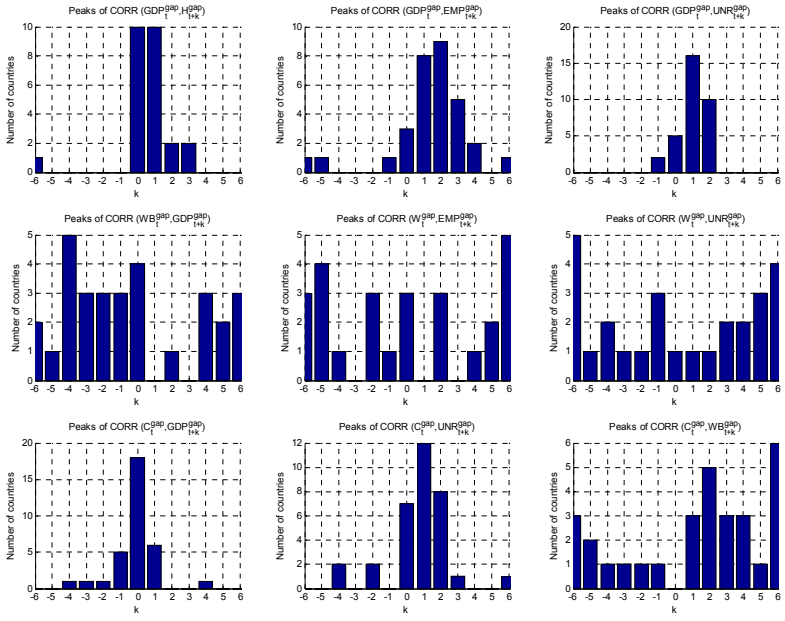


Figure 2 Peak of correlation among labour market data

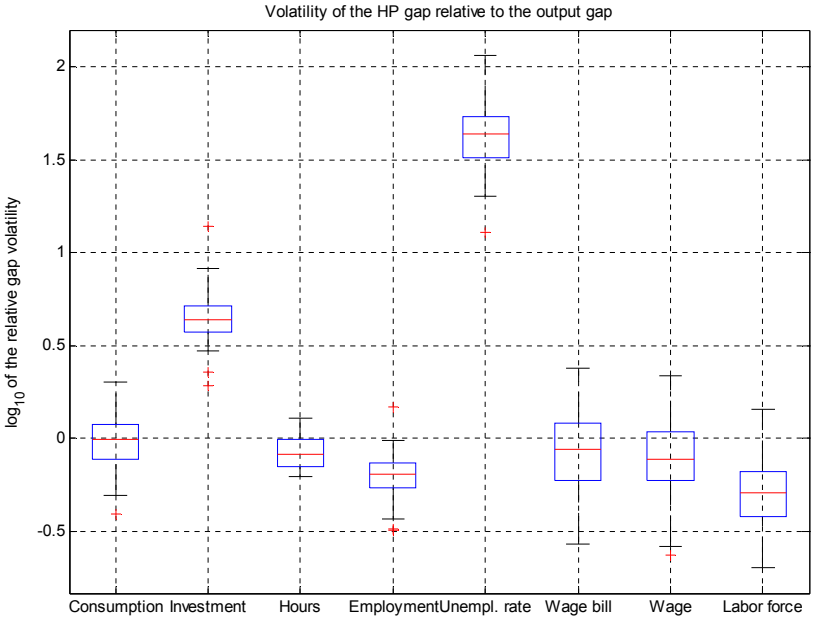


Figure 3 Relative volatilities of the cyclical components

If there were really a time variation after the beginning of the crisis, one would expect a dependency of $\rho_{it}^{s_1 s_2}$ on that indicator z_t.

We did this test for all countries in our sample for the correlation between output gap and unemployment gap. We do not find any evidence of the dependency of this recursive correlation on such an indicator. There is some, but weak, dependency between the correlation in growth rates. Therefore, we conclude that the stylized facts based on cyclical data are stable even during the recent crisis, but that the crisis can slightly affect the correlations in growth rates. This latter finding is probably caused by the change in low frequency component in data and is subject to our future research.

5 Conclusion

In this paper, we seek for robust stylized facts about labour markets in developed countries. Our findings are following. First, there are subsets of variables which correlates with the real output (and consumption) gap at business cycle frequencies. These are the hours worked, employment and unemployment. The correlation in growth rates are weaker, but still present. The gap in these labour market variables typically lag the output gap by one or two quarters. Moreover, contrary to correlations of growth rates, the correlation in gaps between these variables seem to be stable in time, even during the recent recession. The same relationships were not discovered for wages: the correlations of the wage gap and the output gaps, or between the wage gap and the gaps in the rest of labour market variables are small, and different between countries.

We can conclude that we find a set of robust stylized facts that seem to hold in different countries and that are stable in time. Therefore, the modelers dealing with structural models with elaborate labour market blocks should try to replicate these facts as these facts probably represent the true economic mechanisms in advanced economies.

Acknowledgements

The support of the Research Project B2/13 at the Czech National Bank is gratefully acknowledged.

References

[1] Alessi, L., Barigozzi, M. and Capasso, M.: A Review of Nonfundamentalness and Identification in Structural VAR Models, ECB Working Paper Series 922, 2008.

[2] Blasques, F.: On the Phase Dependence in Time-Varying Correlations Between Time-Series, Tinbergen Institute Discussion Papers 13-054/III, 2013.

[3] Christiano, L. J. and Fitzgerald, T. J.: The Band Pass Filter, *International Economic Review* **44**, 2 (2003), 435–465.

[4] Fernández-Villaverde, J,. Rubio-Ramírez, J., Sargent, T. and Watson, M.: ABCs (and Ds) of Understanding VARs, *American Economic Review* **97**, 3 (2007), 1021–1026.

[5] Hodrick, R. J.: Postwar U.S. Business Cycles: An Empirical Investigation, *Journal of Money, Credit and Banking* **29**, 1 (1997), 1–16.

Influence of Demographic Data aggregation on Accuracy of Facility Location

Cebecauer Matej[1]

Abstract. When we are solving location problems in spatially large geographical areas we are dealing with huge numbers of possible candidate locations and serviced customers. The problem is that real infrastructure networks at the state level are too big and it is not possible to compute corresponding decision problems in a reasonable time. The commonly used approach for solving such problem sizes is to use data aggregation where all customers are aggregated into predefined points. The aim of this paper is to investigate influence of aggregation to results of optimization problems. First, we generate primary network based on the method called spatial decomposition. This method does not remove any data from network, but instead it breaks down the network into smaller parts while keeping the connectivity among all adjacent segments. To control the degree of aggregation we vary the size of cells covering the geographical area. The maximal size of cells corresponds to the macroscopic data model. Furthermore, we compare the results of p-median problem for different aggregation levels and evaluate the influence of data aggregation on the accuracy of location decisions.

Keywords: aggregation, network design, networks and graphs, p-median.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The aim of this paper is to investigate and show the influence of spatial aggregation on the accuracy of facilities location. The spatially large realistic geographical areas contain huge numbers of possible candidate locations and customers. The reason of aggregation is that original problem is too big to be solved in a reasonable time. The idea behind aggregation is reduction of size of the problem. The result of commonly used aggregation of geographical areas is the fact that towns and villages are spatially non-dimensional points and that a certain point represents all of inhabitants of a town, but in the reality 20 000 people do not live in one point, but they exists in a certain space. It is obvious that this fact can have influence on solution. Hillmans [5] described aggregation errors in measuring distance between customers and facilities. And we describe aggregation errors with answering the question of influence of spatial aggregation on the accuracy of *p*-median facilities location.

The paper is structured as follows. In section 2 we describe commonly used aggregation, which we call macroscopic level and we propose an algorithm to the layout population and how to obtain different aggregated levels. Section 3 deals with the proposition of experiments and we introduce *price of aggregation* and *distortion of aggregation*, which express the aggregation errors. Finally, in section 4, we shall summarize our conclusions.

2 Data model

For our investigation we need microscopic geographical data. The free source of this kind of data is server OpenStreetMap which we describe in this section. Then we propose in [2] an algorithm for the layout of population to the two-dimensional space and how to obtain different levels of aggregated population used in this paper. The result of creation of the data model in this section is a network or graph usable for optimization problems.

2.1 OpenStreetMap

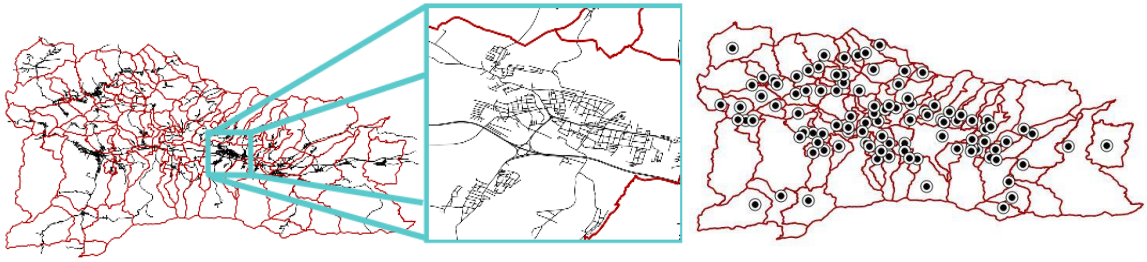
The project OpenStreetMap(OSM), born at University College London in July 2004, was founded by Steve Coast. OSM is an open source map server. All data and added information are available for free and cover the whole world. OSM contains realistic actual data from the whole world an allows to get enormous number of large-scale heterogeneous networks such as roads, railways, shops, restaurants, hospitals, fire/police stations, etc. You can see [1],[4] for more information.

[1] University of Zilina, Faculty of management science and informatics, department of transportation networks, Universitna 8215/5, SK-01026 Zilina, Slovakia, matej.cebecauer@gmail.com.

2.2 Microscopic and Macroscopic level

The commonly used data model in location problems is of macroscopic level with customers represented by towns and villages. In this part we describe process how to generate macroscopic level from microscopic level obtained from OSM.

First, we obtain microscopic data about road infrastructure from OSM, because this infrastructure has the highest density and the best connection to real customers in their houses. This obtained data form *microscopic level*. Further we aggregate this microscopic data to towns and villages. To make this aggregation automatic we have to know areas that belong to each towns or villages. This information about land use of towns can be obtained from OSM as well, just like the main point describing the coordinates and population in the towns. With this data we carry out decomposition method which we explained in more details in [2]. The result of decomposition method are two levels of namely macroscopic level and *microscopic level* divided into segments which represent microscopic data of all towns and villages on the macroscopic level. The only part of decomposition method in this work we need is the macroscopic level, because we work with problems of such a size that we are able to compute them as one entity on a microscopic level, so that we do not need microscopic segments.



(a) (b) (c)

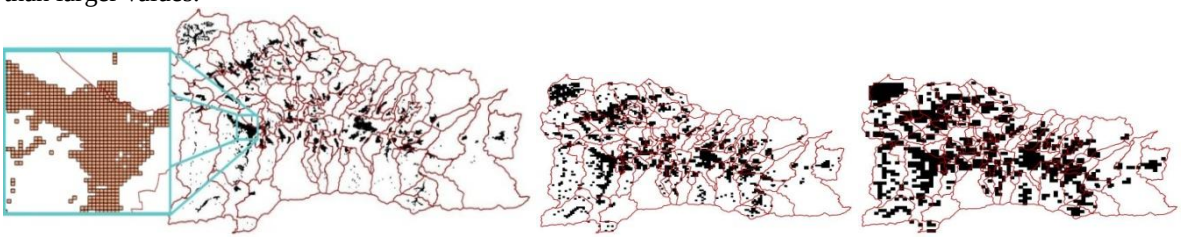
Figure 1 (a) The microscopic level with administrative border of towns and villages. (b) The zoom in microscopic level with roads infrastructure. (c) The macroscopic level with towns and villages points.

2.3 Layout of population

Once we have microscopic level segments, we also need to layout the population. We know the town population as well as the roads in segment. Furthermore, from the OSM we can get buildings and town residential areas that are situated in each segment. There are several different ways of allocating population that were described in [2]. It is not possible to consider every building or person as a node, because we would get a very large graph, and we could not easily find optimal solution in a reasonable time. So we introduce the technique with name *Grid Estimation Of Population* (GEOP) [2]

1. If there are no buildings or residential areas in a segment, then we put all population into main segment node (town point in macroscopic level) and stop. Otherwise we continue to the next step.
2. We create a spatial grid with items of constant size (size is arbitrary = size of square cells).
3. We divide all objects of the segment (buildings, residential areas) into grid cells based on coordinates.
4. We remove all grid cells that are empty (the result is land which we can consider settled).
5. Afterwards we allocate people into grid cells with regard to the number of buildings and fractions of residential areas in these cells.
6. Finally we decide about the location of the main nodes of grid cells and we connect them to the road network.

GEOP allows creating data models with different size of *Side Of Square Cells* (SOSC), so we can simply calibrate the size of grid cells. Each separate cell represents aggregated population, we call that aggregated customers. In the Fig. 3. you can see results for different values of SOSC. Smaller values of SOSC are closer to reality than larger values.



(a) (b) (c)

Figure 2 (a) The aggregated level with aggregated customers of SOSC(100). (b) The aggregated level with aggregated customers of SOSC(500). (c) The aggregated level with aggregated customers of SOSC(1000).

3 Experiments

In this section we propose experiments to illustrate the influence of demographic data aggregation on location of facilities. For these experiments we generate networks which are obtained from data models with different SOSC we described above.

3.1 Networks and input data for experiments

We decided to generate three different microscopic geographical areas based on administrative regions. Each area consists of one macroscopic level and ten microscopic levels with different sizes of aggregated customers (SOSC). It is important to note that towns in macroscopic level are also aggregated customers. The first generated area consists of three administrative regions of Slovak republic (SR), more specifically districts Ružomberok, Dolný Kubín and Liptovský Mikuláš (LM_DK_RK). The second area consists of only one district Žilina (ZA) and the last area consists of two districts Martin and Turčianske Teplice (MN_TT). More details about these areas are in Table 1.

Geographical area	population	number of towns and villages
LM_DK_RK	166 269	105
ZA	156 623	53
MN_TT	118 472	68

Table 1 Information about geographical areas

Further we apply GEOP with different SOSC on the microscopic areas. The results are microscopic levels with a different number of aggregated customers, for more details see Table 2. We decided to use 100 meters as the minimal value of SOSC. We consider this microscopic level, where SOSC equals 100 meters, as the most detailed layout of reality from all SOSC, which we are considering in these experiments and we are able to compute the optimal solution in reasonable time. We denote this SOSC as *SOSC(100)*.

	SOSC										
Geographical area	**2000**	**1000**	**900**	**800**	**700**	**600**	**500**	**400**	**300**	**200**	**100**
LM_RK_DK	505	901	980	1 108	1 271	1 435	1 757	2 178	2 983	4 830	12 119
ZA	239	453	520	574	674	781	989	1 256	1 785	3 307	8 320
MN_TT	284	506	552	607	677	769	965	1 190	1 620	2 647	6 787

Table 2 Information about number of aggregated customers in microscopic levels with different values of SOSC

3.2 Proposition of experiments

To illustrate the influence of the aggregation on the population we propose experiments for locating the facilities in these networks with different aggregated customers. In these networks we are locating different amount of facilities, from ten *(10f)* upwards. For illustration ten facilities represent very realistic number of ambulances for regions of these sizes. Candidates for facilities location are all the aggregated customers for each network. In each location problem for all networks we compute the optimal solution with the exact method that was proposed by Garciá and col. [3]. They called this method, *ZEBRA*. This method is very fast and can compute larger *p*-median problems in a reasonable time, faster than any other exact method before. We compute weighted *p*-median of location problem so that we can investigate the influence on aggregation of population. If we use an aggregation there are two basic errors which we investigate in this paper. These errors have influence to the sum of all distances among customers and their closest facilities that represent the value of the optimal solution.

To describe this errors we define ***F*** as a set of locations of facilities from specific aggregated problem ***X*** and the function ***TO(F,Y)*** that gives back the sum of all distances among aggregated customers from the aggregated problem ***Y*** to their closest facilities from the set ***F*.** This function *TO* makes the mapping from the location of facilities from one aggregated problem to the aggregated customers from another aggregated problem. Next we define function ***SUM(X)*** which returns the sum of all distances among aggregated customers from the aggregated problem *(X)* to their closest facilities located in the aggregated problem *(X)*.

The first is the error in the locating facilities, because in the aggregated problem we have less number of aggregated customers and candidates to locating facilities which are covering the spatial larger areas. Consequence of this fact is that it has *influence on accuracy of facility location.* We introduce ***the price of aggregation (POA)*** which expresses the percentual overpayment with which we overpay *SUM(Y)* of the aggregated problem if we are using the aggregation *X*. The value of POA represents the contrast between *TO(F,Y)* and *SUM(Y)* proportional to *TO(F,Y)*. It is important to note that *Y* aggregation has to be on the lower level of aggregation than *X* aggregation. The POA is defined as following :

$$POA = \frac{TO(F,Y) - SUM(Y)}{TO(F,Y)} \tag{1}$$

The POA is a number between 0 and 1. If the value of POA is closer to 0 then the efficiency measure is closer to *SUM(Y)* and overpayment is minimized. Values closer to 0 are preferred because they offer lower overpayment. We consider *Y* as the lowest aggregation that we denote as SOSC(100). It is important to note that POA also represents how much in percentage we can improve our solution *F* if we use lower aggregation *Y*.

The second error represents the *influence on measuring of distances*. This error is incurred by the aggregation and the fact that the result of aggregation generates spatial larger aggregated customers who are represented as a one point. For example 300 customers in one aggregated customer have the same distance to their closest facility and of course that have the same facility. These circumstances influence the error in the measuring distances. We named this error ***the distortion of aggregation (DOA)*** and it expresses the percentual distortion of the *SUM(X)* with regard to *TO(F,Y)*. The formula of DOA is :

$$DOA = \frac{TO(F,Y) - SUM(X)}{TO(F,Y)} \tag{2}$$

The DOA is a number between 0 and 1. If DOA has a value near 0 then the measure of the distortion is closer to aggregated problem *Y*. Values closer to 0 are preferred because they describe aggregated problem *Y* better. The value 1 is the maximal distortion; it means that the solution of the aggregated problem for all customers is without the costs. This is the lower-bound of the problem and it is not possible to achieve a better solution of this aggregated problem *X*. In this situation the POA can tell us how big the overpayment for this setting is.

The next attribute describing the impact of the aggregation is *the contrast between average distances X* and *Y*. We define function *AD(SUM(X))*, which returns a weighted average distance of the *SUM(X)* We named this attribute ***influence of aggregation (IOA)***. The IOA is the contrast between *AD(X)* and *AD(Y)*.

$$IOA = AD(SUM(X)) - AD(TO(F,Y)) \tag{3}$$

The value of IOA represents how far in the weighted average distance is *SUM(X)* from the *TO(F,Y)*. The aggregated problem *Y* is in our experiment denote as *SOSC(100)*.

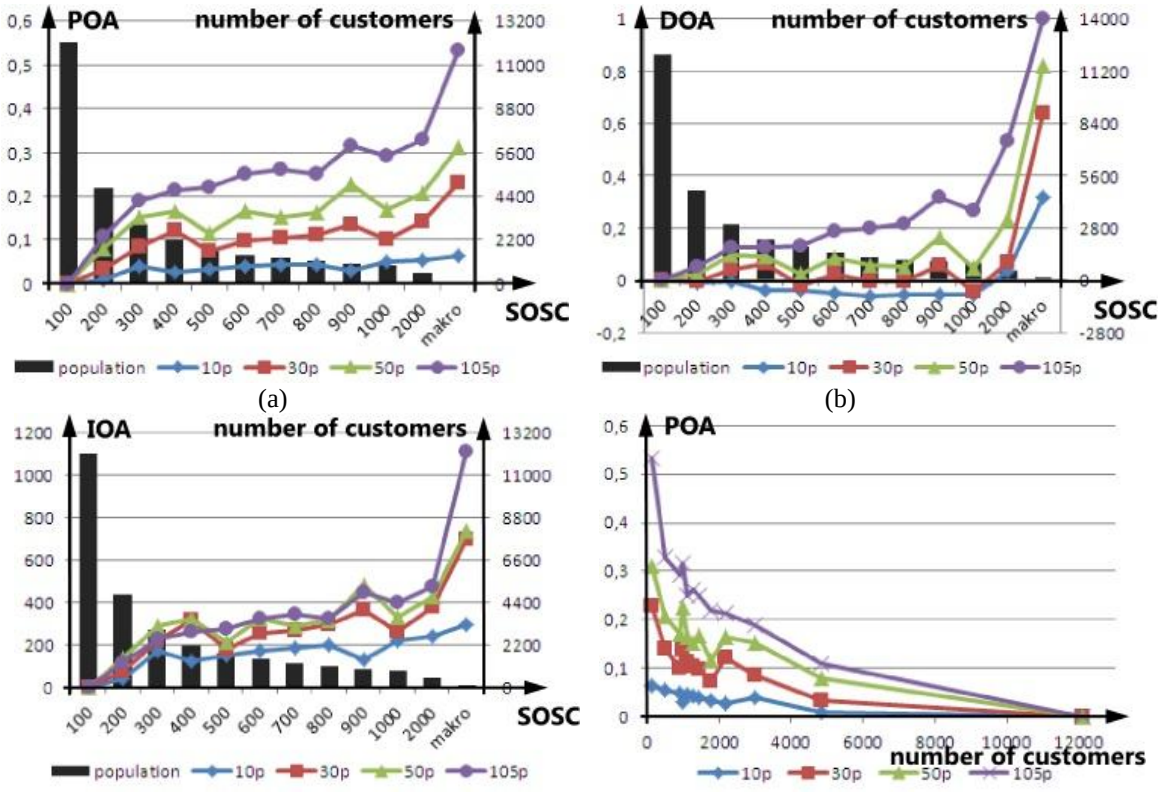


Figure 3 The results of the experiments on network LM_RK_DK. The histograms in graphs represent the number of aggregated customers in each SOSC. (a) The POA, (b) DOA and (c) IOA according to SOSC and the (d) POA according to aggregated customers.

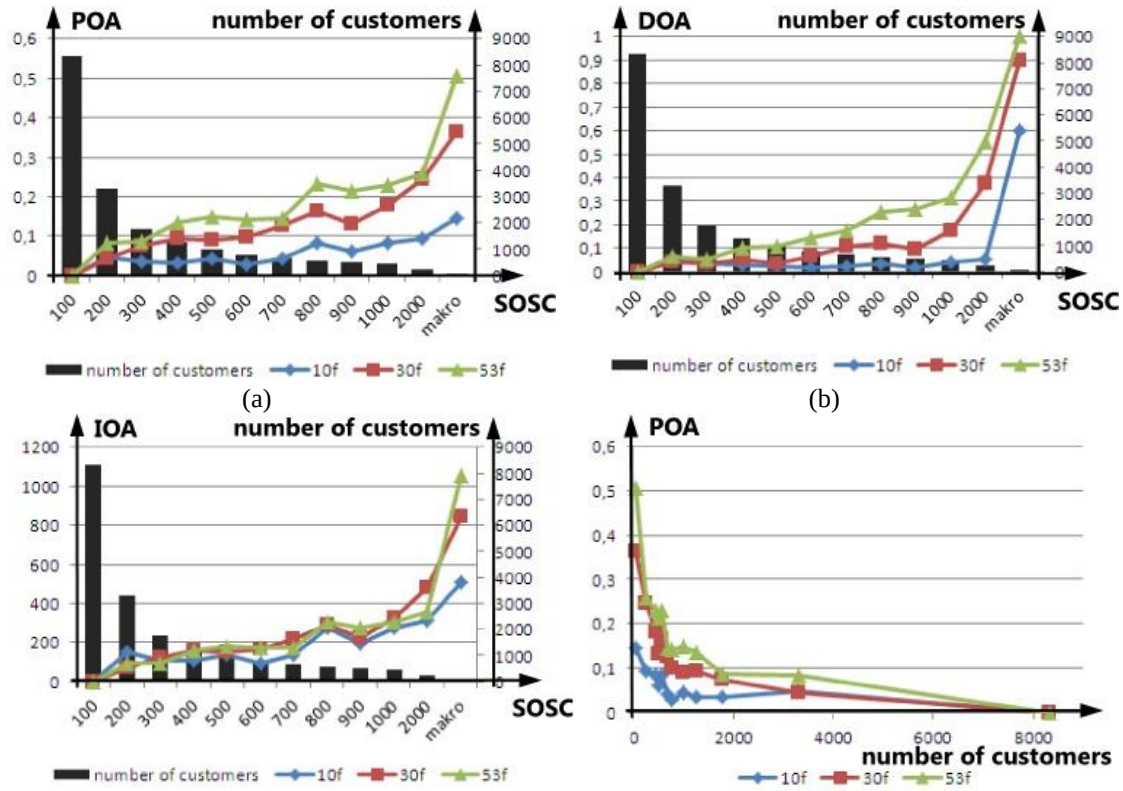


Figure 4 The results of the experiments on network ZA. The histograms in graphs represent the number of aggregated customers in each SOSC. (a) The POA, (b) DOA and (c) IOA according to SOSC and the (d) POA according to aggregated customers.

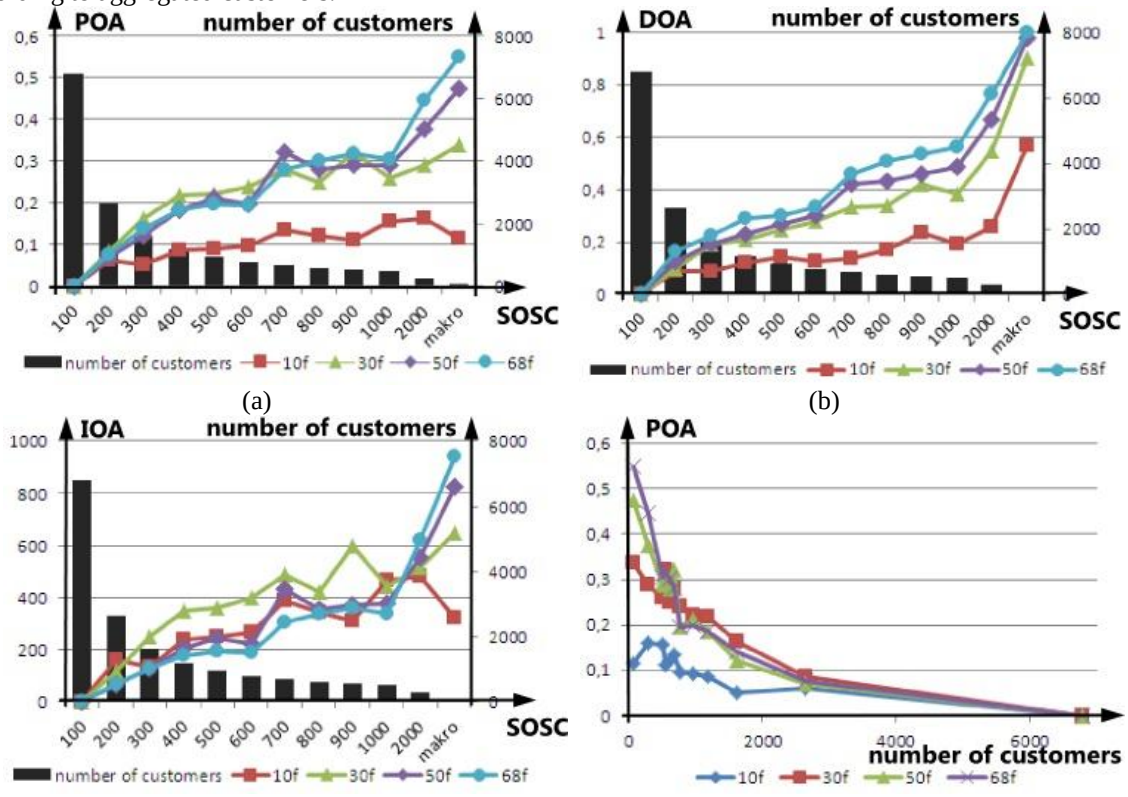


Figure 5 The results of the experiments on network MT_TT. The histograms in graphs represent the number of aggregated customers in each SOSC. (a) The POA, (b) DOA and (c) IOA according to SOSC and the (d) POA according to aggregated customers.

4 Conclusions

This paper was focused on the aggregation impact on the facility layout precision and its overall system effectiveness. We have created three different realistic microscopic models, on which the GEOP method was applied. This method aggregates the population into squares of defined dimension. With different value of the square dimension *SOSC*(100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000) we have created different levels of aggregated customers. Received number of aggregated customers, as well as their layout, depends on the geographical space and gained data for *GEOP*. For this reason it may occur, that for example SOSC(200) does not necessary offer better solution regarding to the SOSC(100), that for example SOSC(500) is giving. This may have impact at the worse solution finding comparing to macroscopic customers aggregation at low numbers of located facilities. The reason of this is depicted in the way of cutting plane into grid using the *GEOP*, where SOSC(200) does not have the same coordinates for aggregated customers as aggregated customers of SOSC(100), but SOSC(500) has. Anyway it is very small degradation according to the SOSC(100).

On the data models generated this way we placed different number of facilities. We observed two errors that can be encroached according to overall passed distance of all customers to their closest facilities. This summed distance represents the task solution. First error is linked to location of facilities, which we have called *price of aggregation*. This error reflects the consequence of the selection from the smaller set of candidates for placement, that represents bigger area compared to the SOSC(100). The price of fairness gets values from 0 to 1, the closer to 0, the closer is to the SOSC(100) solution. We have called the second monitored error *distortion of aggregation* that represents the error of distance measurement. This represents the result distortion that represents the solution of the aggregated task proportionally to the SOSC(100) solution. This error can gain also negative values, when reaching solution worse than SOSC(100) up to the value 1, when it offers the solution, in which is everyone served without costs or traveling no distance. Value 0 of distortion of aggregation stands for no distortion and the solution gained in the aggregated task is equal to the SOSC(100) solution.

Experiments results are depicted in figure 3.-5. From the results presented in the charts one can see, that monitored properties nonlinearly decrease when the number of aggregated customers raises. The decrease is bigger for lower number of aggregated customers. The expected assumption has been confirmed. This fact says, if the bigger number of centers is being placed that the bigger aggregation impact on the errors in the placing and distance measuring will be. On the contrary with the lower number of centers, the aggregation influence is falling. Even some values, the example can be seen in figure 5(d) when placing 10 centers, have found worse solutions than the macroscopic level. Globally we can assume, that on all three considered data models there is exponential decrease stabilizing on the 25% limit of aggregated customers from SOSC(100). Of course, it depends on how big precision has to be provided. But it is possible, that this 25% is a sufficient amount to achieve relatively good precision. We want to improve this estimation in next research using larger experiments for bigger number of different networks of different dimensions and topologies. Also we want to focus on the price of aggregation and distortion of aggregation, but not from the view of the whole system and average distance, as it is in this paper, but from the point of individual customers and how the aggregation influences these distances.

The majority of the Operational Research researchers work on the better and faster algorithms with the goal to solve larger tasks. We want to solve very large tasks by the heuristic based on the results presented in this paper. We assume that the solution by the heuristic on the aggregation better describing the reality can be better solution than the optimal solution obtained on the macroscopic level. In the next work we will also focus on the research of this option and the given heuristic design.

Acknowledgements

This work was supported by the research grants VEGA 1/0339/13 “Advanced microscopic modeling and complex data sources for designing spatially large public services systems”, APVV-0760-11 “Designing of Fair Service Systems on Transportation Networks” and by the institutional research grant of Faculty of Management Science an Informatics.

References

[1] Cebecauer, M. *OpenStreetMap and Optimization Problems*. In Technika, technologies, telematics and informatics in the transport and logistics systems 2012. Pardubice: Univerzita Pardubice., ….
[2] Cebecayer, M. TRANSCOM 2013
[3] García S. Labbé M., Marín A.: *Solving Large p-Median Problems with Radius Formulation*. INFORMS Journal on Comupting, 2010.
[4] Haklay, H., Weber, P.: *OpenStreetMap:User-Generated Street Maps*. Pervasive computing, IEEE CS, 1536-|1268/08, 2008.
[5] Hillsman , E., Rhoda, R. P *Errors in measuring distances from population to service centers,* 1978.

Real-time versus revised Czech data: a DSGE analysis

Jan Čapek[1]

Abstract. The paper investigates the effects of using real-time data instead of revised final data within Dynamic Stochastic General Equilibrium (DSGE) framework. Theoretically, using most recent, updated historical data for ex-post analyses for a historical time-sample may be misleading, because revised data were not available at that time.

The contribution uses a small-scale monetary macroeconomic DSGE model to analyze the importance of real-time data compared to most-recent revised data with a focus on the differences in decision-making of the monetary authority. The analysis proceeds from Bayesian estimation of model parameters in a model with real-time data and in a model with the most recent revised data.

Keywords: real-time data, data revision, recursive estimate, DSGE model

JEL classification: C11, C32, C52, C54, E52, F41
AMS classification: 91B51, 91B64, 91B84, 62P20

1 Introduction

The paper investigates the effects of using real-time data instead of revised final data within Dynamic Stochastic General Equilibrium (DSGE) framework. Theoretically, using most recent, updated historical data for ex-post analyses for a historical time-sample may be misleading, because revised data were not available at that time. In case of e.g. monetary macroeconomic models with the usage of revised data, the monetary authority would be expected to make decisions based on then-unknown data. The contribution uses a small-scale monetary macroeconomic DSGE model to analyze the importance of real-time data compared to most-recent revised data with a focus on the differences in decision-making of the monetary authority. The analysis proceeds from Bayesian estimation of model parameters in a model with real-time data and in a model with the most recent revised data. Statistical significance of the differences in the Bayesian estimates is presented.

Following section introduces the reader to real-time data literature. Literature review focuses on papers that are relevant for this paper's aim. Section 3 presents analysis of the data itself with descriptive statistic. Section 4 presents results of recursive estimation on a SOE DSGE model with real-time and revised data. Final section concludes.

The term "real-time data" addresses data that become available right after collection. In accordance with the literature on similar topics, this paper understands "real-time data" as data that are available 3-4 months after the end of a quarter and are typically the first estimates published for that quarter.

A "vintage" is a quarter at which the data becomes available, or, the time of publishing. For example, if Czech statistical office releases during April of 2013 an estimate of GDP growth for last quarter of 2012, it is said that the data for fourth quarter of 2012 are in vintage of April of 2013.

2 Literature review

The issue of data revising and properties of real-time versus revised data has been investigated in the literature for a long time and until now, it is still a topic for scientific discussion.

The orientation of various literature and methodology differs. Some articles focus on the data itself and investigate properties of real-time data in comparison with revised data. There is also a group of articles that focus on the influence of real-time data on model results, e.g. for forecasting or reactions of monetary policy. This contribution follows both of these lines of research.

[1]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 600 02 Brno, capek@econ.muni.cz

Croushore and Stark [2] investigate properties of a real-time dataset and show that data vintages matter in comparison to revised data. Taylor [8] focuses on forecast quality when real-time data are used for its construction. He shows that forecasting models that use timely data have higher forecasting power. Business cycle analysis in real-time data environment is done e.g. by Lee et al. [4], who construct real-time macroeconomic database for Australia. The database exposes the difficulties in drawing inferences and decision-making based on macroeconomic data that is subsequently revised. Molodtsova et al [6] estimate Taylor rule with revised data and real-time data. There is a difference between the two estimates in Germany but not so much in the USA. Orphanides and van Norden [7] examine the reliability of alternative output detrending methods with special attention to the accuracy of real-time estimates of the output gap. They find out that the revision of published data is not the primary source of revisions in measured output gaps; the bulk of the problem is due to the pervasive unreliability of end-of-sample estimates of the trend in output. Boivin [1] searches for the differences in drifting parameters with real-time data. The findings suggest important but gradual changes in the Taylor rule coefficients, not adequately captured by the usual split-sample estimation.

This contribution follows the trends in literature and it presents analysis on the data itself by descriptive statistics in section 3 and it also calculates model results with recursive estimates in section 4.

3 Data analysis

There is a number of sources of uncertainty concerning the data and its model implementation. This contribution addresses three of them: (i) the differences in the data by source, (ii) the magnitude of data revisions and (iii) the differences in detrended data due to data availability at time of detrending.

3.1 Differences in the data by source

Figure 1 depicts the differences between the data from two widely used databases: OECD and Eurostat. The data should be the same with the exception of inflation, which is only available as CPI in OECD real-time database and HICP in Eurostat database. In most cases, the differences are small or none at all, which is the case of both domestic and foreign interest rates with no error and EuroArea growth with only 0.02 percentage points of average absolute error. On the other hand, average absolute difference in data series for Czech inflation differs by a half of a percentage point, which is not negligible. As the errorbars display, there are no systematic patterns in differences between Eurostat and OECD data.

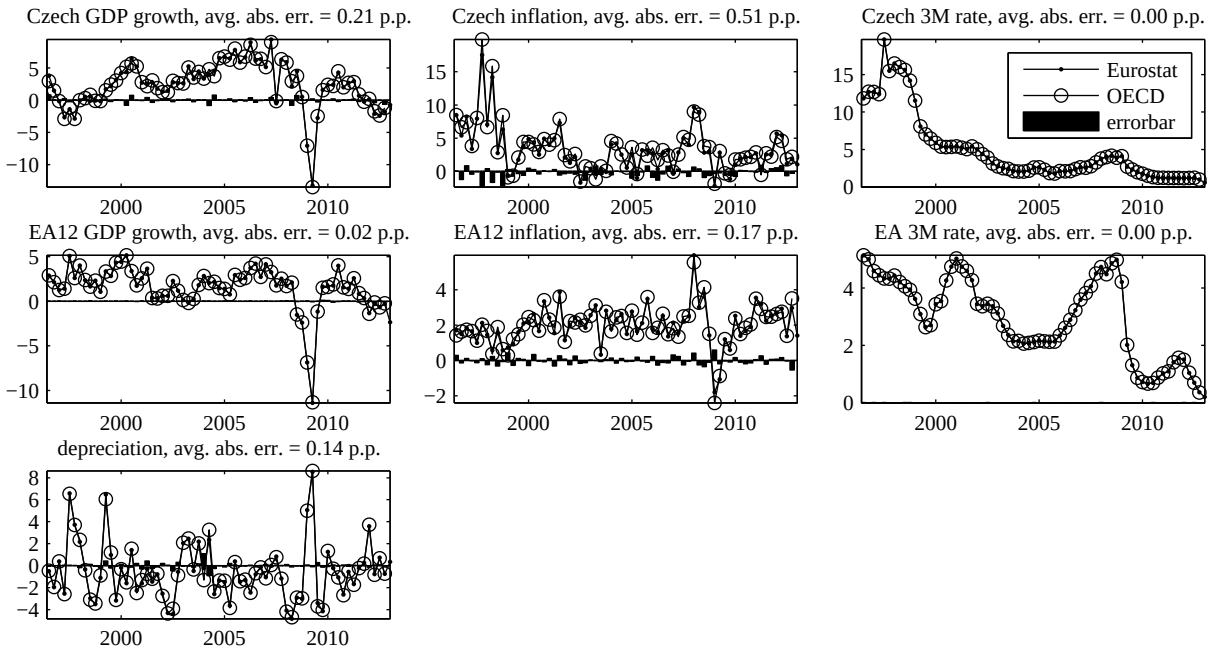


Figure 1 Eurostat and OECD data, *Note*: (error)bar plot around zero = difference between Eurostat and OECD data, "avg. abs. err." = average absolute error in percentage points

3.2 Magnitude of data revisions

Figure 2 displays the series of real-time data and most-recent updated data, all from OECD database. The differences in the series are not insignificant and the errorbars in cases of GDP growths show that the deviations are also non-random. Real-time data tend to underestimate the actual development. For example, in periods of expansion, real-time data underestimate positive growth in the Czech Republic in 2004/2005 and also in EuroArea in 2007/2008. As for recessions, both real-time values for domestic and EuroArea economy underestimate the severity of the crisis of 2009. The average absolute error tops at 1.34 percentage poins (quarterly, per annum) for Czech GDP growth, which means that the average value of the error is approximately 56% of average GDP growth over the sample. *Average error-to-average value* is also very high for EuroArea growth with 39%.

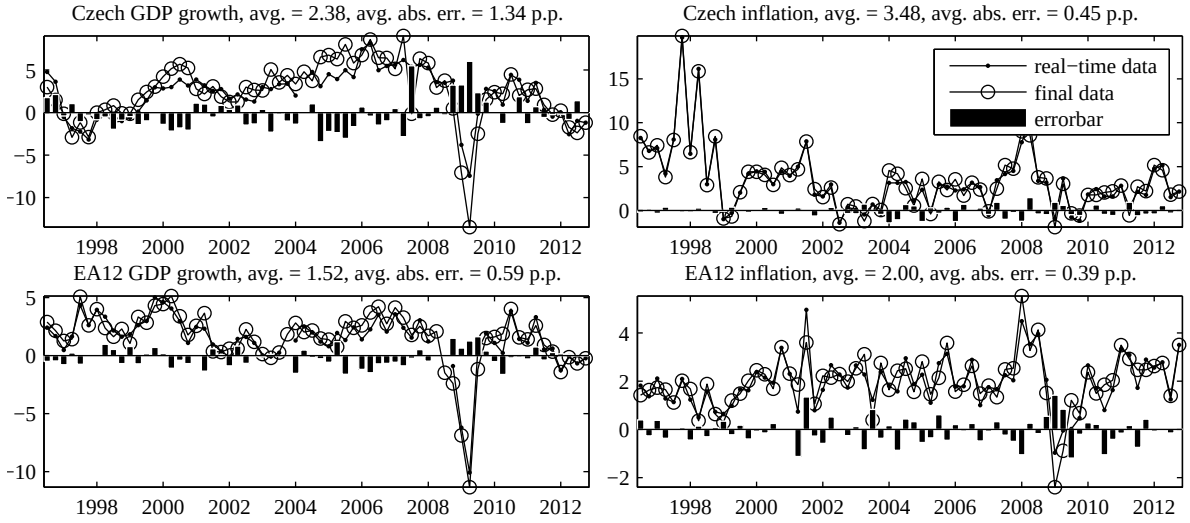


Figure 2 Real-time and final (revised) data, *Note*: (error)bar plot around zero = difference between final and revised data, "avg." = simple average of final data over the sample, "avg. abs. err." = average absolute error in percentage points

3.3 Differences in detrended data due to data availability at time of detrending

This section offers a depiction of a source of uncertainty which is not in the data itself but stems from the need to detrend the time series prior to modeling. In the model used, EuroArea interest rate is detrended with a linear trend and since Czech interest rates display even more curvature, Hodrick Prescott trend is used.

This type of analysis uses only revised data series. At first, the trend is computed on time series truncated to 30 observations. Then, the truncation moves one-by-one to more observations and the trend is recomputed each time. This procedure results in a number of time series of different length, each one detrended on its sample. Such results depict changes in the time series stemming only from the recomputation of the trend.

Figure 3 displays all vintages of detrended data with a solid line and final data with a circled line for comparison. The last values of each vintage form a real-time data in the second row panels. Time series formed this way will be addressed "detrend real-time" data in the remainder of the paper. Note that the difference between "detrend real-time" data and final revised detrended data stems only from the fact that detrending is calculated repeatedly for each length of the series. The third row in Figure 3 displays errorbars that are expectedly systematic.

4 Estimation

4.1 Model and Identification

This paper uses a New Keynesian (NK) Dynamic Stochastic General Equilibrium (DSGE) model. The model is derived from microeconomic behavior of particular economic agents. These include domestic and foreign households, domestic and foreign producers, domestic importers and domestic and foreign monetary authority. The

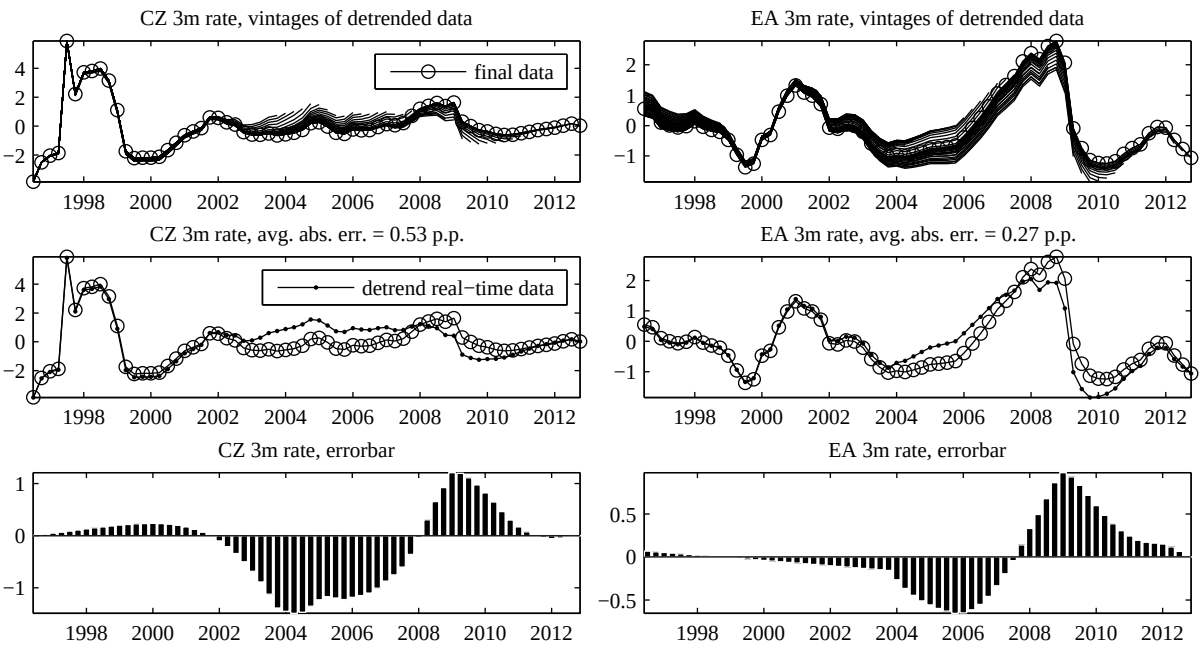


Figure 3 Differences in detrended data due to data availability

model is in small open economy (SOE) setting, so that it presumes two countries - small open economy that is influenced by a big closed economy. The small open economy is the home Czech economy, the big large economy is the foreign EuroArea economy. Most of the model assumptions are adopted from Lubik and Schorfheide [5].

The model is estimated by Bayesian methods in Dynare package[1] on a dataset with 7 variables. All the estimates share the same prior setting. Convergence of each estimate is checked by Brooks and Gelman convergence diagnostics.

4.2 Recursive estimates

Recursive estimates are constructed so that as a new data point is added to the existing set, the model is re-estimated. New values of parameter estimates are then reported. One can argue that if parameters estimates change, it reflects the information in the newly-added data.

Recursive analysis is regularly done by policy institutions if newly-data becomes available to actualize the model for a new round of predictions or forecasting. Recursive analysis is also done on historical sample to analyze the evolution of policy. The former inevitably uses real-time data, the latter mostly uses updated final data, because the estimation is easier to implement and older vintages of real-time data may not be available. This sections shows two examples where the choice of real-time versus final data does matter.

Figures in this section present four estimates to compare. Estimates based on Eurostat and OECD data are standard recursive estimates – only the time series of most updated final data is used, detrended on the whole sample and the detrended series is then truncated to mimic shorter sample. Note that the detrended series do not change at all in this type of estimation, there is just one new data point added in each longer time-frame.

Another approach is a genuine real-time estimate. This sequence of estimates use only data available at the vintage in question. Each time series may undergo revisions and therefore may be different. Also, detrending is made on these potentially different series with different number of observations. Although this approach mimics the historical possibilities of analysis best, it has a drawback that the influence of the data revisions and the influence of potentially changing trends cannot be distinguished.

To address this problem, another recursive estimate is introduced. The original data series for estimation are revised final series, but the truncation of the series is done prior to detrending. This estimation therefore does not reflect any issues of data revisions but it captures the influence of re-detrending in each quarter. It is then possible

[1] `www.dynare.org`

to infer the relevance of data revisions by comparing the genuine real-time estimate and the estimate that covers the problems arising from consecutive detrending. Note that this type of results was earlier addressed to as based on "detrend real-time" data. In order to easily see the significance of the results, a 90 % probability band of the "detrend real-time" estimate is drawn.

The estimates start in second quarter of 1996 and have at least 30 observations. The last observation is third or fourth quarter of 2012 which makes 66 or 67 observations in the most-recent estimates.

Estimates based on Eurostat and OECD data usually behave similarly, although there may be some vertical difference between the series of the estimates. However, on the two presented cases, the "detrend real-time" estimate that covers only the problem of consecutive detrending is different than the former two. This means that knowledge of the trend is significant and not knowing future trend leads to different estimates. Also, the real-time estimates differs from "detrend real-time" estimates, which again means that also the nature of the changes in the data itself due to revisions are significant.

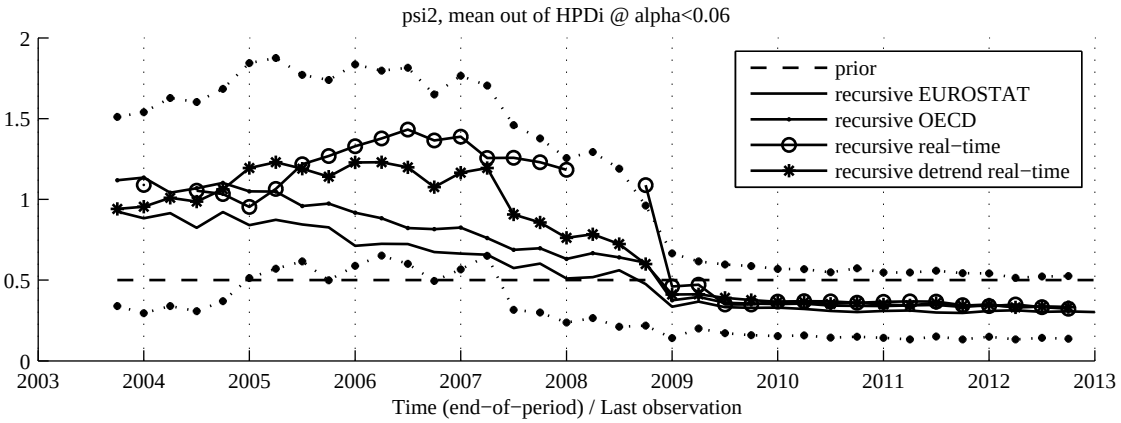


Figure 4 Recursive estimates of the weight on output growth in a domestic Talyor rule

Figure 4 displays the evolution of recursive estimates of a parameter that is the weight on output growth in a domestic Talyor rule. It is therefore a preference parameter of domestic monetary authority. The higher it is, the more important is output growth for the central bank when it set its policy interest rates.

All of the estimates display a tendency of a decline. The tendency is gradual in standard recursive estimates. On the other hand, the real-time estimate shows no gradual tendency but there is a great drop in the estimate in fourth quarter of 2008 (denoted as 2009 in the figure).[2] Since the evolution of the real-time estimate is outside of 90 % probability bands of "detrend real-time" estimate, the difference is statistically significant. Such result may be interpreted intuitivelly – from the point of view of real-time data, the oncoming crisis was a shock that is captured by a major change in the estimate in the quarter in question. Other estimates display gradual decline, because the information may have been included in revised data and the future trend was known.

Figure 5 displays estimation results for the persistence of domestic interest rate and it displays rather different development. Both standard recursive estimates calculated on final revised data tend upwards from 0.6 to final 0.7. However, both remaining estimates also show a decline in years 2006 and 2007 before converging upwards. This results may mean that 2006/2007 period is consistent with lower interest rate smoothing that would seem from the estimates on revised data. Since the evolution of real-time and "detrend real-time" data is similar, the differences are probably mainly due to re-estimation of the trend rather than in data revision itself. Note also that the results are not significant at conventional levels so that we can only discuss tendencies.

5 Conclusion

Data from OECD and Eurostat are very similar and the use in a DSGE model does not make much difference. Comparing revised and real-time data shows that the differences are not insignificant and also with systematic patterns. Also, if the preparation of the data for a model use means detrending (other than demeaning), this makes rise to a problem of trend recomputation, which can also be significant.

[2] Note that missing values of the estimates are due to missing real-time data at respective quarters.

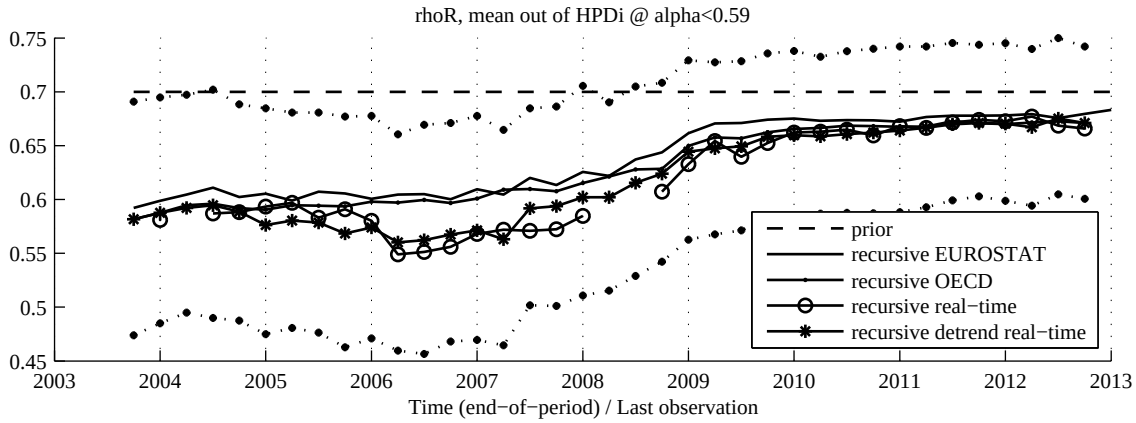


Figure 5 Recursive estimates of the persistence of domestic interest rate

Analysis on a DSGE model shows results where the re-detrending is significant and also where the nature of real-time data in comparison to revised data is significant.

Paper's results are in line with existing literature that finds the use of real-time data significant in various estimation or forecasting exercises. This contribution confirms the results on Czech data.

Acknowledgements

The access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme "Projects of Large Infrastructure for Research, Development, and Innovations" (LM2010005) is highly appreciated.

References

[1] Boivin, J.: Has U.S. Monetary Policy Changed? Evidence from Drifting Coefficients and Real-Time Data. *Journal of Money, Credit and Banking* 38, 5 (August 2006): 1149-1174.

[2] Croushore, D. and Stark, T.: A Real-Time Data Set for Macroeconomists: Does the Data Vintage Matter? *Review of Economics and Statistics* 2003 85:3, 605-617, ISSN 0034-6535

[3] Lee, K., Olekalns, N., Shields, K., and Wang, Z.: Australian Real-Time Database: An Overview and an Illustration of its Use in Business Cycle Analysis, *Economic Record*, Volume 88, Issue 283, pages 495516, December 2012

[4] Lubik, T., and Schorfheide, F.: A Bayesian Look at the New Open Economy Macroeconomics. In *NBER Macroeconomics Annual 2005, Volume 20*, NBER Chapters. National Bureau of Economic Research, Inc, May 2006, pp. 313–382.

[5] Molodtsova, T., Nikolsko-Rzhevskyy, A., and Papell, D. H.: Taylor rules with real-time data: A tale of two countries and one exchange rate, *Journal of Monetary Economics*, Volume 55, Supplement, October 2008, Pages S63-S79, ISSN 0304-3932.

[6] Orphanides, A., and van Norden, S.: The Unreliability of Output-Gap Estimates in Real Time, *Review of Economics and Statistics* 2002 84:4, 569-583, ISSN 0034-6535

[7] Taylor, N.: Economic forecast quality: information timeliness and data vintage effects, *Empirical Economics*, 2013, pp. 1-30, issn 0377-7332.

Partially Joint Transport on Networks

Jan Černý[1], Vladimír Přibyl[2]

Abstract. The paper deals with a road network represented by a graph. It is supposed that there are several objects of the same type (containers, persons etc.) to be transported on the graph from their origins to their destinations at the same time. Any object can pass its route alone or can join some other(s) for passing some segments of the graph together. The cost of joint passing a segment is smaller than the sum of individual passing. In an extreme case, the cost of the joint passing equals to the cost of the individual one. Moreover, the limit of joined objects for transportation together is given.
The goal is to find such routes for all objects that minimize the total cost.
In the paper, several particular cases are studied where all objects have the same destination or the same origin.
Two main approaches to the solution are analyzed. First, several models based on flows on graphs with costs are discussed and afterwards linear programming models are presented.
These problems can be met in practice e.g. in travelling of pupils to schools, of employees to work or of football referees to matches.

Keywords: digraph, routing, joint travelling, costs, optimization.

JEL Classification: C65, L62, O18, R42
MS Classification: 90C15

1 Introduction

In practice, one can meet a situation where some objects are transported partially alone and partially together by the same type of transport. The second is advantageous since the joint transport is more efficient. For instance, two employees of the same employer live in different but not distant places *A* and *B*. Each morning they can drive their cars to their workplace *C* separately. However, they prefer that, in turns, one day the first takes the car, drives from *A* to *B*, picks up the colleague and continues to *C*. The next day, on the contrary, the second colleague takes the car for the trip $B - A - C$. Of course, such a type of partially common transport may relate to more than two persons to more than one destination. The present paper will deal with this type of problems.

1.1 General Problem

Given a network, represented by a graph $G = (V, E)$. A set of objects Q is transported on the network G, an object $q \in Q$ is transported from its origin o_q to its destination d_q. Any object can pass its route alone or it can join some other(s) for passing some edges of the graph together. The cost of joint passing an edge is smaller than the sum of costs of individual passing, i.e. on each edge $e \in E$, besides the price of "solo" passage $c_1(e)$ it is specified the price of "joint" passage of $k > 1$ objects $c_k(e) < kc_1(e)$. The difference between these numbers is significant. Sometimes there is even value $c_k(e)$ so close to the value of $c_1(e)$ that in the simplified model it is expected to be equal. The problem is to find a path p_q from o_q to d_q for each $q \in Q$ in such a manner that the total cost of transport is minimal.

Remark 1. It may seem strange that no requirement concerning time is introduced into the formulation. The authors are convinced that it is not necessary to introduce time-space network in this particular case, where one can suppose that all transportation is going to take place around the same time (e.g. travel to work which starts, say, at 8 a.m. for all employees).

1.2 Cost of Passing an Edge

The cost $c_k(e)$ of passing k objects through the edge e can be defined in several ways, having the common feature that $c_0(e) = 0$ and $c_1(e) > 0$:

[1] University of Economics in Prague/Faculty of Management, Department of Exact Methods, Jarošovská Street 1117/II, 377 01Jindřichův Hradec, Czech Republic, cerny@fm.vse.cz.
[2] University of Economics in Prague/Faculty of Management, Department of Exact Methods, Jarošovská Street 1117/II, 377 01Jindřichův Hradec, Czech Republic, pribyl@fm.vse.cz..

- if no upper bound for the number of together transported objects is given then it may be:
  - constant $c_k(e) = c_1(e)$ for all $k = 2, 3, \ldots$;
  - linear $c_k(e) = c_1(e) + a_e(k-1)$ for all $k = 1, 2, 3, \ldots$ and $e \in E$ where $a_e \geq 0$ is a given constant;
  - nonlinear $c_k(e) = c_1(e) + f_e(k)$ for all $k = 1, 2, 3, \ldots$, where $f_e > 0$ is a given (nonlinear) function;
- if an upper bound b for the number of together transported objects is given then it may be:
  - piecewise constant $c_k(e) = (\text{int}((k-1)/b) + 1)\, c_1(e)$ for all $k = 1, 2, 3, \ldots$;
  - piecewise linear $c_k(e) = (\text{int}((k-1)/b) + 1)c_1(e) + a_e(k - \text{int}(k/b))$ for all $k = 1, 2, 3, \ldots$ and $e \in E$ where $a_e > 0$ is a given constant having the property $a_e b < c_1(e)$;
  - nonlinear $c_k(e) = c_1(e) + f_e(k)$ for all $k = 1, 2, 3, \ldots$ $a > 0$, where $f_e > 0$ is a given function wih "jumps" after each b values of k.

It is necessary to note that saying "linear" means de facto piecewise linear, since for $k \leq 0$ it is $c_k(e) = 0$ and for $k > 0$ it is $c_k(e) = c_1(e) + a_e(k-1)$.

1.3 Variations of the Problem 1.1

From 1.1 it follows that there are at least six variants of the problem 1.1. However, their number is much greater, since it is necessary to distinguish between cases with one or several destinations and, similarly, with one or more origins of trips.

Moreover, as concerns approaches to solution, one should distinguish between different types of nonlinearities. E.g. convex cost is studied in [1], the concave one in [3] the general nonlinear in[6].

One can say that more than ten variants of the problem 1.1 ought to be examined separately. The present paper is focused on one of them.

2 One-to-Many Problem with Linear Cost and Unlimited Joining

In this section it is supposed that the problem 1.1 is specified by the following requirements:

- A vertex $o \in V$ is given such that $o_q = o$ for all $q \in Q$;
- Cost of passing is linear $c_k(e) = c_1(e) + a_e(k-1)$ for all $k = 1, 2, 3, \ldots$ and $e \in E$ where $a_e \geq 0$ is a given constant.

It is obvious that the same mathematical model can be used for the symmetric case "Many-to-One".

First, the case of constant costs, i.e. $a_e \equiv 0$, is tackled.

2.1 Problem 1: Constant Cost

Now, a graph $G = (V, E, c)$ is given where $c(e)$ is the cost of transport of any positive number of objects from the set Q through the edge $e \in E$. The **Problem 1** is to find a path p_q from o to d_q for each $q \in Q$ in such a manner that the total cost of transport is minimal, i.e. the number

$$c(P) = \sum_{e \in P} c(e) \to \min$$

for $P = \{e \in E: e \in p_q \text{ for some } q \in Q\}$.

The solution of the Problem 1 is based on the following propositions:

Proposition 1. Let $G = (V, E, c)$ be a graph. Let $Q \neq \varnothing$, and let $D = \{d_q \in V: q \in Q\}$. Let $G' = (D \cup \{o\}, P)$ be the shortest Steiner tree for $D \cup \{o\} \subset V$ in G. Let p_q be the (uniquely determined) path connecting the vertices o and d_q in G'. Then the paths p_q, $q \in Q$ represent the solution of the Problem 1.

Proof. Indirectly: Let other paths p_q', $q \in Q$ represent the solution of Problem 1, let $P' = \{e \in E: e \in p_q' \text{ for some } q \in Q\}$ and let $c(P') < c(P)$. That is in contradiction with the fact that the shortest Steiner tree represents the cheapest connected sub-graph of the graph G.

Proposition 2. The Problem 1 is NP-hard.

Proof is a consequence of Proposition 1 and the fact (see [2]) that the problem of the shortest Steiner tree on a graph G is NP-complete.

Remark 2. It follows from Proposition 2 that there is no sense in looking for a polynomial algorithm for an exact solution of the Problem 1. Therefore the exact solution can be reached by integer linear programming as seen e.g. in [5], or by some heuristics.

2.2 Problem 2: Linear Cost of Passing

Let a non-empty set of objects Q be given. Let a graph $G = (V, E)$ be given together with a cost function

$$c_k(e) = c_1(e) + a_e(k-1) \text{ for all } k = 1, 2, 3, \ldots \text{ and } e \in E$$

where $c_1(e) > 0$, $a_e \geq 0$.

The **Problem 2** is to find a path p_q from o to d_q for each $q \in Q$ in such a manner that the total cost of transport is minimal, i.e. the number

$$c(P) = \sum_{e \in P} c_1(e) + a_e\left(k_P(e) - 1\right) \to \min$$

for $P = \{e \in E: e \in p_q \text{ for some } q \in Q\}$ and $k_P(e) = \text{card}\{q. \in Q: e \in p_q\}$

Remark 3. Since the Problem 1 as a particular case of the Problem 2 is NP-hard, the same is true for the Problem 2. Therefore, there is no sense in looking for a polynomial algorithm for an exact solution of the Problem 2. The exact solution can be reached by integer linear programming or by some heuristics. A possible approach to linear programming is the following:

2.3 Problem 3: Transformation of the Problem 2 to the Cheapest Flow through a Digraph

Assume that the Problem 2 is fully determined as in 2.2.

Let us denote $V^* = V \cup \{s\}$ where s denotes the sink of the digraph $G^* = (V^*, H^*)$. Let $H^* = \{(o, v) \in E\} \cup \{(u, v) \in E: u \neq o \neq v\} \cup \{(d_q, s): q \in Q\}$. Let each $h \in H^*$ have infinite capacity. Let the cost $c^*(h) = 0$ for $h \in \{(d_q, s): q \in Q\}$ and let $c^*(u, v) = c(u, v) = c_1(u, v) + a_{(u, v)}(f(u, v) - 1)$ for all $(u, v) \in E$, where $f(u, v)$ is the flow through (u, v). The **Problem 3** is to find the flow $f(h) \in \{0, 1, 2, \ldots, \text{card}Q\}$ through the digraph $G^* = (V^*, A^*, c^*)$ meeting the constraints

$$f(d_q, s) = 1 \tag{1}$$

$$\sum_{(o,v) \in H^*} f(o, v) = \text{card}Q \tag{2}$$

$$\sum_{(u,v) \in H^*} f(u, v) = \sum_{(v,w) \in H^*} f(v, w) \quad \text{for each } v \in V^* - \{o, s\} \tag{3}$$

$$\sum_{h \in H^*: f(h) > 0} c_1(h) + a_h\left(f(h) - 1\right) \to \min \tag{4}$$

2.4 Problem 4: Linear Programming Solution of the Problem 3

Assume that the Problem 3 is fully defined. Let us define a binary variable x_h and an integer variable y_h for each $h \in H^*$, where $y_h = f(h)$ is the flow through h and $x_h = 1$ means that $f(h) > 0$. The **Problem 4** is to find values of variables x_h and y_h that meet the following constraints where M is a "very big number" as a substitute for ∞, e.g. $M = (\text{card}Q)^3$:

$$y_h \leq M x_h \tag{5}$$

$$\sum_{h \in \{(o,v) \in H\}} y_h = \text{card}Q \tag{6}$$

$$y_{(d_q, s)} = 1 \quad \text{for each } q \in Q \tag{7}$$

$$\sum_{(u,v) \in H^*} y_{(u,v)} = \sum_{(v,w) \in H^*} y_{(v,w)} \quad \text{for each } u \neq o, v \neq s \tag{8}$$

$$\sum_{h=(u,v) \in H^*, v \neq s} \left(c_1(h) x_h + a_h(y_h - 1)\right) \to \min \tag{9}$$

3 Notes to Practical Application

The problem of partially joint transport of goods or passengers may be found in cases of pupils travelling to schools, employees travelling to work or football referees travelling to matches. In the transport of goods, one can mention the collection of containers dispersed for a "single shot" collection of some material.

By looking at the paper [4] one could argue that application of partially joint transportation problems could be found also in freight train routing.

4 Conclusion

At the beginning a general problem of partially joint transportation was formulated and it was shown that there are many possible variations that differ in the type of costs, in number of origins and destinations.

Afterwards two types of "One-to-many" problem with linear cost are studied. The first one has constant cost, the second one works with linear cost depending on the transported quantity. It was shown that the first problem leads to the Steiner tree problem on graphs, which is known as NP-complete. Since it is a sub-problem of the second one, the later should be NP-hard.

For the sake of solvability, the second problem was reformulated to a minimum cost flow problem and a linear programming model to its solution was presented.

Since the general problem can be applied in wide area of transportation of goods and persons, the authors hope that many not yet solved variant problems will find their solvers in the near future.

Moreover, the set of constraints may be extended. E.g. an incompatibility function g on $Q \times Q$ could be defined such that $g(q, q') = 0$ means that the objects q and q' are fully compatible for joint transportation, whereas $g(q, q') = 1$ means that the common transportation is absolutely impossible and $g(q, q') \in (0, 1)$ means something in between. Adding a member containing g may be an object of further research as well.

Acknowledgements

The authors would like to thank the Czech Science Foundation for the support of the project 402/12/2147 whose results are described in this paper.

References

[1] Bertsekas, D. P., Polymenakos, L. C., and Tseng, P.: An ε-Relaxation Method for Separable Convex Cost Network Flow Problems. *SIAM Journal on Optimization.* **7** (1997), 853-870.

[2] Garey, M. R. and Johnson, D. S.: *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co. New York, NY, USA, 1979

[3] Guisewite, G. M. and Pardalos, P. M.: Minimum concave-cost network flow problems: Applications, complexity, and algorithms. *Annals of Operations Research* **25** (1990), 75-99.

[4] Klug, T., Schlechte, T., Fügenschuh, A. and Borndörfer, R.: Freight Train Routing. *2nd International Conference on Models and Technologies for Intelligent Transportation Systems* 22-24 June, 2011, Leuven, Belgium (pages not known)

[5] Maculan, N., Plateau, G. and Lisser, A.:Integer linear models with a polynomial number of variables and constraints for some classical combinatorial optimization problems. *Pesqui. Oper.* **23** (2004), 161-168.

[6] Moussourakis, J. and Haksever, C.: Project Compression with Nonlinear Cost Functions. *J. Constr. Eng. Manage.*, **136** (2010), 251–259.

Interval data and linear regression: some properties and examples of the possibilistic approach

Michal Černý[1]

Abstract. We consider the linear regression model where we have partial information on its data. Only intervals are available, containing the unobservable values. We deal with the possibilistic approach: we study the ranges of all possible values which a given statistic can attain. In particular, we focus on the OLS-estimator, estimator of the variance of error terms and t-ratios. We also study complexity-theoretic classification of computation of the lower and upper bounds of the ranges. We illustrate the theory by an example motivated by the analysis of production functions, showing that even a small perturbation in data of a regression model can lead to quite wide intervals for some statistics.

Keywords: linear regression, possibilistic regression, interval data

JEL classification: C46
AMS classification: 62J86

1 Introduction

Consider the linear regression model

$$y = X\beta + \varepsilon, \tag{1}$$

where y is the vector of (observations of) the dependent variable, X is the design matrix, β is the vector of regression parameters to be estimated and ε is the vector of disturbances. From now on, n and p will denote the number of observations and the number of regression parameters, respectively. The tuple (X, y) is called *data* for the model (1). In this text we treat the data as fixed constants (i.e., as observations) rather than random variables.

An *interval matrix* $\boldsymbol{X} = [\underline{X}, \overline{X}]$ of the dimension $n \times p$ is a family of matrices $\{X \in \mathbb{R}^{n\times p} : \underline{X} \leq X \leq \overline{X}\}$, where the relation $\leq$ is understood entrywise. An *interval vector* is a one-column interval matrix. Interval matrices and vectors are denoted in boldface.

We are interested in the following problem. Assume that the data (X, y) for the model (1) are unobservable. The only information available to us is a pair $(\boldsymbol{X}, \boldsymbol{y})$ such that we are guaranteed that $X \in \boldsymbol{X}$ and $y \in \boldsymbol{y}$. Then we lose some information: for example, we cannot compute the value of the Ordinary Least Squares (OLS) estimator $\widehat{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y$, since we do not know the values (X, y).

There exist various approaches to this problem. Following [2], we mention two of them here. First assume that the process, which generated intervals $(\boldsymbol{X}, \boldsymbol{y})$ from the real-valued data (X, y), is known. An example is *rounding*: then we have

$$\overline{X} = \lceil X \rceil, \quad \underline{X} = \lfloor X \rfloor, \quad \overline{y} = \lceil y \rceil, \quad \underline{y} = \lfloor y \rfloor,$$

where the operations $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ are understood entrywise. Another example is *additional random error*, that is,

$$\overline{X}_{ij} = X_{ij} + \gamma_{ij}^{+}, \quad \underline{X}_{ij} = X_{ij} - \gamma_{ij}^{-}, \quad \overline{y}_i = y_i + \delta_i^{+}, \quad \underline{y}_i = y_i - \delta_i^{-}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, p,$$

where $\gamma_{ij}^{+}, \gamma_{ij}^{-}, \delta_i^{+}, \delta_i^{-}$ are nonnegative random variables.

Of course, we are interested in computation of various statistics, such as the OLS-estimator $\widehat{\beta}$, Residual Sum of Squares, an estimate of the covariance matrix of $\widehat{\beta}$ etc. In general, such a statistic is a function

[1]University of Economics, Prague, Department of Econometrics, Winston Churchill Square 4, 13067 Prague, Czech Republic, cernym@vse.cz

of data, say $S(X, y)$. Since the data X, y are not observable, the natural approach is to replace the statistic $S(X, y)$ by another statistic $S^*(\underline{X}, \overline{X}, \underline{y}, \overline{y})$, giving us similar information like $S(X, y)$, and study its properties. For example, we can replace the OLS-estimator $\widehat{\beta}(X, y) = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y$ by the central estimator

$$\widehat{\beta}^*(\underline{X}, \overline{X}, \underline{y}, \overline{y}) = (X_C^{\mathrm{T}}X_C)^{-1}X_C^{\mathrm{T}}y_C, \tag{2}$$

where $X_C = \frac{1}{2}(\underline{X} + \overline{X})$ and $y_C = \frac{1}{2}(\underline{y} + \overline{y})$ are the *centers* of $\boldsymbol{X}$ and $\boldsymbol{y}$, respectively. Similarly we can compute e.g. RSS using the centers. Of course, important properties of the statistic S^*, such as consistency, efficiency etc., will depend on the nature of the interval-generating process. This is an interesting direction of study — an example of an important result from this area is Sheppard's Correction [9].

Now we turn to the second approach, which is applicable in case when the interval-generating process is unknown. Then it is natural to consider *all possible values* which the statistic $S(X, y)$ can attain when X ranges over $\boldsymbol{X}$ and y ranges over $\boldsymbol{y}$. Taking the OLS estimator as an example, we are interested in the range of possible values of $\widehat{\beta}_i$ $(i = 1, \ldots, p)$, that is,

$$\overline{\widehat{\beta}_i} = \sup\{b_i : X^{\mathrm{T}}Xb = X^{\mathrm{T}}y \text{ for some } X \in \boldsymbol{X}, y \in \boldsymbol{y}\},$$
$$\underline{\widehat{\beta}_i} = \inf\{b_i : X^{\mathrm{T}}Xb = X^{\mathrm{T}}y \text{ for some } X \in \boldsymbol{X}, y \in \boldsymbol{y}\}.$$

The second approach, called *possibilistic approach*, is complementary to the first one: first we can use e.g. the central estimator (2), and then the range $[\underline{\widehat{\beta}_i}, \overline{\widehat{\beta}_i}]$ quantifies the worst-case error, i.e. the worst possible deviation of the central estimator from the true value $\widehat{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y$.

However, this approach often leads to serious computational problems. One of them is the following.

Proposition 1 (a consequence of Theorem 4 from [1]). *Computation of the ranges* $[\underline{\widehat{\beta}_1}, \overline{\widehat{\beta}_1}], \ldots, [\underline{\widehat{\beta}_p}, \overline{\widehat{\beta}_p}]$ *is an NP-hard problem.* □

It follows that we can expect only exponential-time algorithms for the problem — and this is intractable, especially when the number of observations is large.

In general, it is also interesting to study not only the ranges of possible values of the statistic, but also the simultaneous regions of possible values for two or more dependent statistics. A motivation for this approach can be found in [10], where simultaneous regions for expectation and variance, median and interquartile ratio, mean and Gini coefficient and many other combinations of parameters, estimated from one-dimensional data, are studied. In our setting, it is interesting to consider the simultaneous region for OLS-estimates of all regression coefficients. The region is called OLS-set [4, 1]:

$$B := \{b \in \mathbb{R}^p : \ X^{\mathrm{T}}Xb = X^{\mathrm{T}}y \text{ for some } X \in \boldsymbol{X} \text{ and } y \in \boldsymbol{y}\}.$$

With regard to Proposition 1, we cannot expect nice structural properties (or, at least, computationally testable properties). The set B need not be bounded, it need not be convex. In fact, just testing boundedness is a co-NP-hard problem [1]. Taking an example from [3], the OLS-set B with data

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 \\ 1 & [0;5] \\ 1 & [2;4] \\ 1 & 4 \end{pmatrix}, \qquad y = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

is plotted in Fig. 1.

2 Interval dependent variable

The general model described in the previous section, where both the design matrix $\boldsymbol{X}$ and the dependent variable $\boldsymbol{y}$ is interval, suffers from serious computational problems illustrated by Proposition 1. Hence it makes sense to restrict the problem to less general cases, where more optimistic results could be expected. In this paper we restrict ourselves to the case when the design matrix X is real-valued (i.e., we have $\underline{X} = \overline{X} = X$) and only the output variable is interval. From now on we assume that the matrix X has full column rank.

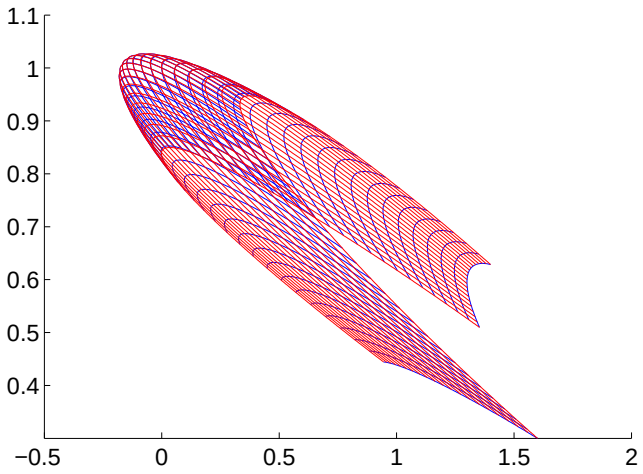


Figure 1 Example of the OLS-set.

Now the bounds for OLS-estimates are easily computable: we have

$$\widehat{\boldsymbol{\beta}} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\boldsymbol{y}, \tag{3}$$

where $\widehat{\boldsymbol{\beta}}_i = [\underline{\widehat{\beta}_i}, \overline{\widehat{\beta}_i}]$, and the expression (3) is evaluated using the *interval arithmetic* [7, 8], defined for two intervals $\boldsymbol{u} = [\underline{u}, \overline{u}]$ and $\boldsymbol{v} = [\underline{v}, \overline{v}]$ as

$$\boldsymbol{u} + \boldsymbol{v} = [\underline{u} + \underline{v}, \overline{u} + \overline{v}], \tag{4}$$

$$\boldsymbol{u} \cdot \boldsymbol{v} = [\min\{\underline{u}\cdot\underline{v}, \underline{u}\cdot\overline{v}, \overline{u}\cdot\underline{v}, \overline{u}\cdot\overline{v}\}, \max\{\underline{u}\cdot\underline{v}, \underline{u}\cdot\overline{v}, \overline{u}\cdot\underline{v}, \overline{u}\cdot\overline{v}\}]. \tag{5}$$

Moreover, the simultaneous region of possible values of the OLS-estimator has the form

$$B = \{(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y : y \in \boldsymbol{y}\}$$

and it is apparent that it is a convex polytope in the parameter space. Moreover, we get a more precise geometric characterization of B.

Proposition 2 ([4, 1]). *The set B is a zonotope centered in the point $\widehat{\beta}_C = \frac{1}{2}(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}(\underline{y} + \overline{y})$ with generators $g_i := q_i(\overline{y}_i - \underline{y}_i)$ with $i = 1, \ldots, n$, where q_i is ith column of the matrix $(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$.* □

Now we turn our attention to another important statistic: the estimator of the standard error of the error terms ε, which has the form

$$\widehat{\sigma} = \sqrt{\frac{1}{n-p}\cdot y^{\mathrm{T}}(I-H)y},$$

where $H = X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$ is the hat matrix. We would like to obtain the range of its possible values, i.e.

$$\overline{\widehat{\sigma}} = \sup\left\{\sqrt{\frac{1}{n-p}\cdot y^{\mathrm{T}}(I-H)y} : y \in \boldsymbol{y}\right\},$$

$$\underline{\widehat{\sigma}} = \inf\left\{\sqrt{\frac{1}{n-p}\cdot y^{\mathrm{T}}(I-H)y} : y \in \boldsymbol{y}\right\}.$$

The value $\underline{\widehat{\sigma}}$ can be computed efficiently: indeed, it suffices to solve the convex quadratic program

$$\min y^{\mathrm{T}}(I-H)y \text{ s.t. } \underline{y} \le y \le \overline{y}. \tag{6}$$

On the other hand, we have the following disappointing result.

Proposition 3 ([11, 5, 2]). *Computation of $\overline{\widehat{\sigma}}$ is an NP-hard problem.* □

t	Y_t	K_t	L_t	t	Y_t	K_t	L_t	t	Y_t	K_t	L_t
1	173.5	125	78	6	193.0	110	115	11	182.2	133	112
2	203.4	120	80	7	190.3	90	145	12	224.9	161	110
3	166.4	115	92	8	208.6	101	143	13	175.3	141	85
4	202.2	135	102	9	172.5	133	99	14	207.4	132	92
5	195.9	132	90	10	190.1	127	89	15	153.3	105	79

Table 1 Data for the Cobb-Douglas function.

Intuitively, computation of $\overline{\widehat{\sigma}}$ is a convex maximization over the n-dimensional cube $\boldsymbol{y}$, and it is easy to see that the maximum is attained in one of its vertices. Therefore, we have the following exponential-time algorithm (and better cannot be expected due to the result of Proposition 3):

$$\overline{\widehat{\sigma}} = \max\left\{\sqrt{\frac{1}{n-p}\cdot(y_C+\operatorname{diag}(s)y_\Delta)^{\mathrm{T}}(I-H)(y_C+\operatorname{diag}(s)y_\Delta)} : s\in\{-1,1\}^n\right\}, \tag{7}$$

where $y_C = \frac{1}{2}(\overline{y}+\underline{y})$ is the center of $\boldsymbol{y}$ and $y_\Delta = \frac{1}{2}(\overline{y}-\underline{y})$ is the *radius* of $\boldsymbol{y}$. The algorithm (7) can be used for small n only; in the next section, we will use it with $n = 15$, which is a situation when the number $2^n = 32768$ is still tractable.

3 Example

As an example we consider the Cobb-Douglas production function of the form

$$\ln Y_t = \beta_0 + \beta_1 \ln K_t + \beta_2 \ln L_t + \varepsilon_t,$$

where Y_t, K_t, L_t stand for output, capital and labor, respectively, and the error terms ε_t are assumed iid with zero mean and finite variance. Then, OLS estimator is applicable. We will use data from Table 1. We have $n = 15$ and $p = 3$. We will study the possible impact of a change in the dependent variable $\ln Y_t$ on the OLS estimates $\widehat{\beta}_1$ and $\widehat{\beta}_2$ and the estimate $\widehat{\sigma}$ (which has further consequences, since it affects the estimate of the covariance matrix of the estimator $\widehat{\beta}$, t-ratios etc.).

We will replace the vector $\ln Y_t$ by the interval vector

$$\boldsymbol{z} := [-\delta + \ln Y_t, \delta + \ln Y_t],$$

where $\delta \in [0, 0.1]$. For example, the choice $\delta = 0.05$ shows us, how the estimates $\widehat{\beta}$ and $\widehat{\sigma}$ can change in the worst case, if the output variable $\ln Y_t$ is measured with an error at most ± 0.05. This can happen, for example, when we round non-integer numbers to one decimal place. Figure 2 shows the resulting intervals $[\underline{\widehat{\beta}_1}, \overline{\widehat{\beta}_1}]$, $[\underline{\widehat{\beta}_2}, \overline{\widehat{\beta}_2}]$, $[\underline{\widehat{\sigma}}, \overline{\widehat{\sigma}}]$. The intervals $[\underline{\widehat{\beta}_1}, \overline{\widehat{\beta}_1}]$ and $[\underline{\widehat{\beta}_2}, \overline{\widehat{\beta}_2}]$ were computed using (3), the values $\underline{\widehat{\sigma}}$ were computed using (6) and the values $\overline{\widehat{\sigma}}$ were computed using (7).

We can see that even with a small δ (i.e. with a narrow interval vector $\boldsymbol{z}$) the ranges of possible values of $\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\sigma}$ are quite large.

Assuming normality of error terms, it is also tempting to have a look at the t-ratio for testing the null hypothesis $c^{\mathrm{T}}\beta = \gamma_0$, where c is a vector of parameters and γ_0 is a constant. The test statistic has the form

$$t = \frac{c^{\mathrm{T}}\widehat{\beta} - \gamma_0}{\sqrt{\widehat{\sigma}^2 c^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}c}}.$$

In fact, it is not easy to compute the upper bound $\overline{t}$ and the lower bound $\underline{t}$ exactly; it amounts to solving the optimization problems

$$\overline{t} = \max\left\{\frac{c^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y - \gamma_0}{\sqrt{\frac{y^{\mathrm{T}}(I-X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}})y}{n-p}c^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}c}} : \underline{y}\le y\le\overline{y}\right\}, \tag{8}$$

$$\underline{t} = \min\left\{\frac{c^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y - \gamma_0}{\sqrt{\frac{y^{\mathrm{T}}(I-X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}})y}{n-p}c^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}c}} : \underline{y}\le y\le\overline{y}\right\}, \tag{9}$$

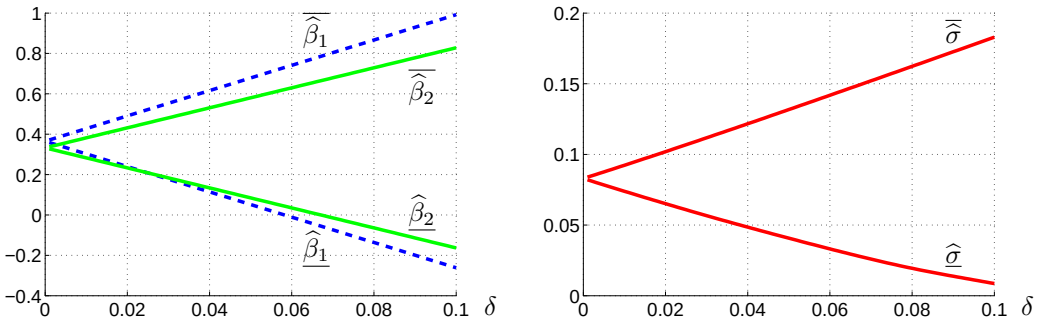


Figure 2 Ranges for $\widehat{\beta}_1$, $\widehat{\beta}_2$ and $\widehat{\sigma}$ with $\delta \in [0, 0.1]$.

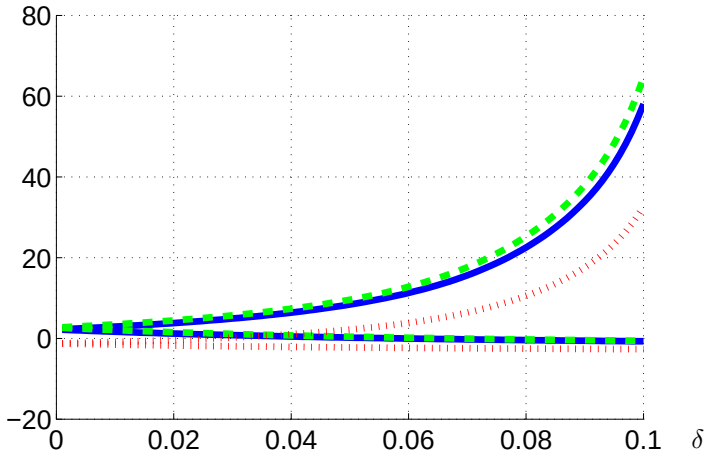


Figure 3 Approximate ranges for t-ratios for testing the hypotheses $\beta_1 = 0$ (solid), $\beta_2 = 0$ (dashed), $\beta_1 + \beta_2 = 1$ (dotted).

where X, c, γ_0, n, p are data and y are variables. The optimization problems seem to be difficult due to the occurrence of y both in the numerator and the denominator. And indeed, the bad news is that (8) is an NP-hard problem [2]. (The complexity-theoretic classification of (9) is not known at the moment.)

However, we will use the following approximation: we set

$$\overline{t}^* = \frac{c^{\mathrm{T}}\overline{\widehat{\beta}} - \gamma_0}{\sqrt{\underline{\widehat{\sigma}}^2 c^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}c}}, \qquad \underline{t}^* = \frac{c^{\mathrm{T}}\underline{\widehat{\beta}} - \gamma_0}{\sqrt{\overline{\widehat{\sigma}}^2 c^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}c}},$$

since the values $\underline{\sigma}$ and $\overline{\sigma}$ are available from Figure 2. We make three choices of c:

$c^{\mathrm{T}} = (0, 1, 0),\ \gamma_0 = 0$ for testing the hypothesis $\beta_1 = 0$,
$c^{\mathrm{T}} = (0, 0, 1),\ \gamma_0 = 0$ for testing the hypothesis $\beta_2 = 0$,
$c^{\mathrm{T}} = (0, 1, 1),\ \gamma_0 = 1$ for testing the hypothesis that returns to scale are constant.

The resulting intervals $[\underline{t}^*, \overline{t}^*]$ are plotted in Figure 3. Observe that even with small values of δ, the interval is quite large. It shows that even with a narrow interval $\boldsymbol{z}$, the conclusions of the t-test must be read carefully — it might happen that the value of the t-statistic can attain values from a quite wide interval.

4 Conclusions

We have studied how a replacement of real-valued data by interval data in a regression model can affect the values of some usual statistics, such as the OLS-estimate of regression parameters, estimate of variance

of error terms or t-ratios. Complexity-theoretic results show that these questions are generally very hard (usually NP-hard or co-NP-hard) in the general setting. The situation is slightly better when we restrict ourselves to the case when only the observations of the dependent variable are affected by interval uncertainty. We have presented an application in modeling the Cobb-Douglas production function.

Acknowledgements

The work was supported by the project P403/12/1947 of the Czech Science Foundation and the project F4/11/2013 of the Internal Grant Agency of University of Economics, Prague.

References

[1] Černý, M., Antoch, J., and Hladík, M.: On the possibilistic approach to linear regression models involving uncertain, indeterminate or interval data. *Information Sciences* **244**, 2013, 26–47.

[2] Černý, M., and Hladík, M.: Complexity of computation and approximation of the t-ratio over one-dimensional interval data. Submitted, 2013.

[3] Černý, M., and Hladík, M.: Interval data, algorithms and computational complexity (In Czech: Intervalová data, algoritmy a výpočetní složitost). *Bulletin of the Czech Statistical Society*, in press, 2013.

[4] Černý, M., and Rada, M.: On the possibilistic approach to linear regression with rounded or interval-censored data. *Measurement Science Review* **11** (2), 2011, 34–40.

[5] Ferson, S., Ginzburg, L., Kreinovich, V., Longpré, L., and Aviles, M.: Exact bounds on finite populations of interval data. *Reliable Computing* **11** (3), 2005, 207–233.

[6] Hladík, M., and Černý, M.: Interval regression by tolerance analysis approach. *Fuzzy Sets and Systems* **193**, 2012, 85–107.

[7] Moore, R., Kearfott, R., and Cloud, M.: *Introduction to Interval Analysis.* SIAM, Philadelphia, USA, 2009.

[8] Neumaier, A.: *Interval Methods for Systems of Equations.* Cambridge University Press, 1990.

[9] Schneeweiss, H., and Komlos, J.: *Probabilistic Rounding and Sheppard's Correction. Technical Report 045.* Department of Statistics, University of Munich, 2008.

[10] Stoye, J.: Partial identification of spread parameters. *Quantitative Economics* **1**, 2010, 323–357.

[11] Vavasis, S. A.: *Nonlinear Optimization. Complexity Issues.* Oxford University Press, 1991.

Modeling Effects of Economic Crisis on Potential Output: Application to the Eurozone

Ondřej Čížek[1]

Abstract. Economic cycles and long run relations are traditionally studied separately, which is not appropriate in the context of the current economic crisis. The goal of this paper is to propose a framework, which links short run fluctuations with long run dynamics within unobserved components methodology. The formulated model is nonlinear and is put in a conditionally Gaussian type of state space form. The model will be analyzed with respect to identifiability and estimated by maximum likelihood method. Finally, estimates and forecasts of potential output of the Euro zone will be calculated for the purpose of illustration and the obtained results will be interpreted.

Keywords: unobserved components method, Kalman filter, maximum likelihood, output gap, potential output, economic crisis.

JEL Classification: C51
AMS Classification: 90C15

1 Introduction

Nowadays, there is a debate among economists and politics whether or not governments should implement active economic policy to stimulate the economy or whether they should take low-budget measures to reduce indebtedness. There is no doubt that indebtedness is a serious economic problem of the Eurozone, but even despite this fact, there is a widespread consensus that now is not a time to tackle this problem. The argument is that contractionary economic policy would cause serious damages to economic activity in a situation of the current economic recession characterized by high unemployment rate. Lots of companies would definitely go bankrupt and many people would loose their change of ever finding a job. A short run transitory recession would thus overgrow into a permanent long run fall in economic activity which describes the relation between economic cycles and long run trends. Traditionally, economic cycles and long run relations are studied separately, which is not appropriate in the context of the current economic crisis. The goal of this paper is to propose a framework, which describes the above mentioned effects of short run fluctuations on long run dynamics.

2 Model

I will use unobserved components methodology, which was pioneered by Watson [6], who separated output into a trend and a cycle. The (natural logarithm of) output y_t is specified as:

$$y_t = \bar{y}_t + \hat{y}_t, \tag{1}$$

where $\bar{y}_t$ is potential output and $\hat{y}_t$ represents output gap.

Watson [6] assumed that trend component follow a random walk with drift and a cyclical component was assumed to follow an AR(2) process. I will differ from this popular specification. Linear trend of the potential output will not be affected by random errors, but by the output gap, which will describe the effect of the business cycle on the long run growth of the potential:

$$\bar{y}_t = \bar{y}_{t-1} + \mu + \alpha_{t-1} \cdot \hat{y}_{t-1}, \tag{2}$$

where

[1] University of Economics, Department of Econometrics, sqr. W. Churchille 4, Prague 3, cizeko@vse.cz.

$$\alpha_{t-1} = \begin{cases} 0, & \hat{y}_{t-1} \geq 0 \\ \alpha, & \hat{y}_{t-1} < 0 \end{cases} \tag{3}$$

Time-varying parameter α_{t-1} becomes zero if $\hat{y}_{t-1} \geq 0$, which describes the fact that increasing output above potential does not have any effects on the dynamics of potential output. On the contrary, $\alpha_{t-1} = \alpha > 0$ when $\hat{y}_{t-1} < 0$.

The output gap is assumed to follow AR(1) process, which is, however, modified by an inertial variable ν_t representing external shock to the European economy caused by the economic crisis in the United States:

$$\hat{y}_t = \lambda \cdot \hat{y}_{t-1} + \nu_t + u_t, \tag{4}$$

$$\nu_t = \lambda \cdot \nu_{t-1} + \varepsilon_t, \tag{5}$$

where $u_t \sim N\left(0, \sigma^u\right)$ is i.i.d. random error and

$$\varepsilon_t = \begin{cases} 0, & t \neq t_0 \\ \varepsilon, & t = t_0 \end{cases} \tag{6}$$

is a time-varying parameter, which is equal to $\varepsilon > 0$ only in time $t = t_0$, where t_0 is the second quarter of the year 2008, when the economic crisis from the United States hit the European economy.

The goal is to estimate the parameters μ, α, λ, ε and σ^u as well as the unobserved variables $\bar{y}_t$, $\hat{y}_t$, which is both done simultaneously by the applied methods.

3 Discussion of the concept of the output gap

The univariate approach can be extended to include other observable variables into the model. This was firstly done by Kuttner [4], who specified potential output as the level of output at which inflation is constant. The advantage of this multiple variable approach is that it uses additional information (e.g. from inflation) to estimate the parameters and especially to decompose output to its components. The other advantage is that the potential output has a precise economic meaning. Disadvantage of the multiple approach is that it is vulnerable to the specification errors of the structural economic equations (e.g. Phillips curve).

The idea of using (observable) inflation as an indicator for the (unobservable) potential output stems from the historical experience with high inflation caused by expansionary economic policy trying to stimulate output. It turned out that high inflation was the only effect of the expansion in the long run. From this point of view, it is then natural to say that the level of output causing rise in inflation is not sustainable in the long run and treat this excess of the output as a short run business cycle fluctuations.

There are, however, problems with defining potential output as the level of output at which inflation is constant. Typical example is the other historical experience with the current economic crisis, which showed that stable inflation is not a sufficient condition for the output to be sustainable in the long run. Typical reaction to this experience was to redefine potential output and to include other observable indicators of the potential output (e.g. indebtedness, prices of other assets) into the structural model (Borio et. al [2]). It is clear that including indebtedness as an indicator will lower the estimated potential output, because indebtedness is very high. It can be then argued that if the potential output had been defined using the indebtedness as an indicator before the economic crisis, it would gave us a signal of output above potential, which caused "overheating" of the economy.

Nevertheless, I do not consider the approach of adding additional observables into the structural model to be helpful as future experience will probably show that lots of other indicators would be needed. Especially the indebtedness is not an appropriate indicator of the level of output, which is sustainable in the long run. In my view, overheating of the US economy was not caused by high output, but by the problems in the financial system, which needed reforms of its institutional framework. In this context, Sims [5] argues that there are gaps in the institutional structure of the Euro area, which will need to be tackled, in order to solve the problem of indebtedness of some countries of the Euro zone. Sims, therefore, proposes institutional reforms to solve high indebt-

edness of the Eurozone and not contractionary economic policy, which would now only cause high damages in the economic activity.

4 State space form

The formulated model is nonlinear because of the time-varying parameter α_{t-1}. It can, however, be put in a conditionally Gaussian type of state space form, to which the linear Kalman filter algorithm can be applied to form a likelihood function.

The transition and the measurement equation of the conditionally Gaussian state space form is as follows:

$$\mathbf{x_t} = \mathbf{A_t}\left(\mathbf{Z_{t-1}}\right)\cdot\mathbf{x_{t-1}} + \mathbf{R_t}\left(\mathbf{Z_{t-1}}\right)\cdot\mathbf{u_t}, \tag{7}$$

$$\mathbf{z_t} = \mathbf{D_t}\left(\mathbf{Z_{t-1}}\right)\cdot\mathbf{x_t} + \boldsymbol{\upsilon}_\mathbf{t}, \tag{8}$$

where $\mathbf{A_t}\left(\mathbf{Z_{t-1}}\right)$ represents that the matrix $\mathbf{A_t}$ depends on the values of the observed variables till the time $t-1$, i.e. on the vector $\mathbf{Z_{t-1}} \equiv \left(\mathbf{z_1}', ..., \mathbf{z_{t-1}}'\right)'$.

In the case of the formulated model, the state vector is defined as $\mathbf{x_t} = \left(\bar{y}_t \quad \hat{y}_t \quad \nu_t \quad 1_t\right)'$ and the matrices from the transition equation (7) take the form:

$$\mathbf{A_t}\left(\mathbf{Z_{t-1}}\right) = \begin{pmatrix} \mathbf{e_1} + \mu\cdot\mathbf{e_4} + \alpha_{t-1}\cdot\mathbf{e_2} \\ \lambda\cdot\mathbf{e_2} - \left(\lambda\cdot\mathbf{e_3} + \varepsilon_t\cdot\mathbf{e_4}\right) \\ \lambda\cdot\mathbf{e_3} + \varepsilon_t\cdot\mathbf{e_4} \\ \mathbf{e_4} \end{pmatrix} \qquad \mathbf{R_t}\left(\mathbf{Z_{t-1}}\right) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

where $\mathbf{e_k}$ is a $\left(1 \text{ x } 4\right)$ vector, which has a number 1 at the k-th position and zeros elsewhere.

The matrix from the measurement equation (8) is as follows:

$$\mathbf{D_t}\left(\mathbf{Z_{t-1}}\right) = \left(\mathbf{e_1} + \mathbf{e_2}\right).$$

Vector of random errors in the transition equation is $\mathbf{u_t} = u_t$ and the vector of measurement errors $\boldsymbol{\upsilon}_\mathbf{t}$ is a vector of zeros.

Because the output gap $\hat{y}_t$ is unobservable, the time-varying parameter α_{t-1} had to be slightly redefined as:

$$\alpha_{t-1} = \begin{cases} 0, & \hat{y}_{t-1|t-1} \geq 0 \\ \alpha, & \hat{y}_{t-1|t-1} < 0 \end{cases}$$

where $\hat{y}_{t-1|t-1}$ represents the estimate of the unobserved variable $\hat{y}_{t-1}$, which was formed by the Kalman filter algorithm on the basis of the information available in $\mathbf{Z_{t-1}} \equiv \left(\mathbf{z_1}', ..., \mathbf{z_{t-1}}'\right)'$.

5 Econometric methodology

Full Information Maximum Likelihood method was used to estimate the parameters. Using the assumption that random error u_t is normally distributed, Kalman filter algorithm was applied to the model written in state space form. The outcomes of this algorithm were then used to construct likelihood function, which was then maximized by standard numerical procedures in Matlab. The construction of the likelihood function is described in Harvey [3]. The square-root version of the Kalman algorithm was used, in order to attain grater numerical precision. Description of this algorithm can be found in Anderson, Moore [1]. Parameters of the model were appropriately transformed, in order to ensure that their values lie within economically reasonable intervals. Because of

the problem with local extremes, lots of initial points were randomly generated, from which the Matlab procedure called fminunc was initialized.

6 Calibration of selected parameters

The applied econometric methodology enables to estimate all the parameters of the model. It is, however, necessary to calibrate some of them because of the limited number of observations in the data set, which is in not large enough to guarantee precise estimates of the parameters characterizing long run relationships. It will be assumed that $\mu = 0.0025$, which means that potential output would grow 1% per year if $\hat{y}_{t-1} \geq 0$. I will also assume that $\bar{y}_0 = 14.4943$, which was obtained as an ordinary least square estimate from a linear regression model $y_t = y_0 + 0.0025 \cdot t$, $t = 1,...,t_0 - 1$.

The econometric estimate of the parameter α was unrealistically high. Potential output closely followed the output, because only small changes in the output gap sufficed for that. Almost all variability in the output was caused by the variability in the potential. It will be, therefore, more reasonable to calibrate the coefficient α. I will base the calibration on the fact that very small value of α will mean almost linear trend for the output gap. It is clear that increasing α will cause higher variability in the growth rate of the potential output. From this point of view, the parameter α is an analogy to the smoothing parameter ω in the generally known Hodrick-Prescott filter defined by the optimization problem:

$$\underset{\{\bar{y}_t\}_{t=1}^{T}}{Min} \sum_{t=1}^{T}\left(y_t - \bar{y}_t\right)^2 + \omega \cdot \sum_{t=2}^{T-1}\left[\left(\bar{y}_{t+1} - \bar{y}_t\right) - \left(\bar{y}_{t+1} - \bar{y}_t\right)\right]^2 .$$

I will calibrate the parameter α in such a way so as to ensure the required (economically sensible) variability of the growth rate of the potential output. Combining the following facts:

$$\mathrm{var}\left(\bar{y}_t - \bar{y}_{t-1}\right) = \mathrm{var}\left(\alpha_{t-1} \cdot \hat{y}_{t-1}\right),$$

$$\frac{\mathrm{var}\left(\alpha_{t-1} \cdot \hat{y}_{t-1}\right)}{\mathrm{var}\left(\alpha \cdot \hat{y}_{t-1}\right)} \approx 0.34,$$

will yield

$$\mathrm{var}\left(\bar{y}_t - \bar{y}_{t-1}\right) = 0.34 \cdot \alpha^2 \cdot \mathrm{var}\left(\hat{y}_{t-1}\right),$$

or

$$\alpha = \frac{\sigma^{\Delta\bar{y}}}{\sigma^{\hat{y}}} \cdot \sqrt{\frac{1}{0.34}},$$

where $\sigma^{\Delta\bar{y}} = \sqrt{\mathrm{var}\left(\bar{y}_t - \bar{y}_{t-1}\right)}$,
$\sigma^{\hat{y}} = \sqrt{\mathrm{var}\left(\hat{y}_{t-1}\right)}$.

The value of the parameter α is determined by the required ratio of the standard error of the growth rate of potential output $\sigma^{\Delta\bar{y}}$ to the standard error of the output gap $\sigma^{\hat{y}}$. I will assume that:

$$\frac{\sigma^{\Delta\bar{y}}}{\sigma^{\hat{y}}} = \frac{0.0625\ \%}{5\ \%} = \frac{1}{80},$$

which gives the value for the parameter $\alpha = 0.0216$.

7 Data

I used the data on quarterly nominal GDP (in millions of Euro) for the Euro zone 17 in fixed composition. The dataset begins at the first quarter of 1995 and ends at the fourth quarter of 2012. The data were already working day and seasonally adjusted. GDP deflator was then used to calculate real GDP. The complete datasets can be found at the following internet addresses:

http://sdw.ecb.europa.eu/quickview.do?node=9484571&SERIES_KEY=119.ESA.Q.I6.Y.0000.B1QG00.1000.TTTT.L.U.A

http://sdw.ecb.europa.eu/browseTable.do?ADJUSTMENT=Y&saf8=1&REF_AREA=566&ESA95_ACCOUNT=B1QG00&node=2120780&FREQ=Q&saf3=3&sfl2=3&saf4=1&sfl1=3&saf5=1&saf6=1&sfl3=4&DATASET=0&saf7=1&advFil=y

8 Interpretation of the parameter estimates

Econometric estimates of the parameters, which were not calibrated is $\lambda = 0.9619$, $\varepsilon = 0.0125$, $\sigma^u = 0.0059$. Values of these parameters are reasonable from an economic point of view. High value of the coefficient λ is in line with the common view that output gap is highly persistent. The value of the parameter ε can be interpreted in such a way that output fell below potential by more than 1 % at time $t = t_0$ at the beginning of the economic crisis in the European union as a consequence of the economic crisis in the USA. This negative economic shock was however persistent due to the high value of λ and caused additional falls in the output below potential in subsequent periods by $1.25 \cdot 0.9619^{t-t_0}$ %.

9 Estimating and forecasting potential output

Estimate of the potential output is depicted at the graph 1. It can be seen that the growth rate of the potential output was zero since the beginning of the economic crisis in 2008. Forecasts 10 years ahead also show that potential output will not rise during this long period. Dynamics of the output is also depicted at this graph. It predicts that output will rise quite rapidly, but this will happen only because the diminishing output gap. The model also estimates that output is now approximately 14 % below potential and predicts that output in 2023 will still be approximately 8 % below potential even after the its growth.

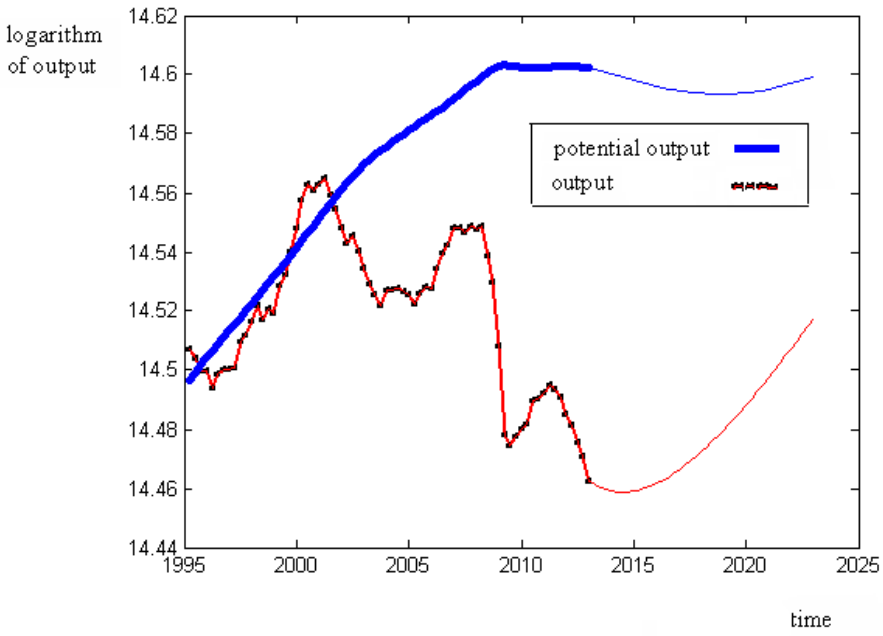


Figure 1 Estimates and forecasts of the (potential) output

10 Conclusion

Economic cycles and long run relations are traditionally studied separately, which is not appropriate in the context of the current economic crisis. In the presented paper, I formulated a model, which links short run fluctuations with long run dynamics within unobserved components methodology. Some of the parameters had to be calibrated, while the others were econometrically estimated by the method of maximum likelihood. Estimation of the parameters was in line with stylized facts. The model was also used to forecast (potential) output of the Eurozone with a forecasting period of 10 years, which showed that potential output will not rise and output will be below potential by 8 % in 2023. The model also predicts that the output in 2023 will still be at a lower level than in pre-crisis periods.

Acknowledgements

Financial support of IGA F4/1/2012 is gratefully acknowledged by the author.

References

[1] Anderson, B. D. O., Moore, J. B.: *Optimal filtering*. Prentice-Hall, 1979.
[2] Borio, C., Disyatat, P., Juselius, M.: Rethinking potential output: Embedding information about the financial cycle. Bank for International Settlements, Working Paper No. 404.
[3] Harvey, A. C.: *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, 1989.
[4] Kuttner, K. N.: Estimating potential output as a latent variable. *Journal of Business and Economic Statistics*, 1994, **12**(3):361-68.
[5] Sims, C.: Gaps in the institutional structure of the euro area. *Financial Stability Review*, Banco de France, April 2012.
[6] Watson, M. W.: Univariate detrending methods with stochastic trends. *Journal of Monetary Economics* **18** (1986), 49-75.

Subsidy Allocation as a Coalition Game

Martin Dlouhý[1]

Abstract. The City of Prague allocates a part of its budget to the city districts in the form of subsidies. The objective of the City of Prague is to find such a set of criteria and their weights that estimate in the best way the expected cost of public administration. The objective of city district is subsidy maximization by influencing the criteria and their weights. The change of allocation rules requires that the city district is able to form a coalition of city districts with majority of votes. The strength of each city districts is evaluated by the number of members of the Prague City Assembly coming from the given city district. The conflict between the city districts is studied as a coalition game. The illustrative calculations for the 2013 subsidy allocation are presented.

Keywords: City of Prague, subsidy allocation, game theory, coalition game.
JEL Classification: H72, C71
AMS Classification: 91A12

1 Introduction

The City of Prague, the capital of the Czech Republic, has an administrative structure of three tiers. The highest tier is the City of Prague itself, represented by the Prague City Assembly and the Magistrate of the City of Prague. At the lowest level, the City of Prague is divided into 57 self-governing city districts. The city districts differ in the size of population and territory. The city districts Prague 1–22 are the so-called "large" city districts with high numbers of inhabitants (from 6 627 to 127 723) and are mostly located in the inner city. The districts Prague 23–57 are the so-called "small" city districts with much lower numbers of inhabitants (from 277 to 10 071) and are mostly located in suburban areas. In total, less than 10% of population live in the city districts Prague 23–57. In 2001, the middle administrative level was introduced. 57 city districts have been grouped into 22 administrative districts for state-administration purposes. One "large" city district in each administrative district has responsibility for providing services of state administration for the entire administrative district. One has to distinguish whether by Prague 4 is meant one of the self-governing city districts Prague 1–57 or one of the administrative districts Prague 1–22.

The City of Prague allocates a part of its budget to the city districts in the form of subsidies. The objective of the City of Prague is to find such a set of criteria and such a set of weights that estimate in the best way the expected cost of public administration. The objective of city district is subsidy maximization by manipulating the criteria and their weights. The decision about the subsidy allocation is made by the Prague City Assembly (63 elected members). During the election period 2010 - 2014, four political parties are represented in the Prague City Assembly: TOP 09 (26 members), ODS (20 members), ČSSD (14 members) and KSČM (3 members). The strength of each city districts is evaluated by the number of members of the Prague City Assembly coming from the given city district. This means that members of the assembly form also coalitions based on the city districts they come from, irrespectively of their political membership and the current governing coalition (TOP 09 and ODS). Not surprisingly, each year the subsidy allocation to city districts is a subject of political struggle among and within the political parties and among the city districts. This process is an interesting topic for the theory of coalition games (Osborne, 2004, Dlouhý, Fiala, 2009).

In brief, we summarize the historical development of the subsidy allocation process for the large city districts Prague 1–22 between years 2001 and 2013. In 2001 and 2002, the city districts received three types of subsidies: 30% of income tax paid by physical persons-entrepreneurs and the so-called "direct subsidy" and "criteria subsidy". The criteria subsidy was based on five criteria: population (50%), the number of pupils at primary schools and kindergartens (35%), the area of urban green (10%), the road maintenance measured as area of roads in square meters (5%). In 2002, the same model as in the year 2001 was used with some minor modifications. In 2003, the special solidarity model was used due to floods in August 2002.

In 2004 - 2006, the index method was used, which was derived from the annual growth of taxes collected by the City of Prague. The minimum average subsidy per inhabitant was introduced in 2004 at the level of 1900 Czech korunas (CZK). This minimum average subsidy was increased to 2000 CZK in 2005 and to 2100 CZK in 2006. In 2007 - 2009, the subsidy allocation processes for the large city districts (Prague 1–22) and the small city districts (Prague 23–57) were separated and varied indicators were used for these two groups of city districts. This separation was used in the following years. The allocation for the large city districts Prague 1–22 was based

[1] University of Economics in Prague, Faculty of Informatics and Statistics, Department of Econometrics, e-mail: dlouhy@vse.cz.

on three criteria: income tax paid by physical persons-entrepreneurs, population and territory. The minimum average subsidy per inhabitant was gradually increased: 2200 CZK in 2007, 2300 CZK in 2008, and 2400 CZK in 2009. The allocation for the small city districts Prague 23–57 was based on six criteria: population (40%), territory (15%), the number of pupils at primary schools and kindergartens (20%), the area urban green (10%), the road maintenance measured as area of roads in square meters (10%), and units of volunteer fire fighters (5%). In year 2010, due to the worsened economic situation in the country, the subsidy for each city districts was indexed by 93.5% of the 2009 subsidy level; however the minimum subsidy per inhabitant 2400 CZK was preserved. In 2011, the 2009 criteria were used again and the minimum average subsidy 2400 CZK per inhabitant was used.

In 2012 - 2013, six allocation criteria were used to calculate subsidies for the large city districts Prague 1–22 (City of Prague, 2011, 2012). The city districts received 30% of income tax paid by physical persons-entrepreneurs (average 2008–2010), which made 9.5% of the total subsidy. The rest of the total subsidy was distributed according to five criteria and their weights: population (30%), territory in square kilometres (10%), the number of pupils at primary schools and kindergartens owned by the city district (30%), the area of urban green served by the city district in hectares (20%), and the road maintenance measured as the area of the roads served by the city district in square meters (10%). It may look as if the income tax has some special position in the allocation. However, it just means that the real weights of criteria in 2013 were as follows: income tax paid by physical persons–entrepreneurs (6.8%), population (28.0%), territory in square kilometres (9.3%), the number of pupils and students at the schools owned by the city district in 2011 (28.0%), the area of urban green served by the city district (18.6%), and the road maintenance measured as the area of the roads served by the city district in square meters (9.3%). For the small city districts Prague 23–57, a different six-criterion subsidy allocation model (see criteria and weights in period 2007-2009). The minimum average subsidy 2400 CZK per inhabitant was again used. Moreover, two city districts received additional direct subsidy for public dumps.

2 Data

In this study, we will focus on the large city districts Prague 1–22 that are responsible for both the self-governing public administration and state administration. The input data on six criteria used for the 2013 subsidy allocation among city districts Prague 1–22 are presented in Table 1 (City of Prague, 2012).

District	Population	Territory (km^2)	Pupils and Students	Urban Green (ha)	Roads (m^2)	Income Tax
Prague 1	29 857	5.51	2 977	43.0	13 413	66 901
Prague 2	49 237	4.18	4 367	58.5	8 862	56 863
Prague 3	71 140	6.48	5 272	68.7	34 289	39 433
Prague 4	127 723	24.20	11 447	248.6	261 144	87 048
Prague 5	81 161	27.50	6 366	87.0	204 290	58 129
Prague 6	97 951	41.56	9 694	209.0	501 522	99 305
Prague 7	41 416	7.12	3 037	7.0	30 630	23 826
Prague 8	103 757	21.80	9 091	207.0	148 652	58 036
Prague 9	53 382	13.31	3 486	120.0	140 641	17 062
Prague 10	108 998	18.60	7 641	145.0	208 085	50 748
Prague 11	77 170	9.79	7 316	227.4	313 640	23 021
Prague 12	54 426	23.32	5 191	136.0	633 960	18 442
Prague 13	59 874	13.23	6 781	166.4	502 047	23 057
Prague 14	45 822	13.53	4 189	91.4	400 229	9 825
Prague 15	30 954	10.25	2 805	53.9	168 983	10 237
Prague 16	8 141	9.30	935	14.2	253 430	4 999
Prague 17	24 973	3.25	2 174	38.9	136 475	7 106
Prague 18	17 323	5.61	1 739	24.0	201 770	1 934
Prague 19	6 627	6.00	763	22.6	137 619	1 988
Prague 20	15 028	16.94	2 094	46.6	663 611	7 898
Prague 21	10 334	10.15	1 144	33.5	313 008	6 562
Prague 22	9 025	15.62	1 159	20.4	256 425	4 035

Table 1: Input Data 2013, Prague 1 – Prague 22

3 Methods

Let us have m city districts and n allocation criteria. In this case, the values are m=22 and n=6. The real values for city district i in each criteria j are y_{ij}. The relative proportion of city district i in dimension j on the total sum for all city districts is denoted r_{ij}, and is calculated as

$$r_{ij} = y_{ij} / \sum_{i=1}^{m} y_{ij} . \tag{1}$$

The total subsidy is denoted B. The subsidy to city district i is denoted s_i. Let us further suppose that each criterion will have the minimum weight at least 0.05. The maximum value of weight is (1-0.05(n-1)). The best case scenario, an upper limit of the subsidy for the ith city district u_i, is that subsidy is calculated as

$$u_i = ((1-0.05n)\max_j r_{ij} + 0.05\sum_{j=1}^{n} r_{ij})B. \tag{2}$$

On the other hand, the worst case scenario, a lower limit of the subsidy l_i, is calculated as

$$l_i = ((1-0.05n)\min_j r_{ij} + 0.05\sum_{j=1}^{n} r_{ij})B. \tag{3}$$

It is certainly good if the city district knows what is upper subsidy limit and optimal weights, however the city district needs to persuade a sufficient number of members of the Prague City Assembly for adoption of its proposal. We assume that the members from other city districts will agree only if the proposed subsidy allocation will also be profitable for the city districts they come from. This is a basic assumption about the players to cooperate in coalition games (Dlouhý, Fiala, 2009). In this case a coalition of city districts C with minimally 32 out of 63 votes in the Prague City Assembly has to be formed. The subsidy from the proposed allocation s_i has to be higher than the existing subsidy allocation s^0_i for the coalition members:

$$s_i \geq s_i^0, \quad i \in C. \tag{4}$$

If the allocation related to the upper limit u_q is profitable also for other city districts with majority of votes, the proposal will adopted. It is however likely that such proposal will not achieve the majority of votes. In such case, the city district has to look for the best available compromise. City district q can achieve the maximum level of subsidy s_q by changing the weights and at the same time having coalition with minimally 32 votes, which can be obtained as a solution of the following program:

$$\begin{aligned}
&\max s_q = \sum_{j=1}^{n} v_{qj} r_{qj} \\
&\text{subject to} \\
&0.05 \leq v_{qj} \qquad \text{for } j = 1, 2, \ldots 6, \\
&\sum_{j=1}^{n} v_{qj} = 1, \\
&\sum_{j=1}^{n} v_{qj} r_{ij} \leq s_i^0 - Mh_i \quad \text{for } i = 1, 2, \ldots, 22, \\
&\sum_{i=1}^{m} w_i h_i \geq 32,
\end{aligned} \tag{5}$$

where h_i is a binary variable representing the vote of each city district, w_i is number of members of the Prague City Assembly from city district i, M is sufficiently high constant, and v_{qj} are optimal weights for the city district q. This program has to have a feasible solution, because the existing subsidy allocation with weights (0.280, 0.093, 0.280, 0.186, 0.093, 0.068), which represent the case of the grand coalition, is included in the feasible set.

Alternatively, we can formulate a model which assumes that the condition of 32 votes is necessary for adoption of the subsidy allocation proposal, however this number of votes has to be achieved by the governing coalition TOP 09 and ODS, not by a random group of members of the Prague City Assembly. The city district i

is a member of coalition *C* if there is at least one member of the Prague City Assembly from the governing coalition (TOP 09 and ODS) and comes the given city district. The city district *i* is not a member of coalition *C* if no member of the Prague City Assembly from the given district is member of the governing coalition or if the given city district is not represented in the Prague City Assembly at all (as of March 2013). The allocation model can be formulated as follows:

$$\begin{aligned}
&\max \sum_{i \in C} s_i = \sum_{i \in C} \sum_{j=1}^{n} v_j r_{ij} \\
&\text{subject to} \\
&0.05 \le v_j \qquad \text{for } j = 1, 2, \ldots, 6; \\
&\sum_{j=1}^{n} v_j = 1; \\
&\sum_{j=1}^{n} v_j r_{ij} \le s_i^0 - M h_i \quad \text{for } i = 1, 2, \ldots, 22; \\
&\sum_{i=1}^{m} w_i h_i \ge 32.
\end{aligned} \tag{6}$$

The similar problem of dual structure can be observed in other situations. For example, Turnovec, Mercik, and Mazurkiewicz (2008) studied the case of the European Parliament. European Parliament has a dual structure because its members represent their own countries and at the same time they are clustered in European political parties, forming clubs in the Parliament.

The performance of the city district in the allocation process can be measured by the ratio of achieved subsidy and existing value of subsidy:

$$s_i / s_i^0 \times 100, \tag{7}$$

or by formula (8), which takes into account that the city district can benefit, but it can also loose, from any new allocation rules:

$$\frac{s_i - l_i}{u_i - l_i} \times 100. \tag{8}$$

4 An Illustrative Example

Firstly, we calculate the lower and upper subsidy limits u_i and l_i with formulas (2) and (3). The difference between the lower and upper limits shows how sensitive the district subsidy is to a change in the weights (Table 2). The last column in Table 2 shows the allocation criterion that is preferred by the given district, i.e. the criterion with the weight that equals to 0.75. Third criterion, the number of pupils and students, is not preferred by any city district. There are six extreme combinations of weights that can be theoretically optimal for the city districts: one of six weights with a value 0.75 and the other five weights with values 0.05. We denote them as *Proposal 1* to *Proposal 6*. *Proposal 1* has weights (0.75, 0.25, 0.25, 0.25, 0.25, 0.25), *Proposal 2* has weights (0.25, 0.75, 0.25, 0.25, 0.25, 0.25), etc.

Secondly, we investigate above mentioned six subsidy allocation proposals by checking the total number of votes for each proposal (Table 3). Two proposals (proposals 3 and 4) can achieve the majority of votes in the Prague City Assembly. As an illustrative example, the subsidy allocation related to *Proposal 3* and its impact on city districts is presented in Table 4. The relative performance of the city district is measured by performance criteria suggested in formulas (7) and (8).

District	s^0_i	l_i	u_i	$u_i - l_i$	Criterion
Prague 1	2.86	1.15	7.91	6.75	6
Prague 2	3.71	1.19	6.96	5.77	6
Prague 3	4.52	1.61	5.60	4.00	1
Prague 4	10.69	6.32	12.02	5.70	6
Prague 5	6.36	4.54	8.22	3.68	2
Prague 6	10.16	9.39	13.57	4.18	6
Prague 7	2.46	0.91	3.25	2.34	1
Prague 8	8.49	4.22	9.34	5.12	4
Prague 9	4.19	2.94	5.23	2.29	4
Prague 10	7.58	4.72	8.87	4.15	1
Prague 11	7.06	4.10	9.57	5.46	4
Prague 12	5.98	3.83	9.94	6.11	5
Prague 13	6.35	4.23	8.20	3.97	5
Prague 14	4.31	2.31	6.35	4.05	5
Prague 15	2.73	1.86	3.14	1.28	2
Prague 16	1.35	1.02	3.74	2.73	5
Prague 17	1.98	1.28	2.27	0.99	5
Prague 18	1.66	0.71	3.06	2.35	5
Prague 19	1.01	0.56	2.10	1.54	5
Prague 20	3.08	2.04	9.61	7.58	5
Prague 21	1.78	1.32	4.64	3.32	5
Prague 22	1.67	1.08	4.22	3.14	2

Table 2: Real, Minimum and Maximum Subsidies as a Percentage of Total Budget

District	Proposal 1	Proposal 2	Proposal 3	Proposal 4	Proposal 5	Proposal 6
Prague 1	0	0	4	0	0	4
Prague 2	3	0	3	0	0	3
Prague 3	6	0	6	0	0	6
Prague 4	5	0	5	5	0	5
Prague 5	2	2	2	0	0	2
Prague 6	0	8	0	8	0	8
Prague 7	1	0	1	0	0	1
Prague 8	6	0	6	6	0	0
Prague 9	1	1	0	1	0	0
Prague 10	7	0	0	0	0	0
Prague 11	0	0	0	2	0	0
Prague 12	0	4	0	4	4	0
Prague 13	0	0	5	5	5	0
Prague 14	0	0	0	0	0	0
Prague 15	0	4	4	0	4	0
Prague 16	0	0	0	0	0	0
Prague 17	0	0	0	0	0	0
Prague 18	0	1	1	0	1	0
Prague 19	0	1	0	1	1	0
Prague 20	0	1	0	0	1	0
Prague 21	0	2	0	2	2	0
Prague 22	0	0	0	0	0	0
Votes	31	24	37	34	18	29

Table 3: Number of Votes for Proposals by City District

District	Real value s^0_i	Proposal 3 s_i	Performance (formula (7))	Performance (formula (8))
Prague 1	2.86	3.07	107.4	28.5
Prague 2	3.71	4.14	111.6	51.2
Prague 3	4.52	4.88	107.8	81.8
Prague 4	10.69	11.06	103.5	83.0
Prague 5	6.36	6.42	101.0	51.3
Prague 6	10.16	10.10	99.4	17.0
Prague 7	2.46	2.81	114.2	81.0
Prague 8	8.49	8.72	102.7	87.9
Prague 9	4.19	3.62	86.4	29.8
Prague 10	7.58	7.45	98.2	65.8
Prague 11	7.06	7.01	99.3	53.2
Prague 12	5.98	5.57	93.1	28.4
Prague 13	6.35	6.61	104.0	59.9
Prague 14	4.31	4.23	98.2	47.6
Prague 15	2.73	2.77	101.4	71.4
Prague 16	1.35	1.19	88.3	6.5
Prague 17	1.98	2.07	104.7	79.8
Prague 18	1.66	1.73	104.5	43.4
Prague 19	1.01	0.89	88.4	21.5
Prague 20	3.08	2.69	87.2	8.6
Prague 21	1.78	1.48	83.5	4.8
Prague 22	1.67	1.48	88.2	12.6

Table 4: Impact of Proposal 3 on Subsidy Allocation

5 Conclusion

We have described the subsidy allocation that is used by the City of Prague in 2013. We have shown that the allocation problem can be seen as a coalition game with a dual structure (city districts and political parties). In the analysis, we do not consider the possibilities that some criterion is abolished (its weight is zero) or that some new criterion was introduced. All the calculations we have made have to be seen as illustrative examples. We are not able to offer any definitive answers about which subsidy allocation model is the best one; however an analysis of the allocation process can help us in understanding the nature of the allocation problem, its advantages and disadvantages (Dlouhý, 2013).

Acknowledgements

The research is supported by the Czech Science Foundation, project no. P403/12/1387 "Models for Efficiency Evaluation in Hierarchical Economic Systems".

References

[1] City of Prague: *Resolution of the City Council no. 1922 on proposal of budget of the City of Prague for year 2012, of subsidy relations to city districts of the City of Prague for year 2012, and budget outlook till 2017.* Prague: City of Prague, 2011.

[2] City of Prague: *Resolution of the City Council no. 1892 on proposal of budget of the City of Prague for year 2013, of subsidy relations to city districts of the City of Prague for year 2013, and budget outlook till 2018.* Prague: City of Prague, 2012.

[3] Dlouhý, M.: Efficiency and resource allocation within a hierarchical organization. In *Proceedings of 30th International Conference Mathematical Methods in Economics 2012*. Karviná: Silesian University, 2012.

[4] Dlouhý, M.: Models of Subsidy Allocation among City Districts. *Prague Economic Papers*, 2014, in print.

[5] Dlouhý, M., Fiala, P.: *Úvod do teorie her*. 2nd Edition. Prague: University of Economics in Prague, 2009.

[6] Osborne, M. J.: *An Introduction to Game Theory*. New York: Oxford University Press, 2004.

[7] Turnovec, F., Mercik, J. W., Mazurkiewicz, M.: National and ideological influence in the European Parliament. *Control and Cybernetics*, 2008, 37(3): 585-606.

RAROI or IRR?

Pavel Doležel[1], Zuzana Dlouhá[2]

Abstract. In this study we deal with estimating profitability of consumer finance loans on a single loan level. We compare two different theoretical approaches to measuring profitability: RAROI (Risk Adjusted Return on Investment) and IRR (Internal Rate of Return). We discuss the usage of these two approaches in portfolio management and we particularly focus on the conditions under which one of the measures prevails over the other in case of profit maximization with constrained or unconstrained capital.

Keywords: Return on Investment, Internal Rate of Return, Consumer finance, Profitability.

JEL classification: G23
AMS classification: 90A09

1 Introduction

Assume we are in the position of a financial institution providing loans to its customers. Our aim is to estimate profitability of a signle loan at the date it enters the approval proces in order to be able to decide whether it is worth of approval. Profitability stands for some measure of net profit that is made on a given exposure. In reality there are few ways how to measure this profit. For example we can take into account time value of money or not, it can be expressed as a percentage of the original amount, as an absolute value or as a profitability rate p.a. We will deal with two approaches:

- RAROI (Risk Adjusted Return on Investment), which stands for the sum of discounted cash flows of the investment divided by the original principal
- IRR (Internal Rate of Return), which stands for the interest rate p.a. that equals the rate we would have to have on a risk-free investment in order to gain the same profit in the same period. For further details on IRR see the articles [4] and [1].

We will assume 90 days default on an exposure level, i.e. application is said to be in default, when it is or ever was at least 90 days past due. This also means, that 90 days default is an absorbing state. As the date of default we will consider the date on which the exposure is 90 days past due the first time. Moreover we assume the German standard 30E/360 for computing number of days between distinct payments and interest being imposed monthly. As a discount rate we use the cost of equity e p.a. as it expresses the best alternative investment. We assume only annuity exposures.

To be formaly correct, we also assume the set of exposures[3] $\mathcal{K}$ and we denote the cardinality of this set K and introduce a bijective (one-to-one) mapping ρ from the set $\mathcal{K}$ to the set of integers $\tilde{K} = \{1, \ldots, K\}$. Each particular exposure κ from the set $\mathcal{K}$ we will refer to as to the integer $k := \rho(\kappa)$ form the set $\tilde{K}$.

We assume that the credit demand in a period Δ is given by the number n of clients applying for a loan during the period Δ. This number of applicants is exogeneous for the risk management. It can be at most boosted by marketing or sales actions, but not by risk department. We wish to compare RAROI with IRR under four different conditions. These conditions reflect four distinct states:

1. We wish to maximize short-term profit while we have a given set of customers, bounded capital and no possibility to reinvest the repaid money.

[1]Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Opletalova 26, 11000, Prague 1, paveldolezel@email.cz
[2]Faculty of Informatics and Statistics, University of Economics, Nám. W. Churchilla 4, 13067, Prague 3, figlova@vse.cz
[3]We will speak of exposures or loans. These two notions are freely substitutable throughout this text.

2. We wish to maximize short-term profit while we have a given set of customers, unbounded capital and no possibility to reinvest the repaid money.
3. We wish to maximize long-term profit while we have bounded capital and we can always reinvest the repaid money.
4. We wish to maximize long-term profit while we have unbounded capital and we can always reinvest the repaid money.

We wish to find out, which of the two proposed measures of profitability are better in which of these situations.

2 Cash flows

Assume the loan k is an annuity with term T_k months, where $T_k \in \mathbb{N}$. Assume in addition we have estimated the probabilities of default on each monthly payment for the loan k. We denote the probability of default on the m-th payment[4] $P_k(m)$. In addition we assume i_k is the interest rate p.a. imposed on the loan k. The original nominal value (original principal) of the loan k we will denote $X_k(0)$. There can be a fee paid monthly which is not part of the annuity. We denote this monthly fee of the loan k f_k.

2.1 Annuity payment

The present value of all the annuity payments discounted by the interest rate i_k (when the payments are performed monthly) must sum up to the nominal value of the loan, i.e.

$$S_k \left(\sum_{t=1}^{T_k} \left(\frac{1}{1+\frac{i_k}{12}} \right)^t \right) = X_k(0), \tag{1}$$

where S_k stands for the annuity payment of the loan k. From (1) we easily derive the equation for annuity payment:

$$S_k = \frac{X_k(0)\,\frac{i_k}{12}}{1-\left(\frac{1}{1+\frac{i_k}{12}}\right)^{T_k}}. \tag{2}$$

2.2 Outstanding balance

We will denote $X_k(t)$ the outstanding balance of the loan k after t-th payment was paid (when repaid on time). It is easy to derive the formula for the computation of $X_k(t)$ as the recursive equality 3 must hold for any $s \in \{0, \ldots, T_{k-1}\}$.

$$X_k(s)\left(1+\frac{i_k}{12}\right) - S_k = X_k(s+1) \tag{3}$$

The equation derived for the outstanding balance is thus

$$X_k(t) = X_k(0)\left(\frac{1-\left(1+\frac{i_k}{12}\right)^{t-T}}{1-\left(1+\frac{i_k}{12}\right)^{-T}}\right). \tag{4}$$

2.3 Cost of money

We assumee that we repay the borrowed loan as a revolving with the same interest rate ϕ p.a. This revolving is set so as to pay just the principal that is to be repaid by the client in each month according to the instalment schedule. Moreover at the date of 90 days default we repay the whole remaining

[4]Probability that the m-th ($m > 1$) payment will not be repaid before 90 days past due while the $(m-1)$-th payment will. For $m = 1$ it is just the probability that no payment will be repaid before 90 days past due.

principal and so the interest after default does not have to be paid. We repay the non-equity $(1 - EQ)$ part of the borrowed loan principal provided by a third party as a revolving payments with interest rate ϕ p.a. The t-th payment equals the t-th scheduled instalment of the loan k, except for the case of default. So the cost of money is the interest paid on this loan before the default occurs

$$NM_k^t = (1 - EQ)\left(1 - \sum_{j=1}^{t} P_k(j)\right)\left(\frac{\phi}{12} \cdot X_k(t-1)\right) + 3(1 - EQ)\, P_k(t)\left(\frac{\phi}{12} \cdot X_k(t-1)\right). \tag{5}$$

2.4 Loss

The expected loss EL_k^t from the period between the scheduled date of $(t-1)$-th and the scheduled date of the t-th payment[5] (including this date) can be computed for the exposure k when the loan is financed through revolving as

$$EL_k^t = P_k(t)\, X_k(t-1). \tag{6}$$

If it is financed through an annuity, the expected loss can be computed as

$$EL_k^t = P_k(t)\, X_k(t-1). \tag{7}$$

2.5 Interest revenue

The interest revenue RE_k from the period between the scheduled date of $(t-1)$-th and the scheduled date of the t-th payment[6] including this date can be computed for the exposure k simply as

$$RE_k^t = \left(1 - \sum_{j=1}^{t} P_k(j)\right) X_k(t-1)\, \frac{i_k}{12}. \tag{8}$$

2.6 Recoveries

For the loan k a recovery rate $r_k(t)$ for a loan defaulted on t-th payment is estimated based on the loan's characteristics. The recovery RC_k^t from exposure defaulted on the t-th payment can be computed for the exposure k simply as

$$RC_k^t = P_k(t)\, \frac{X_k(t-1)\, r_k(t)}{\left(1 + \frac{e}{12}\right)^3}. \tag{9}$$

The recovery rate $r_k(t)$ stands for all the recovered payments after the date of default, no matter whether they are allocated as a principal payment, penalty payment, interest payment or late fee payment. The estimation is always made using the payment number on which the default occured, but it can generaly be estimated using other loan or client characteristics.

2.7 Monthly fees

The loan fees are to be paid monthly together with the annuity payment. The total expected fee paid on exposure k between the scheduled $t-1$-th and t-th payment we denote FE_k^t and it can be computed simply as

$$FE_k^t = \left(1 - \sum_{j=1}^{t} P_k(j)\right) f_k. \tag{10}$$

2.8 Commissions

Let's denote c_k the commissions paid to dealers for the acquisition of the loan k.

[5] For $t = 1$ it is the period before the scheduled date of first payment.
[6] For $t = 1$ it is the period before the scheduled date of first payment.

2.9 Variable operational cost

The variable operational cost should be estimated based on financial data. It can vary accross product types, clients' risk profiles and other characteristics. The variable operational costs (OPEXes) can be generaly divided in at least three different types of costs: origination, servicing and termination. We will not account for termination costs, as the regular termination is more or less costless and the non-regular termination we deal separately as a cost of collections. To be rigorous, we should also include the cost of pre-termination, which is usualy not negligible, but for this, we would have to also estimate the probabilities of pre-termination, which would make the computation too complicated and the value added would be very small. So we denote $v1_k$ the variable operational cost of origination and $v2_k^t$ the monthly variable operational cost of servicing for the month between the scheduled date of $(t-1)$-th and the scheduled date of the t-th payment[7] including this date. The origination cost should depend on the rejection rate, as there are variable costs rather on each application, than just on each contract. The higher is rejection rate, the higher should be the variable costs of origination on the given loan.

2.10 Cost of equity

The cost of equity stands for the interest payments on shareholders' funds which is used in order to provide the given loan. We denote e the cost of equity expressed as a percentage p.a. The sum of equity costs over the life of the loan can be computed as

$$EQ_k^t = EQ\left(1 - \sum_{j=1}^{t} P_k\left(j\right)\right) X_k\left(t-1\right)\frac{e}{12}, \tag{11}$$

where EQ stands for the equity to productive assets ratio. It can be considered to be a constant.

2.11 Costs of collections

The costs of collections should be estimated based on financial data. It can vary accross product types and clients' risk profiles. We denote g_k the costs of collections. The expected total cost of collections is thus

$$CC_k^t = P_k\left(t\right)\frac{g_k}{\left(1+\frac{e}{12}\right)^3} \tag{12}$$

2.12 Insurance

Assume, there is a life insurance and an unemployment insurance offered to the customer. The insurance fee is treated as part of the loan, i.e. it highers the principal. Denote the fee for life insurance considering the exposure k IL_k and the unemployment insurance IU_k. Assume, that the commissions from the insurance are c_{IL} for the life insurance and c_{IU} for the unemployment insurance. When we denote CA the original credit amount that the client would apply for if there was no insurance, then we get $X_k\left(0\right) = CA + c_{IL}IL_k + c_{IU}IU_k$.

2.13 Gross margin

The expected gross margin of loan k from period between the scheduled date of $(t-1)$-th and the scheduled date of the t-th payment[8] including this date we denote GM_k^t and it can be computed as

$$GM_k^t = RE_k^t + RC_k^t + FE_k^t - EL_k^t. \tag{13}$$

[7] For $t = 1$ it is the month before the scheduled date of first payment.
[8] For $t = 1$ it is the period before the scheduled date of first payment.

3 RAROI estimation

RAROI stands for the sum of discounted cash inflows minus the sum of discounted cash outflows divided by the original investment (i.e. the original principal borrowed). Therefore it can be expresed as a percentage of the original principal.

The risk adjusted return on investment $RAROI_k$ for loan k can be computed as

$$RAROI_k = \frac{\sum_{t=1}^{T_k} \frac{GM_k^t - EQ_k^t - NM_k^t - CC_k^t - v2_k^t}{\left(1+\frac{e}{12}\right)^t} + c_{IL} IL_k + c_{IU} IU_k - v1_k - c_k}{X_k(0)} - 1. \tag{14}$$

It is important to note, that this RAROI already contains the cost of equity and so whenever it is positive the loan brings higher interest rate on equity than e p.a. When $RAROI < 0$, it can still be profitable for shareholder, but it is not profitable for the company assuming the cost of capital from shareholder is e p.a.

RAROI can also be considered as profitability index, which stands for profit divided by the original investment.

4 IRR estimation

Estimating IRR requires the same inputs as estimation of RAROI, but the concept of the final profitability measure is quite different. Instead of dividing the cash flows by the original principal, we compute the interest rate which corresponds to the expected cash flows from the investments. Then IRR is the solution for r to the equation $\Pi(r) = 0$, where $\Pi(r)$ is given by

$$\Pi(r) = \sum_{t=1}^{T_k} \frac{GM_k^t - EQ_k^t - NM_k^t - CC_k^t - v2_k^t}{\left(1+\frac{r}{12}\right)^t} - X_k(0) + c_{IL} IL_k + c_{IU} IU_k - v1_k - c_k \tag{15}$$

Finding a solution to the equation $\Pi(r) = 0$ is based on iterative numerical approaches as there is generally no analytical solution to express a root of high degree polynoma. Newton method can be used as well as the method of half intervals. In our case there is just one positive real solution, see [2].

IRR assumes each repaid money can immediately be reinvested for the same rate of return. If this is not true, modified IRR (MIRR) is proposed, see for example [3]. MIRR assumes that positive cash-flows can be reinvested only at the rate of cost of equity e. We can then rewrite the equation for IRR as follows:

$$MIRR = \sqrt[T]{\frac{\sum_{t=1}^{T_k} PCF_k(t)\left(1+\frac{e}{12}\right)^{T-t}}{\sum_{t=1}^{T_k} NCF_k(t)\left(1+\frac{e}{12}\right)^{-t}}} - 1. \tag{16}$$

5 Profitability maximization

Now we show, in which of the four situations described above prevails RAROI and in which prevails IRR or MIRR. Scoring via RAROI is based on rejecting applications with RAROI below some cut-off and scoring via IRR on rejecting applications with IRR below some cut-off.

It can be shown that

$$IRR_k(T_k, RAROI_k) = 12\left(\sqrt[T_k]{RAROI_k\left(1+\frac{e}{12}\right)^{T_k} + 1}\right) - 1. \tag{17}$$

This can be rewritten for $RAROI_k$ as

$$RAROI_k(T_k, IRR_k) = \frac{\left(1+\frac{IRR_k}{12}\right)^{T_k} - 1}{\left(1+\frac{e}{12}\right)^{T_k}}. \tag{18}$$

From (18) we know, that $RAROI_k$ is increasing in the term T_k while IRR_k is fixed and it is also increasing in IRR_k while term T_k is fixed. We can conclude, that IRR is an interest rate p.a. and so it does not directly depend on the term, while RAROI depends on term and is higher for higher terms.

If we are in the first situation of short-term profit maximization and bounded capital it is optimal strategy to score the applicants via RAROI. We do not care about the terms of the provided loans as we have no chance to reinvest the repaid money. Moreover because the capital is bounded, we will approve only $\tilde{k}$ applications with the highest RAROI such that $\forall k > \tilde{k} \sum_{i=1}^{k} X_i(0) > K$, and $\sum_{i=1}^{\tilde{k}} X_i(0) \leq K$, where K is the available capital and X_i is sorted descendingly by $RAROI_i$. The only difference from the second situation is in the cut-off. If we maximize short-term profit, have no chance to reinvest the repaid money, but the capital is unbounded, we will approve all the loans with positive RAROI.

The third situation stands for long-term profit maximization with bounded capital, where we always reinvest the repaid money by providing another loan. In this case, we need to take into account terms, as long term with a given RAROI must be less profitable than short term with the same RAROI, as we make the same money faster and so can faster reinvest them. This is the reason why IRR prevails over RAROI. IRR enables to compare profitability of two loans with different terms, while RAROI does not. Because we have a bounded capital, we set the cut-off such that we approve only $\hat{k}$ applications with the highest IRR such that $\forall k > \hat{k} \sum_{i=1}^{k} X_i(0) > K$, and $\sum_{i=1}^{\hat{k}} X_i(0) \leq K$, where K is the available capital and X_i is sorted descendingly by IRR_i. Again the difference from the fourth situation is just in the cut-off. If we maximize long-term profit, can reinvest the repaid money freely by providing another loan, but the capital is unbounded, we will approve all the loans with positive IRR.

6 Discussion and conclusion

We have provided two distinct ways of measuring profitability of a single loan, including the cost of capital (cost of money representing the debt (emitted bonds, loans provided by third parties, etc.) and cost of equity representing the cost of shareholders' capital (equity)). The first measure, Risk Adjusted Return On Investment was shown to be a measure better reflecting short-term profit maximization in an environment without possibility of reinvesting in the same asset. The second measure, Internal Rate of Return was shown to be a measure better reflecting long-term profit maximization in an environment with possibility of reinvesting in the same asset. The provided results would deserve exact derivation, but for this, there is not enough space in this short text. We have also shown a modified version of Internal Rate of Return, which could be used in the environemnt where it is possible to reinvest the repaid money in the same asset, but it does not necessarily have to be without a waiting period during which the repaid money can not be reinvested in similar loans. This situation would account for relatively low demand for loans.

References

[1] Dudley Jr. C. L., 1972. A note on reinvestment in choosing between net present value and internal rate of return. The Journal of Finance 27 (4), 907–915.

[2] Flemming J. S. and Wright J. F., 1971. Uniqueness of the Internal Rate of Return: A Generalisation. The Economic Journal 81 (322), 256–263.

[3] Kalouda F., 2011. Finanční řízení podniku. Vydavatelství a nakladatelství Aleš Čeněk.

[4] Turvey R., 1963. Present Value Versus Internal Rate of Return-An Essay in the Theory of the Third Best. The Economic Journal 73 (289), 93–98.

On finite single-server queue subject to non-preemptive breakdowns

Michal Dorda[1], Dušan Teichmann[2]

Abstract. The paper deals with modelling and simulation of a finite single-server queueing system with a server subject to breakdowns. We consider that customers come to the queueing system in the Poisson stream. Customers incoming to the system are served according to the FCFS discipline, service times are considered to follow the Erlang distribution defined by the shape parameter and the scale parameter. Customers can wait for the service in the queue which length is limited by (m-1) places, that means the total capacity of the queueing system is equal to m places. Further, we assume that the server can break down. Breakdowns of the server are considered to be non-preemptive that means when a breakdown occurs during customer servicing it is possible to finish it before server repair is started. Times between breakdowns and repair times are assumed to follow the exponential distribution. We model the queue as a quasi-birth death process for which we present steady-state diagram and equation system describing the system behaviour in the steady-state. Solving the equation system in Matlab we get stationary probabilities which are used for computing basic performance measures. The mathematical model is supported by a simulation model in order to validate the outcomes of the mathematical model.

Keywords: $M/E_n/1/m$, queue, method of stages, non-preemptive failures

JEL Classification: C44
AMS Classification: 60K25

1 Introduction

Queuing theory enables us to study the behaviour of such called queueing systems. In general, each queueing system represents a system which serves customers coming in the system. Such types of systems can be met in many sectors, for example in informatics, telecommunications, transport and, of course, economics.

One of the possible ways how queueing system can be sorted it is sorting according to the fact whether the model of the system has incorporated server failures. Failures of the server have an obvious impact on the performance measures of the studied queueing system. Therefore, in many cases it is necessary to incorporate them into the model in order to get relevant outcomes.

Queueing systems subject to server failures, in which inter-arrival or service times are considered to be Erlang distributed, have already been studied in the past. But in comparison with queues assuming exponentially or general distributed inter-arrival and service times the models of queueing systems under assumption of the Erlang distribution are not so common.

Some authors presented queueing models subject to failures under the assumption of the Erlang distribution which was most often applied to model service times. An $M/E_k/1$ queue with server vacation was studied by Jain and Agrawal [2]. The authors assume that the server takes vacations when the system is empty and when it is busy than can break down. It is further considered that the Poisson arrival rate is state dependent. An $M/E_k/1$ queueing system under similar assumptions but with so called multi optional repair was presented in paper [3]. Wang and Kuo [5] solved a finite source $M/E_k/1$- a group of identical machines which is operated under the care of an unreliable service station (so called machine repair problem). The authors employed matrix geometric method to derive the steady-state probabilities and developed the steady-state profit function to find out the optimum number of machines which are repaired by the server. In paper [4] an $M^X/E_k/1$ two-phase queueing system with a single removable server and with gating, server start-up and unpredictable breakdowns is considered.

[1] VŠB – Technical University of Ostrava / Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, Ostrava – Poruba, phone: +420 597 325 754, e-mail: michal.dorda@vsb.cz.
[2] VŠB – Technical University of Ostrava / Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, Ostrava – Poruba, phone: +420 597 324 575, e-mail: dusan.teichmann@vsb.cz.

In paper [1] a finite queueing system with Erlang distributed inter-arrival and service times and with the service station subject to disastrous failures was studied.

The paper is devoted to a modification of a finite single-server queueing system with a server subject to breakdowns which are considered to be non-preemptive. The paper is organized as follows. In Section 2 we will make all necessary assumptions and the mathematical model is presented. In Section 3 we will introduce a simulation model which was created to validate the mathematical model. In Section 4 outcomes of executed experiments are shown.

2 Mathematical model

In this section we will present a mathematical model of a finite queueing system with a server subject to non-preemptive breakdowns. The capacity of the queueing system is equal to m, where $m \geq 2$ - a place in the service and (m-1) places in the queue. Let the arrival process of customers be the Poisson process with the parameter λ. For the Poisson arrival process it holds that customer inter-arrival times are exponentially distributed with the same parameter λ. Costumer service times are supposed to be an Erlang random variable with the shape parameter $n \geq 2$ and the scale parameter $n\mu$. The Erlang distribution with the parameters n and $n\mu$ arises from n mutually independent exponential distribution with the same parameter equal to $n\mu$; thus the mean service time is equal to $\frac{n}{n\mu} = \frac{1}{\mu}$. This property of the Erlang distribution is exploited by method of stages which models Erlang distributed service times as n exponentially distributed service times. Each customer successively goes through n individual phases of service and leaves the system only when the last phase of service has been finished.

Let us assume that server failures also occur according to the Poisson input process with the parameter η. If there already is a failure in the system, then the parameter η is equal to zero. Times to repair are exponentially distributed with the parameter ζ. In the case that a failure arises when a customer is being served, the service of customer is not interrupted (the service continues without any influence of the failure on the server) and after finishing it we begin to repair the server - we can say that server failures are nonpreemptive. Such type of failures can be also interpreted as a special case of server vacations, where the server takes vacation randomly with no dependency on the number of customers finding in the system.

Let us have three discrete random variables denoted as K, P and F. The variable K describes the number of customers finding in the system, where the variable can take the values from the set {0,1,...,m}. The variable P expresses the phase of service which has already been finished, where the variable can take the values from the set {0,1,...,n-1}. The last variable F can take 3 values from the set {0,1,2}, where the meaning of it is as follows:

- If the variable F is equal to 0, there is no failure in the system.
- If the variable F is equal to 1, the failure of the server is waiting for repair because we have to finish servicing of the customer being in the service.
- If the variable F is equal to 2, the failure of the server is being repaired.

It is obvious that the individual states of the system can be described by triplets (k,p,f). The state space of the system is the union of three states subsets:

$$\Omega = \Omega_1 \cup \Omega_2 \cup \Omega_3,$$

where:

- The states in the subset $\Omega_1 = \{(k, p, f): k \in \{1,\ldots,m\}, p \in \{0,\ldots,n-1\}, f = 0\}$ are the states in which there are k customers in the system and the p phases of customer service have already been finished and there is no failure in the system. To these states we have to add the state (0,0,0) which corresponds to the state in which the system is idle (and empty).
- The states in the subset $\Omega_2 = \{(k, p, f), k \in \{1,\ldots,m\}, p \in \{0,\ldots,n-1\}, f = 1\}$ are the states in which there are k customers in the system and the p phases of customer service have already been finished and there is a failure of the server in the system waiting for repair.
- The states in the subset $\Omega_3 = \{(k, p, f), k \in \{0,\ldots,m-1\}, p = 0, f = 2\}$ are the states in which there are k waiting customers in the system and the server is being repaired.

To simplify the formation of equations describing the system state probabilities we created a state transition diagram which depicts transitions between individual system states – see Figure 1. Please notice that vertices represent the states of the system and oriented edges indicate the possible transitions with the corresponding rate.

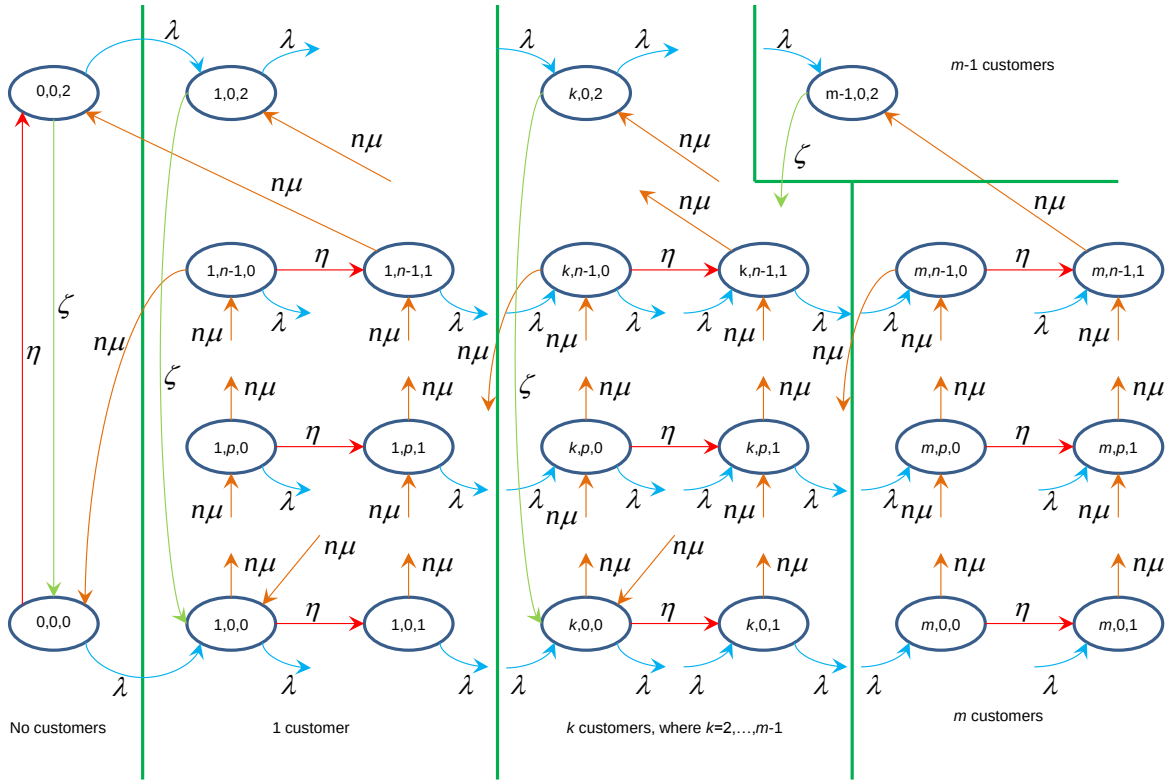


Figure 1 The state transition diagram of the system.

No we are ready to write equations for individual state probabilities. Applying generally known global balance principle we get the following linear equation system which describes the behaviour of the process in steady state:

$$(\lambda+\eta)\cdot P_{(0,0,0)} = n\mu\cdot P_{(1,n-1,0)} + \zeta\cdot P_{(0,0,2)}, \tag{1}$$

$$(\lambda+n\mu+\eta)\cdot P_{(k,0,0)} = \lambda\cdot P_{(k-1,0,0)} + n\mu\cdot P_{(k+1,n-1,0)} + \zeta\cdot P_{(k,0,2)} \text{ for } k=1,\ldots,m-1, \tag{2}$$

$$(n\mu+\eta)\cdot P_{(m,0,0)} = \lambda\cdot P_{(m-1,0,0)}, \tag{3}$$

$$(\lambda+n\mu)\cdot P_{(1,0,1)} = \eta\cdot P_{(1,0,0)}, \tag{4}$$

$$(\lambda+n\mu)\cdot P_{(k,0,1)} = \eta\cdot P_{(k,0,0)} + \lambda\cdot P_{(k-1,0,1)} \text{ for } k=2,\ldots,m-1, \tag{5}$$

$$n\mu\cdot P_{(m,0,1)} = \eta\cdot P_{(m,0,0)} + \lambda\cdot P_{(m-1,0,1)}, \tag{6}$$

$$(\lambda+n\mu+\eta)\cdot P_{(1,p,0)} = n\mu\cdot P_{(1,p-1,0)} \text{ for } p=1,\ldots,n-1, \tag{7}$$

$$(\lambda+n\mu+\eta)\cdot P_{(k,p,0)} = \lambda\cdot P_{(k-1,p,0)} + n\mu\cdot P_{(k,p-1,0)} \text{ for } k=2,\ldots,m-1 \text{ and } p=1,\ldots,n-1, \tag{8}$$

$$(n\mu+\eta)\cdot P_{(m,p,0)} = \lambda\cdot P_{(m-1,p,0)} + n\mu\cdot P_{(m,p-1,0)} \text{ for } p=1,\ldots,n-1, \tag{9}$$

$$(\lambda+n\mu)\cdot P_{(1,p,1)} = \eta\cdot P_{(1,p,0)} + n\mu\cdot P_{(1,p-1,1)} \text{ for } p=1,\ldots,n-1, \tag{10}$$

$$(\lambda+n\mu)\cdot P_{(k,p,1)} = \eta\cdot P_{(k,p,0)} + \lambda\cdot P_{(k-1,p,1)} + n\mu\cdot P_{(k,p-1,1)} \text{ for } k=2,\ldots,m-1 \text{ and } p=1,\ldots,n-1, \tag{11}$$

$$n\mu\cdot P_{(m,p,1)} = \eta\cdot P_{(m,p,0)} + \lambda\cdot P_{(m-1,p,1)} + n\mu\cdot P_{(m,p-1,1)} \text{ for } p=1,\ldots,n-1, \tag{12}$$

$$(\lambda+\zeta)\cdot P_{(0,0,2)} = \eta\cdot P_{(0,0,0)} + n\mu\cdot P_{(1,n-1,1)}, \tag{13}$$

$$(\lambda+\zeta)\cdot P_{(k,0,2)} = n\mu\cdot P_{(k+1,n-1,1)} + \lambda\cdot P_{(k-1,0,2)} \text{ for } k=1,\ldots,m-2, \tag{14}$$

$$\zeta\cdot P_{(m-1,0,2)} = n\mu\cdot P_{(m,n-1,1)} + \lambda\cdot P_{(m-2,0,2)} \tag{15}$$

Because an equation, for example equation (1), is linear combination of all others, we can omit it. To solve the equation system it is necessary to replace the omitted equation by normalization equation (16):

$$P_{(0,0,0)} + \sum_{k=1}^{m}\sum_{p=0}^{n-1}\sum_{f=0}^{1} P_{(k,p,f)} + \sum_{k=0}^{m-1} P_{k,0,2} = 1. \tag{16}$$

After omitting equation (1) we have the equation system of $2mn+m+1$ linear equations formed by equations (2) up to (16) with $2mn+m+1$ unknown stationary probabilities. To solve it using Matlab it is necessary to establish an alternative states description because applied state description in the form of (*k*,*p*,*f*) is three-dimensional and is very good for formation the equation system but is absolutely unsuitable for computations in

Matlab which uses matrices. Therefore we applied an alternative one-dimensional state description in the following form:

- The states (k,p,f) for k=1,...,m, p=0,...,n-1, and f=0,1 can be denoted using a single value $f \cdot m \cdot n + p \cdot m + k$,
- The states (k,0,2) for k=0,...m-1 can be denoted using a single value $2 \cdot m \cdot n + k + 1$.
- The state (0,0,0) is labelled as $2 \cdot m \cdot n + m + 1$.

Applying the alternative one-dimensional state description we are able to transform the equation system in the form we need for using Matlab. After numerical solving of the equation system rewritten in matrix form we obtain the stationary probabilities we need in order to compute performance measures of the studied system. Let us establish three random variables - the number of the customers in the service S, where $S \in \{0,1\}$, the number of customers waiting in the queue L, where $L \in \{0,1,\ldots,m-1\}$ and the number of broken servers F, where $F \in \{0,1\}$. In the case we have computed steady-state probabilities; we are able to compute the following performance measures.

The mean number of the customers in the service *ES* is equal to:

$$ES = \sum_{k=1}^{m} \sum_{p=0}^{n-1} \sum_{f=0}^{1} P_{k,p,f} \,, \tag{17}$$

the mean number of the customers waiting in the queue *EL* is given by expression:

$$EL = \sum_{k=2}^{m} (k-1) \cdot \sum_{p=0}^{n-1} \sum_{f=0}^{1} P_{k,p,f} + \sum_{k=1}^{m-1} k \cdot \sum_{p=0}^{n-1} P_{k,p,2} \tag{18}$$

and for the mean number of the broken servers *EF* it holds

$$EF = \sum_{k=0}^{m-1} P_{k,0,2} \,. \tag{19}$$

Another important performance measures of finite queueing system is the probability that customer will not be served. In our queueing system it means that customer is rejected upon its arrival, it happens when the system is full. Therefore for the probability of rejection P_{rej} we can write:

$$P_{rej} = \sum_{p=0}^{n-1} \sum_{f=0}^{1} P_{m,p,f} + P_{m-1,0,2} \,. \tag{20}$$

3 Simulation model

To validate the proposed mathematical model we created a simulation model of the studied queueing system. The model is based on coloured Petri net, the software CPN Tools in version 3.0.4 was used to create the Petri net model. Figure 2 depicts the created Petri net in the initial marking with the following values of the model parameters: m=5, λ=10 h^{-1}, n=10, $n\mu$=100 h^{-1}, η=0.1 h^{-1}and ζ=0.2 h^{-1}.

To generate concrete values of the random variables used in model we defined a function "*fun ET(k, mi) = round(erlang(k, mi/3600.0))*;" where *k* is the shape parameter and *mi* is the scale parameter of the Erlang probability distribution expressed in [h^{-1}]. Please notice that the values which come from the exponential distribution can be also generated by this function because the exponential distribution is a special case of the Erlang distribution with the shape parameter equal to 1. We apply a second as the basic unit of time for the simulation experiments.

The simulation model consists of 10 places and 9 transitions. The model works with following tokens:

- Tokens "*C*" represent customers; these tokens are defined to be timed in order to model exponentially distributed inter-arrival times and Erlang distributed service times.
- Tokens "*F*" model failures of the server and are declared to be timed as well.
- Auxiliary tokens "*P*" model sources of the queueing system (for example free places in the queue or free places in the service).

To obtain desired simulation outcomes some monitoring functions were defined:

- The monitoring function named "*ES"*, which is bound with the place "*Servicing*", enables estimation of the mean number of the customers in the service.
- The monitoring function named "*EL"*, which is bound with the place "*Queue"*, serves for estimation of the mean number of the waiting customers.

- The monitoring function named "*EF*", which is bound with the place "*Repairing*", was create in order to estimate the mean number of the broken servers.
- To estimate the probability that the customer is rejected upon its arrival we were obliged to define two monitoring functions bound with transitions - the monitoring function named "*Number of Rejected Customers*" is bound with the transition "*Rejection*" and the function "*Number of Arriving Customers*" with the transition "*Customer initialization*". These two functions count how many times the corresponding transition was fired during each simulation run. The probability P_{rej} can be then estimated as ratio of these values - the value of the function "*Number of Rejected Customers*" divided by the value of the function "*Number of Arriving Customers*".

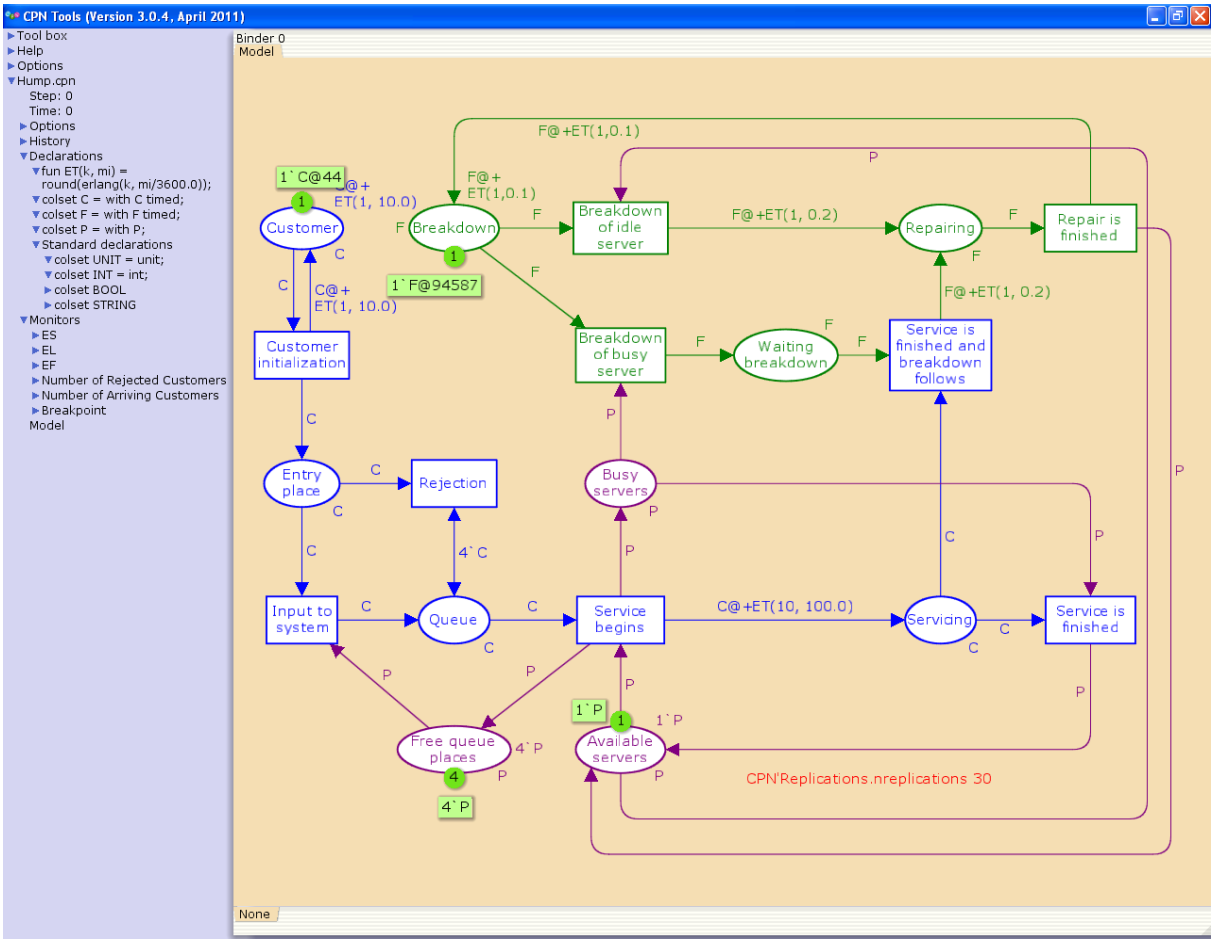


Figure 2 The simulation model created using CPN Tools - Version 3.0.4.

To enable stopping each simulation run after reaching predefined value of simulation time a breakpoint function was defined. The function stops each simulation run after reaching the simulation time equal to 31 536 000 s; that is 365 days. Because the simulation model is stochastic, it is necessary to execute several simulation runs with the model to compute confidence intervals for monitored performance measures. To ensure it we added an auxiliary text "*CPN'Replications.nreplications 30*" into the model. Evaluating the text it is automatically performed 30 simulation runs and simulation outcomes are gained.

4 Results of performed experiments

We executed several experiments with both models to compare analytical and simulation outcomes. Applied values of model parameters are summarized in table 1.

Parameter	m [-]	λ [h^{-1}]	n [-]	$n\mu$ [h^{-1}]	η [h^{-1}]	ζ [h^{-1}]
Applied value	5 up to 10	10	10	100	0.1	0.2

Table 1 Applied values of model parameters.

In table 2 we can see the comparison of analytical and simulation outcomes. Columns corresponding individual performance measures are divided into two sub-columns. Analytical outcomes were obtained using Matlab and the values of observed performance measures were computed according formulas (17) up to (20). The analytical outcomes for individual performance measures are listed in the left sub-columns. On the basis of executed simulation runs (30 runs for each value of the parameter m) we computed 95% confidence intervals for *ES*, *EL*, *EF* and P_{rej}. The bounds of the computed confidence intervals are listed in right sub-columns; the lower bound of the interval is up and the upper bound of the interval is down. We can see that for all executed experiments the analytical value of individual performance measures lies in the corresponding confidence interval. We can see that both models give very similar outcomes, we validated the mathematical model.

m	*ES*		*EL*		*EF*		P_{rej}	
5	0.599	0.596 0.601	2.460	2.454 2.473	0.331	0.330 0.337	0.401	0.399 0.405
6	0.613	0.603 0.614	3.148	3.140 3.170	0.331	0.330 0.342	0.387	0.385 0.396
7	0.622	0.620 0.627	3.851	3.827 3.860	0.331	0.327 0.334	0.378	0.373 0.380
8	0.630	0.623 0.632	4.569	4.553 4.597	0.331	0.330 0.339	0.370	0.368 0.376
9	0.636	0.632 0.641	5.304	5.281 5.335	0.332	0.326 0.336	0.364	0.359 0.368
10	0.640	0.635 0.644	6.055	6.028 6.086	0.332	0.328 0.338	0.360	0.355 0.364

Table 2 Summary of results.

5 Conclusions

In the paper we presented the mathematical and the simulation model of $M/E_n/1/m$ queueing system subject to non-preemptive service station failures. The mathematical model represented by linear equation system can be solved using suitable software, for example Matlab. The mathematical model is supported by the simulation model based on coloured Petri net which was created using software CPN Tools. Both models enable to get the values of performance measures of the studied queueing system.

As regard our future research we would like to extend the model; there are another performance measures we are interested in. The models will be applied to model shunting processes at marshalling yards. Finally we would like to try to derive closed-form formulas for computing steady-state probabilities if it is possible using known methods.

References

[1] Dorda, M. and Teichmann, D.: About a modification of $E_r/E_s/1/m$ queueing system subject to breakdowns. In: *Proceedings of the 30th International Conference Mathematical Methods in Economics 2012, Part I.* (Ramík, J. and Stavárek, D. eds.). Silesian University in Opava, School of Business Administration in Karviná, Karviná, 2012, 117-122.

[2] Jain, M. and Agrawal, P. K.: $M/E_k/1$ queueing system with working vacation. *Quality Technology & Quantitative Management*, **4** (4) (2007), 455-470.

[3] Jain, M., Sharma, G. C. and Sharma, R.: Working vacation queue with service interruption and multi optional repair. *International Journal of Information and Management Sciences*, **22** (2011), 157-175.

[4] Vasanta Kumar, V., Hari Prasad, B. V. S. N. and Rao, K. P. R.: Optimal strategy analysis of an N-policy two-phase $M^x/E_k/1$ gated queueing system with server startup and breakdowns. *International Journal of Mathematical Archive*, **3** (8) (2012), 3016-3027.

[5] Wang, K.-H. and Kuo, M.-Y.: Profit analysis of the $M/E_k/1$ machine repair problem with a non-reliable service station. *Computers and Industrial Engineering*, **32** (3) (1997), 587-594.

Modelling financial flows of development projects subsidized from European funds

Stanislava Dvořáková[1], Petr Jiříček[2]

Abstract. The paper is going to deal with modelling the course of financial flows of projects which are subsidized to a various degree from European funds within the European Union Cohesion Policy. The aim of the paper is to illustrate the influence of subsidies from public financial sources on economic efficiency of investment and development projects. The behaviour of the model is demonstrated by the investment curve which is a summarizing expression of functional dependency of net present value of the project's cash-flows on the project's discount rate, represented by a polynomial function. The degree of the polynomial corresponds to the usual six year period of investment project implementation, the constant term of the polynomial will reflect the share of the subsidy of investment costs from the European funds in the project budget and the coefficients of the polynomial will correspond to the polarity and amount of cash-flows issuing from the project implementation. The model should solve the amount of subsidy from the European funds on the internal return rate of the project, represented in the model by a non-negative root of the polynomial. The sensitivity analysis of the drawn up model will examine the stability of the optimal solution of return on investment in relation to the changes of the anticipated amount of projected cash-flows. The outputs of the model programmed in MAPLE software will be analysed by graphs and complemented with explanatory economic commentary.

Keywords: financial flows, investment curve, polynomial function, sensitivity analysis, conventional projects, economical net present value, social discount rate.

JEL Classification: C20, H43
AMS Classification: 65H04, 68R10

1 Introduction

The scope and importance of the public sector in both Czech and European economy is growing and the effect of replacing private capital with capital originating from public recourses occurs. One of such public resources is also European funds which finance cohesion policy of the European Union, focused on reducing differences among both regions and member countries. Increasing the effectiveness of the cohesion policy depends on efficient financing of development projects of the public sector. For general evaluation of criteria of allocation of investment sources in the public sector we use criteria among which the so called Pareto improvement is the best known, saying that allocation of finance into a development project is optimal when it is impossible to find another allocation of financial sources which would lead to increasing the gain of the project beneficiary without making another stakeholder worse off [8]. The problem of evaluating efficiency of investment in the public sector is the measurement of this very gain, unlike the commercial sphere where the goal of investment projects is maximizing the profit or the market value of the company's equity. For measuring efficiency in the public sector, we use various methods [1], for example-cost minimum analysis (CMA), cost-effectiveness analysis (CEA), cost-utility analysis (CUA). These methods are by their output focused only on minimizing the costs, losses and damage in monetary or natural conception. For complex measurement of the difference between benefits and losses of the project in the resulting quantitative expression the generally accepted method used also for evaluating projects from European funds is the cost-benefit analysis. The aim of the paper is to present the possibilities of modelling financial flows for use of CBA in case of public development projects for determined levels of subsidy from European funds.

[1] College of Polytechnics Jihlava, Department of Mathematics, Tolstého 16, 586 01 Jihlava, Czech Republic, stanislava.dvorakova@vspj.cz.
[2] College of Polytechnics Jihlava, Department of Economic Studies, Tolstého 16, 586 01 Jihlava, Czech Republic, jiricek@vspj.cz.

2 Material and methods

In the paper, we will deal with modelling financial flows of public development projects where we will investigate the sensitivity of change of the height of allocation of the subsidy from European sources to the benefit from the project measured by the criterion of economic net present value, used in CBA according to the EU guide [9]. The behaviour of the model is illustrated by the investment curve which is a summary expression of functional dependence of net present value of the financial flows of the project on the social discount rate of the project, represented by a polynomial function [4]. For modelling financial flows, a model created by MAPLE software was applied. The model was constructed for the parameters of the investment curve of a public development project. The degree of the polynomial corresponds with the usual six year period of development projects implementation. The criteria function for evaluating public project efficiency has the form:

$$ENPV = \sum_{t=0}^{n} \frac{CF_t}{(1+k)^t}, \tag{1}$$

where CF_t is the cash flow of the public development project (i.e. the benefits) minus costs, losses of the project,

k is the common social discount rate (the required return on public investment from the perspective of the investor or donor of the project)

n is the period of economic lifespan of the project (in our case $n = 6$),

$ENPV$ is economic net present value of the public project reflecting the benefit from the project from a wider social economic view of the given investment.

The constant term of the polynomial reflects the share of investment costs subsidy from European funds in the project budget. In the Czech Republic it usually covers 100% of the project budget in non-profit making organizations, 70% in municipalities, and 30% in common projects of commercial companies and the public sector, applying for funding from European funds.

Coefficients of the polynomial correspond with the polarity and height of cash flows (CF_t) issuing from operational and liquidation stages of the project. In this phase, financial flows can generally have both negative and positive values; the positive values, however, should prevail. In the model we consider only the so called conventional projects, containing only one change of financial flow polarity [5], usually corresponding with the change from negative financial flow (from investment costs) to predicted positive financial flows from the operation of the investment.

Thus the created model shows individual stages of the investment process – the investment stage, the operation stage, and the liquidation stage. The pre-investment stage is not included in the project efficiency evaluation due to the so called sunk costs [2]. The sensitivity analysis of the model [7] should solve the influence of the height of subsidy from European funds on the $ENPV$ (using a 3D projection) and its relationship to marginal social return rate of a public project (defined as Economic Rate on Return – ERR, [9]), represented in the model by a non-negative root of the polynomial.

3 Results and discussion

3.1 The course of the $ENPV$ investment curve

The method of internal interest rate is based on the philosophy that we look for the marginal social rate of return ERR (Economic Rate of Return) from the given public investment on condition of return of the investment during the projected life span. It means that we look for the root of the $ENPV$ function (1), i.e.

$$ENPV = \sum_{t=0}^{n} \frac{CF_t}{(1+ERR)^t} = 0 .$$

When calculating the number ERR it is more suitable to convert the $ENPV$ function by means of substitution $x = \frac{1}{1+k}$ to the polynomial

$$f(x) = CF_0 + CF_1 x + \cdots + CF_n x^n .$$

Using some program (e.g. Maple or Matlab) we calculate the roots of this polynomial and by backward substitution $k = \frac{1}{x} - 1$ we convert the real roots back to the *k* variable. We do not consider imaginary roots because they do not have real interpretation in the economic practice. This procedure is more advantageous since there are more algorithms and more elaborate ones for searching for a polynomial root than algorithms for finding a general function root (see e.g. [3] and [6]). Let's also remark that it is necessary to use a calculating program because, as is well known, there are no formulas for calculating polynomial roots of a higher than the third degree.

With the help of the searched number ERR the public project efficiency can be judged according to a certain preference perspective of the investor or donor. However, this is not the aim of this paper. This issue is elaborated in more detail e.g. in [5] or [4]. Let's only say that for determining project efficiency by the method of internal rate of return there must be only one polynomial root in the positive real half-plane.

Hen creating the model for the needs of this paper, a model public project was used in which the project cash flow equals $CF = [CF_0, CF_1, CF_2, CF_3, CF_4, CF_5, CF_6] = [-6,1,1,1,1,1,1]$. It is a conventional project, it has only one change of the sign in the progression of numbers CF. In addition, it holds that $\sum_{t=0}^{n} CF_t = 0$ or also $CF_0 = -\sum_{t=1}^{n} CF_t$. Figure 1 shows the course of the investment curve $ENPV$ in dependence on the common social discount rate *k* of public projects without subsidy ($CF_0 = -6$), with 30% subsidy ($CF_0 = -4.2$), with 70% subsidy ($CF_0 = -1.8$) and with 100% subsidy ($CF_0 = 0$). Here it is clearly visible that the higher the subsidy, the higher ERR is. For 100% subsidy (the black curve) even $ERR \to \infty$, which expresses the reality of an always efficient project at the above mentioned positive financial flows from the project. Let's say about the course of the $ENPV$ curve that the curve has two asymptotes, namely a vertical one with the value of $k = -1$ and a horizontal one $ENPV = CF_0$ for $k \to \infty$. The intersection with the *y* axis, i.e. for $k = 0$, is of the value. $\sum_{t=0}^{n} CF_t$.

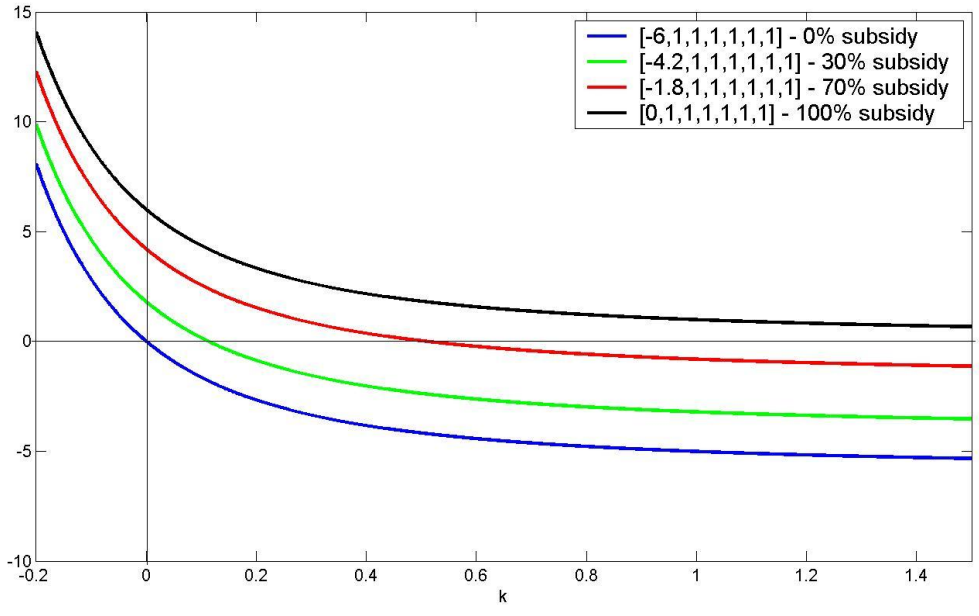


Figure 1 Graph of the course of the $ENPV$ curve

3.2 Sensitivity analysis of the model

On the above described model project, we examined the stability of the optimal solution of the return on investment depending on changes of one input quantities – the anticipated height of the projected cash-flows. We

created a 3D model of the investment curve depending on the social discount rate *k* and on the initial cash flow CF_0, see Figure 2.

For further analysis, let's divide the projection in Figure 2 into three cases. The first case is for $CF_0 \geq 0$. Here, $ENPV$ is always positive, which implies $ERR \to \infty$ and such a project is always efficient. In the real economic practice, this case only corresponds with the considered possibility of a 100% subsidy, where $CF_0 = 0$. The possibility of $CF_0 > 0$ is unreal under the conditions of financing from European funds, because it would have to be a more than 100% subsidy. The second case is for $CF_0 < -6$. Here, $ENPV$ is always negative. Thus the root ERR does not lie in positive values, the social discount rate of the project is negative and therefore the public project is not efficient. In this case it would be a project where the investment costs are higher than the discounted income from the project.

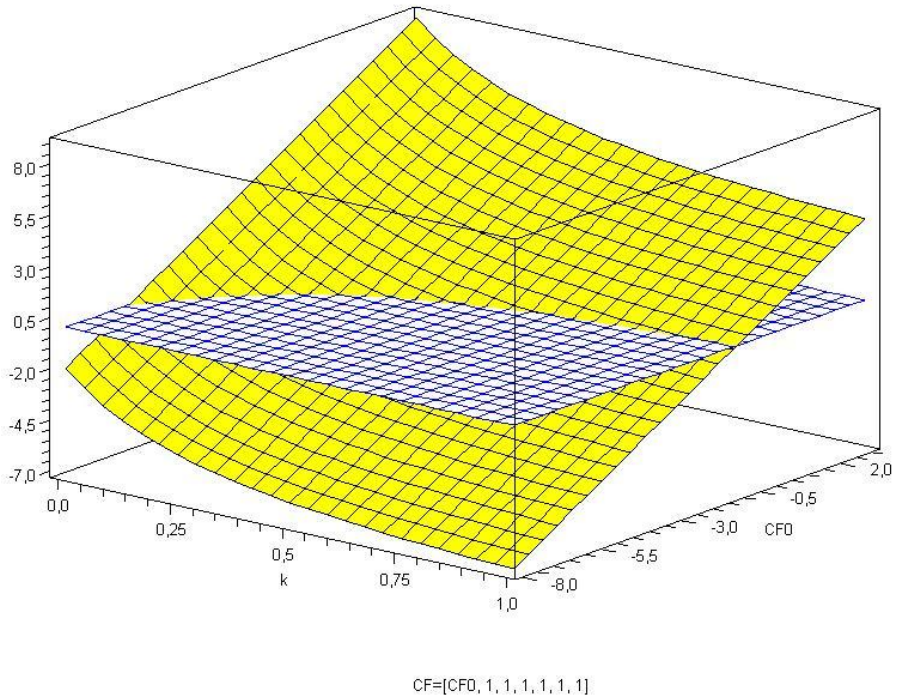


Figure 2 Three-dimensional depiction of $ENPV$

In economic practice, only the third alternative is feasible to use, that is for $CF_0 \in \langle -6, 0)$, i.e. the alternative when the investment costs are subsidized in a less than 100% share. Here, under the conditions of the model situation of positive financial flows caused by the project $ENPV$ passes from positive to negative values and therefore it is possible to find at least one positive root ERR. In Figure 3 there is the curve of the intersection of the surface $ENPV$ with the plane $ENPV = 0$ which corresponds with the roots of the investment curve. In this graph, we can deduct the marginal economic rate of return ERR (on the horizontal axis) if the initial investment is set and the amount of subsidy, i.e. the financial flow CF_0, taken into consideration. For instance, for $CF_0 = -4$, we can estimate $ERR = 0.15$. With an accurate calculation by the Maple system for this case, $ERR = 0.13$. This deviation in estimation could be further eliminated by specifying the interval of the root function display in the range $0 < k < 1$ since the social discount rate as the rate of the project return required by the investor does not excess 100% in practice (see Figure 4).

4 Conclusion

In the paper, we describe a conventional model of solving social economic efficiency of public projects where there is only one change of the cash flow polarity. The conventional model then covers the initial situation when in the submitted projects we predict positive cash flows in the operational and liquidation stages of the development project. The negative cash flow in the investment stage, i.e. $CF = [CF_0, CF_1, CF_2, CF_3, CF_4, CF_5, CF_6]$ is in our case defined by the initial conditions as $CF = [-6,1,1,1,1,1,1]$. The financial flow CF_0 refers to the difference

between investment costs (capital expenditures of a development project) and the subsidy from European funds. In the graph, this model is represented by the $ENPV$ curve which is a representation of a polynomial function of the sixth degree, which corresponds with the chosen six-year average lifetime of development projects.

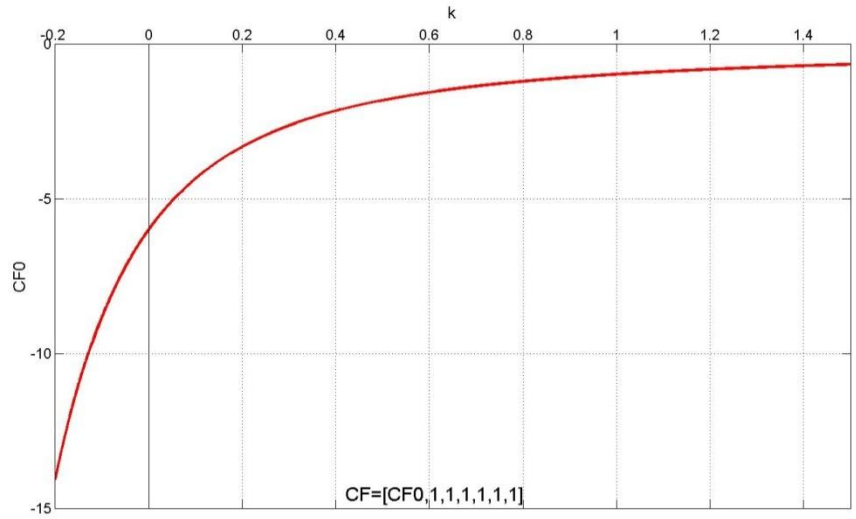


Figure 3 Dependence of the root of ERR on the CF_0 coefficient and the discount rate k

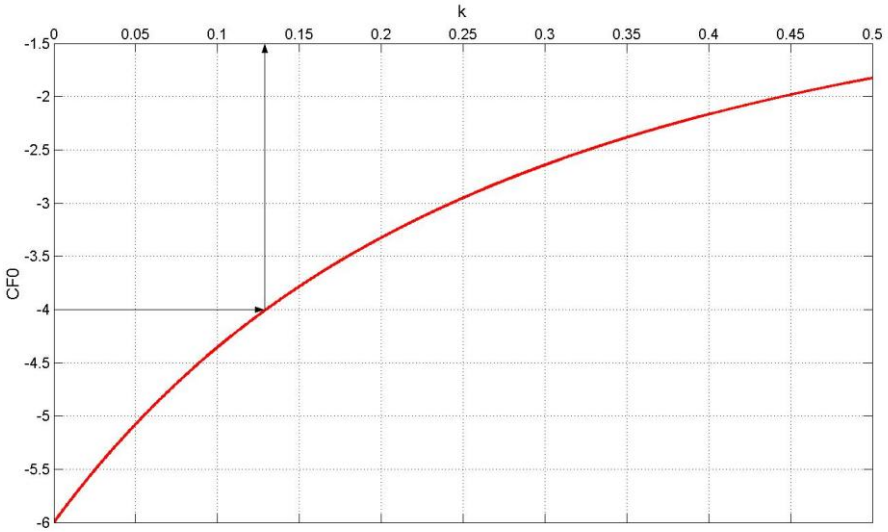


Figure 4 Detail the curve of roots ERR for $0 < k < 0.5$

Sensitivity analysis of the $ENPV$ model is carried out in a situation when the changed input quantity is the financial flow CF_0, i.e. the state when investment costs of the project are gradually subsidized from European funds up to the 100% amount. The other parameters of the project are held constant. In a two-dimensional model, this situation can be seen in the curves in Figure 1 for the states of 0%, 30%, 70%, and 100% subsidy respectively. This situation is better documented by the 3D model created with the MAPLE software (see Figure 2). The marginal situation in the model illustrates the state when the subsidy from the European funds covers 100 % of investment costs (i.e. $CF_0 = 0$). Here, it is not possible to speak rigidly about a conventional project any more since the change of the sign of financial flows does not occur and the investment curve of this marginal situation in the v 3D model does not intersect the plane $ENPV = 0$.

From the solution of the 3D model it theoretically follows that in the situation of predicted positive financial flow in the investment stage ($CF_0 > 0$) in the area to the right of the marginal investment curve the surface $ENPV$ does not intersect the plane $ENPV = 0$ any more for any chosen discount rate of the development project. In a development project constructed this way there is no marginal social discount rate (i.e. economic rate of return ERR) and so for each discount rate required by the investor or donor the project is always economically efficient in this situation. This alternative is not real under the conditions of financing from European funds because it is not possible to obtain a subsidy higher than the investment costs of the project are.

A different situation occurs in the area of real economic solutions, i.e. of negative financial flows in the investment stage of the development project and predicted positive financial flows in the operational and liquidation stages (i.e. $CF_0 < 0$). Here the surface $ENPV$ gradually intersects the plane $ENPV = 0$ along a curve which shows us in 3D depiction the set of roots of the polynomial corresponding to the marginal social discount rates of projects in the growing rate of subsidizing investment costs from European funds.

This root function can be used in 2D depiction simply for estimating ERR under various conditions of the height of investment cost subsidy (see Figure 3), namely with quite an accuracy which for practical use is comparable with a calculation based on solving a polynomial e.g. by the MAPLE program. In the graph, the confirmation of the above mentioned statements is clearly visible when in case of positive financial flow in the investment stage of the project ($CF_0 > 0$) we do not find any polynomial root, i.e. there is no economical rate of return and the project is economically efficient at any discount rate chosen by the investor or donor. In the area $CF_0 \in (-6,6)$ we will find exactly one positive solution of the root at a chosen height of subsidy. On the contrary, in case of too high investment costs of the project (i.e. $CF_0 < -6$) and on condition of zero subsidy the positive cash flows defined in the initial setting of the model will be manifested in a loss rate of the whole project, when the root passes to the negative area of ERR .

The above described model predicts a standard situation of public development project construction for an evaluation from the perspective of granting a subsidy from European funds, when the projects are designed as conventional one. However, in practice a situation which is closer to the real state in economy occurs during the project implementation, i.e. for example the first financial flow in the operational stage (project beginning) is necessarily negative or on the contrary the final financial flow in the liquidation stage of the project (e.g. environmental damage liquidation) is necessarily negative. Here the use of the created 3D model suggests itself for solving such non-predicted situations of a non-conventional project with a view of determining the resulting efficiency of the projects subsidized from European funds.

References

[1] Balážová, E., and Papcunová, E.: *Manažment samospráv*, 1st. edition, Nitra: Municipalia, 2008. 134 s. ISBN 978-80-552-0045-3

[2] Brealey, R., and Myers, S., and Allen, F.: *Principles of Corporate Finance*. 2nd. Edition, New York: McGraw-Hill Companies, 2011. 524 s., ISBN 978-0-07-353074-1

[3] Drábek, J., and Hora, J.: *Algebra Polynomy a rovnice*, Plzeň: Západočeská univerzita, 2001. 125 s. ISBN 80-7082-787-4

[4] Jiříček, P., and Dvořáková, S.: *Závislost průběhu investiční křivky na změně vstupních parametrů pro potřeby hodnocení projektů metodou cost-benefit analýzy*. In Matematika v ekonomické praxi, 1 CD ROM, Jihlava, 2012, s. 119-128, ISBN 978-80-87035-62-7

[5] Levy, Ch., and Marshall, S.: *Kapitálové investice a finanční rozhodování*, 1st. Edition Praha: GRADA, 1999. 917 s., ISBN 80 -7169-504-1

[6] Rektorys, K.: *Přehled užité matematiky*., Praha: Prometheus, 2000. ISBN 80-7196-179-5

[7] Richtarová, D: *The sensitivity analysis application and net present value pyramidal decomposition for investment valuation*, MEKON 2008 : CD příspěvků X. ročníku mezinárodní konference Ekonomické fakulty Vysoké školy báňské - Technické univerzity Ostrava, VŠB - Technická univerzita Ostrava, 2008, s. 1-12. ISBN 978-80-248-1704-0

[8] Stiglitz, J.: *Ekonomie veřejného sektoru*. 1st. Edition. Praha: GRADA, 1999. 661 s., ISBN 80-7169-454-1

[9] European Commission. Directorate General Regional Policy. [online]. 2008 [cit. 2013-4-18]. *Guide to Cost Benefit Analysis of Investment Projects.* http://ec.europa.eu/regional_policy/sources/docgener/guides/cost/guide2008_en.pdf

Capacitated Postman Problem

Jan Fábry[1], Jan Pelikán[2]

Abstract. The paper deals with the Chinese Postman Problem. It has many modifications that are summarized in the first part of the paper. The topic of the work is solving the case study of urban household waste collection, i.e. a Capacitated Postman Problem. In this task, distances and requirements are given for each arc of the network and cyclical routes are searched with the objective to minimize total costs. Each route must respect the vehicle capacity. Moreover, start and finish nodes may be defined for each route. In the paper, two mathematical models of integer programming are formulated. Due to the NP-hardness of the problem, heuristic method is proposed.

Keywords: arc routing, Chinese postman problem, rural postman problem, Eulerian cycle, capacitated postman problem.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The Arc routing problem (ARP) deals with searching for the Eulerian cycle in the graph. In 1763, Leonhard Euler defined the conditions for finding the cycle containing each arc *exactly once*. The problem is polynomially solvable in case all nodes in the connected graph have an even degree, i.e. an even number of incident arcs. In 1962, M. Guan generalized the problem to the Chinese Postman Problem (CPP), in which the objective is to find a minimal-distance cycle containing each arc *at least once*. The problem is also polynomially solvable. First, we search arcs which will be used multiple times to get *Eulerian graph*[3]; minimum-cost matching of all odd-degree nodes is used for this purpose. Then, the *Eulerian cycle* is found using Fleury's algorithm.

The CPP is formulated both for directed and for undirected graphs, and also for mixed graphs consisting of directed and undirected arcs. Solution of the CPP consists in doubling any arcs to obtain the *unicursal graph*. Graph is unicursal if:

a) in an undirected graph, all nodes have even degree,

b) in a directed graph, in-degree of each node is equal to its out-degree,

c) in a mixed graph, each subset of nodes has to satisfy the "balanced set condition".

Searching for the extended graph satisfying the condition of unicursality is polynomial in case of a) and b), and NP-hard in case of c). The special extension of the CPP with the asymmetric matrix of arcs length, i.e. where $c_{ij} = c_{ji}$ is not valid for all arcs in the graph, called the Windy Postman Problem, is also NP-hard (Brucker, 1981 and Guan, 1984).

Another NP-hard problem is the Rural Postman Problem (RPP), in which the set of all arcs contains the subset of arcs that have to be traversed (Eiselt et al., 1995). All other arcs can be also included in the final route. The paper deals with the Garbage Collection Problem that is analogous to the Vehicle Routing Problem. We search for the cyclical routes of vehicles with given capacity Q and their common depot, which has to be included in each route. Each obligatory arc has its collection requirement and therefore, has to be included at least once on any route. Although the arc can be included in multiple routes or it can be traversed several times on the same route, requirements are collected just once. Therefore, we have to consider the arc is serviced on the route or unserviced, i.e. traversed without service (so called deadheading in logistics). Each route has to satisfy the vehicle capacity and the objective is to minimize the total length of all routes. The problem is called the Capacitated Chinese Postman Problem (CCPP) and it is defined both for directed and for undirected graphs (Longo et al., 2006). Further, we consider only undirected graphs. The problem is generalization of the RPP. Because it is the NP-hard problem, heuristics and metaheuristics (Hertz et al., 2000) are used for the solving real large scale instances. In the paper, we offer heuristics very easy to apply for garbage collection problems.

[1] Jan Fábry, University of Economics Prague, Faculty of Informatics and Statistics, Department of Econometrics, Prague 3, W. Churchill sq.4, fabry@ vse.cz
[2] Jan Pelikán, University of Economics Prague, Faculty of Informatics and Statistics, Department of Econometrics, Prague 3, W. Churchill sq.4, pelikan@vse.cz
[3] Eulerian graph is a graph containing an Eulerian cycle

2 The model of the CCPP on undirected graph – directed solution

Let $G = \{V, E\}$ be an undirected graph, where V is a set of n nodes and E is a set of arcs. Let us define $R \subseteq E$ as the set of obligatory arcs. For each arc $(i, j) \in E$, it is given its length c_{ij}, for each arc $(i, j) \in R$, it is known its collection requirement d_{ij}. Depot is situated in node 1, where m vehicles with capacity Q are prepared for garbage collection. Although graph G is undirected, solution of the problem will contain the orientation of the traverse through the arcs. Thus, we will be searching for the directed cyclical routes while in case of undirected graph the direction is not important.

Let binary variable x_{ij}^k is equal to 1 if arc $(i, j) \in E$ is traversed by vehicle k in the direction from node i to node j, 0 otherwise. Binary variable y_{ij}^k is equal to 1 if arc $(i, j) \in R$ is serviced on route k, i.e. requirement d_{ij} is collected by vehicle k, 0 otherwise. If arc (i, j) is serviced on route k it must lie on this route, therefore it is valid:

$$y_{ij}^k \leq x_{ij}^k. \tag{1}$$

If the requirement d_{ij} of arc (i, j) is collected on route k, then either $y_{ij}^k = 1$ or $y_{ji}^k = 1$. Because the arc must be serviced exactly once, i.e. on the only one route, the following equations must be respected:

$$\sum_{k=1}^{m} \left(y_{ij}^k + y_{ji}^k\right) = 1, \quad \forall (i,j) \in R. \tag{2}$$

The sum of requirements d_{ij} serviced on route k must not exceed the vehicle capacity Q:

$$\sum_{(i,j)\in R} d_{ij} y_{ij}^k \leq Q, \quad k = 1,2,\dots,m. \tag{3}$$

Necessary condition for the existence of cyclical routes is unicursality of the graph, i.e. a number of arcs entering each node must be equal to a number of arcs leaving it:

$$\sum_{\substack{i \\ (i,j)\in E}} x_{ij}^k = \sum_{\substack{i \\ (j,i)\in E}} x_{ji}^k, \quad \forall j \in V, \ k = 1,2,\dots,m. \tag{4}$$

Nevertheless, the condition of unicursality does not ensure the route is connected. Therefore, illegal cyclical subtour not containing depot can exist and it is necessary to prevent it.

If a set of arcs (including their multiplicity) for certain k satisfies the condition of unicursality and does not contain node 1, arcs represent a multigraph consisting of nodes subset $S \subset V - \{1\}$ and set of arcs $E(S)$. If a number of arcs (including their multiplicity) is less than $|S|$, then multigraph $\{S, E(S)\}$ does not represent the cyclical route. However, if $|E(S)| > |S| - 1$ then illegal cycle is possible and at least one arc connecting S and $V - S$ has to be added. The number of arcs in $E(S)$ is

$$|E(S)| = \sum_{\substack{(i,j)\in E \\ i,j\in S}} x_{ij}^k + x_{ji}^k, \quad k = 1,2,\dots,m. \tag{5}$$

Thus, avoiding of illegal cycles for tour k can be expressed as follows:

1) If $|E(S)| > |S| - 1$ then at least one arc must connect S and $V - S$, i.e. $x_{ij}^k \geq 1$ for $i \in S$ and $j \in V - S$.

2) If $|E(S)| \leq |S| - 1$ then such arc is not necessary.

These conditions are included in the following mathematical model of CCPP based on Eiselt et al. (1995):

$$\sum_{k=1}^{m} \sum_{(i,j)\in E} x_{ij}^k \, c_{ij} \to min, \tag{6}$$

$$y_{ij}^k \leq x_{ij}^k, \quad (i,j) \in E, k = 1,2,\dots,m, \tag{7}$$

$$\sum_{k=1}^{m} \left(y_{ij}^k + y_{ji}^k\right) = 1, \quad \forall (i,j) \in R, \tag{8}$$

$$\sum_{(i,j)\in R} d_{ij} y_{ij}^k \leq Q, \quad k = 1,2,\dots,m. \tag{9}$$

$$\sum_{\substack{i \\ (i,j)\in E}} x_{ij}^k = \sum_{\substack{i \\ (j,i)\in E}} x_{ji}^k, \quad j = 1,2,\dots,n, \ k = 1,2,\dots,m. \tag{10}$$

$$\sum_{\substack{(i,j)\in E \\ i,j\in S}} x_{ij}^k + x_{ji}^k \ \leq \ |S| - 1 + n^2 u_k^S, \qquad k = 1,2,\dots,m, \ S \subset V - \{1\}, \ S \neq \emptyset, \tag{11}$$

$$\sum_{\substack{i\in S \\ (i,j)\in E}} \sum_{j\notin S} x_{ij}^k \ \geq \ 1 - w_k^S, \qquad k = 1,2,\dots,m, \ S \subset V - \{1\}, \ S \neq \emptyset, \tag{12}$$

$$u_k^S + w_k^S \leq 1, \qquad k = 1,2,\dots,m, \ S \subset V - \{1\}, \ S \neq \emptyset, \tag{13}$$

$$x_{ij}^k \in \{0,1\}, \ y_{ij}^k \in \{0,1\}, \qquad (i,j) \in E, \ k = 1,2,\dots,m, \tag{14}$$

$$u_k^S \in \{0,1\}, \ w_k^S \in \{0,1\}, \qquad k = 1,2,\dots,m, \ S \subset V - \{1\}, \ S \neq \emptyset. \tag{15}$$

The objective (6) is to minimize the total cost of the travel corresponding to the total length of all routes. Illegal subtours are eliminated respecting inequalities (11) – (13). For each S and each k, there are binary variables u_k^S and w_k^S introduced. If $|E(S)| > |S| - 1$, variable u_k^S is equal to 1, and $w_k^S = 0$. Then, inequalities (12) assure subset S must be connected with subset $V - S$ which contains depot.

There are two possibilities for using the mathematical model:

1) At the beginning, all subsets S are generated and then, the mathematical model is applied. The disadvantage of this approach is computational complexity because of huge number of all subsets S.

2) The mathematical model is run without conditions eliminating illegal subtours. In case any illegal subtour is found, inequalities (11) – (13) are added for all vehicles and subset S containing all nodes included in the illegal subtour. Then, extended model is run again. This process is repeating until result is feasible in term of the elimination of illegal cycles. Although the model is not so complex (at the beginning of the process), the disadvantage of this approach is multiple run of the model, and also the gradual increase in the number of subsets S.

The use of the second approach will be demonstrated in Section 5 of the paper.

3 The model of the CCPP on undirected graph – undirected solution

In Section 2, the model is proposed for undirected graph and the optimal solution determines directed cyclical routes. In this section, we propose the model generating sets of arcs laying on each cycle; directed routes are found subsequently. Let integer variable x_{ij}^k represent a number of times arc $(i, j) \in E$ is traversed by vehicle k (in any direction). Variables x_{ij}^k and y_{ij}^k are defined only for $i < j$.

While the objective function (6), constraints (7) and (9) are unchanged, equations (8) and (10) must be modified:

$$\sum_{k=1}^{m} y_{ij}^k = 1, \quad \forall (i,j) \in R, \tag{8a}$$

$$\sum_{\substack{i \\ (i,j)\in E}} x_{ij}^k + \sum_{\substack{i \\ (j,i)\in E}} x_{ji}^k = 2 f_j^k, \quad j = 1,2,\dots,n, \ k = 1,2,\dots,m. \tag{10a}$$

Integer variable f_j^k in (10a) indicates a number of times node j is traversed by vehicle k. Constraints (11) and (13) are excluded from the model, and inequalities (12) eliminating illegal subtours are replaced by:

$$\sum_{\substack{i\in S \\ (i,j)\in E}} \sum_{j\notin S} x_{ij}^k + \sum_{\substack{j\in S \\ (j,i)\in E}} \sum_{i\notin S} x_{ji}^k \geq \frac{1}{M} \sum_{\substack{i\in S \\ (i,j)\in E}} \sum_{j\in S} x_{ij}^k, \qquad k = 1,2,\dots,m,\ S \subset V-\{1\}, S \neq \emptyset, \tag{12a}$$

where M is a high constant. The condition (12a) assures subset S must be connected with subset $V - S$ which contains depot in case there is at least one arc (i, j) with $i, j \in S$, traversed by vehicle k. Apparently, variables (15) are not relevant in the proposed model and integrality constraints for variables x_{ij}^k replace binary ones:

$$x_{ij}^k \in Z_+,\ y_{ij}^k \in \{0,1\}, \qquad (i,j) \in E,\ k = 1,2,\dots,m. \tag{14a}$$

After finding the optimum solution it is necessary to find directed cycles, e. g. with the use of Fleury's algorithm. Similarly to the model (6) – (15), two approaches can be also applied for our modification. We demonstrate the second one in Section 5 of the paper.

4 Heuristics for the CCPP on undirected graph

Because of the NP-hardness of the problem, solution of large scale instances needs the use of heuristics. In the paper, we offer the algorithm using the shortest path approach for finding cyclical routes. Let us assume notation introduced at the beginning of Section 2. Proposed algorithm will find the cyclical routes given by the sequence of nodes $v_1^k, v_2^k, \dots$ (k is the number of the route). The length of the shortest path from node i to node j is denoted p_{ij}. Total length of cyclical route k is F_k, total amount of units loaded on vehicle k is L_k and a number of nodes included in tour k is N_k. The heuristics consists of 5 steps:

Step 1:

$k = 0$.

Step 2:

if $R = \emptyset$ then goto step 5

else $k = k + 1$; $L_k = 0$; $F_k = 0$; $t = 1$; $v_t^k = 1$; $N_k = 1$.

Step 3:

if $R = \emptyset$ then

begin

$v_{t+1}^k = 1$; $N_k = N_k + 1$; $F_k = F_k + p_{v_t^k,1}$;

goto step 5

end.

Step 4:

if there is arc $(v_t^k, j) \in R$ and $L_k + d_{v_t^k,j} \leq Q$ then

begin

$F_k = F_k + c_{v_t^k,j}$; $L_k = L_k + d_{v_t^k,j}$; $N_k = N_k + 1$; $R = R - (v_t^k, j)$; $t = t + 1$; $v_t^k = j$;

if $v_t^k = 1$ then goto step 2 else goto step 3

end

else

if there is $(i, j) \in R$, $i \neq 1$ and $L_k + d_{ij} \leq Q$ then

begin

$F_k = F_k + p_{v_t^k,i} + c_{ij}$; $L_k = L_k + d_{ij}$; $N_k = N_k + 2$; $R = R - (i, j)$; $v_{t+1}^k = i$; $v_{t+2}^k = j$; $t = t + 2$;

if $v_t^k = 1$ then goto step 2 else goto step 3

end

else goto step 2.

Step 5:

End.

Because, at the end of the algorithm run, the value of k corresponds to the number of realized cyclical routes, the total length of them is given as follows:

$$G = \sum_{l=1}^{k} F_l. \tag{16}$$

In Step 4, arc $(i, j) \in R$ is selected randomly. The algorithm can by modified by finding the closest arc to node v_t^k. Therefore, the shortest paths are used for this purpose. In addition, instead of random selection of arc $(v_t^k, j) \in R$, we can use specific rules in case there are more possibilities to select, e.g. the longest arc, or the shortest one. In the demonstration of using the algorithm in the following section, the longest arc is preferred.

5 Computational experiments

Let us demonstrate operating of the mathematical models and the heuristics on the following example. In case of the mathematical models, the second approach mentioned above will be applied. For this purpose, optimization system Lingo and VBA for MS Excel are used.

Example

The following graph is defined (see Fig. 1). Each arc is evaluated by its length and requirement for collection (values in parentheses). Arc (4, 6) has no requirement, i.e. it is optional. Node 1 is the depot, in which 3 vehicles are prepared for collection. Each vehicle has the capacity 20 units. Totally, 21 units have to be collected.

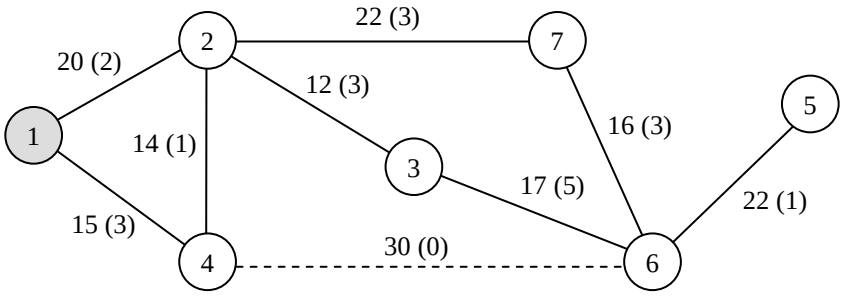


Figure 1 Garbage collection graph

Application of the model generating directed solution

In the first step, we obtained the initial solution without conditions eliminating illegal subtours. In further steps, these conditions are introduced to the model. In each step, subsets *S* of nodes, corresponding to illegal subtours generated in previous steps, are added (see Table 1).

Step	Subset *S*	Tour 1	Tour 2	Objective
1	-	1-2-4-1	2-3-6-5-6-7-2	160
2	{2,3,5,6,7}	1-2-7-6-3-2-4-1	5-6-5	160
3	{5,6}	1-4-2-3-6-5-6-7-2-1	2-3-2	184
4	{2,3}	1-2-3-6-5-6-7-2-4-1	2-4-2	188
5	{2,4}	1-2-4-1	2-4-2-7-6-5-6-3-2	188
6	{2,3,4,5,6,7}	1-4-1	1-4-2-7-6-5-6-3-2-1	190

Table 1 Application of the model – directed solution

Application of the model generating undirected solution

In each step, sets of arcs a vehicle is to traverse are found. Then, directed cycles are generated with the use of Fleury's algorithm. Conditions eliminating illegal subtours are introduced to the model until the feasible solution is obtained (see Table 2).

Step	Subset *S*	Tour 1	Tour 2	Objective
1	-	1-2-4-1	2-3-6-5-6-7-2	160
2	{2,3,5,6,7}	1-4-2-3-6-5-6-7-2-1	2-4-2	188
3	{2,4}	1-2-4-1	2-4-2-7-6-5-6-3-2	188
4	{2,3,4,5,6,7}	1-4-1	1-4-2-7-6-5-6-3-2-1	190

Table 2 Application of the model – undirected solution

Application of the heuristics

The solution obtained with the use of the proposed heuristics generates two cyclical routes: 1-2-7-6-5-6-3-2-4-1 with the length of 160, and 1-4-1 with the length of 30. Total length is 190. Thus, we obtain the alternative optimal solution.

6 Conclusions

The paper is based on the case study of the optimization of routes in garbage collection problem in Poděbrady city. The real problem contains specific conditions, e.g. separated locations of depot and dump. The instance can be classified as the Capacitated Rural Postman Problem. In the paper, two mathematical models and easy construction heuristic algorithm are introduced and demonstrated on the illustration instance. Heuristics can be extended to other steps that can make the algorithm more effective. Other conditions respecting specific requirements of the city could be included. In addition, the solution obtained with the use of the algorithm can be improved using other heuristics and metaheuristics. These are issues of the authors' future research.

Acknowledgements

This research was supported by the Czech Science Foundation, grant No. P403/12/1947 and by the project F4/11/2013 founded by the Internal Grant Agency of the University of Economics Prague.

References

[1] Brucker, P. The chinese postman problem for mixed graphs. In *Graph theoretic Concepts in Computer Science, H. Noltemeier* (ed.), Springer, Berlin, 1981, pp. 354-366.

[2] Eiselt, H. A., Gendreau, M. and Laporte, G. Arc routing problems, Part II: The rural postman problem. *Operations Research*. Vol. 43, No. 3, 1995, pp. 399-414.

[3] Guan, M., G. On the windy postman problem. *Discrete Appl. Math*., Vol. 9, No. 1, 1984, pp. 41-46.

[4] Hertz, A., Laporte, G. and Mittaz, M. A Tabu search heuristic for the capacitated arc routing problem, *Operations Research*. Vol. 48, No. 1, 2000, pp. 129-135.

[5] Longo, H., Poggi de Aragao, M. V. S. and Uchoa, E. Solving capacitated arc routing problems using a transformation to the CVRP. *Computers and Operations Research*, Vol. 33, No. 6, 2006, pp. 1823-1837.

Kuhn-Tucker optimality conditions in Model of a Monopoly Production Price Differentiation

Eleonora Fendeková [1], Michal Fendek [2]

Abstract. One of the effective tools to exercise monopoly's market position or economic strength to define – in accordance with the monopoly's interests – a market price on the level guaranteeing maximum profit is a price differentiation.
In this article we discuss the general aspects of quantitative analysis of the monopoly price differentiation models and we will analyze in more detail the models of consumer's utility maximization in the conditions of price differentiation of the goods and a model of monopoly's profit maximization with price-differentiated production.
A problem of monopoly's profit optimization is built on a specific hypothesis that a monopoly gives a consumer an „all or nothing" offer, meaning that a differentiated price only applies in a case that a consumer buys the whole amount of a good offered with a differentiated price and a corresponding revenue then guarantees a monopoly's maximum profit.
We will analyze a role of consumer's behavior optimization when the consumer's willingness to spend his funds on goods with differentiated prices is related to maximization of his total utility. For the optimization problems we will formulate the Kuhn-Tucker optimality conditions and we will study their interpretation options.

Keywords: price differentiation, Lagrange function Kuhn-Tucker optimality conditions, first, second and third degree price differentiation

JEL Classification: C62, L13, L22
AMS Classification: 90C46, 49M05

1 Introduction

Due to a specific position which the subjects on a supply side have on an imperfect competition market, the producers can promote their interests without immediate danger of a competitor producing the same or similar product entering a relevant market.

We may speak of monopoly price differentiation when a monopoly uses its market position, or economic strength to define – in accordance with its interests – a market price on the level guaranteeing maximum profit. Ultimately the monopoly uses its monopolistic position to generate a monopoly profit over and above the profit achievable in the conditions of perfect competition.

A tool to achieve this goal is a market price and a monopoly has a considerable say in defining the price.

One of the effective tools to use monopoly's market position or economic power to, according to its interests, set a market price to the level which guarantees maximum profit, is a price differentiation. We can speak of a price differentiation in a situation when the identical products are being sold at different prices while this inequality is not due to different production costs. In literature price discrimination is often used as a synonym to price differentiation. We think, however, that this collocation has a slightly negativistic tone which does not correspond to its factual technical meaning, therefore following we will prefer price differentiation to indicate this microeconomic attribute.

In this article we will analyze a role of a consumer behavior optimization, whose willingness to spend his funds to purchase a good with a differentiated price is related to maximization of his total utility.

[1] University of Economics Bratislava, Department of Business Economics, Dolnozemska cesta 1/b, 852 35 Bratslava, Slovakia, e-mail: nfendek@dec.euba.sk.
[2] University of Economics Bratislava, Department of Operations Research and Econometrics, Dolnozemska cesta 1/b, 852 35 Bratislava, Slovakia, e-mail: fendek@dec.euba.sk

2 Conditions for Price Differentiation

Price differentiation is of course possible only under an assumption that a consumer is willing to pay different prices for different amounts of goods. That is, for example, at a lower price he is willing to purchase more – the price-demand function is decreasing. Besides, an effective price differentiation is based on the assumption that a consumer who purchased a product at a lower, favored price would not resell it at a slightly higher price to another consumer, who would otherwise paid a higher price. When such arbitrage deals cannot be ruled out price differentiation cannot be applied.

If a price differentiation can be applied then in fact a monopoly uses its monopolistic position to gain monopolistic profit. Carlton - Perloff [2] show, that a monopoly can use price differentiation in two ways:

A tool for this is a market price while a monopoly can use price differentiation in following areas:

• to set a different price when purchasing different amounts of a same good,

• to set a different price for different consumers or consumer groups.

In a first case it is relatively simple to use a price differentiation. When meeting the conditions of purchasing volume an agreed discount is given. To successfully apply the second method of a price differentiation a mechanism must exist to identify a consumer belonging to a certain consumer group. O'Sullivan - Sheffrin - Perez [7] show, that we can find three types of price differentiation according to traditional classification:

(a) First degree price differentiation. This type of price differentiation of a monopolistic company is sometimes called a perfect price differentiation. A monopoly uses its privileged position on a market to set different prices for different volumes of a same product as well as for different consumer groups. Meaning that a seller in fact sets an individual price for each unit of a product and a price of a certain unit corresponds with a willingness to pay a maximum price by a certain consumer at a certain conditions.

(b) Second degree price differentiation represents a situation when the product prices depend on the purchasing volume of the products but do not depend on any characteristic a consumer may have. This phenomenon of pricing is also interpreted as a nonlinear pricing. Identical pricelists apply to all consumers but the pricelists vary for different purchasing volumes. A monopoly thus does not differentiate prices for particular consumers or consumer groups. The differentiation applies to varying amounts of purchased goods. An example of this approach to price differentiation is a volume rebate.

(c) Third degree price differentiation. With this type of price differentiation a monopolistic company sells any amount of goods at a same unit price. The price varies however for a specifically defined consumer groups. Third degree price differentiation represents probably the most common form of price differentiation. These are for example various types of discounts for students, child or pensioner's travel tickets, different prices on a different days (weekend discounts) and so on.

3 Model of Optimal Consumer Behavior in the Conditions of Price Differentiation

Let's now examine a simplified analytical model of price differentiation of two potential consumers S_1, S_2, who both have a price-differentiated product in their market basket. X represents its purchase in units and w is the total value of purchase of the other goods in monetary units. In [2], we can see, that a consumer can possibly decide:

- to exclusively purchase the studied product with a differentiated price and not to purchase the other goods at all,
- to purchase only the other goods from his market basket and not to purchase the studied product with a differentiated price,
- to purchase a full market basket represented by the studied product with a differentiated price as well as other goods.

Total utility expressed in monetary units, which a consumer feels when purchasing all the goods from the market basket is represented in a form of utility function:

$$f_i(x_i, w_i) = u_i(x_i) + w_i \qquad i = 1,2 \tag{1}$$

where

$u_i(x_i): R \to R$, $i = 1,2$ - is a concave, continuous and differentiable utility function for a product with differentiated price and represents a feeling of utility of a consumer in monetary units corresponding with a purchase of *x* units of the product,

w_i – expenses on other goods in the market basket,

$f_i(x_i, w_i): R^2 \to R$, $i = 1,2$ - consumer's total utility function.

To simplify, we will assume that utility for zero purchase of the studied product has a standardized zero value $u_i(0) = 0$. Maximum willingness of a consumer *i* to pay a certain price in monetary units for a purchase of x_i units of the product is represented by a function $r_i(x_i)$. This function is a solution to an equation:

$$u_i(0) + w_i = u_i(x_i) - r_i(x_i) + w_i \qquad i = 1,2 \tag{2}$$

where on the left side is utility of a purchase of zero units of the product plus a value of the other goods from the market basket and on the right side is utility of a purchase of x_i units of the product reduced by a payment for their purchase plus a value of the other goods from the market basket. When a condition of a standardized zero value of utility at a zero consumption of the product from (2) validates, we get

$$u_i(x_i) \equiv r_i(x_i) \qquad i = 1,2 \tag{3}$$

In other words, a consumer S_i is willing to pay for x_i units of the product a maximum price which correspondents to his feeling of satisfaction from the product purchase represented in monetary units. The function of utility can be thus, with a certain degree of approximation, perceived as a function representing a willingness of a consumer to pay for the studied product the maximum price $r_i(x_i)$. Function $r_i(x_i)$ has after all one more interesting economic interpretation. Its first derivation function $r_i'(x_i)$, which is a function of a marginal willingness of a consumer to pay a relevant price for a certain demand volume, in fact represents how much a consumer is willing to pay for the last unit purchased. So the value of the function of marginal willingness to pay represents the price *p* for which a consumer is willing to buy the whole amount of *x* units of goods:

$$r_i'(x_i) = p \qquad i = 1,2 \tag{4}$$

The function $r_i'(x_i)$ of marginal willingness of a consumer to pay an amount $r_i(x_i)$ de facto ultimately represents an inverse demand function of an *i*-th consumer

$$x_i = (r_i')^{-1}(p) \qquad i = 1,2 \tag{5}$$

Let's now examine a consumer's behavior optimization problem or a total utility maximization problem of a consumer who has the funds m_i at his disposal. He uses these funds to purchase a market basket (x_i, w_i), while he purchases x_i units of the studied product at a market price *p* and the variable w_i represents the total expenses related to the purchase of other goods from the market basket. The problem is to find the values of variables x_i and w_i so that the value of the utility function (1) would be maximal while respecting a budget constraint of a consumer. Mathematical programming problem for *i*-th consumer is analytically represented by:

$$f_i(x_i, w_i) = u_i(x_i) + w_i \to max$$

subject to (6)

$$px_i + w_i = m_i$$

$$x_i, w_i \geq 0$$

This optimization problem of mathematical programming represents maximization constrained extrema problem. Let's modify this problem to a standard form, which is a form of a minimization problem:

$$-f_i(x_i, w_i) = -u_i(x_i) - w_i \to min$$

subject to (7)

$$px_i + w_i = m_i$$

$$x_i, w_i \geq 0$$

Let's formulate a generalized Lagrangian function for this problem. Let us mention that a generalized Lagrangian function does not explicitly include the conditions of non-negative variables, these are accounted for implicitly in a Kuhn-Tucker optimality conditions. A generalized Lagrangian function of a mathematical programming problem (7) is:

$$\mathcal{L}_i(x_i, w_i, \lambda_i) = -u_i(x_i) - w_i + \lambda_i(px_i + w_i - m_i) \tag{8}$$

Kuhn-Tucker optimality conditions for Lagrangian function (8) of the *i*-th consumer S_i are translated in a form:

$$\frac{\partial \mathcal{L}_i(x_i, w_i, \lambda_i)}{\partial x_i} \geq 0 \qquad \frac{\partial \mathcal{L}_i(x_i, w_i, \lambda_i)}{\partial w_i} \geq 0 \qquad \frac{\partial \mathcal{L}_i(x_i, w_i, \lambda_i)}{\partial \lambda_i} = 0$$

$$x_i \frac{\partial \mathcal{L}_i(x_i, w_i, \lambda_i)}{\partial x_i} = 0 \qquad w_i \frac{\partial \mathcal{L}_i(x_i, w_i, \lambda_i)}{\partial w_i} = 0 \tag{9}$$

$$x_i \geq 0 \qquad w_i \geq 0$$

After we substitute an analytical form of the Lagrangian function (8) we can restate the Kuhn-Tucker optimality conditions (9) of the total utility maximization problem as follows:

$$-u_i'(x_i) + \lambda_i p \geq 0 \quad (9.1) \qquad -1 + \lambda_i \geq 0 \quad (9.4) \qquad p x_i + w_i - m_i = 0 \quad (9.7)$$

$$x_i(-u_i'(x_i) + \lambda_i p) = 0 \quad (9.2) \qquad w_i(-1 + \lambda_i) = 0 \quad (9.5)$$

$$x_i \geq 0 \quad (9.3) \qquad w_i \geq 0 \quad (9.6)$$

In other words, if a consumer aspires to identify an optimal consumer strategy $(x_i^\star, w_i^\star)$, meaning that a consumption of $x_i^\star$ units of the product with a differentiated price p and the expenses $w_i^\star$ of consumption of the other goods in a market basket maximize his total utility $f_i(x_i, w_i) = u_i(x_i) + w_i$, then such a Lagrange multiplier λ_i^* must exist, for which the Kuhn-Tucker optimality conditions (9) are met, i.e. the variables vector $(x_i^\star, w_i^\star, \lambda_i^*)$ is a solution to the system of equations and inequalities (9.1),..., (9.7).

We can derive some interesting consequences for the optimal combination of a supply and a price of a price-differentiated product from validation of the Kuhn-Tucker optimality conditions (9) for the total utility maximization problem (6)

a) Validity of the condition (9.7) guarantees that a consumer has precisely such an optimal consumer strategy $(x_i^\star, w_i^\star)$, that means such an optimal consumption of $x_i^\star$ units of the product with a differentiated price and on optimal expenses $w_i^\star$ of consumption of the other goods in a market basket, which he can implement using available financial resources m_i. In other words the optimal consumer strategy $(x_i^\star, w_i^\star)$ meets the condition of a budget constraint.

b) Let's now analyze a structure of an optimal market basket meeting the obvious assumption that a consumer has some other goods in his market basket than the studied product. That means that a variable representing the expenses $w_i^\star$ of consumption of the other goods in the market basket is positive $w_i^\star > 0$. Validity of (9.5) results in the optimal value of Lagrange multiplier $\lambda_i^* = 1$. However if $\lambda_i^* = 1$ and (9.2) and (9.1) validate, then at the same time for the optimal positive volume of consumption $x_i^\star > 0$ stands

$$-u_i'(x_i^\star) + \lambda_i^* p = 0 \quad \wedge \quad \lambda_i^* = 1 \Rightarrow$$

$$p = u_i'(x_i^\star) \tag{10}$$

Relation (10) represents a significant phenomenon of consumer behavior in the conditions of differentiated price. Above all we realize that (10) in fact represents an inverse demand function or a price-demand function and

$$p = p_i(x_i) = u_i'(x_i) \tag{11}$$

Price-demand function of the *i*-th consumer then determines the price $p = p_i(x_i)$, for which a consumer is willing to buy x_i units of goods, while based on (10) this price is equal to marginal utility $u_i'(x_i)$ corresponding to purchase of x_i units of the good.

Therefore a consumer increases his consumption of the price-differentiated product as long as its price is lower than marginal utility $u_i'(x_i)$ corresponding to the last purchased unit of the product. Marginal utility of the last purchased unit

$$u_i'(x_i) = u_i(x_i) - u_i(x_i - 1)$$

in monetary units represents an increase of utility caused by purchase of the last, x_i-th unit of the product. Since marginal utility corresponds to the product price, a consumer spends exactly the same amount to purchase this last unit than an increase of his utility.

c) At last let's address a theoretical yet methodologically interesting hypothesis when a consumer decides to spend all his available funds to purchase solely the product with differentiated price. In this case his optimal expenses $w^\star$ of consumption of the other goods in the market basket are zero $w^\star = 0$. For $x_i^\star$ based on (9.7) stands

$$x_i^{\star} = \frac{m_i}{p} > 0 \tag{12}$$

In this situation the optimal value of Lagrange multiplier λ_i^* resulting from optimality conditions (9.4) and (9.5) belongs to the interval $\langle 1, \infty)$. The case when $\lambda_i^* = 1$ is analyzed in section (b). For its value higher than one $\lambda_i^* > 1$, and for positive consumption of the product $x_i^{\star} > 0$ resulting from (9.2), the optimality condition (9.1) is realized as an equation and

$$-u_i'(x_i^{\star}) + \lambda_i^* p = 0.$$

At the same time $\lambda^* > 1$, so the Lagrange multiplier λ^* can be represented in a form $\lambda_i^* = 1 + \varepsilon, \epsilon > 0$ and therefore

$$-u_i'(x_i^{\star}) + (1+\varepsilon)p = 0 \qquad / \frac{1}{1+\varepsilon}$$

$$p = \frac{u_i'(x_i^{\star})}{1+\varepsilon} \tag{13}$$

Assuming that $1 + \varepsilon > 0$, the relation between marginal utility of the optimal purchasing volume of the product $x_i^{\star}$ and its market price *p* is

$$p < u_i'(x_i^{\star}) \tag{14}$$

This situation shows a favorable position of a consumer on the market of the studied product when market price *p* is lower than marginal utility $u_i'(x_i^{\star})$. In other words, in this specific case the market price is lower than the willingness of a consumer to pay the sum corresponding to his marginal utility. In this situation a consumer dedicates all his funds to purchase the studied product in the volume given by (12) and he doesn't purchase the other goods at all.

4 Conclusion

Based on the formalized analytical tools we showed that if a producer has enough market power to not only accept the market price but to be able to significantly influence and create it, he can quite effectively use his knowledge of consumer behavior to optimize a combination of supply and price of his product. As a matter of fact, it is a rational use of the information complex about the behavior of a consumer with specifically structured market basket, where they separately analyze consumer's utility regarding purchase of optimal volume of a price-differentiated product and this utility is represented in monetary units. Other goods in the market basket are being studied without any further specification of their volumes or range as one "aggregated good" and utility regarding purchase of these other goods is represented in monetary units as a simple sum of expenses spent on the purchase.

Significant is a fact that a company with a substantial market position in order to optimize its behavior at determining a combination of supply and differentiated price of a product, derives from a thorough analysis of consumers behavior while using analytical tools – demand functions and utility functions.

A mathematical programming problem which maximizes utility function of a consumer at budgetary restraints was examined in this article as it is a relevant tool for consumer behavior analysis. We showed that Kuhn-Tucker optimality conditions formulated for this optimization problem confirm the validity of consumer decision making schemes at optimization of his demand in the conditions of differentiated prices.

In a similar way we examined a monopoly profit maximization problem in the conditions of differentiated prices and we showed the fundamental schemes of price differentiation which monopoly can effectively use as a result of its market position to maximize its revenues as well as its profits from selling the products with differentiated price.

References

[1] Avriel, Mordecai: Nonlinear Programming - Analysis and Methods. Dover publications, Inc., New York 2003.
[2] Carlton, D.W.; Perloff, J. M. 2005. *Modern Industrial Organization*. Boston: Addison Wesley, 2005. ISBN 0-321-22341-1.
[3] Fendeková, Eleonora - Fendek, Michal. Podmienky optimálnosti Kuhna-Tuckera v modeloch rovnováhy trhu sieťových odvetví. In *Politická ekonomie : teorie, modelování, aplikace*. - Praha : Vysoká škola ekonomická, 2012. ISSN 0032-3233, 2012, roč. 60, č. 6, s. 801-821.
[4] Fendek, Michal. Modelling of reasonable profit in regulated industries. In *Quantitative methods in economics : multiple criteria decision making XIII : proceedings of the international conference*. - Bratislava : IURA EDITION, 2006. ISBN 80-8078-129-X, s. 16-24.
[5] Martin, S.: Advanced Industrial Econmics. Cambridge: Blackwell, 1996.
[6] Mikušová, N. Mikroekonomické modely teórie hier. In *Aktuálne pohľady na konkurencieschopnosť a podnikanie : [zborník vedeckých prác]*. - Bratislava : Vydavateľstvo EKONÓM, 2010. ISBN 978-80-225-3079-8, s. 145-151.
[7] O'Sullivan, A. - Sheffrin, S. - Perez , P.: Microeconomics: Principles, Applications, and Tools. New York: Prentice Hall, 2006.
[8] Pepall - Richards - Norman: Industrial Organization. Contemporary Theory and Practice. South western publ. 2002.ISBN 0-324-06772-0

Equilibrium analyses in allocation games

Petr Fiala[1]

Abstract. The paper considers a situation where a number of agents are connected in some network relationship. The applications are quite wide and varied. Game theory is a very powerful framework for studying decision making problems, involving a group of agents. Allocation games examine the allocation of value among agents connected by a network. Agents can be seen acting independently and willing to selfishly maximize their utility. Allocation games model the problem of maximizing network utility from the perspective of distributed non-cooperative agents. Many challenging questions arise when one seeks to design a network in such a way so as to optimize a given utility measure. Approaches for searching Nash equilibrium are presented. We provide results with respect to equilibrium existence, computation, convergence, efficiency and quality. Quality results are analyzed by the price of anarchy and the price of stability.

Keywords: networks, game theory, allocation, Nash equilibrium

JEL Classification: C44
AMS Classification: 90B50

1 Introduction

The allocation of tasks in a network of freely accessible shared resources often leads to congestion of single resources and in the worst case to a devaluation of the entire resource network utility. Two main solution approaches suggest themselves in this context:

- To achieve a load-balancing effect the unselfish distribution of tasks between the resources can be targeted by the agents in the sense of cooperative problem solving strategies (see [1]).
- To avoid excessive use of the resources, one can introduce self-interestedly acting economic agents that manage and own individual resources in the network (see [7]).

The use of cooperative agents in resource allocation games for distributed resources is not widely spread. The game resource allocation schemes are often subsumed under the description networking games. Cooperative games for resource sharing often employ Nash bargaining approach, where the bargainers negotiate for a fair contract point from the set of all feasible solutions (see [5]). The outcome is chosen based on a-priori defined fairness criteria: symmetry, Pareto optimality, and invariance with respect to utility transformations.

The concept of using self-interested agents to formulate allocation mechanisms in a game theoretical setting is closer to the classical market concept than solutions employing cooperative strategies. Network resource management is often carried out by using congestion-based pricing and routing capacity allocation. Most non-cooperative allocation strategies in distributed systems consist of following step:

- The formulation of utility functions for the system participants.
- The formulation of best response strategies.
- The existence of Nash equilibrium is proved in the system of multiple agents
- Efficiency is measured compared to achievable welfare.
- Fairness of the simulation result is checked.

In the majority of network allocation games, usually users were modeled as players. In the paper players are associated only with links in the network. It is assumed that network routers may be considered as autonomous entities, which operate independently. Only limited coordinating communication is allowed between them.

The central network resource allocation problem is the network utility maximization problem. Game theory is a very powerful framework for studying decision making problems. In the recent years a subfield known as algorithmic game theory has emerged, combining game theory and algorithms design. The problems deal with the use of algorithmic game theory include establishing the existence of Nash equilibria, designing computation-

[1] University of Economics, Department of Econometrics, W. Churchill Sq. 4, 13067 Prague 3, pfiala@vse.cz.

ally efficient procedures for determining players' strategies, and computing the price of anarchy and price of stability.

The rest of the paper is organized as follows. In Section 2, the problem formulation is provided. Approaches for computing strategies are presented in Section 3. Solution properties are discussed in Section 4. Section 5 presents conclusions.

2 Problem formulation

In the considered problem the network consists of a set of m links, each with capacity $c_i > 0$, $i = 1, 2, \ldots, m$. The state of the network is described by a vector $\mathbf{c} = (c_1, c_2, \ldots, c_m)$ of resource capacities. There are n flows (transmissions), defined by a routing matrix $\mathbf{A} = [a_{ij}]$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, where

$a_{ij} = 1$, if flow j traverses link i, and

$a_{ij} = 0$, otherwise.

The decision vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ represent partitioned allocation of capacity for each of the m flows. Each flow j is characterized by the transmission rate $x_j \geq 0$, and an associated utility measure $u_j(x_j)$, which is assumed to be strictly increasing concave and twice-differentiable function of transmission rate.

The network resource allocation problem is formulated as follows (see [3]):

$$u(\mathbf{x}) = \sum_{j=1}^{n} u_j(x_j) \tag{1}$$

subject to

$$\mathbf{Ax} \leq \mathbf{c}, \tag{2}$$

$$\mathbf{x} \geq \mathbf{0}, \tag{3}$$

In this problem the so-called isoelastic utility functions are used in the form

$$u_j(x_j) = w_j \ln x_j. \tag{4}$$

It was shown (see [4]) that such class of functions leads to proportionally fair allocations of transmission rates thus is typically employed in the analysis of network resource allocation problems.

The interaction between concurrent decision-making agents can be modeled as a network game. Each link in the network is associated with one player. Players must decide how to allocate their total capacities $c_i > 0$, $i = 1, 2, \ldots, m$, among the set of flows traversing their corresponding link. Each player makes a decision individually. The decision of player i, called player's strategy, is denoted $\mathbf{s}_i = (s_{i1}, s_{i2}, \ldots, s_{in})$, where s_{ij} is the fraction of link's capacity i allocated for flow j. The player's choice is restricted only to feasible decisions, satisfying

$$\sum_{j=1}^{n} a_{ij} s_{ij} \leq c_i \tag{5}$$

The transmission rate of a single flow is limited by the minimal allocation of some link along the path of that flow. The path is defined by the routing matrix $\mathbf{A}$. The player's payoff is computed as the value of weighted utility of transmission rates of all the flows passing through the corresponding link. The strategy of the game is defined as $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_1, \ldots, \mathbf{s}_m]$. The player's i payoff is given by:

$$v_i(\mathbf{S}) = \sum_{j=1}^{n} t_j a_{ij} u_j(\min_{k:a_{kj}=1} s_{kj}), \tag{6}$$

where $t_j \geq 0$ is a weight assigned to the flow j.

Two types of game's payoff matrices are considered:

1. A game with uniform payoffs, that is, all weights $t_j = 1$, $j = 1, 2, \ldots, n$.
2. A game with the set of weights defined by

$$t_j = \frac{1}{\sum_{k=1}^{m} a_{kj}}, \tag{7}$$

that is, the reciprocal of length of the path associated with flow j. This weight is equal for all links.

The social welfare is defined as

$$v(\mathbf{S}) = \sum_{i=1}^{m} v_i(\mathbf{S}), \tag{8}$$

that is the sum of all payoffs.

The total utility of the network, given as the objective of problem (1), corresponds to the social welfare with weights (7).

3 Computational procedures

Typical distributed methods of solving such network problems are Lagrangian relaxation-based methods for finding saddle point, or direct interior-point method decentralization. There are also algorithms for computing certain feasible strategies (see [2]). In this section two simple algorithms are presented.

Local allocation algorithm

Each player i, $i = 1, 2, \ldots, m$, solves a local concave optimization problem in the form:

$$\mathbf{s}_i^{(1)} = \arg\max_{\mathbf{s}_i \in D_i} \sum_{j=1}^{n} t_j a_{ij} u_j(s_{ij}), \tag{9}$$

where

$$D_i = \left\{ \mathbf{s}_i \middle| \sum_{j=1}^{n} a_{ij} s_{ij} \le c_i, \forall j, a_{ij} s_{ij} = s_{ij} \right\}, \tag{10}$$

For the class of utility functions (4), the solution can be derived analytically:

$$s_{ij} = a_{ij} c_j \frac{t_j w_j}{\sum_{l=1}^{n} t_l w_{jl} a_{il}} \tag{11}$$

This algorithm is a realization of the simplest rational strategy, which can be computed without any communication between players. Due to this fact, there are no synchronization issues concerning implementation in a networked environment.

Example 1

Consider two links ($m = 2$), the first with capacity $c_1 = 20$ and the second with capacity $c_2 = 80$. There are three flows ($n = 3$); first flow passes through both links, second flow single link 1 and third flow uses link 2. Thus, the routing matrix is

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

All utility functions are assumed to be in form

$$u_j(x_j) = \ln x_j.$$

The local algorithm computes the following strategy vectors for both players (links), $\mathbf{S}^{(1)} = (\mathbf{s}_1^{(1)}, \mathbf{s}_2^{(1)})$:

$$\mathbf{s}_1^{(1)} = (10, 10, 0),$$

$$\mathbf{s}_2^{(1)} = (40, 0, 40).$$

The payoffs of players (i.e. the local utilities of their corresponding flows) are:

$$v_1\left(\mathbf{S}^{(1)}\right) = \ln 10 + \ln 10,$$

$$v_2\left(\mathbf{S}^{(1)}\right) = \ln 10 + \ln 40.$$

If $\mathbf{S}^{(1)}$ were a Nash equilibrium, no player would have an incentive to unilaterally deviate from this allocation. However, since the capacities in both links are uneven, a fair allocation is suboptimal. The high-speed link with $c_2 = 100$ should promote the flow 3, as it does not pass through the bottleneck link with $c_1 = 10$. Thus the following change of player 2 strategy:

$$\mathbf{s}_2^{(1)} = (10, 0, 70)$$

gives a better outcome

$$v_2\left(\mathbf{S}^{(1)}\right) = \ln 10 + \ln 70.$$

Iterative allocation algorithm

The local allocation algorithm can be easily improved, if we allow players to interact in the following way. Initial allocations are computed with the use of the local algorithm. After these allocations are computed, all sources start sending data and transmission rates of all flows gradually increase from zero. The rate of a flow stops accelerating as soon as some link on the path becomes a bottleneck, i.e. the rate reaches minimal allocation of some link along its path. Such flows are called saturated. This means that it is no longer possible to increase its rate. However, other links on the path of such flow may have unused capacity. Thus it is possible to assign this capacity among the non-saturated flows, increase their rates and repeat that until all the flows become saturated.

Following additional notations are used:

- q iteration,
- $\mathbf{S}(q)$ the strategy vector computed in iteration q,
- $\mathbf{S}^{(2)}$ the final strategy vector returned by the algorithm,
- R_i the set of flows traversing link i,
- $R(q)$ all the flows that are not yet saturated in iteration q,
- $L(q)$ all the links that ran out of capacity in in iteration q,
- $i_0(q)$ the smallest index of link that runs out of capacity in iteration q.

The procedure of the iterative allocation algorithm can be summarized in following steps:

Step 1 (Initialization)

$$\text{Let } q = 1.\ R(q) = \{1, 2, \ldots, n\}.$$

$$\text{Execute the local allocation algorithm (solving problem (9)-(10)), } \mathbf{S}(q) = \mathbf{S}^{(1)}.$$

$$L(q) = \left\{ i : \sum_{j=1}^{n} a_{ij} \left(\min_{k:a_{kj}=1} s_{kj}^{(1)} \right) = c_i \right\}.$$

Step 2 (Next iteration)

$$\text{If } q < m, \text{ let } q = q + 1.$$

Step 3 (Saturated flows are removed)

$$R(q) = R(q-1) - R_{i_0(q-1)}$$

Step 4 (Strategy determination)

$$\mathbf{S}(q) = [\mathbf{s}_1(q), \mathbf{s}_2(q), \ldots, \mathbf{s}_m(q)].$$

$$\mathbf{s_i}(q) = \arg \max_{\mathbf{s}_i \in D_i(q)} \sum_{j \in R(q)} t_j a_{ij} u_j(s_{ij})$$

where

$$D_i(q) = \left\{ \mathbf{s}_i \,\middle|\, \sum_{j=1}^{n} a_{ij} s_{ij} \le c_i, \forall j, 0 \le s_{ij} \le a_{ij} s_{ij}, \forall j \in \bigcup_{p=1}^{q-1} R_{i_0(p)}, s_{ij} = \min_{k:a_{kj}=1} s_{ij}(q) \right\}.$$

Step 5 (Stopping rule)

If it holds $R(q) = \emptyset$ then terminate the algorithm and $\mathbf{S}^{(2)} = \mathbf{S}(q)$. Otherwise go to Step 2.

Example 2

Consider the same data as are given in Example 1 and the game with uniform payoff matrix. At the beginning of the execution of the iterative algorithm all flows are not saturated, $R(1) = \{1, 2, 3\}$, so each link allocates capacities maximizing its own objective, and the players' strategies are the same as in Example 1, i.e.:

$$\mathbf{s}_1(1) = (10, 10, 0),$$
$$\mathbf{s}_2(1) = (40, 0, 40).$$

Once all strategies are computed, the sources may transmit data at rates:

$$\mathbf{x} = (10, 10, 40).$$

Only first link becomes filled ($L(1) = \{1\}$), and the first and the second flows are saturated. There is unused capacity in the second link, which may be used in the next step.

For iteration $q = 2$, the first link does not change its strategy since it is filled. The second link sets allocations for saturated flows equal to their current transmission rates (i.e. minimal capacity allocated for these flows on their paths) and calculates the allocations for non-saturated flows (in this case for $j = 3$) to maximize its objective:

$$\mathbf{s}_2(2) = (10, 0, 70).$$

Since both stopping conditions are met in the next iteration, i.e. $q = 3 > m = 2$ and $R(2) = \emptyset$, the iteration algorithm stops with the following players' strategies, which constitute a Nash equilibrium:

$$\mathbf{s}_1^{(2)} = (10, 10, 0),$$

$$\mathbf{s}_2^{(2)} = (10, 0, 70).$$

4 Solution properties

First, the problem of determining Nash equilibria in the formulated network game is considered. The existence of Nash equilibrium for considered game can be concluded from the Rosen's theorem (see [6]), which states that a Nash equilibrium exists for a Nash game if the payoff function for each player is concave with respect to their own strategy and continuous with respect to the strategies of all players and the strategy set for each player is convex and compact.

It is however not obvious whether a given strategy is equilibrium, or how to compute one efficiently. The solution properties are summarized in theorems and corollaries, proofs are given in [2].

In general the local allocation algorithm does not produce a state of equilibrium; a player may be better off changing its allocation without informing the other players (see Example 1).

The strategy computed by iterative allocation algorithm dominates the strategy computed by local allocation algorithm. The iterative allocation algorithm provides even Nash equilibrium.

Theorem 1. *Strategy* $\mathbf{S}^{(2)}$, *provided by the iterative allocation algorithm, constitutes a pure Nash equilibrium.*

The result can be even strengthened, according to the construction of strategy $\mathbf{S}^{(2)}$.

Definition 1. *Strategy* $\boldsymbol{S^0}$ *is* ***non-dominated*** *(strongly Pareto-optimal) if there is no strategy* $\mathbf{S} \neq \mathbf{S^0}$ *such that:*

$v_i(\mathbf{S}) \geq v_i(\mathbf{S^0})$, $i = 1, 2, \ldots, m$, *and*

there exists such k that $v_k(\mathbf{S}) > v_k(\mathbf{S^0})$.

Theorem 2. *Strategy* $\mathbf{S}^{(2)}$ *is non-dominated* (*strongly Pareto-optimal*).

Second, a reversed problem of game design is considered: given an optimal solution of the network utility maximization problem, does it constitute Nash equilibrium of some game variant?

Theorem 3. *For the game with uniform payoffs* ($t_j = 1$, j =1, 2, …, *n*), *the global optimum of the network utility maximization problem (1)–(3) is a Nash equilibrium*.

There is a concept to measuring how efficient Nash equilibrium is in a specific game. The **price of stability** (*PoS*) is the ratio between the value of the best Nash equilibrium to the value of the optimal solution. The **price of anarchy** (*PoA*) is the ratio between the value of the worst Nash equilibrium to the value of the optimal solution.

Corollary 1. *For the game with uniform payoffs the price of stability PoS* = 1.

Theorem 4. *For the game with the set of weights* $t_j = \frac{1}{\sum_{k=1}^{m} a_{kj}}$, *the global optimum of problem* (1)–(3) *in general is not a Nash equilibrium.*

Theorem 5. *If there are no local* (*single-link*) *flows, i.e. for all j, j* = 1, 2, …, *n*, $\sum_{i=1}^{m} a_{ij} > 1$, *any strategy such that*

$$\forall i_1, i_2 \, \forall j\colon a_{i_1 j} = a_{i_2 j}=1,\ s_{i_1 j} = s_{i_2 j}$$

constitutes a pure Nash equilibrium of the considered game.

Remark 1. *Although the strategies characterized in Theorem 5 are points of equilibrium of the considered game, they can be arbitrarily bad in terms of the players' outcomes; the price of anarchy PoA is unbounded.*

5 Conclusions

The network resource allocation problem was formulated. The problem is applicable in several practical fields. The interaction between concurrent decision-making agents can be modeled as network games. Two types of game's payoff matrices are considered: a game with uniform payoffs and a game with the set of weights.

Typical distributed methods of solving such network problems are Lagrangian relaxation-based methods for finding saddle point, or direct interior-point method decentralization. In the paper two simple algorithms are presented and illustrative examples are solved. The algorithms were tested on simulated examples with very good results. Properties of game solutions are discussed. Interesting results for both types of games were presented.

The approach seems to be useful and promising for next research. There are some possible extensions of the approach and some areas for further research. Other types of games can be analyzed.

Acknowledgements

The research project was supported by the grant No. 13-07350S of the Grant Agency of the Czech Republic and by Grant No. IGA F4/19/2013, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Durfee, E. H.: Distributed problem solving and planning. In: *Multiagent Systems*. MIT Press, Cambridge, 1999.

[2] Gasior, D., and Drwal, M.: Pareto-optimal Nash equilibrium in capacity allocation game for self-managed networks. (submitted to journal 2012)

[3] Kelly, F. P., Maulloo, A. K., and Tan, D. K. H.: Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society* **49**(1998), 237–252.

[4] Mo, J., and Walrand, J.: Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking* **8** (2000), 556–567.

[5] Nash, J F.: The bargaining problem. *Econometrica* **18** (1950), 155–162.

[6] Rosen, J.: Existence and uniqueness of the equilibrium points for concave n-person games. *Econometrica* **33** (1965), 520–534.

[7] Wellman, M. P.: Market-oriented programming: Some early lessons. In: Clearwater, S.H. (ed.): *Market-based Control: A Paradigm for Distributed Resource Allocation.* World Scientific Publishing, Singapore, 1995.

Quantitative Evaluation of the EU´s Export Potential to Russia Using the Gravity Model

Lenka Fojtíková[1]

Abstract. The European Union (EU) is the largest trading power in the world. Although its share in the world exports of manufactured goods is gradually declining due to various factors, it still occupies the first place in both exporting industrial goods and commercial services. The EU's share in world merchandise exports is about 15 % and exports of commercial services almost 25 % (WTO, 2011). However, the EU not only exports but also imports some commodities from third countries and its trade balance is negative in the long term. The negative trade balance is created in the import of raw materials and energy such as oil and natural gas. These minerals are imported into the EU mainly from Russia, which therefore occupies the second position among the EU main importers. The dependence of the EU-27 on imports of energy resources from Russia varies in the individual Member States.
The aim of this paper is to evaluate the untapped export potential of the EU to Russia by the member countries in 2000-2011 and to propose appropriate measures to improve export in these countries. The quantitative assessment of the export potential of the member countries of the EU to Russia will be done through a gravity model, which is very often used for the purposes of analysis in this area. Removing the unused export potential to Russia would help to achieve trade balance throughout the EU.

Keywords: foreign trade, export potential, trade balance, gravity model

JEL Classification: C23, F13, F14, F15, F41
AMS Classification: 62M10

1 Introduction

The European Union (EU) is the largest world exporter in the area of merchandise as well as commercial services trade. The larger portion of total EU trade is carried out in the area of merchandise trade (the value of 3.2 billion euros in 2011) than in commercial services trade (1.1 billion euros in 2011) [8]. But to maintain the first position among the leading world traders in the future it is necessary to find new ways leading to the growth of the EU's competitiveness and to strengthen trade relations with fast-growing countries such as Brazil, Russia, India, China and others. The growing incomes in these countries contributed to the creation of a new middle class that is able to consume not only more domestic products, but also products that are imported from other countries, including the EU. But the EU needs not only exports but also imports. From the long term point of view, the largest share of the EU imports belongs to mineral fuels, lubricant and related materials (SITC 3). In 2012, the share of mineral fuels on the total EU imports reached 30.5 % and another 4.5 % belonged to imports of raw materials (SITC 2+4). It means that industrial production in the EU is highly dependent on imports from other countries. The dependence of the EU on energy products has a negative influence on the EU trade balance as well as on the EU's competitiveness in world trade.

The EU has had the largest volume of energy imports with Russia, which is the EU's most important near neighbor. The EU is the first customer of the main Russian export – energy. 80 % of all Russian oil exports, 70 % of all Russian gas exports and 50 % of all Russian coal exports go to the EU. While the EU is by far Russia's biggest overall trade partner, Russia is the European Union's third largest trade partner. In 2012 alone the total volume of trade between the EU and Russia reached 336 billion euros and around 75 % of foreign direct investment in Russia is of European origin [6]. But in the long run, the EU has recorded a trade deficit in merchandise trade with Russia that reached 90.1 billion euros in 2012 [6,7]. Conversely, the EU's trade in services with Russia was in surpluses in the previous years. When the EU is highly dependent on energy imports from Russia, it is necessary to find untapped possibilities in the EU exports to Russia.

The aim of this paper is to evaluate untapped export potential to Russia by the EU's member countries in 2000-2011 using the gravity model. Firstly, we will introduce the theoretical background of the gravity model

[1] VSB-Technical University Ostrava, Faculty of Economics, Department of European Integration, Sokolská 33, Ostrava 1, 701 21, lenka.fojtikova@vsb.cz.

and its application to the EU export flows to Russia. Particular attention will be paid to our empirical findings about real and untapped export potential of the individual member countries of the European Union. Afterwards, the trade balance of the individual member countries with Russia will be presented in the period of 2000-2011. There is a hypothetical assumption that countries that registered an untapped trade potential also achieved a trade deficit with Russia. Finally, the institutional framework for increasing trade cooperation between the EU and Russia will be introduced.

2 Methodology and data

Section 2 concentrates on providing the theoretical foundations of the gravity model, followed by a presentation of the methodology and the data used in the empirical analysis.

2.1 Theoretical development and model identification

In order to find potential export flows of the EU to Russia, the gravity model of trade will be used. Gravity equations are one of the most popular tools in empirical studies in the area of international trade. Gravity models utilize the gravitational force concept that is based on Newtonian physics, as an analogy to explain the volume of trade, capital flows, and migration among the countries of the world. In the area of international trade, the gravity model can be applied to four categories of issues: estimating the cost of the border, explaining trade patterns, identifying effects related to regionalism and, finally, calculating trade potential.

The model was first used by Tinbergen (1962) and Linneman (1966). Although the gravity model was empirically successful in explaining trade flows and the prediction of trade potential, the model lacks strong theoretical foundations. In response to this criticism several studies were written to explain theoretical derivation. Anderson was the first who gave a theoretical explanation based on economic theory. Anderson helped to explain the presence of income variables in the gravity model by specifying demand. According to Anderson, products traded internationally are differentiated by the country of origin (the so called "Armington assumption"). This approach was also adopted by Bergstrand, who specified the supply side of economy more thoroughly. J. H. Bergstrand [3] presents empirical evidence that the gravity equation is a reduced form of a partial equilibrium subsystem of a general equilibrium model with nationally differential products. Deardorf [4] derived a model using the Hecker-Ohlin model. This theory predicts that trade patterns will be based on relative factor abundance. Those countries with a relative abundance of one factor are expected to produce goods that require a relatively large amount of that factor in their production. While the Hecker-Ohlin model and Ricardian's comparative advantage theories are accepted trade theories, they have suffered empirical problems. An alternative theory, first proposed by Staffan Linder, predicts that patterns of trade will be determined by the aggregated preferences of goods within countries. Those countries with similar preferences are expected to develop similar industries and trade in differential but similar industries. But the empirical validity of the Linder hypothesis is unclear. Several studies have found a significant impact of the Linder effect, but others have had weaker results. The monopolistic competition model of a new trade theory is another approach that provides theoretical foundations to the gravity model. In this model, the product differentiation by the country of origin approach is replaced by product differentiation among product firms, and the gravity model serves as an explanation of intra-industry trade. Anderson and Wincoop [1] have also contributed to the theoretical development of the gravity model with the method "the border puzzle". They find that borders reduce bilateral national trade levels by plausible though substantial magnitudes. On the whole, the different theories underlying the gravity model lead to different trade policy implications.

The econometric estimation of gravity equations is connected with using different variables. Tinbergen's original gravity model has been used to estimate the volume of bilateral trade as a function of two main components – the economic size of countries and the distance between them. The traditional gravity model takes a log-linear model form (1):

$$\ln(FT_{ij}) = \beta_0 + \beta_1 ln(GDP_i) + \beta_2 \ln(GDP_j) - \beta_3 \ln(D_{ij}) + \varepsilon_{ij} \quad (1)$$

where FT_{ij} is the volume of trade from country $_i$ to country $_j$, GDP_i and GDP_j present the economic size of countries $_{i,j}$, $D_{i,j}$ denotes the distance between the two countries and $\varepsilon_{i,j}$ is the random component with an expectation equal to 1.

The following empirical studies also included other explanatory variables. Kepaptsoglou et al. [11] presents almost sixty studies that were written during a 10-year period (1999-2009), in which the dependent variables are population, area size, GDP per capita, dummy variables denoting borders, islands, former colonies, membership in a free trade area or customs unions, currency unions, etc. The main dependent variables are bilateral trade flows, exports, imports and foreign direct investments. Kovářová [10] used the gravity model to verify the

Linder hypothesis of foreign trade between the USA and the EU. She includes variables in the gravity model, such as the GDP of the considered countries, the geographical distance between the countries and the absolute value of per capita GDP differential between the two countries. The gravity model has also been recently used to assess the effect of the WTO on multilateral trade patterns. For example, Rose, Subramanian and Wei evaluated the magnitude of the WTO dummy after checking for the customary gravity effects. Lissovolik et al. [12] used the gravity model to evaluate Russia's trade determinants and patterns, with particular reference to the role of the WTO. Benedictis and Salvatici [2] encompass the trade preferences as another component of the gravity model and estimate the effect of trade preferences on the EU trade flows. Despite the fact that the trade data are easily available, they struggle on the problem of missing data when evaluating the impact of the EU's preferential trade policies on developing countries' trade flows. But in many cases, gravity models have significant explanatory power, leading Deardorff [4] to refer to them as a "fact of life".

By defining the gravity model in this way, it is possible to solve the problem cross-sectionally or by using panel data. In the case of cross-sectional analysis, the data for the individual year are used. Panel data analyses enable us to explore the development of international trade among countries in a period of several years. Panel data are used in most studies for periods of at least 5 years. There are generally three types of panel data regression analyses: independently pooled panels, fixed effect models and models with random effect.

2.2 The gravity model of the EU export potential to Russia

The calculation of export potential between the individual EU member states and Russia based on the gravity model firstly supposes the arrangement of the EU member states in alphabetical order, and bilateral export flows with Russia within the sample are then estimated. Such simulated bilateral exports will be compared with the observed ones in order to infer bilateral export potentials. This methodology will be applied at the aggregate level.

The dependent variable is the natural logarithm of the EU´s exports in current USD calculated individually for 27 member countries. The explanatory variables in the model are the natural logarithm of the gross domestic product of the EU member states and Russia measured in current USD, the natural logarithm of the distance between the capitals measured in kilometers. The distance between the EU member states with Russia is measured using the great circle formula, which takes into account the longitude and latitude of the capitals. [13] For each scheme, we estimate the following log-linear equation (2):

$$\ln(EX_{EUMSRt}) = \beta_0 + \beta_1 \ln(GDP_{EUMSt}) + \beta_2 ln(GDP_{Rt}) - \beta_3 D_{EUMSR} + \varepsilon_{EUMSRt} \quad (2)$$

where EX_{EUMSRt} are exports from the EU member state to Russia in period *t*, β_0 are specific effects associated with each bilateral flow, GDP_{EUMSt} indicates the GDP of the EU member countries in period *t*, GDP_{Rt} denotes the GDP of Russia in period *t*, D_{EUMSRt} is the great circle distance between the EU member countries (*EUMS)* and Russia (*R)* and ε_{EUMSRt} is random effect.

Panel data were used for periods 2000-2011. Elasticities were estimated by means of an Ordinary Least Squares (OLS) cross-country regression on 324 observations for 27 countries. All data were obtained from the database of the United Nation Conference on Trade and Development (UNCTAD).

3 Empirical results

The cross-country OLS regression results for the gravity equations (2) are reported in table 1. The overall performance of the model seems to be surprisingly good, with high R^2 value around 0.85. All explanatory variables are found to be highly significant, indicating that the gravity model is appropriate and effective in explaining the EU member countries export flows to Russia.

Variable	Coefficient	t-Statistic	Probability
$\ln(GDP_{EUMSt})$	1.137610	39.18736	0.0000
$\ln(GDP_{Rt})$	0.446382	13.80452	0.0000
$\ln(D_{EUMSRt})$	-2.815240	-26.12049	0.0000

Table 1 Estimated coefficients
Source: self-elaboration using EViews 7

Let us first look at the coefficient on the export structure variable to identify the underlying trade model of the EU member states export flows to Russia. The presented results confirm that bilateral export flows of the individual EU member countries to Russia positively depend on their economic size measured by GDP and

negatively depend on their distance measured by the distance between their capitals. The coefficient β_1 shows a positive value with a high statistical significance. The estimated coefficient shows that, holding other variables constant, a 1 percentage point increase in the GDP of the EU member states will result in a roughly 1.14 percentage point increase in the EU member countries export flows to Russia. The coefficient β_2 also indicates a positive value. When the GDP of Russia increases by 1 percentage point, it causes a roughly 0.45 percentage point increase of the export flows to Russia from the EU member countries. On the contrary, the coefficient β_3 shows a negative value. When the distance increases by 1 percentage point, then the export flows of the EU member countries to Russia decrease by 2.82 percentage points. The estimated coefficients β_1, β_2 and β_3 are consistent with the economic theory. The variable of GDP expresses the economic size of the two countries, in terms of both the production capacity and the size of market. Larger countries, with a greater production capacity, are more likely to achieve economies of scale and increase their exports based on their comparative advantage. They also possess larger domestic markets that are able to absorb more imports. From this point of view, an increase of the GDP exporter as well as importer has a positive impact on bilateral trade among countries. However, in most gravity models the GDP of the exporter has a higher significance for bilateral trade than the GDP of the importer. On the other hand, the distance variable has a negative influence on bilateral trade flows. A decrease in the distance variable indicates that export as well as import with geographically distant countries increases relative to trade with geographically closer countries, whereas an increase indicates that trade with closer countries increases faster than that with distant countries.

In order to find the export potential of the individual member countries, the export potential was compared with the actual export volume. The export potential is simply the predicted export volume from the gravity estimation (2). The difference between the predicted and actual export flows can be interpreted as an untapped export potential. Table 2 shows that the export potential of the EU member countries to Russia ranged from 90.3 % to 105.8 % in 2011. The fifteen EU countries reached a value of more than 100 % and the twelve EU countries recorded an export potential below 100 %. While Slovenia reached the highest use export potential (105.8 %), Malta reached the highest untapped export potential to Russia (90.3 %). Bearing in mind the fact that all the EU member countries use the Common commercial policy that is applied to non EU member countries, the cause of the untapped export potential of the twelve EU member countries cannot be explained by trade barriers, but by the economic structure and comparative advantages that these countries have in the given commodity items.

Country / Code	**Actual export flows (Ex_r)**	**Predicted export flows (Ex_p)**	**Ex_r/Ex_p (%)**
Austria / AT	22,28	22,14	100,61
Belgium / BEL	22,59	21,54	104,88
Bulgaria / BGR	20,39	19,64	103,84
Cyprus / CYP	16,69	18,02	92,63
Czech Republic / CZE	22,24	21,40	103,90
Denmark / DNK	21,41	22,07	97,01
Estonia / EST	21,33	20,66	103,20
Finland / FIN	22,71	23,40	97,05
France / FRA	23,06	23,18	99,44
Germany / DEU	24,62	24,70	99,67
Greece / GRC	20,12	20,95	96,01
Hungary / HUN	21,97	21,08	104,23
Ireland / IRL	20,38	19,97	102,01
Italy / ITA	23,28	23,04	101,01
Latvia / LVA	20,96	21,02	99,70
Lithuania / LTU	22,26	21,66	102,72
Luxembourg / LUX	19,23	19,15	100,44
Malta / MLT	14,72	16,31	90,30
Netherlands / NLD	23,01	22,23	103,51
Poland / POL	22,86	23,43	97,56
Portugal / PRT	19,62	19,12	102,58
Romania / ROU	21,07	21,54	97,82
Slovakia / SVK	21,59	20,55	105,07
Slovenia / SVN	20,44	19,32	105,76
Spain / ESP	21,97	21,55	101,96
Sweden / SWE	22,13	23,29	95,00
United Kingdom / GBR	22,75	23,02	98,85

Table2 The rate utilization of export potential of the EU member countries to Russia in 2011 (%)
Source: own calculations according to data from the UNCTAD Database [13]

4 Trade balance of the EU member states with Russia

The EU has recorded a deficit in merchandise trade with Russia in the long term. It is caused by the dependence of the EU on energy imports that represent the main commodity item of the EU imports from Russia. Table 3 records the trade balance of the individual EU member states with Russia in the period of 2000-2011. The trade balance (export minus import) was calculated from the data obtained from the UNCTAD Database (UnctadStat). The data includes international trade in goods and services and is published in U.S. dollars (USD). Most of the countries recorded a trade deficit with Russia on a different level. The Netherlands, Poland, France, Spain and the United Kingdom reached the highest level of trade deficit in the individual years. Trade surpluses reached in the given years by some member countries are displayed in the grey fields in Table 3. Ireland, Denmark, Luxembourg and Slovenia belong to the countries that reached trade surpluses with Russia most often. Positive changes in trade balance with Russia have been recorded in Austria since 2007. The development of trade balance with Russia was variable especially in Germany. While in 2011 Germany recorded the largest trade surplus among the other EU countries in the amount of 7 950.1 million dollars, in 2001 it reached the largest trade deficit with Russia in the amount of 7 272.4 million dollars. On the whole, the variable level of the trade deficit of EU-27 with Russia in the reported period was influenced by currency fluctuations and changes in commodity prices, especially energy products on the world market. The analysis of the causes of the trade deficit in the individual member countries of the EU was not the object of this paper and presents another area for empirical research.

The hypothesis that has been set in the introduction of this paper was not completely confirmed. Contrary to the results of the trade analysis shown on the untapped export potential in Denmark and Germany in 2011, these countries recorded trade surpluses in merchandise trade with Russia in the same year. On the other hand, the hypothesis was confirmed in Cyprus, Finland, France, Greece, Latvia, Malta, Poland, Romania, Sweden and the United Kingdom. These countries recorded a trade deficit with Russia as well as an untapped export potential to Russia in 2011.

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
AT	-0.548	-0.184	-0.070	-0.213	0.017	-0.697	-0.208	1.008	0.693	0.528	0.304	0.729
BEL	-0.540	-0.615	-0.559	-1.076	-1.880	-2.680	-2.409	-2.185	-2.608	-2.409	-3.685	-5.883
BGR	-1.062	-0.936	-0.831	-0.957	-1.363	-2.316	-2.536	-3.842	-5.007	-2.450	-3.280	-4.036
CYP	-0.104	-0.069	-0.131	-0.129	-0.064	-0.035	-0.044	-0.041	-0.026	-0.016	-0.047	-0.021
CZE	-1.575	-1.365	-1.564	-1.565	-1.234	-2.974	-3.374	-2.053	-4.229	-2.004	-2.466	-1.506
DNK	0.97	0.288	0.208	0.139	-0.060	-0.230	0.002	0.402	0.057	0.642	0.467	0.111
EST	-0.852	-0.775	-0.947	-0.828	-0.926	-1.164	-1.706	-0.867	-0.003	-2.081	-0.051	-0.157
FIN	-1.227	-0.600	-0.477	-1.053	-1.272	-1.054	-2.030	-2.317	-3.769	-4.179	-6.029	-7.950
FRA	-2.714	-2.490	-2.733	-3.772	-5.211	-5.885	-6.666	-7.387	-9.795	-5.869	-7.817	-8.996
DEU	-7.272	0.318	2.043	-1.411	-1.597	-6.265	-8.217	-0.822	-5.261	-6.216	-7.179	7.628
GRC	-0.836	-1.293	-2.004	-2.344	-2.537	-3.878	-3.939	-3.802	-5.939	-3.124	-5.934	-5.227
HUN	-2.171	-1.926	-1.807	-2.329	-2.535	-3.704	-4.165	-3.621	-6.226	-2.780	-3.471	-5.375
IRL	0.152	0.194	0.210	0.234	0.190	0.201	0.212	0.341	0.309	0.224	0.283	0.542
ITA	-2.521	-1.476	-1.141	-1.447	-2.028	-2.253	-0.443	0.294	2.208	-7.969	-8.935	-12.162
LVA	-0.293	-0.205	-0.222	-0.300	-0.395	0.341	-0.372	-0.521	-0.753	-0.373	-0.178	-0.046
LTU	-1.222	-1.102	-0.972	-1.435	-1.990	-2.948	-2.893	-1.822	-5.607	-3.293	-4.397	-5.692
LUX	-0.011	7.920	-1.184	0.020	-0.004	0.018	2.569	0.033	0.209	0.161	0.158	0.212
MLT	-0.035	-0.023	-0.064	-0.041	-0.044	-0.051	-0.139	-0.156	-0.298	-0.119	-0.366	-1.563
NLD	-2.242	-2.191	-3.853	-4.989	-7.852	-14.979	-21.433	-23.826	-32.764	-20.935	-32.232	-25.867
POL	-3.774	-3.382	-3.094	-3.722	-3.548	-5.025	-7.434	-7.921	-11.628	-7.819	-11.583	-16.985
PRT	-0.117	-0.277	-0.200	-0.300	-0.554	-0.883	-0.931	-0.946	-0.096	-0.257	-0.107	-0.138
ROU	-1.032	-1.101	-1.246	-1.929	-2.108	-3.108	-3.648	-3.834	-4.053	-1.378	-1.605	-1.510
SVK	-2.069	-2.050	-1.940	-2.166	-2.354	-3.158	-4.306	-4.008	-4.787	-2.867	-3.780	-6.226
SVN	-0.038	-0.156	0.051	0.026	0.119	0.143	0.268	0.225	0.595	0.401	0.257	0.144
ESP	-1.711	-1.180	-1.815	-2.350	-3.416	-5.086	-7.424	-7.960	-6.885	-4.326	-5.474	-7.845
SWE	0.006	0.491	0.124	0.105	-0.339	-1.163	-1.942	-1.061	-2.391	-2.269	-4.438	-5.630
GBR	-1.884	-2.151	-1.917	-1.909	-4.294	-5.538	-6.121	-4.930	-4.638	-3.038	-2.066	-3.453
EU27	-35.594	-24.092	-24.953	-35.740	-47.280	-75.053	-91.893	-81.617	- 112.690	-81.735	- 113.652	- 115.341

Table 3 Trade balance of the EU´s member states with Russia in 2000-2011 (bil. USD)
Source: own calculations according to data from the UNCTAD Database [13]

5 Conclusion

Taking into account the results of the empirical analysis, there is untapped export potential to Russia in 12 member states of the European Union. Ten of these countries also recorded a trade deficit with Russia in 2011. On the

whole, the EU recorded a trade deficit in the area of merchandise trade throughout the period (2000-2011). Modernization is a strategic objective of today's Russia and it should be a challenge for the EU member countries to use this opportunity as Russian's main import partners. A key step for increasing the EU export growth to Russia is the creation of a proper institutional framework. Political and economic relations between the EU and Russia have been carried out under the Partnership and Cooperation Agreement since 1997. Now, a new EU-Russia Agreement is being negotiated. An ambitious and comprehensive new agreement, which includes a developed regulatory framework with common standards and norms, trade and energy provisions, would help to create wider cooperative approaches with win-win situations. Russia, who joined the WTO in 2012, should also help to lower the tariffs, to remove non-tariff obstacles of trade and make access to the Russian market easier for the EU exporters. Russia participates in EU research and development programmes. The proposal of Jose M. Barosso, which was brought up at the Moscow conference in 2013, to establish a European Union-Russia Strategic Partnership in Research and Innovation is a very important step forward in deepening the EU-Russia trade relations, too. The long-term vision of the EU is a common economic and human space from Lisbon to Vladivostok with free travel of people, free exchange of goods and services and very close overall cooperation [5].

Acknowledgements

This paper was created under the SGS project (SP2013/122) of Faculty of Economics, VSB - Technical University of Ostrava.

References

[1] Anderson, J. E. and Wincoop, E.: *Gravity with Gravitas: A Solution to the Border Puzzle*. [Online], [cit. 2013-05-05]. Available: https://www2.bc.edu/~anderson/BorderEffects.pdf

[2] Benedictis, L. and Salvatici, L: T*he Trade Impact of European Union Preferential Policies. An Analysis Through Gravity Models.* Springer, 2011.

[3] Bergstrand, J. H.: The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence. *Review of Economics and Statistics*. Vol. 67, No. 3, 1985, 474-481.

[4] Deardorf, A., V.: *Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?* [Online], [cit. 2013-05-02]. Available: http://www.nber.org/chapters/c7818.pdf

[5] Europa.eu: *Speech by President Barosso at the Russia-European Union – Potential for Partnership conference: "Moving into a Partnership of Choice".* Moscow, 21 March 2013.Available: http://europa.eu/rapid/press-release_SPEECH-13-249_en.htm

[6] European Commission. Directorat-General for Trade. *Statistics. EU Bilateral Trade and Trade with the World: Russia*. [Online], [cit. 2013-05-02]. Available: http://ec.europa.eu/trade/policy/countries-and-regions/statistics/

[7] Eurostat. *Statistics*. *International Trade*. *Main tables*. [Online], [cit. 2013-05-01]. Available: http://epp.eurostat.ec.europa.eu/portal/page/portal/international_trade/data/main_tables

[8] Eurostat. *Statistics. Economy and finance. Balance of Payments. Main tables*. [Online], [cit. 2013-05-01]. Available:http://epp.eurostat.ec.europa.eu/portal/page/portal/balance_of_payments/data/main_tables

[9] Head, K.: *Gravity for Beginners*. [Online], [cit. 2013-05-07]. Available: http://pioneer.netserv.chula.ac.th/~kkornkar/inter%20trade%20undergrad/term%20paper/gravity%202.pdf

[10] Kovářová, J.: Verification of the Linder hypothesis of the foreign trade between the US and EU. In: Jedlička, Pavel (ed.). *Sborník příspěvků z mezinárodní vědecké konference Hradecké ekonomické dny* (část III). Hradec Králové: Univerzita Hradec Králové, 2013, s. 192-197.

[11] Kepaptsoglou, K. et al.: The Gravity Model Specification for Modeling International Trade Flows and Free Trade Agreement Effects: A 10-Year Review of Empirical Studies. *The Open Economics Journal*, 2010, Vol. 3, 1-13.

[12] Lissovolik, B. and Lissovolik Y.: *Russia and the WTO: The Gravity of Outsider Status*. International Monetary Fund, 2004. Working paper 04/159.

[13] UNCTAD. UnctadStat. [Online], [cit. 2013-05-08]. Available: http://unctadstat.unctad.org/ReportFolders/reportFolders.aspx

Output gap and structural shock estimates for Czech Republic and other selected EU members

Tomáš Formánek[1], Roman Hušek[2]

Abstract. Our paper is focused on different theoretical and practical aspects of estimation of key unobservable macroeconomic variables: the output gap, potential output and supply and demand shocks for Czech Republic and other selected EU countries. We use, evaluate and compare different methods for output gap estimation. Hodrick-Prescott (HP) filtering is used for baseline estimates. We compare those baseline results with improved alternative estimates from other sources: a modified Phillips-curve model, a simple Beveridge-Nelson unobserved components method and the Blanchard-Quah decomposition (based on macroeconomic structural VAR models). Results obtained for data from Czech Republic are compared to other selected EU economies and the overall inter-country homogeneity of macroeconomic behavior and internal (unobservable) dynamics of individual economies are evaluated and compared to each other. Overall, we conclude that prominent heterogeneities exist in the macroeconomic internal dynamics of individual countries as measured and evaluated in this contribution, with potentially negative impacts on the formation and evaluation of economic policy actions. Our results are relatively robust, as they are based on diverse econometric models and estimation methods.

Keywords: output gap, potential product, structural shocks, VAR model.

JEL Classification: C32, C52, E23
AMS Classification: 91B55

1 Introduction

In this paper, we present multiple approaches to consistent estimation of selected key unobservable macroeconomic variables across different economies, with appropriately devised subsequent evaluation of homogeneity across individual countries. The expected (estimated) behavior of both observable and unobservable variables is a key econometric tool for macroeconomic analysis. Namely, this approach may be conveniently used for the evaluation of expected costs and benefits of potential monetary policy actions under consideration by the European Central Bank (ECB). Also, should there be any EU-wide consensus on coordinating individual fiscal policies of the member states, the consistent inter-country comparison techniques presented in our contribution would only gain importance. Although we focus on the estimation and comparison of unobservable data series, our approach is relatively conservative as the estimates of "unobservables" are mainly based on observed data with only a few relatively simple and theoretically well-established (identifying) restrictions. Specifically, in this contribution we do not adopt the DSGE methodology as shown and referenced to in [8].

This article is structured as follows: Next section describes key aspects and general methodology used for estimation of the unobservable variables and the basic principles and constraints for subsequent inter-country analysis of homogeneity, as well as some basic data handling topics. The third chapter contains detailed description of the models used for estimation along with their theoretical background and the empirical results of our contribution. Those results are organized in tables with pair wise correlation coefficients among the estimated unobservable variables for different countries. Evaluation of the results and complementary comments are also provided. Last section and the list of references conclude our contribution.

2 Methodology and data

We aim to estimate and compare a set of relatively small, yet theoretically sound and practical macroeconometric models with intuitively comprehensible estimation outputs, suitable for subsequent comparison of the results among all economies included in our research. This approach has the advantage of providing the reader with multiple shots at some of the most important directly unobservable macroeconomic variables. Individual models contained in this contribution are based on diverse theoretical backgrounds, approaches and estimation methods.

[1] University of Economics, Prague, Dept. of Econometrics, nám. W. Churchilla 4, Praha 3, formanek@vse.cz.
[2] University of Economics, Prague, Dept. of Econometrics, nám. W. Churchilla 4, Praha 3, husek@vse.cz.

Although some of the estimation methods are compatible and generate consistent results, others differ in theoretical framework aspects that may limit our ability to directly compare the results. On the other hand, subsequent homogeneity evaluation based on pair wise correlation analyses may be easily compared as described throughout next chapter. Moreover, many theoretical pitfalls and practical errors may be avoided if fiscal and monetary policy decisions are exercised upon complex evaluation involving *multiple perspectives* rather than based on a single criterion, regardless of its potential sophistication and complexity.

We use quarterly data (2004:Q1 – 2012:Q4) for the following countries: Austria, Czech Republic, Denmark, Estonia, Germany, Hungary, Poland, Slovakia and Slovenia for estimation of all our models, except for specification (3) where we took advantage of available monthly series for the period 2004 to 2012. Our data originate from the International Monetary Fund's IFS database (http://elibrary-data.imf.org/). We use real GDP, inflation (CPI and PPI) and unemployment. All variables were de-seasoned by the authors, unless a series was published by the IMF as already seasonally adjusted. Due to data availability and model significance issues, we had to drop some individual economies from certain model estimations (details may be observed by comparing tables 1, 2 and 3). Fortunately, such individual drop-outs do not hamper significantly with our subsequent comparison and evaluation. EViews6 software was used for all estimations and most of the data handling.

The output gap definition (1) used throughout this paper follows e.g. from [3]. Although this is not a sole specification, it's fairly common to define output gap y_{gap} as the relative deviation of output y from its estimated potential value y^* expressed in percentage points:

$$y_{gap,t} = 100(y_t - y_t^*)/y_t^*. \tag{1}$$

In our paper, we use real GDP base index (2005 = 100%) to measure output. Inflation and unemployment gaps are also constructed and interpreted using the concept defined by equation (1), with the exception of an Expectations Augmented Phillips Curve (EAPC) model described by equation (3) where we closely adhere to the textbook specification used in [10]. Data series based on specification (1) are directly comparable among different economies, avoiding the need for currency translation or additional normalization for the size of individual economies. Apart from model (4), we approximate all potential values y^* using Hodrick and Prescott [5] HP-filter, which is a relatively flexible and widely adopted approach. Technically, we associate the HP-trend component of y with y^*, using (minimizing) the HP-filter expression (where λ determines the smoothness of the trend component):

$$\left(\sum_{t=0}^{T} (y_t - y_t^*)^2 + \lambda \sum_{t=2}^{T-1} \left[(y_{t+1}^* - y_t^*) - (y_t^* - y_{t-1}^*) \right]^2 \right) \rightarrow \min \tag{2}$$

Due to space limitations, in this contribution we only follow-up on data obtained by using the usual λ values (λ=1,600 for quarterly data and λ=14,400 for monthly data). However, we have been able to simulate significantly diverse levels of smoothness in potential product time series by arbitrarily varying the λ parameter, with significant impact on subsequent comparison among individual economies. Generally, higher values of λ put additional weight on smoothing of the trend component and therefore they increase fluctuations in gap values calculated from equations (1) and (2). On the other hand, low λ values may result in spurious cycles in the trend series. While focusing only on the economy of Sweden, Cerra and Saxena [3] provide examples for potential output and output gap using the HP-filter with λ set to 100, 200 and 1,000.

We adopt the approach described by equations (1) and (2) to generate basic output gap estimates, as shown in figure 1. Those estimates may be compared with alternative and arguably more sophisticated results from other models. Again, due to space limitations we only show results obtained from three models, chosen for this contribution so that different theoretical bases and estimation (econometric) approaches may be demonstrated. Alternative methods and models are provided and referenced to in [12], among others. Our first model is a single equation EAPC model based on theory detailed in [10], where the estimated residuals for a specific economy are interpreted as supply shocks. As a second approach, the Beveridge and Nelson [1] BN-decomposition is used to estimate (calculate) the "trend and cycle" of a series, i.e. to decompose an integrated time series into a permanent component (trend) and transitory shocks (cycles). Last but not least, two-variable real GDP and PPI inflation VAR(2) models are estimated for individual economies. Upon applying long-term structural identifying restrictions, the Blanchard and Quah [2] BQ-decomposition is performed. This relatively complex procedure described in detail by [2], [4] and [7] allows us to decompose the estimated VAR model residuals into supply and demand shocks that may be used for further analysis and comparison. Other econometric models for output-gap or supply and demand shock estimation assembled by the authors are either omitted from this contribution or just briefly mentioned. Also, we only have space to show the core of our results. Underlying estimates, econometric verification outputs and generated unobservable macroeconomic series are available from the authors upon request.

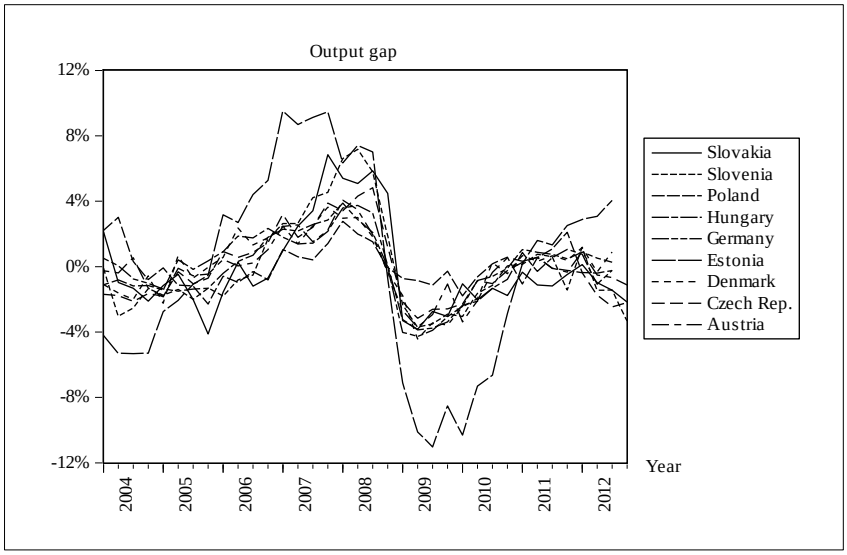


Figure 1 Output gap

Figure 1 shows the baseline HP-filter generated output gap, which is used for most of the following estimations. A prominent 2008-2009 dip in figure 1 reflects the onset of the 2008 economic crisis on different economies. Data for Estonia make for the biggest decline in output gap: from +7.4% in 2008:Q2 to -11.0% in 2009:Q3 (output gap is negative when the actual output lies below potential output). This is in compliance with the fact that Baltic states were among those EU members that were most significantly affected by the crisis onset. Pair wise correlation coefficients for data shown in figure 1 are relatively high and range from approximately +0.75 to +0.92. Surprisingly, Poland is an exception with *corr*(*POL*,*c*) values ranging from +0.35 to +0.60. This might be also observed from figure 1, where output gap for Poland is nearly zero for most of the 2008-2009 period. Whether the 2008 negative shock may have helped synchronizing the business cycles throughout EU economies would be a core question for monetary policy decision making by the ECB and for the potentially upcoming fiscal policy coordination within EU. However, we see no evidence for such synchronizing effect and this question remains open for discussion; see [6] and [7] for details on this topic.

3 Models, empirical results and interpretation

Expectations Augmented Phillips Curve (EAPC) model

Our first model, the EAPC, is a relatively simple modification to the Phillips curve (see [10] for theoretical inference). The corresponding linear regression model (LRM) used in this contribution has been amended by controlling for the autoregressive nature of variables and may be defined as

$$\left(\pi_{ct} - \pi^{e}_{ct}\right) = \beta_0 + \beta_1\left(\pi_{c,t-1} - \pi^{e}_{c,t-1}\right) + \beta_2(u_{ct} - u^{*}_{ct}) + \beta_3(u_{c,t-1} - u^{*}_{c,t-1}) + \varepsilon_{ct} \qquad (3)$$

where π_{ct} and u_{ct} represent CPI inflation and unemployment rate observations for country c at time period t. We use HP-filter generated trends as proxies for the unobserved variables π^{e}_{ct} and u^{*}_{ct}: the expected rate of inflation and the natural rate of unemployment. Therefore, the expressions in brackets may be interpreted as *unanticipated inflation* and *cyclical unemployment* respectively. The estimated residual part $\hat{\varepsilon}_{ct}$ is now defined as the supply shock, according to [10].

	Austria	**Czech Rep.**	**Denmark**	**Germany**	**Hungary**	**Poland**	**Slovenia**	**Slovakia**
Austria	1.000	0.945	0.491	0.329	0.051	0.221	0.601	0.399
Czech Rep.		1.000	0.491	0.335	-0.014	0.234	0.577	0.446
Denmark			1.000	0.342	0.048	0.234	0.381	0.188
Germany				1.000	-0.084	0.322	0.424	0.236
Hungary					1.000	0.144	0.063	-0.002
Poland						1.000	0.395	0.198
Slovenia							1.000	0.261
Slovakia								1.000

Table 1 EAPC-model supply shock correlations

We estimate equation (3) for individual economies and compare the supply shocks obtained. The estimation process may be briefly described as follows: data for all estimated countries were non stationary and cointegrated. Serial correlation of residuals was addressed for all economies by estimating (3) using the Cochrane-Orcutt generalized least squares (GLS) method as described and referenced to in [6]. The R^2 values are satisfactory (above 0.80) and further econometric verifications confirm models' significance and specification. Except for Hungary, every estimated model (3) meets the theoretical constraint $\beta_2 + \beta_3 < 0$ which is a core feature of the Phillips curve and reflects the tradeoff between unanticipated inflation and cyclical unemployment.

Limited theoretical justification of the supply shocks' definition derived from [10] and the relative simplicity of the single equation specification (3) effectively make this model inferior to the following two methods. Yet, the overall heterogeneity in supply shocks as shown in table 1 conforms to the low correlations of the arguably more sophisticated supply shocks as calculated using BQ-decomposition and described in table 3. Also, supply shocks from the EAPC model (3) keep an economically interpretable scale of the dependent variable, whereas shocks from table 3 lose their scale due to orthogonalization.

Beveridge-Nelson (BN) decomposition

This approach belongs to a wider class of *unobserved component* methods. Based on the reasoning performed in [11], we use and interpret the BN-decomposition as a tool providing estimates of unobservable variables instead of treating it as a way of defining a specific type of trend: the BN-trend (as originally suggested in [1]). Lütkepohl [9] and many others provide detailed explanation for the terminology and methods used to describe and estimate model (4). Assuming output y_{ct} has a known ARIMA(p,1,q) structure, we may construct a suitable state-space representation by defining the output as a sum of two unobservable macroeconomic variables: a permanent component (trend) τ_{ct} and a transitory (cycle) component γ_{ct}. For a country c, the state-space representation may be written as

$$y_{ct} = \tau_{ct} + \gamma_{ct}, \tag{4.a}$$

$$\tau_{ct} = \mu_c + \tau_{c,t-1} + \eta_{ct}, \quad \eta_{ct} \sim \mathrm{N}(0, \sigma^2_{c\eta}), \text{ i.i.d.}, \tag{4.b}$$

$$\phi(\mathrm{L})\gamma_{ct} = \theta(\mathrm{L})\omega_{ct}, \quad \omega_{ct} \sim \mathrm{N}(0, \sigma^2_{c\omega}), \text{ i.i.d.}, \tag{4.c}$$

$$corr(\eta_{ct}, \omega_{ct}) = \rho^{(c)}_{\eta\omega}. \tag{4.d}$$

The unobservable permanent component follows a random walk with a drift, while the transitory component is an ARMA (stationary) process with zero mean. Diffuse prior specification (4.d) leaves the correlation between permanent and transitory shocks unconstrained, although various BN-type estimates have been published for both orthogonal ($\rho^{(c)}_{\eta\omega} = 0$) and perfectly correlated ($\rho^{(c)}_{\eta\omega} = 1$) innovations. By enforcing a "smooth trend" prior, we choose ARMA(0,1) structure for all Δy_c series in order to estimate (4) through a slightly modified BNDecomp script, a publicly available EViews Add-in (downloadable from http://www.eviews.com) that calculates the BN-trend as the conditional optimal long-horizon (100 periods ahead) forecast of y_c (with future drift removed). Then, the transitory component may be simply calculated as $\gamma_{ct} = y_{ct} - \tau_{ct}$. Further theoretical background and estimation examples for the state-space model as defined by (4) may be found in [3] and [11].

	Austria	**Czech R.**	**Denmark**	**Estonia**	**Germany**	**Hungary**	**Poland**	**Slovenia**	**Slovakia**
Austria	1.000	0.885′	-0.054	0.159	0.764′	0.989″	0.999″	0.527	0.967″
Czech R.	0.531	1.000	0.362	0.453	0.779′	0.932″	0.870′	0.843′	0.967″
Denmark	0.339	0.807′	1.000	0.903″	0.372	0.026	-0.079	0.729′	0.136
Estonia	0.203	0.710′	0.763′	1.000	0.626	0.190	0.141	0.724′	0.280
Germany	-0.053	-0.207	-0.272	-0.132	1.000	0.741′	0.760′	0.677	0.760′
Hungary	0.916″	0.572	0.315	0.125	-0.021	1.000	0.984″	0.611	0.991″
Poland	0.788′	0.281	0.012	-0.050	0.056	0.771′	1.000	0.500	0.958″
Slovenia	0.342	0.726′	0.594	0.796′	0.034	0.302	0.115	1.000	0.704′
Slovakia	0.731′	0.854′	0.636	0.493	-0.116	0.752′	0.502	0.618	1.000

Table 2 BN-type trend and transitory components: country-wise correlations

Table 2 shows correlations of BN-trends among different countries (upper triangle) as well as the corresponding correlations of transitory components (lower triangle). Hence, for Czech Republic and Slovakia, the coefficient 0.843 refers to $corr(\tau_{CZE}, \tau_{SVK})$ and $corr(\gamma_{CZE}, \gamma_{SVK}) = 0.854$. We use Fisher's z-transformation to evaluate the homogeneity (similarity) among unobservable components for individual economies. Correlations

significantly exceeding arbitrary thresholds of similarity set to $corr(c1,c2) > +0.5$ and $corr(c1,c2) > +0.8$ at the 5% significance level are marked by ′ and ″ respectively. Although correlations in table 2 are significantly higher than correlations for the supply and demand shocks, numerous and significant instances of heterogeneous behavior in table 2 must not be overlooked. At the same time, we need to bear in mind that BN-trend and cycle data may not be *directly* compared to supply and demand shocks from tables 1 and 3 and we only assess the inter-country homogeneity levels obtained through different specification and estimation approaches.

Blanchard-Quah (BQ) structural decomposition, supply and demand shocks

Neoclassical economy provides theoretical grounds for the Blanchard and Quah [2] BQ-decomposition, which is based on *long term identifying restrictions imposed on impulse-response functions* (IRFs) calculated from estimated VAR(*p*) models. It is not a generally applicable procedure and may be utilized only with VAR models composed of certain sets of macroeconomic variables where specific long term economic restrictions are justifiable. As shown in [4] and [7], BQ-decomposition may be also used to untangle the *orthogonal unobservable* neoclassical *supply and demand shocks* (sometimes referred to as structural shocks) from the residuals of an estimated VAR model for each country *c*. BQ-decomposition is performed through implementing additional (identifying) zero-value restriction on the cumulative functions of IRFs: provided demand shocks affect only inflation in the long run, the cumulative response of output to a demand shock may be set to equal zero after an ad-hoc established number of periods representing the long term, whereas reactions to a supply shock are considered permanent (technically, no prior mathematical restriction is imposed on cumulative responses to supply shocks). Lütkepohl [9] provides theoretical description of different aspects of the BQ-decomposition technique (VARs, orthogonalization of residuals, Choleski decomposition and BQ additional long-term identifying conditions).

We have evaluated various specifications of the output and inflation VAR model but neither substituting inflation for unemployment (as suggested by [3]) nor VARX generalization by incorporating dummies (covering the 2008 crisis onset) would yield convincing improvement over the relatively simple specification (5). Also, PPI inflation leads to statistically superior VAR model estimations when compared to our first specification option: the CPI inflation. Finally, a reduced-form second order VAR model was chosen with respect to consistency requirements on VAR estimation (as an input to BQ-decomposition and subsequent comparison among countries) as well as by using the Akaike information criterion. Our model may be written in a simple matrix form as

$$\boldsymbol{y}_{ct} = \boldsymbol{A}_{c1}\boldsymbol{y}_{ct-1} + \boldsymbol{A}_{c2}\boldsymbol{y}_{ct-2} + \boldsymbol{u}_{ct}, \tag{5.a}$$

$$\boldsymbol{\varepsilon}_{ct} = \boldsymbol{G}_c^{-1}\boldsymbol{u}_{ct}, \tag{5.b}$$

where $\boldsymbol{y}_{ct}$ is a 2×1 vector of endogenous variables: output and PPI inflation, both expressed in terms of equation (1). $\boldsymbol{A}_{c1}$ and $\boldsymbol{A}_{c2}$ are 2×2 coefficient matrices for the lagged $\boldsymbol{y}_{ct}$ vectors, $\boldsymbol{u}_{ct}$ is the 2×1 random-element vector and $\boldsymbol{\varepsilon}_{ct}$ is a 2×1 vector containing the estimated supply and demand shocks $\varepsilon_{ct}^{supply}$ and $\varepsilon_{ct}^{demand}$. Details on BQ identification and estimation of the 2×2 decomposition matrix $\boldsymbol{G}_c$ from equation (5.b) are provided in [7]. The main purpose of estimating this model is to generate comparable residual series for subsequent BQ decomposition. Hence, we omit individual tables containing VAR model estimates for individual countries from this contribution. However, proper econometrical verification involving ADF tests for stationarity of endogenous series (only up to the 2008:Q3 shock in some cases) and stationarity of the residuals was performed. As a result of this verification, we had to drop specification (5.a) for Poland from further analysis, as the estimated model would not meet significance criteria (even after relaxing α=5% to α=10%). Such outcome for Poland does seem rather unexpected, yet it may be traced back to the properties of Poland's output gap as in figure 1 and the accompanying description. Further empirically oriented discussion on specification, ordering of variables, model estimation and potential caveats for this type of BQ-decomposition based inter-country analysis is provided in [4] and [7].

	Austria	**Czech R.**	**Denmark**	**Estonia**	**Germany**	**Hungary**	**Slovenia**	**Slovakia**
Austria	1.000	0.572	0.779′	0.264	0.446	0.321	0.494	0.082
Czech R.	0.241	1.000	0.680	0.387	0.466	0.287	0.544	0.172
Denmark	0.224	0.063	1.000	0.361	0.642	0.017	0.457	-0.037
Estonia	-0.001	0.175	0.167	1.000	0.410	-0.268	0.568	0.070
Germany	0.214	0.177	-0.019	-0.259	1.000	-0.049	0.259	0.363
Hungary	0.164	0.322	0.141	0.276	0.311	1.000	-0.095	0.155
Slovenia	0.148	0.369	0.086	0.190	0.195	0.407	1.000	0.154
Slovakia	0.042	0.587	-0.045	-0.030	0.282	0.519	0.072	1.000

Table 3 Supply and demand shocks: country-wise correlations

Table 3 shows the correlations among directly unobservable (decomposed) supply shocks in the upper triangle and demand shocks in the lower triangle as estimated from the residuals for period 2004:Q3 to 2012:Q3. Generally speaking, the level of homogeneity observed among individual economies as in table 3 is significantly lower when compared to the results obtained through Beveridge-Nelson approach. Specifically, only the correlation between supply shocks of Austria and Denmark exceeds our +0.5 similarity threshold at the 5% significance level. Although both unobservable shocks exhibit significant heterogeneity, the lack of homogeneity in supply shocks is more important given the permanent (unconstrained in time) nature of their effects on observable variables. Our results provide evidence against the possibility of reliable EU/EMU-wide unified evaluation of the expected costs and benefits of monetary policy measures under consideration by the ECB and the same unfavorable conclusions must be drawn towards any prospective centralized fiscal policy actions by the EU. Moreover, it should be stressed out that the heterogeneous behavior identified here is not an isolated result, as similar conclusions arise from [7] and from other studies referenced therein.

4 Conclusions

Focusing on some of the core tools and indicators for monetary and fiscal policy decision making process, our contribution finds significant heterogeneity among the potential output, output gap, supply and demand shocks and other unobservable macroeconomic variables that we have consistently estimated and compared across different EU countries. We use diverse models, theoretical backgrounds and disparate estimation methods in order to provide the reader with multiple perspectives on the topic under consideration. Given this approach, the results obtained from different models are not directly nor simply comparable to each other. However, the homogeneity (or the lack of it) that is always evaluated upon consistently estimated and generated data sets may in fact be considered and evaluated across different models.

Our results provide evidence against the synchronization of real business cycles among the EU countries considered. The observed heterogeneity of supply and demand shocks has a strong potential for decreasing the accuracy of consistently evaluated expected impacts of monetary and fiscal policy actions (as opposed to ad-hoc evaluations for individual countries based on diverse and potentially incompatible models). Therefore, we may conclude that prevalent heterogeneity exists between the macroeconomic internal dynamics of Czech Republic and other individual countries as measured and evaluated in this contribution, with potentially negative impacts on the centralized formation of economic policy actions and their unified and consistent evaluation.

Acknowledgements

The research project was supported by Grant No. IGA F4/1/2012, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Beveridge, S., and Nelson, C. R.: A New Approach to Decomposition of Economic Time Series Into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'. *Journal of Monetary Economics* **7** (1981), 151–174.
[2] Blanchard, O. J. and Quah, D.: The dynamic effects of aggregate demand and supply disturbances. *American Economic Review* **79** (1989), 655-673.
[3] Cerra, V., and Saxena S. C.: Alternative Methods of Estimating Potential Output and the Output Gap–An Application to Sweden. *IMF Working Paper* **59**, International Monetary Fund, Washington, 2000.
[4] Clarida, R., and Gali, J.: Sources of Real Exchange Rate Fluctuations: How Important Are Nominal Shocks? *Carnegie-Rochester Conference Series on Public Policy* **41** (1994), 1-56.
[5] Hodrick, R., and Prescott, E. C.: Post War U.S. Business Cycles: an Empirical Investigation. *Journal of Money, Credit and Banking* **29** (1997), 1-16.
[6] Hušek, R.: *Aplikovaná ekonometrie – Teorie a praxe*. VŠE, Nakladatelství Oeconomica, Praha, 2009.
[7] Hušek, R., and Formánek, T.: Srovnání konvergence ekonomik ČR a vybraných zemí eurozóny na základě analýzy funkcí odezvy a nabídkových či poptávkových šoků. *Politická ekonomie* **3** (2011), 291-309.
[8] Juillard, M. et. al.: Measures of Potential Output from an Estimated DSGE Model of the United States. *ČNB Working Paper Series* **11** (2006).
[9] Lütkepohl, H.: *New introduction to multiple time series analysis*. New York: Springer, Berlin, 2005.
[10] Mankiw, N. G.: *Macroeconomics. 2d ed*. New York: Worth, 1994.
[11] Morley, J. C.: The Two Interpretations Of The Beveridge–Nelson Decomposition. *Macroeconomic Dynamics*, **15-03** (2011), 419-439.
[12] Němec, D.: Kvantitativní analýza mezery nezaměstnanosti a výstupu v České republice. *Research Centre for Competitiveness of Czech Economy Working Paper* **22** (2008), 1-22.

A NEW SINGLE-CLASS DEMAND UNCONSTRAINING METHOD

Gregory Fridman[1], Maria Lapina[2]

Abstract.

The authors propose a new statistical unconstraining method which is based on the construction of the distribution function for the censored demand and application of the maximum likelihood approach to estimate distribution parameters. Numerical results are presented of comparative analysis of existing unconstraining methods and the method advocated in the paper. It is demonstrated that the new method has proven to be more efficient in the case of a high percentage of observed censored elements of sample data. Yet another important advantage of the method connected to the fact that it enables one to process the situation of censoring information incompleteness when some elements of the observed sample data are known to be censored or not and for the others this information is not available. Mathematical computer environment Wolfram *Mathematica* has been used for obtaining all the results presented in the paper.

Keywords: demand forecast, unconstraining, statistical distribution, *Mathematica*

JEL classification: C44
AMS classification: 62N

1 Introduction

The theoretical basis and practical applications of unconstraining methods are discussed in many papers, reference to which can be found, for instance, in [1], [2], [3]. It is reported that Projection Detruncation (PD) and Expectation Maximization (EM) methods are the most effective ones. PD was developed at Boeing by Hopperstad [4]. Salch [5] was the first who applied EM approach to the censored data of airline passenger demand. In [6] both iterative statistical methods were compared to four other ones through extensive numerical simulations to demonstrate their advantages and to analyse their impact on forecasting of Revenue Management and therefore on revenue itself.

In the present paper authors advocate a new statistical unconstraining method. It is based on deriving a new statistical distribution which describes a random variable from censored sample data. Unknown parameters of this distribution are then estimated by the maximum likelihood algorithm. Results of numerical calculations are produced for comparative analysis of the proposed method with another three methods.

2 Mathematical formulation and solution to the problem

Consider a normally distributed random variable $X \sim N(\mu_1, \sigma_1)$ with cumulative distribution function $\Phi\Big(\frac{z-\mu_1}{\sigma_1}\Big)$, where

$$\Phi(x) = \frac{1}{2}\left(1 + \operatorname{erf}\Big(\frac{x}{\sqrt{2}}\Big)\right), \quad \varphi(x) = \frac{\mathrm{d}\Phi}{\mathrm{d}x} = \frac{1}{\sqrt{2\,\pi}}\,\mathrm{e}^{-x^2/2}, \quad \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int\limits_0^x \mathrm{e}^{-t^2}\,\mathrm{d}t$$

[1]St.Petersburg State University of Economics, Department of Economical Cybernetics and Mathematical Modeling in Economics, St.Petersburg, Russia, grifri@finec.ru
[2]National Research University Higher School of Economics, St.Petersburg, Russia, lapina.mg@gmail.com

are cumulative distribution function and the probability density function of a normal standard distribution (with zero mean and unit standard deviation) and the error function. Assume that this variable describes a true (unconstrained) demand for a given airticket fare class. Denote $\bar{x} = (x_1, \ldots, x_n)$ to be a sample data for n observations of random variable X.

Let a normally distributed random variable $Y \sim N(\mu_2, \sigma_2)$ characterize restrictions (i.e. booking limits) for this fare class. Its cumulative distribution function $\Phi\left(\frac{z-\mu_2}{\sigma_2}\right)$ depends on parameters (μ_2, σ_2).

Finally, consider a random variable $Z = \min(X, Y)$ which describes the number of bookings for the fare class. Then $z_1 = \min(x_1, y_1)\,, \quad \ldots \quad z_n = \min(x_n, y_n)\,,$ where $\bar{z} = (z_1, \ldots, z_n)$ – a *censored* sample data of the random variable X in which the value x_k is observed only if $x_k \leqslant y_k$. If $y_k < x_k$ then $z_k = y_k$ and in this case the observation x_k is censored:

$$z_k = \begin{cases} x_k\,, & \text{if } \; x_k \leqslant y_k\,; \\ y_k\,, & \text{if } \; x_k > y_k\,, \end{cases} \qquad k = 1, \ldots, n \tag{$*$}$$

From mathematical viewpoint the problem is to estimate distribution parameters μ_1 and σ_1 of random variable X (or both sets of distribution parameters (μ_1, σ_1) and (μ_2, σ_2) of random variables X and Y simultaneously) using censored sample data $\bar{z} = (z_1, \ldots, z_n)$ under assumption that distributions laws of X and Y are known. In general case an additional information is given as well, namely, for some values of $k = 1, \ldots, n$ it is known either an element z_k of the observed sample data $\bar{z}$ represents censored value of the corresponding x_k (i.e. $z_k = y_k$) or not (i.e. $z_k = x_k$).

A straightforward approach to solution to the problem consists of two steps: first, one has to construct a probability density function $f_Z\,(x; \mu_1, \sigma_1; \mu_2, \sigma_2)$ of the random variable $Z = \min(X, Y)$ and then, second, apply maximum likelihood estimation (MLE) technique, originally developed by R.A. Fisher in the 1920s, to estimate all the unknown parameters of derived distribution.

The MLE methods involves the maximization procedure for so-called likelihood function $\mathcal{L}$ which in the considered problem is a joint probability distribution of the n observed values of the censored sample data z_k, $k = 1, \ldots, n$. In the standard case of independent and identically distributed observations z_k, $k = 1, \ldots, n$ the likelihood function $\mathcal{L}$ is given by the product of the individual densities $f_Z\,(z_k; \mu_1, \sigma_1; \mu_2, \sigma_2)$ and takes the form

$$\mathcal{L}(\mu_1, \sigma_1; \mu_2, \sigma_2) = \prod_{k=1}^{n} f_Z\,(z_k; \mu_1, \sigma_1; \mu_2, \sigma_2)\;. \tag{1}$$

The maximum likelihood estimators of the parameters (μ_1, σ_1) and (μ_2, σ_2) are defined as those values $\hat{\mu}_1, \hat{\sigma}_1; \hat{\mu}_2, \hat{\sigma}_2$ which globally maximize function $\mathcal{L}$. For technical and also for theoretical reasons it is easier to work with the logarithm (a monotonically increasing function of its argument) of the likelihood function $\mathcal{L}$.

$$\log \mathcal{L}(\mu_1, \sigma_1; \mu_2, \sigma_2) = \sum_{i=1}^{n} \log f_Z\,(z_k; \mu_1, \sigma_1; \mu_2, \sigma_2) \to \max\;. \tag{2}$$

Consider three cases of the problem formulated above:

• there is no information on which elements of sample data $\bar{z} = (z_1, \ldots, z_n)$ reflects censored or unconstrained values of $\bar{x}$;

• such information can be used for all elements of sample data $\bar{z}$;

• generalized case when such information is partially accessible, i.e. some elements of $\bar{z}$ are known to be censored or not and there is no such information about the others.

It is to be underlined that an assumption as to how X and Y are distributed makes no difference for solution's algorithm.

There is no information on which elements of sample data $\bar{z}$ has been censored

To derive a cumulative distribution function F_Z of random variable $Z = \min\{X, Y\}$ one has to calculate the probability $P(Z > z)$ of the event that random variable Z takes the value greater then z. As a result

one arrives at the formula

$$F_Z(z;\mu_1,\sigma_1,\mu_2,\sigma_2) = 1 - P(Z > z) = 1 - \left(1 - \Phi\Big(\frac{z-\mu_1}{\sigma_1}\Big)\right)\left(1 - \Phi\Big(\frac{z-\mu_2}{\sigma_2}\Big)\right), \tag{3}$$

what readily yields the probability density function $f_Z = \mathrm{d}F_Z/\mathrm{d}Z$ of Z:

$$f_Z(z;\mu_1,\sigma_1,\mu_2,\sigma_2) = \frac{1}{\sigma_1}\,\varphi\Big(\frac{z-\mu_1}{\sigma_1}\Big)\left(1 - \Phi\Big(\frac{z-\mu_2}{\sigma_2}\Big)\right) + \frac{1}{\sigma_2}\,\varphi\Big(\frac{z-\mu_2}{\sigma_2}\Big)\left(1 - \Phi\Big(\frac{z-\mu_1}{\sigma_1}\Big)\right) \tag{4}$$

and the log-likelihood function $\log\mathcal{L}(\mu_1,\sigma_1,\mu_2,\sigma_2)$ takes the form

$$\log\mathcal{L}(\mu_1,\sigma_1,\mu_2,\sigma_2) = \sum_{k=1}^{n}\log\left\{\frac{1}{\sigma_1}\,\varphi\Big(\frac{z_k-\mu_1}{\sigma_1}\Big)\left(1 - \Phi\Big(\frac{z_k-\mu_2}{\sigma_2}\Big)\right) + \right.$$
$$\left. + \frac{1}{\sigma_2}\,\varphi\Big(\frac{z_k-\mu_2}{\sigma_2}\Big)\left(1 - \Phi\Big(\frac{z_k-\mu_1}{\sigma_1}\Big)\right)\right\}. \tag{5}$$

Censoring information can be used for all elements of sample data $\bar{z}$

Following the same analytical procedure, the likelihood function $\mathcal{L}$ can be obtained as follows

$$\mathcal{L}(\mu_1,\sigma_1,\mu_2,\sigma_2) = \prod_{k=1}^{m}\frac{1}{\sigma_1}\varphi\Big(\frac{x_k-\mu_1}{\sigma_1}\Big)\left(1 - \Phi\Big(\frac{x_k-\mu_2}{\sigma_2}\Big)\right)\prod_{j=1}^{r}\frac{1}{\sigma_2}\varphi\Big(\frac{y_j-\mu_2}{\sigma_2}\Big)\left(1 - \Phi\Big(\frac{y_j-\mu_1}{\sigma_1}\Big)\right), \tag{6}$$

where m denotes the number of unconstrained elements and r is the number of censored elements in the sample data $\bar{z}$ and, naturally, $m + r = n$. Introduction of log-likelihood function is effective as well from the viewpoint of the global maximization procedure $\log\mathcal{L}(\mu_1,\sigma_1,\mu_2,\sigma_2) \to \max$.

Generalized case when censoring information is incomplete

Assume that the censoring information for the observed sample data $\bar{z}$ is incomplete, i.e. for some elements z_k it is unknown if they represent x_k or y_k. In this case the likelihood function is constructed as a product of expression for unconstrained and censored elements of $\bar{z}$ and expression for those z_k of which the censoring information is absent:

$$\mathcal{L}(\mu_1,\sigma_1,\mu_2,\sigma_2) = \prod_{k=1}^{m}\frac{1}{\sigma_1}\,\varphi\Big(\frac{x_k-\mu_1}{\sigma_1}\Big)\left(1 - \Phi\Big(\frac{x_k-\mu_2}{\sigma_2}\Big)\right)\times\prod_{j=1}^{r}\frac{1}{\sigma_2}\,\varphi\Big(\frac{y_j-\mu_2}{\sigma_2}\Big)\left(1 - \Phi\Big(\frac{y_j-\mu_1}{\sigma_1}\Big)\right)\times$$
$$\times\prod_{l=1}^{n-m-r}\left\{\frac{1}{\sigma_1}\,\varphi\Big(\frac{z_l-\mu_1}{\sigma_1}\Big)\left(1 - \Phi\Big(\frac{z_l-\mu_2}{\sigma_2}\Big)\right) + \frac{1}{\sigma_2}\,\varphi\Big(\frac{z_l-\mu_2}{\sigma_2}\Big)\left(1 - \Phi\Big(\frac{z_l-\mu_1}{\sigma_1}\Big)\right)\right\}. \tag{7}$$

where, again, m denotes the number of unconstrained elements in observed sample data $\bar{z}$ and r is the number of censored elements in observed sample data $\bar{z}$.

Note that the problem of finding the estimations of parameters of the probability density function f_Z in all three cases can be considered either under the assumption that parameters μ_2 and σ_2 for random variable Y are known or without it. Even if the problem is to find estimations to all four parameters (μ_1,σ_1) and (μ_2,σ_2) it can be subdivided into two independent ones due to the fact that partial derivatives $\log\mathcal{L}$ with respect to μ_1 and σ_1 do not depend on μ_2 and σ_2 and vise versa. The global optimization can be implemented basing on one of the well-known algorithms.

3 Numerical results and discussions

The results of computer numerical simulation are presented in this section to analyse advantages and drawbacks of the proposed unconstraining method and to compare its effectiveness (first of all, accuracy and performance) with those of PD and EM approaches. The simulation methodology was chosen to be partly similar to that used in [6]. Some important alterations were introduced to the process of

unconstrained demand and booking limit generation and additional numerical analysis was performed taking into account the effect of censoring information incompleteness degree in observed sample data on accuracy of the proposed method. Two simulated data sets (i.e. observed sample data $\bar{z}$) were generated. The first one is obtained using initial parameters $\mu_1 = 20$, $\sigma_1 = 4$ and $\sigma_2 = 4$ and the second one using $\mu_1 = 0$, $\sigma_1 = 1$ and $\sigma_2 = 1$. The percentage of unconstraining observations was chosen to be 50%, 75%, 85%, 90%, and 98%.

Figure 1 presents comparative numerical results for unconstrained estimations of mean (left) and standard deviation (right) obtained using N3, PD, EM, and Z1 methods for the first simulated data set in the form of box-and-whisker chart. It is assumed that censoring information can be used for all elements of sample data $\bar{z}$.

Calculations demonstrate that method Z1 gives identical results with EM for moderate and high values of percentage of censored observations and better results for very high censored sample data (more than 96%). Mean absolute errors are calculated in Table 1 for the unconstrained mean and the unconstrained standard deviation.

	N3	PD	EM	Z1
50%	(1.59634, 1.25069)	(0.614932, 0.948666)	(0.229581, 0.207031)	(0.229581, 0.207031)
75%	(3.10603, 1.93417)	(1.66712, 1.8206)	(0.515712, 0.528677)	(0.515687, 0.528612)
85%	(3.91338, 2.34192)	(2.34452, 2.36615)	(0.591423, 0.53379)	(0.591017, 0.533552)
90%	(4.59367, 2.57692)	(2.91175, 2.68328)	(0.855289, 0.713381)	(0.855284, 0.713378)
98%	(6.63266, 3.17802)	(4.5955, 3.46528)	(2.12046, 1.38938)	(1.91494, 1.21637)

Table 1 Mean Absolute Error (MAE) for the first data set, $\mu_1 = 20$, $\sigma_1 = 4$. The first number in each pair gives MAE for the unconstrained mean and the second – for the unconstrained standard deviation.

Note that box-and-whisker charts demonstrated in the paper for presentation of the numerical results are much more informative than simple histograms that usually used, see, for instance [2], [6].

Figure 2 as well as Table 2 illustrate numerical results for the second data set. Again, it is seen that Z1 method is more effective especially for very high values of percentage of observed censored elements.

	N3	PD	EM	Z1
50%	(0.406433, 0.313069)	(0.155855, 0.236075)	(0.053293, 0.052207)	(0.053293, 0.052207)
75%	(0.759536, 0.499272)	(0.406477, 0.471294)	(0.129863, 0.128624)	(0.128468, 0.12785)
85%	(1.01825, 0.574064)	(0.599651, 0.579925)	(0.152037, 0.135692)	(0.152034, 0.135692)
90%	(1.1424, 0.637409)	(0.697426, 0.666965)	(0.181933, 0.160816)	(0.181937, 0.160817)
98%	(1.63486, 0.8068)	(1.14214, 0.882418)	(0.525655, 0.389394)	(0.421494, 0.264834)

Table 2 Mean Absolute Error (MAE) for the second data set, $\mu_1 = 0$, $\sigma_1 = 1$. The first number in each pair gives MAE for the unconstrained mean and the second – for the unconstrained standard deviation.

Finally, figures 3 and 4 contain numerical results of the unconstraining procedure for the first and the second data set correspondingly in the case when some elements of the observed sample data are known to be censored or not and for the others this information is not available. Note again that Z1 method is alone can be used in such a situation of all the unconstraining methods.

Conclusions

This paper presents a new unconstraining approach based on construction of probability density function of the censored random variable and estimation procedure for the distribution parameters using maximum likelihood method. The results of numerical simulations demonstrated that the proposed method has better accuracy than well-known PD and EM approaches at a high percentage of the censoring. Another important advantage of new method is connected to the fact that it enables one to process the situation of censoring information incompleteness when some elements of the observed sample data are known to be censored or not and for the others this information is not available. Other methods cannot be used to work with such kind of data. All the calculations are produced in computer mathematical environment Wolfram *Mathematica*.

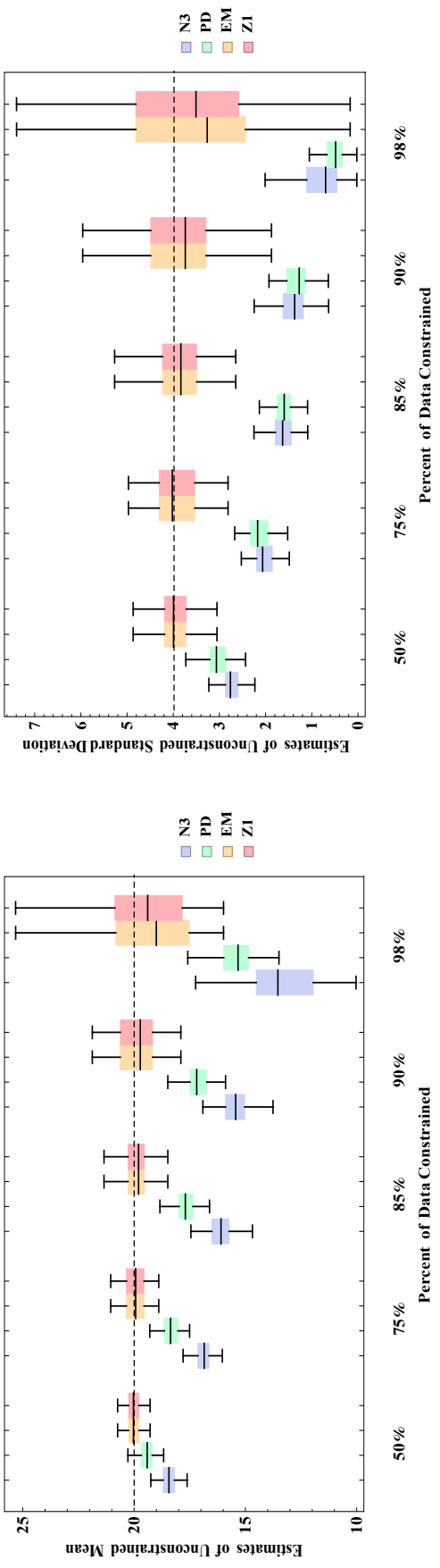


Figure 1 Unconstrained estimations of mean (left) and standard deviation (right) obtained using N3, PD, EM, and Z1 methods for the first simulated data set ($\mu_1 = 20$, $\sigma_1 = 4$ and $\sigma_2 = 4$) for different percentage of censored observations.

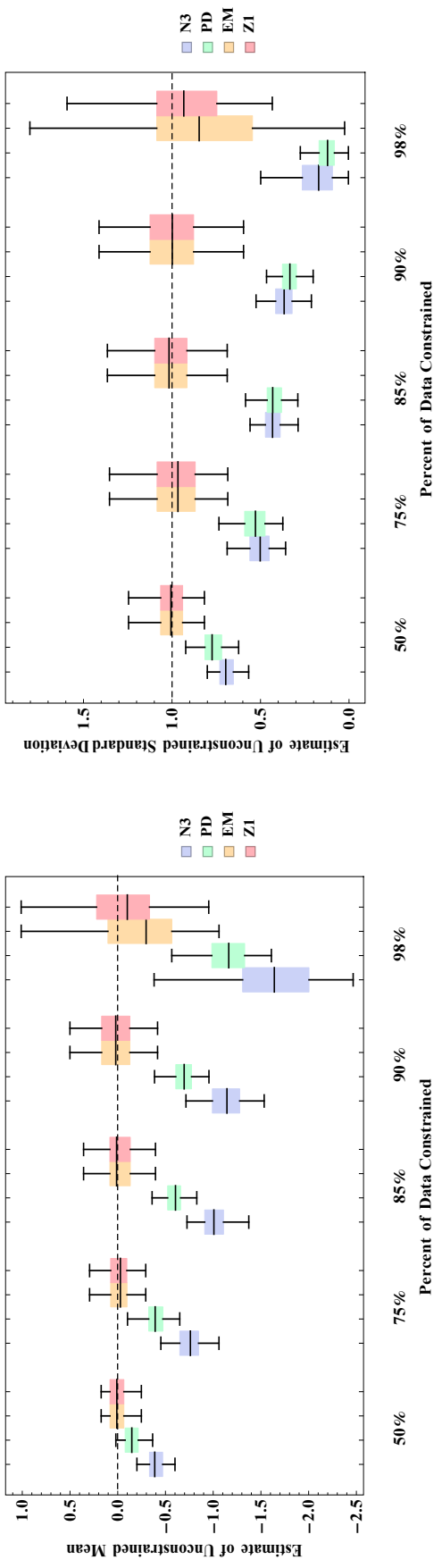


Figure 2 Unconstrained estimations of mean (left) and standard deviation (right) obtained using N3, PD, EM, and Z1 methods for the second simulated data set ($\mu_1 = 0$, $\sigma_1 = 1$ and $\sigma_2 = 1$) for different percentage of censored observations.

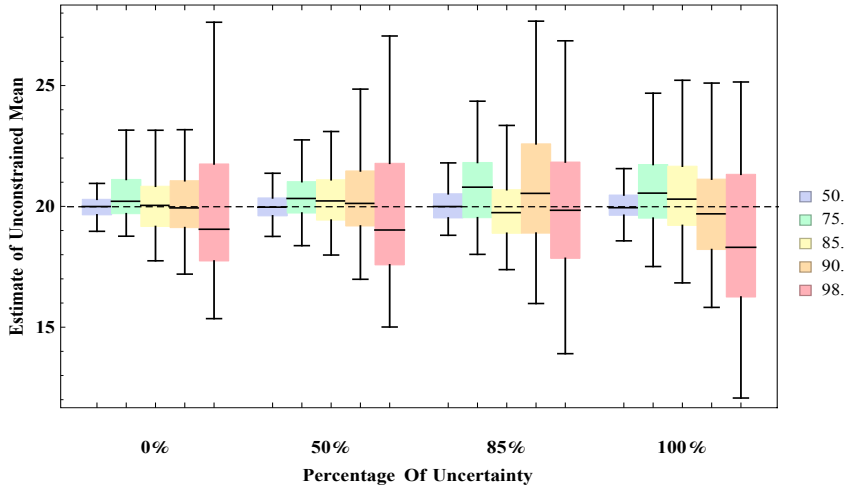


Figure 3 Unconstrained estimations of mean obtained using Z1 method for the first simulated data set ($\mu_1 = 20$, $\sigma_1 = 4$ and $\sigma_2 = 4$) for different percentage of censored observations and information incompleteness.

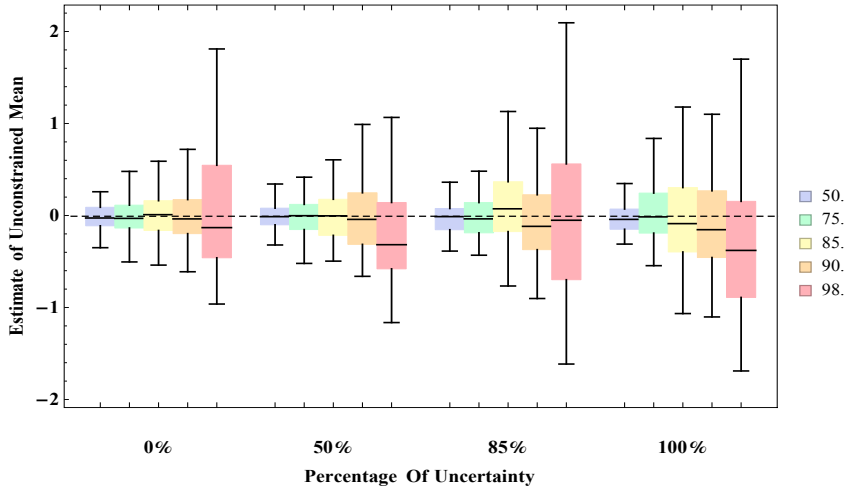


Figure 4 Unconstrained estimations of mean obtained using Z1 method for the second simulated data set ($\mu_1 = 0$, $\sigma_1 = 1$ and $\sigma_2 = 1$) for different percentage of censored observations and information incompleteness.

References

[1] K. Talluri and G. J. van Ryzin 2004 The theory and practice of revenue management, Kluwer Academic Publishers, Boston, USA.

[2] R.H. Zeni 2001 Improved forecast accuracy in revenue management by unconstraining demand estimates from censored data, Ph.D. thesis, Rutgers University, Rutgers, NJ, USA.

[3] P. Guo, B. Xiao, and J. Li 2012 Unconstraining methods in revenue management systems: research overview and prospects, Advances in Operations Research, vol. 2012, article ID 270910.

[4] C. Hopperstad 1995 An alternative detruncation method, Boeing Commercial Aircraft Company Internal Document, Boeing, Renton, Washington, USA.

[5] J. Salch 1997 Unconstraining passenger demand using the EM algorithm, in *Proceedings* of the INFORMS Conference, Dallas, Texas, USA.

[6] L.R. Weatherford and S. Pölt 2002 Better unconstraining of airline demand data in revenue management systems for improved forecast accuracy and greater revenues, Journal of Revenue and Pricing Management, vol. 1, No. 3, pp. 234–254.

Stochastic analysis of costs during the pig breeding process

Ludvík Friebel[1], Jana Friebelová[2], Lukáš Marinov[3]

Abstract: This contribution deals with the economy of agricultural production, specifically in animal farming. The main goal of this paper is to find and apply an efficient simulation model of pig breeding at a particular company in order to improve its economic results. The partial goal is to identify a pig breed that is the most profitable using current technologies. Costs at particular parts of the breeding process were used as the main simulation inputs, together with frequency of litters, length of meantime, number of weaned piglets depending on a sequence of the litter, length of breast-feeding time, and the length of time of pregnancy of sows. Additional inputs were price of produced piglets, price of insemination doses etc. Statistical distributions for the simulations of the inputs were obtained with theoretical curve fitting of sample data, which were provided from research projects focused on piglet production efficiency. There also was obtained the empirical the curve of sow´s profitability depending on number of litters which can help to decide about the culling of sows from the herd. A simulation was performed using the program @RISK. The obtained results and recommendations are further discussed in the paper.

Key words: simulation; pig breeding; probability distribution; weaning of piglets; culling of sows;

JEL Classification: C13, C15, P4
AMS Classification: 60H30

1 Introduction

Many authors have dealt with the economy of piglet production farms, e.g. Pulkrábek [9], Brož [1]. The key factors of economy of piglet production system can be divided into two groups – exogenous and endogenous. Exogenous factors, such as the price of feeding, the price of energy and the price of pork meat, can be influenced negatively. In contrast, the endogenous factors can be managed with success. The optimal replacement policy of sows is described by Rodriguez et al in [10]. Plà et al. have developed a production model using Markow [8] and Semi-Markow chains [7]. A review of mathematical models for sow herd management is described by Plà [6]. Usage of stochastic process models in sow herd management is dealt with by Marín [5].

In this contribution, we focus on the distribution of time spent by farrowing batches at different parts of the production process during meantime. Because of the stochastic character of animal reproduction, we have developed a stochastic simulation model [3] covering costs and incomes arising during the whole reproductive cycle.

2 Methodology

Stochastic analysis of the pig breeding process had to work in the circumstances of a model farm. The pig breeding process is depicted in Figure 1. At the top of Figure 1, we can see the time axis.

Because the duration of production cycles is mandatory (biologically) set, we can only change the duration of the particular stage of meantime of the sows and sows with piglets. However, the possible change of the above mentioned stages have to be ethologically admissible.

The second parameter of the pig breeding process which can be changed with success and which has great economic impact is the culling policy. In Czech farms, it is typical that the culling of sows is induced by veterinary conditions (in our reference farm 68%) and is not regulated statistically as in the other countries.

[1] University of South Bohemia/Faculty of Economics, Department of Applied Mathematics and Informatics, Studentská 13, České Budějovice, friebel@ef.jcu.cz.
[2] University of South Bohemia/Faculty of Economics, Department of Applied Mathematics and Informatics, Studentská 13, České Budějovice, friebelova@ef.jcu.cz.
[3] Milan Král, a.s., marinov@milankral.cz

Placing the inefficient sows into the brewing process from the point of view gestation and fertility has an undesirable impact on the economy of the whole farm.

On the model farm, two breeds of pig were entered into the breeding process. An additional aim of our contribution is to determine which of them is more efficient from the economic point of view.

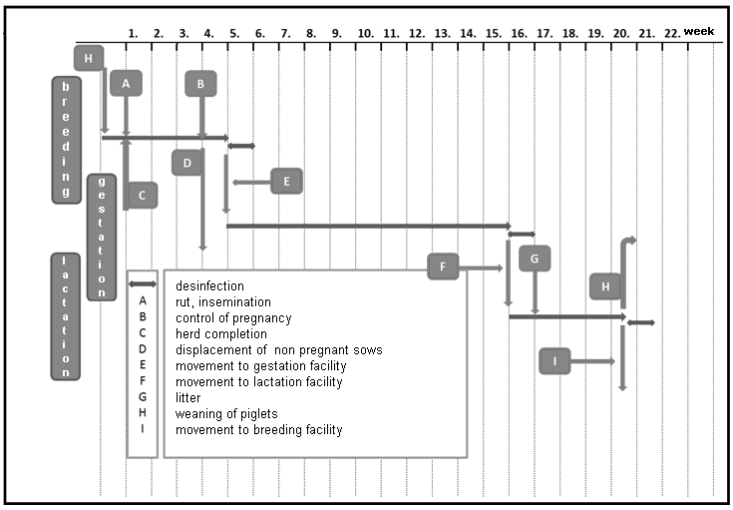


Figure 1: Production diagram

2.1. Data

The data file included 1431 records of Czech Landrase sows (CL) and 3145 of Large White pig (LW) sows. Each record included the birth date, date of culling, reason of culling, date of litter, number of piglets weaned in each litter, number of insemination, litter weight at birth, weaning weight etc. The data had to be processed in order to gain necessary simulation inputs.

2.2. Structure of the herd

The herd structure is a crucial parameter which influences the total efficiency of piglet production. This indicator is primarily given by the veterinary conditions on the farm. The secondary impact is the culling policy depending on sow efficiency. The contemporary structure of the herd is depicted in Table 1.

Litter count	CL-non culled	LW-non culled
0	43	141
1	45	127
2	30	78
3	28	21
4	9	31
5	5	12
6	3	7
7	0	1
Total	163	418

Table 1: Structure of herd depending on frequency of litters

2.3. Costs and incomes

Total costs do not include the purchase of gilts and insemination doses. The highest costs occur in the lactation facility. The level of costs there is 200% higher than elsewhere.

Feeding costs, water costs, wage costs, veterinary costs, energy, purchase of gilts, and insemination doses are quantified in Table 2. For a detailed analysis of costs see http://home.ef.jcu.cz/~ludva/farm1.

Litter count	CZK/head
Lactation	59,09
Gestation	26,53
Breeding	25,88
Gilts	28,34
Contact animals	24,75
Teaser male	29,16
Weaner pen	10,03
Piglets	1,14

Table 2: Total daily costs

The main income of a pig breeding company is from piglets sold. The average price is 72 CZK per kilogram. This price was computed as an average within the last year. The average weight of a weaned piglet is 7.6 kg, the weight of a piglet sold is 30 kg.

2.4. Simulated inputs

Number of inseminations

The price of an insemination dose is one of the important costs. The number of inseminations strongly depends on the number of ruts before becoming pregnant. In order to create this simulation input we used the Distribution Fitting function in the program @RISK. The most suitable were the Binomial and Poisson distributions [4]. Because the probability of becoming pregnant differs from the increasing number of litters, we had to suggest a particular probability distribution function for each litter. In Figure 2, we can see the PDF of becoming pregnant for LW and for the first litter. The number of inseminations was monitored only up to the 15th litter for the same reasons as for the litter size.

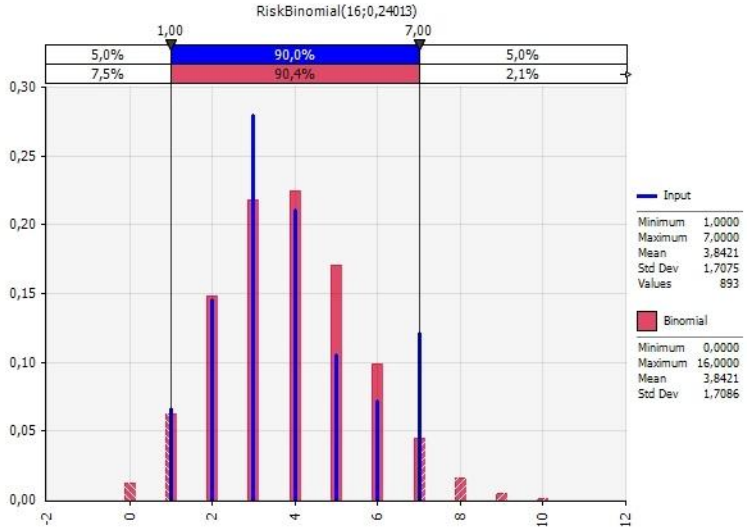


Figure 2: Fitting of number of inseminations

Litter size - number of piglets in a particular litter

The litter size depends on the number of litters. For CL sows, this size increases up to the fourth litter and thereafter decreases. A similar situation exists for LW-sows. In order to create this simulation input we used the Distribution Fitting function in the program @RISK [2]. The most suitable was the Binomial distribution. Because the probability of becoming pregnant differs from litter to litter, we had to suggest a particular PDF for each one. The litter size was monitored only up to 15th litter, thereafter the efficiency of the sow slopes down mainly because of death losses of baby pigs. In Figure 3, we can see the PDF of litter size of CL sows and of the first litter. Because of completeness, we used for the creation of this simulation the input data about culled sows of both breeds.

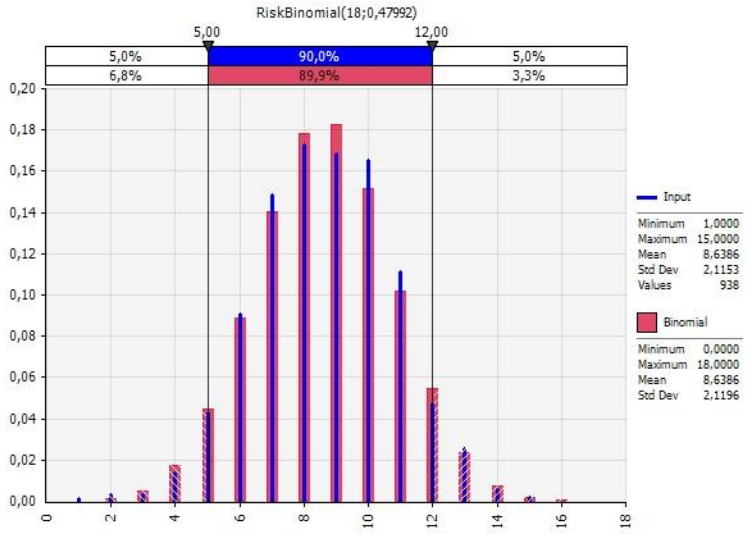


Figure 3: Number of piglets

2.5. Building of a simulation model

Simulation inputs

- Total cost in particular facility;
- Price of weaned piglets;
- Price of insemination dose;
- Litter frequency;
- Meantime (interval between two litters);
- Litter size;
- Number of inseminations before becoming pregnant.

This model was built for each breed separately in order to compare their efficiency. We formed a model batch of sows in accordance with the structure of the total herd. For this model batch we performed a relevant number of iterations with different parameters. The model was moreover divided into models of cost and incomes (see Table 3 and 4).

b	p	v	TB	CB	TG	CG	TL	CLS	PW	CPL	t	TC	CP
1	1	1	98	3 600.74	77		29		9	225.72	204.00	7 532.83	4 062.15
1	2	2	56	1 986.08	77		29		10	250.80	162.00	5 943.25	4 513.50
1	3	3	35	1 178.75	77		29		10	250.80	141.00	5 135.92	4 513.50
1	4	4	35	1 178.75	77		29		10	250.80	141.00	5 135.92	4 513.50
1	5	5	56	1 986.08	77	1 992.76	29	1 713.61	10	250.80	162.00	5 943.25	4 513.50
1	6	6	35	1 178.75	77		29		9	225.72	141.00	5 110.84	4 062.15
1	7	7	35	1 178.75	77		29		9	225.72	141.00	5 110.84	4 062.15
1	8	8	35	1 178.75	77		29		9	225.72	141.00	5 110.84	4 062.15
1	9	9	35	1 178.75	77		29		9	225.72	141.00	5 110.84	4 062.15

b - batch, p - pig, v - number of litter, TB - time spent in breeding facility [day], CB - costs in breeding facility [CZK], TG - time spent in gestation facility [day], CG - cost in gestation facility [CZK], TL - time spent in lactation facility [day], CLS - cost on sow in lactation facility [CZK], PW - number of weaned piglets, CPL - costs on piglets in lactation facility [CZK], t - reproductive cycle [day], TC - Total costs for a sow from piglets to weaning [CZK], CP - cost in weaner pen [CZK]

Table 3: Simulation model costs

PP	PR	SL	TO	PS	NS
14 445.00	2850.02	0.3357	1.7893	1 712.00	5.41
16 050.00	5593.25	0.3046	2.2531	3 838.04	6.86
16 050.00	6400.58	0.1871	2.5887	3 099.21	4.84
16 050.00	6400.58	0.0504	2.5887	834.40	1.30
16 050.00	5593.25	0.0743	2.2531	936.84	1.67
14 445.00	5272.01	0.0288	2.5887	392.73	0.67
14 445.00	5272.01	0.0168	2.5887	229.09	0.39
14 445.00	5272.01	0.0024	2.5887	32.73	0.06
14 445.00	5272.01	0	2.5887	0.00	0.00
			AS	11075.05	
				AP	21.21

PP - Price for piglets sold per litter [CZK], PR - profit per sow during reproductive cycle [CZK], SL - share of particular number of litter in a batch, TO - herd turnover, PS - Profit per sow for a year [CZK], NS - number of piglets, AS - average profit per sow and year [CZK], AP - Average number of piglets per sow and year.

Table 4: simulation model profitability

3 Results

We performed 1000 iterations for different weaning periods. We tested seven possible periods of weaning for 22–28 days for both breeds of pigs and we monitored the total costs and incomes. The total profitability for both breeds is given in Tables 5 and 6. It is evident that the CL pig is more efficient with a weaning time of 22 days, but with a lower risk quantified in the standard deviation. The optimal time to wean is identical for both breeds. Histograms for LW and CL sows are depicted in Figures 5 and 6. It is evident that stochastic CL sows achieve more balanced economic results, the double sided 90% quantile is narrower than for LW stochastic sows. For complete assessment we have to carefully monitor the impact of the time to wean on the weaning weight.

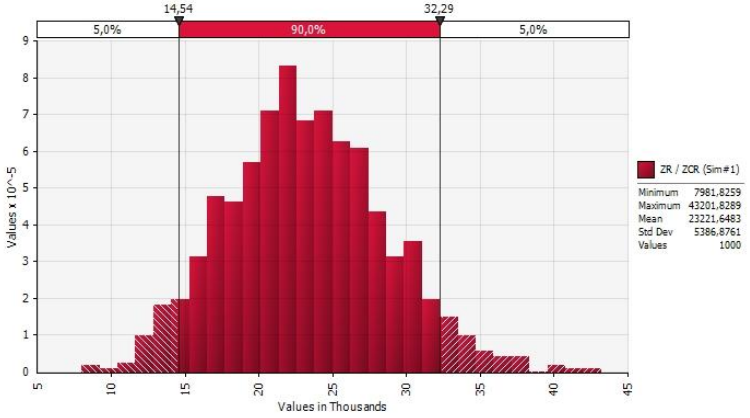


Figure 4: probability distribution function of profit per sow of LW for weaning in 22 days

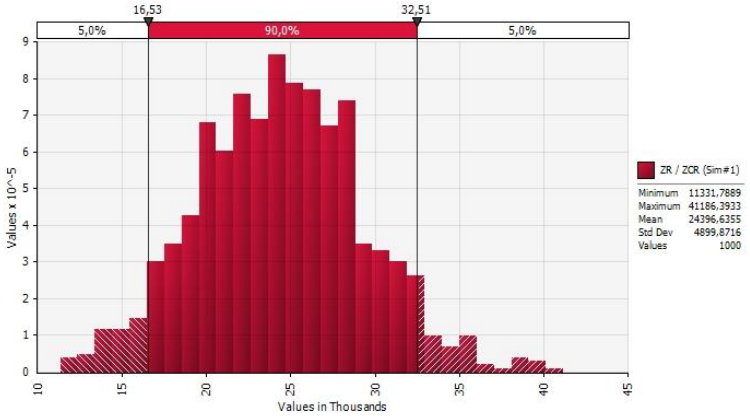


Figure 5: probability distribution function of profit per sow of CL for weaning in 22 days

An interesting fact is that the mean value of piglets weaned per sow for the LW is greater than for the CL although the CL is profitable. This paradox is caused by a greater meantime for the LW, which give birth to more piglets per litter, but the litters occur less frequently.

Days to weaning	**22**	**23**	**24**	**25**	**26**	**27**	**28**
Mean	23222	23124	23027	22931	22837	22745	22653
Std Dev	5387	5367	5347	5328	5309	5290	5272

Table 5: Profitability LW (Profit per sow and year)

Days to weaning	**22**	**23**	**24**	**25**	**26**	**27**	**28**
Mean	24397	24287	24179	24072	23967	23864	23762
Std Dev	4900	4879	4858	4837	4817	4798	4778

Table 6: Profitability CL (Profit per sow and year)

In the presented model, we supposed a uniform distribution of costs during the whole stay for particular parts of production. In further research we will focus on a detailed analysis of the costs during the stay of sows on particular facilities. For example, consumption of feeding there increases during the first 13 days by more than three times, see http://home.ef.jcu.cz/~ludva/farm1. The most expensive feeding occurs at the weaning pen where the situation is very similar.

As was mentioned in the methodology part, one of the most important questions of the pig breeding process with a great economic impact is the culling policy. This policy is closely connected with purchasing gilts and herd replenishing. Revenues for braked sows are economically inconsiderable.

Besides other things, the proposed model can be used for developing pig farms, including dimensioning particular production facilities.

References

[1] Brož, V., Kic, P.: *Technika v dochovu a výkrmu prasat.* 1. vydání. IVVMZ, Praha, 1996. 60 p.
[2] Hardaker, J.B., Winston, W. L.: *Simulation Modeling Using @RISK*. Dusbury, 2001.
[3] Hušek, R., Lauber, J.: *Simulační modely*. Nakladatelství technické literatury, Praha, 1987. 352 p.
[4] Johnson, N. L., Kotz, S.: *Distributions in Statistics* - Vol. I-IV. John Wiley and Sons, New York, 1970.
[5] Marín, J.M., Pla, L.M. and Ríos, D.: Inference for some stochastic process models related with sow management. *Journal of Applied Statistics*. 32(8). 2005. 797-812.
[6] Plà, L.M.: Review of mathematical models for sow herd management. *Livestock Production Sciences*. 106. 2007. 107-119.
[7] Plà, L.M., Faulin, F. J., Rodríguez, S.V.: A Linear Programming Formulation of a Semi-Markov Model to Design Pig Facilities. *Journal of Operational Research Society*. 60: 2009. 619-625.
[8] Plà, L.M., Pomar, C., Pomar, J.: A Markov Decision Sow Model representing the productive lifespan of herd sows. *Agricultural Systems* 76. 2003. 253-272.
[9] Pulkrábek, J., et al.: *Chov prasat*. 1. vyd. Praha : ProfiPress, 2005. 160 p.
[10] Rodriguez, S., Jensen, T. B., Plà, L.M., Kristensen, A. R.: Optimal replacement policies and economic value of clinical observations in sow herds. *Livestock Science*. 138. 2011. 207-219.

On Active Set Method and Second-Order Inclusion Probabilities

Wojciech Gamrot [1]

Abstract. Under various sampling schemes the exact calculation of inclusion probabilities is prohibitively complex even for modest population sizes. Such a difficulty may be tackled by replacing unknown first-order inclusion probabilities in the Horvitz-Thompson expansion estimator of population total with estimates computed in a simulation study. However, one may also desire to estimate the variance of this statistic using simulated second-order inclusion probabilities. To improve the accuracy of their estimates, the available auxiliary information may be employed. Such an information often takes form of partial (hierarchic) ordering inequalities that may be incorporated into estimation using isotonic regresson methods. In this paper such an approach is adopted for a fixed-cost sequential sampling scheme. The active set optimization algorithm is applied to estimate inclusion probabilities and the variance of population total estimates.

Keywords: inclusion probability, restricted estimation, population total.

JEL classification: C83
AMS classification: 62D05

1 Introduction

Design-based estimation of finite population parameters is based on the assumption that population values of the study variable are fixed and the only source of estimator's variability is the sampling scheme. Inclusion probabilities characterizing the sampling scheme play important role in the construction of estimators and assesment of their properties. Sometimes these probabilities are hard to calculate. This problem arises for various sequential sampling and rejective sampling strategies considered among others in [5], [15], [2], [16]. In such circumstances one may focus on estimators that do not utilize inclusion probabilities at all, such as sufficiency-based estimator of [10] or model-based estimators such as those in [12], [17] or [14]. Another possibility is to replace unknown first-order inclusion probabilities with estimates obtained from the simulation experiment. Such a technique has been considered in [6], [13] and [7] for estimation of population totals. It will now be extended to estimating second order inclusion probabilities and the variance of population total estimates. Let the finite population of size N be represented by a set of unit indices $U = \{1, ..., N\}$. Fixed values of the study variable are denoted by $y_1, ..., y_N$. The objective of the survey is to estimate the population total $t = \sum_{i \in U} y_i$ from the sample s. When first-order inclusion probabilities $\pi_1, ..., \pi_N$, $\pi_i = Pr(i \in U)$ for $i \in U$ are known, it is estimated without bias by the well-known expansion estimator:

$$\hat{t} = \sum_{i \in s} \frac{y_i}{\pi_i} \tag{1}$$

Its variance depends on second-order probabilities $\pi_{ij} = Pr(i, j \in s)$ for $i, j \in U$ and it is estimated by:

$$\hat{V}(\hat{t}) = \sum_{i,j \in s} \check{y}_i \check{y}_j \check{\Delta}_{ij} \tag{2}$$

where $\check{y}_i = y_i/\pi_i$, $\check{\Delta}_{ij} = \Delta_{ij}/\pi_{ij}$ and $\Delta_{ij} = \pi_{ij} - \pi_i \pi_j$ for $i, j \in U$. The statistic $\hat{V}(\hat{t})$ may occasionally become negative which is unwelcome but usually unlikely to happen. When inclusion probabilities are

[1] University of Economics, Department of Statistics, 1 Maja 50, 40-287 Katowice

unknown, Fattorini [6] proposes to carry out a simulation to compute estimates $\hat{\pi}_1, ..., \hat{\pi}_N$ of first order inclusion probabilities and to plug them into (1). This results in a simulation-based estimator:

$$\hat{t}_* = \sum_{i \in s} \frac{y_i}{\hat{\pi}_i} \tag{3}$$

To preserve the finiteness of expansion estimator, estimates of first order inclusion probabilities should remain strictly positive. This may be achieved by using Fattorini's formula:

$$\hat{\pi}_{iF} = \frac{k_i + 1}{R + 1} \tag{4}$$

where R is the number of simulated sample replications and k_i is the number of times the i-th unit is drawn for $i \in U$. Other likelihood-based non-negative estimators are also considered in [9]. When suitable auxiliary information in the form of ordering constraints is available, Gamrot [8] proposes to apply isotonic regression procedures to compute improved estimates of first order inclusion probabilities, and studies properties of resulting population total estimator. When variance estimates for (3) are desired, they may be constructed by replacing unknown true second-order probabilities in (2) with Fattorini's estimates:

$$\hat{\pi}_{ijF} = \frac{k_{ij} + 1}{R + 1} \tag{5}$$

where k_{ij} is the number of times the i-th and j-th unit is jointly drawn for $i \neq j \in U$. However, true second-order inclusion probabilities tend to be much lower than first-order ones. Consequently, sampling weights computed as reciprocals of $\hat{\pi}_{ijF}$'s become very unstable. As a result, variance estimates are also highly variable and the probability of obtaining a negative variance estimate increases as well. To mitigate this effect, one may again consider using auxiliary information to improve precision of second-order inclusion probability estimates in the spirit of Gamrot [8]. In the next section this approach is adopted for a certain sampling scheme. An alternative variance estimator is then constructed.

2 Fixed-cost sampling

The fixed-cost sequential sampling scheme of Pathak [11] is characterized by varying inclusion probabilities which are computationally extremely demanding even for very low population sizes. Although sufficiency-based design-unbiased estimates of population totals do exist for this scheme and may be calculated without knowing inclusion probabilities, the simulation approach may be of interest when nonresponse adjustments need to be incorporated or when the original scheme is modified. Overall the scheme constitutes a convenient model for presenting the proposed approach.

Let $c(1), ..., c(N)$ be individual costs of surveying respective population units, known in advance and let L be the fixed survey budget. Budget excesses are unacceptable. The procedure draws units to the sample one by one without replacement and with equal probabilities, until the cumulative cost of the sample becomes greater than or equal L. The element for which it occurs is not included in the sample. Hence the sample selection under such a scheme may be treated as happening in two phases. In the first phase all population units are sorted randomly into a sequence $(A_1, ..., A_N)$ of size N in such a way that all permutations of U are equally likely to occur. In the second phase the cumulative cost of succeeding units in the obtained sequence is computed and first M units are sampled, where M is chosen in such a way that $c(A_1) + ... + c(A_M) < L$ and $c(A_1) + ... + c(A_{M+1}) \geq L$. The sample size is random.

Let $\Psi = \{\psi_1, ..., \psi_{N!}\}$ represent all permutations of U where the k-th permutation is denoted by $\psi_k = (a_1^{(k)}, ..., a_N^{(k)})$ for $k \in \{1, ..., N!\}$. The sequence of individual costs corresponding to elements of ψ_k is $(c(a_1^{(k)}), ..., c(a_N^{(k)}))$ for $k \in \{1, ..., N!\}$. Let

$$M(\psi_k) = \max \left\{ M \in U : \sum_{j=1}^{M} c(a_j^{(k)}) < L \right\} \tag{6}$$

and let $\tau_0(\cdot)$ be a function that assigns to each permutation ψ_i a sequence of its $M(\psi_i)$ first units. Hence $c(a_1^{(k)}) + ... + c(a_{M(\psi_k)}^{(k)}) < L$ and $c(a_1^{(k)}) + ... + c(a_{M(\psi_k)+1}^{(k)}) \geq L$. Moreover, let $\tau_1(\cdot)$ be a reduction function that assigns to any ordered sequence $O = (o_1, ..., o_z)$ an unorderet set $\tau_1(O) = \{o_1, ..., o_z\}$

of its elements. Let $\tau(\cdot) = \tau_1(\tau_0(\cdot))$ represent the composition of τ_1 and τ_0. Hence for any $\psi \in \Psi$, $\tau(\psi)$ represents an unordered sample corresponding to ψ. Let $\Psi_i = \{\psi \in \Psi : a_i \in \tau(\psi)\}$ and let $\Psi_{ij} = \{\psi \in \Psi : a_i, a_j \in \tau(\psi)\}$ for any $i \neq j \in U$. Inclusion probabilities of first two orders may be expressed as

$$\pi_i = \frac{\#\Psi_i}{\#\Psi} \tag{7}$$

and

$$\pi_{ij} = \frac{\#\Psi_{ij}}{\#\Psi} \tag{8}$$

for $i \neq j \in U$ while $\#\Psi = N!$. Calculation of inclusion probabilities through (7) and (8) requires enumerating all possible permutations of U and in most cases is not feasible. However several useful properties of these probabilities may be proven. They are formally stated in following four propositions.

Proposition 1. *If the sample is drawn from a finite population using the Pathak scheme with individual costs $c(1), ..., c(N)$ and if costs of some i-th and i'-th unit $(i, i' \in U)$ satisfy $c(i) \geq c(i')$ then $\pi_i \leq \pi_{i'}$.*

Proof. Let $\psi_k = (a_1^{(k)}, ..., a_N^{(k)}) \in \Psi_i$ and let $a_g^{(k)} = i$ and $a_{g'}^{(k)} = i'$. Let $\psi_k^* = (a_1^{*(k)}, ..., a_N^{*(k)})$ be a permutation of elements in U obtained by exchanging the g-th element and g'-th element in ψ_k, so that $a_g^{*(k)} = i'$ and $a_{g'}^{*(k)} = i$ while $a_z^{*(k)} = a_z^{(k)}$ for $z \in \{1, ..., N\} - \{g, g'\}$. Let $\Psi_i^* = \{\psi_k^* : \psi_k \in \Psi_i\}$. For any $\psi_k \in \Psi_i$, recalling that $c(i) \geq c(i')$ we have: $c(a_1^{*(k)}) + ... + c(a_{M(\psi_k)}^{*(k)}) \leq c(a_1^{(k)}) + ... + c(a_{M(\psi_k)}^{(k)})$ and consequently $M(\psi_k^*) \geq M(\psi_k)$. Meanwhile, among first $M(\psi_k)$ elements in ψ_k^*, one is certainly equal to i' and hence $i' \in \tau(\psi_k^*)$. Consequently $\Psi_i^* \subseteq \Psi_{i'}$ and $\#\Psi_i = \#\Psi_i^* \leq \#\Psi_{i'}$. So from (7) we have $\pi_i \leq \pi_{i'}$. □

Proposition 2. *If the sample is drawn from a finite population using the Pathak scheme with individual costs $c(1), ..., c(N)$ and if costs of some i-th, j-th, i'-th and j'-th unit $(i \neq j, i' \neq j' \in U)$ satisfy $c(i) \geq c(i')$ and $c(j) \geq c(j')$ then $\pi_{ij} \leq \pi_{i'j'}$.*

Proof. Let $\psi_k = (a_1^{(k)}, ..., a_N^{(k)}) \in \Psi_{ij}$ and let $a_g^{(k)} = i$, $a_{g'}^{(k)} = i'$, $a_h^{(k)} = j$, $a_{h'}^{(k)} = j'$. Let $\psi_k^* = (a_1^{*(k)}, ..., a_N^{*(k)})$ be a permutation of elements in U obtained by exchanging the g-th and g'-th element, then exchanging h-th and h'-th element in ψ_k, so that $a_g^{*(k)} = i'$, $a_{g'}^{*(k)} = i$, $a_h^{*(k)} = j'$, $a_{h'}^{*(k)} = j$ while $a_z^{*(k)} = a_z^{(k)}$ for $z \in \{1, ..., N\} - \{g, h, g', h'\}$. Let $\Psi_{ij}^* = \{\psi_k^* : \psi_k \in \Psi_{ij}\}$. For any $\psi_k \in \Psi_{ij}$, recalling that $c(i) \geq c(i')$ and $c(j) \geq c(j')$ we have: $c(a_1^{*(k)}) + ... + c(a_{M(\psi_k)}^{*(k)}) \leq c(a_1^{(k)}) + ... + c(a_{M(\psi_k)}^{(k)})$ and consequently $M(\psi_k^*) \geq M(\psi_k)$. Meanwhile, among first $M(\psi_k)$ elements in ψ_k^*, one element is certainly equal to i' and one element is certainly equal to j' and hence $i', j' \in \tau(\psi_k^*)$. Consequently, $\Psi_{ij}^* \subseteq \Psi_{i'j'}$ and $\#\Psi_{ij} = \#\Psi_{ij}^* \leq \#\Psi_{i'j'}$. From (8) we have $\pi_{ij} \leq \pi_{i'j'}$. □

Proposition 3. *If the sample is drawn from a finite population using the Pathak scheme with individual costs $c(1), ..., c(N)$ and if costs of some i-th and i'-th unit $(i, i' \in U)$ satisfy $c(i) = c(i')$ then $\pi_i = \pi_{i'}$.*

Proof. If $c(i) = c(i')$ then it is true that $c(i) \geq c(i')$ and $c(i) \leq c(i')$. Applying twice proposition 1 we conclude that $\pi_i \leq \pi_{i'}$ and at the same time $\pi_i \geq \pi_{i'}$. Both conditions may simultaneously be true only when $\pi_i = \pi_{i'}$ □

Proposition 4. *If the sample is drawn from a finite population using the Pathak scheme with individual costs $c(1), ..., c(N)$ and if costs of some i-th, j-th, i'-th and j'-th unit $(i \neq j, i' \neq j' \in U)$ satisfy $c(i) = c(i')$ and $c(j) = c(j')$ then $\pi_{ij} = \pi_{i'j'}$.*

Proof. If $c(i) = c(i')$ and $c(j) = c(j')$ then it is also true that $c(i) \geq c(i')$ and $c(j) \geq c(j')$. Applying proposition 2 we get $\pi_{ij} \leq \pi_{i'j'}$. However from assumptions $c(i) = c(i')$ and $c(j) = c(j')$ we may also infer that $c(i) \leq c(i')$ and $c(j) \leq c(j')$, and applying again proposition 2 we get $\pi_{ij} \geq \pi_{i'j'}$. Conditions $\pi_{ij} \leq \pi_{i'j'}$ and $\pi_{ij} \geq \pi_{i'j'}$ are simultaneously satisfied only when $\pi_{ij} = \pi_{i'j'}$. □

Without a loss of generality, let us assume that the finite population is ordered in such a way that $c(1) \leq c(2) \leq ... \leq c(N)$. Proposition 1 leads to a conclusion that first-order inclusion probabilities must satisfy a simple order: $\pi_1 \geq \pi_2 \geq ... \geq \pi_N$. Moreover, according to Proposition 2 second-order inclusion probabilities must satisfy $\pi_{ij} \geq \pi_{i'j'}$ whenever $i \leq i'$ and $j \leq j'$ for $i \neq j, i' \neq j' \in U$. The resulting

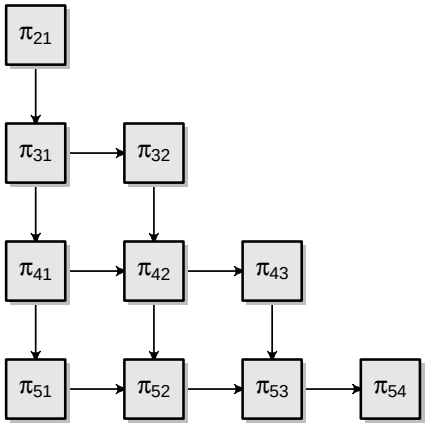


Figure 1 Cover graph for a population of $N = 5$ elements

system of inequalities is illustrated by the Hasse diagram (cover graph) shown in the figure 1. There is no guarantee that Fattorini's estimators of inclusion probabilities satisfy the whole system of inequalities dictated by propositions 1 and 2. Forcing these constraints to be satisfied may improve the accuracy of population total estimates. To preserve constraints one may consider calculating estimates of first-order inclusion probabilities as solutions to the quadratic program

$$\begin{cases} f_1(\hat{\pi}_1, ..., \hat{\pi}_N) \to min \\ \hat{\pi}_1 \geq ... \geq \hat{\pi}_N \\ \hat{\pi}_i \in <0, 1>, i \in U \end{cases} \tag{9}$$

where

$$f_1(\hat{\pi}_1, ..., \hat{\pi}_N) = \sum_{i \in U} R_i (\hat{\hat{\pi}}_{iF} - \hat{\pi}_i)^2 \tag{10}$$

while weights $R_1, ..., R_N$ represent numbers of replications observed for individual units. They will take the same value R when all individual unit costs values are unique, but according to the proposition 3 they may also equal multiples of R in groups of units with exactly the same costs. Let $a^* = a + 1$ for $a = i, j$. Estimates of second-order probabilities are obtained by solving a separate quadratic program:

$$\begin{cases} f_2(\hat{\pi}_{ij}; j < i) \to min \\ \hat{\pi}_{ij} \geq \hat{\pi}_{ij^*}; \;\; j, j^* < i \in U \\ \hat{\pi}_{ij} \geq \hat{\pi}_{i^* j}; \;\; j < i, i^* \in U \\ \hat{\pi}_{ij} \in <0, 1>; \;\; j < i \in U \end{cases} \tag{11}$$

where

$$f_2(\hat{\pi}_{ij}, j < i) = \sum_{j < i \in U} R_{ij} (\hat{\pi}_{ijF} - \hat{\pi}_{ij})^2 \tag{12}$$

while R_{ij} for $i > j \in U$ represent numbers of replications observed for i-th and j-th unit. They will all be equal to R when all unit costs are unique and may be multiples of R in groups of units featuring exactly the same individual costs, on the basis of the proposition 4. Programs (9) and (11) are formulated in terms of inclusion probability estimates $\hat{\pi}_{ij}$ for $i = j$ (first order) and $i < j$ (second order), but knowing that $\pi_{ij} = \pi_{ji}$ for $i, j \in U$ one in fact estimates all inclusion probabilities of first two orders in the whole population. Simulation results applying to units both included and not included in the sample s may then be applied to improve accuracy of finite population totals. It should also be noted that one might also formulate additional constraints involving both first and second inclusion probabilities. Such constraints would however be easier to satisfy and hence less useful to constuct estimates. Both programs may be solved using the active set method discussed in [4]. Plugging obtained estimates $\hat{\pi}_i$ and $\hat{\pi}_{ij}$ into (2) instead of π_i and π_{ij} (for $i, j \in U$) one may construct the variance estimator for (1) in the form:

$$\hat{V}(\hat{t}_*) = \sum_{i,j \in s} \hat{\hat{y}}_i \hat{\hat{y}}_j \hat{\hat{\Delta}}_{ij} \tag{13}$$

where $\hat{\hat{y}}_i = y_i / \hat{\pi}_i$, $\hat{\hat{\Delta}}_{ij} = \hat{\Delta}_{ij} / \hat{\pi}_{ij}$ and $\hat{\Delta}_{ij} = \hat{\pi}_{ij} - \hat{\pi}_i \hat{\pi}_j$ for $i, j \in U$.

3 Numerical illustration

Assume that the finite population consists of $N = 24$ units. Individual costs are given by a vector $\mathbf{c} = [c(1), ..., c(N)]' = [1, 2, ..., 24]'$ and the budget constraint is $L = 90$ which constitutes 30% of the census cost. A simulation involving $R = 500$ sample replications was carried out to estimate unknown inclusion probabilities under Pathak sampling. Estimates of the $N \times N$ inclusion probability matrix $\Pi = [\pi_{ij}]$ (where $\pi_{ii} = \pi_i$ for $i \in U$) obtained using Fattorini's formulae (4), (5), and obtained by solving (9) and (11) through active set method are shown in figure 2. Visual inspection reveals that the introduction of hierarchical constraints reduces volatility of inclusion probability estimates.

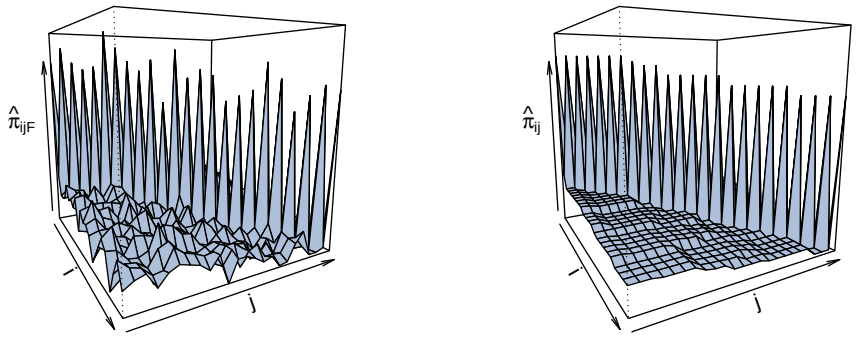


Figure 2 Estimated matrices of inclusion probabilities

To sched some light on properties of the proposed variance estimator a simulation was carried out for values of $\mathbf{c}$ and L introduced above. A total of 5000 realizations of the sample s were drawn with $R = 500$ replications generated for each drawn sample. Two variants of the estimator (3) were computed. The first one denoted by the letter F utilized raw Fattorini's estimates of inclusion probabilities given by (4) and (5). The second one denoted by the letter A utilized inclusion probability estimates obtained by solving (9) and (11) through the active set method. Table 1 shows percentages P_F and P_A of negative variance estimates, their ratio P_A/P_F as well as ratios of variances V_A/V_F and mean square errors M_A/M_F of both estimators, for several value sets $\mathbf{y} = [y_1, ..., y_N]'$ of the study variable.

$\mathbf{y}$	P_F	P_A	P_A/P_F,	V_A/V_F	M_A/M_F
$[1, 2, ..., 24]'$	0.3695	0.2134	0.5776	0.4329	0.5028
$[1, ..., 12, 12, ..., 1]'$	0.0447	0.0026	0.0581	0.6647	0.7269
$[12, ..., 1, 1, ..., 12]'$	0.1365	0.0544	0.3985	0.7076	0.7537
$[24, 23, ..., 1]'$	0.0054	0.0003	0.0645	0.9853	1.0008

Table 1 Selected characteristics of variance estimators' empirical distributions

The results suggest that the proposed estimation method may lead to a substantial reduction in the number of unwelcome negative estimates as compared to Fattorini's approach. Moreover, in three out four investigated cases the variance was considerably reduced and the bias introduced through constraints did not overwhelm the variance reduction effect, which led to a reduced mean squre error as well. In the fourth case both variance and mean square error were comparable.

4 Conclusion

Due to very small expected sample size in the presented example the bias and standard deviation of variance estimates for the estimator $\hat{t}_*$ were large in comparison to its true variance. This had to result in high frequencies of negative variance estimates. In practice samples would be much larger and these frequencies would be very small. Nevertheless, incorporation of hierarchical constraints in the estimation process seems to mitigate risks of such unwelcome developments. The variance reduction is also a benefit. After modifications the proposed approach might potentially be adopted to assess the variance of estimates for population parameters other than the total, such as covariances, ratios, or price indices of [1]. The

main practical challenge seems to be associated with computational complexity of the method as the number of constraints grows with the square of the population size. This might be tackled by solving proposed quadratic programs only for some subset of population units (in particular by collapsing unit groups on the basis of propositions 3 and 4), using parallel computing and/or adopting scalable quadratic programming techniques such as those in [3].

Acknowledgements

The work was supported by the grant No N N111 558540 from the Polish National Science Centre.

References

[1] Białek, J.: Proposition of a General Formula for Price Indices, *Communications in Statistics - Theory and Methods* **41**(2012), 943-952.

[2] Boistard, H., and Lopuhaä, H. P., and Ruiz-Gazen, A.: Approximation of rejective sampling inclusion probabilities and application to high order correlations. *Electronic Journal of Statistics*, **6** (2012), 1967–1983.

[3] Dostál, Z., Horák, D.: Theoretically suported scalable FETI for numerical solution of variational inequalities, *SIAM Journal of Numerical Analysis*, **45** (2007), 500-513.

[4] de Leeuw, J., and Hornik, K., and Mair, P.: Isotone optimization in R: Pool-adjacent-violators algorithm (PAVA) and active set methods, *Journal of Statistical Software* **32** (2009), 1-24.

[5] Fattorini, L., and Ridolfi, G.: A sampling design for areal units based on spatial variability, *Metron* **55** (1997), 59–72.

[6] Fattorini, L.: Applying the Horvitz-Thompson criterion in complex designs: A computer-intensive perspective for estimating inclusion probabilities, *Biometrika* **93** (2006), 269-278.

[7] Fattorini, L.: An adaptive algorithm for estimating inclusion probabilities and performing the Horvitz-Thompson criterion in complex designs, *Computational Statistics* **24** (2009), 623-639.

[8] Gamrot W.: Simulation-assisted Horvitz-Thompson statistic and isotonic regression. In: *Proceedings of 30th International Conference Mathematical Methods in Economics 2012*, (J. Ramik, D. Stavarek eds.). Silesian University of Opava, School of Business Administration, Karvina, 2012, 207-212.

[9] Gamrot, W.: On exact computation of minimum sample size for restricted estimation of a binomial parameter, *Journal of Statistical Planning and Inference* **143** (2013), 852-866.

[10] Kremers, W.K.: Completeness and Unbiased Estimation for Sum-Quota Sampling, *Journal of the American Statistical Association* **81** (1986), 1070-1073.

[11] Pathak, P. K.,: Unbiased estimation in fixed-cost sequential sampling schemes, *Annals of Statistics* **4** (1976), 1012-1017.

[12] Royall, R. M.: The linear least square prediction approach to two-stage sampling, *Journal of the American Statistical Association* **71** (1976), 657–664.

[13] Thompson, M. E., and Wu, C.: Simulation-based randomized systematic PPS sampling under substitution of units, *Survey Methodology*, **34** (2008), 3–10.

[14] Wolny-Dominiak, A.: Modeling of claim counts using data mining procedures in R CRAN. In: *Proceedings of 30th International Conference Mathematical Methods in Economics 2012*, (J. Ramik, D. Stavarek eds.). Silesian University of Opava, School of Business Administration, Karvina, 2012, 980-985.

[15] Wywiał, J. L.: On conditional sampling strategies. *Statistical Papers* **44** (2003), 397–419.

[16] Yu, Y.: On the inclusion probabilities in some unequal probability sampling plans without replacement. *Bernoulli* **18** (2012), 279–289.

[17] Żądło, T.: On MSE of EBLUP, *Statistical Papers* **50** (2009), 101–118.

An Alternative Approach to the Structure Determination of Hierarchical Archimedean Copulas

Jan Górecki[1], Martin Holeňa[2]

Abstract. Copulas offer very a flexible tool for a stochastic dependence modeling. One of the most popular classes of copulas is the class of hierarchical Archimedean copulas, which gained its popularity due to the fact that the models from the class are able to model the stochastic dependencies conveniently even in high dimensions. One critical issue when estimating a hierarchical Archimedean copula is to correctly determine its structure. The paper describes an approach to the problem of the structure determination of a hierarchical Archimedean copula, which is based on the close relationship of the copula structure and the values of measure of concordance computed on all its bivariate margins. The presented approach is conveniently summarized as a simple algorithm.

Keywords: hierarchical Archimedean copula, structure determination, measure of concordance, bivariate margins, nesting condition

JEL classification: C51, C46
AMS classification: 62H99

1 Introduction

Hierarchical Archimedean copulas (HACs), which generalize Archimedean copulas (ACs), overcome some limitations and bring some advantages compared to the most popular class of Gaussian copulas [2]. There already emerged successful applications of HACs in finance, e.g., in collateral debt obligation pricing, see [2, 5]. One critical issue when estimating HAC is to properly determine its structure. Despite the popularity of HACs, there exists only one paper [9] addressing generally the structure determination. The method presented in that paper mainly focus on maximum likelihood estimation (MLE) for the estimation of HAC's parameters, which are later used for the structure determination. The MLE used in the method involves the computation of the density of a HAC that needs up to d derivatives, where d is the data dimension. The authors claim, that the approach is feasible in high dimensions when using numerical method for the density computation and present two examples for $d = 5$, which involves only homogeneous HAC and which incorporates ACs belonging to one Archimedean family. Our approach provides an alternative way to the problem, which completely avoids the need of the HAC's density computation for some Archimedean families, hence is feasible even in very high dimensions.

The paper is structured as follows. The second section recalls some necessary theoretical concepts concerning copulas, the third section presents the proposed approach to the structure determination of HAC and the fourth section concludes the paper.

2 Preliminaries

2.1 Copulas

Definition 1. For every $d \geq 2$, a *d-dimensional copula* (shortly, *d-copula*) is a d-variate distribution function on $\mathbb{I}^d$ ($\mathbb{I}$ is unit interval), whose univariate margins are uniformly distributed on $\mathbb{I}$.

At the first look, copulas (denote the set of all copulas as $\mathcal{C}$) form one of many classes of joint distribution functions (shortly, joint d.f.s). What makes copulas interesting is that they establish a

[1]SBA in Karvina, Silesian University in Opava, Karvina, Czech Republic, Department of informatics, gorecki@opf.slu.cz
[2]Institute of Computer Science, Academy of Sciences of the Czech Republic, Praha, Czech Republic, martin@cs.cas.cz

connection between general joint d.f. and its univariate margins (in text below we use only *margin* for term *univariate margin*).

Theorem 1. ***(Sklar's Theorem)*** *[10] Let H be a d-dimensional d.f. with margins $F_1, ..., F_d$. Let A_j denote the range of F_j, $A_j := F_j(\overline{\mathbb{R}})(j = 1, ..., d), \overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$. Than there exists a copula C such for all $(x_1, ..., x_d) \in \overline{\mathbb{R}}^d$,*

$$H(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)). \tag{1}$$

Such a C is uniquely determined on $A_1 \times ... \times A_d$ and, hence, it is unique if $F_1, ..., F_d$ are all continuous.

Through the Sklar's theorem, one can derive for any d-variate d.f. its copula C using (1). In case that the margins $F_1, ..., F_d$ are all continuous, copula C is given by $C(u_1, ..., u_d) = H(F_1^-(u_1), ..., F_d^-(u_d))$, where $F_i^-, i \in \{1, ..., d\}$ denotes pseudo-inverse of F_i given by $F_i^-(s) = \inf\{t|\ F_i(t) \geq s\}, s \in \mathbb{I}$. Many classes of copulas are derived in this way from popular joint d.f.s, e.g., the most popular class of Gaussian copulas is derived using H corresponding to d-variate Gaussian distribution. But, using this process often results in copula forms not representable in closed form, what can bring difficulties in some applications.

2.2 Archimedean Copulas

This drawback is overcame while using Archimedean copulas due to their different construction process. ACs are not constructed using thr Sklar's theorem, but instead of it, one starts with a given functional form and asks for properties in order to obtain a proper copula. As a result of such a construction, ACs are always expressed in closed form, which is one of the main advantages of this class of copulas [3]. To construct ACs we need a notion of an *Archimedean generator* and a *complete monotonicity*.

Definition 2. *Archimedean generator* (shortly, *generator*) is continuous, nonincreasing function $\psi : [0, \infty] \to [0, 1]$, which satisfies $\psi(0) = 1, \psi(\infty) = \lim_{t\to\infty} \psi(t) = 0$ and is strictly decreasing on $[0, \inf\{t : \psi(t) = 0\}]$.

Remark 1. We denote set of all generators as Ψ.

Definition 3. Function f is called *completely monotone* (shortly, c.m.) on $[a, b]$, if $(-1)^k f^{(k)}(x) \geq 0$ holds for every $k \in \mathbb{N}_0, x \in (a, b)$.

Definition 4. Any d-copula C is called *Archimedean copula* (we denote it d-AC), if it admits the form

$$C(\mathbf{u}) := C(\mathbf{u}; \psi) := \psi(\psi^{-1}(u_1) + ... + \psi^{-1}(u_d)), \mathbf{u} \in \mathbb{I}^d, \tag{2}$$

where $\psi \in \Psi$ and its inverse $\psi^{-1} : [0, 1] \to [0, \infty]$ is defined $\psi^{-1}(0) = \inf\{t : \psi(t) = 0\}$.

For verifying whether function C given by (2) is a proper copula, we can use the property stated in Definiton 3. A condition sufficient[1] for C to be a copula is stated as follows.

Theorem 2. *If $\psi \in \Psi$ is completely monotone, then function C given by (2) is copula.*

We can see from Definition 4 that having a random vector $\mathbf{U}$ distributed according to some AC, all its k-dimensional ($k < d$) marginal copulas have the same marginal distribution. It implies that all multivariate margins of the same dimension are equal, thus, e.g., the dependence among all pairs of components is identical. This symmetry of ACs is often considered to be a rather strong restriction, especially in high dimensional applications.

2.3 Hierarchical Archimedean Copulas

To allow for asymmetries, one may consider the class of HACs[2], recursively defined as follows.

Definition 5. A d-dimensional copula C is called *hierarchical Archimedean copula* if it is an AC with arguments possibly replaced by other hierarchical Archimedean copulas. If C is given recursively by (2) for $d = 2$ and

$$C(\mathbf{u}; \psi_0, ..., \psi_{d-2}) = \psi_0(\psi_0^{-1}(u_1) + \psi_0^{-1}(C(u_2, ..., u_d; \psi_1, ..., \psi_{d-2}))), \mathbf{u} \in \mathbb{I}^d, \tag{3}$$

[1] Necessary and sufficient condition for C to be a copula can be found in [6]

[2] often also called *nested Archimedean copulas*

for $d \geq 3$, C is called *fully-nested hierarchical Archimedean copula* with $d-1$ *nesting levels*. Otherwise C is called *partilally-nested hierarchical Archimedean copula*. [4]

Remark 2. We denote a d-dimensional HAC as d-HAC. We refer to the hierarchical ordering of $C(\cdot;\psi_0), ..., C(\cdot;\psi_{d-2})$ together with the ordering of variables $u_1, ..., u_d$ as the *structure* of a d-HAC.

From the definition, we can see that ACs are special cases of HACs. The most simple proper fully-nested HAC is copula C obtained for $d = 3$ with two nesting levels. The structure of this copula is given by

$$\begin{aligned} C(\mathbf{u};\psi_0,\psi_1) &= C(u_1, C(u_2,u_3;\psi_1);\psi_0) \\ &= \psi_0(\psi_0^{-1}(u_1) + \psi_0^{-1}(\psi_1(\psi_1^{-1}(u_2) + \psi_1^{-1}(u_3)))), \mathbf{u} \in \mathbb{I}^3. \end{aligned} \tag{4}$$

As in the case of ACs we can ask for necessary and sufficient condition for function C given by (3) to be a proper copula. Partial answer for this question in form of sufficient condition is contained in the following theorem [6].

Theorem 3. ***(McNeil (2009))***. *If $\psi_j \in \Psi_\infty, j \in \{0, ..., d-2\}$ such that $\psi_k^{-1} \circ \psi_{k+1}$ have completely monotone derivatives for all $k \in \{0, ..., d-3\}$, then $C(\boldsymbol{u};\psi_0, ..., \psi_{d-2}), \boldsymbol{u} \in \mathbb{I}^d$, given by (3) is a copula.*

If we take the most simple 3-HAC given by (4), we can see that the condition for C to be a proper copula following from McNeil's theorem is $(\psi_0^{-1} \circ \psi_1)'$ to be completely monotone. As this condition will be essential for the rest of this paper we put it in individual definition.

Definition 6. Let $\psi_a, \psi_b \in \Psi_\infty, a, b \in \{0, ..., d-2\}, a \neq b$ and $C(\cdot;\psi_a)$ corresponds to parent of $C(\cdot;\psi_b)$ in the tree structure of C. Then condition for $(\psi_a^{-1} \circ \psi_b)'$ to be compete monotone is called *nesting condition*.

As we can observe, verification of conditions in McNeil's theorem is just $d-2$ verifications of nesting condition for $d-2$ different pairs $\psi_k, \psi_{k+1}, k \in \{0, ..., d-2\}$. McNeil's theorem is stated only for fully-nested HACs, but it can be easily translated also for use with partially-nested HACs.

For the sake of simplicity, assume that each d-HAC structure corresponds to some binary tree t. Each node in t represents one 2-AC. Each 2-AC is determined just by its corresponding generator, so we identify each node in t with one generator and hence we have always nodes $\psi_0, ..., \psi_{d-2}$. For a node ψ denote as $\mathcal{D}_n(\psi)$ the set of all descendant nodes of ψ, $\mathcal{P}(\psi)$ the parent node of ψ, $\mathcal{H}_l(\psi)$ the left child of ψ and $\mathcal{H}_r(\psi)$ the right child of ψ. The leafs of t correspond to the variables $u_1, ..., u_d$.

2.4 Measure of concordance

A measure of concordance (MoC) is a measure, which reflects a degree of dependency between two random variables independently on their univariate distributions. There also exist generalizations for more than two random variables, but we present only pairwise measure of concordance. As $\mathcal{C}$ allows for partial ordering known as *concordance ordering*, a measure of concordance also reflects this ordering (see [3, 7]). One of the most popular measures of concordance is *Kendall's tau*. As we are interested in its relationship with a general bivariate copula, we use its the definition given by (as in [1])

$$\tau(C) = 4 \int_{\mathbb{I}^2} C(u_1,u_2) dC(u_1,u_2) - 1. \tag{5}$$

If C is 2-AC based on a generator ψ and ψdepends on the parameter θ, then (5) states a relationship between θ and τ. This relationship is very important for our approach and is used extensively later in Section 3.

2.5 Okhrin's algorithm for the structure determination of HAC

We recall the algorithm presented in [8] for the structure determination of HAC, which returns for some unknown HAC C its structure using only the known forms of its bivariate margins. The algorithm uses the following definition.

Definition 7. Let C be a d-HAC with generators $\psi_0, ..., \psi_{d-2}$ and $(U_1, ..., U_d) \sim C$. Then denote as $\mathcal{U}_C(\psi_k), k = 0, ..., d-2$, the set of indexes $\mathcal{U}_C(\psi_k) = \{i | (\exists U_j)(U_i, U_j) \sim C(\cdot;\psi_k) \vee (U_j, U_i) \sim C(\cdot;\psi_k), 1 \leq i < j \leq d\}, k = 0, ..., d-2$.

Proposition 4. *Defining $\mathcal{U}_C(u_i) = \{i\}$ for the leaf $i, 1 \leq i \leq d$, there is an unique disjunctive decomposition of $\mathcal{U}_C(\psi_k)$ given by*

$$\mathcal{U}_C(\psi_k) = \mathcal{U}_C(\mathcal{H}_l(\psi_k)) \cup \mathcal{U}_C(\mathcal{H}_r(\psi_k)). \tag{6}$$

Due to space limitations we do not state the proof for the proposition and we refer the reader to the Okhrin's work [8], which includes detailed description of the method and the necessary proofs.

For an unknown d-HAC C, knowing all its bivariate margins, its structure can be easily determined with Algorithm 1, which returns the unknown structure t of C. We start from the sets $\mathcal{U}_C(u_1), ..., \mathcal{U}_C(u_d)$ joining them together through (6) until we reach the node ψ for which $\mathcal{U}_C(\psi) = \{1, ..., d\}$.

Algorithm 1 The HAC structure determination

$\mathcal{I} = \{0, ..., d-2\}$
while $\mathcal{I} \neq \emptyset$ **do**
1. $k = \text{argmin}_{i \in \mathcal{I}}(\#\mathcal{U}_C(\psi_i))$, if there are more minima, then choose as k one of them arbitrarily.
2. Find the nodes ψ_l, ψ_r, for which $\mathcal{U}_C(\psi_k) = \mathcal{U}_C(\psi_l) \cup \mathcal{U}_C(\psi_r)$.
3. $\mathcal{H}_l(\psi_k) := \psi_l, \mathcal{H}_r(\psi_k) := \psi_r$.
4. Set $\mathcal{I} := \mathcal{I} \backslash \{k\}$.

end while

3 Our approach

Recalling Theorem 3, the sufficient condition for C to be a proper copula is, that the nesting condition must hold for each generator and its parent in a HAC structure. As this is the only known condition that assures that C is a proper copula, we concern in this work only the copulas, which fulfill this condition. The nesting condition results in constraints for the parameters θ_0, θ_1 of the involved generators ψ_0, ψ_1 (see [4, 3]). As $\theta_i, i = 1, 2$ is closely related to a MoC, e.g. τ and θ_i relationship established through (5), there is also an important relationship between the MoC and the HAC tree structure following from the nesting condition. This relationship is described for the fully-nested 3-HAC given by the form (4) in Remark 2.3.2 in [3]. There is stated that if the nesting condition holds for the parent-child pair (ψ_0, ψ_1), then $0 \leq \kappa(\psi_0) \leq \kappa(\psi_1)$, where κ is a MoC (as we concern only HACs with binary structures, which incorporates only 2-ACs, which are fully determined only by its generator, we use as domain of κ the set Ψ instead of the usually used set of all 2-copulas). We generalize this statement, using our notion, as follows.

Proposition 5. *Let C be a d-HAC with the structure t and the generators $\psi_0, ..., \psi_{d-2}$, where each parent-child pair satisfy the nesting condition. Let κ be a MoC. Than $\kappa(\psi_i) \leq \kappa(\psi_j)$,* where *$\psi_j \in \mathcal{D}_n(\psi_i)$, holds for each $\psi_i, i = 0, ..., d-2$.*

Proof. If $\psi_i = \mathcal{P}(\psi_j)$, then we get directly $\kappa(\psi_i) \leq \kappa(\psi_j)$ using Remark 2.3.2 from [3]. Otherwise, as $\psi_j \in \mathcal{D}(\psi_i)$, there exists a unique sequence $\psi_{k_1}, ..., \psi_{k_l}$, where $0 \leq k_m \leq d-2, m = 1, ..., l, l \leq d-1, \psi_{k_1} = \psi_i, \psi_{k_l} = \psi_j$ and $\psi_{k-1} = \mathcal{P}(\psi_k)$ for $k = 2, ..., l$. Applying the above mentioned remark for each pair $(\psi_{k-1}, \psi_k), k = 2, ..., l$, we get $\kappa(\psi_{k_1}) \leq ... \leq \kappa(\psi_{k_l})$. □

Thus, having a branch from t, all its nodes are uniquely ordered according to their value of κ assuming unequal values of κ for all parent-child pairs. This provides us an alternative algorithm for the HAC structure determination. We have to assign the generators with the highest values of κ to the lowest levels of the branches in the structure and ascending to higher levels we assign the generators with lower values of κ.

To allow for computation of MoC among m (possibly > 2) random variables (r.v.s) we state the following definition. For simplification, denote the set of pairs of r.v.s as $\mathbf{U}_{IJ} = \{(U_i, U_j) | (i, j) \in I \times J\}$, where $I, J \subset \mathbb{N}, I \neq \emptyset \neq J$.

Definition 8. Let $m \in \mathbb{N}$ and κ be a MoC. Then define an *aggregated* MoC κ^+ as

$$\kappa^+(\mathbf{U}_{IJ}) = \begin{cases} \kappa(U_i, U_j) & \text{if } I = \{i\}, J = \{j\} \\ +(\kappa(U_i, U_j)_{i \in I, j \in J}), & \text{else,} \end{cases} \tag{7}$$

where the non-empty sets $I, J \subset \{1, ..., m\}, I \cap J = \emptyset$ and $+$ denotes an aggregation function[3], for which $+(x, ..., x) = x$ for all $x \in \mathbb{I}$.

Remark 3. $\kappa(\psi_k) = \kappa^+(\mathbf{U}_{\mathcal{U}_C(\mathcal{H}_l(\psi_k))\mathcal{U}_C(\mathcal{H}_r(\psi_k))})$ for a d-HAC C and for each $k = 0, ..., d-2$.

Let us illustrate our approach to the structure determination for $d = 4$. Assume three different structures t_1, t_2, t_3 corresponding to copulas C_1, C_2, C_3. For t_1 let $\mathcal{U}_{C_1}(\psi_2) = \mathcal{U}_{C_1}(\psi_1) \cup \mathcal{U}_{C_1}(\psi_0) = \{3, 4\} \cup \{1, 2\}$. For simplification denote $\{3, 4\} \cup \{1, 2\}$ as $((34)(12))$. For t_2 let $\mathcal{U}_{C_2}(\psi_2) = \{u_4\} \cup (\mathcal{U}_{C_2}(\psi_1) \cup \mathcal{U}_{C_2}(\psi_0)) = \{u_4\} \cup (\{u_3\} \cup \{u_1, u_2\}) = (4(3(21)))$. For t_3 let $\mathcal{U}_{C_3}(\psi_0) = (3(4(12)))$. We see that t_1 is the structure of a partially-nested 4-HAC and t_2, t_3 are the structures of fully-nested 4-HACs. Also assume (without a loss of generality) $\kappa(\psi_2) = \alpha, \kappa(\psi_0) = \gamma$ and $\alpha < \kappa(\psi_1) < \gamma, \alpha, \gamma \in \mathbb{I}$ for all t_1, t_2, t_3. The case when $\alpha = \kappa(\psi_1)$ or $\kappa(\psi_1) = \gamma$ is discussed later for a 3-HAC. Denote $\beta_1 = \kappa(\psi_1)$ for t_1, $\beta_2 = \kappa(\psi_1)$ for t_2 and $\beta_3 = \kappa(\psi_1)$ for t_3. The quantities $\alpha, \beta_1, \beta_2, \beta_3, \gamma$ can be determined from corresponding bivariate distributions as for t_1 is $\alpha = \kappa(\psi_2) = \kappa(U_3, U_1) = \kappa(U_3, U_2) = \kappa(U_4, U_1) = \kappa(U_4, U_2), \beta_1 = \kappa(\psi_1) = \kappa(U_3, U_4), \gamma = \kappa(\psi_0) = \kappa(U_1, U_2)$. For t_2 we have $\alpha = \kappa(\psi_2) = \kappa(U_4, U_3) = \kappa(U_4, U_1) = \kappa(U_4, U_2), \beta_2 = \kappa(\psi_1) = \kappa(U_3, U_1) = \kappa(U_3, U_2), \gamma = \kappa(\psi_0) = \kappa(U_1, U_2)$. For t_3 similarly $\alpha = \kappa(\psi_2) = \kappa(U_4, U_3) = \kappa(U_3, U_1) = \kappa(U_3, U_2), \beta_3 = \kappa(\psi_1) = \kappa(U_4, U_1) = \kappa(U_4, U_2), \gamma = \kappa(\psi_0) = \kappa(U_1, U_2)$.

Now assume a 4-HAC C with unknown structure $t \in \{t_1, t_2, t_3\}$ and $(U_1, U_2, U_3, U_4) \sim C$. Compute κ for all pairs of the r.v.s. If follows from the assumptions that $\kappa(U_1, U_2) = \gamma$ is always (for $t = t_1, t_2, t_3$) the maximum from those values. To satisfy Proposition 5, it is necessarily $\mathcal{U}_C(\psi_0) = \{12\}$, what assures through Algorithm 1 that ψ_0 is assigned to the lowest level of a branch from t. We introduce the a new variable $Z = (U_1, U_2)$, which represents r.v.s U_1, U_2. Once again compute κ for all the pairs of the new r.v.s, which are now r.v.s (U_3, U_4, Z). As Z represents two r.v.s we use generalized κ^+. Thus we get $\beta_1 = \kappa^+(U_3, U_4) = \kappa(U_3, U_4), \beta_2 = \kappa^+(U_3, Z) = \kappa^+(\mathbf{U}_{\{3\}\{12\}})$ and $\beta_3 = \kappa^+(U_4, Z) = \kappa^+(\mathbf{U}_{\{4\}\{12\}})$. Consider that under $t = t_1$ is $\beta_1 > \beta_2 = \beta_3 = \alpha$. Under $t = t_2$ is $\beta_2 > \beta_1 = \beta_3 = \alpha$ and under $t = t_3$ is $\beta_3 > \beta_1 = \beta_2 = \alpha$. The determination of $\mathcal{U}_C(\psi_1)$ in accordance with Proposition 5 is then obvious - $\mathcal{U}_C(\psi_1) = \{3, 4\}$ if $\beta_1 = \max(\beta_1, \beta_2, \beta_3)$ or $\mathcal{U}_C(\psi_1) = \{3, 2, 1\}$ if $\beta_2 = \max(\beta_1, \beta_2, \beta_3)$ or $\mathcal{U}_C(\psi_1) = \{4, 2, 1\}$ if $\beta_3 = \max(\beta_1, \beta_2, \beta_3)$. The set $\mathcal{U}_C(\psi_2) = \{4, 3, 2, 1\}$ for all t_1, t_2, t_3.

The described process is generalized in Algorithm 2 for arbitrary $d > 2$. The algorithm returns the sets $\mathcal{U}_C(z_{d+k+1})$ corresponding to the sets $\mathcal{U}_C(\psi_k), k = 0, ..., d-2$. Passing them to Algorithm 1, we avoid their computation from Definition 7 and we get the requested d-HAC structure without a need of knowing the forms of the bivariate margins.

Algorithm 2 The HAC structure determination based on κ

Input:

1) $\mathcal{I} = \{1, ..., d\}$, 2) $(U_1, ..., U_d) \sim C$, 3) κ^+ ... an aggregated MoC, 4) $z_k = u_k, \mathcal{U}_C(z_k) = \{k\}, k = 1, ..., d$

The structure determination:

for $k = 0, ..., d-2$ **do**

1. $(i, j) := \underset{i^* < j^*, i^* \in \mathcal{I}, j^* \in \mathcal{I}}{\operatorname{argmax}} \kappa^+(\mathbf{U}_{\mathcal{U}_C(z_{i^*})\mathcal{U}_C(z_{j^*})})$
2. $\mathcal{U}_C(z_{d+k+1}) := \mathcal{U}_C(z_i) \cup \mathcal{U}_C(z_j)$
3. $\mathcal{I} := \mathcal{I} \cup \{d+k+1\} \backslash \{i, j\}$

end for

Output:

$\mathcal{U}_C(z_{d+k+1}), k = 0, ..., d-2$

Now consider a fully-nested 3-HAC with two equal generators given by the form $C(u_1, C(u_2, u_3; \psi); \psi) = \psi(\psi^{-1}(u_1) + \psi^{-1}(\psi(\psi^{-1}(u_2) + \psi^{-1}(u_3))))$. As it equals to $\psi(\psi^{-1}(u_1) + \psi^{-1}(u_2) + \psi^{-1}(u_3))$, which is the 3-AC $C(u_1, u_2, u_3; \psi)$, we get for this copula in Step 1. of the algorithm three pairs (1,2), (1,3), (2,3) corresponding to the maximal value of κ^+. This is because all bivariate margins of $C(u_1, u_2, u_3; \psi)$ are distributed equally. Choosing the first pair to be the pair (i, j) we get the result of the algorithm as $\mathcal{U}_C(\psi_0) = \{1, 2\}, \mathcal{U}_C(\psi_1) = \{1, 2, 3\}$. Passing it to Algorithm 1 we get the corresponding structure and denote it as r_1. In the same way we obtain for the second and the third pair the structures we denote as r_2, r_3. But, as $C(u_1, C(u_2, u_3; \psi); \psi) = C(u_1, u_2, u_3; \psi)$, all those structures r_1, r_2, r_3 corresponds to the same copula. Thus, in the case that there are more than one pair corresponding to the maximal value of κ^+ in Step 1., we can choose the pair arbitrarily, because it does not affect the resulting copula, i.e.

[3] like, e.g., max, min or mean

the algorithm return different structures, which however correspond to the same copula. This fact can be also easily generalized for the case when $d > 3$.

4 Conclusions

As the aggregated κ^+ depends only on the pairwise κ and the aggregation function $+$, we can easily derive its empirical version κ_n^+ just by substituting κ in κ^+ by its empirical version κ_n, e.g., by empirical version of Kendall's tau. Using κ_n^+ instead of κ^+ we can easily derive the empirical version of the structure determination process represented by Algorithms 1, 2. Conclude that in this way we base the structure determination only on the values of the pairwise MoC. This is the essential property of our approach, because if the relationship between κ and θ established through (5) is explicitly known, whole HAC, including its structure and its parameters, can be estimated just from κ_n computed on the realizations of $(U_i, U_j), 1 \leq i < j \leq d$ completely avoiding the use of the MLE.

Acknowledgements

This work was supported by the project "Inovace studijních programů na Slezské univerzitě, Obchodně podnikatelské fakultě v Karviné, Reg. č. CZ.1.07/2.2.00/28.0017".

References

[1] C. Genest and A. Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Hydrol. Eng.*, 12:347 – 368, 2007.

[2] J. M. Hofert. *Sampling Nested Archimedean Copulas: with Applications to CDO Pricing*. Suedwestdeutscher Verlag fuer Hochschulschriften, 2010.

[3] M. Hofert. *Sampling Nested Archimedean Copulas with Applications to CDO Pricing*. PhD thesis, Ulm University, 2010.

[4] M. Hofert. Efficiently sampling nested archimedean copulas. *Computational Statistics and Data Analysis*, 55(1):57 – 70, 2011.

[5] M. Hofert and M. Scherer. Cdo pricing with nested archimedean copulas. *Quantitative Finance*, 11(5):775–787, 2011.

[6] A. J. McNeil and J. Nešlehová. "multivariate archimedean copulas, d-monotone functions and l1-norm symmetric distributions. *The Annals of Statistics*, 37:3059 – 3097, 2009.

[7] R. Nelsen. *An Introduction to Copulas*. Springer, 2nd edition, 2006.

[8] O. Okhrin, Y. Okhrin, and W. Schmid. Properties of hierarchical archimedean copulas. Technical report, SFB 649 discussion paper, 2009.

[9] O. Okhrin, Y. Okhrin, and W. Schmid. On the structure and estimation of hierarchical archimedean copulas. *Journal of Econometrics*, 173(2):189 – 204, 2013.

[10] A. Sklar. Fonctions de répartition a n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris*, 8:229 – 231, 1959.

Quantitative modelling of consumption patterns in the CEEC

Ladislava Grochová[1], Luboš Střelec[2]

Abstract. This paper proposes an econometric model of consumption function focusing on macroeconomic determinants under the condition of the CEE countries. Hence, the main aim of the paper is verifying selected models of consumption function and subsequently, the suitability of their application to conditions of real economic situation in the CEE countries. The attention is paid to short- and long-run patterns of behaviour applying a vector error correction model (VECM) that is suitable to study abovementioned problems. The long-run relations are estimated using real private consumption, real national disposable income per capita, real interest rates, anticipated inflation, and dummy for the CEE countries in levels and the short-run fluctuations from equilibrium using the same sample of variables are studied in differences. Hall's hypothesis on the impact of wealth on consumption is also examined including M3 aggregate as a proxy of financial wealth. The research is based on unbalanced panel analysis using selected developed countries data covering the period from 1970 to 2011. Since there is a tendency in changing consumer behaviour during 2008 crisis period it is convenient to use the dummy variable in order to prevent from biased results and to obtain interesting information about the topic.
Keywords: Consumption function, CEEC, unbalanced panel

JEL Classification: C01, C33, E21
AMS Classification: 91G70, 91B42, 91B84

1 Introduction

Private consumption represents an indispensable part of aggregate demand in most of the countries. Not only due to this fact private consumption was, is and will be unimpugnably of a considerable interest of economists or politicians. It represents circa two-thirds of the nation's GDP. For example, in the EU private consumption represents a remarkable 55–57% of GDP on average, in the CEEC the ratio of private consumption to GDP is negligibly higher ranging from 58–60% of GDP, which makes this component of aggregate demand extremely important. This is also reason why both theoretical and empirical approaches dealing with private consumption has evolved rapidly in last decades. In particular, among macroeconomic models we can mention the most influential ones as absolute income hypothesis (Keynes [15]), life-cycle hypothesis (Modigliani [18]), and permanent income hypothesis (Friedman [11]). Among fundamental empirical studies a research based on vector error correction model that models long-run relationship between consumption and income considering also short-run distortions from the equilibrium by Davidson, Hendry, Srba and Yeo [8] must be reminded.

The aim of this contribution is then to develop an econometric model of the consumption function for chosen industrial countries capturing the specifics of CEE countries that is in line with theoretical assumptions. For this purpose we will employ vector error correction model to describe the impact of both short- and long-run determinants as real national disposable income per capita, M3 aggregate, real interest rates, anticipated inflation.

The paper is organized as follows: section 2 brings a brief overview on fundamental and recent literature discussing consumption function, section 3 introduces data and methodology used to estimate private consumption in industrial countries, section 4 dissects results and section 5 concludes.

2 Literature review

Since private consumption is considered to be an important component of aggregate demand it is of a great importance in macroeconomic analysis. Most of recent studies are based on the following fundamental theories and models: absolute income hypothesis (Keynes [15]), life-cycle hypothesis (Modigliani [18]), and permanent income hypothesis (Friedman [11]).

Keynes' absolute income hypothesis (further AIH) introduces static explanation of consumption explained by current disposable income. As only current disposable income is taken into account this theory cannot explain

[1] Mendel University in Brno, Faculty of Business and Economics, Department of Economics, Zemědělská 1, 613 00 Brno, ladislava.grochova@mendelu.cz.
[2] Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, lubos.strelec@mendelu.cz.

the trade-off between present and future income. Nevertheless absolute income hypothesis can be regarded as a good approximation for stable economies.

Fisher's idea [10] including inter-temporal aspects into consumption issues is extended by Modigliani [18] who emphasizes the systematic changes of income during consumer life. Thus, consumer can transfer the money in periods in which the income is high to periods with low expected income. This rationale provides the basis for Modigliani's life-cycle hypothesis (further LCH). According to Modigliani a consumer tries to achieve smooth consumption through savings. That is consumption depends on current income, expected life income and wealth.

Also Friedman's permanent income hypothesis (further PIH) is built on Fisher's theory in terms that consumption does not depend solely on current income. Decisions related to consumption stem from both current income and also expected income in the future. Current income consists of permanent income and transitory income. In general, permanent income is expected to be a long-run income persistent into the future while transitory income is considered as a short-run income that fluctuates around average income.

In terms of econometric testing of consumption function is a major work written by Davidson, Hendry, Serb and Yeo [8] (further DHSY approach). This paper examines the relationship between consumption expenditure and disposable income. Data of Great Britain relating to years 1958 and 1976 were used for empirical research. Using VECM the authors try to point out the discrepancy in the average propensity to consume in the short-term and long-term period, to which economists have to face when they are trying to prove or disprove the theory of Keynesian consumption function (AIH), is not only economic, but also a econometric problem. It was found that an econometric model examining the relationship between consumption and disposable income includes a component of autocorrelation. As autocorrelation manifests itself differently in the short and long time series, which can cause a bias in estimates of the average propensity to consume, authors suggest dynamisation of the model that is in line with the assumptions of LCH and PIH. However, the authors admit that this model does not have to correctly catch the relationship between consumption and disposable income due to many important questions that remain unanswered, such as changes in the distribution of incomes or the direct effect of wealth.

This shortcut inspired Hall [13] who tested the effect of disposable income on consumption and then the effect of wealth on consumption. He links rational expectation hypothesis (further REH) with PIH tests it on time series of the United States in the postwar years. In his work, he concludes that consumption represents a process of random walk monitoring a particular trend. This is due to the fact that consumers try continually to optimize their benefits. Therefore, their wealth develops randomly. It was confirmed that real disposable income has no relation to the formation of future expectations about consumption unless unexpected event occurs. Hall's model finds its support for countries with developed financial markets allowing consumer in position of lender or borrower to maintain a certain level of consumption.

Consequently, the DSHY and Hall's approach has been adopted by many authors like Davis [9], Molana [19] and Chambers [6] who conclude that the model by Davidson, Hendry, Serb and Yeo represents the best specification for United Kingdom, or by Byrne and Davis [2] focusing on variables to be included in the consumption function of G7 countries.

Following the ideas of Hall [13] and basing the model on DSHY approach Byrne and Davis [2] claim that consumption function based on aggregate wealth is somehow instable. Their improvement can be seen in a distinction of liquid and illiquid financial wealth. Using quarterly data over 1972-1998 and SUR estimation of LCH they demonstrate that for G7 countries liquid financial wealth is less important long-term factor determining consumption than illiquid financial wealth.

Among advocates of Keynes' AIH can be listed Campbell and Mankiw [3] that approved the validity of AIH in the US, while Arlt et al. [1] find an evidence of AIH in conditions of Czech economy. In particular, Campbell and Mankiw [3] revise the PIH showing that US consumers violent this hypothesis. They use American postwar data and find that about 50% of consumers in the United States behave according to the Keynesian AIH which means that their consumption decisions depends on their current income rather than permanent income.

The aspects of the consumption function in conditions of the CEE countries were considered in the study written by Arlt et al. [1], Castiglione et al. [5] or Grochová [12] for the Czech Republic or Leszkiewicz-Kędzior and Welfe [16] for Poland. Arlt et al. [1] aimed at testing the theory of PIH. The result of his work was the statement that an excessive sensitivity of real consumption to the change in real disposable income was demonstrated which is in line with AIH, the permanent income hypothesis being rejected. The same conclusions are made by Leszkiewicz-Kędzior and Welfe [16] who base their study on annual polish data between 1970 and 2008.

In contrast to aforementioned studies Carroll [4] using US data suggests that Friedman's PIH including uncertainty matches much better than other models. A strong support of the consumption function under the PIH is found also by Manitsaris [17] who analysed selected fifteen European Union members combining adaptive ex-

pectations with partial adjustment model. A validation of the PIH is also performed by Pánková [21] using an instrumental variables approach under the condition of the Czech economy.

Aiming at the scope of the contribution together with the variables implied from the theoretical discussion, geographical and time aspects as the CEEC and crisis period will be taken into consideration when specifying an appropriate model for industrial countries' private consumption expenditure.

Even if one can claim that private consumption can be determined by country specific factors because of similar abovementioned factors and their determinants like uncertainty, credit constraint, consumption-leisure interactions, habits and durability, alternative to rational expectations, liquidity constraints, the illiquidity of some assets, financial deregulation or demographic factors (Muellbauer [20]), especially in the examined CEE countries, we expect that some common patterns related to consumption are verifiable.

3 Data and Methodology

The research is based on unbalanced panel analysis using selected developed countries data covering the period from 1970 to 2011. The dataset consists of the group of 40 industrial countries over the period 1970–2011, annual data in logarithms. In order to estimate real private consumption, following explanatory variables are used:

- real national disposable income per capita (real GDP per capita as a proxy of real disposable income in case of robustness check),
- real interest rates,
- anticipated inflation,
- M2 aggregate as a proxy of financial wealth,
- dummy for the CEE countries,
- dummy for the global crisis period starting in the 2008.

Real national disposable income per capita (either in current period, lagged or permanent) is the main determinant of consumption. The higher the income, (non-linearly) the higher the consumption, thus, we expect a positive signed coefficient of real national disposable income per capita with an elasticity ranged from 0-1. If current income is significant while lagged insignificant, Keynes' absolute income hypothesis valid and Hall's model is not the appropriate one. This results from the idea that if consumption is dependent on both current and anticipated income, then one might expect that decisions related to consumption are affected by income earned over more time periods and therefore permanent income hypothesis would be confirmed.

Wealth is considered to be one of the most important determinants of consumption from the perspective of the LCH. If the LCH is approved then monetary aggregate (M2) as a proxy of wealth should manifest a positive sign, wealth elasticity being in the interval (0, 1). If significant, the LCH is validated, while an insignificant coefficient foreshadows either AIH or PIH.

Real interest rates are connected with the possibilities deriving from financial market as lending and borrowing. Higher interest rates motivate to reduce loans and make savings more attractive. As this can lead to a reduced consumption a negative sign can be expected. However, higher interest rates increase financial wealth and so even a positively signed coefficient can result.

Anticipated inflation is proxied by 1 period lagged GDP deflator that is well described by first-order autoregressive process. The consumption that is based on anticipated inflation would be in favour of rational expectation hypothesis.

Dummy CEEC is necessary to include because of possible different patterns in consumption behaviour in the CEE countries when compared to the rest of the EU. One would expect a negative sign, since the consumption per capita variable tends to be lower in the CEE countries.

Dummy Crisis is presented in the model to control for the financial and economic crisis period during which the consumption reduced. As a consequence a negatively-signed coefficient can be expected.

Note that the long-run relations are estimated using mentioned variables in levels and the short-run fluctuations from equilibrium are estimated using the same sample of variables (excluding dummy for the CEE countries) are studied in differences. Since there can be a tendency in changing consumer behaviour during 2008 crisis period it is convenient to use the dummy variable in order to prevent from biased results and to obtain interesting information of the topical nowaday situation.

Preliminary to the estimation process, all variables have been tested for unit roots using a test developed by Im, Pesaran and Shin [14] and Fisher type unit root test modified by Choi [7]. All variables have been found to be stationary in differences, i.e. integrated of I(1). In order to estimate the real private consumption and to obtain

robust results, three different techniques could be applied: pooled estimation and fixed effect estimation. To choose appropriate model for real private consumption estimation, we apply the following panel diagnostic tests: F-test for individual effect, Breusch-Pagan LM test and Hausman specification test. Based on results of these tests we decide to use fixed effect estimation, which is also correct from the theoretical point of view due to collection of relatively homogeneous datasets.

In order to model consumption function in chosen developed countries we base the model on abovementioned theoretical and empirical studies. The following relationship is assumed:

$$C_{it} = f\left(Y_{it}, IR_{it}, Inf_{it-1}, CEEC_{it}, CRISIS_{it}\right), \tag{1}$$

where:

- C_{it} is a real private consumption of *i*-th country in time *t*,
- Y_{it} is a real national disposable income per capita (real GDP per capita in the robustness check) of *i*-th country in time *t*,
- IR_{it} is a long term interest rates of *i*-th country in time *t*,
- Inf_{it-1} is an anticipated inflation of *i*-th country in time *t+1*,
- $CEEC_{it}$ is a dummy for the CEE countries,
- $CRISIS_{it}$ is a dummy for the economic crisis period starting in the 2008.

Data source for all mentioned variables is OECD database. To obtain short-run and long-run information on real private consumption we use VECM, i.e. we estimate the following equation:

$$\Delta \ln C_{it} = \alpha_t + \beta_{1t}\Delta \ln C_{it} + \beta_{2t}\Delta \ln Y_{it} + \beta_{3t}\Delta IR_{it} + \beta_{4t}\Delta \ln Inf_{it-1} + \gamma ECT_{it-1} + \eta CEEC_{it} + \lambda CRISIS_{it} + \varepsilon_{it} \tag{2}$$

Since Hall [13] argues that consumption is not only dependent on disposable income but also on wealth that compensates low savings we reflect this including M2 aggregate as a proxy of wealth:

$$C_{it} = f\left(Y_{it}, M2_{it}, IR_{it}, Inf_{it-1}, CEEC_{it}, CRISIS_{it}\right), \tag{3}$$

where $M2_{it}$ is an M2 aggregate being a proxy of wealth of *i*-th country in time *t*. Therefore, we estimate the following equation:

$$\Delta \ln C_{it} = \alpha_t + \beta_{1t}\Delta \ln C_{it} + \beta_{2t}\Delta \ln Y_{it} + \beta_{3t}\Delta IR_{it} + \beta_{4t}\Delta \ln Inf_{it-1} + \beta_{5t}\Delta \ln M2_{it} + \gamma ECT_{it-1} + \eta CEEC_{it} + \\ + \lambda CRISIS_{it} + \varepsilon_{it} \tag{4}$$

4 Results

Firstly, we focus on unit root test. As can be seen from Tab. 1, all variables are non-stationary in levels, but stationary in differences, i.e. all variables are integrated of I(1).

Variable	Levels		Differences	
	statistics	p-value	statistics	p-value
lnC	3.408	0.999	-7.928	0.000
lnIR	-1.044	0.148	-7.984	0.000
LlnInf	-1.317	0.094	-15.506	0.000
lnM2	-0.467	0.320	-2.276	0.011
lnY (GDP variant)	4.493	1.000	-4.868	0.000
lnY (DI variant)	3.461	0.999	-5.576	0.000

Table 1 Im, Pesaran, Shin unit root test[3]

Consequently, pooled ordinary least squares (OLS) and fixed effects are applied to estimate the real private consumption for the group of 40 industrial countries over the period 1970–2011 (see models 1–8 in Tab. II). As can be seen from the mentioned table, some explanatory variables are not statistically significant among which the proxy for wealth and both dummies. The only relevant variables are then income, long-term interest rate and in model 2 monetary aggregate M2. The distortion from equilibrium is caused by lagged income, consumption

[3] Fisher type tests are not reported but are available upon request.

and long-term interest rate. The consumption seems to be unaffected by the crisis period. Even the geographic dummy has not proven the impact on consumption.

	(1) DI pooled	(2) DI,M2 pooled	(3) GDP pooled	(4) GDP,M2 poo~d	(5) DI FE	(6) DI,M2 FE	(7) GDP FE	(8) GDP,M2 FE
D.lnIR	0.0145	0.0226	0.00212	-0.00413	0.0163	0.0133	0.00299	-0.0132
	(0.00669)*	(0.0129)	(0.00641)	(0.0117)	(0.00657)*	(0.0124)	(0.00648)	(0.0127)
D.LlnInf	0.00240	0.0352	-0.00280	0.0869	-0.0730	-0.216	-0.0463	-0.236
	(0.0575)	(0.135)	(0.0547)	(0.122)	(0.0569)	(0.127)	(0.0557)	(0.137)
D.lnY	0.597	0.507	0.659	0.700	0.614	0.560	0.678	0.730
	(0.0317)***	(0.0612)***	(0.0318)***	(0.0648)***	(0.0315)***	(0.0584)***	(0.0329)***	(0.0719)***
L.lnC	-0.0328	-0.0524	-0.0241	-0.0448	-0.201	-0.578	-0.111	-0.395
	(0.00924)***	(0.0242)*	(0.00808)**	(0.0203)*	(0.0237)***	(0.0732)***	(0.0194)***	(0.0739)***
L.lnIR	-0.00694	-0.0239	-0.00690	-0.0225	0.00783	0.00690	0.000871	-0.00584
	(0.00354)	(0.0107)*	(0.00320)*	(0.00950)*	(0.00433)	(0.0142)	(0.00404)	(0.0137)
L.LlnInf	0.0589	0.126	0.00139	0.0512	-0.0249	-0.237	-0.0412	-0.523
	(0.0457)	(0.167)	(0.0436)	(0.147)	(0.0481)	(0.193)	(0.0477)	(0.200)*
L.lnY	0.0165	0.0248	0.00594	0.00880	0.186	0.552	0.0942	0.405
	(0.00929)	(0.0195)	(0.00757)	(0.0165)	(0.0239)***	(0.0650)***	(0.0193)***	(0.0733)***
CEEC	0.00196	-0.00894	-0.000616	-0.0150				
	(0.00327)	(0.0101)	(0.00317)	(0.00867)				
CRISIS	-0.00105	0.00613	0.00331	0.0132	0.000744	0.0106	0.00322	0.0104
	(0.00300)	(0.00541)	(0.00295)	(0.00488)**	(0.00294)	(0.00727)	(0.00299)	(0.00744)
D.lnM2		0.0728		0.0169		0.00623		-0.0297
		(0.0277)**		(0.0263)		(0.0277)		(0.0291)
L.lnM2		-0.00579		-0.00504		-0.0114		-0.0301
		(0.00678)		(0.00542)		(0.0212)		(0.0221)
_cons	-0.0992	-0.196	0.186	0.219	0.155	1.208	0.307	2.545
	(0.223)	(0.777)	(0.211)	(0.679)	(0.235)	(1.008)	(0.232)	(1.017)*
N	519	158	548	161	519	158	548	161
R-sq	0.645	0.600	0.666	0.668	0.651	0.695	0.653	0.692
adj. R-sq	0.639	0.569	0.660	0.644	0.630	0.623	0.632	0.618
BIC	-2505.3	-671.0	-2678.6	-715.6	-2574.0	-758.6	-2716.7	-772.3
F	102.9	19.87	118.9	27.27	113.8	28.93	121.4	28.95

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

. DI..national disposable income variants, GDP..gross domestic product variants

Table 2 Models estimation

Moreover, as showed in Tab. 3, fixed or random effects are preferred to pooled OLS (see F-test and Breusch-Pagan LM test). Therefore, based on Hausman test we favour the fixed-effect model specification to the others

	(1) DI pooled statistics	p-value	(2) DI, M3 pooled statistics	p-value	(3) GDP pooled statistics	p-value	(4) GDP, M3 pooled statistics	p-value
Durbin-Watson test	2.179	0.761	2.202	0.763	2.171	0.761	2.180	0.762
Breusch-Godfrey/Wooldridge test	13.692	0.090	3.164	0.367	15.701	0.047	6.894	0.075
Bera, Sosa-Escudero and Yoon locally robust test	5.169	0.023	2.359	0.125	4.789	0.029	1.798	0.180
F-test for individual effects	3.367	<0.001	4.453	<0.001	1.595	0.046	2.050	0.009
F-test for time effects	2.160	<0.001	3.969	0.001	1.739	0.004	2.337	0.035
LM Test – (Breusch-Pagan)	8.911	0.003	0.333	0.564	6.653	0.010	0.003	0.953
LM Test – time effects (Breusch-Pagan)	429.568	<0.001	10.462	0.001	312.346	<0.001	2.275	0.131

Table 3 Verification of pooled models

	(5) DI FE statistics	p-value	(6) DI, M2 FE statistics	p-value	(7) GDP FE statistics	p-value	(8) GDP, M2 FE statistics	p-value
Durbin-Watson test	2.078	0.785	2.163	0.838	2.129	0.923	2.192	0.880
Breusch-Godfrey/Wooldridge test	8.428	0.393	7.519	0.057	19.939	0.011	11.574	0.009
Bera, Sosa-Escudero and Yoon locally robust test	0.002	0.961	0.204	0.652	0.999	0.318	0.003	0.955
Wooldridge Test for AR(1) Errors in FE Panel Models	0.002	0.969	0.109	0.741	35.102	<0.001	0.021	0.886
Hausman test	93.072	<0.001	96.320	<0.001	49.562	<0.001	43.895	<0.001

Table 4 Verification of fixed effect models

5 Conclusions

The contribution investigates the response of real private consumption per capita to real national disposable income per capita or real GDP per capita, respectively, long-term interest rates, monetary aggregate M2 and anticipated inflation. For the purpose of the private consumption function estimates we used pooled OLS and fixed effects on unbalanced panel data sample. Our results show a strong dependence of real private consump-

tion per capita on income (disposable income per capita or real GDP per capita as proxy of disposable income) without statistical effects of monetary aggregate and anticipated inflation. The results so show strong support for the Keynes' absolute income hypothesis in contrast to permanent income hypothesis, rational expectation hypothesis, and life cycle hypothesis. Similarly, the private consumption seems to be unaffected by the crisis period and geographical country position.

References

[1] Arlt, J., Čutková, J., and Radkovský, Š.: Analýza spotřební funkce v podmínkách ČR. In: *Výzkumná publikace ČNB* **34** (2001).

[2] Byrne, J. P., and Davis, P. E.: Disaggregate wealth and aggregate consumption: an investigation of empirical relationships for the G7. *Oxford Bulletin of Economics and Statistics* **65** (2003), 197–220.

[3] Campbell, J. Y., and Mankiw, N. G.: Permanent Income, Current Income and Consumption. *Journal of Business & Economic Statistics* **8** (1990), 265–279.

[4] Carroll, C. D.: A theory of the consumption function, with and without liquidity constraints. *Journal of Economic Perspectives* **15** (2001), 23–45.

[5] Castiglione, C., Grochová, L., Infante, D., et al.: The demand for beer in presence of past consumption and advertising in the Czech Republic. *Agricultural Economics-Zemědělská ekonomika* **57** (2011), 589–599.

[6] Chambers, M. J.: An alternative time series model of consumption: Some empirical evidence. *Applied Economics* **23** (1991), 1361–1366.

[7] Choi, I.: Unit root tests for panel data. *Journal of International Money and Finance* **20** (2001), 249–272.

[8] Davidson, J. E. H., Hendry, D. F., Srba, F., and Yeo, S.: Econometric Modeling of the Aggregate Time-Series Relationship Between Consumers' Expenditure and Income in the United Kingdom. *The Economic Journal* **88** (1978), 661–692.

[9] Davis, M. A.: The consumption function in macroeconomic models: A comparative study. *Applied Economics* **16** (1984), 799–838.

[10] Fisher, I.: *The Theory of Interest*. The Macmillan Co., New York, 1930.

[11] Friedman, M.: *A theory of the consumption function*. Princeton University Press, Princeton, 1957.

[12] Grochová, L.: Advertising as a possible determinant of private consumption. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* **59** (2011), 97–104.

[13] Hall, R. E.: Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence. *Journal of Political Economy* **86** (1978), 971–987

[14] Im, K. S., Pesaran, M. H., and Shin, Y. Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics* **115** (2003), 53–74.

[15] Keynes, J. M.: *The General Theory of Employment, Interest, and Money*. Harcourt Brace, New York, 1936.

[16] Leszkiewicz-Kędzior, K., and Welfe W. Consumption function for Poland. Is life cycle hypothesis legitimate? *Bank i Kredyt* **43** (2012), 5–20.

[17] Manitsaris, A.: Estimating the European Union Consumption Function under the Permanent Income Hypothesis. *International Research Journal of Finance and Economics* **2** (2006), 131–135.

[18] Modigliani, F.: Fluctuations in the Saving-Income Ratio: A Problem in Economic Forecasting," NBER Chapters. In: *Studies in Income and Wealth, National Bureau of Economic Research, Inc.* 1949, 369–444.

[19] Molana, H.: The Time Series Consumption Function: Error Correction, Random Walk and the Steady-State, *Economic Journal, Royal Economic Society* **101** (1991), 382–403.

[20] Muellbauer, J. N.: The assessment: Consumer expenditure. *Oxford Review of Economic Policy* **10** (1994), 1–41.

[21] Pánková, V.: Permanent Income and Consumption. In: *Proceedings of 30th International Conference Mathematical Methods in Economics*. 2012, 670–674.

Option Pricing and Partial Hedging in the Black-Scholes Model

Haochen Guo[1]

Abstract. Option pricing techniques are often considered on the most mathematically complex of applied areas of finance. The Black-Scholes model is the most widely used models for the option pricing. This paper presents and analyses the option pricing and hedging in the Black-Scholes model. The goal of hedging is against the risk for protect the losses from the investment, that means strategically using instruments in the financial market to offset the risk of any adverse price movements. Inside of paper, it provides example of the partial hedge, which reduces the risk, that using option pricing with the Black-Scholes model.

Keywords: Option pricing, the Black-Scholes model, partial hedging

1. Introduction

This paper is analyses the option pricing and present partial hedging in the Black-Scholes model. The application of hedging strategy in the paper it will use the delta hedging in the Black-Scholes model. Delta hedging involves creating a position with zero (See [4]). There are 2 parts inside of the paper. First part is methodology of option pricing. The binominal model and the Black-Scholes model are two methods which usually to use for the option pricing. In this part, there are 2 examples for calculation both the Black-Scholes model and the binominal model. Second part is static delta partial hedging in the Black-Scholes model. Inside of this part, it will describe the delta hedging and the partial delta hedging. Then will be the conclusion.

2. Methodology of Option Pricing

Option pricing is on no arbitrage argument. If it can add other traded assets or instruments to an option so that the resulting portfolio is risk free, in other words if the option can be perfectly hedged, then the portfolio should return the risk free rate. Since it has no risk, this risk free portfolio will have the same value today for all investors, regardless of their attitude to risk. Because the hedging instruments have market prices, which are the same for all investors, the option must have the same value for investors.

2.1. The Black-Scholes Model

The Black-Scholes model assumes that stock price movements can be described by a statistical process known as geometric Brownian motion. This process is summarized by a volatility factor, σ, which is analogous to the investor's stock price forecasts in the previous models. Formally, the stock price process assumed by Black and Scholes is

[1] Faculty of Economics, VŠB-TUO, Sokolská 33, 701 21 Ostrava 1, Czech Republic haochen.guo.st@vsb.cz

$$\Delta S / S = \mu[\Delta T] + \sigma\eta[\Delta T]^{1/2} \tag{1}$$

So that a stock's return $(\Delta S / S)$ from the present through any future period T has both an expected component $(\mu[\Delta T])$ and a "noise" component $(\sigma\eta[\Delta T]^{1/2})$, where μ is the mean return and η is the standard normally distributed random error term.

Assuming the continuously compounded risk-free rate and the stock's variance remain constant until the expiration date T, Black and Scholes used the riskless hedge intuition to derive the following formula for valuing a call option on a no dividend-paying stock

$$C_0 = SN(d_1) - X(e^{-(r)T})N(d_2) \tag{2}$$

where $e^{-(r)T}$ is the discount function for continuously compounded variables.

$$d_1 = [\ln(S / X) + (r + 0.5\sigma^2)[T]] \div \sigma[T]^{1/2} \tag{3}$$

and

$$d_2 = d_1 - \sigma[T]^{1/2} \tag{4}$$

Properties of the model reveals that the option's value is a function of five variables, there are current security price, exercise price, time to expiration, risk-free rate, security price volatility.

Functionally, the Black-Scholes model holds that $C = f(S, X, T, r, \sigma)$. The first and fourth factors are observable market prices, and the second and third variables are defined by the contract itself. Thus, the only variable an investor must provide is the volatility factor. (See [5])

Example 1. Calculation the Black-Scholes model which use the underlying asset of PetroChina Company Limited, which is the stock from China Stock Exchange Market. PetroChina Company Limited ("PetroChina") is the largest oil and gas producer and distributor, playing a dominant role in the oil and gas industry in China. It is not only one of the companies with the biggest sales revenue in China, but also one of the largest oil companies in the world. It was listed on Shanghai Stock Exchange on November 5, 2007, the stock code is 601857. The parameters for calculation are current stock price (S) is 8.65, strike price (X) is 50, stock volatility (σ) is 0.66, risk free interest rate (r) is 0.1, option time to maturity (T) is 0.5, to compute the Black-Scholes model is based on the normal distribution. After defined the normal distribution, program the Black-Scholes formula is in following table (See [2]).

Normal Distribution	$snormaB[x\_] := Erf[x / Sqrt[2]] / 2 + 0.5$ $sndist = NormalDistribution[0,1]$
Black-Scholes model	$Clear[snormal, d1, d2, BSCALL, BSPUT]$ $d1[S\_, X\_, sigma\_, T\_, r\_] := (Log[S / X] + r + sigma\text{^}2 / 2) * T) / (sigma * Sqrt[t])$ $d2[S\_, X\_, sigma\_, T\_, r\_] := d1[S, X, sigma, T, r] - sigma * Sqrt[T]$ $BSCALL[S\_, X\_, sigma\_, T\_, r\_] := S * snormal[d1[S, X, sigma, T, r]] - X * Exp[-r * T] * snormal[d2[s, x, sigma, T, r]]$ $BSPUT[S\_, X\_, sigma\_, T\_, r\_] := BSCALL[S, X, sigma, T, r] + x * Exp[-r * T] - S$

The result of Black-Scholes model	$BSCALL[8.65,50,0.66,0.5,0.1] = 0.00029176$ $BSPUT[8.65,50,0.66,0.5,0.1] = 38.9118$ $d1[8.65,50,0.66,0.5,0.1] = -3.41889$ $d2[8.65,50,0.66,0.5,0.1] = -3.88558$

Table 1 Calculation of Black-Scholes Model

The result of Black-Scholes and the graph the functions are showing in the following.

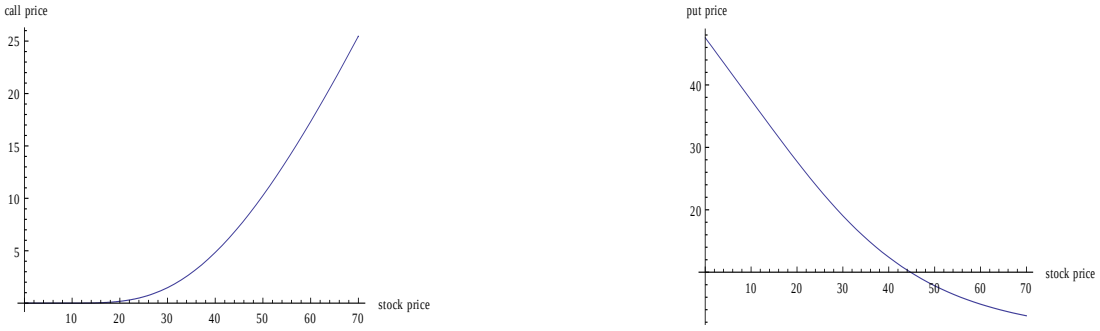


Figure 1 Call and Put Option Price in Black-Scholes Model

2.2. Binomial Model

Cox and Ross (1976) introduced a simple model for pricing an option in which the underlying asset price dynamics are governed by a binomial tree in which the price of an asset can move either upward or downward by a constant multiple, at each time step. It showed how to construct a binomial tree in which there can be no arbitrage, where the probability of an upward move is constant throughout the tree. In a simple binomial tree is assume that the transition probability, for risk neutral valuation it choose *p* so that *S* grows at the risk free rate, that is (See [1])

$$pS_0u + (1-p)S_0d = S_0e^{(r-q)T} \tag{5}$$

$$p = \frac{e^{(r-q)T} - d}{u - d} \tag{6}$$

The total return provided by the stock in a risk-neutral world must be the risk-free interest rate *r*, the dividends provide a return equal to *q*, and the return in the form of capital gains must be *(r-q)*. Because the value of the derivative is the expected payoff in a risk-neutral world discounted at the risk-free rate, that is

$$f = e^{-rT}[pf_u + (1-p)f_d] \tag{7}$$

Example 2. Calculation the binominal model which uses the underlying asset of PetroChina, the parameters data is same with example1. To define binomial option pricing for European options is in the following table (See [2]).

The Binominal Option Pricing Model	$Clear[EuropeanOption, EuropeanCall, EuropeanPut]$ $EuropeanOption[S\_, sigma\_, T\_, r\_, exercise\_Function, n\_] := Module[\{$ $u = N[Exp[Sqrt[T/n] * sigma]],$ $d = N[Exp[-Sqrt[T/n] * sigma]],$ $R = N[Exp[r * T/n]]\},$ $p = (R-d)/(R*(u-d));$ $q = (u-R)/(R*(u-d))]$

```
Sum[exercise[s*u^j*d^(n-j)]*Binomial[m,j]*p^j*q^(n-j),{j,0,n}]
```

$$\sum_{j=0}^{n} p^{j} q^{-j+n} Binomial[m, j] exercise[d^{-j+n} s u^{j}]$$

```
EuropeanCall[S_,X_,sigma_,T_,r_,n_]:=EuropeanCall[S,X,sigma,T,r,n]=EuropeanOption[S,sigma,T,r,Max[#1-x,0]&,n]
EuropeanPut[S_,X_,sigma_,T_,r_,n_]:=EuropeanOption[S,sigma,T,r,Max[S-#1,0]&,n]
```

The result of binominal model for call option	0.00024110850

Table 2 Calculation of Binominal Model

There is the graph shows option pricing in binominal model and the Black-Scholes model.

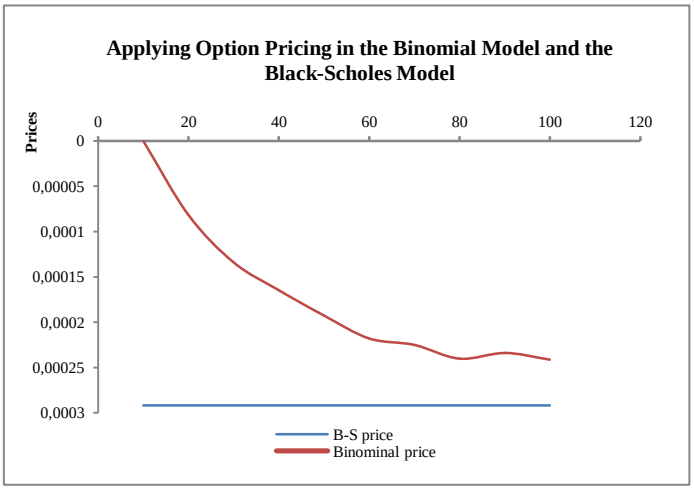


Figure 2 Applying Option Pricing in the Binomial Model and the Black-Scholes Model

3. Static Delta Partial Hedging in the Black-Scholes Model

Delta measures how volatile an option on futures premium is relative to the underlying, expressed by the change in option premium to change in futures price. The greater the extent to which the option is in the money, the greater its delta and vice versa. Delta is measured on a continuum form 0 to 1. High delta options are close to one. Delta is a metric for hedgers to determine how volatile the underlying is that they are attempting to hedge and the degree to which a hedge might be effective. In the delta hedging, it hedges the call position by purchasing $\frac{\partial C}{\partial S}$, where C is the call price. Following table shows the formulation of delta hedging for options. (See [9])

Parameters	Call Option	Put Option
Delta $\frac{\partial C}{\partial S}$	$N(d_1)$	$-N(-d_1) \equiv N(d_1) - 1$

Table 3 Formulation of Delta Hedging for Options

To hedging the position, the hedge ratio can be calculated as following table (See[3]).

$Clear[delta]$

$delta[S\_, X\_, sigma\_, T\_, r\_] := snormal[d1[S, X, sigma, T, r]]$

$delta[8.65,50,0.66,0.5,0.1] = 0.000314385$

Table 4 Calculation of Static Delta Hedging

The result of the delta hedge ratio is 0.000314385.Partial hedge is a position that has been hedged in part and not in its entirety. This will reduce, but not eliminate, adverse movements to the position being hedged. A partial hedge is undertaken if the hedger wishes to be still able to benefit from some upside but is concerned about the amount of the exposure being taken. Following table is distinguished between full hedging and partial hedging in the delta hedging strategy.

Full Hedging	$\Delta\Pi = h \cdot \Delta S - \frac{\partial C}{\partial S} \cdot \Delta S = 0$ $h = \frac{\partial C}{\partial S} = delta$
Partial Hedging	$\Delta\Pi = \Delta S \cdot (h \cdot x - \frac{\partial C}{\partial S}) = 0$ $h = \frac{\frac{\partial C}{\partial S}}{x} = \frac{delta}{x}$

Table 5 Distinguish Between Full Hedging and Parting Hedging

In the partial delta hedging which the $\prod$ is position value, S is underlying asset value, h is hedge ratio, x is determine the partial underlying asset, the $x = [0;100\%]$. When $x = 1\%$ means the full hedging,

Example 3. Calculation the partial hedging, underlying asset is PetroChina which the same computation data with previews examples. Following table is to distinguish between the full hedging and partial hedging.

x	0.01	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
h ratio	0.03143850	0.00314385	0.00157193	0.00104795	0.00078596	0.00062877	0.00052398	0.00044912	0.00039298	0.00034932	0.00031439

Table 6 Full Hedging and Partial Hedging

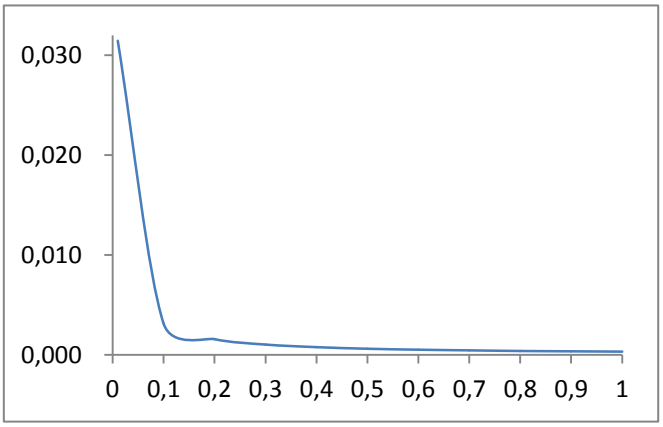


Figure 3 Partial Hedging Trends

4. Conclusion

The paper describes partial hedging strategy using delta hedging applied. It is describes about distinguish of basic static delta hedging strategy. In the paper, there are 3 examples, which are to calculation of Black-Scholes model and binominal model, static delta hedging and partial hedging in the Black-Scholes model. The Black-Scholes option-pricing model is the first successful option-pricing model, published in 1973 and based on stochastic calculus. It focuses on the pricing of European options, in which the underlying does not pay a dividend in the option period. The option is priced according to the value of the underlying, the volatility of the value of the underlying, the exercise price, the time to maturity, and the risk-free rate of interest. The model provides a general approach to option pricing and has given rise to a number of other option-pricing models. As a result, for European options, the binomial model converges on the Black-Scholes formula as the number of binomial calculation steps increases. In fact, the Black-Scholes model for European options is really a special case of the binomial model where the number of binomial steps is infinite. In other words, the binomial model provides discrete approximations to the continuous process underlying the Black-Scholes model.

Acknowledgements

The research of the second author was supported by SGS project of VSB-TU Ostrava under No.SP2013/3.

References

[1] Alexander,C.:*Market risk analysis volume III: pricing, hedging and trading financial instruments.* John Wiley & Sons,Ltd. 2008.

[2] Benninga,S,.Wiener,Z.:*Binomial Option Pricing, the Black-Scholes Option Pricing Formula, and Exotic Options*. Mathematica in Education and Research, 1997.

[3] Benninga,S,.Wiener,Z.;*Dynamic Hedging Strategies*. Mathematica in Education and Research, 1998.

[4] Hull,C,J.:*Option, futures, and other derivatives.* Prentice-Hall International, Inc. 2006.

[5] Jorion,P.:*Financial risk manager handbook 2001-2002.* John Wiley & Sons,Inc. 2001.

[6] Tichý,T.:*Replication methods in the pricing and hedging of barrier options.* Finance a Uver-Czech Journal of Economics and Finance.54(2004).305-324.

[7] Zmeškal,Z.:*Application of the American real flexible switch options methodology a gereralized approach.* Financial a Uver-Czech Journal of Economics and Finance.58(2008).261-275.

[8] Zmeškal,Z.:*Application of the fuzzy-stochastic methodology to appraising the firm value as a European call option.* European Journal of Operational Research.135(2001).303-310.

[9] Zmeškal,Z., Dluhošová,D.,Tichý,T.:*Financial models.* VŠB-Technická Univeryita Ostrava, 2004.

[10] Zmeškal,Z.:*Fuzzy-stochastic estimation of a firm value as a call option.* Finance a Uver.49(1999).168-175.

[11] Zmeškal,Z.:*Generalised soft binomial American real option pricing model (fuzzy-stochastic approach).*European Journal of Operational Research.207(2010).1096-1103.

[12] Zmeškal,Z.:*Hedging strategies and financial risks.* Finance a Uver-Czech Journal of Economics and Finance.54(2004).50-63.

[13] Zmeškal,Z.:*Real compound sequential and learning options modeling.*28th International Conference on Mathematical Methods in Economics 2010.688-693.

LMD1 decomposition of the changes in the industrial energy consumption in the Czech Republic

Vladimír Hajko[1]

Abstract. This article applies logarithmic mean Divisia index decomposition method to analyze the changes in the energy consumption of the industrial sector in the Czech Republic, EU-27, EU-15 and EU-12 during the period of 1997 to 2009. The results indicate several differences between the EU-12 and EU-15 countries. The major differences are: the magnitude of the activity effect; how influential was the period of the recession for the energy consumption in the respective groups; the amount of reduction in the energy consumption and the development of the energy intensity effect in the period of crisis.

Keywords: LMD1, Energy, Energy consumption, Industry

JEL classification: Q43
AMS classification: 11Y05

1 INTRODUCTION

In the last two decades (following the transition from the centrally planned economy) the industrial sector in the Czech Republic (generally in the transforming countries) has gone through significant changes. The original rather ineffective structure and technology has been gradually improving and becoming far more energy and market efficient. Given the increasing competition and general market tendency to improve efficiency, it is of little surprise that the industrial sector has been the sector that gone under most intense transformation, not only in the Czech Republic but in most European countries. For instance in the EU-12 countries[1], we can observe nearly 20 percentage points reduction in the overall industrial energy consumption share in 2011 compared to 1990. The EU-15 countries[2] recorded a similar shift from industry to transport, though with only about 5 p. p. decrease in the share of industry and somewhat less than 5 p.p. increase in the share of transport).

Despite the significant shift of the energy consumption from the industry to other sectors (as hinted above, mainly to transport), the industrial sector still represents about a quarter (in EU average) or a third (in CZ) of the total energy consumption and remains in the live interest of many government policies (including the energy conservation and CO_2 reduction scenarios such as the Energy Policy for Europe (also known as 20-20-20 plan)).

This article aims to unfold these changes, with the main focus on the energy consumption development in the industrial sector. For comparative reasons, the following results also cover, apart from the Czech Republic, the aggregates of EU-27, EU-15 (original member countries) and EU-12 (new member countries). The main tool of the analysis is the application of the so-called logarithmic mean Divisia 1 index method (LMD1), proposed by [4]. The history of application of the index decomposition analysis (IDA) to decompose an aggregate ranges back to 1970s (see [6] for comprehensive review). Nowadays, the IDA methods are widely accepted analytical tool for policy making [2], and it is indeed evidenced not only by the number of studies employing the IDA methods but also by its acceptance by the renowned international bodies such as the International Energy Agency [10].

However, rather little attention has been given to these methods not only in the Czech Republic, but also in most countries of the former Eastern Bloc. One significant burden is the data comparability across different countries. Not only for a cross-country comparison, but also for the meaningful correspondence of

[1]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 507/41a, Brno, Czech Republic, vladimir.hajko@mail.muni.cz

[1]Bulgaria, Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Romania, Slovakia, Slovenia

[2]Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, United Kingdom.

the individual country analyses, it is necessary to have a unified methodological structure and collection of the data (mainly regarding the delimitation of the individual sub-sectors how the value added is measured). Eurostat databases are useful in their methodological homogeneity, but in certain cases do not provide the latest (and, given the late economic troubles the Europe and the world have been going through, probably interesting) data.

2 DATA

This section is devoted to providing a detailed information of the data used in the analysis as well as to show some of the interesting facts regarding the data. All the figures and the tables presented in this article are my own. Two main indicators necessary for this IDA application are the final energy consumption and the gross value added (GVA). The data on the energy consumption are available under Eurostat's table [nrg100a] , and are measured in thousand tonnes of oil equivalent[3]. The data on the gross value can be found under the table [nama_nace60_c] and are measured in current millions of euro (from January 1, 1999) or current millions of ECU (up to Decemeber 31, 1998). As hinted above, the reason for the selection of the data source is specifically the availability of the indicators for multiple countries in exactly the same structure and measurement methods. This should allow any further researcher to directly compare the results across different countries and allows for extensibility of the results.

The first important thing that needs to be noted is that a meaningful analysis requires that the data for individual sub-sectors are related to the same set of activities. In other words, it is of no use if we collect data that do not correspond to the identical underlying sub-sector. As the Table 1 shows, the source data structure allows for the good match in only 11 of the 13 energy subsectors, meaning that this analysis is necessarily limited (in European average) to analyze only about the 90% of the consumption that took place in all of the industrial sub-sectors.

Due to certain peculiarities in the data, one might be lead to a certain misunderstanding of the results if it is not kept in in mind that "Industry" in this analysis is (albeit unfortunately) limited to the sum of the 11 energy subsectors corresponding to the 14 NACE codes as presented in the Table 1. An example of such peculiarity can be the atypical behavior of the energy consumption in the elusive "Not elsewhere specified" sub-sector in several EU-12 countries. For instance, in the Czech Republic this sub-sector accounted for more than half of the total Czech energy consumption in the early 1990s. This however can be attributed to the somewhat imperfect statistical collection and classification of the energy consumption, especially during the first half of the 1990s. An overall picture of the all sub-sectors indicates the overall pattern is not very different from the examined part of the industry. Yet I feel it is important to distinguish the overall industry and the examined sub-sectors delimitation. Furthermore, such unhelpful statistical classification in the early 1990s also speaks for the selection of the time frame "after the dust has settled" - in case of this analysis the time period was chosen from the 1997 onwards.

Figure 1 shows the development of both the industrial (final) energy consumption and the Gross Value Added in the examined industrial sector. We can see that the average behavior in the European countries shows a downward trend in both energy consumption share and the GVA share. However, the heavier presence (compared to EU-15) of industrial sub-sector can be clearly seen in EU-12 aggregate and Czech Republic. Unlike the aggregate of EU-15, the GVA share of the industry in the Czech Republic is decreasing. Despite that the Czech Republic still has above average (compared both to EU-27 and EU-15) level of industry's contribution on the GVA.

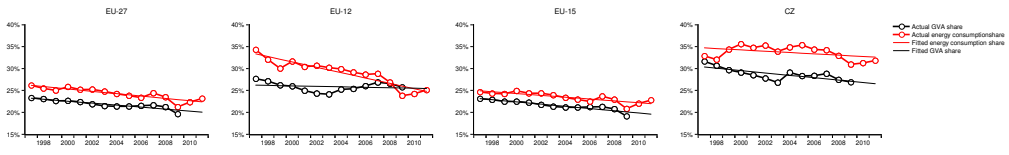


Figure 1 Gross value added and energy consumption shares, examined sub-sectors, EU-27 countries, 1990-2010

[3]The TOE unit is defined by the International Energy Agency as the amount of energy released by burning one ton of crude oil and represents 41.868 GJ.

Energy code	Energy description	NACE code	NACE description
B_101805	Iron and steel	DJ	Manufacture of basic metals and fabricated metal products
B_101810	Non-ferrous metals	#N/A	
B_101815	Chemical, including petrochemical	DF	Manufacture of coke, refined petroleum products and nuclear fuel
		DG	Manufacture of chemicals, chemical products and man-made fibres
		DH	Manufacture of rubber and plastic products
B_101820	Non-metallic mineral products	DI	Manufacture of other non-metallic mineral products
B_101825	Mining (excluding energy producing industries) and quarrying	C	Mining and quarrying
B_101830	Food processing, beverages and tobacco	DA	Manufacture of food products, beverages and tobacco
B_101835	Textile and leather	DB	Manufacture of textiles and textile products
		DC	Manufacture of leather and leather products
B_101840	Pulp, paper and printing	DE	Manufacture of pulp, paper and paper products; publishing and printing
B_101846	Transport equipment	DM	Manufacture of transport equipment
B_101847	Machinery;	DK	Manufacture of machinery and equipment n.e.c.
B_101851	Wood and wood products (other than pulp and paper);	DD	Manufacture of wood and wood products
B_101852	Construction;	F	Construction
B_101853	Not elsewhere specified.	#N/A	
#N/A		DL	Manufacture of electrical and optical equipment
#N/A		DN	Manufacturing n.e.c.

Table 1 Matching energy consumption and GVA sub-sectors

3 METHODOLOGY

The index decomposition methods are used to identify the individual contributions of several factors to the combined total change. It is especially useful method in analyzing a change[4] in an overall aggregate V (where $V = \sum_i V_i$), that is composed of multiple factors ($V_i = x_{1,i}$,$x_{2,i}, \ldots, x_{n,i}$), with non-zero changes in factors (which means the overall change is the results of multiple interacting factors). In other words, the decomposition method allows to quantify the relative contributions of the pre-defined factors to the change in the examined aggregate.

The LMD1 method applied in this article exhibits several desirable properties. First, the LMD1 fulfills the so-called Fisher's [8] tests (except the circular test[5]) - namely the time-reversal test[6] and the factor-reversal test[7]. Furthermore it is zero-value robust[8]. Detailed comparison of the properties of the decomposition methods can be found in [6].

It should be noted the additive refined Laspeyer's index decomposition methods (see e.g. [11], [7] or [1]; an example of the application of the additive refined Laspeyer's method on the cross-country energy consumption can be found in [9]) also fulfill the Fisher's factor-reversal and time-reversal tests. However, the application of the logarithmic-mean index decomposition method has advantages over the refined Laspeyer's methods. One of them is that LMD methods can be easily constructed for any number of factors, while the complexity of the interaction terms in the formulas of the refined Laspeyer's methods grows very quickly.

It is important to point out a strong advantage of the application of the decomposition methods that satisfy the factor-reversal test over the other decomposition methods. It is well known the magnitude of the unexplained residual of the actual change can be very high (even in orders of tens or hundreds of % of the actual change (though the Divisia based methods usually result in only marginal residual terms)).

[4]The change can be expressed as$\Delta V_{total} = V^T - V^0$ for the additive form, or $D_{total} = \frac{V^T}{V^0}$ for the multiplicative form

[5]The circular test has the form of $1 = D^{0S} D^{ST} D^{T0}$,where S is a point between time $t = 0$ and $t = T$ (this test can also be expressed in the (perhaps more clear) form of $D^{0T} = D^{0S} D^{ST}$). However, as shown by [8], the circular test is not met by any weighted aggregative with *changing weights*. Constant weights would, however, impose a serious questionability of the index method's usefulness.

[6]As the name suggests, to pass the time-reversal test, the method must exhibit a property, that an index number for a change from the period $t = 0$ to period $t = T$ must be a reciprocal value (for multiplicative form) of the index number for a change from period $t = T$ to period $t = 0$, i.e. $D^{0T} = \frac{1}{D^{T0}}$. For the additive form, the property is: $\Delta V^{0T} = -\Delta V^{T0}$.

[7]To pass the factor-reversal test (in the multiplicative form), the product of the individual factor contributions must equal the observed ratio of the aggregate, i.e. $D^{0T} = \frac{V^T}{V^0} = \prod_{k=1}^{n} D_k$ (for the additive form, the property is: $\Delta V^{0T} = V^T - V^0 = \sum_{k=1}^{n} \Delta V_k$). The decomposition method that satisfies this test is called *a perfect decomposition method* as it does not contain an unexplained residual.

[8]Which means the method can be used when there are 0 values in the dataset. In this article the occasional zero values in the dataset were handled in accordance with the recommendation suggested in [3], i.e. by replacing the zero values by a sufficiently small number ($\delta = 10^{-20}$). The methodological discussion of this type of zero values' handling is thoroughly discussed in [5].

As such, they leave an open question on accuracy of the results and provide a significant burden to the interpretation of the results (the more detailed discussion of the issue can be found in [2]).

Another advantage of the LMD1 method is its consistency in aggregation [4], a very desirable property that is not met by the alternative LMD2 decomposition (the difference between LMD1 and LMD2 is in the way the weights are computed) proposed by [3] (albeit LMD2 is also perfect in decomposition). Furthermore, the results from multiplicative LMD1 and additive LMD1 can be mutually linked[9], thus eliminating the need to estimate both multiplicative and additive forms, as one can be transformed into the other.

The relationship examined in the rest of the article focuses on the changes in the energy consumption E, using the relationship $E = Q \cdot EI$, where the energy consumption (measured in TOE) is defined as the product of continuous variables of the economic activity Q (measured by the value added) and the energy intensityEI (measured in TOE / 1000 Euro).

This basic relationship can be further rewritten as

$$E = Q \sum_i \frac{E_i Q_i}{Q_i Q} = Q \sum_i EI_i S_i \tag{1}$$

where $EI_i = \frac{E_i}{Q_i}$ and $S_i = \frac{Q_i}{Q}$. If we differentiate the Equation 1 by time, we can write:

$$\frac{\partial E}{\partial t} = \frac{\partial Q}{\partial t} \sum EI_i S_i + Q \sum_i \frac{\partial EI_i}{\partial t} S_i + Q \sum_i EI_i \frac{\partial S_i}{\partial t} \tag{2}$$

and if we divide this equation by $E = Q \sum_i EI_i S_i$, and rearrange:

$$\frac{\partial}{\partial t} ln\left(E\right) = \left(\frac{\partial Q}{Q \partial t}\right) Q \sum \frac{EI_i S_i}{Q \sum_i EI_i S_i} + Q \sum_i \frac{EI_i S_i}{Q \sum_i EI_i S_i} \left(\frac{\partial EI_i}{EI_i \partial t}\right) + Q \sum_i \frac{EI_i S_i}{Q \sum_i EI_i S_i} \left(\frac{\partial S_i}{S_i \partial t}\right) \tag{3}$$

The integration of Equation 3 yields:

$$ln\left(\frac{E^T}{E^0}\right) = \int_0^T \left(\sum_i w_i\left(t\right) \left(\frac{\partial}{\partial t} ln\left(Q\left(t\right)\right)\right) + \sum_i w_i\left(t\right) \left(\frac{\partial}{\partial t} ln\left(EI_i\left(t\right)\right)\right) + \sum_i w_i\left(t\right) \left(\frac{\partial}{\partial t} ln\left(S_i\left(t\right)\right)\right) \right) dt \tag{4}$$

where $w_i\left(t\right) = \frac{Q(t) EI_i(t) S_i(t)}{Q(t) \sum_i EI_i(t) S_i(t)}$. And exponentiating Equation 4 results in:

$$\frac{E^T}{E^0} = \underbrace{e^{\int_0^T \left(\sum_i w_i(t) \left(\frac{\partial}{\partial t} ln(Q(t))\right)\right) dt}}_{D_{activity}} \underbrace{e^{\int_0^T \left(\sum_i w_i(t) \left(\frac{\partial}{\partial t} ln(EI_i(t))\right)\right) dt}}_{D_{intensity}} \underbrace{e^{\int_0^T \left(\sum_i w_i(t) \left(\frac{\partial}{\partial t} ln(S_i(t))\right)\right) dt}}_{D_{structural}} \tag{5}$$

Since in practice, we can only observe discrete data, we need to work with the discrete version of the equations in question (and since the variables that constitute weights are also continuous, it motivates the use of some form of mean of two discrete values set apart by non-infinitesimal time). The discretization of Equation 5 results in:

$$\frac{E^T}{E^0} \approxeq \underbrace{e^{\left(\sum_i w_i(t^*) \left(ln\left(\frac{Q^T}{Q^0}\right)\right)\right)}}_{D_{activity}} \underbrace{e^{\left(\sum_i w_i(t^*) \left(ln\left(\frac{EI_i^T}{EI_i^0}\right)\right)\right)}}_{D_{intensity}} \underbrace{e^{\left(\sum_i w_i(t^*) \left(ln\left(\frac{S_i^T}{S_i^0}\right)\right)\right)}}_{D_{structural}} \tag{6}$$

with $t^* \in (0, T)$. A several possible options to determine the t^* can be used. The motivation to use logarithmic mean of weights was explained earlier, though one can indeed use a different scheme, such as e.g. arithmetic mean (resulting in the so called Arithmetic mean Divisia index), etc.

[9] Using the formula $\frac{\Delta V_{total}}{ln(D_{total})} = \frac{\Delta V_{effect_1}}{ln\left(D_{effect_1}\right)} = \frac{\Delta V_{effect_2}}{ln\left(D_{effect_2}\right)} \ldots = \frac{\Delta V_{effect_k}}{ln\left(D_{effect_k}\right)}$ (since $\frac{\Delta V_{x_k}}{ln\left(D_{x_k}\right)} = L\left(V^T, V^0\right)$) where ΔV terms represent the additive form of the effects (with $\Delta V_{total} = V^T - V^0 = \Delta V_{x_1} + \Delta V_{x_1} + \ldots + \Delta V_{x_n}$), D terms represent the multiplicative form of the effects (with $D_{total} = \frac{V^T}{V^0} = D_{x_1} D_{x_2} \ldots D_{x_n}$), where $L\left(V^T, V^0\right)$ is the logarithmic mean of V^T and V^0 .

However, we can recall that certain schemes (such as LMD1[10] used in this article) will result in perfect decomposition, thus:

$$\underbrace{\frac{E^T}{E^0}}_{D_{total}} = \underbrace{e^{\left(\sum_i \widetilde{w}_i(t^*)\left(ln\left(\frac{Q^T}{Q^0}\right)\right)\right)}}_{D_{activity}} \underbrace{e^{\left(\sum_i \widetilde{w}_i(t^*)\left(ln\left(\frac{EI_i^T}{EI_i^0}\right)\right)\right)}}_{D_{intensity}} \underbrace{e^{\left(\sum_i \widetilde{w}_i(t^*)\left(ln\left(\frac{S_i^T}{S_i^0}\right)\right)\right)}}_{D_{structural}} \tag{7}$$

The individual effects (activity, energy intensity, and structural) represent how the industry's energy consumption would change if the other factors did not change and therefore allow us e.g. to examine the actual energy efficiency improvements (via the energy intensity effect), in contrast to the simple computation of the energy intensity indicator.

4 RESULTS

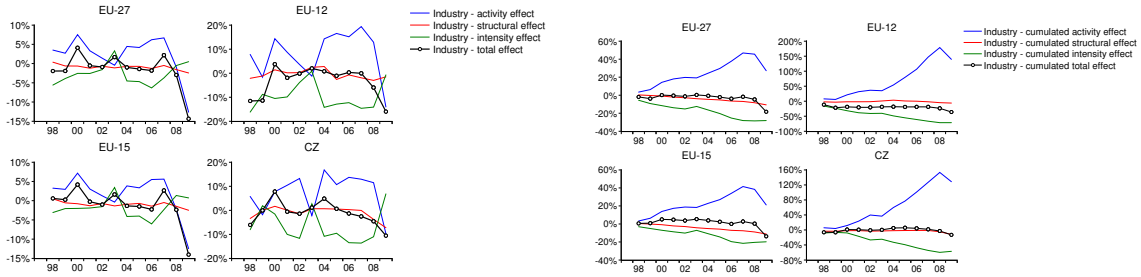
The results of decomposition (per Equation 7) are summarized in the Figure 2 (note the different scales of axes). While the relative picture might seem similar, the actual changes in the energy consumption were much more prevalent in the new member countries than in the original EU-15 countries. In 2008 (compared to the base year 1997), the overall change of the energy consumption was only +0.3% in the EU-15, but −23.7% in EU-12 (and $-2,84\%$ in CZ). We might note the cumulated impact of the period of crisis on the energy consumption in the industry. The changes in the energy consumption (compared to the level of 2007) amounted approx. to −16% in EU-15 and −21% in EU-12 (and −14.6% in CZ). An interesting point is that in the period of the crisis, unlike the both the EU-12 and CZ, only the EU-15 experienced (though rather marginal) increase in the energy intensity effect by the combined factor of 1.02.

The differences in the individual effects are also very prominent. For instance, in the peak in 2008 (comparing to the base year 1997), the change in the energy consumption due to the increase of the economic activity of the industry sector would (ceteris paribus) result in the increase by the factor of 1.38 in EU-15, but by the factor of 2.79 in EU-12 and 2.53 in CZ. The structural effect (representing the impact of changing the industry's structure, i.e. shares of individual sub-sectors) played only a minor role, as it contributed to the −8.9% change in EU-15, and −4.7% in EU-12 (−5% in CZ). The energy intensity effect, indicates that the energy efficiency improvements (not internal industrial restructuring towards inherently different sub-sectors) taking place in the new member countries were the major player in the energy intensity convergence[11]. Ceteris paribus this effect would lead to the change of the industrial energy consumption by the factor of 0.8 (≈ a reduction in consumption by 20%) in EU-15, by the factor of 0.29 in EU-12 (and by the factor of 0.40 in CZ).

As for the most influential sectors in the examined industry (due to the limited space of the article, the effects for the individual sub-sectors are not plotted here, but the figures are available from the author), it seems there are 4 commonly influential (from the energy consumption viewpoint) sub-sectors in both the original and new member countries. These sectors are (in decreasing order) Iron and steel (DJ), Petrochemical (DFGH), Non-metallic mineral products (DI) and Food processing (DA). For instance, these sectors are among top 5 sectors with both the highest activity effects (which would result in the increase of the energy consumption) and the most significant energy intensity effects (which would result in the decrease of the energy consumption) in *all* examined country groups. Overall, the major (from the industry's viewpoint) efficiency improvements took place in the Iron and steel sub-sector both in original and new member countries (though the decrease in energy consumption in 2008 and 2009 was mostly due to the decrease in the activity effect).

[10] In LMD1 the weights are given by $\widetilde{w}_i(t^*) = \frac{L(E_i^T, E_i^0)}{L(E^T, E^0)} = \frac{\frac{(E_i^T - E_i^0)}{(ln(E_i^T) - ln(E_i^0))}}{\frac{(E^T - E^0)}{(ln(E^T) - ln(E^0))}}$, where $L(a,b)$ is the logarithmic mean of a and b in the usual definition, i.e. $L(a,b) = \frac{(a-b)}{(ln(a)-ln(b))}$, with $L(a,a) = a$.

[11] Though it should be noted the energy intensity indicator (TOE / 1000 EUR in value added) is still about twice as high in EU-12 (0.19) or CZ (0.22) than in EU-15 (0.11).



(a) Individual year-by-year effects

(b) Cumulated effects

Figure 2 The examined industry's activity, structural and intensity effects

5 CONCLUSION

The results indicate the changes in energy consumption that can be attributed to the changes in economic activity, were almost 5 times as high (in terms of the relative change against the base period) in EU-12 and 4 times as high in CZ when compared to the EU-15 aggregate. The period of recession was more influential (in terms of the relative drop) in the energy consumption for the EU-12 aggregate than for the EU-15 aggregate. Despite the activity effect, the EU-12 countries managed to reduce the energy consumption by 36% (though CZ managed only 12% reduction), as opposed to the 14% reduction in EU-15. Unlike both EU-12 and CZ, the EU-15 experienced (albeit marginal) increase in the energy intensity effect by the combined factor of 1.02 during the period of crisis (2008 and 2009). In all examined groups, the major activity and intensity effects took place in the Iron and steel sub-sector.

References

[1] J. Albrecht, D. Francois, and K. Schoors, *A Shapley decomposition of carbon emissions without residuals*, Energy Policy **30** (2002), 727–736.

[2] B. W. Ang, *Analysis for policy making in energy: which is the preferred method?*, Energy Policy **32** (2004), 1131–1139.

[3] B. W. Ang and K. H. Choi, *Decomposition of aggregate energy and gas emission intensities for industry: a refined Divisia index method*, The Energy Journal **18** (1997), 59–73.

[4] B. W. Ang and F. L.. Liu, *A new energy decomposition method: perfect in decomposition and consistent in aggregation*, Energy **26** (2001), 537–548.

[5] B. W. Ang and N. Liu, *Handling zero values in the logarithmic mean Divisia index decomposition approach*, Energy Policy **35** (2007), 238–246.

[6] B. W. Ang and F. Q. Zhang, *A survey of index decomposition analysis in energy and environmental studies*, Energy **25** (2000), 1149–1176.

[7] E. Diezenbacher and B. Los, *Structural Decomposition Techniques: Sense and Sensitivity*, Economic Systems Research **10** (1998), 307–323.

[8] I. Fisher, *The making of index numbers; a study of their varieties, tests, and reliability*, Augustus M Kelley Pubs, 1967.

[9] V. Hajko, *Changes in the Energy Consumption in EU-27 Countries*, Review of Economic Perspectives **12** (2012), 3–21.

[10] IEA, *World Energy Outlook 2012*, International Energy Agency, 2012.

[11] J. W. Sun, *Changes in energy consumption and energy intensity: A complete decomposition model*, Energy Economics **20** (1998), 85–100.

Individual and panel modeling of foreign business cycle transmission on small open economics

Jana Hančlová[1]

Abstract. The purpose of this paper is to investigate a panel model of transmission of foreign business cycle on small open economics for a group of new member states in the EU (EU12) over the period 1995Q1 – 2012Q4. There are two different channels of business cycle transmission – financial markets and foreign trade. This paper will be investigated second one. We examine the relationship among the variables of domestic output growth, foreign output growth and the real exchange rate. We provide the panel unit roots tests using Levin, Lin and Chu, Breitung and Im, Pesaran, Shin tests assuming common unit root process and the ADF or PP Fisher-type tests for individual unit root process. Once the existence of a panel unit root has been established, we examine a long-run equilibrium relationship among the variables using panel cointegration test without structural breaks (seven Pedroni's tests, Kao (Engle-Granger based) tests and also combined Fisher - Johansen tests. There is evidence that both variables $\ln GDP_{it}$ and $\ln ERT_{it}$ are integrated of order one for the homogenous and the heterogeneous alternative of panel unit root tests. In our application we find evidence of cointegration between domestic output growth, foreign output growth and real exchange rate for most new EU countries. Our analysis is based on quarterly data using software Eviews 7.

Keywords: foreign trade, business cycle, DD schedule, panel unit root test, panel cointegration test.

JEL Classification: C23, C12, E32
AMS Classification: 91B26

1 Introduction

One of the most striking features of the business cycles across countries are the pattern of co-movement of output, inflation, interest rate and real equity prices. There are two different channels of business cycle transmission – financial markets and foreign trade. This paper will be investigated second one. We examine the relationship among the variables of domestic output growth, foreign output growth and the real exchange rate.

To analyze how output is determined we introduce the concept of aggregate demand for a country's output in an open economy. Aggregate demand is the amount of a country's goods and services demanded by households and firms throughout the world (Krugman, [10]). Just as the output of an individual good or services depends in part on the demand for it, a country's overall short-run output level depends on the average demand for its products. Corresponding aggregate demand for an open economy's output depends on various factors – consumption demand, investment demand, government demand, and the current account (net export demand), see [9], [14]. An important determinant of the current account is the real exchange rate, the ratio of the foreign price level measured in domestic currency to the domestic price level. We examine output market equilibrium - i.e. the relationship between output and the exchange rate (the DD schedule). We expect that any rise in the real exchange rate will cause an upward shift in the aggregate demand function and expansion of output, all else equal.

This paper examine the relationship among the variables of domestic output growth, foreign output growth and the real exchange rate (DD schedule) as transmission of foreign business cycle on small open economics for a group of new member states in the EU (EU12) over the period 1995Q1 – 2012Q4. We provide the panel unit roots tests for common or individual panel unit root process. We examine a long-run equilibrium relationship using panel cointegration tests without structural breaks. The paper is divided into four parts. Based on the introduction of the theoretical concepts of two different channels of business cycle, the second part specifies an empirical equilibrium model, panel unit roots tests and also panel cointegration tests without structural breaks. The third part deals with testing of panel unit roots and panel cointegration tests for a group of new member countries EU12 during the period 1995Q1 – 2012Q4. The final part summarizes the empirical results.

[1] VSB-Technical University of Ostrava/Faculty of Economics, Department of Systems Engineering, Sokolská 33, Ostrava 1,701 21, Czech Republic, jana.hanclova@vsb.cz.

2 Econometric methodology

One problem with results for individual countries is that they are often impaired by short data span that lowers the power of the unit root and cointegration test. In response recent studies have employed a panel unit root test combined with the panel cointegration test proposed by Pedroni [15] to exploit the extra power from combining cross-sectional and time series data.

2.1 An empirical model and data selection

As the goal to investigate the impact of foreign cycle on domestic business cyclical development we concentrate according to Juriová [7] to one transmission channel – foreign trade. The foreign business cycle has an impact on domestic economy through its exports (through changes in export demand and indirectly through changes in terms of trade). We can propose our empirical panel model as follows:

$$\ln GDP_{it} = \alpha_i + \gamma_i \cdot t + \beta_{1i} \cdot \ln GDP\_EU27_t + \beta_{2i} \ln ERT_{it} + \varepsilon_{it}, \quad (1)$$

where GDP_{it} is domestic demand (in millions of national currency, chain-linked volumes, reference year 2005, including 'euro fixed' series for euro area countries), GDP_EU27_t is real gross output EU27 countries and ERT_{it} is real effective exchange rate (deflator: consumer price indices - 27 trading partners, index, 2005=100). Subscript *t* denotes time ($t = 1,\ldots,T$) and subscript *i* is country $i = 1,\ldots, N$. We provide our analysis for a group of new member states in the EU (BG-Bulgaria, CZ-Czech Republic, EE–Estonia, CY-Cyprus, LV-Latvia, LT-Lithuania, HU-Hungary, MT-Malta, PL-Poland, RO-Romania, SK-Slovak Republic and SI-Slovenia). All variables are seasonally adjusted and transformed into natural logs. We use quarterly data over the period 1995Q1 – 2012Q4. Data are obtained from the EUROSTAT database.

2.2 Panel unit root tests

Panel unit root tests are similar, but not identical, to unit root tests carried out on a single series(see, [5]. We compute one or more of the following tests: Levin, Lin and Chu (LLC, [11]), Breitung (BR, [1]), Im, Pesaran and Shin (IPS, [6]), Fisher-type tests using ADF (FADF) and PP tests (see Maddala and Wu (FPP, [12])). Next we briefly describe the five panel unit root tests which are included in EViews 7.

For purposes of panel unit root testing, there are two natural assumptions that we can make about the ρ_i (the autoregressive coefficients in AR(1) process of y_{it} (i.e. $\ln GDP_{it}$ or $\ln ERT_{it}$). First, one can assume that the persistence parameters are common cross-sections so that $\rho_i = \rho$ for all i. The LLC, BR tests all employ this assumption. Alternatively, one can allow ρ_i to vary freely across cross-sections. The IPS, FADF and FPP tests are of this form.

Tests with common unit root process

LLC and BR employ a null hypothesis of a unit root. The LLC and BR tests consider the following basic augmented Dickey-Fuller ADF specification with deterministic variables (intercept and trend to vary across individuals) for each cross-section augmented process with the lag order s_i:

$$\Delta y_{it} = \rho_i \cdot y_{i,t-1} + \sum_{j=1}^{s_i} \omega_{ij} \Delta y_{it-j} + \alpha_i + \gamma_i \cdot t + \varepsilon_{it} \quad (2)$$

where $\rho_i = \rho$, ε_{it} are assumed to be independently distributed across *i* and *t* (*i=1, 2, ..., N; t=1, 2, ..., T*). It is worth noticing that throughout the paper we consider specification that includes a linear time trend. The null hypothesis of panel unit root is: $H_0 : \rho_1 = \rho_2 = \cdots = \rho_N = \rho = 0$ (there is unit root) against the alternative hypothesis $H_1 : \rho_1 = \rho_2 = \cdots = \rho_N = \rho < 0$ (there is no unit root). The Breitung method differs from LLC. Breitung suggests in [1] a test statistic that does not employ a bias adjustment whose power is substantially higher than that of LLC test. (First adjustment, only the autoregressive portion is removed when constructing the standardized proxies and second adjustment, the proxies are transformed and detrended).

Tests with individual unit root processes

The Im, Pesaran, and Shin (IPS), and the Fisher-ADF and Fisher-PP tests all allow for individual unit root processes so that ρ_i may vary across cross-sections. IPS test begin by specifying a separate ADF regression for

each cross section in equation (2). The null hypothesis may be written as, $H_0 : \rho_i = 0$ for all i and the alternative hypothesis is given by:

$$H_1 : \begin{cases} \rho_i = 0 & for\ i = 1,\ldots,N_1 \\ \rho_i < 0 & for\ i = N_1+1, N_1+2,\ldots,N, \end{cases} \tag{3}$$

where the i may be reordered as necessary. The IPS t-bar statistic is defined as the average of individual ADF statistics. In Monte Carlo experiments it was shown that if a large enough lag order is selected for the underlying ADF regressions, then the small sample performance of the t-bar test is reasonable and generally better than the LLC test. Fisher ADF and PP tests combine the *p*-values from individual unit root tests. This idea has been proposed by Maddala and Wu in [12]. The null and alternative hypotheses are the same as for the IPS in equation (3). In fact, the power of the Z-test is in some cases more than three times that of the IPS test. It seems that this test outperforms the other tests and is recommended.

2.3 Panel cointegration testing

There is a number of procedures for computing panel cointegration tests, see [5]. We provide a brief description of the cointegration tests supported by Eviews 7. The Pedroni and Kao tests are based on Engle-Granger two-step (residual-based) cointegration tests. The Fischer test is a combined Johansen test.

Pedroni (Engle-Granger based) cointegration tests

The Engle-Granger [2] cointegration test is based on an examination of the residuals of a spurious regression performed using I(1) variables. If the variables are cointegrated then the residuals should be I(0). On the other hand if the variables are not cointegrated then the residuals will be I(1). Pedroni [15] and Kao [8] extend the Engle-Granger framework to tests involving panel data. Pedroni proposes several tests for cointegration that allow for heterogeneous intercepts and trend coefficients across cross-sections. Consider the following regression according to [4]:

$$\ln GDP_{it} = \alpha_i + \gamma_i \cdot t + \beta_{1i} \cdot \ln GDP\_EU27_t + \beta_{2i} \ln ERT_{it} + \varepsilon_{it}, \tag{4}$$

where variables are assumed to be integrated of order one (I(1)). Parameters α_i and γ_i are individual and trend effects. Under the null hypothesis of no cointegration the residuals $\hat{\varepsilon}_{it} = e_{it}$ should be I(1). The general approach is to obtain residuals e_{it} and then to test whether they are I(1) by running the auxiliary regression

$$e_{it} = \rho_i e_{it-1} + \sum_{j=1}^{s_i} \omega_{ij} \Delta e_{it-j} + \xi_{it} \tag{5}$$

for each cross-section. Pedroni describes various methods of constructing statistics for testing for null hypothesis of no cointegration ($\rho_i = 1$). There are two alternative hypotheses: the **homogenous alternative** $(\rho_i = \rho) < 1$ for all *i*, and **the heterogeneous alternative** $(\rho_i < 1)$ for all *i*.

Kao (Engle-Granger based) cointegration tests

The Kao test follows the same basic approach as the Pedroni tests, but specifies cross-section specific intercepts and homogeneous coefficients on the first-stage regressors. In the bivariate case described in Kao [8], we have

$$\ln GDP_{it} = \alpha_i + \beta_2 \ln ERT_{it} + \varepsilon_{it} \qquad \ln GDP_{it} = \ln GDP_{it-1} + u_{it}, \qquad \ln ERT_{it} = \ln ERT_{it-1} + \nu_{it}, \tag{6}$$

More generally, we may consider running the first stage regression equation (2), requiring α_i *to be heterogeneous*, β_{2i} *to be homogeneous across cross-sections*, and setting all of the trend coefficients γ_i and β_{1i} to zero. Kao then runs either the pooled auxiliary regression

$$e_{it} = \tilde{\rho}_i e_{it-1} + \sum_{j=1}^{s} \omega_j \Delta e_{it-j} + \xi_{it}. \tag{5}$$

Under H_0: no cointegration, Kao shows defined variants statistics (Eviews 7, p. 702) converging to N(0; 1) asymptotically.

Combined individual Fisher (combined Johansen) cointegration tests

Fisher derives in [3] a combined test that uses the results of the individual independent tests. Maddala and Wu [12] use Fisher's result to propose an alternative approach to testing for cointegration in panel data by combining tests from individual cross-sections to obtain at test statistic for the full panel. If π_i is the *p*-value from an individual cointegration test for cross-section *i*, then under the null hypothesis for the panel, $-2\sum_{i=1}^{N}\log(\pi_i)\chi^2_{2N}$.

Eviews reports χ^2 value base on MacKinnon-Haug-Michelis *p*-values for Johansen's cointegration trace test and maximum eigenvalue test.

3 Empirical results

We provide the panel unit roots tests for common or individual panel unit root process. We examine a long-run equilibrium relationship among the variables $\ln GDP_{it}, \ln ERT_{it}$ and $\ln GDP\_EU27_t$ using panel cointegration tests.

3.1 Panel unit root tests

Table 1 reports the results concerning panel unit root tests for variables $\ln GDP_{it}$ and $\ln ERT_{it}$ in level or 1st difference (Δ) including deterministic components (intercept (c) or trend and intercept (c+t)) with automatic lag length selection based on Schwarz information criterion and Bartlett kernel bandwidth selection. The probabilities for Fisher tests are computed using an asymptotic Chi-square distribution. All the other tests assume asymptotic normality.

variables	**c** or **c+t**	H_0: Unit root (assumes **common** unit root)		H_0: Unit root (assumes **individual** unit root process)		
		LLC	**BRE**	**IPS**	**FADF**	**FPP**
$\ln GDP_{it}$	c+t	2.732 (0.997)	2.361 (0.991)	4.352 (1.000)	10.718 (0.991)	7.185 (0.999)
$\Delta\ln GDP_{it}$	c	-15.198*** (0.000)	X	-16.328*** (0.000)	239.929*** (0.000)	355.047*** (0.000)
$\ln ERT_{it}$	c+t	-3.154*** (0.001)	0.607 (0.728)	-3.825*** (0.0001)	54.986*** (0.0003)	51.766*** (0.0008)
$\Delta\ln ERT_{it}$	c	-18.688*** (0.000)	X	-18.445*** (0.000)	296.286*** (0.000)	331.575*** (0.000)
individual variables		**ADF**	**PP**			
$\ln GDP\_EU27_t$	c+t	-0.360 (0.987)	-0.003 (0.996)			
$\Delta\ln GDP\_EU27_t$	c	-3.393** (0.015)	-3.440** (0.013)			

Table 1 Panel unit root tests

Probabilities values are in brackets. *** or ** denotes statistical significance level at the 1% or 5%, respectively.

Our results provide evidence that we cannot reject the null hypothesis at the 5% level of significance for our variables $\ln GDP_{it}$ in level and we can reject null hypothesis for these variables in first difference i.e. *there is evidence that both variables* $\ln GDP_{it}$ *are integrated of order one for the homogenous and the heterogeneous alternative of panel unit root tests. These tests also suggest that* $\ln ERT_{it}$ *variables do not contain common or individual panel unit root after excluding deterministic linear trend.* We can also conclude that exogenous $\ln GDP\_EU27_t$ variable is I(1) using simple ADF and PP unit root tests at the 5% level of significance. We proceed on this basis to test for panel cointegration.

3.2 Panel cointegration testing

Once the existence of a panel unit root has been established, the issue arises whether there exists a long-run equilibrium relationship among the variables $\ln GDP_{it}, \ln ERT_{it}$ and $\ln GDP\_EU27_t$ for the period 1995Q1 – 2012Q4 in new member the EU countries. Given that each variable is integrated of order one, we first test for

Pedroni panel cointegration test which evaluate the null against both the homogeneous and the heterogeneous alternatives. The results are reported in table 2 for deterministic individual intercept and individual trend. None of the seven Pedroni tests reveal any evidence of common or individual cointegration. Thus, *we can conclude that the variables* $\ln GDP_{it}, \ln ERT_{it}$ *and* $\ln GDP\_EU27_t$ *do not share a long-run equilibrium relationship at 5% level of significance using both the homogeneous and the heterogeneous alternatives of seven Pedroni tests.*

Alternative hypothesis: **common** AR coefs.			Weighted Statistic		Alternative hypothesis: **individual** AR coefs.		
	Statistic	Prob.		Prob.		Statistic	Prob.
Panel v-Statistic	0.633	0.263	0.073	0.471	Group rho-Statistic	0.162	0.564
Panel rho-Statistic	-0.318	0.375	-0.806	0.210	Group PP-Statistic	-0.936	0.175
Panel PP-Statistic	-0.719	0.236	-1.552	0.060	Group ADF-Statistic	2.102	0.982
Panel ADF-Statistic	1.781	0.963	1.429	0.923			

Table 2 Pedroni's panel unit roots test

The result of the *Kao test* was tested under assumption – no deterministic trend. Kao ADF t-statistic was -2.646*** with *p*-value 0.004, which means that we reject the null hypothesis of no panel cointegration with common estimated parameter for natural logs the exchange rate. *The Kao test produces significant evidence cointegration among the variables* $\ln GDP_{it}, \ln ERT_{it}$ *and* $\ln GDP\_EU27_t$ *at 5% level of significance using the heterogeneous intercept and the homogeneous* β_{1i} and β_{2i} *across cross-sections.*

The results of the *Johansen/Fisher panel cointegration test* for the full panel presents table 3. We also assume the trend effect. To determine the number of cointegration relationship in our panel DD model we proceed sequentially from number of CE(s) =0 to 2 until we fail to reject. The Fisher trace statistics, which are reported in the second column in table 3 with p-value in the third column, recommend one cointegrating relationship for the full panel model. The second part of the output in table 3 provides the same results for the Fisher maximum eigenvalue statistic. The alternative combined *individual Fisher/Johansen test provide evidence of individual cross-section cointegration among the variables for BG, CZ, LT, HU, MT, SK and SI countries at 10% level of significance.*

Hypothesized No. of CE(s)	Fisher Stat. (from trace test)	Prob.	Fisher Stat. (from max-eigen. test)	Prob.
None	68.08	0.000	79.07	0.000
At most 1	15.57	0.903	19.48	0.726
At most 2	7.318	0.999	7.318	0.999

Table 3 Unrestricted Cointegration Rank Test (Trace and Maximum Eigenvalue)

4 Conclusions

In this paper examine output market equilibrium - i.e. the relationship between output and the exchange rate (the DD schedule in new EU12 countries over the period 1995Q1 – 2012Q4. We provide the panel unit roots tests for common or individual panel unit root process. We examine a long-run equilibrium relationship among the variables $\ln GDP_{it}, \ln ERT_{it}$ *and* $\ln GDP\_EU27_t$ using panel cointegration tests without structural breaks. The main results obtained are summarized below:

- There is evidence that *both variables* $\ln GDP_{it}$ *and* $\ln ERT_{it}$ *are integrated of order one for the homogenous and the heterogeneous alternative of panel unit root tests*. We can also conclude that exogenous $\ln GDP\_EU27_t$ variable is I(1) at the 5% level of significance. We proceed on this basis to test for panel cointegration.
- None of the seven *Pedroni tests reveal any evidence of common or individual cointegration*. Thus, we can conclude that the variables $\ln GDP_{it}, \ln ERT_{it}$ and $\ln GDP\_EU27_t$ do not share a long-run equilibrium relationship at 5% level of significance using both the homogeneous and the heterogeneous alternatives of seven Pedroni tests.

- *The Kao test produces significant evidence cointegration* among the variables $\ln GDP_{it}$, $\ln ERT_{it}$ and $\ln GDP\_EU27_t$ at 5% level of significance *using the heterogeneous intercept and the homogeneous* β_{1i} *and* β_{2i} *across cross-sections.*
- *The Fisher trace and eigenvalue statistics recommend one cointegrating relationship for the full panel model. The alternative combined individual Fisher/Johansen tests provide evidence of individual cross-section cointegration for seven countries at 10% level of significance.*

The results of panel cointegration testing support our conclusion that there is the evidence of panel long-term equilibrium in our output markets (i.e. the relationship between output and the exchange rate) with individual deterministic trend, which is probably indicator of the convergence process. Fisher/Johansen test recommended due to the verification of the results heterogeneous alternative, which can be associated with the conditional convergence process rather than absolute convergence in the most of new EU Member States in the period 1995-2012. In the next part of our research we compare results for estimation of our panel cointegration relationship in an DD model n new EU12 countries. We use estimators using Fully Modified Ordinary Least Squares (FMOLS) panel estimators. We also used Granger Causality test to perform panel data specific testing. We also include structural breaks to panel unit root tests and panel cointegration tests.

Acknowledgements

This research was supported by the Student Grant Competition of Faculty of Economics, VSB-Technical University of Ostrava within project (SP2013/91) and also by the European Social Fund within the projects CZ.1.07/2.3.00/20.0296.

References

[1] Breitung, J.: The Local Power of Some Unit Root Tests for Panel Data. *Advances in Econometrics* **15** (2000), 161-178.
[2] Engle, R. F. and Granger C. W. J.: Co-integration and Error Correction: Representation, Estimation, and Testing. *Econometrica* **55** (1987), 251-276.
[3] Fisher, R. A.: *Statistical Methods for Research Workers.* Oliver & Boyd, Edinburgh, 1932.
[4] Hančlová, J., Lukáčik, M. and Szomolányi, K.: A VEC Model of the Czech and Slovak Economies. In: *Proceedings of 28th International Conference on Mathematical Methods in Economics 2010.* University of South Bohemia, České Budějovice, 2010, 208-213.
[5] Hančlová, J.: Output convergence amongst the NUTS2 regions of Visegrad countries. In: *Proceedings of 28th International Conference on Mathematical Methods in Economics 2010.* University of South Bohemia, České Budějovice, 2010. 201-207.
[6] Im, K. S., Pesaran, M. H. and Shin, Y.: Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics* **115** (2003), 53-74.
[7] Juriová, J. Influence of cyclical development of the most significant foreign-trade partners on small open economy (VAR approach). In: *Proceedings of the 30th international conference Mathematical methods in economics 2012* (Ramík, J. and Stavárek, D.). Silesian University, School of Business Administration, Karviná, 2012, 390-395.
[8] Kao, Ch.D.: Spurious Regression and Residual-Based Tests for Cointegration in Panel Data. *Journal of Econometrics*, **90** (1999), 1-44.
[9] Kiszová, Z., and Nevima, J.: Usage of analytic hierarchy process for evaluating of regional competitiveness in case of the Czech Republic. In: *Proceedings of the 30th international conference Mathematical methods in economics 2012* (Ramík, J. and Stavárek, D.). Silesian University, School of Business Administration, Karviná, 2012, 402-407.
[10] Krugman, P. R., and Obsfeld, M.: *International Economics – Theory and Policy.* Pearson International Edition, New York, 2009.
[11] Levin, A., Lin, C. F. and Chu, C.: (2002). Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties. *Journal of Econometrics* **108** (2002), 1-24.
[12] Maddala, G. S. and Wu, S.: A Comparative Study of Unit Root Tests with Panel Data and A New Simple Test. *Oxford Bulletin of Economics and Statistics* **61** (1999), 631-52.
[13] Narayan, P. K. and Smyth, R.: Energy consumption and real GDP in G7 countries: New evidence form panel cointegration with structural breaks. *Energy Economics* **30** (2008), 2331-2341.
[14] Pankova, V.: Permanent Income and Consumption. In: *Proceedings of the 30th international conference Mathematical methods in economics 2012* (Ramík, J. and Stavárek, D.). Silesian University, School of Business Administration, Karviná, 2012, 670-674.
[15] Pedroni, P.: Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors. *Oxford Bulletin of Economics and Statistics* **61** (1999), 653-70.

Research of the convergence of unemployment rate of university graduates in the Czech Republic

Jana Hanclova[1], Milan Simek[2]

Abstract. The article is concerned with modeling and assessment of the process of convergence of the unemployment rate of the graduates of Bachelor's and Master's Degree study programs at the Faculties of Economics in the Czech Republic within the period of 2002 – 2012. The unemployment rate of university graduates is the ratio of the number of unemployed graduates to the difference between the total number of the graduates and the number of the graduates who continue studying at university. The theoretical concept of convergence is based on the absolute (unconditional) respectively relative (conditional) form and the method of beta convergence is used within the article. The paper empirically estimates the panel model by the means of the pooled least squares method with fixed overview effects for the Bachelor's and Master's degrees of studies separately. The results show that there are vast differences in the employment of graduates on the labor market especially after the completion of the Bachelor's degree. The results of the estimated panel models further document the (non)presence of convergence, but also the speed of the process. There are also significant differences in the convergence behavior of the unemployed graduates from individual Faculties of Economics of Czech universities.

Keywords: unemployment rate of graduates, absolute and conditional convergence, Faculties of Economics, the Czech Republic, panel unit roots tests, ADF models.

JEL Classification: C23, E24
AMS Classification: 91B40

1 Introduction

The paper focuses on the use of the concept of convergence while examining the level of unemployment of university graduates in the Czech Republic. The unemployment rate of the target group is very closely related to the issues of their chances at the labor market and to the so-called employability.

Employability may be defined as a combination of the factors that enable individuals to progress toward obtaining employment or enter into employment, continuance of employment and progress in their careers. It is a complex concept, involving not only personality traits, skills, attitudes and motivation of each individual, but also other external factors, which go beyond policy in the area of education and expert training, such as labor market regulation, demographics, economic structure and economic situation in general [6].

The European Commission considers the group of young people up to 24 years of age, including university graduates to be a vulnerable group at the labor market. The problems in employment at the labor market compared with other groups of individuals there often lead to increased levels of unemployment. Unemployment of young people is influenced by many factors [14] and [11]. For example Balcar [1] states that employers consider the reached level of soft skills to be one of the important factors of good employability of graduates. At the same time the studies of the European labor market [2] show the importance of soft skills for employment at the labor market and recommend systemic development of soft skills within the education system. The position of the graduates at the labor market and their employability in the Czech Republic is often monitored by means of indicators focused on the signs of their success in getting employment. From a quantitative point of view, it is a specific unemployment rate of university graduates (e.g. according to the area of studies, faculty and type of school).

Unemployed graduate is defined by the Ministry of Labour and Social Affairs [13] as a job seeker registered at labor offices of the Czech Republic by their permanent residence at a certain date, whose time after the successful completion of his studies did not exceed 2 years. *The unemployment rate of university graduates* is then defined as the number of unemployed university graduates to the difference between the total number of gradu-

[1] VSB-Technical University of Ostrava/Faculty of Economics, Department of Systems Engineering, Sokolská 33, Ostrava 1,701 21, Czech Republic, jana.hanclova@vsb.cz.
[2] VSB-Technical University of Ostrava/Faculty of Economics, Department of National Economy, Sokolská 33, Ostrava 1, 70121, Czech Republic, milan.simek@vsb.cz.

ates and the number of graduates who continue to study at a university. In practice, there is also used the standardized unemployment rate of graduates which is the unemployment rate of graduates without the different unemployment rate in different regional labor markets [13].

In the Czech Republic the number of unemployed graduates can be tracked always on the 30th April and 30th September. Koucký and Zelenka [13] reported that "With the possibility to split the graduates into three periods also depending on how long before the survey they completed the study, it is also possible to map in a detail the development of their employability in the transition from school to the labor market. This is the period of 0-0.5 year, 0.5 - 1 year and 1-2 years." Though in this context, the authors draw attention to the varying length of time between the individual surveys. When expressing the condition for a longer period the so-called long-term standardized unemployment rate is used, which is the average rate calculated from the standardized unemployment rates of the last 8 years [13].

Taking into account the above-mentioned facts for assessing employability of university graduates, it seems preferable to follow unemployment rate over time and based on the results obtained, and only then to consider the changes in the level of employment (employability) of graduates of a particular university, faculty, or area of studies. For example the principle of convergence may be used to monitor this. Žďárek [19] states that the concept of convergence which had been created in connection with the economic theory of growth and subsequently applied in many areas of modern economics is used for example to investigate price changes and price levels over time. The concept of convergence exists as absolute or relative one. *The absolute (unconditional)* convergence is a process where e.g. countries converge to one of the (common) values of the given parameter (the so-called steady state), regardless of the initial starting position. *The conditional (weak, relative)* convergence is a situation where the countries with different initial states of the given indicator are approaching each other, but the single common state is not reached (similar economies should converge). This approach analyzes so-called beta and sigma convergence.

This article explores the beta convergence by testing the panel stationarity of the unemployment rate of graduates of the Faculties of Economics in the Czech Republic for a group of Bachelor's as well as the following Master's Degree programs in 2002-2008 (2012). The paper is structured into four parts. Based on the introduction of the theoretical concepts of beta convergence with focus on the unemployment of university graduates, the second part specifies panel unit roots tests. The third part deals with data analysis, testing of panel unit roots and estimation of the adequate econometric ADF model including verification and economic interpretation. The final part summarizes the empirical results.

2 Panel unit roots tests for convergence

Panel unit root tests for convergence among series, or group-wise convergence, utilize Bernard and Durlauf's definition of time series convergence for long-run output movements [5], where two (or more) countries have converged when long-run forecasts of per capita output differences tend to zero as the forecasting horizon tends to infinity. In the bivariate context, tests for time series convergence require cross-country per capita output differences to be stationary. In the multivariate or panel context, a group of countries have converged if the null hypothesis that the difference between each country's output and the cross-sectional mean has a unit root can be rejected in favor of the alternative hypothesis that each difference is stationary, see Pesaran [17] and [10]. Several papers use panel methods to investigate output convergence (Ben-David [3], Evans and Karras [8], and Fleissig and Strauss [9], among others) or inflation convergence (Lee and Wu [15], Kočenda and Papell [12]).

In the panel framework, testing for (stochastic) convergence of a group of N time series requires studying the dynamic properties of the series differential with respect to the cross-sectional mean. Group-wise (stochastic) convergence implies according to Lopez and Papell [16] that:

$$\lim_{t\to\infty} E\left(y_{i,t+k} + \sum_{j=1}^{N} \frac{y_{i,t+k}}{N} \middle| I_t \right) = \eta_i \quad for\ i = 1,\ldots,N, \tag{1}$$

where I_t represents the information set available at time r. If $\eta_i = 0,$ the convergence follows Bernard and Durlauf's definition (see in [4]) of absolute convergence $\eta_i \neq 0$ the convergence is conditional or relative as defined by Durlauf and Quah [7], which implies that the series have converged toward a time-invariant equilibrium differential. We use the standard panel unit root tests to account for restriction on the intercepts when testing for group-wise convergence. Let consider the following system of ADF regressions:

$$\Delta y_{it} = \alpha_i + \beta_i y_{i,t-1} + \delta \cdot trend + \sum_{j=1}^{k_i} \phi_{ij} \Delta y_{i,t-j} + \varepsilon_{it} \quad for\ i = 1,\ldots,N, t = 1,\ldots,T, \tag{2}$$

where $\beta = \beta_i$ means the homogenous rate of convergence, k_i the lagged first differences that account for serial correlation and $\varepsilon_{it} \sim N(0, \sum)$, where $\sum$ is the non-diagonal covariance matrix. The null and alternative hypotheses tested are $\beta = 0$ and $\beta < 0$. While it would be desirable to allow for heterogeneous rates of convergence, the choices are problematic. The alternative hypothesis for these tests, however, is that $\beta_i < 0$ for at least one *i*, which is not economically relevant for investigating convergence among a group of countries.

3 The empirical results

When assessing the employability of graduates we will examine the convergence of the unemployment rate of graduates from the Faculties of Economics in the Czech Republic for the group of Bachelors and Engineers (Ing.).

Analysis of the unemployment rate of graduates

Unemployment rate of graduates of Bachelor's (URF_bc) or Master's (URF_mgr) study programs of the Faculties of Economics in the Czech Republic from 2002 to 2012 is shown in Figure 1.

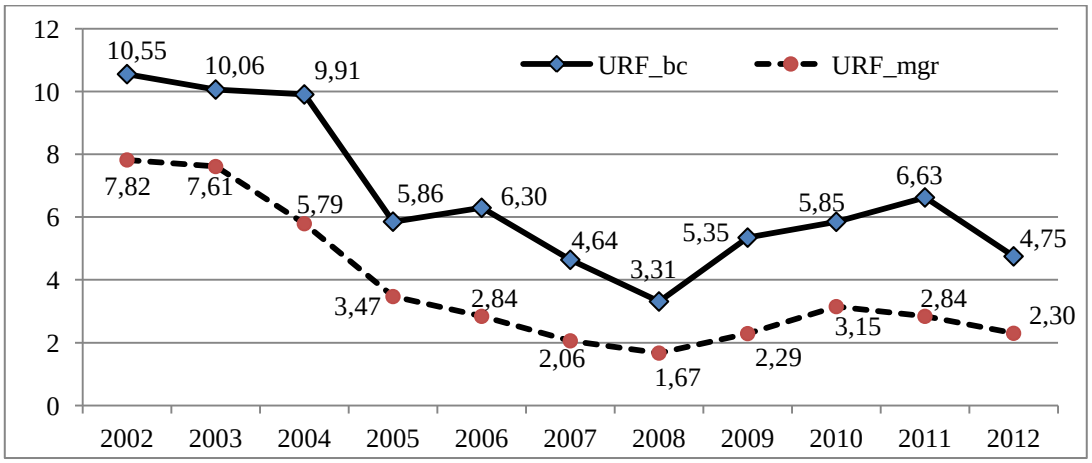


Figure 1 The development of the unemployment rate of graduates from the Faculties of Economics in the Czech Republic (%).

Source: The Education Policy Centre [18]

The level of the unemployment rate of the graduates of the surveyed faculties with economic studies was different for graduates from Bachelor's and Master's Degree programs throughout the analyzed period. In the case of the Bachelor's Degree of studies the unemployment rate was by 2-3 percentage points higher than for Master's Degree graduates. This situation has probably been caused by the reluctance of employers to employ the graduates with Bachelor's Degrees because of the lack of suitable jobs for such type of graduates. Some employers also did not consider this type of the university education to be a full-value one.

In the period under survey there have also been changes in the level of the unemployment of the graduates from the Faculties of Economics. In the period 2002-2008 there was a gradual decrease of the values of the unemployment rate of the Bachelor's Degree graduates from 10.6% to 3.3% and in the case of Master's Degree study programs from 7.8% to 1.7%. These conditions are related to the development of the Czech economy and labor market requirements as in this period the employers demanded large volume of manpower. After 2008 there was a significant reduction in the employability of these graduates, as evidenced for example by the unemployment rate 6.6% in the case of the graduates from Bachelor's Degree programs and 2.8% in the case of the graduates from Master's Degree programs. This development was influenced by the factors of the economic crisis of the Czech labor market. For the purpose of further research of convergence the time period 2002 - 2008 will be selected when there was a long-term decline in the observed unemployment rate.

Research of the beta convergence for the group of Faculties of Economics with Bachelor's programs

When estimating the panel model in equation (2) we proceed on the basis of the following indication of the variables - URF_{it} the unemployment rate of Bachelors *i* of the faculty in the year *t*, where *i=1,...,25* and *t=2002,..., 2008*; URF_{At} - the unemployment rate of Bachelors from all the Faculties of Economics in the Czech Republic in the year *t*. Furthermore, we define the difference:

$$y_{it} = \left|\ln(URF_{it}) - \ln(URF_{At})\right| \quad for\ i = 1,\ldots,N, t = 1,\ldots,T. \tag{3}$$

The Estimation of the panel model (2) has been performed alternatively by SUR (Seemingly Unrelated Regression) by the method of pooled least squares with cross-section fixed effects (PLS_FE). The verification of the estimated model (autocorrelation, heteroskedasticity, normality of residual components) highlighted the more robust estimation using the PLS_FE method. The trend component has also not been statistically significant at 5% level of significance. The resulting estimation of the modified model in the years 2002 – 2008 was the following:

$$\Delta y_{it} = \hat{\alpha}_i - 1,07\, y_{i,t-1} + \sum_{j=1}^{k_i} \hat{\phi}_{ij} \Delta y_{i,t-j} \qquad R^2 = 0,60 \quad N = 26,\ T = 7. \tag{4}$$

In the equation step of testing convergence using the ADF model (4) we test the hypothesis $H_0 : \beta = 0,$ that we reject at 5% level of Significance (prob. = 0,000) in favor of the alternative hypothesis $H_0 : \beta < 0.$ We come to the conclusion that y_{it} is stationary in terms of trends. This indicates the presence of convergence in the period examined in case of the faculties in question. In the next step, we tested the hypothesis of statistical significance of intercept. If we reject $H_0 : \alpha_i = 0,$ it is the case of conditional convergence, otherwise it is the case of absolute convergence. Conditional convergence of the unemployment rate of graduates from the Faculties of Economics with Bachelor Degree level has been proved for the majority of them - 18 faculties (VSEUJEP, ESFMU, OPFSU, FEZCU, FESUPA, FPVUT, FMEUTB, FFUVSE, FMVVSE, FPVSE, NFVSE, PEFMENDELU, VSH, VSFS, UNYP, VSEKS, VSO and VSKE) and it represents that the above-mentioned faculties converge to a various stable state in comparison with all the faculties and after reaching its stable state grow at the same pace as all the other faculties. Absolute convergence was confirmed for 8 faculties (FIMUHK, EFVSBTUO, FMKUTB, FISVSE, FMJHVSE, VSKV, SAVS and SCVI) and we may state that these faculties have the same steady state (long-term equilibrium level) of the unemployment rate of graduates as all the Faculties of Economics with Bachelor's study program (absolute variations with increasing time generate stationary process with zero medium value).

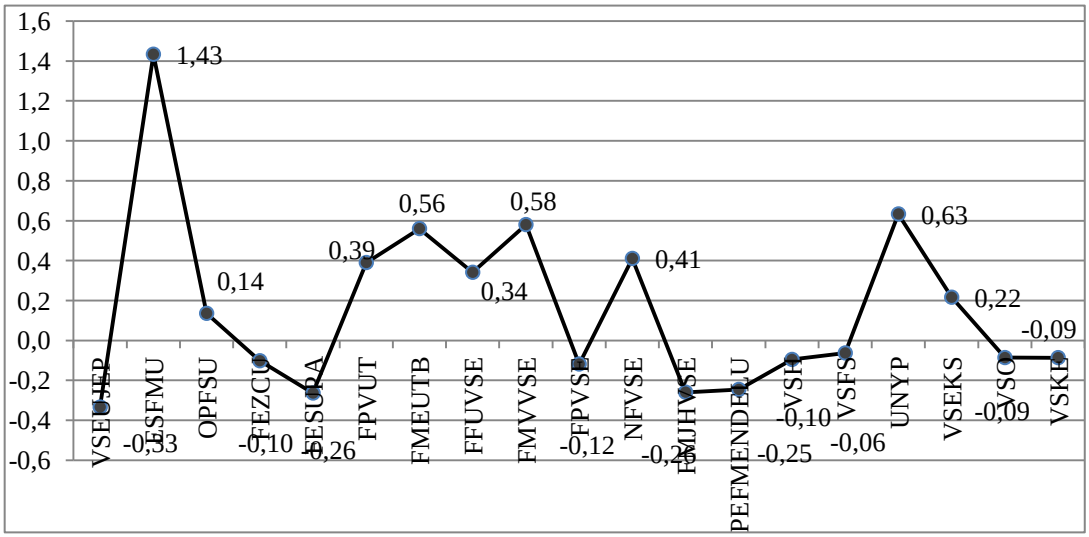


Figure 2 Conditional convergence *URF* for Bachelors in terms of the direction of approaching

The estimated level constant $\hat{\alpha}_i$ was further decomposed $\hat{\alpha}_i = \hat{\alpha}_0 + \hat{\alpha}_{1i} = \underset{sign.=0,00}{0,766} + \hat{\alpha}_{1i}.$ The development of all the estimated parameters $\hat{\alpha}_{1i}$ statistically different from zero is depicted in Figure 2 and we may observe

the convergence "from the top", i.e. the approaching of the faculty unemployment rate of Bachelors towards the national level by reducing unemployment (*positive trend*) or "from the bottom", i.e. the approximation of the faculty unemployment rate of graduates towards the national level by negative increase of the unemployment rate. We have ranked namely the faculties ESFMU, UNYP, FMVVSE and FMEUTB to the group approaching from the top. The ones converging from the bottom are namely VSEUJEP, FESUPA, FMJHVSE a PEFMENDELU.

Research of beta convergence for the group with Master's Degrees Faculties of Economics

The same procedure has been implemented for the group of the 20 Faculties of Economics with Master's Degree studies. The testing of panel unit root rejected the null hypothesis of unit root (assuming common or individual unit root process) including individual effects and linear trends. We use Levin, Lin & Chu or Breitung t-test and also Im, Pesaran and Shin W/test, ADF or PP Fisher Chi-square tests at 5% level of significance.

The estimation of the adequate ADF model according to the equations (2) by the method of pooled least squares with cross-section of fixed effects (PLS_FE) in the years 2002 – 2008 is summed up by the equation:

$$\Delta y_{it} = \hat{\alpha}_i - 0,88 y_{i,t-1} - 0,88 \cdot trend + \sum_{j=1}^{k_i} \hat{\phi}_{ij} \Delta y_{i,t-j} \qquad R^2 = 0,58 \quad N = 20,\ T = 7. \tag{5}$$

The results indicate that the trend stationary process with the inclusion of a deterministic trend, which is statistically significant at the 5% level of significance is according to the expectations negative. The unemployment rate of graduates of Master's Degree from the Faculties of Economics converges to the national level of this indicator. The estimation and decomposition of the level constant $\hat{\alpha}_i = \hat{\alpha}_0 + \hat{\alpha}_{1i} = \underset{sign.=0,00}{6,27} + \hat{\alpha}_{1i}$ indicate a higher baseline average level of the unemployment rate of Master's Degree graduates with respect to the national level. Figure 3 presents the development of the constant $\hat{\alpha}_{1i}$ for statistically significant values. From the top especially the faculties EFVSBTUO, FPVUT, OPFSU and private faculty VSH are approaching the national level. Conversely, the undesirable direction of convergence increasing the difference in the unemployment rate of Master's Degree students on the national level is shown primarily by VSEUJEP.

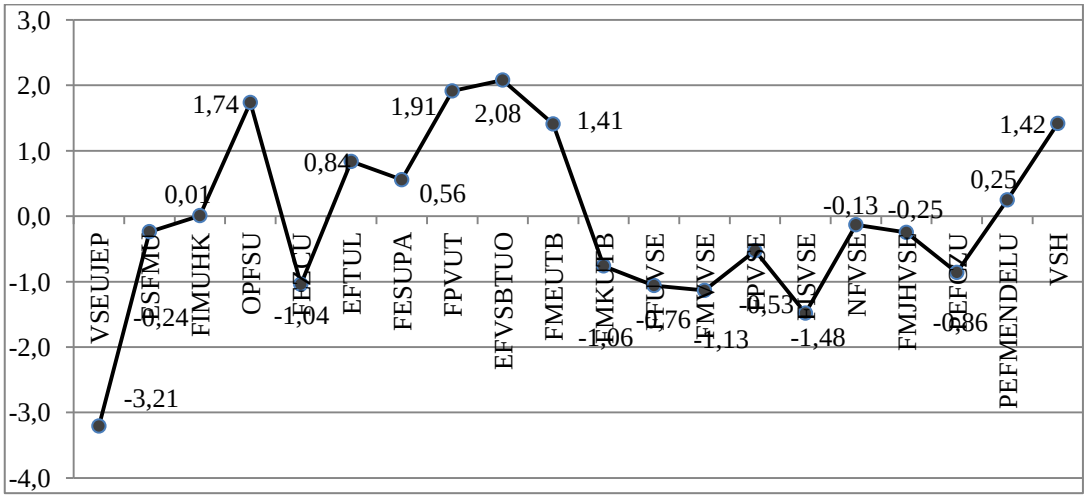


Figure 3 Conditional convergence *URF* for Master's in terms of the direction of approaching.

4 Conclusions

This paper investigates the behavior of the rate of unemployment of the graduates from the Faculties of Economics in the Czech Republic in 2002 - 2008. The main results obtained are summarized below:

- The research studies are mostly devoted to the convergence of output or prices for individual economies or panel models, but this article deals with the convergence of the unemployment rate of university graduates.
- A time series analysis showed that the unemployment rate for the entire reporting period is higher for the graduates from the Faculties of Economics with Bachelor's Degree programs, in comparison with Master's Degree programs. In the period 2002 - 2008 there was a decrease in the unemployment rate in both groups of the observed graduates. In connection with the effects of the economic crisis in the years 2009 - 2011 there had been an increase of the level of the unemployment of the graduates from the Faculties of Economics in both types of studies.

- The testing and estimating the panel unit root for the group of 26 Bachelor's programs Faculties of Economics has proved the existence of conditional convergence for the majority of 18 faculties. The other faculties converge absolutely. The development in the unemployment rate compared to the national level did not include the statistically significant negative deterministic trend.
- The results of the research of the convergence of the unemployment rate of the Master's Degree studies of 20 Faculties of Economics show that in the course of time there has been a mutual conditional convergence to the different steady states for all faculties and the differences are trend stationary with the inclusion of the deterministic linear trend.

Acknowledgements

This research was supported by the Student Grant Competition of Faculty of Economics, VSB-Technical University of Ostrava within project (SP2013/148) and also by the European Social Fund within the projects CZ.1.07/2.3.00/20.0296 and CZ.1.07/4.1.00/22.0001.

References

[1] Balcar, J.: The „Soft Five" in Romania. *The Romanian Economic Journal* **15** (2012), 21-41.
[2] Balcar, J., Homolová, E. and Karásek, Z. et al.: *Transferable competences across economic sectors: Role and importance for Employment at European level.* Publications Office of the European Union, Luxembourg, 2011.
[3] Ben-David, D.: Trade and convergence among countries. *Journal of International Economics* **40** 3–4 (1996), 279–298.
[4] Bernard, A. B. and Durlauf, S. N.: Convergence in international output. *Journal of Applied Econometric* **10** (1995), 97–108.
[5] Bernard, A. B. and Durlauf, S. N.: Interpreting tests of the convergence hypothesis. *Journal of Econometrics* **71** (1996),161–174.
[6] EU: Council conclusions of 11 May 2012 on the employability of graduates from education and training. *Official Journal of European Union* C 169/11.
[7] Durlauf, S. N. and Quah, D.: The new empirics of economic growth. In: *Handbook of Macroeconomics* (Taylor, J. B. and Woodford, M., eds.) Elsevier, 1999, 235–308.
[8] Evans, P. and Karras, G.: Do economies converge? Evidence from a panel of U.S. states. *The Review of Economics and Statistics* **78** (1996), 384–388.
[9] Fleissig, A. R. and Strauss, J.: Panel unit-root tests of OECD stochastic convergence. *Review of International Economics* **9** (2001), 153–162.
[10] Hanclova, J., Lukacik, M. and Szomolanyi, K.: A VEC Model of the Czech and Slovak Economies. In: P*roceedings of 28th International Conference on Mathematical Methods in Economics 2010.* University of South Bohemia, České Budějovice, 208-213.
[11] Hanclova, J.: Comparative properties of structural macroeconomic modelling. In: *Proceedings of International Conference Quantitative Methods in Economics (Multiple Criteria Decision Making XIV*). Iura Edition, Bratislava, 2008, 70-75.
[12] Kocenda, E. and Papell, D. H.: Inflation convergence within the European Union: a panel data analysis. *International Journal of Finance and Economics* **2** (1997), 189–198.
[13] Koucký, J. and Zelenka, M.: *Postavení vysokoškoláků a uplatnění absolventů vysokých škol na pracovním trhu 2010*. The Education Policy Centre, Prague, 2011.
[14] EU: *Labour Market Developments in Europe 2012. European Economy 5/12.* European Commission, 2012.
[15] Lee, H.Y. and Wu, J.L.: Mean reversion of inflation rates: evidence from 13 OECD countries.*Journal of Macroeconomics* **23** (2001), 477–487.
[16] Lopez, C. and Papell, D, H.: Convergence of EURO area inflation rates. *Journal of International Money and Finance* **31 (**2012), 1440-1458.
[17] Pesaran, M. H.: A pair-wise approach to testing for output and growth convergence. *Journal of Econometrics* **138** (2007), 312–355.
[18] The Education Policy Centre: *Unemployed University Graduates.* http://www.strediskovzdelavacipolitiky.info/app/navs2010/
[19] Žďárek, V. Price Convergence of the Czech Republic and the New EU Member States Toward the EURO Area. *Working paper 04/2009.* Research Centre for Competitiveness of Czech Economy, Prague, 2009.

Microeconomic view on the inventory optimization problem

Simona Hašková[1], Pavel Kolář[2]

Abstract. The paper deals with the optimization of inventory stock. The fundamental questions of logistics are: in what quantity and when to resupply the stock. In order to answer these questions the efficient methods are derived that offer a relatively accurate and satisfactory response (including its economic interpretation) even in terms of uncertainty. The discrete case, when the unit of supply is indivisible, is studied in detail and the results are then extrapolated by the limiting transition to the continuous case, when the unit of inventory is further arbitrarily divisible. In both cases the method relies on the value of the authors own established indicator of the storage efficiency and on the fragment of the probability distribution of the demand during the delivery time, which depends on it.

Keywords: case based reasoning, logistics, daily demand, delivery time, storage costs, ordering, transferring and handling costs, reorder level, safety stock, efficiency storage indicator.
JEL Classification: C51

AMS Classification: 91B32, 91B82

1 Introduction

In practice most sellers or manufacturers, who can not manage without storing the retailed goods or components required in manufacturing, are looking for the right answer to the question "in what quantity and when to resupply their stock?". The criteria of "the correct" answer here is the minimization of the total cost associated with it. The parameters, on which the correct answer depends, are:

- JP – The variable daily demand with the jp values showing the number of removed units of the considered inventory items from storage in one working day of the current year.
- D – The variable annual consumption of the considered inventory items with values $d = \Sigma jp_i$ (summing up all the working days in the current year).
- H – The variable part of the annual costs associated with the storage of the unit of the inventory item.
- C - Costs associated with one supply of the inventory item (ordering, transferring and handling costs of one delivery that within given specific logistic terms do not depend on the number of X units of supplement delivery.
- DD – The variable delivery time of the ordered additional supply (its values dd indicate the number of working days that pass from placing the order of the additional supply to its acceptance into stock).
- Z – The loss from each considered inventory item that would not satisfy the current demand due to an empty storehouse (the loss from customer dissatisfaction in the case a retailer or the failure of production in the case of a manufacturer, e.g. the lost margin).

If we assume from the above unlimited time duration of business, it is customary to consider the possible variants within one accounting period, which is usually a calendar year. It is also suitable to split the initial question into two questions – "How much to reorder?" and "When to reorder?". If X is the answer to the question "How much to reorder?", then the values of the variable D / X are the annual numbers of orders of additional supply. Let us denote by HD (reorder level) the level of the considered inventory item ($HD \geq 0$), at which the storehouse gives the order for an additional supply. Then we can answer the question "When to reorder?" by pointing to the selected HD (reorder level).

In the following we proceed from the standard textbook model (see e.g. [1], [2], [5], [7]) and we precise the solution in terms of certainty. We give reasons for the adaptation and generalization of this model on conditions of uncertainty and then by the "case based reasoning" approach we derive a general procedure for solving the task.

[1] VŠE Praha/Fakulta managementu JH, katedra managementu, Jarošovská 1117/II, 377 01 Jindřichův Hradec, haskovas@post.cz.
[2] VŠE Praha/Fakulta managementu JH, katedra managementu, Jarošovská 1117/II, 377 01 Jindřichův Hradec, kolar@fm.vse.cz.

2 The model and the solution of the problem of even removal from storage in terms of certainty

Ideally, in the case of even storage removal where jp and dd are the constants (JP = jp and DD = dd), the dynamics of stocks (i.e. the course of the state of stock, depending on the time measured by the number of working days) can be approximated by the continuous model, which is the "saw" in Figure 1 (see e.g. [4], [10]). This is exactly valid, for example, for the state of stock of even flowing fluid refilled to the initial level of X at the moment of emptying the tank.

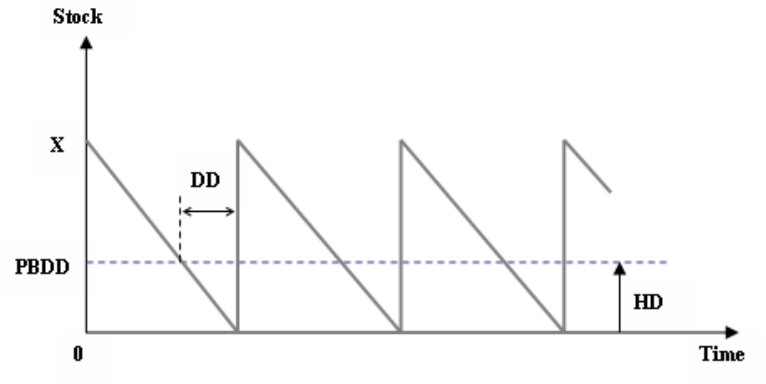


Figure 1 Dynamics of stock in the ideal case

The variable PBDD (the demand from the placing of the order to the time of delivery) is given by the product of PBDD = DD · JP, and therefore in this case has a constant value PBDD = pbdd = dd · jp. For HD_{opt} (i.e. HD minimizing the total costs associated with the storage of the inventory item and thus correctly answering the question "When to reorder?") it applies that HD_{opt} = PBDD independently of X ≥ PBDD.

It is obvious that the deviation of the HD value from the value PBDD on this or that side results in the increase of loss (costs) associated with the storage of the inventory item. Therefore HD_{opt} = PBDD. If HD = PBDD then the resulting dependence of the value of the total annual costs N(X) associated with the storage of the inventory item on the amount X of the additional supply can be described by:

$$N(X) = H \cdot X / 2 + C \cdot D / X \tag{1}$$

where H · X / 2 is the annual costs of storing and C · D / X is the annual amount of ordering, transferring and handling costs associated with replenishment of stock. Let us denote EOQ (economic order of quantity) as the volume of supplemental supply (i.e. the value of variable X), minimizing the expression (1) (for more details see [6]). Then EOQ can be obtained by solving the system $dN(X)/dX = H / 2 – C \cdot D / X^2 = 0$, $d^2N(X)/dX^2 = 2 \cdot C \cdot D / X^3 > 0$ due to the variable X:

$$EOQ = (2 \cdot C \cdot D / H)^{1/2} \tag{2}$$

By substituting EOQ for X in (1) and by the subsequent substitution of the expression $(2 \cdot C \cdot D / H)^{1/2}$ for EOQ to the right side of the equation we successively get:

$$N(EOQ) = H \cdot EOQ / 2 + C \cdot D / EOQ = (2 \cdot C \cdot D \cdot H)^{1/2} \tag{3}$$

3 The adaptation of the model to the conditions of uncertainty

In the examined ideal case (Figure 1) when HD = PBDD the last unit of the inventory item is contracted out at the time of arrival of a new supply (for more details see [9]). If HD > PBDD then the decrease of its stock would regularly stop on the level PZ = HD – PBDD. For the value PZ the teeth of the "saw" shown in Figure 1(starting from the second tooth) would move upward. The level of PZ is a safety stock, which in a case of need (an unexpected increase in the value of DD or an unexpected occurrence of jp_i > jp within the time of delivery) acts as a buffer contained in HD. Its aim is to eliminate partially or completely the threat of customer dissatisfaction due to the unexpected presence of PBDD above the usual value. In the described ideal case where this is excluded, this buffer is not needed. It would only increase the storage cost by H · PZ. Therefore, in Figure 1 PZ = 0 is selected.

The practice sometimes substantially deviates from the examined ideal. Both the daily demand (JP) and the length of delivery time (DD) may vary over time and their fluctuations make their product variable, thus the value of the variable PBDD. If values of JP and DD can, in this context, be regarded as independent random variables with known probability distributions of their values, then PBDD will be the random variable with probability distribution inferred from the distribution of variables JP and DD. Under certain assumptions, we can consider every "saw teeth" from the perspective of a long time horizon as the results of repeating of the same random process consisting of elementary random experiments that take place within individual working days. If a number of repetitions is large enough so that the law of large numbers could apply (see [3]), we can represent the variable JP, DD and PBDD as the unbiased predictions of their values, i.e. as the expected values E[JP], E[DD] a E[PBDD]. Completely analogous to Figure 1 is then Figure 2:

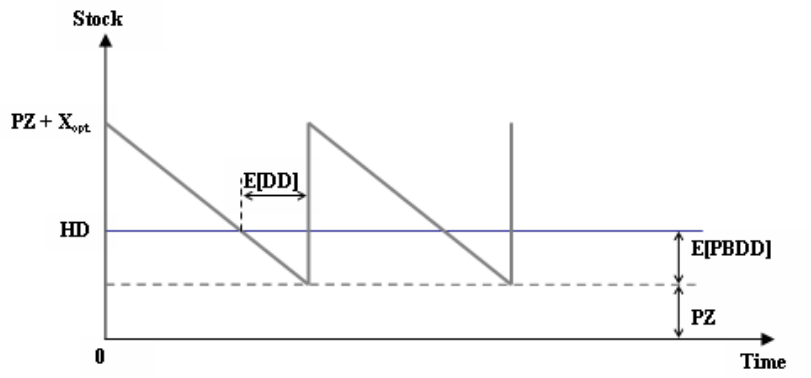


Figure 2 Model of dynamics of stock in terms of uncertainty

In it (unlike in Figure 1) the safety stock level PZ is an essential part of HD used to eliminate the unavoidable overhangs of the variable values PBDD over E[PBDD]. Therefore, it makes sense to consider PZ only within the limits $0 \leq PZ \leq (pbdd_{max} - E[PBDD])$, where for HD applies $E[PBDD] \leq HD \leq pbdd_{max}$. This model can, in a certain sense, be regarded as a statistical average of all possible random realizations of a long-term process of "development inventory stock" with the additional supply in the amount of X_{opt} units and with HD = E[PBDD] + PZ. For determination X_{opt} (the correct answer to the question "How much to reorder?") of this model it applies all that has already been mentioned in connection with the model in Figure 1. Much more difficult is finding HD_{opt}, i.e., the correct answer to the question "When to reorder?" The following part of this paper deals with the solution to this problem.

4 The general solution of the problem by the "case based reasoning" approach

Both considered models here (i.e. the model in terms of certainty - see Figure 1 and the model in terms of uncertainty - see Figure 2) answer the question "How much to reorder?" in the same way. In both cases we obtain X_{opt} by the procedure described in Article 2. The certain annual consumption of D will be replaced by the expected value $E[D] = E[\Sigma jp_i]$ in terms of uncertainty. The same thing can be said of the total annual cost (1), induced by the "saw" component of the examined fluctuations of the stock. The only difference is that while in the first case we deal with the real costs, in the second case we deal with the expected cost (see [3] and [8]).

The source of the substantial difference between the two cases is the existence of the expected annual cost E[NPZ(HD)] of the safety stock that depends on the chosen HD level under conditions of uncertainty. By analogy with equation (1), valid in terms of certainty, at $X = X_{opt}$ for the expected total cost E[N(HD)] we can write:

$$E[N(HD)] = H \cdot X_{opt} / 2 + C \cdot E[D] / X_{opt} + E[NPZ(HD)] \tag{4}$$

In the general case, E[NPZ(HD)] has two components: the expected annual storage costs of the safety stock level in the amount $H \cdot PZ = H \cdot (HD - E[PBDD])$ and the expected annual loss E[ZNZ(HD)] from a lack of stock. Therefore

$$E[NPZ(HD)] = H \cdot PZ + E[ZNZ(HD)] \tag{5}$$

To minimize E[N(HD)] according to HD means to minimize E[NPZ(HD)] according to HD. In the following we derive an algorithm to minimize E[NPZ(HD)] according to HD in the discrete case and then we generalize the idea of a solution to the continuous case.

4.1 Minimizing E [NPZ (HD)] in the discrete case

By "the discrete case" we mean the case when the unit of inventory is no more arbitrarily divisible (e.g. the washing machine in the warehouse store of washing machines, etc.). If assuming $HD_{max} = pbdd_{max}$, then $E[NPZ(HD_{max})] = H \cdot PZ_{max} = H \cdot (pbdd_{max} - E[PBDD])$. Because there is no shortage of stock, $E[ZNZ(pbdd_{max})] = 0$. A gradual reduction of HD from HD_{max} to HD – 1 gradually decreases PZ to PZ – 1 (see Figure 2). This leads to a constant marginal saving of the annual storage costs of safety stock by $H \cdot PZ - H \cdot (PZ - 1) = H$. At the same time it comes to a gradual increase of the marginal expected annual losses from the lack of stock by

$$\Delta E[ZNZ(HD)] = E[ZNZ(HD-1)] - E[ZNZ(HD)] \tag{6}$$

The idea of the algorithm of minimization E[NPZ(HD)] can be described as follows: As the first current value HD we choose the value equal to $pbdd_{max}$ and then for each current HD we count according to (6) $\Delta E[ZNZ(HD)]$. If $\Delta E[ZNZ(HD)] \geq H$, then by the transition to HD – 1 we are no better off. Thus $HD_{opt} = HD$ and the process ends. In the opposite case when $\Delta E[ZNZ(HD)] < H$ the new current HD becomes HD – 1 and the calculation according to (6) is repeated for the new current HD. The flowchart of the algorithm of this idea, in which the designation ΔE(HD) is an abbreviation for $\Delta E[ZNZ(HD)]$, is shown in Figure 3:

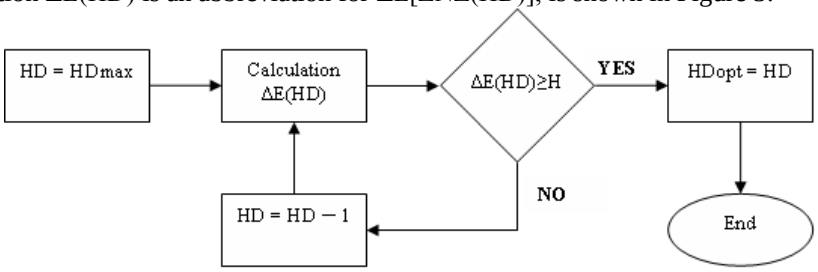


Figure 3 The flowchart of the optimization algorithm idea

The algorithm elaborates from HD_{max} to HD_{opt} from above and HD_{opt} is the one HD, to which the algorithm arrived by its recursive steps, and that as a result of (6) satisfies the condition

$$\Delta E[ZNZ(HD)] \geq H > \Delta E[ZNZ(HD+1)] \tag{7}$$

The calculation of E[ZNZ(HD)] for various HD is based on the probability distribution of the discrete random variable PBDD and on other terms, which are therefore necessary to define.

- Given a set $\Omega = \{0, 1, \ldots, pbdd_{max}\}$ of integer values of the variable PBDD and the projection $p: \Omega \rightarrow \langle 0, 1\rangle$ that for the sum of p(i) over all $i \in \Omega$ holds $\Sigma\, p(i) = 1$. The projection p defining the set of pairs $\{(i, p(i)): i \in \Omega\}$, where i is the current value of pbdd and p(i) its probability occurrence, is called probability distribution of a discrete random variable PBDD.
- The excess of demand over the HD is a discrete random variable PP(HD) with pp values dependent on the values of pbdd, for which it follows: pp = 0 for pbdd = i ≤ HD and pp = i – HD for pbdd = i > HD. Hence for the value pp = 0 we get (summing in the range from i = 0 to i = HD) its probability occurrence of Σ p(i); for pp = i – HD we get the probability occurrence of p(i). Furthermore, it is obvious that the expected value of the demand excess $E[PP(HD)] = 0 \cdot \Sigma\, p(i) + \Sigma\, (i - HD) \cdot p(i) = \Sigma\, (i - HD) \cdot p(i)$, where in the result it is summed over all i > HD.

$E[ZNZ(HD)] = E[PP(HD)] \cdot Z \cdot E[D] / X_{opt}$, from which after substitution to (6) and after adjustment we get $\Delta E[ZNZ(HD)] = (E[PP(HD-1)] - E[PP(HD)]) \cdot Z \cdot E[D] / X_{opt} = (\Sigma\, p(i)) \cdot Z \cdot E[D] / X_{opt}$, where in the result it is summed over all i ≥ HD. Analogously $\Delta E[ZNZ(HD+1)] = (\Sigma\, p(i)) \cdot Z \cdot E[D] / X_{opt}$, where in the result it is summed over all i ≥ HD + 1. Substituting both results in (7) we obtain after adjustment

$$p(HD) + \sum p(i) \geq H \cdot X_{opt} / (Z \cdot E[D]) > \sum p(i) \tag{8}$$

where in both summations it is summed over all i > HD. Sought HD_{opt} is the HD, which satisfies the condition (8). This condition can be made more transparent by designation of the constant $H \cdot X_{opt} / (Z \cdot E[D])$ by the symbol Φ, by which it takes the shape of

$$p(HD) + \sum p(i) \geq \Phi > \sum p(i) \tag{9}$$

The constant $\Phi = H \cdot X_{opt} / (Z \cdot E[D])$, sandwiched in (9) between the inequalities is the inverse value of the efficiency storage indicator.

$$\Psi = Z \cdot E[D] / (H \cdot X_{opt}) = (Z / H) \cdot (E[D] / X_{opt}) \tag{10}$$

This indicator is the product of the ratio of the unit loss from the lost margin to the annual cost of the unit storage (Z / H) and the annual number of orders ($E[D] / X_{opt}$). It measures the loss of the dissatisfaction of the annual consumption of the inventory item with the annual costs of the storing of the optimum additional supply. The condition (9) in combination with Figure 2 then says that the higher this indicator is (and therefore Φ lower), the higher the safety stock level is worth holding. In the case of the limited storage capacity it can be decided on the basis of the analysis of this indicator, which inventory items are worth storing and which are not.

4.2 Minimization of E[NPZ(HD)] in the continuous case

By the continuous case we mean the case when the unit of stock is arbitrarily divisible (as it is in the case of coal in the boiler room, the fluid reserve in the reservoir, etc.). The transition from discrete to continuous case can be reached by the "from below" unbounded physical division of the original unit. In terms of mathematical analysis it corresponds to the possibility of the limit transition from differential $\Delta x = 1$ in our discrete model to an arbitrarily small value of the differential $dx \to 0$ in the continuous model. This is related to the transition from a finite number of elements in the set of values $\Omega = \{0, 1, \dots, pbdd_{max}\}$ of the discrete random variable PBDD to the interval $\langle 0, pbdd_{max})$ of countless values of pbdd of the continuous variable PBDD.

The initial distribution $p: \Omega \to \langle 0, 1\rangle$ of the discrete PBDD in the form $\{(i, p(i)): i \in \Omega\}$, where p(i) is the probability occurrence of pbdd = i, passes in the case of the continuous PBDD in the continuous function $f: \langle 0, pbdd_{max}) \to \langle 0, 1\rangle$, in which for each $x \in \langle 0, pbdd_{max})$ and for a sufficiently short interval $dx \subset \langle 0, pbdd_{max})$ the product of $f(x) \cdot dx$ gives the probability that $x \le pbdd \le x + dx$. The function *f* is the density of probability of a continuous random variable PBDD. By assigning each value $HD \in \langle 0, pbdd_{max}\rangle$ to its corresponding value *F(HD)* obtained by integrating the density of probability *f(x)* in the range of 0 to HD, i.e. by the rule $F(HD) = \int f(x) \cdot dx$, we define the distribution function $F: \langle 0, pbdd_{max}\rangle \to \langle 0, 1\rangle$ of the continuous random variable PBDD with the extreme values of $F(0) = 0$ and $F(pbdd_{max}) = 1$. In the space between these extreme values *F(HD)* value gives the probability that pbdd < HD.

The discrete version of the continuous distribution function is the expression $F(HD) = \Sigma\, p(i)$, where is summed over all i < HD. Taking into account the fact that, if summing over the entire field Ω, Σ p(i) = 1 applies, then by substituting *F(HD)* to the equation (9) we obtain

$$1 - F(HD) \ge \Phi > 1 - F(HD) - p(HD) \tag{11}$$

The condition (11) can be achieved only when

$$\Phi = 1 - F(HD) \text{ or (which says the same thing), } F(HD) = 1 - \Phi \tag{12}$$

It is obvious that (12) holds for both the discrete and continuous case. What it means in the continuous case specifically is apparent from Figure 4:

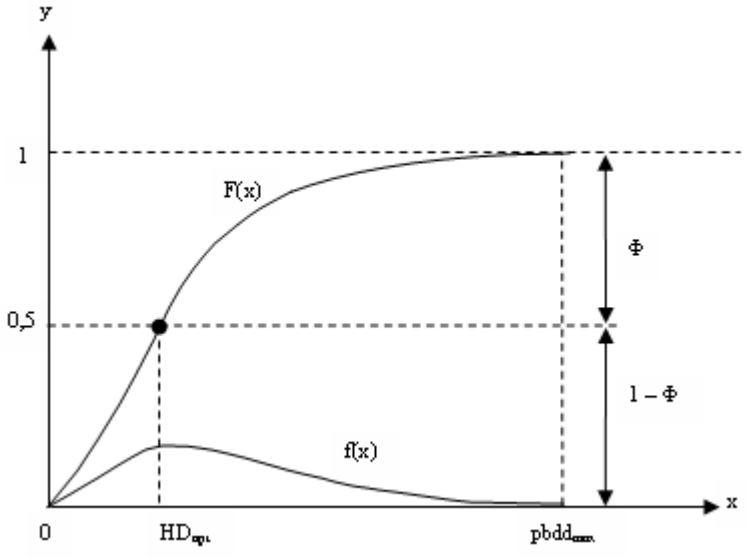


Figure 4 Identification HD_{opt} from the graph of the distribution function of a continuous random variable PBDD

We see that the sought HD_{opt} is placed under the intersection of the distribution function *F(x)* of a random variable PBDD with the level of $y = 1 - \Phi$.

5 Summary and conclusion

The paper focuses in detail on solving one of the fundamental problems of logistics, which is the question of "in what quantity?" and "when?" to resupply stock. While the literature gives a relatively accurate and satisfactory response both in terms of certainty and uncertainty to a question "how much?", the question of "when?" is often answered more or less vaguely in terms of uncertainty. The first part of the paper clarifies the standard definition of a model of the dynamics of the stock state with the steady storage removal in terms of certainty (Figure 1), draws conclusions from this, which further extends and generalizes on the conditions in terms of uncertainty.

The main contribution to the paper can be considered the derivation of a simple and effective methodology for solving this kind of task, including its economic interpretation. It turned out that for the answer to the more problematic question "when?" it suffices to know the probability distribution of the demand during the delivery time and the readily ascertainable value of, in this paper, the established indicator of the storage efficiency (10). The higher this indicator is, the higher the safety stock level is worth holding. In the case of the limited storage capacity it can be then decided on the basis of the analysis of the efficiency indicator, which inventory items are worth storing and which not.

Acknowledgements

The costs associated with the participation and the presentation of this paper on MME 2013 conference was supported by the institutionally support of the long-term strategic development FMVŠE; the contract number is IP632182.

References

[1] Brealey, R. A., Myers, S. C. & Marcus, A. J.: *Fundamentals of Corporate Finance* (7th Rev.Ed). United States of America: Mcgraw-Hill Education, 2011, ISBN 978-0-07-803464-0.
[2] Horáková, H., Kubát, J.: *Řízení zásob*. Profess Consulting, 1999, ISBN 80-85235-55-2.
[3] Jiroušek, R., Scozzafava, R.: *Basic Probability*. Nakladatelství Oeconomica, VŠE Praha 2003, ISBN 80-245-0471-5.
[4] Kodera, J., Marek, P.: Dynamic Inventory Model. *Industrial Engineering & Management Systems*, (2006), 43–47.
[5] Kolář, P.: *Manažerské finance*. Nakl. Bilance, Praha 1997.
[6] Maddah, B., Jaber, M.Y.:Economic order quantity for items with imperfect quality: Revisited. *International Journal of Production Economics.*Vol. 112, Issue 2 (2008), 808–815.
[7] Maness, T. S., Zietlow, J. T.: *Short-Term Financial Management*. Belmont, CA, U.S.A., South Western College, 2004, 736 s., ISBN 0-3242-0293-8.
[8] Shapiro,J.F., Wagner, S.N.**:** Strategic Inventory Optimization. *Journal of Business Logistics*, Vol. 30, No. 2 (2009).
[9] Schniederjans, M. J.: *Just-in-time Management*. Needham Heights, MA, U.S.A., Allyn and Bacon, 1993, ISBN 0-205-13600-1.
[10] Williams, B.D,Tokar, T.: A review of inventory management research in major logistics journals: Themes and future directions. *International Journal of Logistics Management*, Vol. 19, Issue 2 (2008), 212. DOI: 10.1108/09574090810895960.

Comparison the Gender Wage Differences in the Selected Czech Hospitals

Veronika Hedija[1]

Abstract. The article is devoted the gender wage differences in the Czech public sector. It aims to determine and compare the amount of the gender pay gap in selected two Czech hospitals and estimate the gender pay gap which cannot be explained by different characteristics of men and women and which may be the result of wage discrimination against women.

Both hospitals are located in the same town. Data comes from the year 2010 and is administrative. Data set contains data of more than 8500 employees. We use the average treatment effect on the treated estimation to extract the part of gender pay gap, which cannot be explained by different characteristics of men and women. To reduce the potential selection bias, we supplemented this by matching as preprocessing using exact matching and coarsened exact matching.

We concluded that the gender pay gap varied in our two compared hospitals. In hospital 1, female employees took about 77.1 percent of men wage and in the hospital 2 about 88 percent. Nevertheless, unexplained gender pay gap was relative low in both hospitals and it ranged from 5 to 6 percent in favor of men. Most of the wage disparity between male and female employees could be explained by the difference in existing personal and firm characteristics of men and women working in examined hospitals.

Keywords: average treatment effect on the treated, matching, gender pay gap, wage differences, labour market

JEL Classification: J16, J24
AMS Classification: 91G70

1 Introduction

According to the Eurostat data, the Czech Republic belongs to the countries with the highest gender pay gap in the European Union. In 2010, the gender pay gap gained 25.5 percent, what was about 9 percentage points more than the gender pay gap of the whole European Union. (Eurostat [2]) In this context, there is a series of questions. Are this high wage differences between men and women due to wage discrimination and what extent? Could the gender wage differences be explained by different characteristics of men and women?

The issue of gender wage differentials in the Czech Republic and the gender wage gap decomposition is devoted a number of papers. Most of these attempts to purify the overall gender wage gap of the part that can be explained by different personal or firm characteristics of men and women and to extract the part which remains unexplained. And it is this part that can be described as potentially discriminatory.

The vast majority of papers focus on the wage differences in the private sector (for example Hedija and Musil [4], Jurajda and Paligorova [7], Mysíková [10]), but there are also studies where attention is paid to the public sector, or comparing gender pay gaps in both sectors (see Eriksson, Gottwald and Mrazek [1] or Jurajda [8]). The vast majority of these studies used data from Information System on Average Earnings (ISPV) or European Union Statistics on Income and Living Conditions (EU-SILC).

Hedija and Musil [3] and [4] presented the analysis of gender pay gap within selected firms, used data from selected Czech private and public sector companies. The conclusions of these studies cannot be generalized to the whole Czech Republic however these provide interesting information. The analysis based on the data of a particular company provides a number of advantages. It allows eliminate any bias arising from classification workers into general categories. Using data from one company enables better comparison, especially, of actual workload and performance of individual staff. This article builds on these studies and strives for the maximum micro view of gender wage differences. Unlike the above studies to calculate the unexplained gender pay gap, the average treatment effect on the treated (ATT) estimation will be used supplemented by matching procedure as preprocessing.

[1] Vysoká škola polytechnická Jihlava, katedra ekonomických studií, Tolstého 16, 586 01 Jihlava, veronika.hedija@vspj.cz

The aims of the article is to determine the amount of the gender pay gap in selected two Czech hospitals and estimate and compare the gender pay gap which cannot be explained by different characteristics of men and women and which may be the result of wage discrimination against women.

The first section of the paper maps existing knowledge in the area of gender pay gap in the Czech Republic. Attention is given to the conclusions of selected studies analyzing the gender wage differences in the Czech public sector. The second section describes the methodology and data set. Unexplained portion of the pay gap is estimated by calculating the average treatment effect on the treated. In order to improve the quality of the estimation, matching procedure is used. To minimize bias arising from estimation and fundamental uncertainty, the simulation is used when estimating ATT. The final section of the article summarizes the obtained results and compares the unexplained gender pay gap in examined hospitals.

2 Gender Wage Differences: results of the selected studies

Wage differences between men and women exist not only in private but also in public sector. Empirical studies show that part of the existing gender wage gap in the Czech public sector can be explained by different characteristics of men and women. The part of wage differences remains unexplained and should be attributed to wage discrimination against women (Jurajda [8], Eriksson, Gottwald and Mrázek [1], Stupnytskyy [11]). These studies confirm that the private sector reported a higher unexplained gender pay gap compared with public sector. This finding is not surprising and can be attributed to the fact that the wage formation in public sector wages is subject to some degree of wage regulation.

Jurajda [8] used data from the Information System on Average Earnings and examined wage differences in the Czech Republic and Slovakia in 1998. He investigated the gender pay gap separately in the private and public sector using Oaxaca-Blinder Decomposition. He concluded that main source of wage differences in both sectors are various forms of employment segregation which explained over one third of the overall gender pay gap. The unexplained gender pay gaps differed dramatically between the public and private sector. In the private sector it represented more than 60 percent of the gender pay gap in comparison to the public sector, where it created just under 40 percent of the gender pay gap.

Eriksson, Gottwald and Mrázek [1] examined the gender pay gap in the public and private sector using managerial pay. They used the data set combining information from three separate sources: Information System on Average Earnings, Survey of Managerial Staff Earnings and firm register for 1998. They estimated the managerial pay function where age, education, gender, industry, region, firm size and position within firm are the explanatory variables. Their conclusions about the unexplained gender pay differences in the private and public sector are different in both periods. The wage differences due to gender were about 10.8 percent in state-owned firms and 20.1 percent in private-owned firms.

The managerial wage differences in the Czech Republic are also focused the study of Stupnytskyy [11]. Author took data from Information System on Average Earnings using Oaxaca-Blinder decomposition estimated the unexplained part of managerial gender pay gap. Gender pay gap was 0.57 in private firms and 0.26 in the stated-owned firms in the 2005. The unexplained part of gender pay gap reached 37 percent of GPG in private and only 15 percent of GPG in budgetary sector. The main factors explaining wage differences were education and position. These factors elucidated about two third of gender pay gap in budgetary sector.

3 Data

The unexplained wage difference between men and women is estimated in two selected Czech hospitals, which were willing to provide the necessary data. Both hospitals are located in the same town. The data comes from the year 2010 and is administrative.

Employees with a long-term illness, on maternity or parental leave or working on a business agreement were excluded from the sample. The final data set covers the data of 8662 employees, of which 1636 are male and 7026 female. In hospital 1, women represent about 77 percent of employees and in hospital 2 almost 83 percent.

In the hospitals the following characteristics of employees were measured: gender, age, highest level of education, occupation, number of years that the employee works in the hospital, working time (full-time means 1), department where the employee works, hours worked per year, overtime hours, days of sick leave, days of leave and annual gross wage. In the article, the hourly gross wage is used as the explained variable. The hourly gross wage for every employee is calculated as a ratio of the annual gross wage (including bonuses) and the sum of worked hours (including overtime) and hours of annual leave (days of annual leave*8*working time).

Table 1 shows the average characteristics of male and female employees in both hospitals. In the hospital 1, women earn on average 77 percent of male wage and in the hospital 2, average female wage reach 88 percent of average male wage.

	Hospital 1		Hospital 2	
Characteristic	**Men**	**Women**	**Men**	**Women**
Age (year)	43.060	41.800	41.995	38.518
Education				
primary	0.079	0.083	0.042	0.038
lower secondary vocational	0.005	0.004	0.000	0.001
secondary vocational	0.187	0.073	0.240	0.094
secondary school diploma	0.151	0.571	0.191	0.555
higher professional	0.022	0.086	0.035	0.084
bachelor's degree	0.010	0.038	0.019	0.049
tertiary	0.546	0.145	0.472	0.179
Occupation				
worker	0.190	0.038	0.276	0.038
technical and economic worker	0.128	0.115	0.105	0.084
orderly	0.130	0.097	0.103	0.111
lower medical worker	0.012	0.036	0.013	0.029
paramedical worker	0.024	0.022	0.033	0.147
medical laboratory technician	0.026	0.067	0.007	0.054
pharmaceutical assistant	0.002	0.017	0.000	0.008
midwife	0.000	0.001	0.000	0.033
nurse	0.029	0.491	0.048	0.346
another professional	0.036	0.017	0.030	0.029
pharmacist	0.000	0.003	0.003	0.006
doctor	0.423	0.096	0.383	0.115
Working time	0.896	0.959	0.910	0.934
Overtime (hour per year)	242.092	66.759	140.434	55.257
Sick leave (hour per year)	28.634	68.809	6.510	25.190
Years in firm	8.357	9.330	10.057	10.624

Table 1 Average characteristics of the employees of hospitals

A very important factor when examining the potential wage discrimination against women is the wage setting-power of managers in firm. The greater the flexibility in wage formation, the larger the effect of the different approach towards the remuneration of male and female subordinates. In 2010, wages in the Czech public health sector were regulated by Government Regulation No. 564/2006 on the salaries of employees in public service and administration. It defined the wage classes and grades and assigns workers to a wage class and to a grade level and set the base gross wage for the individual classes and grades. But the maximum amount of permitted bonuses was not regulated ant the maximum amount was regulated neither at the level of hospitals. This means that the final gross wage of an employee including bonuses depended to a large extent on the decision of the managers, who were limited by the Government Regulation, which regulates only the minimum wage, and of cause by the size of the budget of the hospital. This implies that the management of hospital have a relative flexibility in wage formation.

4 Methodology

There are several methods we can use to estimate the amount of wage differences between men and women, which cannot be explained by different characteristics of these. We estimate the unexplained part of gender pay gap using estimation of average treatment effect on the treated (ATT). Treatment literature defines the average treatment effect this way: '*The average treatment effect on the treated is the mean effect for those who actually participated in the program*'. (Wooldridge [12], p. 605) In our case the ATT is the mean effect for women in the form of a lower wage due to them being women. And this way we use it for estimating the unexplained part of gender pay gap in the two selected Czech hospitals.

We use the following equation for the calculation of the average treatment effect on the treated

$$ATT = E(y_1 - y_0 | w = 1) \quad (1)$$

Where w is the binary treatment indicator, w = 1 denotes treatment and w = 0 otherwise, y_1 is the potential

outcome with treatment and y_0 is the potential outcome without treatment. In our case, be treated means be the women. We can rewrite the ATT as

$$ATT = E(y_1|w = 1) - E(y_0|w = 1) \quad (2)$$

Where ATT represents the gender pay gap which cannot be explained different characteristics of men and women. The term $E(y_1|w = 1)$ is the sample average of logarithm of gross hourly wage of women and the term $E(y_0|w = 1)$ is the sample average of logarithm of gross hourly wage of women, if they were men. It seems to be very easy to compute the ATT bud in reality only one of the ride-side terms is known. From our sample, we can compute the first term on the right-side of the equation (2). The second term, the average of logarithm of women gross hourly wage if they were men, we have to estimate some way. There is more way to estimate this.[2] We estimate it using the regression model and simulation.

First, we estimate the coefficient of wage function of men form the regression model using least squares method.

$$y_0 = \beta_0 . X_0 + u \quad (3)$$

Where, y_0 is the logarithm of gross hourly men´s wages, β_0 is the vector of coefficients of wage function, X_0 is the vector of the chosen characteristics of women and u is a disturbance term. The explanatory variables are used: age, age squared, education, working time, years in the firm, number of overtime hours, sick leave, occupation and department.

We use the estimated coefficients of male wage function to compute the average wage of women, if they were men. We used simulation with R program to reduce the estimation and fundamental uncertainty. The simulation of estimating the parameters of the wage function and stochastic component is made, where the first step involves estimating the parameters, the second step the simulation of the random component. We use thousand simulations and receive thousand expected values of where the final expected value is calculated as the average of thousand simulations. (King et al. [9]) Than, we compute the average wage of women, if they were men using this formula

$$E(y_0|X, w = 1) = E(\beta_0 . X_{1i}). \quad (4)$$

Where $E(\beta_0 . X_{1i})$ is the mean of the expected wage of every woman in the sample after the simulation.

Finally, we estimate the average treatment effect on the treated.

$$ATT = E(y_1|w = 1) - E(\beta_0 . X_{1i}). \quad (5)$$

In this case, the ATT expresses the unexplained part of gender pay gap. But the obtained results can be biased to some extent. If the characteristics of treated and control groups are too different the coefficient of female wage function will be probably far enough the coefficients of hypothetical male wage function. To reduce bias and model dependence, we will use matching as a preprocessing procedure as proposed Ho et al. [5]. The matching is based on the idea that the simplest and least controversial way to determine the discriminatory component of the wage gap is to compare wages between men and women who show exactly the same characteristics. The aim of matching procedure is creating as much as possible homogeneous sample of men and women. There are more matching methods: one-to-one matching, exact matching, propensity score matching, monotonic imbalance bounding. The most used matching methods is exact matching which exclude women whose characteristics do not match with any man into sample and vice versa. The disadvantage of this method is that their conditions are very restrictive and leads often to selection of too small sample. This is why we use here coarsened exact matching too. Coarsened exact matching is based on exact matching but trying to overcome its shortcomings. The main idea of this method is to coarsen variables into groups, and then exact match. At the end the original values of matched data are retained (Iacus et al. [6]). Both methods offer freely available program MatchIt (Ho et al. [5]).

Application of exact matching very often caused the big loss of data and the sample of employees after matching procedure is very small. This causes the obstacles using regression to estimate potential wage of women if they are men. In the case of exact matching, we use this formula for ATT calculation

$$ATT = (\sum_{i=1}^{N} w_i)^{-1}\{\sum_{i=1}^{N} w_i \, [E(y_1|x, w = 1) - E(y_0|x, w = 0)]\} \quad (6)$$

The term $E(y_1|x, w = 1)$ is the sample average of logarithm of gross hourly wage of women in the subsample and the term $E(y_0|x, w = 0)$ is the sample average of logarithm of gross hourly wage of men in the subsample. The subsamples contain the employees with similar characteristics excluded gender.

[2] More Wooldridge [12] and Ho et al. [6].

The ATT computing without and with matching as preprocessing let as known, whether there exists the wage differences which cannot be explained by different known characteristics of men and women.

5 Empirical Results

Firstly, we estimated the ATT in both hospitals without matching. We estimated the wage function of male employees in the hospitals using equation 3 and the least square method and then calculated the average treatment effect on the treated using equation 5.

The ATT differed between examined two hospitals. In both hospitals the ATT reached negative values. The female earnings were lower than male and this difference could not be explained by known different characteristics of men and women. In the hospital 1, ATT reached -0.08945 and in the hospital 2 only -0.0293. The unexplained wage differences between men and women were higher in the hospital 1. Nevertheless, if we assess the gender wage gap we have to take into account, that not all characteristics of male and female employee in the hospital are known and quantifiable. So the part of gender wage gap can be explained for example by the higher talent or work commitment of men. The conclusion can also be partially distorted by too differences in characteristics of men and women in the sample, where the wage function of men could be far enough of hypothetical wage function of women, if they were a men.

Then we used matching procedure to get more homogenous sample of men and women and estimated ATT for this modify sample of employees. Using exact matching, we selected the subsamples of women and men with exactly identical characteristics. We worked with these characteristics: age, education, occupation, number of years in this hospital, working time, department, overtime hours and sick leave. Applying exact matching led, as expected, to the great loss of data. The resulting sample of employees was very small in the case of both hospitals and results are more informative. In hospital 1 there were chosen only 5 pairs of employees reporting exactly same characteristics, in the hospital 2 only 2 pairs. Because of very small sample of data, the ATT was calculated using equation 6. The unexplained gender pay gap reached approximately 8.4 percent in favor of men in hospital 1and 1.6 percent in favor of women in hospital 2.

Finally, we use coarsened exact matching to select subsamples of men and women with the most similar characteristics. Also in this case the sample of employees decreased significantly. After matching procedure, the sample of employees was narrowed to 255 employees in hospital 1 and to 450 employees in hospital 2. Then the wage function of men in the sample was estimated using least square method and ATT calculated. Using coarsened exact matching led to decline the ATT in hospital 1 and increase the ATT in hospital 2. However, the ATT was very similar in both hospitals after application of matching procedure it was -0.0619 in hospital 1 and -0.0531 in hospital 2.

	Without matching	**Exact matching 1)**	**Coarsened exact matching**
Hospital 1	-0.08945*	-0.08401	-0.0619*
	(0.0336)	(0.3153)	(0.0304)
N	2592	10	255
Men	584	5	103
Women	2008	5	152
Hospital 2	-0.0293	0.0158*	-0.0531*
	(0.0208)	(0.0048)	(0.0202)
N	6070	4	450
Men	1052	2	178
Women	5018	2	272

*significant at the 5% level, standard deviation in parentheses, 1) ATT is calculated using equation 6

Table 2 Average treatment effect on the treated

6 Discussion and conclusion

The main aim of this paper was analyze the amount of the gender pay gap in selected Czech public sector firms, find out and compare the unexplained part of the gender pay gap. The gender pay gap was examined in two Czech hospitals using administrative data and more methods.

The gender pay gap was different in our two compared hospitals. In hospital 1, female employees took about 77.1 percent of men wage and in the hospital 2 about 88 percent. To extract the part of the gender pay gap that could not be explained be different characteristics of male and female employee, the average treatment effect on the treated was estimated.

When we used whole sample of employees, the unexplained part of gender pay gap differed in both hospitals. The ATT reached value approximately -0.09 in hospital 1 and -0.03 in hospital 2. It means that the unexplained gender pay gap was higher in hospital 1, which reported the higher overall gender pay gap. The unexplained wage gap was approximately 9 percent in favor of men. Results for hospital 1 were very similar to Jurajda [8] who examined the unexplained part of gender pay gap in the Czech public sector in 1998. According to his conclusions, the gender pay gap in the public sector was about 0.24 and over a third of the overall gender wage gap stayed unexplained and could be due to discrimination against women. It represented the wage gap about 0.08 in favor of men.

To get more homogenous sample of men and women we applied the matching procedure as preprocessing. After using coarsen exact matching the unexplained gender pay gap was very similar in both hospitals. The ATT reached approximately -0.06 in hospital 1 and -0.05 in hospital 2.

We can conclude that identified unexplained gender pay gaps in both hospitals depend on using estimation method. Using more homogenous sample of men and women led to a convergence of values found for both hospitals. In this case, the unexplained gender pay gap was very similar in both hospitals and relative small. Most of the wage disparity between male and female employees could be explained by the difference in existing personal and firm characteristics of men and women working in examined hospitals.

References

[1] Eriksson, T., Gottwald, J., Mrázek, P.: *Determinants of Managerial Pay in the Czech Republic*. WDI Working paper No. 310, Michigan, 2000. http://wdi.umich.edu/files/publications/workingpapers/wp310.pdf (1.10.2012)

[2] Eurostat. Statistical Database, 2013. http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&init=1&language=en&pcode=tsdsc340&plugin=1 (10.1.2013).

[3] Hedija, V., Musil, P.: Gender Wage Differences in the Selected Czech Public Sector Company. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis **LX** (2012), 81-88.

[4] Hedija, V., Musil, P.: Jsou ženy v České republice mzdově diskriminovány? Závěry z vybraného podniku. *Ekonomická revue* 15 (2012), 27-40.

[5] Ho, D. E., Imai, K., King, G., Stuart, E.A.: Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis* **15** (2007), 199-236.

[6] Iacus, S. M., King, G., Porro, G.: Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis* **20** (2012), 1-24.

[7] Jurajda, S., Paligorova, T.: Czech Female Managers and their Wages. *Labour Economics* **16** (2009), 342-351.

[8] Jurajda, S.: Gender Wage Gap and Segregation in Enterprises and the Public Sector in Late Transition Countries. *Journal of Comparative Economics* **31** (2003), 199-222.

[9] King, G., Tomz, M., Wittenberg, J.: Making the Most of Statistical Analyses: Improving Interpretation and Presentation. *American Journal of Political Science* **44** (2000), 341–355.

[10] Mysíková, M.: Gender Wage Gap in the Czech Republic and Central European Countries. *Prague Economic Papers* **3/2012** (2012), 328-346.

[11] Stupnytskyy, O.: *Struktura rozdílů ve výdělcích mužů a žen v managementu.* VÚPSV, Praha, 2007.

[12] Wooldridge, J.: *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, 2002.

Another view on time-varying correlations: The case of stocks and bonds

Radek Hendrych [1]

Abstract. The aim of the contribution is to introduce an innovative approach to conditional covariance and correlation modelling. This can be obviously useful in multivariate financial time series analysis, e.g. in the multivariate GARCH context. The proposed method consists of two steps. The first one is based on the LDL factorization of the conditional covariance matrix, state space modelling and associated Kalman recursions. Moreover, it is able to deliver a dynamic orthogonal transformation of given stochastic vector data. The second step of the suggested technique analyses conditional covariances of transformed time series which is indeed simplified due to its simultaneously uncorrelated components. In the paper, performance of the introduced procedure is tested in an empirical financial framework. Namely, the daily correlation links between logarithmic returns on stocks and bonds are investigated and compared with other estimated dynamic correlations gained by several common methods, e.g. the moving averages, the diagonal BEKK model or the dynamic conditional correlation (DCC) models.

Keywords: conditional correlation, conditional covariance, dynamic conditional correlation, GARCH, state space modelling.

JEL classification: C32
AMS classification: 91B84

1 Introduction

Analysis of time-varying correlations is undoubtedly an important part of multivariate time series modelling. In particular, it is worth of interest from both the theoretical and the practical point of view. Correlations are crucial inputs for many tasks of financial, portfolio and risk management, e.g. an asset allocation, a construction of an optimal portfolio or a hedging problem. This issue and related topics are discussed in many academically or practically oriented publications, see e.g. the comprehensive works [1] or [4] and the references given therein.

From a general perspective, the main task is to capture time-varying behaviour of conditional covariances with special regard to modelling of conditional correlations in the given model framework. Moreover, such a class of models is indeed worthy of interest from the mathematical point of view. In particular, the conditional covariance (correlation) matrix must be symmetric and positive definite. Additionally, the conditional correlation matrix must have unit diagonal elements. Indisputably, such requirements might bring really tough constraints into estimation, especially in the case of higher dimension. Hence, it is obviously more effective to consider various representations which naturally simplify or completely eliminate these restrictions. In general, one can distinguish between two approaches: (i) *direct* - it contains an explicit modelling expression for correlations, (ii) *indirect* - time-varying correlations are simply obtained by a normalization of conditional covariances.

Section 2 introduces a general model framework (a straightforward multivariate analogy of univariate conditional heteroscedastic models). Section 3 gives a brief overview of common estimators of dynamic correlation links. Section 4 presents an innovative approach to conditional covariance and correlation modelling in details. Section 5 considers an empirical financial application which examines bivariate correlations between logarithmic returns on stocks and bonds and compares various achieved results with each other. Finally, Section 6 contains conclusions.

[1] Charles University in Prague, Faculty of Mathematics and Physics, Dept. of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Prague 8, Czech Republic, hendrych@karlin.mff.cuni.cz

2 Model framework

Consider a multivariate stochastic vector process $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ of dimension $(n\times 1)$. Denote $\mathcal{F}_t$ the σ-algebra generated by observed time series up to and including time t, i.e. $\mathcal{F}_t = \sigma(\mathbf{X}_s, s \le t)$ is the smallest σ-algebra with respect to which $\mathbf{X}_s$ is measurable for all $s \le t$, $s,t \in \mathbb{Z}$.

In this framework, assume the following model

$$\mathbf{X}_t = \mathbf{H}_t^{1/2}\mathbf{Z}_t, \tag{1}$$

where $\mathbf{H}_t$ is the $(n\times n)$ positive definite conditional covariance matrix of $\mathbf{X}_t$ given $\mathcal{F}_{t-1}$. Furthermore, one supposes that $\{\mathbf{Z}_t\}$ is an $(n\times 1)$ i.i.d. stochastic vector process with the following first two moments: $\mathsf{E}(\mathbf{Z}_t) = \mathbf{0}$ and $\mathsf{cov}(\mathbf{Z}_t) = \mathbf{I}_n$, where $\mathbf{I}_n$ denotes the identity matrix of order n.

In the model (1), the conditional and the unconditional moments of $\mathbf{X}_t$ can be easily derived:

$$\mathsf{E}(\mathbf{X}_t|\mathcal{F}_{t-1}) = \mathbf{0}, \quad \mathsf{cov}(\mathbf{X}_t|\mathcal{F}_{t-1}) = \mathbf{H}_t^{1/2}(\mathbf{H}_t^{1/2})^\top = \mathbf{H}_t, \tag{2}$$

$$\mathsf{E}(\mathbf{X}_t) = \mathbf{0}, \quad \mathsf{cov}(\mathbf{X}_t) = \mathsf{E}(\mathbf{H}_t), \quad \mathsf{cov}(\mathbf{X}_t, \mathbf{X}_{t+h}) = \mathbf{0},\ h \neq 0. \tag{3}$$

Hence, from (2), it is evident that $\mathbf{H}_t^{1/2}$ is any $(n\times n)$ positive definite matrix such that $\mathbf{H}_t$ is the conditional covariance matrix of $\mathbf{X}_t$ given $\mathcal{F}_{t-1}$. From the theoretical point of view, $\mathbf{R}_t$ (the conditional correlation matrix of $\mathbf{X}_t$ given $\mathcal{F}_{t-1}$) can be obtained by the straightforward normalization of the conditional covariance matrix $\mathbf{H}_t$.

3 Models of conditional covariances and correlations

3.1 Basic approaches

One should start with simple multivariate moving averages (MA) in their general form, i.e.

$$\mathbf{H}_t = \frac{1}{M}\sum_{s=t-M}^{t-1}\mathbf{X}_s\mathbf{X}_s^\top, \quad M \ge 2. \tag{4}$$

Multivariate exponentially weighted moving averages (EWMA) are defined as

$$\mathbf{H}_t = (1-\lambda)\mathbf{X}_{t-1}\mathbf{X}_{t-1}^\top + \lambda\mathbf{H}_{t-1}, \quad \lambda \in (0,1). \tag{5}$$

Both models (4) and (5) are evidently simple and, therefore, they find broad practical applications. Each of them contains only one unknown parameter; e.g. $M = 100$ or $\lambda = 0.94$ are commonly set for daily data, see [4]. The given forms of (4) and (5) guarantee the positive semidefiniteness.

3.2 BEKK model

A BEKK model is a particular case of a more general VEC model, see e.g. [4]. The simplest BEKK(1,1) model can be generally represented as

$$\mathbf{H}_t = \mathbf{C}\mathbf{C}^\top + \mathbf{A}\mathbf{X}_{t-1}\mathbf{X}_{t-1}^\top\mathbf{A}^\top + \mathbf{B}\mathbf{H}_{t-1}\mathbf{B}^\top, \tag{6}$$

where $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are $(n\times n)$ parameter matrices. The positive semidefiniteness is guaranteed by the construction. On the other hand, such models include lots of unknown parameters, and thus their calibration (usually via (quasi) maximum likelihood) can be challenging. Therefore, a simpler form may be used, e.g. the so-called *diagonal* BEKK(1,1), i.e. the matrices $\mathbf{A}$ and $\mathbf{B}$ are restricted to be diagonal.

3.3 Constant and dynamic conditional correlations

A *constant conditional correlation* (CCC) model by Bollerslev [2] decomposes the matrix $\mathbf{H}_t$ as

$$\mathbf{H}_t = \mathbf{C}_t \mathbf{R} \mathbf{C}_t, \tag{7}$$

where $\mathbf{C}_t$ is a diagonal matrix formed by time-varying standard deviations $\sqrt{h_{ii,t}}$ and $\mathbf{R}$ is an $(n \times n)$ constant conditional correlation matrix. The diagonal elements of $\mathbf{C}_t$ can be modelled by common (univariate) techniques for conditional variances. Moreover, the matrix $\mathbf{R}$ is usually estimated by a sample correlation matrix of standardized errors $\boldsymbol{\gamma}_t = \mathbf{C}_t^{-1}\mathbf{X}_t$. However, the assumption of constant conditional correlations may seem unrealistic and too restrictive.

Engle [5] has extended the model (7) in a natural way to a more general case of *dynamic conditional correlations* (DCC) which are defined as

$$\begin{aligned} \mathbf{H}_t &= \mathbf{C}_t \mathbf{R}_t \mathbf{C}_t, & (8)\\ \mathbf{R}_t &= \mathsf{diag}\{\mathbf{Q}_t\}^{-1/2} \mathbf{Q}_t \mathsf{diag}\{\mathbf{Q}_t\}^{-1/2}, & (9)\\ \mathbf{Q}_t &= (1-\alpha-\beta)\mathbf{S} + \alpha \boldsymbol{\gamma}_{t-1}\boldsymbol{\gamma}_{t-1}^\top + \beta \mathbf{Q}_{t-1}, & (10) \end{aligned}$$

where $\mathbf{R}_t$ is a matrix of time-varying conditional correlations with unit diagonal elements, α and β are scalars and $\mathbf{S}$ is a parameter matrix and $\mathsf{diag}\{\mathbf{Q}_t\}$ is a diagonal matrix with $q_{11,t}, \ldots, q_{nn,t}$ on the main diagonal. If $\mathbf{Q}_t$ is positive definite, then so is $\mathbf{R}_t$. To ensure that $\mathbf{Q}_t$ is positive definite, it is sufficient to suppose that $\alpha \geq 0$, $\beta \geq 0$, $\alpha + \beta < 1$ and $\mathbf{S}$ is positive definite, see [5]. It is also usual to assume that $s_{ii} = 1$ in order to guarantee the unique specification of $(\alpha, \beta, \mathbf{S})$. The so-called mean-reverting DCC model (8)-(10) is an analogy of the scalar diagonal GARCH model.

There exist various alternatives to $\mathbf{Q}_t$ in (10), e.g. *integrated* DCC defined as

$$\mathbf{Q}_t = (1-\lambda)\boldsymbol{\gamma}_{t-1}\boldsymbol{\gamma}_{t-1}^\top + \lambda \mathbf{Q}_{t-1}, \quad \lambda \in (0,1). \tag{11}$$

This model is a direct analogy of the multivariate exponentially weighted moving averages, see above. The process for $\mathbf{Q}_t$ has the unit root and the covariances have no tendency to revert to a constant value. In general, the DCC models can be calibrated by two-step (quasi) maximum likelihood, i.e. the first step calibrates conditional variance terms and the second one fits conditional correlations, see [4] and [5].

3.4 Orthogonal GARCH model

An orthogonal GARCH model (OGARCH) is based on a time-invariant orthogonal transformation of multivariate time series $\{\mathbf{X}_t\}$, see [1]. Namely, assume that $\mathsf{cov}(\mathbf{X}_t) = \boldsymbol{\Sigma}$ has the spectral decomposition:

$$\boldsymbol{\Sigma} = \mathbf{P}^{-1}\boldsymbol{\Lambda}(\mathbf{P}^{-1})^\top, \tag{12}$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues of the unconditional covariance matrix $\boldsymbol{\Sigma}$ and $\mathbf{P}^{-1}$ is the orthonormal matrix of associated eigenvectors. Further, suppose that the conditional covariance matrix of the transformation $\mathbf{Y}_t = \mathbf{P}\mathbf{X}_t$ given $\mathcal{F}_{t-1}$, say $\mathbf{G}_t$, is diagonal and that each of its diagonal elements follows a univariate conditional variance process, e.g. a GARCH process. According to the declared assumptions, the conditional and the unconditional covariance matrix of $\mathbf{X}_t$ are

$$\mathsf{cov}(\mathbf{X}_t|\mathcal{F}_{t-1}) = \mathbf{P}^{-1}\mathbf{G}_t(\mathbf{P}^{-1})^\top, \quad \mathsf{cov}(\mathbf{X}_t) = \mathbf{P}^{-1}\mathsf{E}(\mathbf{G}_t)(\mathbf{P}^{-1})^\top. \tag{13}$$

From the practical point of view, the calibration follows a two-step procedure, see [4]. Firstly, one extracts the spectral decomposition of $\mathbf{S}$ (i.e. a sample counterpart of $\boldsymbol{\Sigma}$). Secondly, one estimates suitable univariate conditional covariance models for each element of the transformed stochastic vector process $\{\mathbf{Y}_t\}$, and thus delivers the conditional covariances (or correlations) of $\{\mathbf{X}_t\}$.

4 Conditional covariances and correlations via state space modelling

The idea of an orthogonal transformation, which has been introduced in the context of the OGARCH technique, can be further extended. In particular, one can assume a dynamic modification of this approach.

Following the algebraic theory, each real symmetric positive definite matrix has a unique LDL decomposition, see e.g. [7]. Namely, let the conditional covariance matrix $\mathbf{H}_t$ have the LDL reparametrization in the standard form, i.e.

$$\mathbf{H}_t = \mathbf{L}_t\mathbf{D}_t\mathbf{L}_t^\top \quad [= (\mathbf{L}_t\mathbf{D}_t^{1/2})(\mathbf{L}_t\mathbf{D}_t^{1/2})^\top = \mathbf{H}_t^{1/2}(\mathbf{H}_t^{1/2})^\top], \tag{14}$$

where $\mathbf{L}_t$ is a $(n \times n)$ lower triangular matrix with the unit diagonal and $\mathbf{D}_t$ is a $(n \times n)$ diagonal matrix with positive elements $d_{ii,t}$ on its diagonal. In particular, $\det(\mathbf{L}_t) = 1$, $\mathbf{L}_t$ is invertible, and the inverted matrix $\mathbf{L}_t^{-1}$ is also a $(n \times n)$ lower triangular matrix with unit diagonal elements. Point out that the decomposition (14) requires no parameter constraints for $\mathbf{H}_t$ being symmetric and positive definite since this is guaranteed by the structure.

The form of the matrix $\mathbf{L}_t$ provides a natural orthogonal transformation of $\mathbf{X}_t$:

$$\mathbf{Y}_t = \mathbf{L}_t^{-1}\mathbf{X}_t \quad [= \mathbf{L}_t^{-1}\mathbf{H}_t^{1/2}\mathbf{Z}_t = \mathbf{D}_t^{1/2}\mathbf{Z}_t]. \tag{15}$$

With respect to the declared assumptions and relations (2), (3) and (15), the transformation $\mathbf{Y}_t$ has the following conditional and unconditional moments:

$$\mathsf{E}(\mathbf{Y}_t|\mathcal{F}_{t-1}) = \mathbf{0}, \quad \mathsf{cov}(\mathbf{Y}_t|\mathcal{F}_{t-1}) = \mathbf{D}_t, \tag{16}$$

$$\mathsf{E}(\mathbf{Y}_t) = \mathbf{0}, \quad \mathsf{cov}(\mathbf{Y}_t) = \mathsf{E}(\mathbf{D}_t), \quad \mathsf{cov}(\mathbf{Y}_t, \mathbf{Y}_{t+h}) = \mathbf{0},\ h \neq 0. \tag{17}$$

For a dynamic estimation of unknown quantities in the LDL decomposition (14) for the model (1) with a given entire sample $\{\mathbf{X}_1, \ldots, \mathbf{X}_T\}$, the state space modelling seems to be truly useful. The issue of state space models and associated Kalman recursions is elaborated in various publications, see e.g. [3].

With regard to (15), assume the following dynamic discrete-time linear state space representation (a generalized analogy of the recursive OLS estimator):

$$\boldsymbol{\beta}_{t+1} = \boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t, \tag{18}$$

$$\mathbf{X}_t = \mathbf{G}_t\boldsymbol{\beta}_t + \mathbf{Y}_t,\ t = 1, \ldots, T. \tag{19}$$

Denote $\boldsymbol{\beta}_t$ the $(N \times 1)$ vector containing all unknown row elements of $\mathbf{L}_t$. The $(n \times N)$ matrix $\mathbf{G}_t$ clearly includes only zeroes and the elements of $\mathbf{X}_t$ due to (15), $N = n(n-1)/2$. The state equation (18) is a pure multivariate random walk, but one can clearly suppose some richer versions.

Moreover, recapitulate crucial assumptions of this model: $\left\{\boldsymbol{\beta}_1, ((\boldsymbol{\varepsilon}_t)^\top, \mathbf{Y}_t^\top)^\top\right\}_t$ is a sequence of uncorrelated random vectors with finite second moments and $\mathsf{E}\boldsymbol{\varepsilon}_t = \mathbf{0}$, $\mathsf{cov}(\boldsymbol{\varepsilon}_t) = \mathbf{M}_t$, $\mathsf{E}\mathbf{Y}_t = \mathbf{0}$, $\mathsf{cov}(\mathbf{Y}_t) = \mathbf{N}_t$ and also $\mathsf{cov}(\boldsymbol{\varepsilon}_t, \mathbf{Y}_t) = \mathbf{0}$. The matrix $\mathbf{N}_t$ is supposed to be diagonal due to (17). Further, the initial state vector $\boldsymbol{\beta}_1$ is assumed to be random with the expected value $\mathsf{E}(\boldsymbol{\beta}_1) = \mathbf{0}$ and the variance $\mathsf{var}(\boldsymbol{\beta}_1) = \kappa\mathbf{I}_N$, $\kappa \to \infty$, i.e. the so-called standard diffuse prior. The covariance matrices $\mathbf{M}_t$ and $\mathbf{N}_t$ could be captured essentially by constant parameter matrices which are estimated via a (quasi) maximum likelihood procedure, see [3]. In the given framework, the standard Kalman recursive formulas for predicting, filtering and smoothing can be used to obtain corresponding estimators of $\boldsymbol{\beta}_t$ and consequently also the transformed vector $\mathbf{Y}_t$ with simultaneously uncorrelated components. The conditional variances $d_{ii,t}$ of $Y_{i,t}$ can be proceed by some advanced (univariate) methods, e.g. by means of GARCH models, see [8].

5 Empirical study: Stock and bonds

To examine empirical performance of the suggested approach to conditional covariance and correlation modelling based on the LDL decomposition and the associated state space representation (see Section 4), an empirical application is considered. Namely, the daily bivariate correlations between log-returns on stocks and bonds are investigated. All mentioned estimators of conditional correlations are compared from various modelling perspectives. In general, there is no consensus about how stocks and long term bonds are related. Short-run correlations are obviously affected, e.g., by new announcements. Long-run correlations between these two types of assets should be state dependent, e.g. driven by macroeconomic factors. The way how the correlation links respond to these factors might be changed over time, see [4]. The daily logarithmic returns on the S&P 500 index and 30-year bond futures from 3 January 1990 to 30 April 2013 are observed, see [9] (get the quotes for ^GSPC and ^TYX, respectively).

The estimated conditional correlations are graphically presented in Figure 1. It indicates numerous similarities among the introduced models of conditional correlations, e.g. analogies among the state space technique (see Section 4) and both previously mentioned DCC procedures. The diagonal BEKK(1,1), the EWMA and especially the OGARCH estimators seem to be more volatile, but the main trends in the correlations are comparable with the other estimators. Note that the CCC approach is not truly competitive in this comparison due to the fact that this correlation estimator is constant, i.e. $\hat{\rho}_t = 0.029$. Generally, the time-varying correlations are mostly negative during the 90's, rather positive after the year 2000 and positive at the end of the observed period. Table 1 contains the sample correlation matrix of the estimated time-varying conditional correlations (the used abbreviations clearly follow the labels in Figure 1. In particular, one can see that the delivered dynamic correlations are strongly positively correlated with the others, which is also obvious from Figure 1.

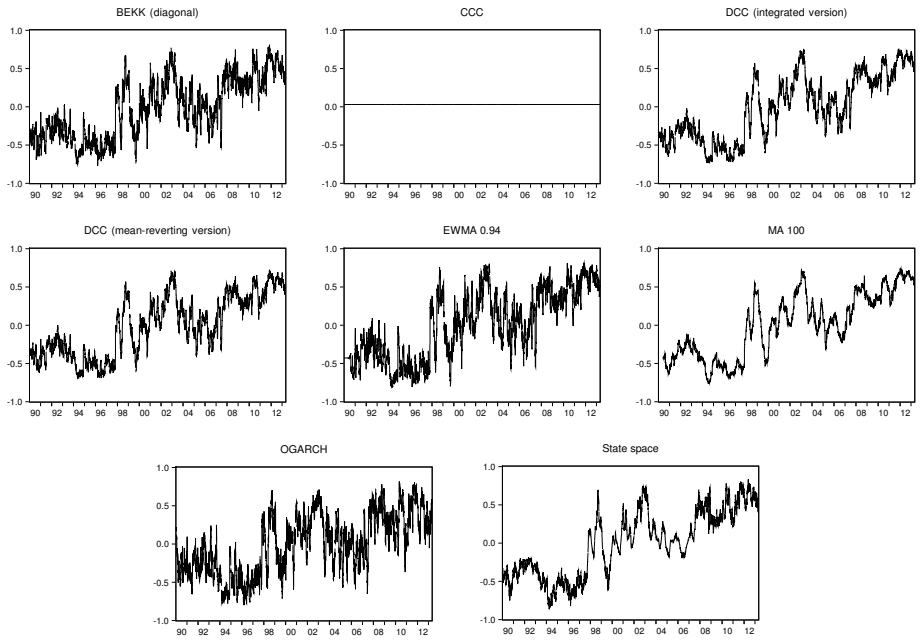


Figure 1 The estimated conditional correlations between the S&P 500 and the 30Y bond futures.

	BEKK (diag)	DCC (int)	DCC (mr)	EWMA 0.94	MA 100	OGARCH	Sspace
BEKK (diag)	1.000						
DCC (int)	0.979	1.000					
DCC (mr)	0.987	0.999	1.000				
EWMA 0.94	0.996	0.980	0.987	1.000			
MA 100	0.929	0.976	0.966	0.929	1.000		
OGARCH	0.949	0.918	0.929	0.944	0.858	1.000	
Sspace	0.947	0.956	0.957	0.948	0.937	0.906	1.000

Table 1 The sample correlations of the estimated conditional correlations.

For other performance measures, one can calculate various regression-based tests on portfolio returns, $\mathbf{w}_t\mathbf{X}_t$, where $\mathbf{w}_t$ is a vector of portfolio weights. Note that the conditional variance of $\mathbf{w}_t\mathbf{X}_t$ is $\mathbf{w}_t^\top\mathbf{H}_t\mathbf{w}_t$. Two types of portfolio weights are considered: the *equally weighted portfolio* (EWP), i.e. $\mathbf{w}_t = \mathbf{1}/n$, $\mathbf{1}$ is the $(n \times 1)$ vector of ones, and the *minimum variance portfolio* (MVP), i.e. $\mathbf{w}_t = (\mathbf{H}_t^{-1}\mathbf{1})/(\mathbf{1}^\top\mathbf{H}_t^{-1}\mathbf{1})$.

First, the Engle-Colacito regression is defined as $\{(\mathbf{w}_t\mathbf{X}_t)^2/(\mathbf{w}_t^\top\hat{\mathbf{H}}_t\mathbf{w}_t)\} - 1 = \lambda + \xi_t$, where ξ_t is an error term. The null hypothesis $\lambda = 0$ is verified (in the presence of an HAC robust estimator of the standard deviation of ξ_t), see [6]. Second, the LM test of ARCH effects is based on the property that the series $\{(\mathbf{w}_t\mathbf{X}_t)^2/(\mathbf{w}_t^\top\mathbf{H}_t\mathbf{w}_t)\}$ does not exhibit serial correlation. The null hypothesis that $\{(\mathbf{w}_t\mathbf{X}_t)^2/(\mathbf{w}_t^\top\hat{\mathbf{H}}_t\mathbf{w}_t)\}$ is serially uncorrelated is tested as in [8]. In addition, the Ljung-Box test statistics finding serial correlations in the standardized residual series are computed, see again [8].

Table 2 delivers the previously mentioned performance measures of the particular conditional correlations, i.e. the sample standard deviations $\hat{\sigma}$ and p-values corresponding to the Ljung-Box statistics Q with 9 lags, the ARCH (LM) tests with 5 lags and the Engle-Colacito (EC) tests for both portfolios.

	$Q(9)$	$\hat{\sigma}_{EWP}$	LM_{EWP}	EC_{EWP}	$\hat{\sigma}_{MVP}$	LM_{MVP}	EC_{MVP}
BEKK (diag)	0.2654	0.0090	0.0676	0.5108	0.0082	0.9865	0.1363
CCC	0.0430	0.0090	0.0000	0.2513	0.0083	0.3025	0.8774
DCC (int)	0.3911	0.0090	0.0023	0.9277	0.0082	0.7927	0.0750
DCC (mr)	0.3808	0.0090	0.0082	0.7824	0.0081	0.9035	0.2410
EWMA 0.94	0.0755	0.0090	0.2928	0.0158	0.0083	0.9993	0.0000
MA 100	0.0196	0.0091	0.0000	0.0062	0.0085	0.0030	0.0000
OGARCH	0.3821	0.0090	0.2977	0.9847	0.0091	0.3537	0.9403
Sspace	0.3382	0.0090	0.7164	0.7810	0.0080	0.9978	0.6716

Table 2 The comparison of the different conditional correlation models.

6 Conclusion

An innovative approach to conditional covariance and correlation modelling was proposed. The suggested two-step technique motivated by the principle of the OGARCH method is based on the LDL factorization of the conditional covariance matrix and the associated discrete linear state space modelling. Jointly, they can deliver a dynamic orthogonal transformation of given stochastic vector data. This indeed simplifies further conditional covariance analysis due to its simultaneously uncorrelated components. In the considered empirical financial framework, the proposed method demonstrated its capabilities. Namely, it is at least comparable with other introduced models, see e.g. Figure 1 and Table 2, and it is numerically comfortable due to the state space modelling with the associated Kalman recursions. Finally, further research will be focused on a construction of a one-step method using augmented discrete linear state space modelling.

Acknowledgements

This work was supported by the grant GA P402/12/G097.

References

[1] Bauwens, L., Laurent, S. and Rombouts, J. V. K.: Multivariate GARCH Models: A Survey, *Journal of Applied Econometrics* **21** (2006), 79–109.

[2] Bollerslev, T.: Modeling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model, *Review of Economics and Statistics* **72** (2000), 498–505.

[3] Brockwell, P. J. and Davis, R. A.:*Introduction to Time Series and Forecasting*. Springer, New York, 2002.

[4] Engle, R.: *Anticipating Correlations: A New Paradigm for Risk Management*. Princeton University Press, Princeton, 2009.

[5] Engle, R.: Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroscedasticity Models, *Journal of Business & Economic Statistics* **20** (2002), 339–350.

[6] Engle, R. and Colacito, R.: Testing and Valuing Dynamic Correlations for Asset Allocation, *Journal of Business & Economic Statistics* **24** (2006), 238-253.

[7] Harville, D. A.: *Matrix Algebra from a Statistician's Perspective*. Springer, Secaucus (NJ), 1997.

[8] Tsay, R. S.: *Analysis of Financial Time Series*. Wiley-Interscience, Hoboken (NJ), 2005.

[9] *Yahoo! Finance* [on-line]. Available on http://finance.yahoo.com, accessed 5 May 2013.

On alternative remedy for multicollinearity in data set

Robert Hlavatý[1]

Abstract. The paper deals with multicollinearity in terms of econometrics model. This undesirable phenomenon affects results of mathematical modelling adversely wherever there is a need for expression of dependency of endogenous and exogenous variables. Whether there is a situation with occurrence of more than one exogenous variable it is possible this unwanted dependency could occur and it is necessary to eliminate it.
Following text brings a new way of resampling for variable causing multicollinearity. An analogical approach for eliminating multicollinearity has been already known sometimes it does not bring good results though. The proposed approach offers brand new approach for multicollinearity elimination and brings various advantages compared to approaches known before. Functionality of procedures described is validated with the help of simulation model.

Keywords: multicollinearity, simulation, Markov chain, ergodicity.

JEL Classification: C15
AMS Classification: 91G70

1 Introduction

According to Hušek [5], the multicollinearity is an undesirable phenomenon found in data sets with purpose of describing dependence between the response variables and explanatory variables. It is possible to come across such data sets usually in the field of statistics and similar disciplines like econometrics, economic forecasting, stochastic modelling etc. Spanos and McGuirk [9] state that there are basically two types of problems with multicollinearity. Either it can be unwanted dependence between regressors or it can regard conditionality of correlation matrix$X^T X$. This paper will be dedicated to the first class of these two problems, particularly to multicollinearity in data set and to possibilities of its remedy while eliminating a loss of original input information. York [12] says there is main reason for multicollinearity existence – a lack of input information. If one is not able to improve data set in this regard, it is still possible to eliminate this unwanted phenomenon. There are several ways of remedy for multicollinearity in econometrics model. The utmost way is excluding a variable (variables) causing multicollinearity from the model [11]. This is only possible when the variable is not of great importance for the model. In good quality model, all of variables should be of great importance however, that is why this way of remedy is not quite advisable. The other option is excluding only some observations of the time series because it was found that some observations might affect multicollinearity to greater extent than other observations [7].

To avoid devaluation of input data by such ways there is an option of transforming the values of a variable to different values which would lead to reduction of dependency inside the model to acceptable level. By the acceptable level the low dependency is meant. From a subjective point of view it is sometimes meant as medium dependency. According to Hendl [4] low dependency is characterized by correlation coefficient $|r|$ lying in interval $|r| \in \langle 0{,}1\ \ 0{,}3\rangle$. Medium dependency is then for $|r| \in \langle 0{,}3\ \ 0{,}7\rangle$.

There are various ways of remedy for multicollinearity using data transformation. There is *ridge regression* among those oldest design by Hoerl and Kennard in 1970. Ranjit [8] proposes solution of this problem by *principal component regression*. There is also very popular approach called *bootstrapping* described by Efron and Tibshirani [2].

The purpose of following experiment is to design different way of replacing original values by n-valued sequence that will help to reduce unwanted correlation while preserving properties of original data set in best possible way. For this, the instrument of Markov chain will be used. This apparatus can be used very well for dummy variables because Markov chains are based on a sequence of independent events where every event value is dependent only on previous event value [1]. Transforming original variable to dummy variable undoubtedly carries such property. In case of dummy variable, this chain can consist of particular number of values. Creating a transition matrix of the Markov chain, it is then possible to define matrix of steady states if the transition matrix is regular [10].

[1] Czech university of life sciences, Faculty of economics, Department of systems engineering.
Kamýcká 129, 165 21, Praha 6 – Suchdol. hlavaty@pef.czu.cz

The aim of the paper is to create, upon steady states probabilities, new different n-valued vector and from this vector create back a whole new vector consisting of new values. This experiment should prove that new vector values keep properties of the original values. Simultaneously, the multicollinearity should not be present in the data set anymore.

2 Experiment design

2.1 Input data

In the beginning it is necessary to create representative experimental data set, on which the multicollinearity presence will be tested followed by eliminating the multicollinearity. The multicollinearity means correlation between two variables that is why there is no need to create large model. A pair of two variables x_1, x_2 will be sufficient enough. These variables are not real but artificially created. It is also necessary to have sufficient number of observation, so the multicollinearity could be created artificially between these two variables. Both of them will consist of 15 observations. According to Hančlová and Tvrdý [3] it is recommended to have at least 30 observations when operating with any time series. It is, however, not necessary for the simulation model. It would also bring complications for creating the artificial multicollinearity between variables. The more observations we have, the less chance of strong multicollinearity existence there is. It is then more difficult to create artificial multicollinearity. For better understanding of principles of this transformation the 15 values are enough for preserving clarity of this illustrative example.

The presence of the multicollinearity is then desired in the model, so it can be further eliminated. Creating artificial multicollinearity is based on a fact that there sure will be strong dependency between the variables if the trend functions of variables will be nearly identical in slope and different in constant. This is ensured by generating values of variables in the following way:

Let x_1 be random integer number from interval $\langle 40 \;\; 60 \rangle$.

Then x_2 is integer number defined by

$$x_2 = 2x_1 + a \,;\, a \in Z \,;\, a \in \langle -2; 2 \rangle \tag{1}$$

Variable x_2 is then integer number created as a double of x_1 with the maximum deviation of 2. Slopes of both functions will not be identical due to this deviation. This would not be desirable for it would cause so called perfect multicollineaerity (correlation coefficient would equal 1; see formula 4). That is why the values of x_2 are varied by the deviation. The number 2 was derived empirically. For lower deviation, the correlation was becoming almost too perfect, for higher deviation there was not always strong correlation between the variables. A critical value of strong correlation in this experiment is considered 0,8.

To calculate a degree of dependency the Farrar-Glauber test is used. It is based on creating a correlation matrix

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1k} \\ r_{21} & r_{22} & \cdots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \cdots & r_{kk} \end{pmatrix} \tag{2}$$

Its elements are calculated by multiplication of matrices $X^T X$. Elements of matrix X are calculated from input data by the formula

$$\mathrm{x_{ij}} = \frac{\mathrm{x} - \bar{\mathrm{x}}_\mathrm{j}}{\mathrm{s}.\sqrt{\mathrm{n}}} \tag{3}$$

where x represents real original values and $\bar{x}_j$ is mean of the variable, n is number of observations and s standard deviation of the variable. Using this formula we receive the correlation matrix

$$\mathrm{R} = \begin{pmatrix} 1 & \mathrm{r_{12}} \\ \mathrm{r_{21}} & 1 \end{pmatrix} \tag{4}$$

Diagonal values will always be equal 1 and the matrix is triangular so every value will be expressed twice. Test criterion here can be expressed as determinant of the matrix. However its value has no economical interpretation. That is why this criterion will be value $r_{12} = r_{21}$ from the matrix. Let us say that the degree of dependency is absolute value of $r_{12} = r_{21}$. Should this value reach over 0,8 then there is a multicollinearity between x_1 and x_2.

2.2 Bases for eliminating multicollinearity

The least appropriate solution would surely be omitting multicollinearity-causing variable(s) from a model. This solution would make impossible further examination of how omitted variable affected dependent variable. The fact of multicollinearity existence in a model does not mean that its estimated parameters are not statistically significant. On the contrary, the omitted explanatory variable could have highest degree of explanation for endogenous variable. Possible omitting the variable gives us whole new model and interpretation of this model is inappropriate because it does not consider influence of the omitted variable anymore. Therefore omitting any variable is quite irresponsible since in well-constructed model, every variable should have its importance.

The different solution is replacing original values by new values; let's say binary values 0 and 1. If the value increased compared to the previous value, the value 1 will be assigned instead. If the value decreased, the value 0 will be assigned. In this case there is a great loss of original information because reducing any data to 0 or 1 is quite vague. It is not even clear which of these two values should be assigned if there was no change compared to previous observation. It is then reasonable to use broader spectrum of values instead of two. Number of new replacing values depends on one particular trend of the variable. If there are larger fluctuations in the trend of the variable, it is then recommended to use more of new values for replacing the original. In this experiment the range of 5 values was chosen to replace the original values. The way of replacing the values is as follows:

Change compared to previous observation	New value
Significant decrease $(20\%, \infty)$	0
Slight decrease $(10\%, 20\%)$	0,25
Small change $(0\%, 10\%)$	0,5
Slight increase $(10\%, 20\%)$	0,75
Significant increase $(20\%, \infty)$	1

Table 1 Replacing original values with the new ones

Given interval boundaries of ∞ are merely theoretical. It actually means all changes greater than 20% considering the original data set was ridden of extreme values that would otherwise distort interpretation of the model. In our experimental data, let there be maximum change of 50% compared to previous observation. Still the probability of such change is very low. Probability density of change compared to previous observation is equivalent to *normal distribution* here. That is why these intervals are concentrated more to small change than to maximum possible change.

2.3 Using Markov chains

The basic assumption here says there is a state n in discrete time dependent only on state $n-1$. Knowing a whole sequence of transitions of states we are able to define transition matrix of Markov chain P. Let us first define frequencies of each type of transition from one observation to another. Number of these frequencies is 25. Those are actually all pairs of new values including the transition of one state to the same state. If any type of transition is not present, the matrix must be reduced of this type. The matrix of frequencies of transitions will have the following structure:

$$\mathrm{N} = \begin{pmatrix} 0 \to 0 & 0 \to 0{,}25 & 0 \to 0{,}5 & 0 \to 0{,}75 & 0 \to 1 \\ 0{,}25 \to 0 & 0{,}25 \to 0{,}25 & 0{,}25 \to 0{,}5 & 0{,}25 \to 0{,}75 & 0{,}25 \to 1 \\ 0{,}5 \to 0 & 0{,}5 \to 0{,}25 & 0{,}5 \to 0{,}5 & 0{,}5 \to 0{,}75 & 0{,}5 \to 1 \\ 0{,}75 \to 0 & 0{,}75 \to 0{,}25 & 0{,}75 \to 0{,}5 & 0{,}75 \to 0{,}75 & 0{,}75 \to 1 \\ 1 \to 0 & 1 \to 0{,}25 & 1 \to 0{,}5 & 1 \to 0{,}75 & 1 \to 1 \end{pmatrix} \tag{5}$$

So every element of matrix corresponds to one type of transition from one state to another. Let us define numbers of individual transition types in the following way:

$$\mathrm{N} = \begin{pmatrix} \mathrm{n}_{11} & \mathrm{n}_{12} & \cdots & \mathrm{n}_{1\mathrm{k}} \\ \mathrm{n}_{21} & \mathrm{n}_{22} & \cdots & \mathrm{n}_{2\mathrm{k}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{n}_{\mathrm{k}1} & \mathrm{n}_{\mathrm{k}2} & \cdots & \mathrm{n}_{\mathrm{kk}} \end{pmatrix} \tag{6}$$

To gain transition matrix of Markov chain it is necessary to have row sums equal 1 (100%). This can be achieved by normalizing of matrix N:

$$\mathrm{p_{kl}} = \frac{n_{kl}}{\sum_{k=1}^{l} n_{kl}} \tag{7}$$

Then the transition matrix of Markov chain is received:

$$\mathrm{P} = \begin{pmatrix} \mathrm{p_{11}} & \mathrm{p_{12}} & \cdots & \mathrm{p_{1k}} \\ \mathrm{p_{21}} & \mathrm{p_{22}} & \cdots & \mathrm{p_{2k}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{p_{k1}} & \mathrm{p_{k2}} & \cdots & \mathrm{p_{kk}} \end{pmatrix} \tag{8}$$

The transition matrix expresses probabilities of transition from one state to another. For example it is possible to say that probability of transition from state 0 to state 0 would be here p_{11}. Practically, this means if the value of variable x_1 significantly decreased in observation $n-1$, it will significantly decrease again in observation n with the probability p_{11}. These probabilities are expressed only for the following observation. It is, however, possible to calculate limiting probabilities of the states. This can be done by solving following the system of linear equations (expressed as an augmented matrix of system of linear equations using coefficients from transition matrix):

$$\left(\begin{array}{cccc|c} \mathrm{p_{11}}-1 & \mathrm{p_{21}} & \cdots & \mathrm{p_{k1}} & 0 \\ \mathrm{p_{12}} & \mathrm{p_{22}}-1 & \cdots & \mathrm{p_{k2}} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathrm{p_{1(k-1)}} & \mathrm{p_{2(k-2)}} & \cdots & \mathrm{p_{k(k-1)}}-1 & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{array}\right) \tag{9}$$

Solving the system of linear equations a basic solution vector is received. This vector expresses limiting probabilities

$$p^{\infty} = (p1, p2, \ldots, pk) \tag{10}$$

where – in case of this experiment – the $k = 5$, because we have 5 different states of Markov chain. These values can be interpreted in the following way. From the limiting (long term) point of view, the variable will lie in state **0** (significant decrease) with the probability p_1, in state **0,25** (slight decrease) with probability p_2 etc. It is hereby possible to gain general structure of the variable (=time series). This structure is based on frequencies of transitions and reflects the structure of original data.

2.4 Creating new values of variable

With the help of the limiting structure **(10)** it is now possible to create new vector of the variable. Values from figure **(10)** will divide the probability spectrum into five differing sections where each of this section represents one of the values $\{0; 0{,}25; 0{,}5; 0{,}75; 1\}$. Boundaries of these sections will be created as an accumulation of values of limiting probabilities. This situation is described in figure 1:

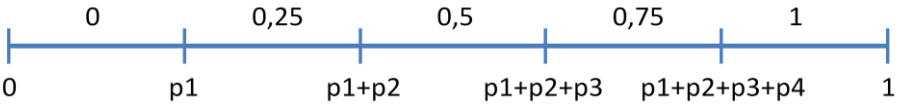


Figure 1 Probability spectrum

To get new values of the variable, a random number from $(0\,,1)$ will be generated for every observation. Generated number will always belong into one of the sections (intervals). Depending on which section it is, we will assign the appropriate value of section $\{0; 0{,}25; 0{,}5; 0{,}75; 1\}$ instead of original values. Thus we receive new variable consisting of values $\{0; 0{,}25; 0{,}5; 0{,}75; 1\}$ instead original values causing multicollinearity. Performing the test **(3)** we find out that multicollinearity is not present anymore. The new five-valued vector will be, of course, different every time we resample the original values. This is caused by randomly generated numbers. Therefore only one replacing of values cannot be considered conclusive. Performing this operation 100 000 times (i.e. 100 000 replacing of values and 100 000 multicollinearity tests) the correlation coefficient of the new variables is 74% lower than of the original variables. This eliminates multicollinearity reliably.

It is not possible to use the new vector practically for regression coefficient estimation because the new values represent original values only in terms of the limiting structure. These values, however, does not really correspond to original values in terms of their meaning. E.g. if the original values actually expressed crop yields in t/ha in a range 5-10 t/ha, then replacing these values with values $\{0; 0{,}25; 0{,}5; 0{,}75; 1\}$ would change the whole meaning of the variable and further estimation of the model parameters would be pointless.

2.5 Backwards transformation

For the obvious reasons it is necessary to transform five-valued vector into original ranges in which the original values were found. This can be done by backward approach of the first transformation of the data. This backward transformation will work with mean of the original data. For every observation, it is necessary to multiply original mean value by appropriate coefficient depending on which value of $\{0; 0{,}25; 0{,}5; 0{,}75; 1\}$ was assigned to the observation in the first transformation. E.g. in *n*-th observation the value **0,75** was assigned. This value originally expressed slight increase (see table 1). The slight increase meant change between 10% and 20%. Let us then generate random change from this interval and add it to the mean value of the original data. In other words we will multiply mean value of the original data by random coefficient generated in $\langle 1{,}1; 1{,}2\rangle$. All five ways of backwards transformation can be found in the following table:

Value assigned after first transformation	Coefficient for multiplying the mean value
0	Random number $\in \langle \mathbf{0,7}; \mathbf{0,8}\rangle$
0,25	Random number $\in \langle \mathbf{0,8}; \mathbf{0,9}\rangle$
0,5	Random number $\in \langle \mathbf{0,9}; \mathbf{1,1}\rangle$
0,75	Random number $\in \langle \mathbf{1,1}; \mathbf{1,2}\rangle$
1	Random number $\in \langle \mathbf{1,2}; \mathbf{1,3}\rangle$

Table 2 Replacing original values with the new ones

Boundary values 0,7 and 1,3 limit the new value from both left and right side to maximum change of 30% to prevent generating of extreme values.

3 Illustrative example

By illustrative example we understand only **one cycle (sample)** of creating new vector of values for **one** variable. Original values of variable where the multicollinearity is present:

Original values to be transformed (x_1)	56	60	59	44	42	50	48	59	59	55	47	40	56	48	60
Variable having unwanted correlation with variable in the first row (x_2)	111	119	116	90	86	98	98	119	117	112	92	79	114	95	121
Transforming x_1 to five-valued function	-	0,5	0,5	0	0,5	0,75	0,5	1	0	0,5	0,25	0,25	1	0,25	1

Table 3 Original variables and transformation to five-valued function

Upon frequencies of the values in the third row it is possible to create the matrix of frequencies and the transition matrix. From the transition matrix we receive liming probabilities vector (rounded) according to figure **(10)**:

$$p^{\infty} = (0{,}17; 0{,}26; 0{,}28; 0{,}06; 0{,}23) \tag{11}$$

By accumulating these limiting probabilities we gain 5 intervals (rounded): $(0; 0{,}17)$, $(0{,}17; 0{,}43)$, $(0{,}43; 0{,}71)$, $(0{,}71; 0{,}77)$, $(0{,}77; 1)$. The generated number from $(0\,; 1)$ will always belong to one of these intervals. According to this, the appropriate value $\{0; 0{,}25; 0{,}5; 0{,}75; 1\}$ will be assigned. The fact that the random number would be exactly the same as one of the interval boundaries is not taken into consideration because the probability of such phenomenon is practically zero. Row of generated random numbers for every observation and appropriate assigned values and new values after backwards transformation are in table 4 (with the original values of x_1 for comparison):

Random number (rounded)	0,58	0,02	0,94	0,57	0,15	0,28	0,84	0,74	0,13	0,46	0,65	0,47	0,77	0,44
Assigned value	0,5	0	1	0,5	0	0,25	1	0,75	0	0,5	0,5	0,5	0,75	0,5
New value	57	41	66	53	37	45	66	58	41	47	53	55	62	48
Original value (without 1st observation)	56	60	59	44	42	50	48	59	59	55	47	40	56	48

Table 4 Transformation of values

In this case the 1st original observation was omitted since there are only 14 transitions between 15 values. Performing the backwards transformation the missing value cannot be easily made up from nothing. It would be

possible to think of the 1st original value as a transition from non-existent previous value, thus the change would be marked as *no change* (0%).

Original correlation coefficient was 99% which means almost perfect multicollinearity. After resampling there is a correlation only 12%. Multicollinearity is therefore eliminated. Means of both original vector and new vector are 51,64 and 52,07 respectively. Performing 100 000 cycles of transformation leads to almost the same values of means.

4 Results

Using the backwards transformation we receive values of the variable very similar to original values. Performing 100 000 cycles of primary and backwards transformation we will receive 100 000 versions of values of new variable. All of these samples bring 77% improvement compared to original correlation coefficient. These samples alone do not have a meaning and it is necessary to work with these samples altogether. One sample alone (i.e. one resampling of 14 values to new values) is based on generating 14 random values and is not sufficiently conclusive. Limiting probabilities can be used only when the number of random experiments is large enough.

Primary objective of regression models is not eliminating multicollinearity but it is estimation of parameters (regression coefficients). It is necessary to identify which of the 100 000 versions should be used for parameter estimation. By the nature of the simulation model it is not possible to choose only one sample, but it is crucial to work with the values altogether. It is possible to use gradually all the samples instead of original values. The mean of values transformed differs from the mean of original values only by $\pm 0{,}03\%$ after 100 000 cycles.

Compared to some other approaches, this simulation brings several differences in terms of remedy for multicollinearity. It is undoubtedly better than those ways based upon reduction of input data. In the proposed solution there is also loss of original data – namely the first observation of the variable. This observation, nevertheless, affects the structure of the new vector anyway. It could be even possible to place this variable back to the vector after the transformation. Placing the one value will not cause the multicollinearity again. Proposed approach does not add any extraneous elements to the original data; it only dissolves the data and then solves them again. This process is repeated 100 000 times which makes original static data somehow dynamic.

5 Conclusion

The proposed method can be used for multicollinearity elimination while not changing original structure of the data in any harsh way. In this algorithm, only two kinds of loss can be noticed. Firstly it is loss of the first value of the variable – which can be, however, acquired again – as described in *results*. The second loss is deviation from original data caused by resampling which is actually negligible. The proposed simulation approach is opened, only the algorithm is consistent. Its parameters can be changed as one desires according to character of input data. This algorithm cannot be used for very small data sets. Such data sets are not suitable for parameters estimation either. The number of values assigned can be changed depending on the structure of the data set. In this experiment, five values were used. Future extension of this experiment is about to show, how many replacing values should be chosen in general and how much this transformation affects further estimation of parameters e.g. in econometrics model.

References

[1] Borovkov, K.: *Elements of stochastic modelling*. New Jersey: World Scientific, 2003.

[2] Efron, B. and Tibshirani, R.: A introduction to the bootstrap. Boca Raton: Chapman, c1994, xvi, 436 s Vallander, S. S.: Calculation of the Wasserstein distance between probability distributions on the line. *Theory of Probability and its Applications* 18 (1973), 784–786.

[3] Hančlová, J. and Tvrdý, L.: *Úvod do analýzy časových řad*. Ekonomická fakulta, VŠB-TU, Ostrava, 2003.

[4] Hendl, J.: *Přehled statistickým metod zpracování dat*. Praha: Portál, 2004.

[5] Hušek, R.: *Ekonometrická analýza*. Vyd. 1. Praha: Oeconomica, 2007.

[6] Kába, B.: *Teorie pravděpodobnosti*. Státní pedagogické nakladatelství, Praha, 1977.

[7] Öztürk F. and Agdenis F.: Ill-conditioning and multicollinearity. Linear Algebra and its Applications (2003), vol. 321, no. 1-3, p. 295–305.

[8] Ranjit K.P.: *Multicollinearity - Causes, effects and remedies*. I.A.S.R.I., Library Avenue, New Delhi, 2008.

[9] Spanos A. and McGuirk A.: The problem of near-multicollinearity revisited: erratic vs systematic volatility. *Journal of Econometrics* (2002), vol. 108, no. 2, p. 365–393.

[10] Tijms H.C.: A first course in stochastic models. 2nd ed, 2003.

[11] Tvrdoň, J.: Ekonometrie. Provozně ekonomická fakulta, Praha, 2003.

[12] York R.: Residualization is not the answer: Rethinking how to address multicollinearity. *Social Science Research* (2002), vol. 108, no. 2, p. 1379–1386.

DSGE model with housing sector: application to the Czech economy

Miroslav Hloušek[1]

Abstract. This paper deals with multi-sector DSGE model that is estimated on Czech economy data. The model is taken from Iacoviello and Neri (2010). There are two production sectors: consumption/investment goods sector and housing sector. These sectors contain various types of nominal and real rigidities and also different technological trends. On the demand side, there is financial friction in the form of collateral constraint which affects borrowing capacity and consumer spending. The model is estimated using Bayesian techniques. The parameters are economically interpreted, the model moments are compared to moments from data and dynamical properties of the model are studied using impulse responses and variance and shock decompositions. Results show that monetary policy has more pronounced effect on consumption and output when houses are better collateralizable. Consumption shocks, housing technology and housing preference shocks played important role in fluctuation of the real variables while inflation target shocks and cost-push shocks influenced mostly nominal variables. However, recent boom and bust in housing prices was caused primary by housing preference shocks (demand side shocks). Supply shocks were also significant but to much less extent.

Keywords: housing, DSGE model, collateral constraint, Bayesian estimation.

JEL classification: E37
AMS classification: 91B64

1 Introduction

Development in housing market in recent years attracted widespread attention, especially in U.S. where it was thought as trigger of the financial crises. Situation in the Czech Republic was not so severe but connection between housing market and macroeconomy also deserves more detailed examination. It is purpose of this paper. Multi-sector DSGE model with housing market is estimated on Czech data using Bayesian techniques, the data fit of the model is assessed and dynamical properties of the model are examined. The results from variance decomposition show that consumption shocks, shocks to housing technology and preferences and inflation target shocks were the most important driving forces for fluctuations of key macro variables. On the other hand, shock decomposition shows that high prices of houses were caused primary by housing preference shocks; housing technology shocks and consumption shocks contributed only partly. Capability of monetary policy to influence consumption and output heavily depends on loan-to-value ratio. If households have better access to credit, impact of monetary policy to above mentioned variables is substantially increased while impact to inflation is changed only slightly.

The rest of the paper is organized as follows. Section 2 describes the structure of the model, Section 3 briefly comments data and estimation technique. Results of the estimation, data fit of the model and dynamical properties are discussed in Section 4. Final section concludes.

2 Model

The model is borrowed from Iacoviello and Neri [4] and ranks among medium-scale models. Due to shortage of space I describe the model only verbally and refer the reader to the original paper for details. Figure 1 helps for better orientation. There are two types of households: patient (lenders) and impatient (borrowers). Patient households work, consume and accumulate housing. They also own capital and land and supply funds to firms and

[1]Masaryk University, Department of Economics, Lipová 41a, 602 00 Brno, hlousek@econ.muni.cz

to impatient households. Impatient households also work, consume and accumulate housing, but they are credit constrained and their houses serve as collateral. Variations in housing values affect their borrowing and spending.

Production side of the model economy is divided into two sectors with different rates of technological progress. The non-housing sector uses capital and labor for production of wholesale goods that are subsequently used for production of consumption, business investment goods and intermediate goods. The housing sector produces new houses that are added to existing stock. This sector uses capital, labor, land and intermediate goods. There are nominal wage rigidities in both housing and non-housing sectors and price rigidity in the retail sector. The rigidities are enabled by existence of labor unions and retailers that has some market power and can influence their wages and prices. The rigidities are modeled in Calvo [3] style, the production functions have standard Cobb-Douglas form. Monetary policy follows Taylor rule with interest rate smoothing and attention to inflation and output gap. Several exogenous shocks (mostly following AR(1) processes) are added so that the model can be taken to data.

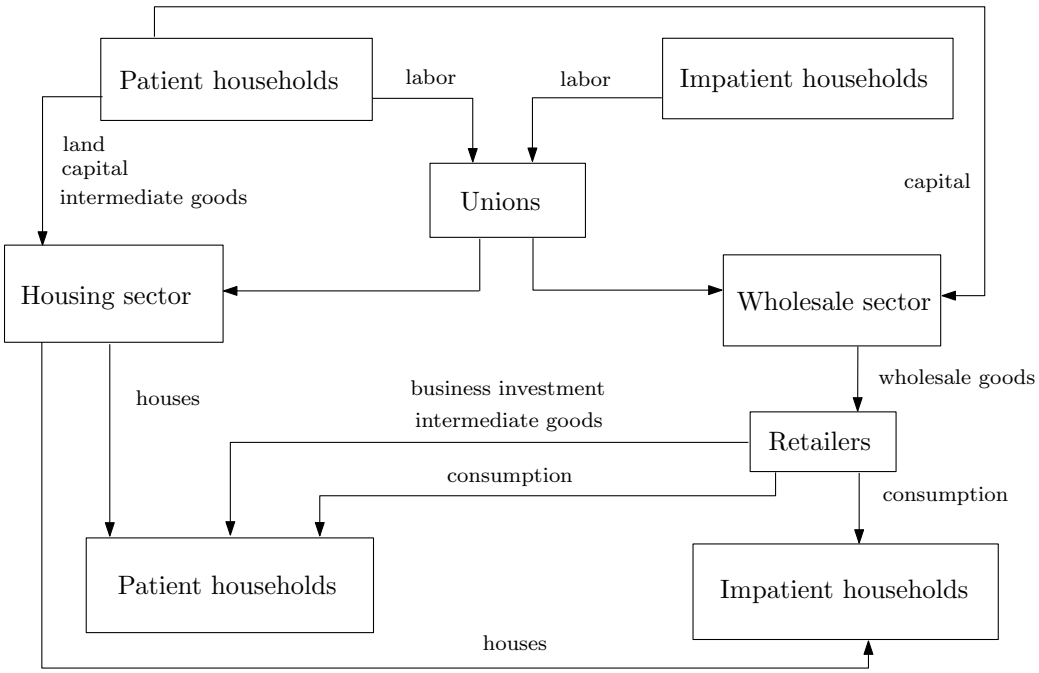


Figure 1 Model structure

3 Data and estimation

The model is estimated using data for following model variables: consumption (C_t), residential investment (IH_t), non-residential investment (IK_t), real house prices (q_t), inflation (π_t) nominal interest rate (R_t), worked hours and wage inflation in housing (NH_t, WH_t) and wholesale sector (NC_t, WC_t) respectively. Time series are quarterly, they are obtained from the Czech Statistical Office and the Czech National Bank databases and cover time period 1998:Q1 – 2012:Q4. The detailed description can be found on web appendix [7].

Some of the model parameters are calibrated according to Iacoviello and Neri [4] and data from national accounts. Specifically, discount factor of patient households was set to 0.9957 which corresponds to the real interest rate (1.7 %) calculated from data. Discount factor of impatient households was set to much lower value 0.97. Loan-to-value ratio (LTV) was calibrated to 0.75 as a combination of estimated values of LTV's for households and entrepreneurs in Hloušek [6]. Full description of calibrated parameters and their values are quoted in [7].

The rest of the model parameters was estimated using Bayesian techniques. Posterior distribution of the parameters was obtained by Random Walk Chain Metropolis-Hastings algorithm. It was generated 2,000,000 draws in two chains with 1,000,000 replications each, 90 % of replications were discarded so as to avoid influence of initial conditions. MCMC diagnostics were used for verification of the algorithm. All computations were carried out using Dynare toolbox (Adjemian et al. [1]) in Matlab software.

4 Results of estimation

This section discusses results of estimation and studies behavior of the model. Table 1 shows prior means, standard deviations and posterior means together with 95 % confidence intervals of selected estimated parameters. Rest of the estimated parameters can be found in [7].

	Prior distribution			Posterior distribution		
Parameter	Density	Mean	S.D.	Mean	2.5 %	97.5 %
Habit formation						
ε	beta	0.50	0.08	0.40	0.31	0.50
ε'	beta	0.50	0.08	0.53	0.40	0.66
Labor income share						
α	beta	0.65	0.05	0.71	0.63	0.78
Calvo parameters						
θ_π	beta	0.67	0.05	0.72	0.67	0.78
$\theta_{w,c}$	beta	0.67	0.05	0.75	0.70	0.79
$\theta_{w,h}$	beta	0.67	0.05	0.69	0.62	0.75
Taylor rule						
r_R	beta	0.75	0.10	0.91	0.89	0.93
r_π	normal	1.50	0.10	1.39	1.23	1.56
r_Y	normal	0.00	0.10	0.22	0.08	0.36
Technology growth rates						
$100\gamma_{AC}$	normal	0.50	1.00	0.42	0.37	0.47
$100\gamma_{AH}$	normal	0.50	1.00	-0.61	-1.00	-0.24
$100\gamma_{AK}$	normal	0.50	1.00	0.13	0.08	0.19

Table 1 Prior and posterior distribution of structural parameters

Parameters ε and ε' express habit consumption of patient and impatient households. The posterior mean of ε is lower than prior and the posterior mean of ε' is slightly higher than prior. Altogether, it indicates quite weak habit in consumption and does not correspond to values usually used for the Czech economy. The labor income share of constrained households $(1-\alpha)$ was estimated to 0.29. This is slightly higher than values found in empirical studies for U.S. economy (0.21) or Sweden (0.18); see Iacoviello and Neri [4] and Wallentin and Sellin [9]. Much higher estimate was obtained by Hloušek [6] for the Czech economy (0.55) and Christensen et al. [8] for Canada (0.38). However, these last two papers used different model structure. Estimated values of Calvo parameters indicate that price and wage rigidities are almost equally important. This is in contrast to empirical studies which found that wages are more rigid than prices; see e.g. Hloušek and Vašíček [5] or Andrle et.al. [2]. The reason can be again different sector structure of the model. Parameters in Taylor rule show that the Czech National Bank pays large attention to interest rate smoothing and also output gap. On the other hand, posterior mean of parameter of inflation r_π is slightly lower than mean of the prior which is usually used in calibrated models.

Estimated parameters γ's together with several model equations[2] can be used for computation of trends in the model variables. The quarterly growth rates for consumption, business and residential investment and real house prices are respectively 0.49, 0.62, -0.33 and 0.82. According to the model, the steep trend in house prices was mainly caused by negative technological progress in the housing sector.

Next step is evaluation of data fit of the model. Table 2 shows moments calculated from data and moments obtained from model simulations (with 90 % probability intervals).[3] Data for the real variables were linearly detrended so that it corresponds to treatment of the variables in the model. The outcome of the model is very good. The volatility of the variables is matched quite precisely, all volatilities fall into probability bands. The model has only minor problems to match high volatility of real house prices q. It is understandable because boom and bust in house prices during recent years was unusual. Regarding correlations the data fit is again very good. The correlations are within the probability intervals although by narrow margin in some cases. The model produces higher correlation between residential investment and output (IH, Y) and residential investment and house prices (q, IH) than the data. On the other hand, correlations of house prices with output and consumption $(q,Y$ and $q,C)$ are lower in the model than in the data. Autocorrelations from model simulation and from estimated VAR(1) model are compared in Figure 2 (again with 90 % probability intervals). The model outcome is quite satisfactory, only

[2] Equations (12) to (15) in Iacoviello and Neri [4].

[3] The model definition of output Y is the sum of consumption and business and residential investment. Corresponding definition is used for the data.

	Data	Model				Data	Model		
		Mean	5 %	95 %			Mean	5 %	95 %
Volatility					Correlations				
C	3.29	4.22	2.70	6.79	C,Y	0.94	0.89	0.73	0.96
IH	11.31	14.63	9.34	20.88	IH,Y	0.03	0.40	−0.14	0.80
IK	8.06	8.00	5.61	11.40	IK,Y	0.95	0.93	0.89	0.97
q	13.49	8.46	4.83	13.52	q,Y	0.86	0.50	−0.16	0.87
π	1.01	1.20	0.98	1.41	q,C	0.84	0.37	−0.35	0.85
R	0.29	0.38	0.23	0.65	q,IH	−0.15	0.40	−0.29	0.83
Y	4.27	5.07	3.30	7.59	q,π	−0.03	0.01	−0.37	0.47

Table 2 Moments from data and model

autocorrelations of consumption up to three lags are much higher in the model than in the data. Also empirical autocorrelations of interest rate and of inflation for one lag lies at margin of lower probability interval.

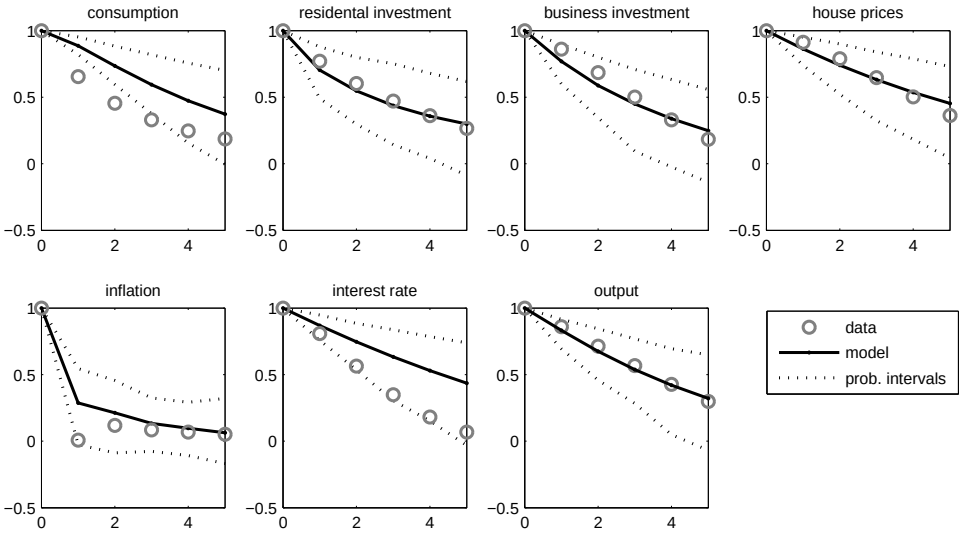


Figure 2 Autocorrelations

The dynamical properties of the model are further studied using impulse responses, variance decomposition and shock decomposition. Figure 3 depicts reaction of model variables to monetary policy shock (increase by one percentage point) for three versions of the model. In all three specifications the estimated parameters are kept at their posterior means. In the benchmark model the loan-to-value ratio (parameter m) is calibrated to value 0.75, in the "high collateral" model the value of LTV is 0.95 which means that constrained households have better position to get loan and in the "no collateral" specification the LTV is set to 0.001 which means that houses are not collateralizable and impatient households are excluded from financial markets. Reactions of the model variables for all three specifications are qualitatively same but they differ in its magnitude, especially for consumption and output. Higher LTV causes larger drop in consumption and output by 4.6 and 2.1 percentage points relative to the benchmark. On the other hand, impact for behavior of inflation is very similar across specifications.[4] It indicates that disinflation policy is more costly when collateral effect is present and is high.

Table 3 shows variance of the model variables explained by each shock. The results are quite intuitive. Consumption technology shocks explain most of the volatility of consumption and housing technology and preference shocks are important for behavior of residential investment (IH) and housing prices (q). Inflation target shocks are mainly responsible for variance of interest rate (R) and together with cost-push shocks also for variance of inflation (π). On the other hand, investment technology shocks are unimportant even for business investment (IK). Output is thus driven primarily by consumption technology shocks.

Finally, Figure 4 depicts historical decomposition of the real house prices into shocks during the estimated period. This figure shows that housing preference shocks became more important since the end of 2001; from

[4]The difference in drop of inflation between specification with high LTV ratio and the benchmark was 0.8 percentage points.

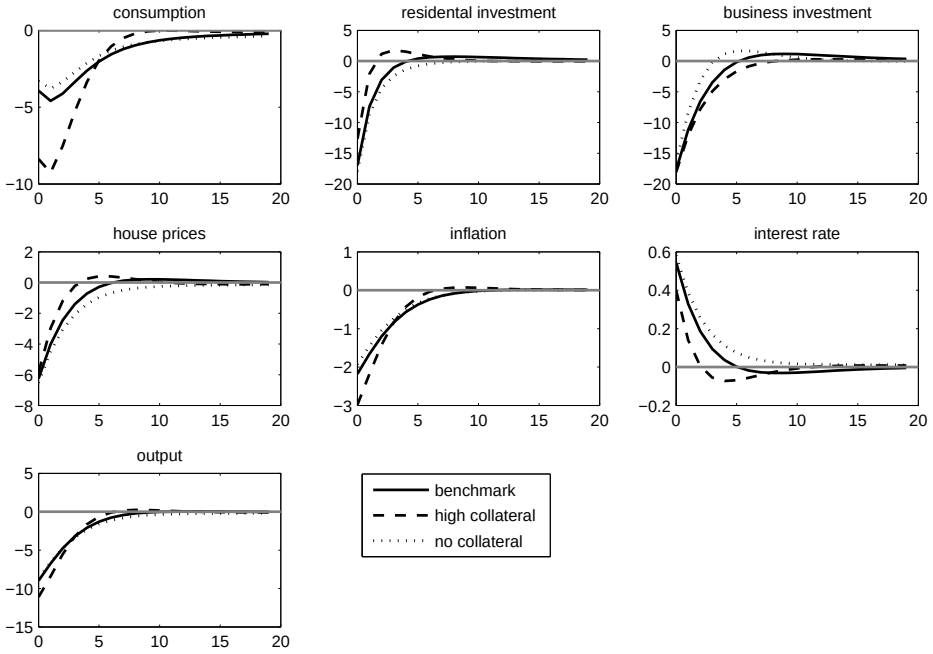


Figure 3 Impulse responses to monetary policy shock

	cons. tech.	monet.	housing tech.	housing pref.	invest. tech.	cost- -push	infl. targ.	labor supply	inter- temp.
C	79.4	4.0	0.1	0.1	0.0	1.6	3.3	4.7	6.8
IH	0.5	0.8	67.9	27.8	0.0	0.1	0.5	2.3	0.3
IK	65.6	9.8	0.6	0.5	0.2	7.3	7.2	6.6	2.3
q	8.5	0.4	38.3	51.6	0.0	0.3	0.3	0.4	0.2
π	8.2	11.5	0.1	0.2	0.0	37.0	38.2	2.5	2.4
R	14.4	2.4	0.1	0.7	0.0	2.0	76.5	1.9	2.0
Y	67.7	5.9	7.1	3.9	0.0	3.1	4.6	6.7	1.1

Table 3 Variance decomposition

this year onwards it was the main determinant of rising prices of houses. The same shock was responsible for subsequent decline during and after the crises. Housing technology shocks also contributed to the development of house prices but in more stable manner. Consumption technology shocks also increased their importance in explaining house prices behavior mainly from 2002. After the peak in 2008 consumption shocks diminish together with decline of house prices. This analysis shows that both demand and supply shocks played important role but mostly demand shocks were responsible for wild behavior of house prices.

5 Conclusion

This paper presented results of estimation of medium-scale DSGE model with housing sector on Czech data. The model fits the data in many aspects quite successfully. Detailed analysis of the model dynamics revealed which shocks explain behavior of the key macroeconomic variables. Consumption shocks, housing technology and housing preference shocks turned out as important shocks for development of the real variables in corresponding sectors. Nominal variables were primarily influenced by inflation target and cost-push shocks. If we look at behavior of house prices, shocks to housing preferences were the main driving force, especially during turbulent period in last ten years. Consumption and housing technology shocks also contributed but they were more stable. Looking at single monetary shock, reaction of consumption and output is much more pronounced if loans to constrained households are more accessible. On the other hand, impact on inflation and other variables can be considered as negligible. These results should be taken into consideration for formation of monetary policy.

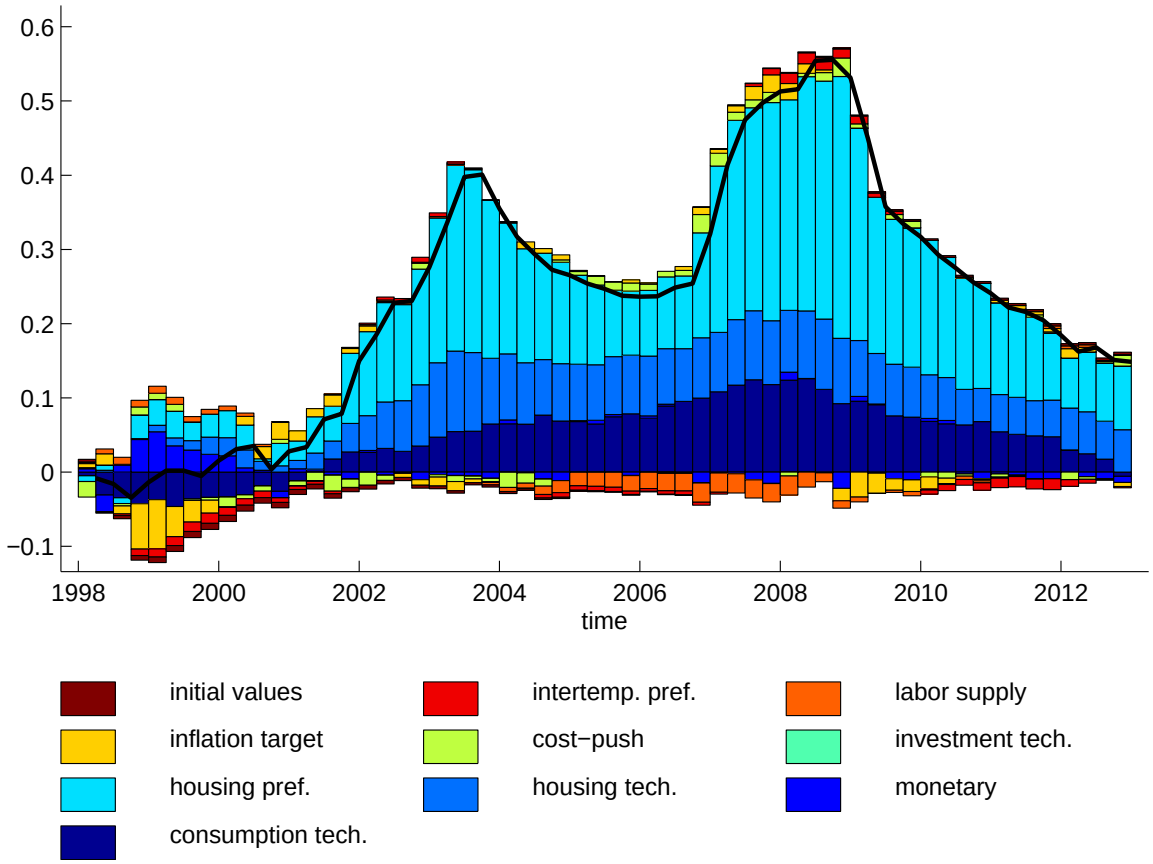


Figure 4 Shock decomposition of housing prices

References

[1] Adjemian, S., Bastani, H., Juillard, M., Mihoubi, F., Perendia, G., Ratto, M. and Villemot S.: Dynare: Reference Manual, Version 4, *Dynare Working Papers,* 1, Cepremap, 2011.

[2] Andrle, M., Hlédik, T., Kameník, O. and Vlček, J.: Implementing the New Structural Model of the Czech National Bank, CNB: *WP series* 2, 2009.

[3] Calvo, G.: Staggered prices in a utility-maximizing framework, *Journal of Monetary Economics 12*, 383-398, 1983.

[4] Iacoviello, M. and Neri, S.: Housing Market Spillovers: Evidence from an Estimated DSGE Model, *American Economic Journal: Macroeconomics*, Vol. 2, No. 2, pp. 125-164, 2010.

[5] Hloušek, M. and Vašíček, O.: Sensitivity Analysis of DSGE Model of Open Economy with Nominal Rigidities. In *Mathematical Methods in Economics*, Liberec: Technická univerzita v Liberci, pp. 192-197, 2008.

[6] Hloušek, M. DSGE model with collateral constraint: estimation on Czech data, In Ramík, J., Stavárek, D. *Proceedings of 30th International Conference Mathematical Methods in Economics.* Karviná: Silesian University, School of Business Administration, pp. 296-301, 2012.

[7] Hloušek, M. DSGE model with housing market: Application to the Czech economy, Appendix, *mimeo*, 2013. Available at: `www.econ.muni.cz/~hlousek/papers/housing_appendix.pdf`.

[8] Christensen, I., Corrigan, P., Mendicino, C. and Nishiyama, S.: Consumption, Housing Collateral, and the Canadian Business Cycle, Bank of Canada, *Working paper*, 2009-26, 2009.

[9] Walentin, K. and Sellin, P.: Housing collateral and the monetary transmission mechanism, Sveriges Riksbank: *Working paper series 239*, 2010.

Measurement of operational complexity of supplier-customer system using entropy – a case study

Jiří Hofman [1], Ladislav Lukáš [2]

Abstract. The paper concerns with analysis of operational complexity of company supplier-customer relations. Well-known approach for measuring that operational complexity is based upon entropy. However, there are several approaches thereon. In the first part, we discuss various general measures of uncertainty of states, the power entropies in particular. In the second part, we use Shannon entropy as a base framework for the case study – supplier system of the most important commodity in brewery industry, the malted barley. We assume a problem-oriented database exists, which contains detailed records of all product orders, deliveries and forecasts both in quantity and time being scheduled and realized, within time period given. We concern ourselves with quantity perturbations since they play more crucial role as the time variations. The general procedure elaborated consists of three basic steps – preprocessing of data with consistency checks in Java, calculation of histograms and empirical distribution functions, and finally, evaluation of conditional entropy. These two last steps are realized by Mathematica modules. Illustrative results of operational complexity of malted barley supplier system focused on volume quantity variations are presented in detail.

Keywords: business economics, supplier-customer systems, firm performance, complexity measures, information and entropy, power entropy.

JEL Classification: C63, C81, L25, M21
AMS Classification: 91B42

1 Introduction

Business economics knows two types of complexity of supplier-customer systems, a structural complexity and an operational one, in principle. As usual, the structural complexity is defined as static variety of system and their design dimensions, and it describes links among various business units and their hierarchies. It has dominantly a static representation and undergoes time changes usually in long-term periods.

On the contrary, the operational complexity can be defined by uncertainties associated with dynamics of system. Hence, it reflects temporal changes in supplier-customer system, and an operational complexity measure should express behavioral uncertainties of the system during the time with respect to specified levels of control. We know that operational complexity of supplier-customer system is associated with specific data provided by inventory management. It has to record all possible types of flow variations within and across companies in detail, e.g. replenishment time disturbances, deviations of material in/out flows, etc. We refer to [3] for more details relating inventory management.

2 Theoretical background

The theoretical framework for quantification of any system complexity is provided by information theory, in general. We may refer [1] and [4] for more details about Shannon information-theoretic measure and corresponding entropy, and [5] for some generalizations, power information in particular. In [2], we can find more elaborated derivation of classic entropy, which represents the most known quantitative measure of expected amount of information required to describe the state of a system. In general, the complexity of a system increases with increasing levels of disorder and uncertainty of its states.

In [5], we can find some motivating ideas. First of all, the generalization of Shannon entropy which is called power information there, and we shall call it *α-power entropy* $H_\alpha(\pi)$, for any $\alpha>0$. It is given by expression (1).

[1] Univ. of West Bohemia, Faculty of Econ., Dept. of Business Admin. and Mngmt, Husova 11, 30614 Plzeň, hofman@kpm.zcu.cz.
[2] Univ. of West Bohemia, Faculty of Econ., Dept. of Econ. and Quantitative Methods, Husova 11, 30614 Plzeň, lukasl@kem.zcu.cz.

$$H_\alpha(\pi)=\sum_{i=1}^{N} \phi_\alpha(p_i)\,,\ \alpha>0,\ \alpha\neq 1 \quad , \quad H_1(\pi)=\sum_{i=1}^{N} \phi_1(p_i)\,,\ \alpha=1 \tag{1}$$

Where $H_1(\pi)$ is the Shannon entropy (here measured in *nats* instead of *bits*, since the binary logarithm $\log_2$ is replaced by natural one). Function $\phi_\alpha(t)$, $0 \le t \le 1$ is so called α-power information function, which is strict concave one with boundary values $\phi_\alpha(0)=\phi_\alpha(1)=0$. Finally, π stands for probability distribution of system, which states are completely described by mutually disjoint events $\{A_1,...,A_N\}$ with probabilities $(p_1,...,p_N)$. Functions $\phi_\alpha(t)$ and $\phi_1(t)$ are given by following expressions.

$$\phi_\alpha(t) = [t(1-t^{\alpha-1})]/(\alpha-1),\ \alpha>0,\ \alpha\neq 1 \quad , \quad \phi_1(t) = \lim_{\alpha\to 1}\phi_\alpha(t) = -t\ln t,\ \alpha=1 \tag{2}$$

Figure 1 shows some graphs of $\phi_\alpha(t)$, for α=0.5, 1, 2, 3, 4, which may be identified in a consequent way being depicted by plain line, thick one, dashed one, dotted one, and plain one, as well.

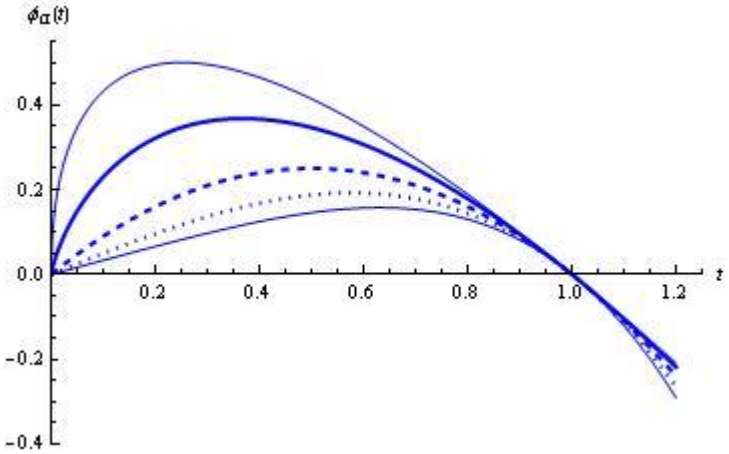


Figure 1 Power information functions, α = 0.5, 1, 2, 3, 4

The strict concavity and zero boundary conditions are preserved by passing to the adjoint α-power information functions $\psi_\alpha(t) = \phi_\alpha(1-t)$, which are plotted on Figure 2 for the same values α=0.5, 1, 2, 3, 4.

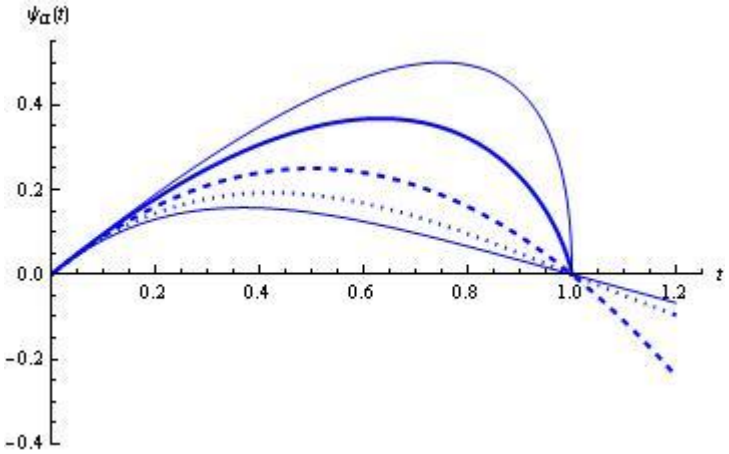


Figure 2 Adjoint power information functions, α = 0.5, 1, 2, 3, 4

As mentioned above, the complexity of a system should intuitively increase with increasing level of disorder and uncertainty of its states, and we know that $H_1(\pi)$ complies therewith. However, such characteristic property is maintained by α-power entropy $H_\alpha(\pi)$, $\alpha > 0$, as well, which is known as the information preservation law:

$$0 = H_\alpha(\pi_D) \le H_\alpha(\pi) \le H_\alpha(\pi_U) = (1-N^{1-\alpha})/(\alpha-1)\,,\ \alpha>0 \tag{3}$$

where symbols π_D and π_U stand for the singular (Dirac) and finite uniform probability distributions of system states, and $H_1(\pi_U)=\ln N$, in particular. The classic entropy, i.e. original Shannon entropy, of π_U is $H(\pi_U)=\log_2 N$, and in the sequel, we denote it simply $I_U = H(\pi_U)=\log_2 N$, too, being already measured in *bits*.

In a similar way to quantity $H_\alpha(\pi)$, one defines an *adjoint α-power entropy* $G_\alpha(\pi)$, $\alpha > 0$ by adopting adjoint α-power information functions $\psi_\alpha(t) = \phi_\alpha(1–t)$.

$$G_\alpha(\pi) = \sum_{i=1}^{N} \psi_\alpha(p_i) = \sum_{i=1}^{N} \phi_\alpha(1-p_i),\ \alpha > 0,\ \alpha \neq 1\ , \quad G_1(\pi) = \sum_{i=1}^{N} \psi_1(p_i) = \sum_{i=1}^{N} \phi_1(1-p_i),\ \alpha = 1 \tag{4}$$

Using (4) we can calculate formulas (5) and (6) for $G_\alpha(\pi_U)$ and $G_1(\pi_U)$, respectively, provided that $p_i = 1/N$, i=1,..,N. We use symbolic calculation power of Mathematica thereon, in particular

```
Sum[(1-1/N)*(1-(1-1/N)^(α-1)),{n,1,N}]//FullSimplify
Limit[(N-1-N*((N-1)/N)^α)/(α-1),α→1]//FullSimplify
```

$$\begin{aligned} G_\alpha(\pi_U) &= \sum_{i=1}^{N} \psi_\alpha(1/N) = \sum_{i=1}^{N} \phi_\alpha(1-1/N) = \left(\sum_{i=1}^{N} (1-1/N)(1-(1-1/N)^{\alpha-1})\right)/(\alpha-1) \\ &= (N-1-N)((N-1)/N)^{\alpha})/(\alpha-1)\ ,\ \alpha > 0,\ \alpha \neq 1 \end{aligned} \tag{5}$$

$$G_1(\pi_U) = \lim_{\alpha\to 1} G_\alpha(\pi_U) = \lim_{\alpha\to 1} (N-1-N)((N-1)/N)^{\alpha})/(\alpha-1) = (N-1)\ln(N/(N-1)),\ \alpha = 1 \tag{6}$$

Using (3), (5), Figures 3 and 4 show plots of values $H_\alpha(\pi_U)$ and $G_\alpha(\pi_U)$ when parameter α ranges [0.1,4.0] and for various number of system states N = 10, 50, 100, 1000. These functions are simply identifiable by plain line, thick one, dashed one, and dotted one, respectively.

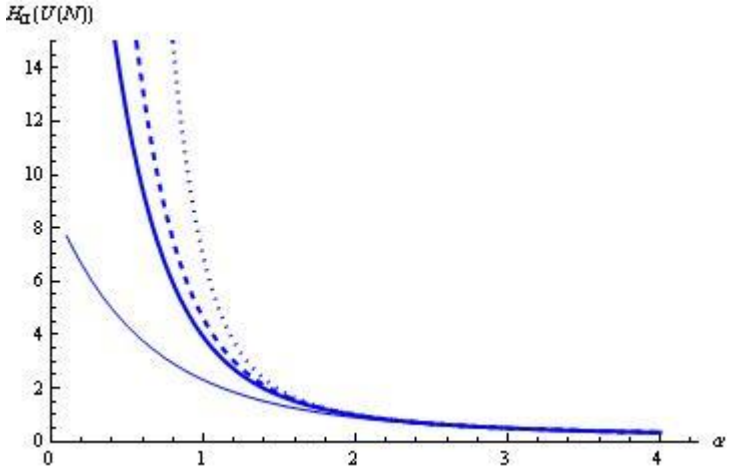


Figure 3 Power entropies of discrete uniform distributions π_U, for N = 10, 50, 100, 1000

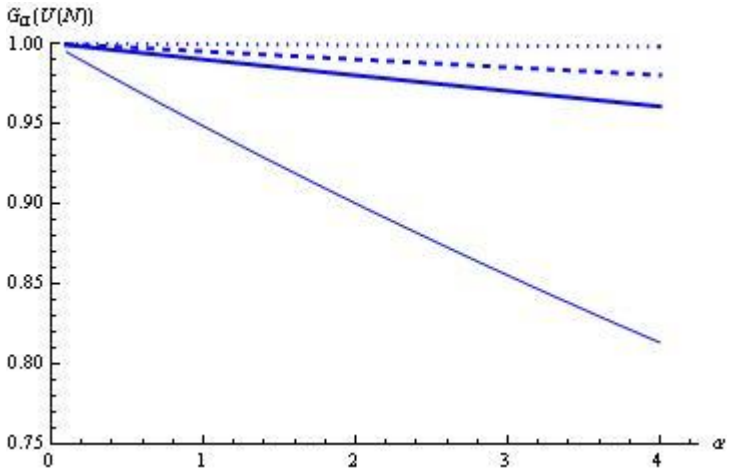


Figure 4 Adjoint power entropies of discrete uniform distributions π_U, for N = 10, 50, 100, 1000

Using Mathematica, we have also calculated several values of $H_\alpha(\pi_U)$ and $G_\alpha(\pi_U)$ for different values of parameter α and small sample space having just N = 10 different system states, see the Table 1. It shows also an

interesting property that $H_2(\pi_U) = G_2(\pi_U)$, which holds for any N. In this case the $H_2(\pi)$ is known as the *quadratic entropy*, see [5]. We can also inspect that functions $\phi_2(t)$ and $\psi_2(t)$ are equal each other, i.e. $\phi_2(t) = \psi_2(t) = t(1-t)$.

α	**0.50**	**0.99**	**1**	**1.01**	**2**	**3**	**4**
$H_\alpha(\pi_U)$	4.32456	2.3293	2.30259	2.27628	0.9	0.495	0.333
$G_\alpha(\pi_U)$	0.973666	0.948744	0.948245	0.947745	0.9	0.855	0.813

Table 1 Selected values of $H_\alpha(\pi_U)$ and $G_\alpha(\pi_U)$, for N=10

3 Operational complexity of supplier-customer system – case study

We refer to [2] and [7] for general framework and more details relating building problem-oriented database and construction of flow variation quantities denoted in general ${}_{(e,r)}Q_k$, and ${}_{(e,r)}T_k$, k=1,...,n, for set of n different products $\{P_1, ..., P_n\}$ being considered. Prefix indices e and r stand for entity and production phase, respectively, (e equals to s for supplier, i for interface, and c for customer, and r equals to s for scheduled, p for actual production, and f, o, d for forecast, order and delivery, in particular).

Our case study is focused on measuring operational complexity of malted barley supplies into known brewery in Pilsen during period 2008Q1-2011Q1. In [4], there are collected the corresponding data and building of problem-oriented database consisting of 20 main suppliers. Since we concern ourselves with volumetric variations between scheduled Q_s and actually produced and delivered Q_p volumes exclusively, we denote this quantity $\Delta Q_{p,s} = Q_p - Q_s$, and we tackle it as a random variable with empirical distribution for each supplier Di, i=1,..,20.

At present, we have two programs for operational complexity analysis. First, `EnComP1mma.java` checks consistency of input data files and generates outputs, which are fetched into the Mathematica notebook `EnComP2mma.nb` for numerical calculations and generation of graphical outputs, as well. The present version implements calculation of entropy and other related quantities by formulas (7).

$$H(\pi) = -\sum_{i=1}^{N} p_i \log_2(p_i) ,$$
$$I_u = H(\pi_U) = -\sum_{i=1}^{N} (1/N) \log_2(1/N) = \log_2(N) , \qquad h(\pi) = H(\pi)/I_u \tag{7}$$

For illustration, we present typical outputs for the suppliers D_3 and D_5 each, which are two largest ones. Figures 5 and 6 show plots of their malted barley delivery variations $\Delta Q_{p,s}$ during 2008Q1-2011Q1 and the corresponding empirical distribution functions with resized definition interval [0,1.1] because of comparison purposes with other ones in general.

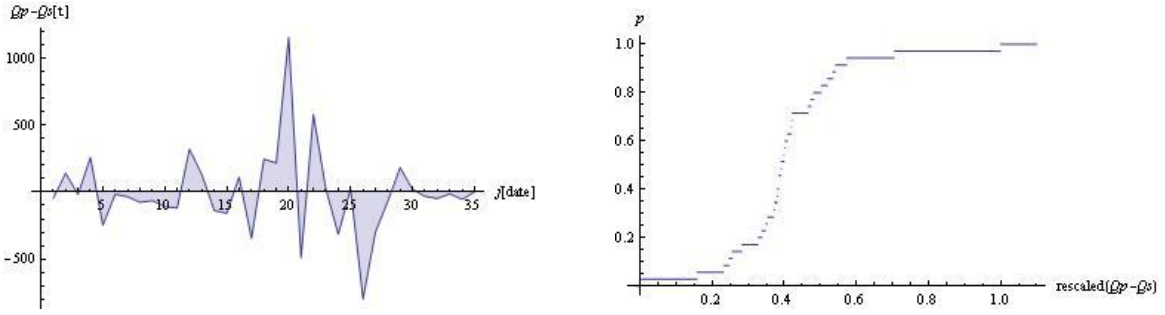


Figure 5 Supplier D_3: Q_p-Q_s – left: delivery variation $\Delta Q_{p,s}$ – right: empirical distribution function

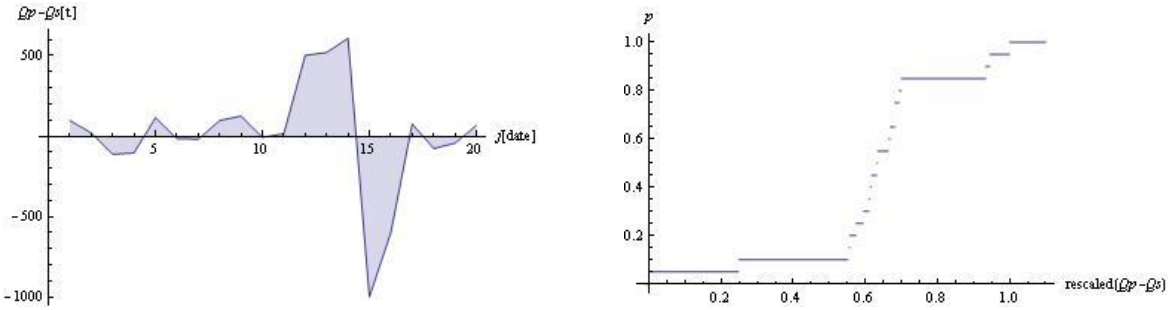


Figure 6 Supplier D_5: Q_p-Q_s – left: delivery variation $\Delta Q_{p,s}$ – right: empirical distribution function

We have selected five major suppliers of malted barley (D_1, D_3, D_5, D_{13}, D_{18}) during 2008Q1-2011Q1 period in particular, which are scheduled and have been ordered to supply more than 58 % of the total quantity desired by the brewery, and which delivered almost 60 % of the total quantity received. The calculated entropies and entropy ratios for this subset of suppliers using formulas (7) are depicted on the Table 2.

supplier	**D_1**	**D_3**	**D_5**	**D_{13}**	**D_{18}**
$H(\pi)$	4.48386	5.015	4.12193	4.31582	4.1066
$h(\pi)=H(\pi)/I_U$	0.965546	0.977719	0.953724	0.918175	0.934951

Table 2 Entropies $H(\pi)$ and entropy ratios $h(\pi)$ for major suppliers (D_1, D_3, D_5, D_{13}, D_{18}) in 2008Q1-2011Q1

The rather high values of entropy ratios $h(\pi)$ express large variability of deliveries, i.e. uncertainties in supplies of such important commodity of any brewery as malted barley. However, relatively small number of particular deliveries of each supplier could also cause such warning values by low size of samples for construction of empirical distributions. Sure, and last but not at least, there is a fact, that top quality malted barley as an important agricultural commodity depends upon seasonal weather, as well, which may support relatively high number of suppliers on another side. Hence, we summarized deliveries of all twenty suppliers D*i*, *i*=1,..,20, in order to analyze their particular delivery contributions.

Figure 7 show plots of total delivery volumes scheduled Q_s and their excesses Q_p-Q_s in [t] of all suppliers D*i*, *i*=1,..,20 in the period investigated, where Q_p denotes volumes received.

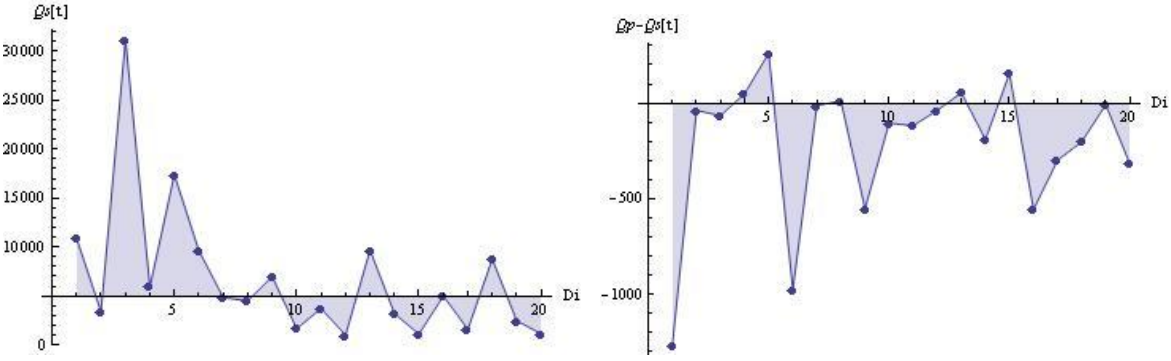


Figure 7 Suppliers D_1-D_{20}: – left: Q_s total volumes scheduled, – right: $\Delta Q_{p,s} = Q_p$-Q_s delivery variations

First, we can clearly confirm an identification of malted barley major suppliers (D_1, D_3, D_5, D_{13}, D_{18}), and further, we directly inspect large variability of $\Delta Q_{p,s}$ if being tackled as a global random variable describing uncertainty of delivery that commodity from the brewery management point of view.

Making total balances of malted barley both scheduled & ordered Q_s and received Q_p by brewery from all deliveries of all suppliers during 2008Q1-2011Q1, we can conclude that only 96.87 % of the requested volume was really delivered, which means that brewery faces a global demand excess 3.13 % of global scheduled & ordered quantity. The global results together with results of major suppliers are summarized on the Table 3 below.

Balance [t]	**Ω_s**	**%**	**Ω_p**	**%**	**$\Delta = \Omega_p - \Omega_s$**	**%**
Sum D*i*, *i*=1,..,20	134 200	100.00	129 998	100.00	–4 202	100.00
$D_1+D_3+D_5+D_{13}+D_{18}$	77 990	58.11	76 776	59.06	–1 214	28.89

Table 3 Malted barley: total volume scheduled Ω_s, received Ω_p, and excess $\Delta = \Omega_p - \Omega_s$ in 2008Q1-2011Q1

Finally, Figure 8 shows relative variations $(Q_p\text{-}Q_s)/Q_s$ given in percents for all suppliers together with 8 % tolerance bands, which are usually applied by the brewery management when making negotiation and settlement of particular delivery agreement with each supplier. We can directly inspect also five out-of-tolerance suppliers during the investigated period: D_1, D_6, D_{16}, D_{17}, and D_{20}.

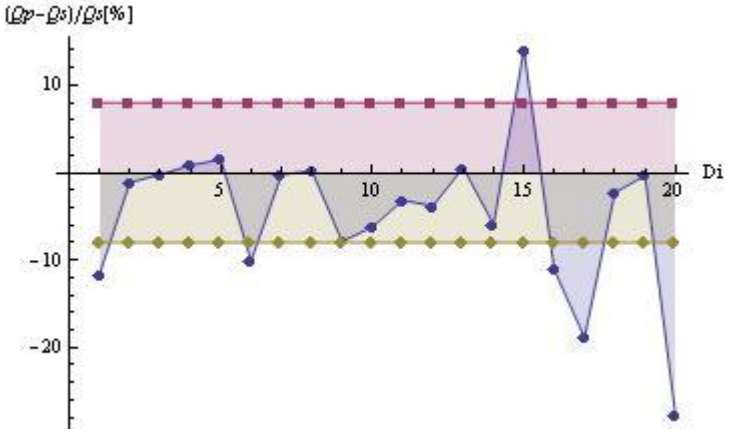


Figure 8 $(Q_p\text{-}Q_s)/Q_s$ [%] total volumes scheduled, suppliers D*i*, *i*=1,..,20

4 Conclusion

The measure of operational complexity based upon entropy provides versatile instrument for supplier-customer system analysis. We have presented not only well-known Shannon entropy, but also a more general approach based upon α-power entropy, which as we hope can open new fields of applications, e.g. case dependent quantitative measuring of various flow variations of as volume as term oriented quantities in supplier-customer systems. In the case study presented, we have analyzed uncertainty of malted barley deliveries into brewery and some aspects related therewith. We focused our research primarily to volume variations of deliveries, since they play a dominant role from the brewery management point of view. In the on-going research, we will try first, to couple volume and term variations together in order to get complex quantitative characteristics, and second, to implement the α-power entropy in our Mathematica notebooks in order to enrich our tools for managerial analysis and decision-making.

Acknowledgements

The research project was supported by Grant SGS2012-036 of the Faculty of Economics, University of West Bohemia in Pilsen.

References

[1] Feder, M., and Merhav, N.: Relations between entropy and error probability. *IEEE Trans. Inform. Theory* **40** (1994), 259-266.

[2] Hofman, J., and Lukáš, L.: Quantitative measuring of operational complexity of supplier-customer system with control thresholds. In: *Proceedings of 30th Int. Conf. Mathematical Methods in Economics* (Ramík, J., and Stavárek, D., eds.). Karviná: Silesian Univ., School of Business Administration, 2012, pp. 302-308, ISBN 978-80-7248-779-0.

[3] Lukáš, L.: *Pravděpodobnostní modely v managementu Teorie zásob a statistický popis poptávky* (*Probabilistic models in management Inventory theory and statistical description of demand*, in czech). Academia, Prague, 2012, ISBN 978-80-200-2005-5.

[4] Malá, A.: Měření složitosti dodavatelsko-odběratelských vztahů pomocí kvantitativních měr v konkrétním podniku (Measuring of supplier-customer relations complexity using quantitative measures in specific company, in czech) Plzeň: ZČU/FEK, Bachelor thesis, 2013, pp.79.

[5] Morales, D., and Vajda, I.: Generalized information criteria for Bayes decisions. *Kybernetika* **48** (2012), 714–749.

[6] Prochorov, Ju. B., Rozanov, Ju., A.: *Teorija verojatnostej* (*Theory of probability*, in russian). Nauka, Moskva 1967.

[7] Sivadasan, S., Efstathiou J., Calinescu A., Huaccho Huatuco L.: Advances on measuring the operational complexity of supplier-customer systems. EJOR **171** (2006), pp. 208-226.

Imprecise input data and option valuation problem

Michal Holčapek[1], Tomáš Tichý[2]

Abstract. During last decades the stochastic simulation approach, both via MC and QMC has been vastly applied and subsequently analyzed in almost all branches of science. Very nice applications can be found in areas that rely on modeling via stochastic processes, such as finance. However, as for any other approach, the most crucial step is feeding the assumed model with data and estimation of model parameters. It is a matter of fact that financial processes are instable in time and often switch their regimes. Several scholars therefor suggest to specify some parts of financial models by means of fuzzy set theory. In this contribution the recent knowledge of fuzzy numbers and their approximation is utilized in order to suggest fuzzy-MC simulation to option price modeling in terms of fuzzy-random variables. In particular, we suggest three distinct fuzzy-random processes as an alternative to a standard crisp model and show application possibilities of one of them on illustrative example.

Keywords: Fuzzy numbers, fuzzy random variable, option, simulation

JEL classification: C46, E37, G17, G24
AMS classification: 90C15

1 Introduction

Options, a specific nonlinear type of a financial derivative, play an important role in the economy. In particular, the usage of options allows one to reach a higher level of efficiency in terms of risk-return trade-off, whether through speculation or hedging strategy. The options holder can exercise his right, eg. buy or sell an underlying asset, when he finds it useful. Obviously, it is the case of positive cash flowing from the option exercising. Otherwise the option matures unexploited. By contrast, the seller of the option has to act according to the instructions of the holder. This asymmetry of buyer/seller rights implies the needs of advanced technique for option pricing and hedging. Denoting the underlying asset price at maturity time as S_T we can write the payoff function for European call and put options as $\Psi_T^{(vanilla\ call)} = (S_T - K)^+$ and $\Psi_T^{(vanilla\ put)} = (K - S_T)^+$, respectively.

For example, for the call option price f_t at time $t < T$ it generally holds that $f_t = e^{d_\tau} E[(S_T - K)^+]$, where a discount factor d_τ relates to the probability measure under which the expectation operator E is evaluated and $\tau = T - t$ is the remaining time to maturity. Commonly, E^P denotes the real world expectation (under physical probability measure), while E^Q is used within the risk-neutral world, i.e., where r is a riskless rate valid over time interval τ. Since financial asset prices are often restricted to positive values only, geometric processes are commonly preferred. If, for example, $Z(t)$ denotes a stochastic process for log-returns of financial asset S, e.g., a non-dividend paying stock, to model its price in time we have to evaluate the exponential function of $Z(t)$. It follows that under Q the key formula above can be rewritten into:

$$f_t = e^{-r\tau} E^Q[(S_T e^{r\tau + Z_\tau^Q} - K)^+],$$

where Z_τ^Q is a (potentially compensated) realization of a suitable stochastic process over ? such that it is ensured that S is a martingale.

Actually, the assumption about Z determines suitable approaches for option pricing. The standard ways to option pricing, as well as replication and hedging, dates back to 70s to the seminal papers of Black

[1]Centre of Excellence IT4Innovations, division of University of Ostrava, 30. dubna 22, 701 03 Ostrava 1, Czech Republic, michal.holcapek@osu.cz
[2]Technical University Ostrava, Faculty of Economics, Department of Finance, Sokolsk 33, 701 21 Ostrava, Czech Republic, tomas.tichy@osu.cz

and Scholes [2] and Merton [11] or Boyle [3]. While Black and Scholes [2] and Merton [11](1973) derived their respective models within continuous time by solving partial differential equations (and thereafter called Black-Scholes-Merton partial differential equations) for risk free portfolio consisting of option itself and its underlying asset, Boyle [3] suggested that in order to obtain the (discounted) expectation of the option payoff function the Monte Carlo simulation technique can be useful, i.e., instead of riskless portfolio construction and utilization of no-arbitrage principle the risk-neutral world is assumed. It is a well known result of quantitative finance that these approaches are equivalent under the assumption of complete markets, or, at least, when an equivalent martingale measure exists.

Although the original approaches slightly differ in details the underlying process is derived from Gaussian distribution and all parameters are either deterministic or probabilistic, ie. particular probabilities are assigned to the set of real numbers. However, in the real world, it is often difficult to obtain reliable estimates to input parameters. The reason can be that sufficiently long time series of data is lacking or the data are too heterogenous. Several research papers collected by Ribeira et al. [13] suggested that the fuzzy set theory proposed by Zadeh [16] can be useful for financial engineering problems of such kind.

One of the first attempts to utilize the fuzzy set theory in option pricing dates back to Cherubini [4]. Further results were reviewed eg. by Holčapek and Tichý [8]. However, except recent, but brief contribution of Nowak and Romaniuk [12] there was no attempt to value an option with fuzzy parameters via Monte Carlo simulation approach.

In this paper, we try to fill the gap by suggesting three distinct types of potential underlying processes defined on the basis of fuzzy-random variables. More particularly, we assume (geometric) Brownian motion with fuzzy volatility and (geometric) Brownian motion with time t replaced either by fuzzy process or by fuzzified gamma subordinator, which allows us to redefine in finance well known and commonly used Lévy type variance gamma model [5] in terms of fuzzy-random subordinator. In the next section, we provide brief details about option pricing via plain Monte Carlo simulation within the risk neutral setting. Next, LU-fuzzy numbers and relevant operations with them are defined. After that, three potential candidates to option underlying asset price model are suggested. Finally, a European option price is evaluated assuming all three processes.

2 Cubic rational spline approximation of LU-fuzzy numbers

In the solving of practical problems it is useful to approximate fuzzy numbers by simpler functions. The main reason is that the definition of arithmetic with fuzzy numbers naturally based on Zadeh's extension principle (see [6]) is complex from the computation perspective. Among the basic approximations of fuzzy numbers there belong well known triangular or rectangular approximations (see, e.g., [1] and references therein). Note that the arithmetic with such kind of fuzzy numbers is computationally simple, but often does not correspond to the results obtained by Zadeh's extension principle. Therefore, in recent works on fuzzy numbers (more precisely, LU-fuzzy numbers that are expressed in terms of α-cuts systems), a very popular approach to the approximation of fuzzy numbers is based on rational splines (see [14, 15])) that helps to overcome these difficulties. The idea of approximation of fuzzy numbers using rational splines is based on an advanced system of parameters in contrast to three or four applied in the case of triangular or rectangular models, respectively. In the sequel we briefly recall the basic construction of parametrized LU- fuzzy numbers, for details, we refer to [7] (see also [14]).

Let $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_N = 1$ be real numbers for a finite decomposition of the unit interval. Considering the differentiable case, a parametrized LU-fuzzy number A is represented by the following system of vectors of parameters

$$A = (\alpha_i; u_i^-, \delta u_i^-, u_i^+, \delta u_i^+)_{i=0,\ldots,N}, \tag{1}$$

with the data $u_0^- \le u_1^- \le \cdots \le u_N^- \le u_N^+ \le u_{N-1}^+ \le \cdots \le u_0^+$ and the slopes $\delta u_i^- \ge 0$ and $\delta u_i^+ \le 0$, for all $i = 0, \ldots, N$. So, each parametrized fuzzy number is expressed as a finite system of intervals (α_i-cuts) with slopes defined for each endpoint. Note that the use of slopes is crucial here and enables us to complete the remaining α-cuts using the rational spline interpolation and to obtain a special case of LU-fuzzy number. It is easy to see that each LU-fuzzy numbers expressed as an infinite system of α-cuts can be successfully approximated by a finite system of α_i-cuts with suitable slopes for endpoints.

In this paper, we consider the rational cubic splines. Recall that *a piecewise rational cubic Hermite parametric function (spline) $P \in C^1[\alpha_0, \alpha_n]$ with parameters v_i, w_i, $i = 0, \ldots, n-1$* is defined for any

$\alpha \in [\alpha_i, \alpha_{i+1}]$, $i = 0, \ldots, n-1$, by

$$P(\alpha) = P_i(\alpha, v_i, w_i) = \frac{(1-\theta)^3 u_i + \theta(1-\theta)^2(v_i u_i + h_i \delta u_i) + \theta^2(1-\theta)(w_i u_{i+1} - h_i \delta u_{i+1}) + \theta^3 u_{i+1}}{(1-\theta)^3 + v_i \theta(1-\theta)^2 + w_i \theta^2(1-\theta) + \theta^3},$$

where the notations u_i and δu_i are, respectively, the data values and the first derivative values (slopes) at the nodes α_i, $i = 0, \ldots, n$ with $\alpha_0 < \cdots < \alpha_n$, $h_i = \alpha_{i+1} - \alpha_i$, $\theta = (\alpha - \alpha_i)/h_i$ and $v_i, w_i \geq 0$. Note that v_i and w_i are called tension parameters, and if $v_i = w_i$, we obtain the ordinary cubic spline. Here, we suppose

$$v_i = w_i = \begin{cases} \frac{\delta u_{i+1} + \delta u_i}{u_{i+1} - u_i}, & \text{for } u_{i+1} \neq u_i, \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

which guarantees the global monotonicity (see [14]). Further, we restrict ourselves to uniform partitions of the unit interval, i.e., $\alpha_{i+1} - \alpha_i$ is a constant for any $i = 0, \ldots, N-1$. Finally, we express the system (1) in terms of special matrices as follows

$$A = \begin{pmatrix} (u_0^-, \delta u_0^-) & \cdots & (u_N^-, \delta u_N^-) \\ (u_0^+, \delta u_0^+) & \cdots & (u_N^+, \delta u_N^+) \end{pmatrix}.$$

The matrix expression allows us to use some of the basic operations with matrices to introduce operations with fuzzy numbers. First, let us define the operations of addition and multiplication on the set $\mathbb{R}^2$ by

$$(u_1, u_2) \oplus (v_1, v_2) = (u_1 + v_1, u_2 + v_2) \quad \text{and} \quad (u_1, u_2) \otimes (v_1, v_2) = (u_1 v_1, u_1 v_2 + u_2 v_1),$$

where the standard operations of sum and multiplication of reals are used on the right side of the equality. Further, let us denote by $\leq$ the lexicographical ordering on $\mathbb{R}^2$. One could demonstrate that the algebraic structure $\mathbb{R}^2 = (\mathbb{R}^2, \oplus, \otimes, \leq)$ possesses nearly all properties of fields. The only difference is that $(\mathbb{R}^2 \setminus \{(0,0)\}, \oplus)$ does not form an abelian group, because there is no inversion for elements of type $(0, r)$.[1] Obviously, $\mathbb{R}$ can be embedded into $\mathbb{R}^2$ using $f(r) = (r, 0)$.

Thus, parametrized LU-fuzzy numbers can be naturally expressed using matrices over $\mathbb{R}^2$ in the same respect as over an ordered field. Now, we can define the addition and multiplication of parametrized LU-fuzzy numbers as follows (consider the i-th column):

$$(A \oplus B)_i = \begin{pmatrix} (u_i^-, \delta u_i^-) \oplus (v_i^-, \delta v_i^-) \\ (u_i^+, \delta u_i^+) \oplus (v_i^+, \delta v_i^+) \end{pmatrix}_i \quad \text{and} \quad (A \otimes B)_i = \begin{pmatrix} \min\limits_{p,q \in \{+,-\}} ((u_i^p, \delta u_i^p) \otimes (v_i^q, \delta v_i^q)) \\ \max\limits_{p,q \in \{+,-\}} (((u_i^p, \delta u_i^p) \otimes (v_i^q, \delta v_i^q)) \end{pmatrix}_i$$

where min and max are defined with respect to the lexicographic ordering $\leq$ on $\mathbb{P}$. The scalar multiplication (by reals) can be derived using the multiplication of matrices, where each real number r is interpreted as a matrix of the form

$$r = \begin{pmatrix} (r, 0), \ldots, (r, 0) \\ (r, 0), \ldots, (r, 0) \end{pmatrix}.$$

Hence, we simply obtain (consider the i-th column)

$$(r \otimes A)_i = \begin{pmatrix} (r u_i^{sgn(r)-}, r \delta u_i^{sgn(r)-}) \\ (r u_i^{sgn(r)+}, r \delta u_i^{sgn(r)+}) \end{pmatrix}_i$$

where $sgn(r) = +$, if $r \geq 0$, $sgn(r) = -$, otherwise, and the following rules are applied: $++ = -- = -$ and $+- = -+ = -$. It should be noted that all operation introduced above coincide with the operations defined in [14].

[1] Note that $(\mathbb{R}^2, \oplus, \otimes)$ is an MI-field, where MI-fields generalize the concept of field in the respect that a structure of so-called pseudoidentities is admitted. In the case of $\mathbb{R}^2$, the elements of type $(0, r)$ are pseudoindentities and $(\mathbb{R} \setminus E, \otimes)$, where $E = \{(0, r \mid r \in \mathbb{R})\}$, forms an abelian group. For details about MI-algebras, we refere to [9, 10].

3 Random parametrized LU-fuzzy numbers

To apply the Monte-Carlo techniques on problems with uncertainty modeled using parametrized fuzzy numbers, we need the concept of fuzzy random variable which values are expressed by parametrized LU-fuzzy numbers. For simplicity, such kind of fuzzy random variables will be called as random parametrized LU-fuzzy numbers. In [14], the authors proposed to defined random prametrized LU-fuzzy numbers in such a way that both the data and slopes are random variables satisfying all conditions stated for the parametrized LU-fuzzy numbers (see (1). Similarly, we can defined a random matrix (i.e., its values are random variables) with values expressed by parametrized LU-fuzzy numbers.

Definition 1. A random matrix

$$X = \begin{pmatrix} (X_0^-, \delta X_0^-) \cdots (X_N^-, \delta X_N^-) \\ (X_0^+, \delta X_0^+) \cdots (X_N^+, \delta X_N^+) \end{pmatrix} \tag{3}$$

where $X_0^- \leq X_1^- \leq \cdots \leq X_N^- \leq X_N^+ \leq X_{N-1}^+ \leq \cdots \leq X_0^+$ and $\delta X_i^- \geq 0$ and $\delta X_i^+ \leq 0$, $i = 0, \ldots, N$, is called a *random parametrized LU-fuzzy number*.

4 Potential candidates for price modelling

As we have already argued, it can be very difficult to obtain reliable estimates for the parameters (e.g., volatility or intensity of jumps) of the stochastic process $Z(t)$. It is the reason why many researchers suggest to define the underlying process in terms of fuzzy or fuzzy-random variables. In this section, three distinct fuzzy-random models are suggested as potential candidates to describe the option underlying asset price process; in particular, we assume (i) standard market model (Brownian motion) with fuzzy parameter, (ii) Brownian motion with fuzzy subordinator, and (iii) Brownian motion with fuzzified gamma subordinator.

Model 1 (standard market model with fuzzy parameter) Let σ_{LU} be an LU-fuzzy number defined around crisp estimation of σ. Then, we can model price returns by the following fuzzy-stochastic model:

$$Z(t) = \mu t + \sigma_{LU}\sqrt{t\varepsilon}.$$

Model 2 (Brownian motion with fuzzy subordinator) Let x_{LU} be a non-negative LU-fuzzy number centered around t so that it can be a subordinator. Then, we get the following alternative to the common assumption of Brownian motion:

$$Z(t) = \theta g(t) + \xi\sqrt{g(t)\varepsilon}.$$

Model 3 (Brownian motion with fuzzified gamma subordinator) Let g_{LU} be an LU-fuzzy number centered around a random gamma variable. Then, we can get another alternative model by using g_{LU} as a subordinator to the Brownian motion:

$$Z(t) = \theta x_{LU}(t) + \sigma\sqrt{x_{LU}(t)\varepsilon}.$$

5 Results

In order to evaluate the risk-neutral expectation via Monte Carlo simulation, we need to get models of the preceding section into the exponential and choose a proper ω_{LU} such that the complex process will be martingale when discounted by the riskless rate.

$$f_t = e^{-r\tau}E^Q\left[\left(S_T e^{r\tau+Z_\tau^Q} - K\right)^+\right] \approx \frac{e^{-r\tau}}{N}\sum_{i=1}^{N}\left(S_T e^{r\tau+Z_\tau^{Q^{(i)}}} - K\right)^+,$$

where the superscript (i) refers to the i-th scenario from a given probability space.

Comparative results of particular models for various input data are provided in Table 1. Let us assume put options written on stock price index in the form of a mutual fund price. For the illustrative example we derive the input data from the price observations of a Pioneer stock fund over 5 years.

Generally, we assume crisp values of initial price of the underlying asset ($S_0 = 100$), exercise price ($K = 100$), riskless rate ($r = 0$) and maturity ($T = 1$). Model 1 is similar to BS model, except that the volatility of underlying asset price returns is defined as a fuzzy random number over normal distribution $N(0.15; 0.1)$, with $s = 0.15$ being the most commonly observed value. For sensitivity reasons, we also consider $N(0.15; 0.05)$ and $N(0.20; 0.1)$. By contrast, Model 2 (first panel) provides us the results of Brownian motion with subordinator defined as a fuzzy random number over uniform distribution $U(0.5; 1.5)$. Within the model, a symmetry of log-returns can be assumed or it can be relaxed by setting suitable θ to obtain either the positive and the negative skew. Similarly, in the second panel, the fuzzy-option price assuming Model 3 is depicted for fuzzy random gamma process with variance parameter 0.85 allowing again both, symmetry and asymmetry of log returns by setting suitable θ.

Model 1 (BS model with fuzzy volatility σ_{LU}) $S_0 = 100$, $K = 100$, $r = 0$, $T = 1$		
	$(\sigma_m = 0.15)$, $(\sigma_m = 0.25)$	
$\sigma_s = 0.1$	$\begin{pmatrix} (1.22, 227) & (3.10, 209) & (6.76, 7023) \\ (22.56, -138) & (14.66, -161) & (6.76, -7023) \end{pmatrix}$, $\begin{pmatrix} (1.15, 235) & (3.78, 202) & (8.85, 8773) \\ (26.75, -128) & (17.22, -148) & (8.85, -8773) \end{pmatrix}$	
$\sigma_s = 0.05$	$\begin{pmatrix} (2.89, 221) & (5.98, 194) & (10.05, 9849) \\ (19.64, -148) & (14.69, -156) & (10.05, -9849) \end{pmatrix}$, $\begin{pmatrix} (1.04, 208) & (2.90, 207) & (5.73, 35241) \\ (14.33, -155) & (9.28, -180) & (5.73, -35241) \end{pmatrix}$	

Table 1 Output table of pricing algorithm for put options when various models are considered

More complex figure about the distribution of option values due to the particular models can be obtained by inspection of fuzzy-histograms (see Figure 1). Apparently, both subordinators lead to very different results. While fuzz-subordinator makes the distribution clearly decreasing with slight discontinuity for very low values, the fuzzyfication of gamma subordinator leads to some sort of normal distribution.

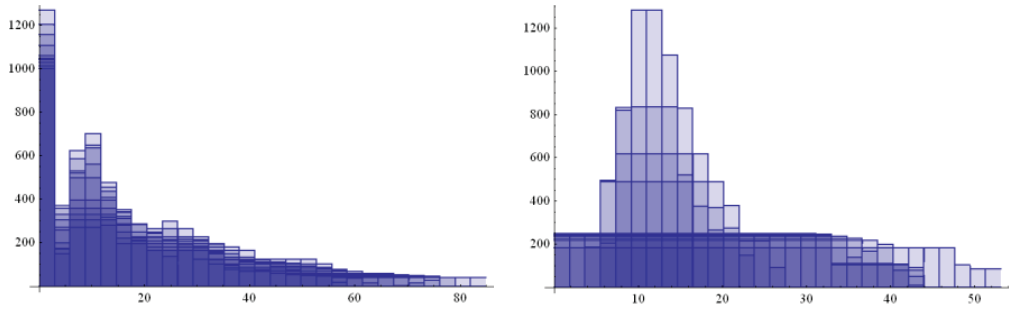


Figure 1 Comparison of fuzzy histograms of option price for both models fuzzy subordinator on the left and fuzzyfied gamma subordinator on the right

6 Conclusion

Many issues of financial modeling and decision making require some knowledge about the future states. However, sometimes it is very difficult to get reliable parametrization of stochastic models. In this contribution we suggested an alternative approach to option valuation problem via Monte Carlo simulation by specifying three distinct types of fuzzy-random processes. Suggested models of financial returns can have very interesting impact on option pricing and hedging. First we should note that the BS option price should always be around the midpoint. The results there-fore indicated that increasing the fuzzy-volatility, we get wider spread of fuzzy-option price. Similarly, we could observed the same effect for additional skewness. Obviously, positive skewness had higher impact. In subsequent research It can be interesting to study the effect of particular parameters on fuzzy-option price, compare it to real market data as well analyze the convergence of fuzzy-Monte Carlo simulation.

Acknowledgements

This work was supported by the European Regional Development Fund in the IT4Innovations Centre of Excel-lence project (CZ.1.05/1.1.00/02.0070) and the European Social Fund (CZ.1.07/2.3.00/20.0296). The research of the second author was supported by SGS project of VSB-TU Ostrava under No. SP2013/3 and Czech Science Foundation through project No. 13-13142S. The support is greatly acknowledged.

References

[1] A. Ban. Approximation of fuzzy numbers by trapezoidal fuzzy numbers preserving the expected interval. *Fuzzy Sets and Systems*, 159(11):1327 – 1344, 2008.

[2] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637 – 659, 1973.

[3] P. Boyle. Options: a monte carlo approach. *Journal of Financial Economics*, 4:323–338, 1977.

[4] U. Cherubini. Fuzzy measures and asset prices: Accounting for information ambiguity. *Applied Mathematical Finance*, 4:135 –149, 1997.

[5] R. Cont and P. Tankov. *Financial Modelling with Jump Processes*. CRC Press, 2010.

[6] D. Dubois and H. Prade. Operations on fuzzy numbers. *International Journal of Systems Science*, 9:613–626, 1978.

[7] M.L. Guerra and L. Stefanini. Approximate fuzzy arithmetic operations using monotonic interpolations. *Fuzzy Sets Systems*, 150(1):5–33, 2005.

[8] M. Holčapek and T. Tichý. Option pricing with fuzzy parameters via monte carlo simulation. *Communications in Computer and Information Science*, 211:25–33, 2011.

[9] M. Holčapek and M. Štěpnička. Arithmetics of extensional fuzzy numbers – part I: Introduction. In *Proc. IEEE Int. Conf. on Fuzzy Systems*, page submitted, Brisbane, 2012.

[10] M. Holčapek and M. Štěpnička. Arithmetics of extensional fuzzy numbers – part II: Algebraic framework. In *Proc. IEEE Int. Conf. on Fuzzy Systems*, page submitted, Brisbane, 2012.

[11] R. Merton. Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4:141–183, 1973.

[12] P. Nowak and M. Romaniuk. Computing option price for lévy process with fuzzy parameters. *European Journal of Operational Research*, 201:206–210, 2010.

[13] R.A. Riberio, H.J. Zimmermann, R.R. Yager, and J. Kacprzyk. *Soft Computing in Financial Engineering*. Springer, 1999.

[14] L. Stefanini, L. Sorini, and M.L. Guerra. Parametric representation of fuzzy numbers and application to fuzzy calculus. *Fuzzy Sets Systems*, 157(18):2423–2455, 2006.

[15] L. Stefanini, L. Sorini, and M.L. Guerra. Simulation of fuzzy dynamical systems using the LU-representation of fuzzy numbers. *Chaos, Solitons & Fractals*, 29(3):638–652, 2006.

[16] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

The SGP - faulty by design or made faulty by politicians? An assessment based on simulation model.

Šárka Horáková[1], Robert Jahoda[2]

Abstract. By joining the European Monetary Union (the "EMU") member countries lost the ability to use the monetary policy as a tool of macroeconomic regulation. The attention was than focused on the regulation of the fiscal policy and the agreed instrument was the Stability and Growth Pact (the "SGP"). States of the EMU have agreed to satisfy the 3% of GDP for the maximum annual public budgets deficit. Based on the evolution of the public debt in member countries we can say that the SGP has failed as a tool of fiscal discipline. In this paper we answer the question whether the failure was given due to incorrect concept of the SGP or whether the development of the debt was affected more by arbitrary disrespect of agreed rules. The two reasons mentioned above are interdependent. To separate them, we construct a dynamic model of EU countries public debt. By using real data we simulate the potential values of public debt in a situation where the SGP rules had been respected in recent years. Comparing the results for the potential debt given by simulation of the model with the current real values we are able to quantify the impact of non-compliance for each country. The initial results indicate that there are both EU states where non-compliance has led to a negligible increase in public debt - up to 5% of GDP, as well as other states where this factor caused the growth of public debt by more than 25% of GDP.

Keywords: Fiscal policy, primary balance, public debt, European Monetary Union, Stability and Growth Pact, fiscal sustainability

JEL Classification: C63, E62, H62, H63
AMS Classification: 65C20, 68U20

Introduction

One of the motives of the Maastricht Treaty was the establishment of procedures for the functioning of the single currency and the European Monetary Union ("EMU"). By joining the EMU member countries dress surrendered part of their sovereignty and entrust monetary policy in the hands of the European Central Bank (ECB). Fiscal policy remained to them as a tool for macroeconomic stabilization. The need for coordination of fiscal policies is necessary to ensure the functioning of the common currency and the Stability and Growth Pact (SGP) was drawn as a tool of fiscal discipline. It was expected that the Maastricht criteria together with the SGP could ensure fiscal discipline in the EU. However, as we saw in last years, the expectations were not met. Von Hagen and Wolff [7] claim that the introduction of SGP together with Excessive Deficit Procedure (EDP) caused the use of "creative accountability" when member countries were falsifying the statistics or using nonstandard fiscal operations to meet the reference values.

The financial crisis, which has grown into a debt crisis in the euro area, clearly shows how important it is to pay more attention to the state of public debt. Despite Maastricht criteria and the SGP, many states failed to reduce the value of the debt to a sustainable level (Van Nieuwenhuyze [9]). Stoian and Alves [11] in their study also showed that the euro area does not perform such a fiscal policy that would be compatible with the requirements of the SGP.

One of the consequences of the debt crisis is that risk can be determined for each country separately, while bonds are denominated in the common currency. This is in a contradiction with the principles of the EMU, which were established in the 90-ties (Buiter [3]). The economy of the member countries is very heterogeneous and for more, we cannot talk about fiscal federalism in EU. The Euro seems overvalued for the southern European countries (including France) and underestimated for countries in Northern Europe, especially Germany (Duwicquet [5] or Coudert [4]). Financial markets may decide to refuse a loan for some countries, simply be-

[1] Masaryk University, Faculty of Economics and Administration, Department of Public Economics, Lipová 41a Brno 62100, sarka.horakova@econ.muni.cz
[2] Masaryk University, Faculty of Economics and Administration, Department of Public Economics, Lipová 41a Brno 62100, robert.jahoda@econ.muni.cz

cause the operation would be too risky. Consequently heavily indebted countries may find themselves close to default.

The concept of fiscal sustainability can be defined variously. Blanchard et al. [2] for example, states that sustainable fiscal policy is when the ratio of debt / GDP after some excessive deflection returns to its original value. Afonso [1] simply claims that for sustainable fiscal policy, the government deficits have to be compensated by deficit surpluses in the future. ECB [6] notes that fiscal sustainability is *a country's ability to service all accumulated government debt at any point in time*.

Reference values in the SGP have been established specifically to ensure the sustainable fiscal discipline of the signatory countries. The value of 60% debt / GDP ratio was determined based on the average of members. 3% threshold for the deficit / GDP ratio is based on a simple dynamic model for the public debt. The respect of this value is necessary to avoid a constant growth of debt and the snowball effect.

In our paper, we use the simple public debt dynamic model to answer the question whether the failure was given due to incorrect concept of the SGP or whether the development of the debt was affected more by arbitrary disrespect of agreed rules. The two reasons mentioned above are interdependent. To separate them, we construct a dynamic model of EU countries public debt. By using real data we simulate the potential values of public debt in a situation where the SGP rules had been respected in recent years. Comparing the results for the potential debt given by simulation of the model with the current real values we are able to quantify the impact of non-compliance for each country.

Sotian and Alves [16] used similar model to calculate what level of government surplus would be necessary in order to stabilize public debt. Stoian [12] believes that the economic crisis may be launched due to improper fiscal policy. He uses the dynamic model of debt to determine whether the conduct of fiscal policy is vulnerable or not. Afonso [1] uses the model of a stationary test to count fiscal sustainability of all Eurozone member countries and like most authors he agrees that the implementation of fiscal policy is not sustainable. By using the model Izák [8] confirms that the government that runs budget deficit faces significantly higher costs of borrowing. Van den Noord [10] attempts to quantify the costs of financial crisis and the impact on the sustainability of public finances.

1 Methodology and data

In the following part of the paper we present a simulation model to assess the impact of deviating the terms agreed in SGP. When constructing the model, we assume the literature of fiscal sustainability, especially the already mentioned works by Blanchard et al. [2] or Izák [8]. This chapter contains also a brief description of the data used in the model.

We use two basic indices for the variables. The upper always marks a year for which the value is calculated or simulated. For the lower index we have two possibilities. In the first case, the lower index of variables is missing; than the variable is calculated from the available data (see next section). In the second case the lower index S is used to describe the simulated variable which is determined as described below.

The root of model (1) is based on the equation of time for the public debt, which stipulates that the level of public debt in current year (marked as B^i) is equal to the level of debt in the previous year and this year's change in debt.

Ideally, the change in this year's debt (equation 2) is equivalent to the level of public deficit in given year (D^i). Afonso [1] draws attention to the fact that the transfer of public deficit into debt is not necessarily time-consistent within one year. He also notes that the government policy can raise the debt by extra budgetary ways. Therefore we introduce the variable ADJ^i that captures all situations in which the growth of public debt in the reference year is different from the size of the public deficit.

The public deficit (in equation 3a) can be expressed as the sum of the primary balance (PD^i) and interest on the public debt (IB^i). At the same time we follow the standard practice of using capital letters to indicate nominal values of the monitored variables and lowercase letters denote their share in gross domestic product (see equation 3b).

$$B^i = B^{i-1} + \Delta B^i \tag{1}$$

$$\Delta B^i = D^i + ADJ^i \tag{2}$$

$$D^i = PD^i + IB^i \tag{3a}$$

$$d^i = \frac{D^i}{GDP^i} \tag{3b}$$

In the real economy the amount of interest on public debt depends on a combination of instruments used by the government to finance the debt and their cost. We use a simplified relationship in the model (equation 4) to show that the amount of interests in a given year depends on the amount of the debt at the end of the previous year and the average implicit interest rate (r^i) of the debt. In equation 5 is expressed in annual real growth of gross domestic product (y^i) a nominal growth adjusted for inflation (π^i).

$$IB^i = r^i * B^{i-1} \tag{4}$$

$$y^i = \frac{GDP^i}{GDP^{i-1}} - 1 - \pi^i \tag{5}$$

The rules of the SGP assume that the public deficit in the medium term has to converge to zero and must not exceed 3% of GDP in a given year. While we cannot clearly interpret what „the public deficit in the medium term has to converge to zero „ means, the 3% limit is easily controllable.

If the conduct of budgetary policy does not respect this limit, we cannot blame the approved rules of the SGP for the consequences. The equation 6 therefore provides a basic assumption on which the model stands, namely that the value of the public deficit in each year should be lower than 3% of GDP.

As such, we simulate this policy in the form of the ratio indicator (d_S^i) or nominal indicators (D_S^i) in the equation 7. The S indicates that these values are simulated to satisfy the rules of the SGP and at the same time to not reach worse values than achieved in actual policy.

$$d_S^i \leq 3\% \tag{6}$$

$$d_S^i = \frac{D_S^i}{GDP^i} \tag{7}$$

Assuming that the simulated policy of the state in each year complied with the rules of the SGP and it would not achieve the worse values than realpolitik, it is expected that the simulated level of public debt B_S^{i-1} should be reduced, or at least at the same level as the actual debt. In this situation, however, the level of simulated interests on the public debt IB_S^i was also on the lower or the same level as for the actual level of public debt. We assume (see equation 8) that the simulated level of interests on public debt is due to the simulated level of public debt and the amount of the implicit interest rate.

$$IB_S^i = r^i * B_S^{i-1} \tag{8}$$

Simulated level of public deficit in a given year is determined by simulated interest on the public debt and the amount of simulated primary deficit (equation 9a and 9b). If the government succeeds to maintain the public deficit below the reference value of 3% of GDP, there is no reason for the simulated primary balance to be different from the real state. However, if the budgetary policy of the government does not comply with the rules of the SGP we simulate the primary balance so that it reaches together with the simulated interest on public debt the worst permissible level - 3% of GDP. Another constraint for the simulated primary balance says that it cannot be worse than the actual primary balance (such term is described in equation 9a).

$$if\,(d^i \leq 3\%; PD_S^i = PD^i; PD_S^i = MIN(PD^i; 3\% * GDP^i - IB_S^i)) \tag{9a}$$

In this situation, the overall deficit is equal to the sum of the simulated primary balance and simulated interest on public debt. This deficit never exceeds 3% of GDP in the given year and is in accordance with the rules of the SGP. The rules, however, allow an exception for the 3% reference limit in case when the real GDP growth is negative or the annual growth is slowed by 2 percentage points. For such situations the model allows that simulated value of public budget deficit is equal to the actual value of the primary balance and simulated interest on the public debt (this exception is described in equation 9b).

$$if\ ((y^i < 0 \vee (y^i - y^{i-1}) < -2\%); D_S^i = PD^i + IB_S^i; D_S^i = PD_S^i + IB_S^i) \tag{9b}$$

Equations 1 and 2 are used to describe the decomposition of actual government debt, equations 10 and 11, then return to compose the simulated level of debt. Equation 10 therefore says that the simulated increase in public debt is equal to the simulated deficit and "time discrepancies" that we have excluded in equation 2 from other calculations. Finally, equation 11 describes the simulated amount of debt at the end of the budget period is equal to the simulated value of the debt at the beginning of the period and the simulated growth of public debt in the given period.

$$\Delta B_S^i = D_S^i + ADJ^i \tag{10}$$

$$B_S^i = B_S^{i-1} + \Delta B_S^i \tag{11}$$

In order to perform the above mentioned calculations, it is necessary to set the period from when the quality of budgetary policy of the government will be followed up. Due to the date of the introduction of the Stability and Growth Pact and the availability of the necessary data for each country, the first simulation for the public deficit is for the year 1996. At the same time, we set of the same level for simulated and actual indebtedness in 1995 (equation 12). The most recent data in our model used from the Eurostat server are for the year 2012. We evaluate the quality of fiscal policy of governments in the past 17 years, when we compare the amount of real and simulated level of public debt (equation 13). For reasons of comparability we use always a GDP ratio.

$$B_S^{1995} = B^{1995} \tag{12}$$

$$A1_j = \frac{B^{2012} - B_S^{2012}}{GDP^{2012}} \tag{13}$$

As already mentioned the model is based on Eurostat data for the period 1995 - 2012 and assesses the quality of budgetary policies of national governments in 1996 - 2012. Unfortunately, not all data are available in time series 1995 - 2012. In the case of Romania the first data are available in 1996, in Bulgaria in 1998 and in Denmark, Poland and Greece up from 2000. In these cases, the evaluation period of fiscal policy is shorter. In the model, the following data from the Eurostat server were used: gross domestic product, public debt, public deficit and public debt interest costs. All other values are calculated according to the equations mentioned above. The values for inflation are calculated from the development of the HICP annual average.

2 Results

The main results are shown in the following Figure 1 The dark column reveals the amount of debt in 1995, the hatched column shows the change in the amount of debt up to 2012 in the situation when states would implement responsible fiscal policy in the sense that they follow the rules of the SGP. The black indicator A1 (values on the secondary axis) indicates the effect of failure to comply with these rules to change of the level of deb. The states in the chart are sorted according to this indicator. Finally, the sum of all three values (green dot) gives the actual level of debt in 2012. All values are given in% of GDP.

Even in 1995, the debt of the most of the countries was below or near the level of the Maastricht convergence criteria (60% of GDP). The average value (weighted average with weights according to the economic performance of each state) for the above-mentioned 27 countries reached 66.2% of GDP. Even then the states such as Greece (94.0% GDP), Italy (120.9% of GDP) and Belgium (130.2% of GDP) protruded above that average.

If states apply responsible budgetary policy, it is surprising that the greater part of them would meet the increase of public debt. The average increase in the value of the above-mentioned 27 countries is 9.0% of GDP. In some countries the application of responsible budgetary policy brings excepted results i.e. the decrease in public debt. The indicator A1 shows the effect of non-compliance with the agreed rules on the increase in the debt level. A1 for the budgetary responsible countries reaches up to 10% of GDP (countries on the right of the chart), for less fiscally responsible countries the indicator reaches 30% and more. (the left part of the chart). The average value of this indicator monitored for 27 countries is 11.4% of GDP. This indicator thus explains the 56% of increase in the average debt of the monitored countries for the period 1995-2012. Let's assume that the rules of the SGP do not allow derogation from the responsible budgetary policy in case of economic decline. The average value of A1indicator would be 15.5% of GDP and explains 76% of increase in the average debt.

As we said the indicator A1 is able to quantify the extent to which the budgetary policy was responsible in the period 1995 – 2012. It may be surprising that in the case of disciplined countries (the right side of the chart) the impact on the reduction of the debt was greater than in undisciplined countries. The graph shows that these states would not succeed to reduce the public debt even if their budgetary policies were responsible. In contrary, the level of public debt would decrease in all of them except Hungary who is not a Eurozone member. It is hard to explain this paradoxical situation, our next paper will be focused on detailed analysis. Hagen and Wolff [7] offer a partial explanation by saying that: "SGP rules have induced governments to use stock-flow adjustments. This tendency is especially strong for situations when reducing the deficit is particularly large". Governments that had problems to deal with public deficits (and therefore with the rules of the SGP), probably solved this situation with using the stock-flow adjustment (creative accounting) to avoid the accusation of not fulfilling the agreed rules.

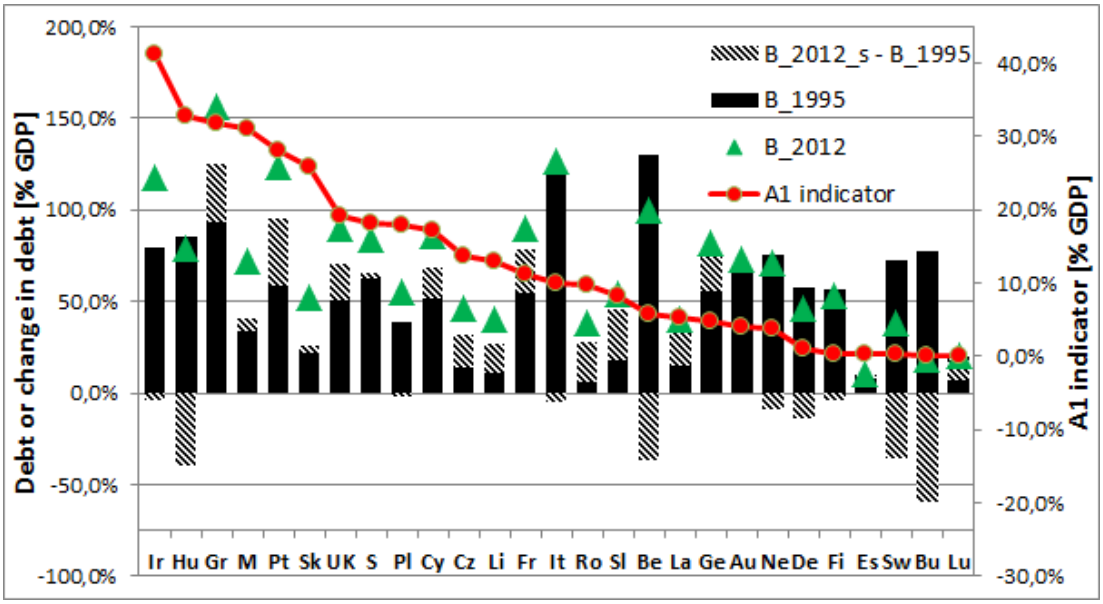


Figure 1 Values for debt, change of the debt and A1 indicator of EU member countries (% of GDP)

Another explanation that can be given is the existence of exceptions to the SGP. We take in to account only the influence of the economic downturn. The annual decline of more than 2% does not require maintaining the deficit below the 3% of GDP. The results of our model show that this is the case especially for the counties from the left side of the chart. Our model shows that the average implicit interest rate on public debt was still close to 8% in 1996 (see Figure 2).The greater frequency of situations when the countries can avoid the agreed rules causes that the effect of budgetary responsible policy will be outweighed by influence of the period when the states do not pursue the level of deficit.

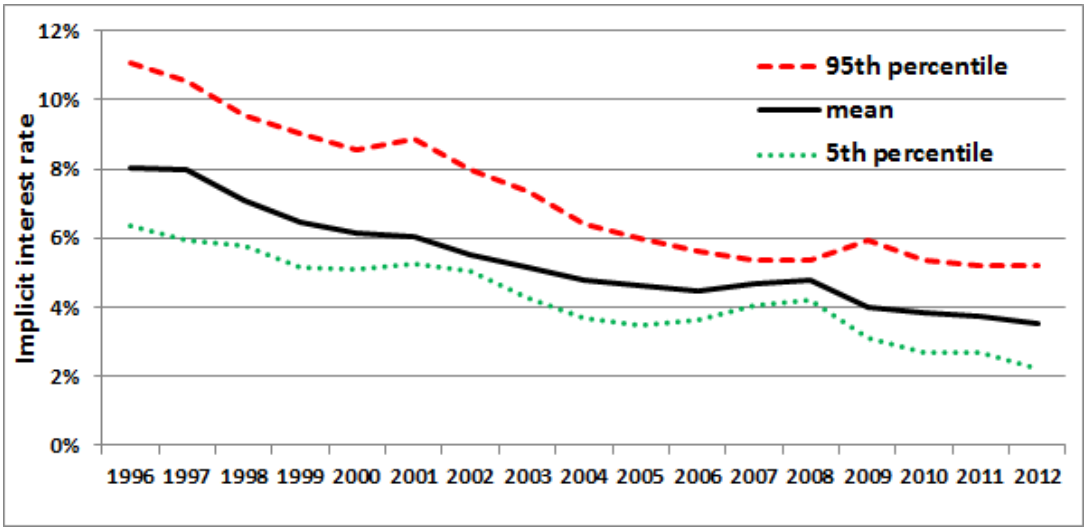


Figure 2 The average implicit interest rate and the development of values corresponding to the 5th and 95th percentile

Along with the average public debt around 60% of GDP, this meant that countries with higher levels of debt had to achieve primary surpluses to thrive comply with the Maastricht convergence criteria for entry into the Eurozone. Their settings - maximum 3% deficit and 60% debt of GDP - were adopted with regard to the implicit interest rate on public debt in the amount of 5%. One of the positive effects of the introduction of the common currency was the convergence of interest rates and their reduction to below 4%. As the figure shows, the greatest degree of convergence was achieved in 2008, from that year due to the financial and economic crisis, interest rates diverge more. The situation of lower interest cost and primary surpluses in the countries could be used to sharp reduction of fiscal imbalances. But it turned out that one component of the deficit (interest costs) was continuously substituted by another component of the deficit (primary deficit).

The results of our previous research indicated that good intentions about themselves are not sufficient to ensure good outcomes of the budgetary policy. The rules of the Stability and Growth Pact allowed the short-term deficit financing. In the medium term the public finances should be balanced, however it never happened in the most of the countries. Even the application of the maximum allowable limit for the deficit was not observed by some states. According to our calculations, even if the fiscally responsible policies took place between 1995 and 2012 we would observe an average increase in the level of public debt by 9.0% of GDP. The disrespect of the 3% rule caused the increase in the level of public debt of 11.4% of GDP in the reference period. We should also notice the influence of variable stock-flow adjustment, which in some countries was used to conceal the real state of public finances. The rules of SGP were set simply and uniformly so they led to the decline in interest costs on public debt. But the countries are not forced to profit from this favourable situation and did not reduce their public deficit. If we summarize the above, we must conclude that the actual setting of the SGP not guarantee adequate decrease or stabilization in the level of public debt. This effect was further enhanced by non-compliance with agreed rules - both by Member States and by the European Commission, respectively representatives of ECOFIN who did not enforce the respect of SGP.

Acknowledgements

Supporting people and institutions (including Science Foundations) go here.

References

[1] AFONSO, António. *Fiscal policy sustainability: some unpleasant European evidence.* [online]. SSRN [cit. 2013-04-20]. Available on World Wide Web: <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=3174834>

[2] BLANCHARD, Olivier. et al. *The sustainability of fiscal policy: new answers to an old question.* [online]. OECD economic studies 15, autumn 1990

[3] BUITER, Willem., *The Sense and Nonsense of Maastricht' revisited: What have we learnt about stabilization in EMU?* . [online]. [cit. 2013-04-20]. Available on World Wide Web: < http://www.willembuiter.com/sense.pdf>

[4] COUDERT, Virginie et al. *Les mésalignements de taux de change réels à l'intérieur de la zone euro*, Revue de l'OFCE, 2013/1 N° 127,, p.35-56. [cit. 2013-04-20]. Available on World Wide Web:http://www.cairn.info/article.php?ID_REVUE=REOF&ID_NUMPUBLIE=REOF_127&ID_ARTICLE=REOF_127_0035&WT.mc_id=REOF_127#s1n6

[5] DUWICQUET, Vincent, et al. *Désajustement de change fédéralisme budgétaire et redistribution*. [online]. Revue de l'OFCE, 2013/1 N° 127,, p.57-96. [cit. 2013-04-20]. Available on World Wide Web: <http://www.cairn.info/article.php?ID_REVUE=REOF&ID_NUMPUBLIE=REOF_127&ID_ARTICLE=REOF_127_0057&WT.mc_id=REOF_127#citation>

[6] ECB monthly bulletion, April 2012

[7] von HAGEN, Jürgen-WOLFF, Guntram, B. *What do deficits tell us about debt? Empirical evidence on creative accounting with fiscal rules in the EU*. [online]. GESY [cit. 2013-04-20]. Available on World Wide Web: <http://www.sfbtr15.de/uploads/media/148.pdf>

[8] IZÁK, Vratislav. Primary balance, public debt and fiscal variables in postsocialist members of the european union. [online]. *Prague Economic Papers* 2/2009

[9] Van NIEUWENHUYZE, Christophe. *Debt, assets and imbalances in the euro area.* [online]. Revue de l'OFCE, 2013/1 N° 127, p. 123-152. [cit. 2013-04-20]. Available on World Wide Web: http://www.cairn.info/revue-de-l-ofce-2013-1-page-123.htm

[10] Van den NOORD, Paul Joseph, Turning the page? EU fiscal consolidation in the wake of the crisis, *Empirica*, 38, issue 1, p. 19-51.

[11] STOIAN, Andreea-ALVES, Rui, Henrique. *Can EU High Indebted Countries Manage to Fulfill Fiscal Sustainability?Some Evidence from the Solvency Constraint*. FEP working papers

[12] STOIAN, Andreea, A Retrospective Approach on Assessing Fiscal Vulnerability: Empirical Evidence for Overindebted European Countries, *Timisoara Journal of Economics*, vol. 4(3(15)), p. 183-188.

The spectral properties of the lag operator

Richard Horský[1]

Abstract. In the theory of stochastic processes some operators are employed. Among them the lag operator plays remarkable role. It enables us to get formally familiar descriptions of the stochastic processes. On the other hand we can deal with this operator as the object of the functional analysis. This approach provides more particular view on the problems connected with the convergence of stochastic processes. We will analyze the spectral properties of the lag operator in different spaces.

Keywords: stochastic process, stationarity, lag operator, spaces of sequences, spectral analysis, spectral radius.

1 The stochastic process, its convergence and stationarity

The space $L_2(\Omega,\pi)$ is the space of all functions defined (almost everywhere) on the measurable space Ω equipped by the probability measure π. It is a complete space with respect to the norm derived from the scalar product $(X,Y)=\int_\Omega XYd\pi$, i.e. the Hilbert space. The elements of this space are random variables with a finite mean and a finite variance. The space of all measurable functions defined (almost everywhere) on the measurable space Ω with the probability measure π is usually denoted as $L_0(\Omega,\pi)$. The convergence structure is given by the convergence in probability.

Stochastic process is a mapping $\mathbf{X}: T \to L_0(\Omega,\pi)$, domain of which is the set of all integers *T*. The values of this mapping are random variables $X_t, t=0,\pm1,\pm2,\ldots$ We will write $\mathbf{X}=(X_t)$. The domain *T* is usually interpreted as time set. On the other hand one can only observe results of a stochastic process. The *time series* is a sample realization of such a process from an infinite population of possible realizations. We often talk about a model of time series. The model denotes the same as the stochastic process. It is a form in which the time series is generated.

1.1 Stationary stochastic process

We usually take some assumptions on a stochastic process $\mathbf{X}=(X_t)$ to ensure its „good"properties:

1) All the random variables X_t have the same finite mean: $EX_t=\mu$.
2) All the random variables X_t have the same finite variance: $\sigma^2=DX_t=EX_t^2-\mu^2$.
3) The covariance is invariant with respect to any time shift, i.e. $\mathrm{cov}(X_t,X_s)=\mathrm{cov}(X_{t+k},X_{s+k})$ for all integers *t, s, k*.

Such a process is called *stationary stochastic process*. One often assumes the case of the *normal distribution* within a process. Then the stochastic process is called *normal* and stationarity is satisfied. It is convenient to define the *autocovariance function*

$$\gamma_k=\mathrm{cov}(X_t,X_{t+k}).$$

This function, the domain of which is the time set *T*, is even and so we can regard it only for nonnegative integers. Its value at zero is obviously the variance $\gamma_0=DX_t=\sigma^2$. We can notice that the requirement 1) means that $X_t\in L_1(\Omega,\pi)$ and the requirement (2) that $X_t\in L_2(\Omega,\pi)$. The autocovariance is the scalar product of so called *centered variables* $X_t-\mu$ (its mean is zero) in the Hilbert space $L_2(\Omega,\pi)$.

1.2 The white noise process and general linear process

If a stochastic process has zero mean, a finite positive variance σ_ε^2 and consists of noncorrelated random variables (shocks) ε_t (it means $\gamma_k=0$ for any positive *k*), then it is called the *white noise process*. This is the basic

[1] University/of Economics, department of mathematics, Ekonomická 957, 148 00 Praha 4, rhorsky15@seznam.cz.

stochastic process, from which other models are derived. This process is obviously stationary. The white noise process is an orthogonal system in $L_2(\Omega,\pi)$.

The *general linear process* has the form

$$X_t = \mu + \sum_{n=0}^{\infty} \psi_n \varepsilon_{t-n}, \tag{1}$$

where μ is the common mean of all X_t and (ψ_n) is a sequence of real numbers that are called *weights of the process*, $\psi_0 = 1$. We can suppose w.l.o.g. that $\mu = 0$. The stochastic process with $\mu = 0$ is called the *centered process.* The partial sums of the series in (1) are random variables

$$S_t(n) = \sum_{j=0}^{n} \psi_j \varepsilon_{t-j}. \tag{2}$$

The convergence of the series (1) is regarded as the *convergence in the square mean.* It means

$$E(S_t(n) - X_t)^2 \to 0, \text{ for } n \to \infty. \tag{3}$$

The convergence (3) is derived from the norm of the Hilbert space $L_2(\Omega,\pi)$. We will show that this type of convergence is equivalent to the stationarity of the process (1). If the process is not stationary then the series on the right side in (1), or equivalently the sequence (2) respectively, are not convergent.

Let us take the function $S_t(n) - S_t(m)$, $n > m$. The partial sums (2) are elements of the space $L_2(\Omega,\pi)$ that is complete with respect to the norm $\|X\|_2^2 = E(X^2) = \int_{\Omega} X^2 d\pi$. Hence we obtain

$$E(S_t(n) - S_t(m))^2 = \sum_{i=m+1}^{n} \sum_{j=m+1}^{n} \psi_i \psi_j E(\varepsilon_{t-i}\varepsilon_{t-j}) = \sum_{i=m+1}^{n} \psi_i^2 E(\varepsilon_{t-i}^2) = \sigma_\varepsilon^2 \sum_{i=m+1}^{n} \psi_i^2. \tag{4}$$

Thus if

$$\sum_{n=0}^{\infty} \psi_n^2 < \infty \tag{5}$$

holds, the right side of (4) tends to zero, $\sum_{i=m+1}^{n} \psi_i^2 \to 0$ for $m, n \to \infty$. It follows the sequence (2) is Cauchy in $L_2(\Omega,\pi)$ and hence the series in (1) is convergent. If the condition (5) is satisfied, the linear process is stationary. We can express the autocovariance function of the process (1) in terms of its weights

$$\gamma_k = E(X_t X_{t+k}) = E\left(\sum_{n=0}^{\infty} \psi_n \varepsilon_{t-n}\right)\left(\sum_{n=0}^{\infty} \psi_n \varepsilon_{t+k-n}\right) = \sigma_\varepsilon^2 \sum_{n=0}^{\infty} \psi_n \psi_{n+k}. \tag{6}$$

The condition (5) is equivalent to the fact that the sequence of the weights lies in the Hilbert space l_2. The series on the right side in (6) is a scalar product of the sequences (ψ_n) and (ψ_{n+k}) in the space l_2. With respect to the well-known Schwartz inequality and the fact the l_2-norm of (ψ_n) is equal or greater than the l_2-norm of (ψ_{n+k}) we obtain

$$|\gamma_k| = \sigma_\varepsilon^2 \left|\sum_{n=0}^{\infty} \psi_n \psi_{n+k}\right| \le \sigma_\varepsilon^2 \sum_{n=0}^{\infty} \psi_n^2 = \gamma_0. \tag{7}$$

The inequality (7) proves that the centered process (1) is stationary the autocovariances being upper bounded by the finite variance $\gamma_0 = EX_t^2$ of the centered process (1).

In the end of this section we can notice that a sufficient condition for (5) is the absolute convergence of the series $\sum_{n=0}^{\infty}\psi_n$, in other words

$$\sum_{n=0}^{\infty}|\psi_n|<\infty. \tag{8}$$

The condition (8) is equivalent to the fact the sequence (ψ_n) lies in l_1-space. (8) is only sufficient but not necessary for (5). It is obvious from the trivial example $\psi_n=\frac{1}{n}$.

2 The lag operator

The introduction of the lag operator provides a simplification for formal writing of the stochastic processes at least in the case of models of the type (1). But as we said it is not only formal abbreviation. This operator bears the typical properties of any linear operator similarly to operators of derivative, difference, summation or integral and so on. That shows it is an object which may be studied by the means of the functional analysis.

2.1 The algebra of the lag operator

As to the formal introduction of the notion of the lag operator it is not necessary to solve the problem of its domain and range (see [2] or [3]). However, we cannot avoid to specify these sets since the operator is a mapping after all. The natural access with respect to the definition of the stochastic process or its special case the general linear process seems to define

$$B:\Pi\to\Pi,\quad B(X_t)=(X_{t-1}) \tag{9}$$

for any stochastic process $\mathbf{X}=(X_t)\in\Pi$, where Π denotes the set of all stochastic processes. The mapping B is the so called *lag operator*. We will use this operator in the form $BX_t=X_{t-1}$, which means its application on the particular random variable in the stochastic process. It allows to write some formulae in more comfortable way.

The set Π is in actually the linear space. The addition and scalar multiplication are defined in the analogical way as in the case of the spaces of sequences or functions, it means pointwise. It can be easily seen that the mapping (9) is a linear operator. Its powers, defined by the induction

$$B^0=I\ ,\ B^{n+1}=B\left[B^n\right] \tag{10}$$

for all nonnegative integers n, are also linear operators. The symbol I in (10) denotes the identity on Π. Thus by (10) we have $B^k(X_t)=(X_{t-k})$ or $BX_t=X_{t-k}$. The space of all linear operators on Π forms the algebra (see [1], p.9). The algebra of all (finite) linear combinations of the operators (10) is its subalgebra.

2.2 Algebraic and analytical properties of the lag operator

As to the infinite linear combinations, i.e. the power series in B, we can apply the results from the previous part 1.2. However, the infinite linear combinations are connected with the structure of convergence. We restrict to the stochastic processes which are bounded, i.e.

$$\|(X_t)\|=\sup\{|X_t|:t\in T\}<\infty. \tag{11}$$

The stochastic processes satisfying (11) forms a space that we denote B(T).The function (11) is the norm in this space and this space is complete (Banach) space with respect to this norm. B(T) is a subspace of Π. To avoid formal difficulties we denote the restriction of the operator (9) on B(T) by the same symbol, i.e. B. The lag operator is now continuous (or bounded) operator which norm is

$$\|B\|=\sup_{\|(X_t)\|=1}\|B(X_t)\|=1. \tag{12}$$

To see this it is sufficient to take a constant process $\mathbf{E}=(1)$ which norm (11) is obviously 1 and $B\mathbf{E}=\mathbf{E}$. The same argument may be used to show that any operator (10) has the norm equal to 1. Thus all of the operators

(10) lie in the space [B(T)] of all continuous operators on the space B(T). The algebra [B(T)] contains a subalgebra of all finite linear combinations of the operators (10), i.e. the polynomials in B.

Let us draw our attention to the case of the general linear process (1). Using the lag operator B we can formally write

$$X_t = \sum_{n=0}^{\infty} \psi_n B^n \varepsilon_t = \psi(B)\varepsilon_t$$

at any point $t \in T$. The abbreviation $\psi(B)$ stands for $\sum_{n=0}^{\infty} \psi_n B^n$. Now we realize whether this abbreviation is an operator from [B(T)]. Let us compute

$$\|\psi(B)\| \le \sum_{n=0}^{\infty} |\psi_n| \|B^n\| = \sum_{n=0}^{\infty} |\psi_n|. \tag{13}$$

Hence if the condition (8) is satisfied then

$$\psi(B) = \sum_{n=0}^{\infty} \psi_n B^n \in [\mathrm{B}(T)]. \tag{14}$$

The operator (14) is called *linear filter*. The condition (8) is substantial.

Example 1. If we set $\psi_n = \frac{1}{n}$ then $(\psi_n) \in l_2 - l_1$ and $\psi(B)$ is not continuous operator on B(T). It is obvious if we take $\mathbf{E} = (1)$ then $\psi(B)\mathbf{E} \notin \mathrm{B}(T)$.

Example 2. If we set $\psi_n = \lambda^n$ then $(\psi_n) \in l_1$ iff

$$|\lambda| < 1. \tag{15}$$

The condition (15) is equivalent to the fact $\psi(B) = \sum_{n=0}^{\infty} \lambda^n B^n \in [\mathrm{B}(T)]$. $\psi(B)$ is so called *geometric lag operator*. The name is derived from the fact the sequence of the weights is geometric. Let us notice that the choose $\lambda = 1$ leads to the divergent series (in [B(T)]) $\sum_{n=0}^{\infty} B^n$. This process is called *random walk process.*

Example 3. If we set $\psi_n = 0$ for any $n > q$, q *is* fixed nonnegative integer, then $(\psi_n) \in l_1$ and $\psi(B)$ is obviously continuous operator on B(T). Here $\psi(B) = \sum_{n=0}^{q} \psi_n B^n$ is so called *polynomial lag operator (of the degree q).*

Example 4. Suppose $\psi(B) = \frac{\Theta(B)}{\Phi(B)}$, where $\Phi(B)$ is a polynomial lag operator of positive degree p, $\Theta(B)$ is a polynomial lag operator of degree q and polynomials $\Phi(B)$ and $\Theta(B)$ are coprime. The fraction means the composition of the inverse to $\Phi(B)$ and the operator $\Theta(B)$. The problem of the existence of the inverse operator $\Phi^{-1}(B)$ is meaningful. It is shown e.g. in [4], p.27, that $\Phi^{-1}(B)$ exists iff the polynomial $\Phi(z)$ of the complex variable z has all its roots outside the unit circle. This linear filter is called *rational lag operator*.

In terminology of the classical Box Jenkins methodology the linear filter in the Example 3 defines the process denoted MA(q) (*moving average process of the order q*) or that in the Example 4 defines the process ARMA(p,q) (*autoregressive moving average process of the order (p, q)*) respectively.

Besides the convergence of (1) there is another important problem: problem of its invertibility. Using the terminology of the lag operator it is a question of the invertibility of the operator (14). The operator

$$\pi(B)=\sum_{n=0}^{\infty}\pi_n B^n \tag{16}$$

is the inverse to the linear filter (14) iff $\pi(B)\psi(B)=I$ or equivalently

$$\pi_0=1,\ \sum_{k=0}^{n}\psi_k\pi_{n-k}=0, \tag{17}$$

n positive integer. The problem (17) is a difference equation with the initial condition. It determinates the sequence (π_n) uniquely. The condition (8) formulated for (π_n) assures $\pi(B)\in[\mathrm{B}(T)]$. The important case is that in Example 4 when $\Theta(B)=I$. Then $\pi(B)=\Phi(B)$.The model of the form $\Phi(B)X_t=\varepsilon_t$ is well-known as the *autoregressive model of the order p*, abbreviation AR(p). If (14) is of the type from Example 3 and the polynomial $\Theta(z)$ has all roots outside the unit circle then $\pi(B)\in[\mathrm{B}(T)]$ is of the type of rational lag operator $\pi(B)=\Theta^{-1}(B)$.

3 Spectral properties of the lag operator

The lag operator (9) can be also interpreted in another way. The matrix representation of the lag operator requires another conception of the stochastic process. From now on we will regard the stochastic process as the sequence $\mathbf{X}=(X_t,X_{t-1},\ldots)=(x_1,x_2,\ldots)$. This access is justified by the knowledge of the process up to the time *t* (future values are unknown at present time *t* while the previous one are available). We will consider these process as the elements of the spaces $l_p(L_2(\Omega,\pi)),1\le p\le\infty$, i.e. the spaces as the classical l_p (see [7], p.94) but there is a difference: members of the sequences from $l_p(L_2(\Omega,\pi))$ are functions (random variables) from $L_2(\Omega,\pi)$. We will denote them, nevertheless briefly l_p. These spaces are complete and (except l_∞) separable.

In the final dimension the linear operators are represented by matrices. The situation in ifinite dimension is rather more complicated. However if the spaces are complete and have a countable basis the matrix representation is usual (see [7], p.208).

The lag operator *B* is in this context expressed by and coincided with the matrix

$$\mathbf{B}=\begin{bmatrix}0&1&0&\cdots\\0&0&1&\cdots\\0&0&0&\cdots\\\vdots&\vdots&\vdots&\ddots\end{bmatrix}. \tag{18}$$

We will make a classification of the spectrum of (18).

3.1 The elements of the spectral analysis

If we deal with a spectrum of a linear operator $A:V\to V$ we study the properties of the operator

$$\lambda I-A, \tag{19}$$

where λ is a scalar (real or complex). The set of all values λ for which the inverse of the operator (19) exists and its range is dense in *V* is called the *resolvent set of the operator A,* and is denoted $\rho(A)$. Its complement (in the field of scalars) is called the *spectrum of the operator A*, and is denoted $\sigma(A)$. The basic properties of these sets are those $\rho(A)$ is an open subset of the field of scalars, $\{\lambda:|\lambda|>\|A\|\}\subset\rho(A)$, and $\sigma(A)$ is a closed set for which $\{\lambda:|\lambda|\le\|A\|\}\subset\sigma(A)$. The number $r(A)=\sup_{\lambda\in\sigma(A)}|\lambda|$ is called the *spectral radius of the operator A.* It generally holds $r(A)\le\|A\|$. The classification of the spectrum of the operator *A* is as follows:

1. The *point spectrum* $\sigma_P(A)$ is created by the points λ for which $(\lambda I-A)^{-1}$ does not exist. Its elements are so called *eigenvalues of the operator A.*
2. The *continuous spectrum* $\sigma_C(A)$: $(\lambda I-A)^{-1}$ exists but it is not continuous and the range of (19) is dense in *V.*

3. The *residual spectrum* $\sigma_R(A)$: $(\lambda I - A)^{-1}$ exists and the range of (19) is not dense in *V*.

3.2 The spectrum of the lag operator

The operator *B* is coincided with the matrix (18). Suppose $V = l_p$, $1 \le p \le \infty$. The kernel of *B*, i.e. $\{x \in V : B\mathbf{X} = \mathbf{0}\}$, consists of the vectors $\mathbf{X} = x_1\mathbf{E}_1$, where $\mathbf{E}_1 = (1,0,0,\ldots)$. If $\mathbf{Y} = (y_n) \in V$ then $\mathbf{Y} = B\mathbf{X}$ for any $\mathbf{X} = (x_1, y_1, y_2, \ldots)$, x_1 any scalar, and thus the range of *B* is the whole space *V*. Spectral analysis starts by the claim that $\lambda \in \rho(B)$ for any $|\lambda| > 1$. Let us consider the geometric sequence

$$\Lambda = (1, \lambda, \lambda^2, \ldots). \tag{20}$$

The sequence (20) is the eigenvector of *B* whenever $|\lambda| < 1$. The sequence (20) belongs to the eigenvalue λ. It follows that $r(B) = \|B\| = 1$. If $|\lambda| = 1$ then $\Lambda \in l_p$ iff $p = \infty$. It means that $\sigma(B) = \sigma_P(B) = \{\lambda : |\lambda| \le 1\}$ iff $V = l_\infty$. In any case the spectrum of the lag operator *B* is the compact unit circle (in the field of complex numbers).

In the following consideration we will suppose $V = l_p, 1 \le p < \infty$ and $|\lambda| = 1$. We describe the range of $\lambda I - B$. If $\mathbf{Y} = (y_n) \in V$ is given then the equation $(\lambda I - B)\mathbf{X} = \mathbf{Y}$ has in the components form

$$\lambda x_n - x_{n+1} = y_n, \tag{21}$$

The general solution of the linear difference equation (21) is

$$x_n = .C\lambda^{n-1} - y_1\lambda^{n-2} - \ldots - y_1, \tag{22}$$

where *C* is a constant. The space *W* of sequences with a finite nonzero elements is a dense subspace in l_p, $p < \infty$. If $\mathbf{Y} = (y_n) \in W$, i.e. there is a positive integer N such that $y_N = y_{N+1} = \ldots = 0$, we can find an element $\mathbf{X} = (x_n) \in V$ like this: we choose the constant *C* so that $C\lambda^{N-1} - y_1\lambda^{N-2} - \ldots - y_{N-1} = 0$. This choice leads with respect to (22) and (21) to the sequence $\mathbf{X} = (x_n) \in W$ for $x_N = x_{N+1} = \ldots = 0$. This consideration shows that $\sigma_C(B) = \{\lambda : |\lambda| = 1\}$ in l_p, $1 \le p < \infty$.

3.3 Conclusion

The spectral analysis presented in previous section is in accord with the results given in such monographies as [2], [3] or [4]. There is an interesting result as to the existence (and continuity) of the inverse to the difference operator $\Delta = I - B$ which is employed in the theory of stochastic processes. As we have found $\Delta^{-1} = (I - B)^{-1}$ does not exist on the space l_∞. It shows that the AR(1) process with Δ as the autoregressive operator, i.e. $\Delta X_t = \varepsilon_t$, is not stationary linear process. If we restrict Δ on a l_p, $1 \le p < \infty$, Δ^{-1} exists but it is not continuous. That shows a small perturbation of ε_t has generally a great influence on X_t. However the considered restriction means to give up the concept of stationarity. The stochastic process in l_p, $1 \le p < \infty$, may have a constant variance up to some point in the past, further its variance has to tend to zero. However it is not a problem since an observed time series in distant past is not available.

References

[1] Anděl, J.: *Statistická analýza časových řad.* SNTL, Praha 1976
[2] Arlt, J.: *Moderní metody modelování ekonomických časových řad*. Grada Publishing 1999
[3] Cipra, T.: *Analýza časových řad s aplikacemi v ekonomii*. SNTL, Alfa, Praha 1986
[4] Dhrymes, J.P.:*Distributed lags. Problems of Estimation and Formulation*, NORTH HOLLAND, Amsterdam 1985
[5] Marek, I., Žitný, K.: *Matrix Analysis for Applied Science, Vol. 1*. Teubner Verlagsgesselschaft Leipzig 1983
[6] Rényi, A.: *Teorie pravděpodobnosti*. Academia Praha 1972
[7] Taylor, A.E.: *Úvod do funkcionální analýzy*. Academia Praha 1973

The Impact of the Technological Progress on the Development of the Slovak Economy – Neoclassical approach

Peter Horvát[1], Brian König[2], Filip Ostrihoň[3]

Abstract. The following paper introduces a simple RBC model, as a tool for evaluation of the technological progress in the conditions of Slovakia, and its consequent effect on the development of the Slovak economy.
The first section outlines the model itself and contains the derivation of the equilibrium conditions and the steady state values. The second section describes the estimation procedure and the data used for estimation. The Bayesian method was used to estimate the model according to the real values of Slovak GDP and Consumption in order to capture, to the highest possible extent, the complexity of the economic reality with the DSGE model. The final section describes the use of estimated DSGE model in order to assess the impact of technological progress and further evaluate the effect of technological shocks on the selected variables.
According to the estimated model the immediate response of the Slovak GDP to a technological shock, at the magnitude of 1% positive deviation is approximately 1.4% growth of GDP. The impact of the one period shock lasts at least next 12 years, and its aftermath on GDP can be visible for another 13 years.

Keywords: Technological progress, DSGE model, Neoclassical approach, the Bayesian estimation, Slovakia.

JEL Classification: C11, E13
AMS Classification: 91B16

1 Introduction

Based on the neoclassical assumptions the *technological progress* was the main driving force of the per capita GDP growth in developed countries through last several decades. Since Slovakia, as a transitive economy, is seemingly catching up with mentioned countries we can expect stronger influence of the *technological progress* on the economy of the country. The aim of this paper is not to examine the change of the impact of *technological progress* throughout the years, but merely to capture the current state and describe it for possible future analysis.

In the second section we will introduce the model itself, and derive the equilibrium conditions and the steady state values. In following two sections we will describe the estimation procedure and the data used for estimation. We have used the *Bayesian method* to estimate the model according to the actual values of Slovak gross domestic product (GDP) and final household consumption (FHC), in order to capture the complexity of the economic reality, with the mentioned *Dynamic Stochastic General Equilibrium* (DSGE) model to the highest possible extent. In the fifth section we will describe the utilization of the estimated DSGE model to assess the impact of *technological progress* and further evaluate the effect of *technological shocks* on the selected variables. The results of the described analysis are discussed in the final chapter.

2 Neoclassical RBC model

The presented model was inspired by the Kydland and Prescots *real business cycle* (RBC) framework, further methodically reprocessed and described by DeJong and Dave [1]. Therefore we have extensively drawn from DeJong and Dave, particularly relations (1) - (8) and (11) – (16) are contained in the source literature [1]. The model features the labor / leisure trade-off decision process, with the notion of economic growth.

The model assumes a large number of households, which may be aggregated as a representative household. Their aim is to maximize the utility *U*, which yields from the leisure and the consumption of goods. The con-

[1] Economic University in Bratislava, Department of Operations Research and Econometrics, Dolnozemská cesta 1/b, 852 35 Bratislava, Slovakia, 1peter.horvat@gmail.com.
[2] Economic University in Bratislava, Department of Operations Research and Econometrics, Dolnozemská cesta 1/b, 852 35 Bratislava, Slovakia. konigbrian@gmail.com
[3] Economic University in Bratislava, Department of Operations Research and Econometrics, Dolnozemská cesta 1/b, 852 35 Bratislava, Slovakia, filip.ostrihon@euba.sk

sumer also assigns higher utility to the more recent consumption instead of the future consumption. In the model the total consumers' utility is modeled as a sum of discounted instantaneous utility from present to infinity.

$$U = E_0 \sum_{t=0}^{\infty} \beta^t \left(\frac{{c_t}^{\varphi} {l_t}^{1-\varphi}}{1-\phi} \right)^{1-\phi} \rightarrow \max_{c_t, l_t} \tag{1}$$

Where E_0 is the expectations operator, conditioned to the present information (time $t = 0$), β is the discount factor, used to asses the future utility, c_t is the particular consumption of goods and l_t is the amount of leisure at time t. The utility itself is distributed according to *Constant relative risk aversion function* (CRRA), where parameter φ determines the ratio of both the consumption and the leisure on the consumer's immediate utility. The parameter ϕ denotes the risk aversion and the intertemporal elasticity of substitution.

The representative household is also capable to produce goods, according to the production technology at its disposal. As an abstraction goods are modeled through single representative good. The production technology is described, according to the framework of neoclassical modeling, as a *Cobb-Douglas production function*.

$$y_t = z_t {k_t}^{\alpha} {n_t}^{1-\alpha} \tag{2}$$

Where the y_t is the level of output or the production of a single aggregated good, z_t is denoting the level of productivity or *technological progress*. The k_t and n_t are the level of capital stock and the level of labor input respectively. The coefficient α denotes the elasticity of capital and, due to constant returns of scale, also the elasticity of labor of the output. Furthermore we assume that the representative household has to its disposal, at each given moment, one unit of time, which the household can distribute into work or leisure.

$$1 = n_t + l_t \tag{3}$$

Where the n_t denotes the amount of time spent with labor and l_t represents the time left for the leisure. Similarly the output generated at time t can be consumed (c_t) or left for the future, to increase the stock of capital, as an investment (i_t).

$$y_t = c_t + i_t \tag{4}$$

The capital stock itself diminishes over time, through deprecation. That's why it has to be replenished through investment.

$$k_{t+1} = i_t + (1-\delta) k_t \tag{5}$$

Where δ denotes the depreciation rate of the capital (k_t). The evolution of *technological progress* is described as a first order autoregressive process (AR1), in order to capture persistence of the technological shock, to some degree.

$$\log z_t = (1-\rho) \log \bar{z} + \rho \log z_{t-1} + \varepsilon_t \tag{6}$$

Where ρ is the autoregressive coefficient and ε_t is the innovation or *technological shock* at the time t. We assume that ε_t can be described as independently and identically distributed IID (0,σ).

2.1 Optimality conditions

The optimality conditions may be derived through rearranging the constraints (2) – (5) to denote the capital stock in the next period (time $t + 1$) as a function of current capital stock, leisure and consumption. As a solution to consumers constrained optimization problem, we've found the set of Euler's equations or intertemporal optimality conditions.

$${c_t}^{(1-\phi)\varphi-1} {l_t}^{(1-\phi)(1-\varphi)} = \beta E_t \left\{ {c_{t+1}}^{(1-\phi)\varphi-1} {l_{t+1}}^{(1-\phi)(1-\varphi)} \left[1 - \delta + \alpha z_{t+1} \left(\frac{k_{t+1}}{1-l_{t+1}} \right)^{\alpha-1} \right] \right\} \tag{7}$$

$$\frac{(1-\varphi)}{\varphi} \frac{c_t}{l_t} = (1-\alpha) z_t \left(\frac{k_t}{1-l_t} \right)^{\alpha} \tag{8}$$

$$k_t = z_{t-1} {k_{t-1}}^{\alpha} (1-l_{t-1})^{1-\alpha} + (1-\delta) k_{t-1} - c_{t-1} \tag{9}$$

If we expand the third intertemporal optimality condition (9) back into constraints (2) – (5), we will obtain the first outline of the used model. For purpose of estimation we've introduced another stochastic variable in equation (7), denoting the expectation error among the observations. The equation (7) may be than denoted as follows.

$$c_t^{(1-\phi)\varphi-1} l_t^{(1-\phi)(1-\varphi)} = \beta \left\{ c_{t+1}^{(1-\phi)\varphi-1} l_{t+1}^{(1-\phi)(1-\varphi)} \left[1 - \delta + \alpha z_{t+1} \left(\frac{k_{t+1}}{1 - l_{t+1}} \right)^{\alpha-1} \right] \right\} \cdot e^{\epsilon_t} \quad (10)$$

2.2 The Steady state

Stable equilibrium may not be achieved for every model, but for this particular model (7), (8), (2) – (6) the steady state does exist and we may compute it analytically when we assume that the steady state value of the $\bar{z}$ is equal to 1. From the first intertemporal optimality condition we may find (under the assumption that in steady state the variables are no longer changing) the steady states value of the quotient of capital and labor and subsequently derive steady state values for the other variables.

$$\frac{\bar{k}}{\bar{n}} = \left(\frac{\alpha}{\frac{1}{\beta} - 1 + \delta} \right)^{\frac{1}{1-\alpha}} \quad (11)$$

$$\frac{\bar{y}}{\bar{n}} = \left(\frac{\bar{k}}{\bar{n}} \right)^{\alpha} = \left(\frac{\alpha}{\frac{1}{\beta} - 1 + \delta} \right)^{\frac{\alpha}{1-\alpha}} \quad (12)$$

$$\frac{\bar{\imath}}{\bar{n}} = \delta \frac{\bar{k}}{\bar{n}} = \delta \left(\frac{\alpha}{\frac{1}{\beta} - 1 + \delta} \right)^{\frac{1}{1-\alpha}} \quad (13)$$

$$\frac{\bar{c}}{\bar{n}} = \left(\frac{\alpha}{\frac{1}{\beta} - 1 + \delta} \right)^{\frac{\alpha}{1-\alpha}} - \delta \left(\frac{\alpha}{\frac{1}{\beta} - 1 + \delta} \right)^{\frac{1}{1-\alpha}} \quad (14)$$

$$\bar{n} = \frac{1}{1 + \frac{1}{1-\alpha} \cdot \frac{(1-\varphi)}{\varphi} \cdot \left[1 - \delta \left(\frac{\bar{k}}{\bar{n}} \right)^{\alpha} \right]} \quad (15)$$

$$\bar{l} = 1 - \bar{n} \quad (16)$$

2.3 Log – linearized model

As a final step of this section we've performed logarithmic approximation of the model, and modeled the denoted variables as the deviations from the steady state. Consequently, we've obtained the final form of the model as followed.

$$0 = \tilde{c}_t - \tilde{l}_t - \tilde{z}_t - \alpha \tilde{k}_t - \alpha \tilde{n}_t \quad (17)$$

$$\begin{aligned} 0 = {} & \big((1-\phi)\varphi - 1\big)\tilde{c}_t + (1-\phi)(1-\varphi)\tilde{l}_t - \big((1-\phi)\varphi - 1\big)\tilde{c}_{t+1} - (1-\phi)(1-\varphi)\tilde{l}_{t+1} \\ & - \log\left[\alpha e^{\tilde{z}_{t+1}} \frac{e^{(1-\alpha)(\tilde{n}_{t+1} + \log \bar{n})}}{e^{(1-\alpha)(\tilde{k}_{t+1} + \log \bar{k})}} + (1-\delta) \right] + \log\left[\alpha \frac{e^{(1-\alpha)\log \bar{n}}}{e^{(1-\alpha)\log \bar{k}}} + (1-\delta) \right] \end{aligned} \quad (18)$$

$$0 = \tilde{y}_t - \tilde{z}_t - \alpha \tilde{k}_t - (1-\alpha)\tilde{n}_t \quad (19)$$

$$0 = \log\left(\bar{n} e^{\tilde{n}_t} + \bar{l} e^{\tilde{l}_t} \right) \quad (20)$$

$$0 = \tilde{k}_t - \log\left(\bar{\imath} e^{\tilde{\imath}_t} + (1-\delta)\bar{k} e^{\tilde{k}_{t-1}} \right) + \log\left(\bar{\imath} + (1-\delta)\bar{k} \right) \quad (21)$$

$$0 = \tilde{y}_t - \log\left(\bar{\imath} e^{\tilde{\imath}_t} + \bar{c} e^{\tilde{c}_t} \right) - \log(\bar{\imath} + \bar{c}) \quad (22)$$

$$0 = \tilde{z}_t - \rho \tilde{z}_{t-1} \tag{23}$$

Where arbitrary variable $\tilde{x}_t$ denotes the deviation of that variable from its steady state $\tilde{x}_t = \log(x_t / \bar{x})$.

3 The Data and their transformation

We have used the data available at Slovak statistical office [4] for the estimation of the model, particularly the time series of real gross domestic product (GDP), real government consumption, real export, real import and real final household consumption (FHC). All mentioned series were available quarterly, from 1996Q1 to 2012Q4. Thus our sample consists out of 68 observations.

According to the model, stated in the previous section, we are abstracting from any governmental or foreign influence. In order to represent this abstraction in the data, we had to remove the governmental and the external balance component of the GDP. Thus we've added import and subtracted government spending and export from the GDP. This way we've obtained GDP component representing only the consumption of households and firms. As a second observed variable we have used FHC, which didn't need any additional adjustments.

Even after obtaining the amended GDP and the FHC values, we still weren't able to utilize them to estimate derived model. As we may see, the model (17) – (23) uses logarithms of the deviations from the steady state values of GDP and FHC, instead of the actual values. Therefore we took logarithms of mentioned variables, and applied seasonal filter (Census X12). Afterwards we've filtered the data using *Hodrick – Prescot* (HP) filter, set to the default settings for quarterly data ($\lambda = 1600$). Through mentioned step we intent to capture the dynamic evolution of the steady state values represented in our model as the trend component of the GDP and FHC series. Further the modeled logarithms of the deviations from the steady state values were represented as the cyclical component, obtained from the HP filter.

4 Model estimation

For the estimation procedure we've used the *Metropolis – Hastings Markov chain Monte Carlo* (MH-MCMC) algorithm, preprogrammed for the software environment of Octave in so called *Dynare* [2] computational tool.

MH-MCMC, which may be utilized for *Bayesian estimation* of the parameters of particular model, requests prior information about the distribution of parameters, hereinafter denoted as the "priors". In order to obtain such distributions we made an assumption partially based on the information contained in literature, such as Szomolányi, Lukáčik and Lukáčiková [5] and partially on our own insight. After a series of trials and errors we have identified following set of priors (Table 1) for which we have computed the corresponding steady states, which we've obtained from the relations (11) – (16).

Parameters			Steady state values	
Notion	**Mean**	**Distribution**	**Notion**	**Value**
A	0.56	Beta	$\bar{n}$	0.771729886
B	0.95	Beta	$\bar{l}$	0.228270114
Δ	0.005	Beta	$\bar{c}$	13.2644347
Φ	0.3	Beta	$\bar{i}$	0.677354508
Φ	1.27	Beta	$\bar{y}$	13.94178921
P	0.95	Beta	$\bar{k}$	135.4709016
σ_ε	0.011	Inv_gamma		
σ_ε	0.05	Inv_gamma		

Table 1 Priors and corresponding Steady state

In the final run of the estimation we have used the following settings of the MH-MCMC algorithm. We have set the algorithm to perform 50 000 replications of the Monte Carlo simulation, the number of Markov chains blocks were 4, the algorithm's drop rate was set to 0.45 and the jump scale to 0.25. We have used the Sims computational method. As indicated above, we have also included the observed series of deviation of GDP and FHC from theirs steady state values, which contained 68 observations (1996Q1 – 2012Q4). The results of the last run of the Bayesian (MH-MCMC) algorithm are presented in the Table 2.

Parameters	Prior mean	Post. mean	Conf. inter.		Prior	Post. Dev.
α	0.56	0.56	0.5584	0.5617	Beta	0.001
β	0.95	0.9498	0.9482	0.9515	Beta	0.001
φ	0.3	0.3138	0.2812	0.3465	Beta	0.02
ϕ	1.27	1.2825	1.2486	1.3167	Beta	0.02
δ	0.005	0.0051	0.0033	0.0068	Beta	0.001
ρ	0.95	0.9487	0.9327	0.965	Beta	0.01
σ_ε	0.011	0.0112	0.01	0.0123	Inv_gamma	0.001
σ_ε	0.05	0.0488	0.0473	0.0503	Inv_gamma	0.001

Table 2 Estimation results

Since all of the prior means are contained in the estimated 90% confidential intervals, we can conclude that set priors were chosen suitably and we will proceed with describing them instead of the posteriors. The parameter α with value of approximately 0.56 says that the production technology is more demanding on the stock of capital than labor force. The value of β indicates that the consumption of identical good consumed one year from now, would yield only 81.45% utility compared to the current consumption. The φ parameter with the value 0.3 indicates that household strongly prefers leisure compared to consumption. According to the estimated value of ϕ the overall intertemporal substitution rate is equal to -0.27. Estimated parameter δ, at the level of 0.005 indicates that the stock of capital is mainly composed of long-term durable items, which fully depreciate after more than 50 years. We are aware that estimated value of δ is merely a tenth of standardly used value for depreciation, but for reasonable interpretation of steady state values we weren't able to use greater values. The last parameter ρ at the level of 0.95 signalizes a strong persistence of the effect of possible technological shock.

When looking closer at estimated steady state values of various variables we see another picture. First it is important to point out that these aren't the actual values of the steady state. Due to the application of the HP filter in the section 3 we set the steady state values of the *technological progress* equal to the evolving trend component of the HP filter at each given moment. Therefore the steady state values obtained from the estimated model are deviations of the evolution of selected variable relative to the evolution of the *technological progress*. With this in mind we start to interpret particular parameters of the model. The steady state of labor indicates that approximately 77.17% of the time in given moment is spent with labor. That leaves 22.83% of the time for leisure. The steady state value of the consumption indicates that the actual consumption is oscillating around 95.14% of the overall production. Analogically the steady state of investment is relatively at 4.86% of the GDP. Such result corresponds with relatively low rate of depreciation, which implies high durability of capital goods and thus only very small need for additional investment. Therefore almost all of the production of the country may be consumed by the household. Finally the steady state value of the capital stock is oscillating around the 9.72 times the level of quarterly GDP of Slovakia. Thus capital stock of Slovakia may in steady state correspond to the 2.43 multiple of annual GDP, which is in line with work of Kubíček [3]

5 The impact of the technological progress

In the next step we have used the parameters, confirmed through the Bayesian estimation, to analyze the effect of *technological progress* on Slovak economy. In order to do so we have performed stochastic simulations with the estimated model. We have set the number of replications to 100 000. This way we have obtained the impulse response function (IRF), which depicts the adjustment of particular variable to the effect of a given shock. In our case we have set the number of observations, during which we will be tracking the IRF to 200 periods. Consequently, we have obtained the evolution of IRF for all of the variables present in the model, visible at Figure 1.

Conclusively, we can interpret the obtained results. The highest impact of the *technological shock* or an innovation to productivity can be seen on investment. Slovak households temporarily substitute consumption with investment, while they are motivated with higher future returns. Although the effect is only short-lived it is definitely of highest magnitude. Through the investment we can also explain the effect on the stock of capital, which is positive as well, but keeps longer persistence than investment. Mentioned phenomenon can be explained by the multiplicative effect of both investment and *technological progress*, which in turn are responsible for higher output, of which greater share was transformed into capital. Even after the retraction of the investment, the original investment keeps the growth of the capital positive. Similar is the interpretation of consumption, which rises moderately due to the growth in overall output.

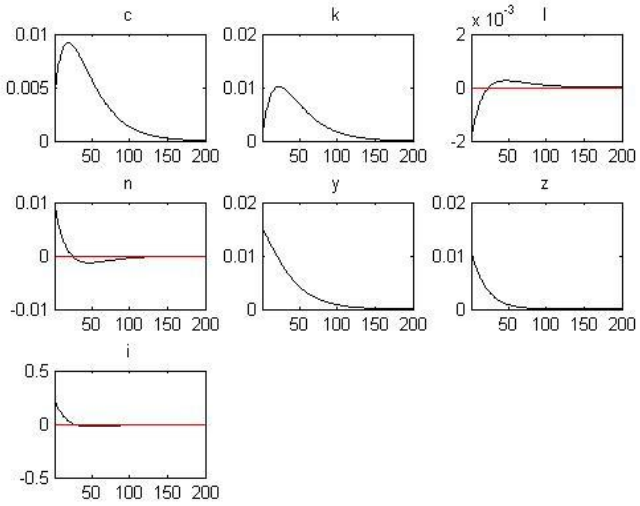


Figure 1 Impulse response function off Technological shock

Speaking of the overall output, rise of its immediate value is firstly fueled by the instant growth of productivity, which translates into rise of output both directly, through the higher output at the same level of inputs and indirectly, through the immediate growth in the rate of consumers substation, for whom is the labor (and with it the future consumption) far more interesting than leisure. When immediate substation effect wares off the overall output is still fueled with the increase of capital (as explained before). The leisure naturally evolves precisely in the opposite direction than labor does, but with different magnitude. And finally the persistence of the technical shock in the technical progress variable is determined solely by the value of autoregressive coefficient ρ.

6 Conclusion

In the presented paper we have described application of the RBC model in the conditions of Slovakia, as a tool to evaluate the impact of arbitrary *technological shock* on the selected variables of the country. The immediate impact of 1% positive shock on the total output is rise in approximately 1.382%. Similarly the 1% positive *technological shock* affects all other model variables. In case of FHC it results in approximately 0.4% increase. The greatest increase was recorded for the investment, where the magnitude is approximately 20.709%. In contrary the lowest increase was identified for the stock of capital, at the approximate magnitude of 0.1%. As mentioned in section 5 the impact on the labor market demonstrates the substitution effect among the time spent in work and leisure. The magnitude of 1% earlier mentioned shock results in 0.873% increase in labor input and 0.173% decrease in the leisure. The direct effect on the economy of the *technological shock* will diminish over time and it is significant for at least 47 periods (which corresponds with approximately 12 years), after which the value of deviation from the steady state of the technological progress is less than 0.001. The momentum of the shock lasts for another 49 periods (which corresponds with approximately 13 years), after which the deviation from the steady state of the GDP is insignificant (less than 0.001).

Acknowledgements

This paper is supported by the Grant Agency of Slovak Republic - VEGA, grant no. 1/0595/11 "Analysis of Business Cycles in the Economies of the Euro Area (with Regards to the Specifics of the Slovak Economy) Using Econometric and Optimization Methods"

References

[1] DeJong, D. and Dave, C.: *Structural Macroeconometrics.* Princeton University Press, Princeton, 2007.
[2] Griffoli, T. M.: *DYNARE User Guide: An introduction to the solution & estimation of DSGE models.* [online] 2011.
[3] Kubíček, .J: Dlouhodobý ekonomický růst, produktivita a konvergence. Karlova Univerzita, Praha, 2005.
[4] Slovak Statistical Office: Source of data [online] <http://www.statistics.sk/pls/elisw/vbd>
[5] Szomolanyi, K., Lukáčik, M. and Lukáčiková, A.: Estimate of Parameters of Production Function of Slovak Economy: Econometric Analysis of Nominal Wages. *Quantitative Methods in Economics (Multiple Criteria Decision Making XVI)* **16** (2012), 210-215.

Implementation of logistic regression into technical analysis

Juraj Hruška[1]

Abstract. Most of the investment strategies based on technical analysis are based on the principle that your strategy should be as easy as it is possible; in order to simplify your decision making. Goal of this paper is to use more sophisticated methods to combine the signals from indicators of technical analysis to create advanced form of investing strategy, which will be sustainable in long run. This can be ensured by self-correcting mechanisms build-in the strategy itself. Econometrical methods will be used to determine whether some kind of indicator has it relevance on the chosen type of asset. All input variables will be time series of dummy variables showing whether the indicator is suggesting taking a long position or not and of course their lags. Explained variable will be the successful trade (the price movement upwards is greater then spread and commissions). For this kind of purposes logistic regression seems to be essential, which is widely used in credit scoring. Basically the problem whether to invest or not is the same issue as whether to give a customer a loan or not. The only difference will be in the type of data. Credit scoring use mostly panel data, however it will be handled solely with time series in this paper.
Keywords: logistic regression, technical analysis, moving averages, automated trading.

JEL Classification: G11
AMS Classification: 62J02

1 Introduction

Technical analysis is one of the most popular and most efficient tools in trading business. It uses quantum of price-based information and indicators, which should bring at least trustworthy hints when to buy or sell certain asset. In this paper I will focus on one specific indicator of technical analysis; moving average. It is one of the easiest and most basic indicators. On the other hand its performance in the past has proved its significance. Many traders build their trading strategies on this indicator, or on the indicators derived from it, such as Moving average convergence divergence (MACD) or more complicated types of moving averages and their combinations. I have chosen the simple moving averages for my analysis. I used logistic regression to determine which combinations of moving averages give the most suitable signals for automated trading strategies.

Moving averages are basic indicator, but they are still used as a benchmark in many research analysis. Gradojevic and Gençay [4] analyze fuzzy technical indicators and moving averages implemented on the EUR/USD exchange rate. Pavlov [7] used moving averages in bootstrap simulations as a tool in portfolio selection process. Shintani, Yabu and Nagakura[9] have used indicators based on moving averages as predictors in time series modeling and testing hypothesis that indicators are falsely supported and regression is spurious.

2 Methodology and data

I based this paper on historically confirmed assumptions about indicators. I did my analysis on time series of 5 minute close spot exchange rate of EUR/USD from 15.10.2012 till 5.5.2013 (all data available from Bloomberg database with such high frequency). And I validated the results on the data since 15.4.2013 to 10.5.2013. Intervals are overlapping due to necessity of creating lagged variables and rather small original validating sample. Consequently, I tested all possible combinations of two simple moving averages.

Explained variable for purposes of this paper, so called successful trade i.e. system, was able to close the trade with take profit order with profit level 0,1% without being canceled out by stop-loss order. If any of the levels had been triggered, system waited for 100 minutes and then the trade was automatically considered as unsuccessful. Logistic regression was used to test, how combinations of these moving averages can explain the trade being successful.

[1] Masaryk university, Faculty of Economics and administration, Finance department, Lipová 4, 60300,206887@mail.muni.cz

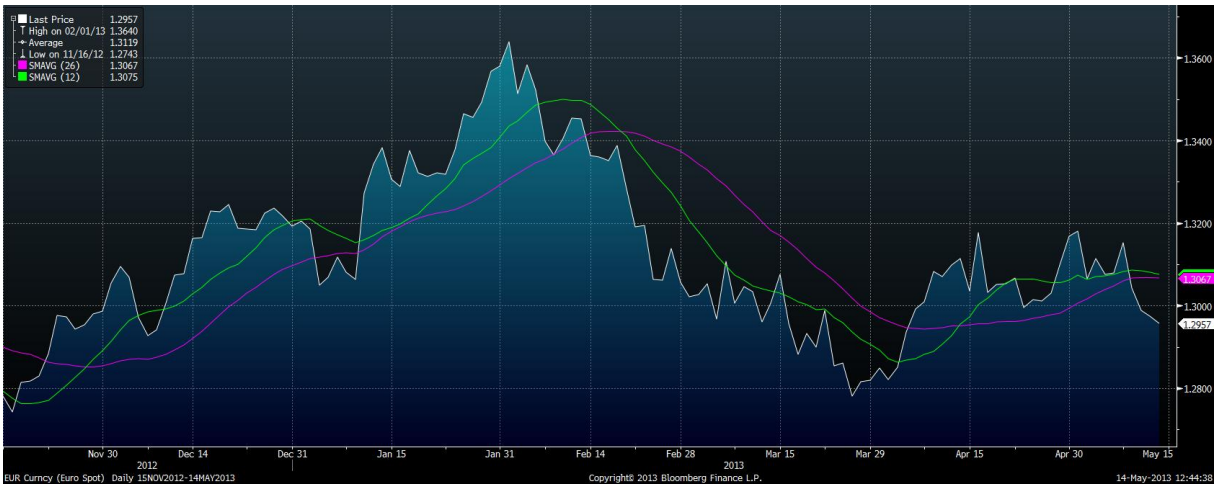


Figure 1 EUR/USD exchange rate development (15.10.2012-5.5.2013) from Bloomberg

Moving averages are used in combination with the line of spot prices. When the spot price crosses the line of moving average upwards, it is considered as a signal to take a long position. On the other hand when the line of the spot price crosses the moving average downwards, it is a signal to take a short position or to close a long position. The same logic holds in case we do not use spot price but another moving average (it should be shorter than the first one, or the positions should change).

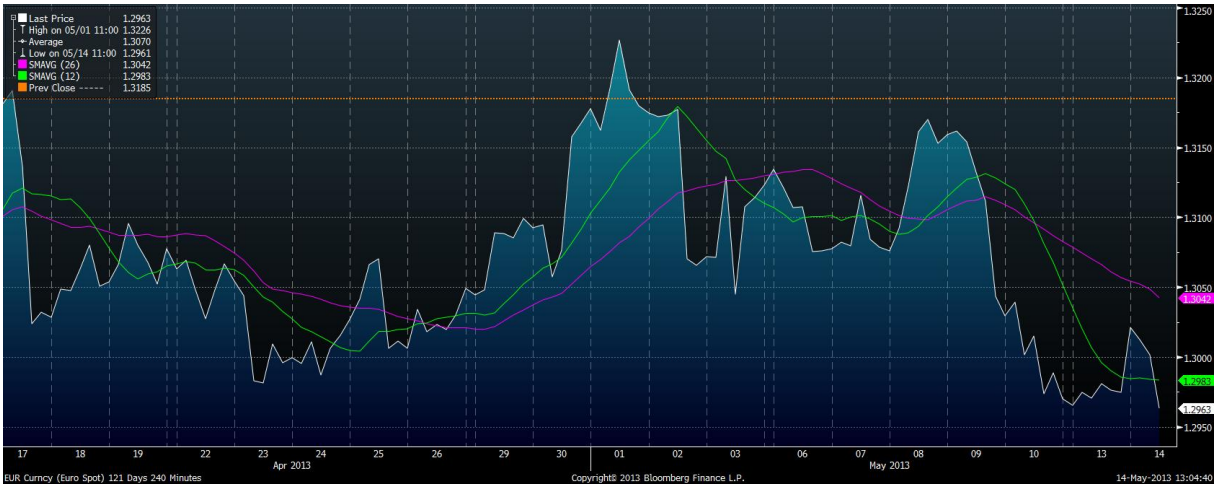


Figure 2 EUR/USD exchange rate development (1.4.2013-15.5.2013)-validating data from Bloomberg

Moving average is calculated as the average of current price of asset and several spot prices directly following each other:

$$MA\{n\} = \frac{1}{n}\sum_{i=0}^{n} S_{t-i} \tag{1}$$

Usually the Combination of 25 and 12 prices and 12 and 5 prices are considered as the most efficient ones according to Brada[1]. But it always depends on the frequency of evaluated data. For closing price of the day different combinations of moving averages could be used, than for the tick prices. Also the number of signals is an important factor when choosing a type of data we want to analyze. If we want to use them for automated trading, it would be efficient to use averages that produce more signals. This model tested all combinations of moving averages up to length of 40.

Logistic regression was used for analysis of the successful trades by the moving averages as explaining variables. It is based on maximum likelihood function and does not have such strong requirements for the variables as Ordinary least square estimation. The general formula of the model is:

$$y_i = \beta_1 + \beta_2 X + \epsilon_i \tag{2}$$

Where, y_i is the dependent variable and X is the matrix of explaining variables and $\beta_!$ is intercept and β_2 is a vector of coefficient of explaining variables and ϵ_i is residual.

Logistic function is derived from the formula of normal standard distribution and has form of a maximum likelihood function:

$$f(y_i|X;\beta,\sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)}} exp\left\{-\frac{(y_i-\beta_1-\beta_2X)^2}{2\sigma^2}\right\} \tag{3}$$

The maximum likelihood function is transformed into a log-likelihood function, with unknown parameters β and σ^2, which we are trying to estimate:

$$logL(y_i|X;\beta,\sigma^2) = -\frac{N}{2}log(2\pi\sigma^2) - \sum_{i=1}^{n}\frac{(y_i-\beta_i-\beta_2X)^2}{2\sigma^2} \tag{4}$$

After differentiating the log-likelihood function with respect to all parameters and equaling them to 0, we get estimations of all unknown coefficients in form:

$$\frac{\partial logL(y_i|X;\beta,\sigma^2)}{\partial\beta} = 0 \tag{5}$$

Logistic regression is mostly used for binary choice models. Model can be enhanced to the form, where the explaining variable would express also signal to take short position. This would certainly require some multinomial form of the model. In practice we are modeling a probability that the explaining variable is 1. Mostly the cut-off probability is 0.5, but it can be modified due to requirements to the significance of the model. Logistic regression transforms probabilities to *logits*.

$$logit(p_i) = ln\left(\frac{p_i}{1-p_i}\right) \tag{6}$$

Where p_i is the probability that observation *i* will be evaluated as 1. So we are transforming the explained variable to the *logits*.

$$logit(p_i) = \beta_1 + \beta_2X \tag{7}$$

These parameters are estimated by the Maximum likelihood estimation. Probabilities would be obtained from the function:

$$p_i = \frac{1}{1+e^{-\beta_1-\beta_2X}} \tag{8}$$

More about logistic regression theory can be found in Heij [3].

3 Definition of the model

At first I created explaining variable a "successful trade" (Trade) which is the situation, when profit from acquisition in time *t* is higher than 0,1% in following 20 observations (100 minutes), without triggering the stop-loss set on the level of 99% of Ask price. Then I calculated moving averages up to length of 40. Then I used them to create explaining variables in a form of dummy variables that should signalize profit if we take a long position in the asset. Theory says that moving averages suggest submitting buying order if the faster moving average crosses the slower moving average upwards. After the comparison of all possible combinations 780 explaining variables were created. Then 5 lags of every variable were added, which sum up to over 5000 regressors altogether. That is rather extensive amount of variables for regression, but if I wanted to identify as much successful trades as possible, I simply could not omit any variable. The longer moving averages are not that interesting for the model, because since certain level they all behave approximately the same and give the same information. One of the solutions might be adding moving averages with higher sequence.
Model can be extended with different types of indicators of technical analysis such as, Moving average Convergence/Divergence, Relative Strength Index, CCI, or some trend lines or any kind of indicator based on the volume of trading.

The resulting model after, ignoring all insignificant variables consisted of around 250 predicators. The model can be accepted according to the global tests.

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	939.6593	276	<.0001
Score	916.9869	276	<.0001
Wald	891.9390	276	<.0001

Table 1 Significance statistics of model

According to the summary statistics of the model, it can be considered as rather weak model, because its Somer's D, calculated as ratio of concordant and disconcordant pairs, has small value and the value of the area under the ROC curve is small as well, compared to the area under the curve of a random model. The area under the curve is only 0.5973 which means only 19.46 % of cases are explained by chosen variables. Transformation of the explained might be helpful in a way, that it would contain more trend profits and less random profits.

ROC Association Statistics							
	Mann-Whitney						
ROC Model	**Area**	**Standard Error**	**95% Wald Confidence Limits**		**Somers' D (Gini)**	**Gamma**	**Tau-a**
Model	0.5973	0.00325	0.5909	0.6036	0.1946	0.1946	0.0691
ROC1	0.5000	0	0.5000	0.5000	0	.	0

Table 2 ROC statistics for the model with main data

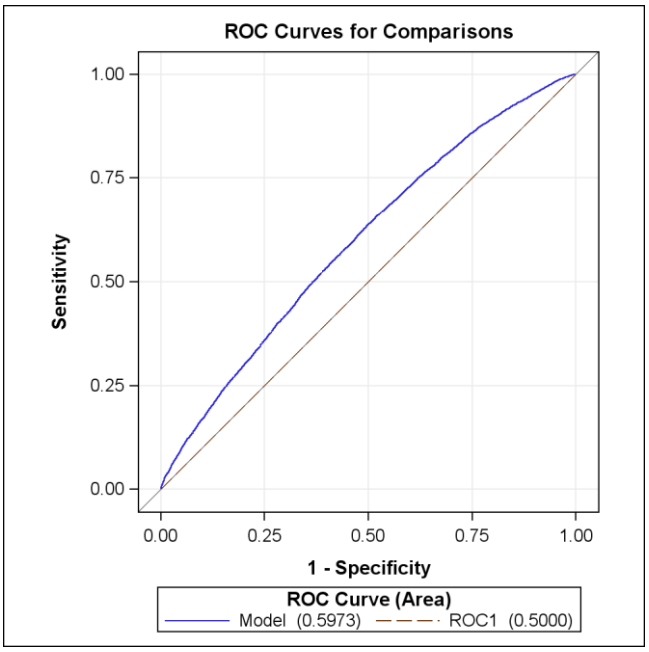


Figure 3 ROC curve for the model with main data

Nevertheless, after analyzing the frequency table of Trade observed values and fitted values, we can see that in spite of the fact that model found only a small fraction of successful trades; it was quite effective in preferring profitable trades to trades with loss. The rate of success was nearly 65%. If we kept investing according to this model we would gain 2.94% after selected period of time.

Table of Trade by Model			
	Model-fitted values		
Trade	0	1	Total
0	31876 76.85 99.95 76.94	16 0.04 0.05 35.56	31892 76.89
1	9556 23.04 99.70 23.06	29 0.07 0.30 64.44	9585 23.11
Total	41432 99.89	45 0.11	41477 100.00

Table 3 Frequency table of Trade and fitted values

If we decrease the cut-off ratio of the log-likelihood function from 0.5 to 0.45 the number of trades will increase to 170, which is nearly four times as much. However, the successfulness of trades will also fall to 53.22%. In any of these cases the unsuccessful trade was created by triggering stop-loss barrier. On the other hand the gain after this period was around 3.59 %, which is even more than in the previous model. That is caused by the fact, that many trades did not cross the profit level 0.1%, but they were still profitable.

After testing the model on validating data set of EUR/USD exchange rate, slightly better results were obtained.

ROC Association Statistics							
	Mann-Whitney						
ROC Model	Area	Standard Error	95% Wald Confidence Limits		Somers' D (Gini)	Gamma	Tau-a
Model	0.6240	0.00675	0.6108	0.6372	0.2480	0.2480	0.0883
ROC1	0.5000	0	0.5000	0.5000	0	.	0

Table 4 ROC statistics for the model with validating data

With area under ROC curve 0.624 and Gini coefficient 0.248, calculated model seems to have medium quality on validating data. That suggests that the model is not overlearnt even with such a number of explaining variables.

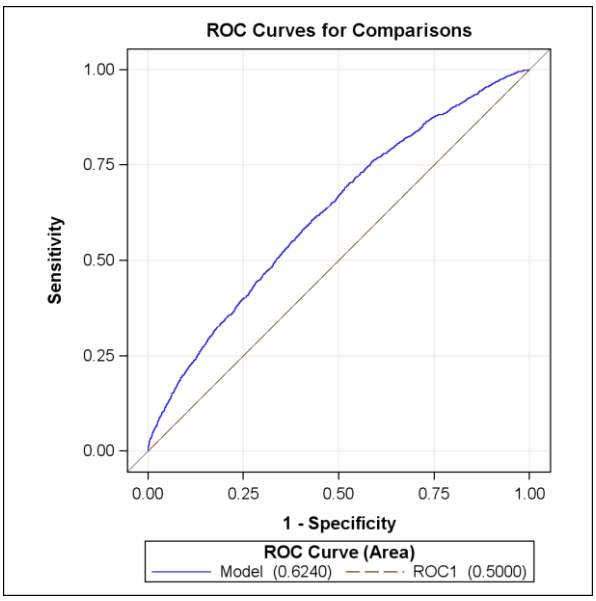


Figure 4 ROC curve for the model with validating data

The frequency table shows that even the effectiveness of predicating trades was somewhat better (65.63% of the trades were correctly selected). These results are for the model with cut-off with basic level of 0.5. An interesting fact is that more trades were picked from quarter amount of observations. The overall of profit after the validating period (month and half) was 4.287%. With the cut-off level 0.45, the model was effective on 53 %, so it can be concluded that it is more efficient to stick with stricter model. Profit was only 3.684% for the more benevolent model with cut-off 0.45.

Table of Trade by Model			
	Model-fitted values		
Trade	0	1	Total
0	7238 76.61 99.70 77.13	22 0.23 0.30 34.38	7260 76.84
1	2146 22.71 98.08 22.87	42 0.44 1.92 65.63	2188 23.16
Total	9384 99.32	64 0.68	9448 100.00

Table 5 Frequency table of Trade and fitted values - validation data

4 Conclusion

After these calculations I can proclaim that logistic regression can be successfully used as a decision-making tool in automated trading. Profits gained by this strategy are higher than using just ordinary moving averages strategies. My model produced only few signals, but successfully avoided many of false signals, that would be created by moving averages themselves. These models were weak, but still profitable. They can be improved by using more indicators as explaining variables and even moving averages can be exchanged for the exponential moving averages. In view of automated trading this model was creating only few signals, compared to what would be necessary for application in such a field; it would also require higher frequency of observations.

Acknowledgements

Support of Masaryk University within the project MUNI/A/0753/2012 "European financial systems 2013" (Student Project Grant at MU, Faculty of Economics and Administration, Department of Finance) is gratefully acknowledged.

References

[1] Brada, J. *Technická analýza*. Vysoká škola ekonomická v Praze, Praha, 2000.
[2] Blommberg database – data source
[3] Heij, Ch. *Econometric methods with applications in business and economics*. Oxford University Press, Oxford 2004, 438-462.
[4] Gradojevic, N., & Gençay, R. *Fuzzy logic, trading uncertainty and technical trading*. Journal of Banking and Finance, *2013*, 578-586.
[5] Musílek, P.: *Trhy cenných papírů*.Ekopress, Praha, 2002, 327-338.
[6] Luo, J., Si, Y. -., & Fong, S. *Trend following with float-encoding genetic algorithm.* Paper presented at the *7th International Conference on Digital Information Management, ICDIM 2012,* 173-176.
[7] Pavlov, V., & Hurn, S. *Testing the profitability of moving-average rules as a portfolio selection strategy*. *Pacific Basin* Finance Journal, 2012, 825-842.
[8] Sekerka, B. *Matematické a statistické metody ve financování, cenných papírech a pojištění*. Profess Consulting, Praha, 2002, 198-202.
[9] Shintani, M., Yabu, T., & Nagakura, D. *Spurious regressions in technical trading*. Journal of Econometrics, 2012, 301-309.

Comparison of the AHP and ANP approach to Investment Decision Making

Eva Chalúpková[1], Jiří Franek[2]

Abstract. The decision making can be performed according to perspectives, which may also have a conflicting nature. Models that are helpful in dealing with such problems can be found in decomposition multiple attribute decision making methods (MADM). In this paper the authors are dealing with an investment decision making problem that can be divided into hierarchy of levels thus MADM methods the AHP (Analytic Hierarchy Process) and the ANP (Analytic Network Process) are applied. The aim of the paper is to compare both approaches their results and to analyse which of them has a better fit for actual investment decision making from entrepreneurs perspective. The sensitivity analysis is performed on results from AHP to assess sensitivity of alternatives ranking and weights values. The ANP method includes also the interrelationships among selected criteria. Results are compared.

Keywords: leasing, analytic hierarchy process, analytic network process, MADM

JEL Classification: M10, C44
AMS Classification: 90B50, 90C29, 91B06

1 Introduction

Decision making belongs to day to day activity of entrepreneurs. It represents a process in which an individual or a group selects the best or optimal alternative with regard to a set of attributes (criteria). Capital investment decisions are among the most important decision making activities in business. Most of these capital investments are focused towards renewal of fixtures or expansion of business tangible and intangible assets [3]. Investment decisions have a profound and frequently long term impact on the business. It is therefore necessary to consider more than few decision criteria. That is why this multiple attribute decision making methods can be used. This paper is focused on the application of multi level decomposition decision making methods of Analytic Hierarchy Process (AHP) and Analytic Network Process (ANP) developed by [7], [8]. These methods were largely studied and analysed in various business decision making problems [2], [6], [10]. The goal of this paper is to present both methodologies (AHP and ANP) in comparison manner and to assess their appropriate utilization for entrepreneurs. The AHP decision making process is extended by sensitivity analysis of its results.

2 Multi level decomposition decision making methods

The problem of multi-criteria evaluation of alternatives is foremost a task of finding the best (optimal, compromise) alternative, raking of the alternatives from the best to the worst conceivable, alignment of alternatives to hierarchical clusters, separation of alternatives into two groups of acceptable and not acceptable, determination of the effectives (non dominated, pareto) set of alternatives or exclusion of the ineffective alternative [12]. In short it is the optimization problem. The rank of alternatives and selection of the optimal one is based on weighted sum criteria (total weighted utility) of the alternative that can be calculated using to the eq. (1)

$$U(a_i) = \frac{\sum_{j=1}^{m} x_{i,j}.v_j}{\sum_{j=1}^{m} v_j}, \tag{1}$$

where v_j represents non-normalized weight of the j^{th} criterion, and stands for the sum of all (non-normalized) criteria weights, x_{ij} represents the evaluation of the i^{th} alternative according to the j^{th} criterion. Then for the weighted sum criteria of normalized weights following formula (2) can be applied:

$$U\left(a_i\right) = \sum_{j=1}^{m} w_j.x_{i,j}, \tag{2}$$

[1] VSB-Technical University of Ostrava/Faculty of Economics, Department of Accounting, Sokolska tř. 33, 701 21 Ostrava, eva.chalupkova@vsb.cz.
[2] VSB-Technical University of Ostrava/Faculty of Economics, Department of Business Administration, Sokolska tř. 33, 701 21 Ostrava, jiri.franek@vsb.cz.

where the w_j represents the normalized weight of the j^{th} criterion.

The advantages of multi level decomposition decision making methods can be found in the decision maker's ability to evaluate each alternative using a large number of criteria. Methods can be accordingly classified based on the information about the criteria or alternatives preferences into following categories: methods with cardinal information, methods with ordinal information about criteria preferences, methods with nominal information [5]. Multi level decomposition methods are among the most convenient when it comes to evaluation of finite number of alternatives. One of the basic and most widely used methods is the analytic hierarchy process (AHP) [9].

2.1 Analytic Hierarchy Process (AHP)

The AHP solves multi-criteria decision making problem based on hierarchy. Generally the hierarchy has three levels: the goal, criteria and alternatives. Criteria can be broken down to sub-criteria to make lower level. The theoretical procedure of the AHP method consists of four steps: hierarchy design, identification of priorities, combination and evaluation. The method itself is based on Saaty method of pair-wise comparison [7]. The actual weight estimation can be calculated according to several approaches. The most frequent and appropriate are the Row Geometric Mean Method (RGMM) and Eigenvalue method (EVM). In order to provide sufficient close results in most situations, [9] suggested the approach of the geometric mean of axis' row. The geometric consistency index suggested is to measure the individual consistency of judgment matrices. For an approximation method that provides sufficiently close results in most situations. In the RGMM the weights w_j can be obtained through an algorithm based on the geometric mean method (method of least logarithmic squares) under the same necessary condition then the solution is a normalized geometrical mean of the matrix as follows

$$w_j = \left[\prod_{j=1}^{i} s_{ij}\right]^{\frac{1}{i}} \Bigg/ \sum_{j=1}^{i} \left[\prod_{j=1}^{i} s_{ij}\right]^{\frac{1}{i}}. \tag{3}$$

The geometrical mean can be calculated using MS Excel function *GEOMEAN*. This function will be employed for calculations in the application part. On the other hand Saaty applied the main of eigenvalue S as a desirable vectors priority of ω, where the linear system (4) determines the vector as it is stated below,

$$S\omega = \lambda\omega, \;\; e^T\omega = 1, \tag{4}$$

where λ is the principal of eigenvalue towards A. If the decision maker is constant, the value of λ should be equal to the n. We can also say that the value of λ is greater to the value of n ($\lambda > n$). The estimation of eigenvalues for the main vector inconsistent matrix can be done by normalizing the number of lines of each element. Various researchers have proved that for small deviations around consistent ratio ω_i/ω_j of the EVM method gives a good approximation of the priority vector. The difference between above mentioned methods is evident when estimated utilities of alternatives are closer to each other and small change in values can influence the ranking, then it is recommended to calculate and compare both methods. Results from the EVM can be considered more precise. However, the evaluation requires a certain level of matrix consistency, i.e. that the elements are linear independent. That can be assessed employing consistency ratio C.R. as follows

$$C.R. = \frac{C.I.}{R.I.} = \frac{\frac{\lambda_{\max} - m}{m-1}}{R.I.}, \quad \text{where } \lambda_{\max} = \sum_{j=1}^{m} \frac{(\mathbf{S}\cdot\mathbf{v})_j}{m \cdot v_j} \tag{5}$$

where *C.I.* is the consistency index, λ_{max} is the highest eigenvalue of the matrix and m represents the number of independent rows of the matrix. The λ_{max} can be calculated using the matrix **S** that represents pair-wise comparison matrix and **v** means the matrix eigenvector. *R.I.* represents the random index that can be obtained from i.e. [1]. Solution of AHP can be found using supermatrices as follows

$$W = \begin{matrix} \\ goal \\ criteria \\ alternatives \end{matrix} \begin{bmatrix} 0 & 0 & 0 \\ W_{21} & 0 & 0 \\ 0 & W_{32} & I \end{bmatrix}, \tag{6}$$

where $\boldsymbol{W_{21}}$ is the matrix of criteria's pair-wise comparisons, $\boldsymbol{W_{32}}$ is the matrix of alternatives and criteria pair-wise comparisons and $\boldsymbol{I}$ is the identity matrix. We take these comparisons as independent. This leads us to the more complex ANP method where the decision maker has to consider criteria and alternatives interdependency.

2.2 Analytic Network Process (ANP)

The ANP structures the problem related to options in reverse logistics in a hierarchical form. With the ANP, the interdependencies among criteria, sub-criteria and determinants for the options can be considered. The original

analytical network process (ANP) was proposed in [8]. ANP is the extension of analytic hierarchy process (AHP) and is a more general form of AHP. Many decision problems cannot be structured hierarchically because they involve the interaction and dependence of higher level elements on low level elements. Saaty in [8] applied ANP to handle dependence among criteria and alternatives without assuming independent decision criteria. The ANP feedback approach replaces hierarchies with networks, and emphasizes interdependent relationships among various decision-making and also interdependencies among the decision criteria and allows more systematic analysis. This model can be defined by a supermatrix (7) see the differences in relation to (6)

$$W = \begin{matrix} goal \\ criteria \\ sub-criteria \\ alternatives \end{matrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ W_{21} & 0 & 0 & 0 \\ 0 & W_{32} & W_{33} & W_{34} \\ 0 & 0 & W_{43} & W_{44} \end{bmatrix}, \tag{7}$$

where $\boldsymbol{W_{ij}}$ represents all possible and logical pair-wise comparison weights. The actual method of ANP can be characterized in three basic steps: (i) composition of the initial supermatrix $\boldsymbol{W}$, where the normalized weights w_{ij} are put in columns; (ii) then the initial supermatrix is transformed into weighted supermatrix $\overline{\boldsymbol{W}}$ where

$$\sum_{i=1}^{j} w_{ij} = 1, \tag{8}$$

that derives from the need to find the convergent solution and the results would then represent the global criteria weights; (iii) the third step involves a calculation of the limited (final) supermatrix $\overline{\boldsymbol{W}}^{\infty}$ that can be processed in the case of the non cyclic weighted matrix according to [12] as follows

$$\overline{W}^{\infty} = \lim_{k \to \infty} \overline{W}^{k}, \tag{9}$$

$$\overline{W}^{N} = \frac{1}{N} \sum_{k}^{N} \overline{W^{k}}, \tag{10}$$

where $\overline{W}^{\infty}$ is the limited supermatrix, $\overline{W}^{k}$ is the supermatrix without a cycle powered *k*-times (9). In the case of the cyclic matrix the formula can described as in (10).

2.3 Sensitivity analysis of investment project alternatives

Based on acquired values of the weighted sum criterion $U(a_i)$ of particular alternatives, and having in mind the rule of the highest value, the best alternative can be found. According to results all alternatives can be organized by different measures (e.g. best to worst alternative, etc.). Then it is possible to assess created ranking with various exclusion methods and select a number of alternatives to be considered further. To assess the stability of the rank of alternatives the sensitivity analysis of weights changes of alternatives can be used. With the aforementioned information in mind (the rank of alternatives) that was calculated by employing weighted sum criterion $U(a_i)$ with formulas (1) and (2), the following sensitivity analysis will be dealing with evaluation of alternatives according to weights estimated for this criterion [11]. The weighted sum criterion $U(a_i)$ has to be calculated for particular alternatives. The aim is to find a limited value that would cause a change in the ranking of alternatives *m* and *n*. Actual procedure of the sensitivity analysis that was used in this study can be found in [12] pp 48-49.

3 Capital investment decision making using AHP and ANP and its comparison

Following chapter deals with the application of both aforementioned methods on a capital investment example. The hierarchical decision-making process is drafted in the following Figure 1. It has 4 levels (goal, criteria, sub-criteria, alternatives). This hierarchy has two loops in the third level. In the AHP those loops were not used.

The decision making process with AHP/ANP has been done through following procedure:

1. Selection of alternatives for capital investment financing (a vehicle - fixed asset);
2. Financial and non-financial criteria perspectives were taken into consideration (C_1 and C_2);
3. Definition of sub-criteria (f_1,...,f_6);
4. Pair-wise comparisons of criteria where in the purchase of fixed assets the costs of whole transaction were considered. In this case the financial criteria are preferred before the non-financial.
5. Pair-wise comparison of sub-criteria among themselves.

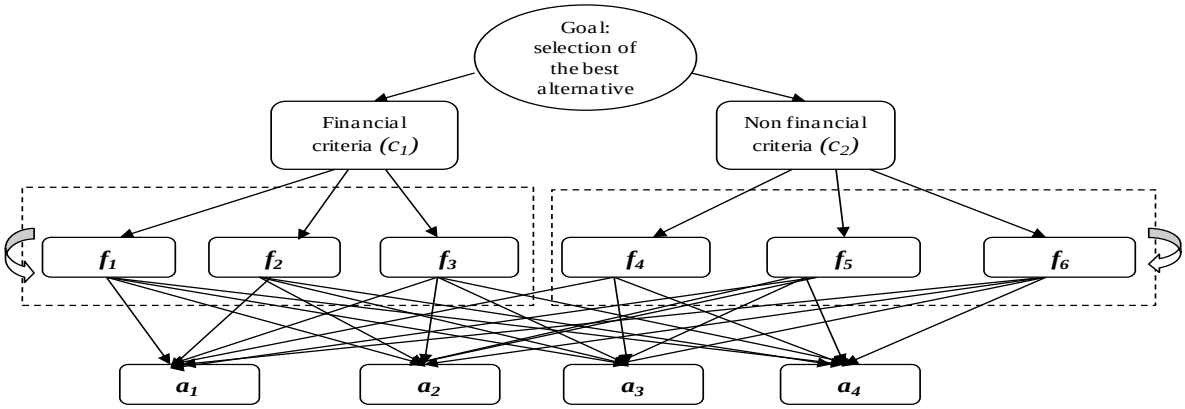


Figure 1 AHP/ANP decision making hierarchy

Data were gathered from several financial products that are available for assessment. These products serve as alternatives. The data are shown in the Table 1.

Criteria/ Alternatives	Financial criteria (c_1)			Non-financial criteria (c_2)		
	Max	Min	Min	Max	Min	Min
	Total cost [EUR]	Monthly payment [EUR]	Other fees [EUR]	Fixture period [months]	Administration	Office distance
Consumer loan 1 (a_1)	24 000	200	698	120	12	4
Consumer loan 2 (a_2)	18 000	288	9 010	36	13	9
Capital lease 1 (a_3)	21 600	285	4 600	60	5	8
Capital lease 2 (a_4)	18 000	250	6 800	48	5	6
Criteria symbol	f_1	f_2	f_3	f_4	f_5	f_6

Table 1 Capital investment decision data of alternatives

Table 2 shows the unweighted ANP supermatrix that consists of weights estimated by pair-wise comparisons and calculated using (3). All consistencies were within limits.

	goal	c_1	c_2	f_1	f_2	f_3	f_4	f_5	f_6	a_1	a_2	a_3	a_4
goal	0	0	0	0	0	0	0	0	0	0	0	0	0
c_1	**0,875**	**1**	0	0	0	0	0	0	0	0	0	0	0
c_2	**0,125**	0	**1**	0	0	0	0	0	0	0	0	0	0
f_1	0	**0,731**	0	0	**0,900**	**0,833**	0	0	0	**0,733**	**0,669**	**0,699**	**0,644**
f_2	0	**0,188**	0	**0,821**	0	**0,167**	0	0	0	**0,199**	**0,243**	**0,230**	**0,271**
f_3	0	**0,081**	0	**0,179**	**0,100**	0	0	0	0	**0,068**	**0,088**	**0,070**	**0,085**
f_4	0	0	**0,740**	0	0	0	0	**0,900**	**0,875**	**0,687**	**0,637**	**0,731**	**0,731**
f_5	0	0	**0,167**	0	0	0	**0,821**	0	**0,125**	**0,186**	**0,258**	**0,188**	**0,188**
f_6	0	0	**0,094**	0	0	0	**0,179**	**0,100**	0	**0,127**	**0,105**	**0,081**	**0,081**
a_1	0	0	0	**0,535**	**0,622**	**0,645**	**0,655**	**0,105**	**0,678**	0	**0,731**	**0,793**	**0,735**
a_2	0	0	0	**0,113**	**0,063**	**0,041**	**0,048**	**0,056**	**0,066**	**0,122**	0	**0,076**	**0,058**
a_3	0	0	0	**0,239**	**0,100**	**0,219**	**0,208**	**0,419**	**0,090**	**0,648**	**0,188**	0	**0,207**
a_4	0	0	0	**0,113**	**0,215**	**0,094**	**0,089**	**0,419**	**0,166**	**0,230**	**0,081**	**0,131**	0

Table 2 Unweighted ANP supermatrix

4 Results and discussion

In the case of pair-wise comparisons of financial criteria (c_1) the preference of the given sub-criteria the sub-criterion f_1 was selected (total cost) as the most important. It represents a value of money that has to be borrowed from the financial institution. It is a decisive factor because it has an influence on the decision maker's (entrepreneur's) sum of liabilities. The decision maker's (DM) goal is to purchase particular fixed asset with minimum use of liabilities. The amount of monthly payment (f_2) represents smaller financial demands from the decision maker than for a payment of the whole value of the asset. The DM's goal is to pay/have monthly low as possible. The actual purchase of the fixed asset comes together with miscellaneous costs (f_3) like other fees or insurance that changes the final price of the asset but not as much as the other sub-criteria.

The non-financial criteria (c_2) represent less preferred group of criteria. The sub-criterion f4 (fixture period) can be perceived as the most important, because this represents the payback time in total with fees that are bound to the period. Generally, longer period is preferred to shorter because the real value of money is decreasing.

Various financial institutions have different demands for client's credit score (in the case of loans or leasing) which has to be proclaimed by specific documents and makes it more difficult to administer. In this case these documents are represented in numbers (sub-criterion f5) so the less documents has to be delivered the more is this option preferred. The last sub-criterion that is considered is the distance to particular institution's office.

The evaluation of alternatives

In this step the DM has to evaluate alternatives according to given criteria. All estimated weights calculated using RGMM (3) method in excel can be put in the supermatrix (6) to calculate $U(a_i)$ of the AHP approach using equations (9) and (10). Results are shown in the table 4. In the next step the looping comparisons are made according to the supermatrix (7). Weights were again calculated using the RGMM approach (3). Application of ANP should provide an alternative perspective on dependencies among given criteria and alternatives. The ANP unweighted supermatrix is shown in Table 2. Then the limited ANP supermatrix in Table 3 is calculated using equations (9) and (10). The results were stable after 7th iteration. Actual results of the realized decision making using AHP and ANP approaches are presented in Table 4. Results show that the ranking of criteria has not changed except sub-criteria f_3 and f_4 that have swapped places. Priorities estimated by ANP are less variable and tend to be more equivalent. Small changes in priorities estimated by ANP had no impact on their ranking.

	Goal	c_1	c_2	f_1	f_2	f_3	f_4	f_5	f_6	a_1	a_2	a_3	a_4
Goal	0	0	0	0	0	0	0	0	0	0	0	0	0
c_1	0	0	0	0	0	0	0	0	0	0	0	0	0
c_2	0	0	0	0	0	0	0	0	0	0	0	0	0
f_1	**0.160**	0.160	0.160	0.160	0.160	0.160	0.160	0.160	0.160	0.160	0.160	0.160	0.160
f_2	**0.095**	0.095	0.095	0.095	0.095	0.095	0.095	0.095	0.095	0.095	0.095	0.095	0.095
f_3	**0.031**	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031
f_4	**0.160**	0.160	0.160	0.160	0.160	0.160	0.160	0.160	0.160	0.160	0.160	0.160	0.160
f_5	**0.095**	0.095	0.095	0.095	0.095	0.095	0.095	0.095	0.095	0.095	0.095	0.095	0.095
f_6	**0.031**	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031
a_1	**0.207**	0.207	0.207	0.207	0.207	0.207	0.207	0.207	0.207	0.207	0.207	0.207	0.207
a_2	**0.033**	0.033	0.033	0.033	0.033	0.033	0.033	0.033	0.033	0.033	0.033	0.033	0.033
a_3	**0.117**	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117	0.117
a_4	**0.072**	0.072	0.072	0.072	0.072	0.072	0.072	0.072	0.072	0.072	0.072	0.072	0.072

Table 3 Final limited ANP supermatrix

Criteria		**Local priorities**		**Global priorities**	
		AHP	**ANP**	**AHP**	**ANP**
c_1	f_1	0,731	0,560	0.639	0.280
0,875	f_2	0,188	0,333	0.165	0.166
	f_3	0,081	0,107	0.071	0.054
c_2	f_4	0,740	0,561	0.092	0.281
0,125	f_5	0,167	0,331	0.021	0.166
	f_6	0,094	0,107	0.012	0.054
Alternatives	a_1	0,561	0,482		
	a_2	0,092	0,077		
	a_3	0,214	0,273		
	a_4	0,133	0,168		

Table 4 Comparison of AHP and ANP results

The Figure 2 shows significant changes in values of criteria priorities and values of utility function of given alternatives. Priorities estimated by ANP are more evenly distributed. This is a result of interdependencies.

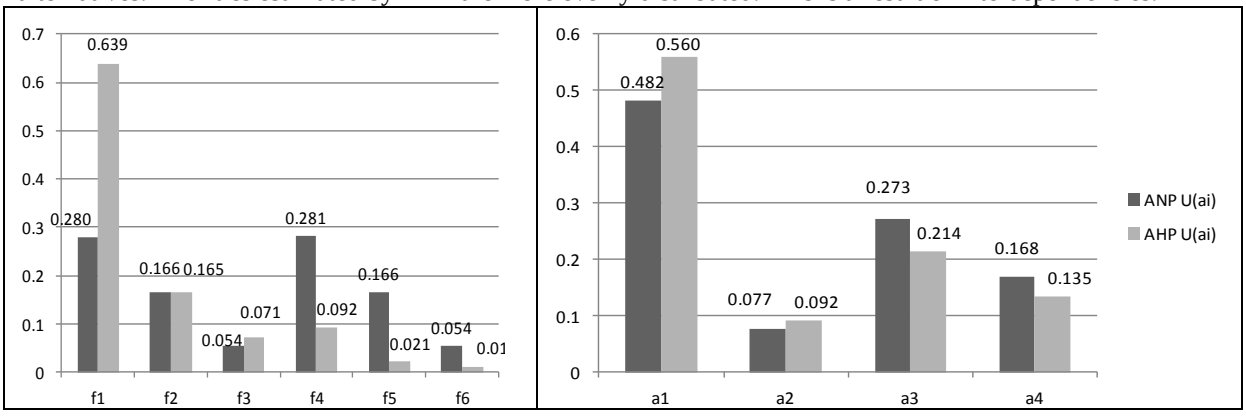


Figure 2 Comparison of criteria global weights and alternative's utility of AHP/ANP results

Results of the comparison led to a suggested conclusion that utilization of ANP on small scale and less complicated decision making problems is applicable, whereas AHP shows relative robust weight distribution that can be evaluated using sensitivity analysis as in [12] pp 48-49, thus it is not very efficient to spend time and make computations that are necessary for ANP method. In the case of decision making about alternatives the AHP is sufficient tool. When the estimation of criteria weights is considered, the ANP can be applied to evaluate possible interdependencies and their influences. AHP can be easily exercised using for example MS Excel add-on DAME [4].

5 Conclusion

The aim of the paper was to compare two MADM approaches to decision making, the analytic hierarchy process and analytic network process that were applied on a decision making problem of financial banking of a capital investment. The problem had four levels. Financial and nonfinancial decision-making criteria were considered and their weights estimated using RGMM method. Final results were calculated by limited supermatrices. Based on the results gathered from the application in a decision making problem example it can be suggested that AHP method is sufficient and more efficient that ANP when dealing with less complicated decision making with a goal to select the best alternative. It is also advisable to apply the sensitivity analysis to evaluate the robustness of results. In the case of criteria and alternatives weight distribution (importance) the ANP can be considered for further evaluation to cover the consideration of possible interdependencies among given criteria and alternatives. Employment of MS Excel also demonstrated its practical use for small firms that can not afford expensive decision support software. Results of this study were verified by Super Decision software and have not revealed any significant differences.

Acknowledgements

This paper was elaborated with support from the Student Grant Competition of the Faculty of Economics, VŠB-Technical University of Ostrava project registration number SP2013/173 “Aplikace dekompozičních vícekriteriálních metod MADM v oblasti podnikové ekonomiky a managementu”.

References

[1] Alonso, J. A., Lamata, M., T.: Consistency In The Analytic Hierarchy Process:: A New Approach. *International Journal of Uncertainty, Fuzziness & Knowledge-Based Systems*. **4** (2006), 445-459.

[2] Brožová, H.: AHP analysis of teacher's managerial competencies. In: *Proceedings Of 30th International Conference Mathematical Methods In Economics, pts I and II*. (Ramik, J; Stavarek, D. eds.). Opava, Silesian University Opava, 2012, 67-72.

[3] Fotr, J., Souček, I.: *Investiční rozhodování a řízení projektů*. Grada Publishing, Praha, 2011.

[4] Perzina, R., Ramik, J.: DAME - Microsoft Excel add-in for solving multicriteria decision problems with scenarios. In: *Proceedings Of 30th International Conference Mathematical Methods In Economics, pts I and II*. (Ramik, J; Stavarek, D. eds.). Opava, Silesian University Opava, 2012, 697-702.

[5] Ramík, J.: *Vícekriteriální rozhodování - analytický hierarchický proces (AHP).* Slezská univerzita v Opavě, Obchodně podnikatelská fakulta, Karviná, 1999.

[6] Rydval, J.: Quantification of framing effect using ANP. In: *Proceedings Of 30th International Conference Mathematical Methods In Economics, pts I and II*. (Ramik, J; Stavarek, D. eds.). Opava, Silesian University Opava, 2012, 774-780.

[7] Saaty, T., L.: *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. McGraw-Hill, New York, 1980.

[8] Saaty, T., L.: *Decision making with dependence and feedback: the analytic network process*. RWS Publications, Pittsburgh, 1996.

[9] Saaty, T., L., Vargas, G., L.: *Models, Methods, Concepts & Applications of the Analytic Hierarchy Process*. Kluwer Academic Publishers, Boston, 2001.

[10] Zmeškal, Z.: Vícekriteriální hodnocení variant a analýza citlivosti při výběru produktů finančních institucí. In: *Sborník příspěvků ze 7. mezinárodní vědecké konference Finanční řízení podniků a finančních institucí*. Ostrava, Vysoká škola báňská – Technická univerzita Ostrava, 2009, 485-490.

[11] Zmeškal, Z., Dluhošová, D., Richtarová, D.: Application of sensitivity analysis when’s financial decisions. In: *Managing And Modelling Of Financial Risks* (Dluhošová, D.). Ostrava, Vysoká škola báňská – Technická univerzita Ostrava, 2010, 55-68.

[12] Zmeškal, Z., Dluhošová, D., Tichý, T.: *Finanční modely: koncepty, metody, aplikace*. 3. Vyd. EKOPRESS, Praha, 2013.

Using Pearson distribution system to monitor a process, which distribution is unknown

Magdalena Chmielińska[1]

Abstract. A control chart is a tool of statistical quality control, which is commonly used in factories. Meeting its basic assumptions, guarantee faultless estimation of correctness of monitored production process. Infringement of classical control charts assumptions may cause false signals in case of regulated process, or lack of signal as well as signal delayed in time in case of inregulated process.
In this paper, based on true data, the trial of calculation the control limits for production process, whose controlled feature has a distribution, which is essentially different to a normal distribution, was made. A proposed method of calculation the control limits based on the Pearson distribution system, regards in first step to selection from the Pearson family distribution, the correct one, which can by used to estimation the data empirical distribution. Next the level of control limits are fitted on based on the quantile of this distribution. The features of the proposed method were testing by the simulations carried out in R.

Keywords: Control charts, the Pearson's System of Distributions

JEL Classification: C44
AMS Classification: 90C15

1 The control chart

Control chart is a tool, which enables statistical monitoring of production process. It gives information about run of process, as well as it contributes to the rise of a production quality. Its characteristic features are simplicity of construction and effectiveness.

The classical control charts are base on an assumption, that the diagnostic variable, which is observed during process monitoring has a normal distribution with an expected value μ and a standard dividation σ [6]. They assume also, that the measurements in successives time periods are independent. The control chart is a graph dedicated to register the results of current quality control of products. On its are plotted the central line – defining the expected value of the monitored characteristicts, the control limits (upper and lower) and the warning limits (upper and lower).

The control lines are set at such a level that the statistical probability of the values below the lower and above the upper control line, when the process is in – control, was appropriately small. Classically, assuming a normal distribution of the characteristic, these lines are three standard deviations of the monitored manufacturing process away from the central line. Warning lines, and are usually placed at the level of the center line plus and minus 2σ [4]. Due to the violation of the assumption of normal distribution of monitored characteristic, the position of the control lines is increasingly set on the basis of the values of the appropriate quantile of known theoretical distributions that can be used to model the actual data.

Points configuration on a control chart in relation to the level of control and warning limits is the basis for reasoning about the correctness of the course of the controlled process. In the literature [9] the many signs of a deregulation process are considered, while in this paper as a signal of process disturbance just a point above (below) the upper (lower) control line is considered.

The control charts are characterized by the ARL (Average Run Length). In case of considering only exceeded upper or lower control lines, it is the inverse of the probability of the signal (p_s) for a single observation (sample) [6]:

$$ARL = \frac{1}{p_s} \tag{1}$$

ARL for regulated process is marked as ARL_0 and for process beeing out – of – control as ARL_1.

[1] University of Economics in Katowice/Department of Statistic, ul. Bogucicka 14, 40-226 Katowice, chmielinska.magda@gmail.com.

2 The Pearson's system of distributions

Elaborated toward the end of XIX century, by Carl Pearson a system of distributions of random variability, came into begin by solve a differential equation in the form [3]:

$$\frac{df(x)}{dx} = -\frac{a+x}{c_0 + c_1 x + c_2 x^2} f(x) \tag{2}$$

As results of this equation are density functions in the form:

$$f(x) = Ce^{-\int \frac{a+x}{c_0+c_1x+c_2x^2}dx}. \tag{3}$$

Miszczak [8] announces that, the atributtes of those density functions depend on squer trinomial in the form of:

$$M(x) = c_0 + c_1 x + c_2 x^2 \tag{4}$$

Defining the square of the skewness coefficient (β_1) and flattening coefficient (β_2) as:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}, \qquad \beta_2 = \frac{\mu_4}{\mu_2^2},$$

where μ_i – central moment of *i*-th order,

classification of seven of the 12 awarded by the Pearson type of greatest practical use, can be represented graphically (Fig. 1). It is considered that the main types are types I, IV and VI, the types of transition are, types III (straight line) and type V (almost in a straight line). There are also types of symmetry, which include:

- Type 0 (in the literature referred to as type N) at the point (0,3) to denote the normal distribution,
- Type II includes a vertical line on the Y axis for $\beta_2 < 3$ of which is a special case of type I,
- Type VII covering a vertical line on the Y axis for $\beta_2 > 3$ of which is a special case of type IV.

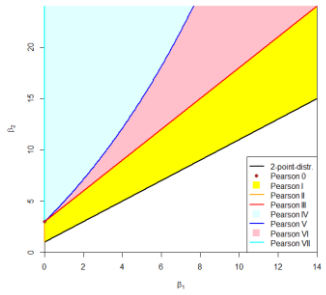


Figure 1 Pearson types of distributions of the system according to β_1, β_2
Source: software R (R i386 3.0.0)

The normal distribution, designated on the figure 1 as "Pearson 0" is a limiting distribution of all designed by Pearson types [5]. I type the Pearson's distribution is obtained when the two elements trinomial square (4) are real and have different signs. Denoting elements as x_1 and x_2, probability density function is:

$$f(x) = C(x - x_1)^{\frac{a+x_1}{c_2x_2 - c_2x_1}}(x - x_2)^{\frac{a+x_1}{c_2x_1 - c_2x_2}} \tag{5}$$

The solution (5) may be distributions U – shaped, when both exponent are negative, or distributions J – shaped, as one of the powers is negative and the other is positive. Distributions belonging to this family are distributions of beta class.

Type II, which is a particular case of distribution of type I, is characterized in that the module of value of square roots of the trinomial are equal. Distribution of this type is symmetrical, and its density function is:

$$f(x) = C(x^2 - x_1^2)^{-\frac{1}{2c_2}} \tag{6}$$

Another type of Pearson's distributions obtained when $c_2 = 0$, is the type III. Density function of this distribution is given by:

$$f(x) = Ce^{-\frac{x}{c_1}}(c_0 + c_1 x)^{\frac{c_0 - ac_1}{c_1^2}} \tag{7}$$

Since the reduction $c_2 = 0$ is equivalent to the union $\beta_2 = 1{,}5\beta_1 + 3$, the density distribution of type III expressed by the parameter β_1 can be written as:

$$f(x)=C\left(\frac{12+6x\sqrt{\beta_1}+3\beta_1+1{,}5x\beta_1\sqrt{\beta_1}}{12+3\beta_1}\right)^{-\frac{\beta_1-4}{\beta_1}} e^{-\frac{2x}{\sqrt{\beta_1}}} \tag{8}$$

Particular cases of this type are distributions of gamma distributions.

Type IV of Pearson's distribution system, which is a basic type, we get when the trinomial square (4) does not have real roots. Its density function is given by:

$$f(x)=Ce^{\frac{(c_1-2ac_2)arctg\left(\frac{c_1+2c_2x}{\sqrt{4c_0c_2-c_1^2}}\right)}{c_2\sqrt{4c_0c_2-c_1^2}}}\left(c_0+x(c_1+c_{2x})\right)^{-\frac{1}{2c_2}} \tag{9}$$

This type is in fact an asymmetric version of the Student's t distribution. His special case, according to Heinrich [11], the Cauchy distribution is.

Another type of Pearson's curve (Type V) is obtained when trinomial square (4) has only one root. The density distribution of this type is:

$$f(x)=Ce^{\frac{2ac_2-c_1}{c_2(c_1+2c_2x)}}\left(c_1+2c_2x\right)^{--\frac{1}{c_2}} \tag{10}$$

Type VI of distribution is obtained, however, the real roots of trinomial square (4) exist and have the same sign. Density of Pearson's distributions type VI is as follows:

$$f(x)=C(x-x_1)^{\frac{a+x_1}{c_2x_2-c_2x_1}}(x-x_2)^{\frac{a+x_2}{c_2x_1-c_2x_2}} \tag{11}$$

Symmetrical type VII of distribution is a special case of Pearson's type VI distribution. It is obtained when $c_1=a=0$. Its density function is expressed by the formula:

$$f(x)=C\left(c_0+c_2x^2\right)^{--\frac{1}{2c_2}} \tag{12}$$

A special case of this family of distributions is the Student – t distribution.

A full list of all 12 types of curves Pearson's Magiera [7].

3 Description of the data

Presented in this paper considerations was based on available data collected and processed by a plant representing the automotive industry leading its activities in the province of Silesia. The analyzed data derived from a middle – operating control of chosen characteristic of produced there element. Shared data cover a range of information about their controls selected characteristics of each batch produced the selected item in the period from 10.09.2012 to 15.09.2012. In total, this is 10969 records.

During this period, there was no disturbance of the production process, since all obtained during the audit, the values were situated in determined for the controlled characteristic standards. The analyzed data can be treated so as data from a process of correct course. Therefore, this data can help determine the control limits of the controlled characteristic.

These data were previously similar analysis. In [2] tested the compatibility of the empirical distribution of the analyzed data with normal distribution and gamma distribution. In both cases verified hypothesis was rejected.

4 Analysis of the correctness of the manufacturing process

The analysis of the production process should start by checking (usually based on historical data) assumptions of control chart. As the Chmielińska [2] analyzed the empirical distribution of the data in this study can not be approximated by known theoretical distributions, including the normal distribution.

Due to the lack of normal distribution of the controlled diagnostic variable classical control charts, requiring compliance of the empirical distribution of the data with a normal distribution, can lead to an incorrect assessment of the correctness of the production process. This situation occurs in the case of the analyzed data. The classical control chart (Figure 2) indicates that the production process is out – of – control. ARL_0 is equal $ARL_0=79$, while it is known that the process runs correctly, that the ARL_0 should be the order of 370.

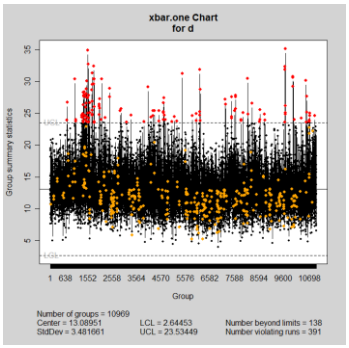


Figure 2 The classical control card removed for the analyzed data

4.1 Non-classical method for determining the control limits

In view of the risk of an erroneous assessment of the production process for the empirical data on the distribution significantly different from a normal distribution the control limits on the control chart should be set in a different way. The article [2] studied the consequences of monitoring the analyzed production process using control charts based on the quantile of the theoretical distribution, whose sharp is close to the shape of the distribution of empirical data and based on the quantile of estimated by a kernel estimation density function of empirical data. Also Chakraborti, van der Laan, van de Wiel [1] provide an effective solution to this problem. In this paper a method for determining the control lines based on quantiles suitably selected curve of the Pearson family of distributions is proposed.

The preparation of the proposed chard, should start by selecting the type of curve that approximates the empirical distribution of the data the best from among the family of Pearson's distributions. It is necessary to determine the parameters of all functions and their mutual comparison. In this case, the analysis carried out in the R. In the package PearsonDS, which is dedicated Pearson curves, is a function *pearsonMSC* that sets the parameters for all seven basic Pearson curves and makes their comparison using information criteria (ML, AIC, AICC, BIC, HQC). All benchmarks show that the type IV curve of Pearson (Pearson IV) models the empirical distribution of the data the best of all.

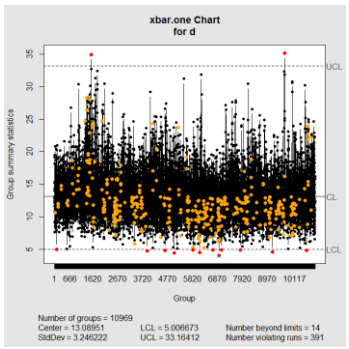


Figure 3 Control chard for the analyzed data plotted where control lines were determined on the basis of quantile of Pearson curve type IV

Figure 3 shows the control chart where the control limits are based on the values of the corresponding quantiles fitted to the data Pearson curve type IV. On this basis it can be concluded that the proposed method is effective for determining the control limits for the data of non-normal probability distribution. It is determined at a such level that the probability of obtaining a signal of deregulation for in – control process is of the order of $p_s = 0{,}001$. ARL_0 is equal $ARL_0 = 784$.

4.2 Verification of the proposed method for determining the control limits based on the quantile of corresponding curve form the Pearson family of distributions

In order to verify the accuracy of the proposed method for determining the control lines based on the quantile of suitable Pearson's curve the computer simulations was carried out. The simulation procedure, including two variants proceed in stages. The first option assumes property run of process. In the first step, from the available data samples consist of n – element were randomized. This samples form the basis of the determining of levels of control lines. Considered discussed two methods of determining the control limits, i.e. the classical method and a method based on the values of the quantile of corresponding Pearson distribution curve. In case of the

proposed method parameters for all types of distributions were determined, their fitting to empirical data assessed using the information criteria (ML, AIC, AICC, the BIC, HQC) to select the best type, then of the type indicated by the five considered information criteria chosen this dominant recognizing it as the type, which is the best suited to the empirical data. In the next step, control charts were plotted with the specified control limits and for the available data, the actual number of exceedances of the control line, and the signal probability and ARL_0 were counted. The procedure was repeated 1000 times, averaging the results. Control lines determined based on samples consist of 100, 500, 1000, 2000, 4000 and 5000 elements.

In the second option assuming the out – of – control process, as in the first variant *n* - elements samples were used for the determination of the control lines in accordance with an algorithms of the analyzed methods. Then the generated values from a normal distribution with parameters $\mu = 11$ and $\sigma = 2{,}9$. These values were applied to the prepared control charts, then number of exceedances of control lines were counted, the probability of a signal and ARL_1 were calculated. The procedure was repeated 1000 times, averaging the results.

Sample size *n*	The in – control process		The out – of – control process	
	The classical control chart	The chart based on the quantile of corresponding Pearson's curve	The classical control chart	The chart based on the quantile of corresponding Pearson's curve
100	69,54	104,67	318,77	23,43
500	76,86	334,73	454,58	39,83
1000	78,68	495,10	474,56	47,54
2000	79,31	569,94	487,34	50,54
4000	79,45	626,44	494,23	51,06
5000	79,47	614,44	498,00	52,46

Table 1 ARL_0 and ARL_1 for the analyzed data according to the method of determining the level of control lines and the number of initial sample

Table 1 shows the average run length in depend on the method for determining the control lines and the number of the initial sample of analyzed data for a process of correct run and data derived from of the normal distribution with parameters $\mu = 11$ and $\sigma = 2{,}9$ for the out – of – control process. On this basis it can be concluded that regardless of the method for determining the control lines and accuracy of the process, along with an increase in the initial sample size ARL increases. Assuming correct run of manufacturing process, this means that with the increase in the sample size precision of the assumed methods increase. For the in – control process ARL_0 value should be large enough (about 370) to false signals about the deregulation process did not occur very often, and when the process is out – of – control ARL_1 value should be small enough that the deregulation can be readily detected [10]. Comparing the ARL values obtained by the classical control chart and the proposed chart should recognize that the non – classical method, based on the corresponding curve form the Pearson's family of distributions guarantee the ARL_0 and ARL_1 at the expected level. Classic control chart for the analyzed data generates the exact opposite – low ARL_0 value to regulated process and high ARL_1 value to process deregulated. This means that in case of regulated process, numerous false signals of deregulation are occurred, while, in case of deregulated process this chart, based on the assumption that the monitored characteristic has a normal distribution with parameters similar to the values set at the sample, the value of which, in the present case are at a level of $\bar{x} = 13{,}09$ and $s = 3{,}48$, regardless of the initial sample size is insensitive to the disorder introduced in the process. Adverse results for classical control chart are the result of the lack of fulfillment of its basic assumption – the assumption of normal distribution of the controlled characteristic.

Analyzing the ARL values generated by the proposed control chart should be aware that they depend on the goodness of fit of Pearson's distribution to the empirical distribution of the data. This goodness depends on the choice of the appropriate type curve and the value of the estimated parameters. Choosing Pearson's curve used to estimate the accuracy of the analyzed process and at the same time the value of the parameters is determined by the initial sample drawn out by lot. Analyzing the information contained in Table 2, showing the distribution of a function of the Pearson's family of curves depending on the size of the initial sample, it should be noted that the correct (same as for the totality of the available data) type of curve with probability $p = 0{,}9$ is chosen from among the available options for sample numbering a minimum 1,000 items. For 500 – elemental initial sample, which is a true dominant type IV, but the probability of selecting is row $p = 0{,}64$. Too small initial sample size makes the appropriate type (for the analyzed data – type IV) is chosen very rare. For the 100 – elemental initial sample type IV is considered to be the most appropriate for modeling the analyzed data average fifteen times for each hundred attempts made.

Sample size *n*	The Pearson's cave type							
	0	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**
100	0,0715	0,0285	0,0045	0,256	0,152	0,471	0	0,0165
500	0	0	0	0,0165	0,641	0,3425	0	0
1000	0	0	0	0,0005	0,907	0,0925	0	0
2000	0	0	0	0	0,9895	0,0105	0	0
4000	0	0	0	0	0,995	0,005	0	0
5000	0	0	0	0	0,9955	0,0045	0	0

Table 2 Distribution of Pearson's curves type for the analyzed data, depending on the size of the initial sample

Analyzing the information contained in the tables above, it should be noted that the proposed method to with the assumed high probability (at least $p = 0{,}9$) chose the correct Pearson's distribution giving acceptable results (comparable to the results achieved by classical control chart with the fulfillment of assumptions) requires a relatively large initial batch of samples least 1,000 items.

5 Conclusions

Presented in this paper reflections on the monitoring of production processes characterized by a non – normal distribution of the controlled characteristics lead to the conclusion that the use of the classical approach based on the common infringed assumption about the normality of distribution of controlled characteristics, leading to an incorrect assessment of the correctness of the process being monitored. Proposed in this paper method for determining control limits for the monitored process based on modeling the empirical distribution of diagnostic variable using Pearson distributions and estimating about proper conduct of the process being monitored on the basis of the resulting distribution, belongs to distribution – free methods.

On the basis of carried out simulation analysis can be concluded that the proposed method has the desirable features of good control chart. The proposed chart responds quickly to increase of defective in manufacturing process and does not indicate too often the disturbance, when the process is in – control. The proposed method seems to be an effective method for determining the control lines for a process, which monitored characteristic has unknown distribution. The huge advantage of it is to not assuming any kind of distribution. The disadvantage of the proposed method is a need for a large initial sample required to estimate the levels of control lines.

Moved in this paper problem of process monitoring, which distribution is unknown, is important from the control of the production process point of view. Therefore, the presented analysis will be further developed. The use of stable distributions to approximate the unknown distribution of the monitored characteristic is considered.

References

[1] Chakraborti S, van der Laan P., van de Wiel M. A.: A Class of Distribution-free Control Charts. "Journal of the Royal Statistical Society" 2004, No. 3, Vol. 53, s. 443-462.

[2] Chmielińska M. : *Wykorzystanie estymacji jądrowej do monitorowania procesów o nieznanym rozkładzie*. *(Using the kernel estimation to monitor a process, which distribution is unknown).* In printting.

[3] Goc M: *Modelowanie stopy bezrobocia za pomocą krzywych Pearsona*. *(Modeling the unemployment rate using Pearson curves).* In: FOLIA POMERANAE UNIVERSITATIS TECHNOLOGIAE STETINENSIS 2010, Oeconomica 280 (59), 45–52

[4] Grant E. L. (1972), Statystyczna kontrola jakości, *(Statistical quality control)*, PWE, Warszawa.

[5] Johnson N., L., Cotz S., Balakrisshman N. (1994): *Continuous Univariate Distributions (Volume I)*, John Wiley&Sons, Second Edition, New York.

[6] Kończak G. (2007) Metody statystyczne w sterowaniu jakością produkcji, *(Statistical methods in quality control of production)*. Wydawnictwo Akademii Ekonomicznej w Katowicach, Katowice.

[7] Magiera R.: *Modele I metody statystyki matematycznej. Część I Rozkłady i symulacja stochastyczna.* Wydanie drugie rozszerzone, Oficyna Wydawnicza Gis, Wrocław, 2005.

[8] Miszczak W.: *System rozkładów Pearsona I ich zastosowania. In: Zastosowania statystyki i matematyki w ekonomii, (The Pearson distributions and their applications. In: Applications of Statistics and Mathematics in Economics),* (Ostasiewicz W. eds), Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław, 2004, 15 – 43.

[9] Montgomery D. C. (1997), Introduction to Statistical Quality Control, John Wiley&Sons, Inc, Third Edition, New York.

[10] Reynolds M. R., Amin R. W., Arnold J. C.: *CUSUM charts whit variable sampling intervals*. Technometrics, Vol. 32, November 1990, 371 – 384.

[11] http://www-cdf.fnal.gov/physics/statistics/notes/cdf6820_pearson4.pdf [30.04.2013].

The Global Financial Crisis and Stock Returns: Evidence from the Czech Republic, Hungary and Poland

Michaela Chocholatá[1]

Abstract. The paper deals with the analysis of the stock market co-movements of the Central Eastern European (CEE) stock markets (Czech, Hungarian and Polish) vis-à-vis the Western European stock market based on BEKK-GARCH models using daily closing values of individual stock market indices during the period from January 8, 1996 to January 21, 2013, and studies also the impact of the current global financial crisis on the corresponding conditional correlations.
The existence of the co-movement of CEE stock markets vis-à-vis Western European markets was proved. The average conditional correlations varied in the whole analysed period as follows: Hungary 0,438 – 0,470; the Czech Republic 0,415 – 0,472 and Poland 0,496 – 0,525. Although the stock market integration strengthened during the whole analysed period, it surprisingly weakened in the crisis period (the only exception was the co-movement of the Polish WIG with German DAX).

Keywords: financial integration, BEKK-GARCH, global financial crisis, conditional correlation

JEL Classification: C58, G15, G01, C22
AMS Classification: 91G70, 91B84

1 Introduction

Nowadays the term "crisis" can be heard almost everywhere. Although during the recent history we can find different periods of various crises, e. g. Asian currency crisis of 1997, Russian crises of 1998, the Brazilian devaluation in 1999, sub-prime mortgages crises in US in 2007, the current global financial crisis is often characterized to be the worst one since the Great Depression. Since the question of financial integration between economies of different countries has recently become more popular (in order to diversify the risks effectively and to improve the capital allocation efficiency), therefore it seems to be very attractive for analysts to assess the impact of current financial crisis on various especially emerging markets. There are plenty of studies dealing with this issue using various ways and methods of analysis in order to capture how shocks from one market can be transmitted to another market(s). Schmukler [18] tries to present a balanced view of financial globalization, dealing with the pros and cons that globalization entails for developing countries. Asongu [1] deals with the linkages between financial integration (globalization) and crises and also provides the survey of various definitions of contagion. We will use the definition of Forbes and Rigobon [14] who distinguish the stock market co-movement during the periods of stability and during the periods after a shock or crisis. They use the term contagion to define „a significant increase in cross-market linkages after a shock to one country (or group of countries)". So, in case that the co-movement does not increase significantly after a shock or crisis, they speak about interdependence. In their paper [14] they also present different methodologies which can be used to analyse the stock market co-movement, e. g. cross-market correlation coefficients, Autoregressive Conditional Heteroscedasticity (ARCH) and Generalized ARCH (GARCH) models, cointegration techniques and direct estimation of specific transmission mechanisms. In recent analyses the dominant role play various variants of the multivariate GARCH (MGARCH) models which are superior to univariate GARCH versions, because they are able to capture also the development of covariances over time. Different types of multivariate GARCH models can be used, e. g. VECH model [5], CCC model [6], BEKK model [12], DCC model [13] and AG-DCC model [10].[2]

The main aim of this paper is to study the stock market co-movements of the Central Eastern European (CEE) stock markets (Czech, Hungarian and Polish) vis-à-vis the Western European stock market (as a benchmark the French CAC40, German DAX and STOXX Europe 600[3] were used) based on BEKK-GARCH models[4] and also to assess the impact of the current global financial crisis on the corresponding conditional correlations.

[1] University of Economics/Faculty of Economic Informatics, Department of Operations Research and Econometrics, Dolnozemská cesta 1, 852 35 Bratislava, e-mail: michaela.chocholata@euba.sk
[2] For an extensive survey of multivariate GARCH models see e. g. Bauwens et al. [3].
[3] The STOXX Europe 600 Index represents large, mid and small capitalisation companies across 18 countries of the European region: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom.

The CEE stock markets can be characterized as emerging stock markets and it is commonly known that the emerging stock markets are usually more volatile than the developed stock markets. Concerning the impact of the crisis, majority of the emerging markets are easily affected. Although there exist some studies dealing with the co-movements of CEE stock markets and their integration with the Western European stock markets using various benchmark stock markets as well as various techniques and also giving mixed results, not so many studies dealt with the impact of the crisis on the degree of the stock market integration. Some studies based on multivariate GARCH models are as follows: Égert and Kočenda [11] studied co-movements between three developed stock markets (France, Germany, the UK) and the three CEE stock markets based on five-minute tick intraday stock price data. They detected very little systematic positive correlation between the Western European stock markets and the three CEE stock markets. Wang and Moore [19] investigated the extent of integration of three CEE stock markets with the aggregate eurozone market. They proved a higher level of the stock market correlation during and after the Asian and Russian crisis and also during the period after integration of the CEE countries into the EU. Baumöhl et al. [2] analysed the integration of the stock markets of V4 countries with the German market and confirmed that the correlations of the stock market indices of V4 countries (the only exception was the Slovak SAX) with German DAX became higher during the analyzed period. Horvath and Petrovski [15] proved a quite high level of stock market integration between the analysed CEE countries and Western Europe (corresponding conditional correlations were around 0,6). They did not confirm the impact of the crisis on the degree of the stock market integration between the analysed group of countries.

The paper is organised as follows: section 2 deals with the methodology of MGARCH models, section 3 describes the data used for analysis and estimation results, section 4 concludes.

2 Methodology

The methodology in this paper is based on MGARCH models in order to capture the market interdependencies in the conditional mean and/or conditional variance equations. As it was already mentioned in the introduction, nowadays there exist several various multivariate GARCH model specifications with certain advantages and disadvantages, and with different restrictions and specifications concerning the conditional variance. The basic MGARCH model was firstly introduced by Bollerslev et al. [5] who extended the univariate GARCH model [4] to include a vectorised conditional covariance matrix. To estimate the unrestricted model, known as the VECH-GARCH model, it is quite a hard issue because of the large number of parameters contained within it. Since that time several other modifications of MGARCH models have been developed (some of them were mentioned in the introduction) in order to reduce the number of parameters needed to be estimated. The most widely used models for modelling of the conditional covariances and correlations are the BEKK model [12] and the DCC model [13]. There exist also some studies in order to determine whether the BEKK or the DCC model should be applied in practical analyses, e. g. Caporin and McAleer [9] and Huang et al. [16] came to the conclusion that each of these models has its advantages and disadvantages, but it is not possible to judge in general about which one of these models to prefer. In this paper we will apply the bivariate BEKK-GARCH (1,1) model. To capture the dynamic relationship in returns we use the vector autoregressive (VAR) model with *k* lags:

$$\mathbf{r_t} = \boldsymbol{\omega} + \sum_{i=1}^{k} \boldsymbol{\Gamma_i} \mathbf{r_{t-i}} + \boldsymbol{\varepsilon_t} \tag{1}$$

where $\mathbf{r_t}$ is a 2×1 dimensional vector of daily stock returns, $\boldsymbol{\omega}$ is a 2×1 dimensional vector of constants, $\boldsymbol{\Gamma_i}$ ($i = 1, 2, \ldots k$) are a 2×2 dimensional matrices of parameters and $\boldsymbol{\varepsilon_t}$ is a vector of innovations (disturbances) conditional on information at time $t-1$. The conditional distribution of $\boldsymbol{\varepsilon_t}$ is assumed to be multivariate normal with the mean zero and the conditional variance-covariance matrix $\mathbf{H_t}$, i. e.

$$\boldsymbol{\varepsilon_t} | \Omega_{t-1} \sim N(0, \mathbf{H_t}) \tag{2}$$

where Ω_{t-1} represents the information set at time $t-1$.

It is possible to specify $\mathbf{H_t}$ in different ways. Taking into account that $\mathbf{H_t}$ is a conditional variance-covariance matrix, positive definiteness has to be ensured. We will concentrate on the BEKK-GARCH model in which the matrix $\mathbf{H_t}$ can be received through the generalization of the univariate GARCH model of Bollerslev [4]. The matrix $\mathbf{H_t}$ of the BEKK-GARCH (1,1) model is defined as follows:

[4] For simplicity we will consider only the GARCH (1,1) models without further emphasising of this fact.

$$\mathbf{H_t} = \mathbf{C}'\mathbf{C} + \mathbf{A}'\boldsymbol{\varepsilon}_{\mathbf{t-1}}\boldsymbol{\varepsilon}'_{\mathbf{t-1}}\mathbf{A} + \mathbf{B}'\mathbf{H_{t-1}}\mathbf{B} \tag{3}$$

where **C** denotes a 2×2 dimensional upper triangular matrix of parameters and **A** and **B** are 2×2 dimensional matrices of parameters.

Since the presented model is an extension of a univariate one, the parameters can be similarly estimated based on the maximum likelihood method by replacing the one-dimensional function of the sample by its multi-dimensional counterpart. Taking into account the fact, that the conditional distribution of $\boldsymbol{\varepsilon}_\mathbf{t}$ is assumed to be multivariate normal, the parameters can be estimated by maximising of the log-likelihood function which is in case of bivariate models as follows (see e. g. [7], [8]):

$$L(\boldsymbol{\theta}) = -T\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\left(\ln|\mathbf{H_t}| + \hat{\boldsymbol{\varepsilon}}_{\mathbf{t-1}}\mathbf{H_t^{-1}}\hat{\boldsymbol{\varepsilon}}'_{\mathbf{t-1}}\right) \tag{4}$$

where T represents the number of observations, $\boldsymbol{\theta}$ represents the vector of all the unknown parameters and all the other symbols are the same as above.

3 Data and Estimation Results

Daily stock price indices of Central Eastern European (CEE) countries - the Czech PX, Hungarian BUX and Polish WIG20 were used for analysis. As a benchmark for Western European stock markets the French CAC40, German DAX and STOXX Europe 600 Index were used. The analysis in this paper is based on daily closing values of individual stock market indices during the period from January 8, 1996 to January 21, 2013 (totally 4088 observations)[5] and was carried out in econometric software EViews. The source of data is as follows: Czech PX [23], Hungarian BUX and Polish WIG20 [21], French CAC40 and German DAX [22], STOXX Europe 600 Index [20].

At the beginning of the analysis it is necessary to be aware of the fact that in the stock market analysis we have to do with the financial time series, one of the basic features of which is the non-stationarity. The non-stationarity in variance can be quite easily solved e. g. by using the logarithmic transformation of the corresponding time series. Since the non-stationarity in mean also represents a serious problem in econometric analysis, it is necessary to deal with it. We also started by testing the existence of the unit root in individual logarithmic stock price indices based on the Augmented Dickey – Fuller (ADF) test. At 1 % significance level we failed to reject the null hypothesis about the existence of unit root, i. e. all the series were identified to be non-stationary. The first differences of all analysed logarithmic stock indices, i. e. logarithmic stock returns, were already stationary[6]. In further analysis we also concentrated on modelling of logarithmic returns.

Descriptive statistics for logarithmic returns[7] together with the Jarque-Bera test statistics testing the normality are summarized in Table 1. The mean values vary around zero, concerning the values of standard deviations in CEE countries, the Hungarian market is most volatile with standard deviation around 1,87 %, followed by the Polish market with standard deviation 1,83 % and the Czech market with volatility of around 1,52%. Taking into account the Western European logarithmic stock returns, the highest volatility was identified in case of German DAX (1,62%) and less volatile were the logarithmic stock returns of STOXX Europe 600 Index (1,31%). All the distributions were negatively skewed and leptokurtic, the normality hypothesis can be rejected (see the Jarque-Bera statistics and corresponding probability values).

	DLBUX	DLPX	DLWIG	DLCAC	DLDAX	DLSTOXX
Mean	0.058789	0.020921	0.026020	0.016508	0.029471	0.017154
Median	0.092686	0.052739	0.022334	0.056160	0.099445	0.079673
Maximum	22.01627	12.36405	10.89614	10.59459	10.79747	9.409957
Minimum	-15.52589	-19.90195	-11.68553	-9.471537	-8.396307	-7.929709
Std. Dev.	1.875661	1.521893	1.829898	1.557932	1.616628	1.310741
Skewness	-0.183046	-0.938140	-0.140966	-0.031006	-0.059671	-0.152245
Kurtosis	14.39155	19.64185	6.544326	7.811220	6.874363	8.458169
Jarque-Bera	22121.15	47761.98	2152.782	3942.539	2558.621	5089.052
Probability	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Observations	4087	4087	4087	4087	4087	4087

Table 1 Descriptive statistics of logarithmic stock returns

[5] Although the data span available was longer, we started from January 8, 1996 in order to overcome the infrequent trading of the Czech PX index. The number of available observations was in individual cases different, but only those data were used which were defined for each of the analysed index series.

[6] The results are available from the author upon request.

[7] Logarithmic returns were calculated as the first difference of logarithmic index series and thereafter multiplied by 100.

First of all the stock market interdependencies were measured by unconditional correlation coefficients. Since we wanted also to assess the impact of the crisis onto the market co-movements, the calculation was done for the whole analysed period as well as for the crisis period from October 10, 2008 till the end[8]. The corresponding unconditional correlations are presented in Table 2. Since the unconditional correlations of CEE markets vis-à-vis Western European markets varied between 0,487 and 0,567 during the whole analysed period (the upper triangular' portion of a matrix in Table 2), the values became higher during the crisis period October 10, 2008 – January 21, 2013 (the lower triangular' portion of a matrix) and ranged between 0,628 and 0,699. Also the unconditional mutual correlations between the Western European markets became higher during the crisis period.

	DLBUX	**DLPX**	**DLWIG**	**DLCAC**	**DLDAX**	**DLSTOXX**
DLBUX	1.000000	0.545360	0.561432	0.510925	0.493172	0.545071
DLPX	***0.653513***	1.000000	0.559377	0.531238	0.486659	0.566797
DLWIG	***0.660533***	***0.717716***	1.000000	0.526481	0.510933	0.556419
DLCAC	***0.660280***	***0.642111***	***0.683735***	1.000000	0.875543	0.946946
DLDAX	***0.650396***	***0.627895***	***0.695947***	***0.951720***	1.000000	0.900765
DLSTOXX	***0.661247***	***0.672521***	***0.698551***	***0.969245***	***0.950260***	1.000000

Table 2 Unconditional correlation matrix – the upper triangular' portion of a matrix (period: January 8, 1996 – January 21, 2013), the lower triangular' portion of a matrix (period: October 10, 2008 – January 21, 2013)

In order to assess the development of conditional correlations against time, the VAR(*k*)-BEKK-GARCH(1,1) models were estimated by non-linear maximum likelihood method assuming the multivariate normal distribution for the innovations. The number of lags *k* used in individual VAR models is in Table 3 and was specified by the Schwarz information criterion, but in case that the serial correlation was detected, more lags were used in order to ensure the uncorrelatedness.

	DLCAC	**DLDAX**	**DLSTOXX**
DLBUX	9	9	6
DLPX	10	4	10
DLWIG	10	1	8

Table 3 Number of lags *k* used in individual VAR(*k*) models

Concerning the estimation results[9], based on the statistically significant parameters of the individual VAR(*k*) models we can conclude, that there exists a price spillover effect from the Western European stock market (using each of CAC, DAX and STOXX) to Hungary and the Czech Republic but not vice versa. For Poland the price spillover effect in neither direction was confirmed. Concerning the variance equation parameters, i. e. the parameters of the BEKK-GARCH(1,1) models, all of them were statistically significant in all analysed cases which confirms the adequate use of these models.[10] Figure 1 plots the conditional correlations from the estimation of BEKK-GARCH(1,1) models for the logarithmic stock returns of CEE countries vis-à-vis Western European markets. The average conditional correlations for individual CEE countries vis-à-vis Western European markets were in the whole analysed period as follows: Hungary 0,438 – 0,470; the Czech Republic 0,415 – 0,472 and Poland 0,496 – 0,525. Similarly as Horvath and Petrovski [15] we can conclude that although the values of the conditional correlations fluctuate, it is evident that there is the co-movement of CEE stock markets vis-à-vis Western European markets.

Furthermore we tested the character of market co-movement (i. e. the intensity of the stock market integration) in individual cases by estimating the trend models for the whole analysed period and for the crisis period. The estimated values of the corresponding trend parameters are summarized in Table 4. Taking into account the results, we can conclude, that although the stock market integration strengthened during the whole analysed period, it surprisingly weakened in the crisis period (the only exception was the co-movement of the Polish WIG with German DAX). It means that the contagion effect [14] was not confirmed.

[8] It is problematic to identify the start date of the current global financial crisis. We chose the date October 10, 2008, since the analysed stock market indices rapidly declined for the first time on this day.
[9] The estimation results are not reported from the space reasons, but can be provided by the author upon request.
[10] Standardized residuals of all models were successfully tested for autocorrelation using the Portmanteau statistics.

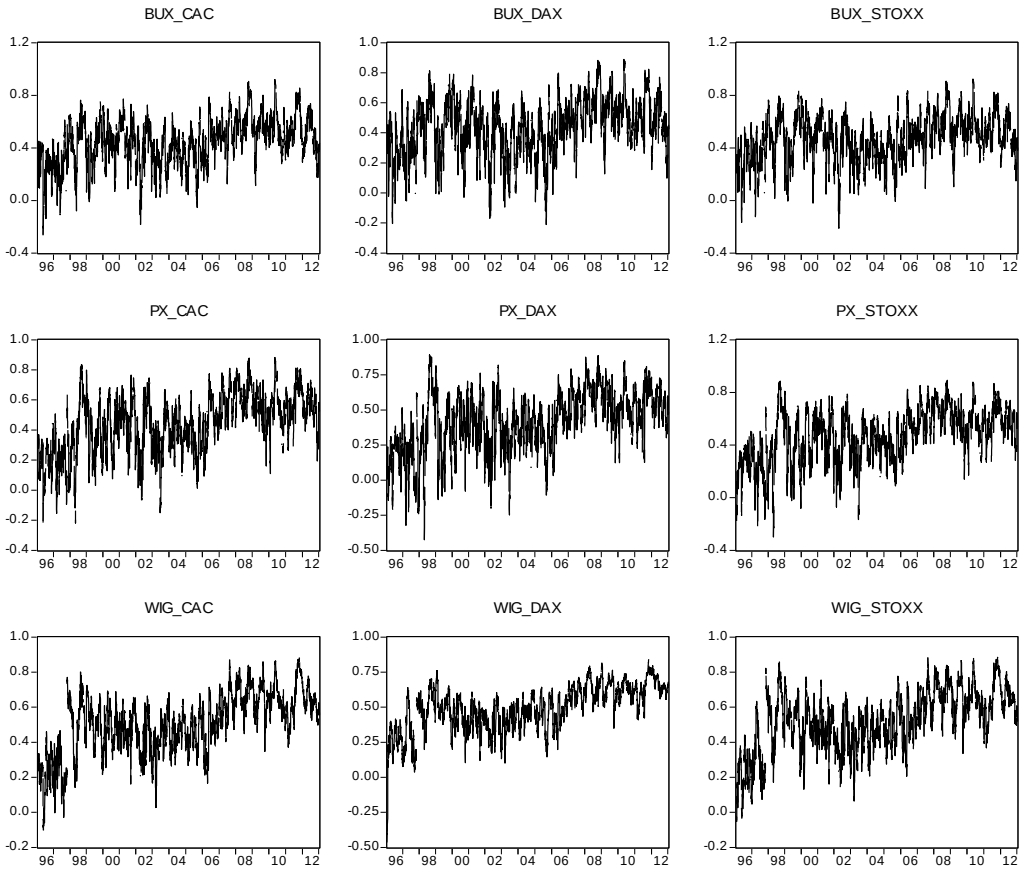


Figure 1 Conditional correlations from the BEKK-GARCH(1,1) model

	DLCAC		DLDAX		DLSTOXX	
DLBUX	$6{,}36.10^{-5}$***	-0,000123***	$5{,}08.10^{-5}$***	-0,000148***	$4{,}44.10^{-5}$***	-0,000124***
DLPX	$8{,}79.10^{-5}$***	$-3{,}09.10^{-5}$**	$9{,}26.10^{-5}$***	$-5{,}83.10^{-5}$***	$8{,}45.10^{-5}$***	$-4{,}22.10^{-5}$***
DLWIG	$8{,}97.10^{-5}$***	$-3{,}38.10^{-5}$***	$8{,}82.10^{-5}$***	$2{,}95.10^{-5}$***	$8{,}24.10^{-5}$***	$-1{,}10.10^{-5}$

Table 4 Trend in conditional correlations – the first columns (period: January 8, 1996 – January 21, 2013), the second columns (period: October 10, 2008 – January 21, 2013)[11]

4 Conclusion

In this paper we analysed the stock market integration of the Central Eastern European (CEE) stock markets (Czech, Hungarian and Polish) vis-à-vis the Western European stock market based on BEKK-GARCH models using daily closing values of individual stock market indices during the period from January 8, 1996 to January 21, 2013. We also investigated the impact of the current global financial crisis on the corresponding conditional correlations.

The whole analysis was based on logarithmic stock returns using the unconditional correlations as well as the conditional correlations from the BEKK-GARCH(1,1) models. Concerning the unconditional correlation coefficients of CEE markets vis-à-vis Western European markets these varied between 0,487 and 0,567 during the whole analysed period, but became higher during the crisis period (0,628 – 0,699). Forbes and Rigobon [14] showed that although this is the typical conclusion of the studies based on correlation coefficients, these coefficients are biased and inaccurate due to heteroscedasticity in market returns. In case of conditional correlations from the BEKK-GARCH(1,1) models the results were different. Since the values ranged during the whole analysed period between 0,415 and 0,525, no higher intensity of the stock market integration was proved during the crisis period. The results concerning the height of conditional correlation coefficients are more or less comparable to those of other studies (see e. g. [2], [15], [19]). In order to assess the impact of the crisis period, the con-

[11] Note: The symbols *** and ** denote the statistical significance of the parameter at the 0,01 and 0,05 significance level.

cluding remarks of Horvath and Petrovski [15] are similar as ours. Although they did not investigate separately the pre-crisis and crisis periods, just based on the graphical illustration they came to the conclusion that the conditional correlations in the Czech Republic and Hungary decreased somewhat during the crisis period and in general the crisis does not have impact on the strength of stock market integration between CEE countries and Western Europe. On the other side, Wang and Moore in [19] tested the impact of various crises and confirmed strengthened integration between CEE countries and Western European countries. They also provide the survey of several other studies which confirmed that international correlation increased in crisis periods.

Acknowledgements

This paper is supported by the Grant Agency of Slovak Republic - VEGA, grant no. 1/0595/11 "Analysis of Business Cycles in the Economies of the Euro Area (with Regards to the Specifics of the Slovak Economy) Using Econometric and Optimization Methods".

References

[1] Asongu, S. A.: The 2011 Japanese earthquake, tsunami and nuclear crisis: evidence of contagion from international financial markets. *MPRA Paper* No. 31174, 2011.
[2] Baumöhl, E., and Farkašovská, M., and Výrost, T.: Integrácia akciových trhov: DCC MV-GARCH model. *Politická ekonomie* **4** (2010), 488-503.
[3] Bauwens, L., and Laurent, S., and Rombouts, J. V. K.: Multivariate GARCH Models: A Survey, *Journal of Applied Econometrics* **21** (2006), 79-109.
[4] Bollerslev, T.: Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics* **31** (1986), No. 3.
[5] Bollerslev, T., and Engle, R. F., and Wooldridge, J. M.: A Capital-Asset Pricing Model with Time-Varying Covariances, *Journal of Political Economy* **96** (1988), No. 1, 116-131.
[6] Bollerslev, T.: Modeling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model, *Review of Economics and Statistics* **72** (1990), 498-505.
[7] Brooks, C.: *Introductory Econometrics for Finance*. Cambridge University Press, Cambridge, 2008.
[8] Canarella, G., and Miller, S. M., and Pollard, S. K.: *NAFTA Stock Markets: Dynamic Return and Volatility Linkages*. Nova Science Publishers, New York, 2010.
[9] Caporin, M., and McAleer, M.: Do We Really Need Both BEKK and DCC? A Tale of Two Covariance Models. Available at: http://eprints.ucm.es/8590/1/0904.pdf (cit. 2013-02-15)
[10] Cappiello, L., and Engle, R.F., and Sheppard, K.: Asymmetric Dynamics in the Correlations Of Global Equity and Bond Returns. *Journal of Financial Econometrics* **4** (2006), No. 4, 537-572.
[11] Égert, B., and Kočenda, E.: Time-Varying Comovements in Developed and Emerging European Stock Markets: Evidence from Intraday Data. *William Davidson Institute*, Working Paper Number 861 (2007).
[12] Engle, R. F., Kroner, K. F.: Multivariate Simultaneous Generalised GARCH, *Econometric Theory* **11** (1995), 122-150.
[13] Engle, R. F.: Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business and Economic Statistics* **20** (2002), 339–350.
[14] Forbes, K. J., and Rigobon, R.: No Contagion, Only Interdependence: Measuring Stock Market Comovements. *Journal of Finance* **57** (2002), No. 5, 2223–2262.
[15] Horvath, R., and Petrovski, D.: International Stock Market Integration: Central and South Eastern Europe Compared. *William Davidson Institute Working Papers Series*, wp 1028 (2012).
[16] Huang, Y., and Su, W., and Li, X.: Comparison of BEKK GARCH and DCC GARCH Models: An Empirical Study. In: *Proceedings of the 6th in international conference on Advanced data mining and applications Part II*, Springer-Verlag, Berlin Heidelberg, 2010, 99-110.
[17] Mills, T. C., and Markellos, R.N.: *The Econometric Modelling of Financial Time Series*. Cambridge University Press, Cambridge, 2008.
[18] Schmukler, S.L.: Financial Globalization: Gain and Pain for Developing Countries. In: *Federal Reserve Bank of Atlanta* ECONOMIC REVIEW, Second Quarter 2004.
[19] Wang, P., and Moore, T.: Stock Market Integration for the Transition Economies: Time-Varying Conditional Correlation Approach. *The Manchester School* **76** (2008), No. 1, 116–133.
[20] http://www.stoxx.com
[21] http://stooq.com
[22] http://finance.yahoo.com
[23] http://www.pse.cz

Production Function and Human Capital: The Case of Visegrad Countries

Silvie Chudárková[1], Tomáš Verner[2]

Abstract. The Classical economists already emphasized the effect of education, more precisely skills and knowledge (human capital), to increase individual productivity and wealth of nation. Comprehensive concepts of theory of human capital appeared in the early Sixties of 20th century. In the last two decades, the human capital influence on economic growth and productivity was confirmed by many economic studies. The aim of this paper is to find out whether there is significant human capital effect on economic output in Visegrad group countries during the period 1999-2011. Augmented standard aggregate production function linking output to productive inputs – labour, physical capital, human capital and total factor productivity is used according to the new theories of economic growth. Many researchers express human capital as investment in education, schooling enrollments or average length of education. From our point of view, these variables do not seem to be appropriate indicators of human capital. The value of it is expressed by tertiary graduates. Annual data collected from Eurostat were employed. Due to shortcoming of observations panel data model was performed. We found out negligible, but positive and statistically significant influence of human capital on economic growth.

Keywords: economic growth, human capital, panel data model, production function.

JEL Classification: J24, C30
AMS Classification: 62P20

1 Introduction

The production function determines the interdependence of input production factors and output. The production function specifies the maximum volume of output which could be produced for a certain time period and for a given amount of inputs and with the combination of inputs. According to the considered number of factors which indicate the volume of production, the production functions are divided into single-factor and two-factor, more precisely multifactor production function. The aggregate production function describes the relationship of the volume of used input production factors, technological progress and output. The aggregate production function became the core of models of economic growth. The issue of economic growth accompanies economic theory for its whole existence. In the 20th century we can distinguish two important milestones in the development of the theories of economic growth. Firstly, Forties and Fifties brought Keynesian models of growth followed by the neoclassical models of growth. Solow's model is the best known neoclassical growth model and became the subject of the further research. Assumption that labour is identical in all countries belongs to the contentious issues of Solow's model. This model assumes that education and qualifications are coincident in all countries. Equally qualified labour allows apply the best available technologies to all countries, which is not really true. The poorer countries often have worse educational systems and suffer from a lack of qualified workers, who hinder the use of advanced technologies. This shortcoming should be removed by including the human capital in the production function, as an additional explanatory variable. The second wave of interest in the issue of growth appeared at the end of Eighties and Nineties. In particular this was reflected by developing models of endogenous growth, which contain human capital, inter alia.

The main aim of this paper is to find out whether there is significant positive human capital effect on economic output in Visegrad group countries. Furthermore, we can consider the higher effect of human capital the greater improvement in the competitiveness at the national level. Education, especially higher education, and training represents the fifth of twelve pillars of competitiveness at macroeconomic level [17]. Besides, national competitiveness the increasing significance of regions in concept of European Union deserves more attention

[1] Silesian University in Opava, School of Business Administration in Karviná, Department of Economics, Univerzitní nám. 1934/3, 733 40 Karviná, Czech Republic, chudarkova@opf.slu.cz.
[2] Silesian University in Opava, School of Business Administration in Karviná, Department of Economics, Univerzitní nám. 1934/3, 733 40 Karviná, Czech Republic, verner@opf.slu.cz.

especially because of the economic efficiency of regions representing the basis of competitiveness of the country [15].

The paper is structured as follows. In Section 2, we define production function and the position and importance of human capital in this function. In Section 3, we introduce and describe the dataset and specify the used methods. In Section 4, we present and discuss the results obtained from testing. Section 5 concludes the paper.

2 Human capital as a production factor

Becker [7] defined human capital as skills and adequate motivation to apply these skills. The main premise, which the human capital theory is based on, postulates that education increases the productivity of the individual. Each individual tries to optimise the return on their investment in education and will continue in the study until the rate of return on her investment in education will exceed the rate of returns of alternative investments. The benefit to the society is then the increased labor productivity of better educated members of society and also technological progress. Generally, economists of the 1960's were trying to determine how various factors contribute to economic growth.

As already been mentioned, we distinguish single-factor and two-factor, more precisely multifactor production functions. The single-factor production function expresses the relationship between one variable input used in the production (usually labour) and a range of output. However, two-factor production function is examined and discussed in the economic theory more often. The two-factor production function can be represented graphically by using isoquant. Generally the two-factor production function can be expressed as follows:

$$Y = f(L, K) \tag{1}$$

where Y is the real product, L is the labour and K is the capital.

A Cobb-Douglas production function became a significant and widely used two-factor production function. Its authors belong to representatives of the neoclassical growth theory. Cobb-Douglas production function thus involves two input production factors – labour and capital. Labour and capital are included in the production function in various proportions and combinations, with the possibility of their mutual and unlimited substitution and with a free pricing of production factors depending on the dynamics of labour and capital. In basic expression is constructed the Cobb-Douglas production function so that the production effect of one production factor could be represented by the effect of another [12]:

$$Y = f(A, L^{\alpha}, K^{\beta}) \tag{2}$$

where Y is the real product, A is transformation parameter, L is the labour, K is the physical capital, α is labour elasticity coefficient and β is capital elasticity coefficient. If $\alpha + \beta < 1$, then there are decreasing returns to scale (if $\alpha + \beta = 1$, constant returns are present, and $\alpha + \beta > 1$, increasing returns are observed). The transformation parameter A is a coefficient of the aggregate productivity of production factors, which refers to both factors (also called total factor productivity). We discover size of the parameters α and β through the statistical methods.

Human capital as a production factor is included in the new growth theory models. Total product in the endogenous growth models is determined by both physical capital and labor and human capital which is accumulated in through education in every individual. The new growth theory applies the extended Solow's model (3) [12] and production function presented in the following form (4) [5]:

$$Y = A \cdot L^{\alpha} \cdot K^{\beta} \cdot e^{rt} \tag{3}$$

where Y is the real product, A is transformation parameter, L is the labour, K is the physical capital, α is labour elasticity coefficient and β is capital elasticity coefficient and e^{rt} is time factor, which reflects the influence of quantitative changes in production, including technological progress,

$$Y = A_t (L_t)^{1-\alpha-\beta} K_t^{\alpha} H_t^{\beta} \tag{4}$$

where Y is the product, A shows the level of technology, L is the labour, K is the physical capital, H is the human capital, α and β determine the proportions of individual factors on the overall product.

P. M. Romer, R. E. Lucas (see [13]) and also N. G. Mankiw, D. Romer and D. Weil [14] contributed significantly to the development of endogenous growth models. Overall, the research in this area confirms the existence of a relationship between the development of education and economic growth. One of the conclusions of endogenous growth models is that economic growth depends partly on the level of human capital. It assumes that human capital is the source of production of new ideas. It is true that the more developed economy, the stronger the relationship of education to the economic growth. While in less developed countries the primary task of starting economic growth nationwide is to ensure primary education, in the developed countries on the other hand is to drive further economic growth primarily on ensuring tertiary education. Romer [16] in his work addresses the issue of differences between the education and experience on the one hand, and technological progress on the other. The main source of economic growth is technological progress, in his opinion. Mankiw, Romer and Weil [14] in their work tried to eliminate shortcomings of the Solow's model by including the human capital expressed as an investment in education. Simplified representation of the value of human capital, respectively identification of human capital investment in education, with the achieved level of education or the number of students in various stages of study, is often a prerequisite in empirical studies examining the human capital at the macroeconomic level.

Barro [3] and others find a strong positive correlation between schooling enrollment and the subsequent growth rate gross domestic product (GDP) per capita. Barro [4] states that the growth of human capital expressed as an average length of education by one year corresponds to an increase of GDP growth by four percentage points a year. Bassanini and Scarpetta [6] states, that their results point to a positive and significant impact of human capital accumulation to output per capita growth. If the average length of study period is ten years, one additional year of study will increase production by six per cent. The existence of correlations between human capital, in this case the number of university graduates, and economic growth in their work was also confirmed by De la Fuente and Donénech [9]. Through that research the need for investment in human capital can be justified. However, there are views that refute or do not confirm the influence of human capital on economic growth. Bils and Klenow [8] in their study do not disprove any correlation between economic growth and human capital. However, they concluded that it is the level of gross domestic product, respectively its growth, leading to a higher level of human capital in the economy. Unlike previous studies on the causality of these variables this one is seen in the reverse order. Söderbom and Teal [18] came to the conclusion that human capital has a small, and not statistically significant effect, on the level of output. In addition, a more detailed analysis of the labour market during the period 1999 – 2010 in Visegrad group countries can be found in Tvrdon [19].

3 Data and econometric methodology

We employ annual data for Visegrad Group countries between 1999 and 2011. GDP at constant prices (2005) in EUR was used as a proxy variable of real product (Y), gross fixed capital formation at constant prices (2005) in EUR for physical capital (K), while for human capital (H) we used sum of graduates in upper secondary education, post-secondary non-tertiary education, first and second stages of tertiary education (i.e. ISCED'97 levels 3, 4, 5, 6). Worked hours were used for labour (L). All of the variables were collected from Eurostat database.

To examine the above mentioned relationship we perform panel data analysis. Panel data (or longitudinal data) cover both a time series and a cross-sectional dimension compared to pure time series or cross-sectional data [20]. Panel data models have become more and more popular among researchers because of their capacity for capturing the complexity of human behavior as contrasted to cross-sectional or time series data models [10]. Klevmarken [11], Hsiao [10] and Baltagi [2] list a number of panel data's benefits, e.g. (i) controlling for individual heterogeneity, (ii) give more informative data, more variability, less collinearity among the variables, more degrees of freedom and more efficiency, (iii) are better able to study dynamics of adjustment, (iv) are better capable to identify and measure effects that are simply not detectable in pure cross-section or pure time series data, (v) allow to construct and test more complicated behavioral models than purely cross-section or time series data and thus allow a researcher to analyze a number of important economic questions that con not be addressed using one dimensional data, and limitations, e.g. (i) design and data collection problem, (ii) distortions of measurement errors or (iii) selectivity problem.

A panel data set is formulated by a sample that contains N cross-sectional units (individuals, firms, households, countries etc.) that are observed at different time periods T [1], i.e. N = 4 and T = 13 in our case. Simple linear panel data model can be written as (5):

$$y_{it} = \alpha + \beta X'_{it} + u_{it} \tag{5}$$

where y represents the dependent variable, X vector of explanatory variables and subscript i denotes cross-section dimension (V4 countries) whereas t time series dimension (1999-2011), α, β are coefficients and u is a random disturbance term. In general, three different methods can be used to estimate linear panel data models by means of ordinary least squares: (i) common constant as in equation (5), (ii) fixed effects and (iii) random effects. The common constant method implies that there are no differences among variables of the cross-sectional dimension, so-called homogenous panel. Fixed or random effects allow us to capture the differences among units; hence the random disturbance term u is given by (6):

$$u_{it} = \mu_i + \nu_{it} \tag{6}$$

where μ_i denotes unobservable individual-specific effect which is time-invariant and is responsible for any individual-specific effect that is not contained in the regression. The term ν_{it} denotes remainder disturbance which varies over individuals and time [2], [10]. But the question, which model is more appropriate still remains. For common constant and fixed effect model we can apply standard F-test under the null hypothesis (H_0 thereinafter) that all the constants are the same [1]. In random effect model we assume zero correlation between explanatory variables and the unobserved effect. Hausman test can be employed to find out if this assumption is fulfilled under H_0: random effects are consistent and efficient.

Moreover, it should fulfill the assumptions for standard ordinary least squares error terms, i.e. the remained disturbance is homoskedastic, serially and spatial uncorrelated. In particular, the Cobb-Douglas production function is nonlinear in the parameters; hence we take the natural logarithm of (4) and obtain (7):

$$\ln Y = \ln A_t + (1-\alpha-\beta)\ln(L_t) + \alpha\ln(K_t) + \beta\ln(H_t) + \mu_i + \nu_{it} \tag{7}$$

4 Empirical results

In this section we present and discuss the results from Cobb-Douglas production function with human capital in the case of Visegrad countries. At first we estimate step by step all three models by means of least squares method. According to above mentioned tests the fixed effect model seems to be the most appropriate model to identify the effect of labour, physical and human capital on real product.

The coefficient of determination (R^2) = 0.98, the regression model and coefficients are statistically significant at 5 % level. The remained disturbance ν_{it} fulfills the assumptions for used methods.

When we estimate (7), anti-log, and we get (8) for the Czech Republic, (9) for the Hungary, (10) for the Poland and (11) for the Slovakia. The unobservable country-specific effect μ_i is represented by changes in the intercept.

$$Y_{CZ} = 16.62(L_t)^{0.09}(K_t)^{0.7}(H_t)^{0.28}\nu_{it} \tag{8}$$

$$Y_{HU} = 17.71(L_t)^{0.09}(K_t)^{0.7}(H_t)^{0.28}\nu_{it} \tag{9}$$

$$Y_{PL} = 14.12(L_t)^{0.09}(K_t)^{0.7}(H_t)^{0.28}\nu_{it} \tag{10}$$

$$Y_{SK} = 16.14(L_t)^{0.09}(K_t)^{0.7}(H_t)^{0.28}\nu_{it} \tag{11}$$

The estimated coefficients and signs are in accordance with expectations. The intercept represents total factor productivity. It differs across countries, it includes, e.g. technology, resources, climate and institutions. If we increase worked hours by 1 %, output (gross domestic product) goes up by about 0.09 %, holding the other inputs constant. Similarly, one more per cent of physical capital (gross fixed capital formation) causes output in-

creasing about 0.7 % (ceteris paribus). Increasing human capital (secondary and tertiary graduates) is associated with raising the output about 0.28 per cent.

5 Conclusion

The aim of this paper was to find out whether human capital positively affected economic output. The augmented Cobb-Douglas production function was used. It expresses relations among inputs and output. Annual data of Visegrad group countries during 1999 – 2011 were employed. Gross domestic product for output, gross fixed capital formation for physical capital, sum of graduates in upper secondary education, post-secondary non-tertiary education, first and second stages of tertiary education for human capital and worked hours for labour.

The panel data analysis was conducted to examine that relationship. Fixed effect method was chosen as the most appropriate, and total factor productivity differs across countries. The positive link among inputs and output was confirmed for Visegrad countries. If we increase physical capital input by one per cent, respectively human capital, respectively worked hours, output goes up by about 0.7 %, 0.28 % or 0.09 % ceteris paribus.

Acknowledgements

This paper was created in the students´ grant project "Seeking Factors and Barriers of the Competitiveness by Using of Selected Quantitative Methods". Project registration number is SGS/1/2012.

References

[1] Asteriou, D., and Hall, S. G. *Applied Econometrics: A modern Approach using EViews and Microfit*. Palgrave Macmillan, New York, 2007.
[2] Baltagi, B. H. *Econometric Analysis of Panel Data*. John Wiley & Sons, Chichester, 2005.
[3] Barro, R. J.: *Economic Growth in a Cross Section of Countries*. NBER Working Paper No. 3120, 1989.
[4] Barro, R. J: Human Capital and Growth. *The American Economic Review* **91**, No. 2 (2001), 12-17.
[5] Barták, M., and Vomáčková, H.: *Ekonomický rozměr lidského kapitálu*. Universita J. E. Purkyně v Ústí nad Labem, Ústí nad Labem, 2007.
[6] Bassanini, A., and Scarpetta, S.: *Does Human Capital Matter for Growth in OECD Countries?: Evidence from Pooled Mean-Group Estimates*. OECD Economics Department Working Papers No. 282, 2001.
[7] Becker, G. S.: *Human Capital*. Columbia University Press, New York, 1964.
[8] Bils, M., and Klenow, P. J.: *Does Schooling Cause Growth or the Other Way Around?* NBER Working Paper No. 6393, 1998.
[9] De la Fuente, A., and Donénech, R.: *Human Capital in Growth Regression: How much Difference Does Data Quality Make?* OECD Economics Department Working Papers No. 262, 2000.
[10] Hsiao, C. *Analysis of Panel Data*.Cambridge University Press, Cambridge, 2003.
[11] Klevmarken, N. A. Panel studies: What can we learn from them? *European Economic Review* **33**, No. 2-3 (1989), 523-259.
[12] Kucharčíková, A.: The Human Capital as the Production Factor. *Scientific Papers of the University of Pardubice, Series D* **16**, No. 19 (1/2011), 125-133.
[13] Lucas, R. E.: On the Mechanics of Economic Development. *Journal of Monetary Economics* **22** (1988), 3-42.
[14] Mankiw, N. G., Romer, D., and Weil, D. N.: A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics* **107**, No. 2 (1992), 407-437.
[15] Nevima, J., Ramík, J.: Application of multicriteria decision making for evaluation of regional competitiveness. *27th International Conference on Mathematical Methods in Economics*. Prague, 2009, 239-244.
[16] Romer, P. M.: Endogenous Technological Change. *Journal of Political Economy* **98**, No. 5 (1990), S71-102.
[17] Schwab, K., et al.: *The Global Competitiveness Report 2011-2012*. World Economic Forum, Geneva, 2011.
[18] Söderbom, M., and Teal, F.: *Trade and Human Capital as Determinations of Growth*. CSAE Working Paper Series 2001-10, 2001.
[19] Tvrdon, M.: Impacts of the global economic crisis on the European labor market performance: the case of Visegrad Group countries. *Recent Advances in Business Administration. 5th WSEAS International Conference on Business Administration.* WSEAS Press, Puerto Morelos, 2011.
[20] Wooldridge, J. M.: *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Massachusetts, 2002.

Analysis of efficiency in multi-period systems: an application to performance evaluation in education

Josef Jablonsky[1]

Abstract. Data envelopment analysis (DEA) is a non-parametric method for efficiency and performance analysis of decision making units. The paper deals with production systems where decision making units are described by their inputs and outputs in several consecutive periods. The paper presents multi-period DEA models that allow evaluation the efficiency of DMUs within the whole production chain. Efficiency and super-efficiency DEA models for multi-period systems are formulated and illustrated on a case study. The study consists in analysis of research and teaching performance of Czech economic faculties in four years period from 2007 until 2010. The model considers two inputs (number of academic employees and labour costs), two outputs for teaching efficiency (number of students and number of graduated) and the number of publications in various important categories and the number of RIV points for research efficiency.

Keywords: data envelopment analysis, performance, efficiency, multi-period models

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Data envelopment analysis (DEA) is a non-parametric technique for evaluation of relative efficiency of decision making units described by multiple inputs and outputs. This approach that was firstly introduced in (Charnes et al., 1978) is based on solving a sequence of linear programming problems. Conventional DEA models analyze relative technical efficiency of the set of n decision making units (DMUs) that are characterized by m inputs and r outputs in one period. The efficiency score θ_q of the DMU_q is defined as the weighted sum of outputs divided by the weighted sum of inputs as follows:

$$\theta_q = \frac{\sum_{k=1}^{r} u_k y_{kq}}{\sum_{i=1}^{m} v_i x_{iq}}, \tag{1}$$

where u_k, $k = 1, 2, \ldots, r$ is the positive weight of the k-th output, v_i, $i = 1, 2, \ldots, m$ is the positive weight of the i-th input, and x_{ij}, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$ and y_{kj}, $k = 1, 2, \ldots, r$, $j = 1, 2, \ldots, n$ are non-negative values for the DMU_j of the i-th input and the k-th output respectively. Conventional DEA models maximize the efficiency score (1) under the assumption that the efficiency scores of all other DMUs do not exceed 1 (100%). This problem must be solved for each DMU separately, i.e. in order to evaluate the efficiency of all DMUs the set of n optimization problems must be solved. The presented problem is not linear in objective function but it can be modified using Charnes-Cooper transformation into a linear optimization problem and then solved easily. The transformation consists in maximization of the nominator or minimization of the denominator in expression (1). The constraints of this LP optimization problem express the upper bound for efficiency scores of all DMUs except the DMU_q and the unit sum of the denominator/nominator in (1). The model that maximizes the nominator in (1) is referenced as DEA input oriented model, the model that minimizes the denominator is DEA output oriented model. In both cases the DMUs with $\theta_q = 1$ are lying on the efficient frontier estimated by the model and denoted as efficient units. Otherwise the units are inefficient and the efficiency score can be explained as a rate for increasing inputs or reduction outputs for reaching maximum efficiency.

The above mentioned models are conventional DEA models with input or output orientation. They can be further modified according to returns to scales and other assumptions. Among them are super-efficiency models that aim with ranking of efficient units as they receive the same maximum efficiency score in conventional

[1] University of Economics, Prague, Department of Econometrics, W. Churchill Sq. 4, 13067 Praha 3, Czech Repubic, e-mail: jablon@vse.cz

models. Another category of DEA models are network models that evaluate efficiency of production units with parallel or serial structure. Another stream in theory and practice of DEA models deals with evaluation of efficiency in multi-period production systems. In this case the models estimate the total efficiency in the context of time serial data. There were proposed various models how to deal with time serial data in DEA models in the past. Malmquist index (Fare and Grosskopf, 1996) and dynamic analysis (Sueyoshi and Sekitani, 2005) are only two of them that are relatively widely applied. Sam Park and Park (2009) formulated a two-stage DEA model that evaluates aggregative efficiency in multi-period case (further referenced as PP model).

The paper aims at multi-period DEA models, formulates a modification of the model presented in (Sam Park and Park, 2009) and applies the presented models to evaluation of teaching and research efficiency of Czech economic faculties. The paper is organized as follows. Section 2 formulates PP model and its modification. Section 3 presents the application of the models presented in previous parts of the paper on the data set of Czech economic faculties. The results of numerical experiments are analyzed and discussed. The final section of the paper summarizes presented results.

2 Multi-period DEA model

Conventional DEA models can be formulated either in their primal or dual form - see e.g. (Zhu, 2003). In this section we will use the dual form that has several advantages with respect to formulation of multi-period models. Let us suppose that the DMUs are described by the same set of inputs and outputs in T consecutive time periods $t = 1, 2, \ldots, T$, and assume that x^t_{ij}, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, and y^t_{kj}, $k = 1, 2, \ldots, r$, $j = 1, 2, \ldots, n$ are the values of the i-th input and the k-th output in the t-th period of the DMUj. The first phase of the output oriented PP model with constant (variable) returns to scale assumption that evaluates the aggregative efficiency of the DMU$_q$ is formulated as follows:

$$
\begin{aligned}
&\text{Maximize} && \theta_q && \\
&\text{subject to} && \sum_{j=1}^{n} x^t_{ij}\lambda^t_j \le x^t_{iq}, && i=1,2,\ldots,m, t=1,2,\ldots,T \\
& && \sum_{j=1}^{n} y^t_{ij}\lambda^t_j \ge \theta_q y^t_{iq}, && k=1,2,\ldots,r, t=1,2,\ldots,T \\
& && \left(\sum_{j=1}^{n} \lambda^t_j = 1, \right. && \left. t=1,2,\ldots,T\right) \\
& && \lambda^t_j \ge 0, && j=1,2,\ldots,n, t=1,2,\ldots,T,
\end{aligned}
\qquad (2)
$$

where λ^t_j $j = 1,2,\ldots,n$, $t = 1, 2,\ldots, T$ are variables of the model – coefficients of linear (convex) combination of the DMUs in time period t, and θ_q is the total aggregative efficiency score of the DMU$_q$. Let θ^*_q is the optimal objective function value of model (2). Then the second phase of the PP model is as follows:

$$
\begin{aligned}
&\text{Maximize} && \sum_{t=1}^{T}\sum_{i=1}^{m} s^{t-}_i / x^t_{iq} + \sum_{t=1}^{T}\sum_{k=1}^{r} s^{t+}_k / y^t_{iq} && \\
&\text{subject to} && \sum_{j=1}^{n} x^t_{ij}\lambda^t_j + s^{t-}_i = x^t_{iq}, && i=1,2,\ldots,m, t=1,2,\ldots,T \\
& && \sum_{j=1}^{n} y^t_{ij}\lambda^t_j - s^{t+}_k = \theta^*_q y^t_{iq}, && k=1,2,\ldots,r, t=1,2,\ldots,T \\
& && \left(\sum_{j=1}^{n} \lambda^t_j = 1, \right. && \left. t=1,2,\ldots,T\right) \\
& && s^{t-}_i \ge 0, && i=1,2,\ldots,m, t=1,2,\ldots,T, \\
& && s^{t+}_k \ge 0, && k=1,2,\ldots,r, t=1,2,\ldots,T, \\
& && \lambda^t_j \ge 0, && j=1,2,\ldots,n, t=1,2,\ldots,T,
\end{aligned}
\qquad (3)
$$

where s_i^{t-} and s_k^{t+} are slack/surplus variables belonging to constraints of model (2). According to the results of both stages the DMUs can be classified as follows:

- (Fully) efficient DMUs – the optimal objective function value of model (2) $\theta^*_q = 1$ and matrices of slack and surplus variables are zero matrices.
- Weakly efficient DMUs - the optimal objective function value of model (2) $\theta^*_q = 1$ and at least one element of matrices of slack/surplus variables is positive.
- Inefficient units - the optimal objective function value of model (2) $\theta^*_q > 1$ (for output oriented models).

The second phase of the PP model is slightly modified comparing to its original formulation. The original formulation contains in its objective function a simple sum of slack and surplus variables which seems to be meaningless due to possible high differences in input/output values. That is why a sum of relative slack and surplus variables is used in objective function of model (3).

In order to evaluate the DMUs, model (2) must be solved in the first phase and depending on the optimal objective function value the second phase follows. This phase must be applied when the optimal objective function value of the first phase is $\theta^*_q = 1$, i.e. the DMU under evaluation is recognized as fully or weakly efficient. In this case the second phase allows determination whether this unit is fully or weakly efficient. The inefficient or weakly efficient DMUs can be ranked according to the optimal objective function values from the first or second phase respectively. Nevertheless, the efficient units cannot be ranked at all using the PP model as proposed in (Sam Park and Park, 2009). In single period systems, this problem is solved by using of super-efficiency models or any other models that allow ranking of efficient DMUs. Information about the most important models of this class can be found e.g. in (Jablonsky, 2009).

First super-efficiency model was introduced by Andersen and Petersen (1993). Their model is based on the idea to remove the unit under evaluation from the set of units and then measure its distance from the new efficient frontier. The super-efficiency score of an originally efficient DMU is lower than one (taking into account the output oriented model) and can be explained as a rate for possible worsening of outputs in order to keep the efficiency status. A similar approach as in Andersen and Petersen model can be used for multi-period models too. For these purposes model (2) can be extended by additional constraints that ensure zero values for all weights of the DMU under evaluation, i.e.

$$\lambda_q^t = 0, \qquad t = 1,2,\ldots,T, \tag{4}$$

The advantages of this modified model consist in a possibility to discriminate better among fully efficient DMUs and in its higher computational efficiency because the second phase, i.e. the model (3), has to be solved for weakly efficient units only. It is clear that the modified model must be solved for each DMU ones in its first phase and usually only in a few cases (weakly efficiency) in its second phase in order to get the complete ranking of DMUs.

Let us denote φ^*_q optimal objective function value of model (2) with the constraint set extended by (4). The DMUs can be ranked according to φ^*_q values as follows:

- (Fully) efficient DMUs, i.e. the units with $\varphi^*_q < 1$ (for output oriented models), can be ranked according to this value – lower values indicate higher rank.
- Weakly efficient DMUs, i.e. the units with $\varphi^*_q = 1$, can be ranked according to the results of model (3) – of course $\theta^*_q = \varphi^*_q = 1$ in this second phase model.
- Inefficient DMUs, i.e. units with $\varphi^*_q > 1$, can be ranked as the fully efficient units – lower values of φ^*_q indicate higher rank.

3 Multi-period efficiency analysis of Czech economic faculties

Evaluation of efficiency of teaching process and research outputs of higher educational institutions belongs to important and widely discussed problems. The importance of this task consists in the necessity to evaluate the performance of the institution in both educational and research areas in order to have appropriate records for allocation of funds for future periods. Due to continuing discussion about future modifications of educational system in higher education in the Czech Republic modelling in this field is of a high importance. There are many open problems connected with the mentioned task. One of them is the selection of appropriate indicators that describe the performance of the evaluated institutions − they can be both of quantitative and qualitative nature and it is not always easy to get their numerical expression. Another problem consists in the selection of a modelling tool for analysis of the given data set. Various econometric and statistical models that are usually based on estimation of parameters of production functions that explain the performance by several defined indicators are widely used in this context. In this section of the paper we are trying to apply DEA models presented in previous

section for teaching and research efficiency evaluation of 19 economic faculties of Czech public universities. Data set available for the analysis contains information about various economic, teaching and research characteristics in a four years period starting in 2007. Any attempt to apply DEA models in any field including applications in education has to contain several steps:

1. Definition of inputs. Two input variables used further in analysis are the number of academic staffs and labour cost of the faculty.
2. Definition of outputs. There are considered two groups of output variables - one for teaching efficiency and one for research efficiency. Teaching efficiency is described by the total number of students and the number of graduated students. Research efficiency is influenced by the number of publications in various categories (books, papers in journals with a positive impact factor - Jimp, and papers in proceedings indexed in Web of Science - CPCI). The overall quality of publication is measured by so called RIV points. Finally we decided to work with two sets of outputs. The first set contains two teaching outputs and research is measured by RIV points only. The second set of outputs consists of two teaching outputs as in the first set and three research outputs – numbers of publications in three important categories as mentioned above.
3. Selection of an appropriate DEA model and its assumptions (returns to scale, weight restrictions, etc.). The multi-period DEA model with constant returns to scale assumption was further applied in numerical experiments concerning evaluation of overall efficiency of Czech economic faculties from 2007 until 2010.
4. Analysis of the results given by the model(s).

The data set (two inputs, two "teaching" outputs and four "research" outputs) for all 19 economic faculties for the last year of the given time period (2010) is presented in Table 1. The complete data set for the remaining years cannot be presented due to the space limits.

Faculty	Academic staff	Labour costs	# of students	# of graduates	Books	Jimp	CPCI	RIV points
FSV UK	137	57831	4105	821	37.39	5.58	24.35	3632
EkF JČU	69	26842	1764	522	2.74	7.00	1.50	578
FSE UJEP	67	26246	2200	559	2.08	2.13	0.00	122
ESF MU	93	49739	4453	738	12.66	11.67	8.00	1064
OPF SU	108	44908	4385	882	3.77	14.07	9.33	853
FE ZČU	61	20063	2312	519	3.04	3.67	3.17	367
HF TUL	83	32510	2081	600	12.83	6.00	5.67	988
FES UP	78	35977	2639	556	5.39	9.17	3.07	933
FP VUT	81	30280	2758	821	4.53	6.50	2.63	538
EkF VŠB	175	71448	6539	1701	24.13	15.67	7.51	1684
FME Zlín	84	28277	3419	970	5.99	29.03	3.50	889
FFU VŠE	82	42899	3176	805	2.84	1.50	10.50	946
FMV VŠE	172	71074	4713	1301	13.30	0.00	0.00	902
FPH VŠE	106	47113	3778	1022	4.61	3.67	3.33	422
FIS VŠE	100	43880	3332	686	4.35	48.22	11.54	1274
NH VŠE	65	28621	2572	462	5.41	2.83	15.12	1064
FM VŠE	39	16542	1437	321	0.16	8.84	2.00	277
PEF ČZU	186	121546	9462	2822	12.79	15.00	5.62	1648
PEF MZLU	114	49361	3658	958	4.48	4.00	6.50	1151

Table 1 Data set – 2010.

Computational experiments were realized with two set of outputs (I and II) as mentioned above with two four-period DEA models under the assumption of constant returns to scale – the original PP model (2) and (3) and its super-efficiency modification that is extended by constraints (4). All calculations were performed using our own codes written in LINGO modelling language. The results are presented in Table 2. This table contains the following information:

- Efficiency scores of all DMUs for both sets of outputs given by the original PP model – as its output oriented version is used the higher values of the scores mean lower rank.
- Rank of the units (faculties) according to the efficiency scores given by the PP model.

- Super-efficiency scores of all DMUs for both sets of outputs given by modification of the PP model (2) – due to output orientation of models the lowest score means the highest rank.
- Rank of the units (faculties) according to the super-efficiency scores given by modification of the PP model.

Faculty	Eff.score I PP model	Eff.score II PP model	Rank I	Rank II	Eff.score I Super PP	Eff.score II Super PP	Rank I	Rank II
FSV UK	1.0000	1.0000	1	1	0.4536	0.4474	1	2
EkF JČU	1.0798	1.1439	12	17	1.0798	1.1439	12	17
FSE UJEP	1.0413	1.0413	8	12	1.0413	1.0413	8	12
ESF MU	1.0000	1.0000	1	1	0.9611	0.9043	6	8
OPF SU	1.0731	1.0506	10	13	1.0731	1.0506	10	13
FE ZČU	1.0492	1.0239	9	11	1.0492	1.0239	9	11
HF TUL	1.3595	1.1518	18	18	1.3595	1.1518	18	18
FES UP	1.2344	1.1175	17	16	1.2344	1.1175	17	16
FP VUT	1.0834	1.0565	13	14	1.0834	1.0565	13	14
EkF VŠB	1.1257	1.0000	14	1	1.1257	0.8934	14	7
FME Zlín	1.0000	1.0000	1	1	0.7002	0.6251	2	3
FFU VŠE	1.0000	1.0000	1	1	0.8167	0.7035	5	4
FMV VŠE	1.4650	1.3068	19	19	1.4650	1.3068	19	19
FPH VŠE	1.0767	1.0000	11	1	1.0767	0.9596	11	9
FIS VŠE	1.2142	1.0000	16	1	1.2142	0.4014	16	1
NH VŠE	1.0000	1.0000	1	1	0.7993	0.7355	4	5
FM VŠE	1.1907	1.0000	15	1	1.1907	0.9737	15	10
PEF ČZU	1.0000	1.0000	1	1	0.7580	0.7391	3	6
PEF MZLU	1.0298	1.1072	7	15	1.0298	1.1072	7	15

Table 2 Efficiency scores and rank of faculties given by the PP model

The comparison of results given by the original PP model and its super-efficiency modification is quite clear. The original procedure cannot rank fully efficient DMUs and as it is shown, among the DMUs 6 of them are efficient (10 of them in case of the second set of outputs), i.e. 6 or even 10 units cannot be ranked by using the original model. The modified procedure allows ranking of all units easily. Another question is a discussion about the contribution of the presented PP model and its modification. As it is clear the efficiency (super-efficiency) score of this multi-period model is given as the best (minimum in case of output oriented model) efficiency score of particular periods. That is why the advantage of the PP model consists in a possibility to get the multi-period efficiency score by solving one optimization problem for each DMU instead of solving T optimization problems for each DMU.

The results of models with two different sets of outputs presented in Table 2 are quite contradictory. It shows that it is very important to pay attention to a proper selection of input and output variables of the model. The models "I" contain only one output – the number of RIV points. This number is given directly as the weighted sum of the numbers of publications in particular categories (each category has its weight reflecting its importance). The models "II" contain three outputs – books, Jimp and CPCI proceedings. The results in Table 2 show that all DMUs that are fully efficient in models "I" are fully efficient in models "II" as well. Nevertheless, the opposite relation does not hold. Some of inefficient units in models "I" are efficient in models with three outputs. An extreme situation occurs for the Faculty of Informatics and Statistics, University of Economics, Prague (FIS VSE). This faculty is inefficient in models "I" and its efficiency score is quite high and that is why its rank is quite low (16). In the contrary FIS VSE is efficient in models "II" and its super-efficiency score is the lowest, this means that it is on the highest rank among all faculties. This difference can be explained quite easily by analysis of the source data set presented in Table 1. FIS VSE has the highest number of publication in CPCI proceedings in 2010, it is much higher than the other faculties have. This fact causes that the faculty is fully efficient and leads to a quite low value of the super-efficiency score as well. This example demonstrates very well that it is necessary to build the model carefully and explain its results in the context of reality.

4 Conclusions

Analysis of efficiency of DMUs within multiple periods is an important task and many various models were proposed for these purposes in the past. The models presented in this paper and applied for evaluation of efficiency of Czech economic faculties is one of the last contributions in this area. The advantage of the proposed super-efficiency multi-period DEA model consists in its computational efficiency because it is necessary to solve only one optimization problem for each DMU in order to get its aggregative (super-) efficiency score. Disadvantage is the fact that the final results depend on the efficiencies of the DMUs in particular periods without any interconnections among periods. Another problem consists in a possible infeasibility of the optimization problem (2) with additional constraints (4) under the assumption of variable returns to scale. This problem is discussed e.g. in (Lee and Zhu, 2012). A future research can be focused on analysis of multi-period production systems with interconnections among the periods and using other DEA models than the conventional ones are, e.g. SBM models proposed in (Tone, 2002).

Acknowledgements

The research is supported by the Czech Science Foundation, project no. P403/12/1387.

References

[1] Andersen, P., and Petersen, N.C.: A procedure for ranking efficient units in data envelopment analysis. *Management Science*, **39**, 1261-1264, (1993).

[2] Charnes, A., Cooper, W. W. and Rhodes, E., Measuring the efficiency of decision making units. *European Journal of Operational Research*, **2**, 6, 429–444 (1978).

[3] Fare, R. and Grosskopf, S., *Intertemporal production frontiers with dynamic DEA*. Boston: Kluwer Academic Publishers (1996).

[4] Jablonsky, J., Comparison of ranking methods in data envelopment analysis models. In: *Mathematical Methods in Economics.* University of South Bohemia, Ceske Budejovice, 302–308 (2010).

[5] Lee, H.S. and Zhu, J., Super-efficiency infeasibility and zero data in DEA. *European Journal of Operational Research*, **216**, 429-433 (2012).

[6] Sam Park, K. and Park, K., Measurement of multiperiod aggregative efficiency. *European Journal of Operational Research*, **193**, 567-580 (2009).

[7] Sueyoshi, T. and Sekitani K., Returns to scale in dynamic DEA. *European Journal of Operational Research*, **161**, 536-544 (2005).

[8] Tone, K.: A slack-based measure of super-efficiency in data envelopment analysis. *European Journal of Operational Research*, **143**, 32-41 (2002).

[9] Zhu, J. *Quantitative models for performance evaluation and benchmarking: Data envelopment analysis with spreadsheets and DEA Excel Solver*. Boston: Kluwer Academic Publishers (2003).

Public service system design with disutility relevance estimation

Jaroslav Janáček[1], Marek Kvet[2]

Abstract. This paper deals with the problem of designing the optimal structure of most public service systems such that the users´ discomfort is minimized. The discomfort is usually proportional to the sum of demand-weighted network distances between the users and the nearest source of provided service. We assume that all facilities have equal setup cost and enough capacity to serve all users. Real instances of the weighted *p*-median problem are characterized by big number of possible service center locations, which can take the value of several thousands. In such cases the location-allocation approach usually fails due to enormous computational time or huge memory demands. This weakness has led to the approximate covering approach based on specific model reformulation. The disutility values are here estimated by some upper and lower bounds given by so called dividing points. Deployment of dividing points influences the solution accuracy. It is based on the idea that some disutility values can be considered relevant and are expected to obtain in the optimal solution. Hereby, we are studying various approaches to the relevance with their impact on the solution accuracy and saved computational time.

Keywords: weighted *p*-median problem, approximate covering model, disutility relevance estimation

JEL Classification: C61
AMS Classification: 90C06, 90C10, 90C27

1 Introduction

Design of almost any public service system [2], [4], [10] includes determination of center locations, from which the associated service is distributed to all users of the system. The service providing facilities must be usually concentrated to a limited number of centers due to economic and technological reasons [8], [11]. Regardless of the case whether the service is delivered to the users or the users travel for the service to the nearest center along the shortest path on the transportation network, which covers the serviced area. Thus the public service system structure is formed by the deployment of limited number of service centers and the associated objective in the standard formulation is to minimize the social costs, which are proportional to the distances between serviced objects and the nearest service centers. The social costs can be denoted as a disutility, to which a system user is exposed. In other words, the standard approach to a public service system design assumes that the user is serviced from the nearest located service center or from the center, which offers the smallest disutility to the user. Thus a mathematical model of the public service system design problem often takes the form of a large weighted *p*-median problem, where the numbers of serviced users and possible service center locations take the value of several thousands. The number of possible service center locations seriously impacts the computational time [9]. The necessity of solving large instances of the *p*-median problem as a predecessor to the weighted *p*-median problem has led to the approximate approach, which enables to solve real-sized problems in admissible time [1], [3], [5], [6]. The approximate approach is based on the upper bound minimization and performs as a heuristic, where the lower bound of the optimal value of the objective function is easy to obtain. The accuracy of the approach is given by a convenient determination of so-called dividing points, which are used for the upper approximation of the objective function. Based on the radial formulation of the classical *p*-median problem, we suggest the approach to the design of a public service system. The approach generalizes the radial approaches to the classical *p*-median problem to obtain a solving technique for the large weighted *p*-median problem, which is the core of the approach to the public service system design, where the system disutility is to be minimized.

[1] Prof. RNDr. Jaroslav Janáček, CSc., University of Žilina, Faculty of Management Science and Informatics, Univerzitná 1, 010 26 Žilina, Slovak Republic, e-mail: Jaroslav.Janacek@fri.uniza.sk
[2] Ing. Marek Kvet, University of Žilina, Faculty of Management Science and Informatics, Univerzitná 1, 010 26 Žilina, Slovak Republic, e-mail: Marek.Kvet@fri.uniza.sk

2 Public service system design model

The public service system design with minimal disutility is a task of location of at most p service centers so that the sum of individual disutilities of each user coming only from the nearest located service center is minimal. To describe the problem, we denote I a set of possible service center locations and J the set of possible users' locations. The symbol b_j denotes the number of users located at the location j. The disutility for a user at the location j following from the possible center location i is denoted as d_{ij}. Under assumption that the user's disutility originates only from the nearest service center, we can state the problem as follows:

$$\textit{Minimize} \qquad \left\{ \sum_{j \in J} b_j \min\left\{ d_{ij} : i \in I_1 \right\} : I_1 \subset I, |I_1| \le p \right\} \tag{1}$$

The symbol I_1 denotes a subset of the former set of all possible service center locations. The problem (1) is also known as the weighted p-median problem, which is broadly discussed in [1], [3], [5], [6] from the viewpoint of solving techniques suggested for fast solving of the huge instances. Within this paper, we focus on so called radial formulation used in the above papers and we explore the approaches based on the set of dividing points.

The strategic decision in the problem (1) concerns location of centers at possible center locations from the set I. To model this decision at particular location, we introduce a zero-one variable $y_i \in \{0, 1\}$, which takes the value of 1, if a center should be located at the location i, and it takes the value of 0 otherwise. To obtain an upper or a lower bound of the original objective function, the range $[d_0, d_m]$ of all possible disutilities $d_0 < d_1 < \ldots < d_m$ from the matrix $\{d_{ij}\}$ is partitioned into $r+1$ zones. The zones are separated by a finite ascending sequence of so called *dividing points* $D_1, D_2 \ldots D_r$ chosen from the sequence $d_0 < d_1 < \ldots < d_m$, where $0 = d_0 = D_0 < D_1$ and also $D_r < D_m = d_m$. The zone k corresponds with the interval $(D_k, D_{k+1}\rangle$. The length of the k-th interval is denoted by e_k for $k = 0 \ldots r$. In addition, auxiliary zero-one variables x_{jk} for $k = 0 \ldots r$ are introduced. The variable x_{jk} takes the value of 1, if the disutility of the user at $j \in J$ from the nearest located center is greater than D_k and it takes the value of 0 otherwise. Then the expression $e_0x_{j0} + e_1x_{j1} + e_2x_{j2} + \ldots + e_rx_{jr}$ constitutes an upper approximation of the disutility d_{j*} from user location j to the nearest located service center. If the disutility d_{j*} belongs to the interval $(D_k, D_{k+1}\rangle$, then the value of D_{k+1} is the upper estimation of d_{j*} with the maximal possible deviation e_k. Let us introduce a zero-one constant a_{ij}^k for each triple $[i, j, k] \in I \times J \times \{0 \ldots r\}$. The constant a_{ij}^k is equal to 1, if the disutility d_{ij} between the user location j and the possible center location i is less or equal to D_k, otherwise a_{ij}^k is equal to 0. Then the radial-type weighted covering model can be formulated according to [5], [6] as follows:

$$\textit{Minimize} \qquad \sum_{j \in J} b_j \sum_{k=0}^{r} e_k x_{jk} \tag{2}$$

$$\textit{Subject to:} \quad x_{jk} + \sum_{i \in I} a_{ij}^k y_i \ge 1 \quad \forall\, j \in J,\ \forall\, k = 0, 1, \ldots, r \tag{3}$$

$$\sum_{i \in I} y_i \le p \tag{4}$$

$$x_{jk} \ge 0 \quad \forall\, j \in J,\ \forall\, k = 0, 1, \ldots, r \tag{5}$$

$$y_i \in \{0, 1\} \quad \forall\, i \in I \tag{6}$$

The objective function (2) gives the upper bound of the sum of original disutility values. The constraints (3) ensure that the variables x_{jk} are allowed to take the value of 0, if there is at least one center located in radius D_k from the user location j. The constraint (4) puts a limit p on the number of located facilities.

To obtain a lower bound of the objective function value of the optimal solution of the original problem, several approaches can be employed. We present here the simplest one, which can be introduced as the expression $e_0x_{j1} + e_1x_{j2} + e_2x_{j3} + \ldots + e_{r-1}x_{jr}$.

3 Dividing points deployment and disutility relevance

Deployment of the dividing points influences the size of the covering model (2) – (6) and the accuracy of the result. The dividing points can be chosen only from the set of values $d_0 < d_1 < \ldots < d_m$ of the disutility matrix

$\{d_{ij}\}$, where $D_0 = d_0$ and $D_m = d_m$. Let the value d_h have a frequency N_h of its occurrence in the matrix $\{d_{ij}\}$. In the suggested approaches, we start from the hypothesis that the disutility d_h from the sequence $d_0 < d_1 <\ldots< d_m$ occurs in the resulting solution n_h times and that is why the deviation of this disutility value from its approximation encumbers the total deviation proportionally to n_h. The disutility d for a user from the nearest located service center can be only estimated taking into account that it belongs to the interval $(D_k, D_{k+1}\rangle$. The maximal deviation of the upper estimation D_{k+1} from the exact value d is $D_{k+1} - D_k^1$, where D_k^1 is the minimal element of the sequence $d_0 < d_1 <\ldots< d_m$, which belongs to the interval $(D_k, D_{k+1}\rangle$. If we were able to anticipate a frequency n_h of each d_h in the unknown optimal solution, we could minimize the total deviation of the upper approximation from the unknown optimal solution by convenient deployment of dividing points. The dividing points for the upper approximation follow from the optimal solution of the problem described by (7) – (11). If the disutility value d_h belongs to the interval ending by the dividing point d_t then the decision variable z_{ht} takes the value of 1.

$$\textit{Minimize} \quad \sum_{t=1}^{m}\sum_{h=1}^{t}(d_t - d_h)n_h z_{ht} \tag{7}$$

$$\textit{Subject to}: \quad z_{(h-1)t} \le z_{ht} \quad \forall\, t = 2, 3, \ldots, m, \ \forall\, h = 2, 3, \ldots, t \tag{8}$$

$$\sum_{t=h}^{m} z_{ht} = 1 \quad \forall\, h = 1, 2, \ldots, m \tag{9}$$

$$\sum_{t=1}^{m-1} z_{tt} = r \tag{10}$$

$$z_{ht} \in \{0, 1\} \quad \forall\, t = 1, 2, \ldots, m, \ \forall\, h = 1, 2, \ldots, t \tag{11}$$

The link-up constraints (8) ensure that the disutility value d_{h-1} belongs to the interval ending with d_t only if each other disutility between d_{h-1} and d_t belongs to this interval. Constraints (9) assure that each disutility value d_h belongs to some interval and the constraint (10) enables only r dividing points to be chosen. After the problem (7) – (11) is solved, the nonzero values of z_{tt} indicate the disutility values d_t corresponding with dividing points.

The associated solving technique consists of the estimation of the relevance n_h, solving the dividing points deployment problem (7) – (11) and subsequently solving the radial-type weighted covering model (2) – (6).

The above approach is obviously based on the "relevance" of a disutility d_h, which expresses the strength of our expectation that the disutility value d_h will be a part of the unknown optimal solution, which is searched for. We suggested and explored several ways of the relevance estimation. In the first one, we started with the frequency N_h of d_h occurrence in the disutility matrix $\{d_{ij}\}$ and set the initial relevance at the value n_h given by the expression (12), where T is a positive shaping parameter.

$$n_h = N_h e^{\frac{-d_h}{T}} \tag{12}$$

The approach (12) follows the hypothesis that the relevance decreases exponentially with increasing disutility value [6]. The other approach to the relevance described in [7] uses the column ranking evaluation $L^{ts}_j(d_{ij})$ of the disutility value d_{ij} and the relevance n_h is defined according to (13).

$$n_h = L^{ts}_h(d_h) = \sum_{j\in J}\sum_{\substack{i\in I \\ d_h = d_{ij}}} L^{ts}_j(d_{ij}) \tag{13}$$

The linear column ranking function $L^{ts}_j(d_{ij})$ is defined as follows: Let $P_j(d_{ij})$ be the position of d_{ij} in the ascending sequence of the j-th column items of the disutility matrix $\{d_{ij}\}$ and let a denote the cardinality of I. Then $L^{ts}_j(d_{ij}) = a + s * (1 - P_j(d_{ij}))$ for $P_j(d_{ij}) < a + 1 - t$ and $L^{ts}_j(d_{ij}) = 0$ otherwise. The parameters t and s represent a *threshold* and a *step* respectively. The *threshold* influences the number of $t - 1$ largest disutility values of the j-th column, which are not taken into account and the *step* gives the difference between the contributions of the k-th and $(k - 1)$-th item of the ascending sequence of the j-th column items. The parameter t can vary over the range $[p \ \ldots \ a\text{-}1]$ of integers and the step s can take the value from the interval $\langle 0, a/(a\text{-}t)\rangle$. This ranking expression of the relevance can be modified using the exponential reduction. Then, the associated relevance n_h is defined in accordance to (14).

$$n_h = L^{ts}_h(d_h) e^{\frac{-d_h}{T}} \tag{14}$$

The last suggested approach to the relevance estimation comes from the exponential approach, but it takes into account that the slope of the exponential function is too steep in the neighborhood of zero disutility and that the relevant disutility values can be sparsely distributed in this neighborhood. To avoid the groundless reduction of the relevant disutility values, we moved the exponential function to the range of biggest disutility values. The shifted exponential approach computes the relevance n_h according to (15).

$$n_h = N_h \, g(h) \tag{15}$$

The function $g(h)$ is equal to 1 for each $h \leq h_{crit}$ and it is defined by (16) for $h > h_{crit}$.

$$g(h) = e^{-\frac{h-h_{crit}}{T}} \tag{16}$$

The constants T and h_{crit} are parameters of the approach. T is the shaping parameter and h_{crit} is the critical value, which can be determined according to (17) for some given parameter q.

$$h_{crit} = min\left\{h \in Z^+ : \sum_{u=0}^{h} N_u \geq \frac{q}{p} \sum_{t=0}^{m} N_t\right\} \tag{17}$$

4 Computational study

So far, we have presented the approximate covering method for public service system design problem. This method is based on the disutility relevance estimation, which can be performed in different ways. Hereby, we are presenting the results of numerical experiments, which were aimed at the comparison of four suggested approaches from the viewpoint of computational time and the solution accuracy. All experiments were performed using the optimization software FICO Xpress 7.3 (64-bit, release 2012). The associated code was run on a PC equipped with the Intel® Core™ i7 2630QM processor with the parameters: 2.0 GHz and 8 GB RAM.

Particular approaches to the disutility relevance estimation were tested on the pool of benchmarks, whose size varied from 300 to 1500 possible service center locations. The number of user locations of the designed system was the same as the cardinality of the set I. For each size of the problem, six instances were solved. These instances differ in the value of parameter p, which limits the number of located service centers. The value of p was set in such a way, that the ratio of $|I|$ to p equals 2, 3, 4, 5, 10 and 20 respectively. The reported results are the average values grouped by the number of possible service center locations. As the system disutility of the users located at j the network distance between the location and the nearest located service center was taken. The weight b_j of each $j \in J$ corresponds with the number of users located at the location j. Furthermore, we assume that each possible location of a service center has enough capacity to serve all users.

The optimal solution of the problem was obtained by solving the well-known location-allocation model in the optimization environment Xpress. Since the covering model provides only the approximation of the former objective function, its real value must be computed according to the values of location variables y_i, $i \in I$ and the disutility matrix $\{d_{ij}\}$. Mathematical formulation can be expressed by (18).

$$\sum_{j \in J} \min\left\{d_{ij} : i \in I,\, y_i = 1\right\} \tag{18}$$

In the covering model (2) – (6), the number r of dividing points was set to the value of 20 in all solved instances and the parameter T was set to the value of 1. The threshold t used in the expressions (13) and (14) was set to the quarter of its possible range, which means $t = (3p + a - 1)$ *div* 4, where the value of a is equal to $|I|$. The step s was set to the value of $\alpha / (\alpha - t)$. The value of q in the expression (17) was 2.

The quality of the solution obtained by the covering approach with any of the disutility relevance estimation is measured by *gap*, which is defined as follows: Let *ES* denote the objective function value of the **e**xact **s**olution of the problem and let *CS* be the real objective function value of the **c**overing **s**olution computed according to (18). The gap represents the difference between these values expressed in percentage of the exact solution.

$$gap = \frac{|ES - CS|}{ES} * 100 \tag{19}$$

Time comparison of the suggested approaches is plotted in Table 1, which contains average computational time in seconds for each approach and each size of the solved problems. It is important to note, that the reported value of the covering approach includes two optimization processes – dividing points deployment by model (7) –

(11) and solving covering problem (2) – (6). As we can see in the table, the approximate covering method provides the solution much faster than the exact method, which can be considered very useful mainly in large-scaled weighted *p*-median instances.

$\|I\|$	Exact solution	Exponential relevance (12)	Ranking relevance (13)	Ranking-exponential relevance (14)	Shifted exponential relevance (15)
300	4,79	4,20	12,57	4,79	5,04
500	23,10	4,38	16,75	4,96	5,72
700	75,20	5,67	19,66	7,08	5,64
900	169,25	5,90	30,79	7,18	6,78
1100	326,80	7,36	37,72	8,70	6,68
1300	519,53	8,79	41,95	10,08	7,05
1500	855,29	6,83	49,13	8,84	7,55

Table 1 Time comparison of different approaches to the disutility relevance estimation. Computational time is given in seconds

Quality of the covering solution is evaluated by *gap*, which is computed according to the formula (19). Average results of numerical experiments are reported in Table 2, which has the same structure as Table 1. As concerns the solution accuracy, the experiments have indicated, that the most appropriate way of the disutility relevance estimation is the shifted exponential approach described by (15).

$\|I\|$	Exponential relevance (12)	Ranking relevance (13)	Ranking-exponential relevance (14)	Shifted exponential relevance (15)
300	7,55	6,53	10,89	1,78
500	6,29	9,81	6,62	0,66
700	6,28	11,36	6,28	0,32
900	5,43	15,52	5,57	0,47
1100	4,49	17,47	4,49	0,14
1300	3,84	19,24	3,84	0,17
1500	1,73	21,95	1,73	0,11

Table 2 Average gaps between the approximate covering solutions and the exact ones expressed in percent of the exact solution

5 Conclusions

The main goal of this paper was to present and compare different approaches to the disutility relevance estimation, which constitutes the core problem of the approximate approach to large public service system design problems. Comparing the reported results in Table 1 and Table 2, we can conclude that the studied approaches can be successfully applied for the public service system design even if they were originally developed to solve the classical p-median problem. It follows from the presented results that the computational time of the approximate approaches is hundred times smaller in comparison with the exact method used for bigger problem instances. The accuracy of the approximate approaches is very good, especially when the shifted exponential estimation of the disutility relevance is employed. Thus we can conclude that we present useful tool for the large public service system design, which can be implemented using common commercial optimization software.

Acknowledgement

This work was supported by the research grants VEGA 1/0296/12 “Public Service Systems with Fair Access to Service” and APVV-0760-11 “Designing of Fair Service Systems on Transportation Networks”. We would also like to thank to „*Centre of excellence for systems and services of intelligent transport*“ **(**ITMS 26220120050) for built up the infrastructure, which was used.

References

[1] Avella, P., Sassano, A., Vassil'ev, I..: *Computational study of large scale p-median problems*. In: Mathematical Programming 109, pp. 89-114, (2007).

[2] Current, J., Daskin, M., Schilling, D.: *Discrete network location models*. In: Drezner Z (ed) et al. Facility location. Applications and theory. Berlin: Springer, pp. 81-118, (2002).
[3] García, S., Labbé, M., Marín, A.: *Solving large p-median problems with a radius formulation*. In: INFORMS Journal on Computing 23 (4), pp. 546-556, (2011).
[4] Ingolfsson, A., Budge, S., Erkut, E.: *Optimal ambulance location with random delays and travel times*, In: Health Care management science, 2008, 11 (3), pp. 262-274, (2008).
[5] Janáček, J.: *Approximate Covering Models of Location Problems*. In: Lecture Notes in Management Science: Proceedings of the 1st International Conference ICAOR ´08, Vol. 1, Sept. 2008, Yerevan, Armenia, pp. 53-61, ISSN 2008-0050, (2008).
[6] Janáček, J., Kvet, M.: *Approximate solving of large p-median problems*. In: ORP3-Operational research peripatetic post-graduate programme: proceedings, Cádiz, Spain, September 13-17, 2011, Cádiz: Servicio de Publicaciones de la Universidad de Cádiz, 2011, ISBN 978-84-9828-348-8, pp. 221-225, (2011).
[7] Janáček, J., Kvet, M.: *Relevant Network Distances for Approximate Approach to Large p-Median Problems*. In: Operations research proceedings 2012: selected papers of the international conference on operations research: September 4-7, 2012, Hannover, to appear
[8] Janáček, J., Linda, B., Ritschelová, I.: *Optimization of Municipalities with Extended Competence Selection*. In: Prager Economic Papers-Quarterly Journal of Economic Theory and Policy, Vol. 19, No 1, 2010, pp. 21-34, (2010).
[9] Janáčková, M., Szendreyová, A.: *An impact of set of centers of distribution system design problem on computational time of solving algorithm*. In: Mikulski J (ed). Advances in Transport Systems Telematics, Katowice, pp. 109-112, (2006).
[10] Jánošíková, L.: *Emergency Medical Service Planning*. In: Communications –Scientific Letters of the University of Žilina, Vol. 9, No 2, pp. 64-68, (2007).
[11] Marianov, V., Serra, D.: *Location problems in the public sector*. In: Drezner Z (ed.) et al. Facility location. Applications and theory. Berlin: Springer, pp. 119-150, (2002).

Allocation of Trains to Platforms at Railway Station

Ľudmila Jánošíková[1], Michal Krempl[2]

Abstract. This paper deals with the organization of train traffic in a passenger railway stations. One of the important problems which a dispatcher has to solve every day is track allocation for every approaching train. There is a tool helping him in his job called the Track Occupancy Plan, which is attached to the train traffic diagram. The plans specify the scheduled times (of arrivals and departures), as well as the assignment of platform tracks to arriving or departing trains.
In this paper the problem of allocation of trains to platforms is formulated as a bi-criterion mixed integer programming problem. The first objective is to minimise the deviations of the arrival and departure times proposed by the model from the times specified by the timetable. The second criterion maximises the desirability of the platform tracks to be assigned to the trains. For finding a solution there is used a lexicographic approach and the local branching algorithm.

Keywords*:* routing trains, scheduling, mixed integer programming, multiple-objective programming

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The problem of routing and scheduling trains at a station is a subproblem of the generation of a timetable for a railway company. The generation of a timetable is a hierarchical process. At the first stage, a preliminary timetable for the whole network is proposed. In this phase, a macroscopic viewpoint at the railway network is applied. Stations are considered as black boxes. Capacity limits of particular stations and the movement of trains inside the stations are not taken into account. Then, at the second stage, a microscopic viewpoint related to stations is applied. At every station, the network timetable is checked whether it is feasible with respect to capacity, safety and train operators' preferences. To prove the feasibility, detailed routes and schedules for the trains are generated. If desired arrival and departure times are not feasible at the microscopic level, the process returns to the first stage, where the timetable must be adjusted.

In the Czech and Slovak Republic, planning train movements through the station is done by hand, using planner's experience and a set of rules determined by a railway company. The main goal of this research is to design a more sophisticated approach which would serve as a planner's decision supporting tool and result in better routing and scheduling plans. Such an approach can play an important role especially at large, busy stations with multiple platforms and multiple in-lines and out-lines. Improvement in the plan quality results in

1. better management of train operation in the station, namely:
   a) shorter times of routes occupation by arriving and departing trains,
   b) uniform workload of the infrastructure elements, such as tracks, switches, and platforms,
   which leads to a more robust plan resistant to random disturbances;
2. higher service quality perceived by passengers, namely:
   a) shorter distances needed for changing trains,
   b) more appropriate platforms (platforms near to ticket sales points and to the station entrance, platforms equipped by station shops or catering etc.),
   c) less probability of changing the planned platform when the train delays,
   which leads to a higher share of railway in public transport;
3. meeting train operators' requirements on arrival and departure times and platforms assigned to trains.

Routing and scheduling trains at a station has been studied by researchers in countries, where large, busy stations with capacity constraints can be found. Billionet addressed only the routing problem [2]. The problem was modelled using a graph theory and the integer programming formulation of the resulting graph colouring problem was solved. However, the *k* colouring problem is not indeed an optimisation problem, it means any

[1] University of Žilina, Faculty of Management Science and Informatics, Department of Transportation Networks
Univerzitná 1, 010 26 Žilina, Slovak Republic, e-mail: Ludmila.Janosikova@fri.uniza.sk
[2] VŠB - Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport
17. listopadu 15, 708 33 Ostrava-Poruba, Czech Republic, e-mail: michal.krempl.st@vsb.cz

feasible solution is acceptable and the problem formulation does not reflect the solution quality, such as route lengths or platform preferences for individual trains. In [7, 8] the problem of train routing was described as a weighted node packing problem, using bivalent programming, while the solution algorithm applied the branch-and-cut method. A disadvantage of the above presented models is that the calculations connected with them are computationally too complex and time consuming. Another, practically oriented approach has given up on applying the integer programming methods, and replaced them by the heuristics, solving the scheduling and routing problems at a time [3]. The algorithm incorporates, or considers, the operational rules, costs, preferences and trade-offs, which are applied by experts creating plans manually. The shortcoming of this approach is obvious: since it is a heuristics, the optimality of the resulting plan is not guaranteed.

Other way of research, e.g. [1, 4], has been directed at operational train management. In real time it is necessary to reflect the requirements of the operation burdened with irregularities, i.e. to re-schedule the arrivals and departures times, and/or re-route trains.

In this paper we propose a mixed integer programming (MIP), multiple criteria model of the routing and scheduling problem. The problem can be solved by a lexicographic approach, where particular criteria are ranked according to their importance.

2 Problem formulation

The problem of routing and scheduling trains, as described in the previous section, consists of the following partial issues subject to decision-making process. For each train,

- a platform track must be specified at which the train should arrive; the platform track assignment determines the route, on which the train approaches from an in-line (or from a depot) to the platform, or departs from the platform to an out-line (or to a depot),
- arrival time at the platform and departure time from the platform need to be determined.

The solution should minimise deviations from the planned arrival and departure times and maximise the total preferences for platforms and routes.

The inputs to the mathematical programming model are as follow:

1. track layout of the station, which is necessary for determining feasible platform tracks for a train and conflicting routes,
2. list of trains, where the data required for each train include:
   a) planned time of its arrival at the platform,
   b) planned time of its departure from the platform,
   c) line on which the train arrives (in-line) and departs (out-line),
   d) list of feasible platform tracks with their desirability for the train,
   e) category of the train.

All time data are given in minutes.

Further on we present the formulation of the MIP model. First we need to explain the symbols used:

Subscripts which in the mathematical model represent objects

i, i', j train

k, k' platform track

Input parameters (constants)

t_i^{Pa} planned arrival time of train i at the platform

t_i^{Pd} planned departure time of train i

t^{Cn} standard amount of time passengers take to change trains (depends on particular railway station)

I_i arrival line track (in-line) for train i

O_i departure line track (out-line) for train i

c_i category of train i; $c_i = 1$ for regional stopping trains and increases with the speed and distance travelled by the train

t^{min} minimum dwell time of a train at the platform

t^{max} maximum time interval, in which two train movements are tested for a conflict

p_{ik} preference coefficient; it reflects the desirability of the assignment of platform track k to train i

s_{ik} number of switches on the route of train i from the arrival line track to platform track k and from platform track k to the departure line track

$s_i^{\min}$ number of switches on the shortest train route in the station

$s_i^{\max}$ number of switches on the longest train route in the station

$a(l,k,l',k')$ coefficient, which has value *true*, if the route connecting line l to platform track k conflicts with the route connecting line l' to platform track k'; if there exists any route connecting line l to track k and any route connecting line l' to track k' such that these two routes do not conflict, then $a(l,k,l',k') = false$. $a(l,k,l',k') = true$ for $k = k'$ or $l = l'$. The existence of route conflicts can be identified in advance from a detailed map of the track layout.

We adopted the concept of conflicting routes and conflict solving from the source [3]. If two trains are on conflicting routes we must ensure that there is at least a required minimum headway (time interval) between them, for safety and signalling reasons. The minimum headway depends on the order, types, and lengths of the trains, on whether the trains are arriving or departing from the station, and on the platform track and line used by each train. For example, let $h(i,k,i',k')^{da}$ is the minimum headway required between train i departing from track k and the next train i' arriving at track k'. The superscripts d and a denote departure and arrival, and the order of the superscripts indicates the order of the trains, i.e., train i is followed by i'. Similarly we have $h(i,k,i',k')^{aa}$, $h(i,k,i',k')^{ad}$ and $h(i,k,i',k')^{dd}$ for combinations arrival – arrival, arrival – departure, departure – departure. We need not introduce subscripts to denote the in-lines or out-lines used by trains since for an arriving train i the in-line is already specified by I_i, and for a departing train i the out-line is specified by O_i.

The preference coefficient p_{ik} may reflect:

- operator's preferences of platforms,
- the distance of the track k to the connecting trains,
- the length of the route used by train i arriving to or departing from platform track k. The smoother and shorter the route is, the less the possibility of a conflict with other trains is, hence the probability of delay propagation decreases.

In our model, coefficient p_{ik} is set according to the following formula:

$$p_{ik} = \begin{cases} 1 & \text{if track } k \text{ is the planned (or desired) track for train } i \\ 0.9 & \text{if track } k \text{ is located at the same platform as the planned track} \\ 0.8\left(\dfrac{s_i^{\max} - s_{ik}}{s_i^{\max} - s_i^{\min}}\right) & \text{otherwise} \end{cases}$$

Sets of objects

K set of all platform tracks

$K(i)$ set of feasible platform tracks for train i

U set of all arriving, departing, and transit trains

$W(j)$ set of all connecting trains, which has to wait for train j

$V^{aa} = \{(i,j) : i,j \in U, i < j, |t_i^{Pa} - t_j^{Pa}| \le t^{\max}\}$ set of ordered pairs of those trains that may arrive concurrently

$V^{ad} = \{(i,j) : i,j \in U, i < j, |t_i^{Pa} - t_j^{Pd}| \le t^{\max}\}$ set of ordered pairs of those trains that arriving train i and departing train j may travel concurrently

$V^{da} = \{(i,j) : i,j \in U, i < j, |t_i^{Pd} - t_j^{Pa}| \le t^{\max}\}$ set of ordered pairs of those trains that departing train i and arriving train j may travel concurrently

$V^{dd} = \{(i,j) : i,j \in U, i < j, |t_i^{Pd} - t_j^{Pd}| \le t^{\max}\}$ set of ordered pairs of those trains that may depart concurrently

Decision and auxiliary variables of the model

$$\text{for } i \in U, k \in K(i): x_{ik} = \begin{cases} 1 & \text{if track } k \text{ is assigned to train } i \\ 0 & \text{otherwise} \end{cases}$$

u_i difference between the planned and real arrival time of train i at a platform, $i \in U$

v_i difference between the planned and real departure time of train i from a platform, $i \in U$

The following auxiliary variables $\boldsymbol{y}$ are introduced for the couple of those trains i and j that may travel concurrently. They enable to express safety headways between conflicting trains.

$$\text{for } i,j \in U, i<j: y_{ij}^{aa} = \begin{cases} 1 & \text{if train } i \text{ arrives before train } j \text{ arrives} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{for } (i,j) \in V^{ad}: y_{ij}^{ad} = \begin{cases} 1 & \text{if train } i \text{ arrives before train } j \text{ departs} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{for } (i,j) \in V^{da}: y_{ij}^{da} = \begin{cases} 1 & \text{if train } i \text{ departs before train } j \text{ arrives} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{for } (i,j) \in V^{dd}: y_{ij}^{dd} = \begin{cases} 1 & \text{if train } i \text{ departs before train } j \text{ departs} \\ 0 & \text{otherwise} \end{cases}$$

Model formulation

minimise
$$c_i \sum_{i \in U} (u_i + v_i) \tag{1}$$

maximise
$$\sum_{i \in U} \sum_{k \in K(i)} p_{ik} x_{ik} \tag{2}$$

subject to

$$v_i + t_i^{Pd} \geq u_i + t_i^{Pa} + t^{\min} \qquad \forall i \in U \tag{3}$$

$$v_i + t_i^{Pd} \geq u_j + t_j^{Pa} + t^{Cn} \qquad \forall j \in U; i \in W(j) \tag{4}$$

$$u_{i'} + t_{i'}^{Pa} \geq u_i + t_i^{Pa} + h(i,k,i',k')^{aa} - M\left(1 - y_{ii'}^{aa}\right) - M(1 - x_{ik}) - M(1 - x_{i'k'})$$
$$\forall (i,i') \in V^{aa}, k \in K(i), k' \in K(i'): a(I_i, k, I_{i'}, k') \tag{5}$$

$$u_i + t_i^{Pa} \geq u_{i'} + t_{i'}^{Pa} + h(i',k',i,k)^{aa} - My_{ii'}^{aa} - M(1 - x_{ik}) - M(1 - x_{i'k'})$$
$$\forall (i,i') \in V^{aa}, k \in K(i), k' \in K(i'): a(I_i, k, I_{i'}, k') \tag{6}$$

$$v_{i'} + t_{i'}^{Pd} \geq u_i + t_i^{Pa} + h(i,k,i',k')^{ad} - M\left(1 - y_{ii'}^{ad}\right) - M(1 - x_{ik}) - M(1 - x_{i'k'})$$
$$\forall (i,i') \in V^{ad}, k \in K(i), k' \in K(i'): a(I_i, k, O_{i'}, k') \tag{7}$$

$$u_i + t_i^{Pa} \geq v_{i'} + t_{i'}^{Pd} + h(i',k',i,k)^{da} - My_{ii'}^{ad} - M(1 - x_{ik}) - M(1 - x_{i'k'})$$
$$\forall (i,i') \in V^{ad}, k \in K(i), k' \in K(i'): a(I_i, k, O_{i'}, k') \tag{8}$$

$$u_{i'} + t_{i'}^{Pa} \geq v_i + t_i^{Pd} + h(i,k,i',k')^{da} - M\left(1 - y_{ii'}^{da}\right) - M(1 - x_{ik}) - M(1 - x_{i'k'})$$
$$\forall (i,i') \in V^{da}, k \in K(i), k' \in K(i'): a(O_i, k, I_{i'}, k') \tag{9}$$

$$v_i + t_i^{Pd} \geq u_{i'} + t_{i'}^{Pa} + h(i',k',i,k)^{ad} - My_{ii'}^{da} - M(1 - x_{ik}) - M(1 - x_{i'k'})$$
$$\forall (i,i') \in V^{da}, k \in K(i), k' \in K(i'): a(O_i, k, I_{i'}, k') \tag{10}$$

$$v_{i'} + t_{i'}^{Pd} \geq v_i + t_i^{Pd} + h(i,k,i',k')^{dd} - M\left(1 - y_{ii'}^{dd}\right) - M(1 - x_{ik}) - M(1 - x_{i'k'})$$
$$\forall (i,i') \in V^{dd}, k \in K(i), k' \in K(i'): a(O_i, k, O_{i'}, k') \tag{11}$$

$$v_i + t_i^{Pd} \geq v_{i'} + t_{i'}^{Pd} + h(i',k',i,k)^{dd} - My_{ii'}^{dd} - M(1 - x_{ik}) - M(1 - x_{i'k'})$$
$$\forall (i,i') \in V^{dd}, k \in K(i), k' \in K(i'): a(O_i, k, O_{i'}, k') \tag{12}$$

$$u_i + t_i^{Pa} \geq v_{i'} + t_{i'}^{Pd} + h(i',k,i,k)^{da} - My_{ii'}^{aa} - M(1 - x_{ik}) - M(1 - x_{i'k})$$
$$\forall i, j \in U, i < j, k \in K(i) \cap K(i') \tag{13}$$

$$u_{i'} + t_{i'}^{Pa} \geq v_i + t_i^{Pd} + h(i,k,i',k)^{da} - M\left(1 - y_{ii'}^{aa}\right) - M(1 - x_{ik}) - M(1 - x_{i'k})$$
$$\forall i, j \in U, i < j, k \in K(i) \cap K(i') \tag{14}$$

$$y_{ij}^{aa} = 1 \quad \forall i, j \in U, i \neq j, I_i = I_j, t_i^{Pa} \leq t_j^{Pa} \tag{15}$$

$$\sum_{k \in K(i)} x_{ik} = 1 \qquad \forall i \in U \tag{16}$$

$$u_i, v_i \geq 0 \qquad \forall i \in U \tag{17}$$

$$x_{ik} \in \{0,1\} \qquad \forall i \in U \ \forall k \in K(i) \tag{18}$$

$$y_{ij}^{aa} \in \{0,1\} \qquad \forall i, j \in U, i < j \tag{19}$$

$$y_{ij}^{ad} \in \{0,1\} \qquad \forall (i, j) \in V^{ad} \tag{20}$$

$$y_{ij}^{da} \in \{0,1\} \qquad \forall (i, j) \in V^{da} \tag{21}$$

$$y_{ij}^{dd} \in \{0,1\} \qquad \forall (i, j) \in V^{dd} \tag{22}$$

Model description

Objective function (1) minimises the weighted deviations of the arrival and departure times proposed by the model from the times specified by the timetable. The weights cause that long-distance/high- speed trains will respect planned times and regional trains will be postponed if necessary. The second criterion maximises the desirability of the platform tracks to be assigned to the trains.

Constraint (3) ensures that a minimum dwell time needed for boarding and alighting must be kept.

Constraint (4) states that connecting train i with real departure $v_i + t_i^{Pd}$ has to wait in station to time at least $u_j + t_j^{Pa} + t^{Cn}$.

Constraints (5) – (12) ensure that a minimum headway will be kept between conflicting trains. More precisely, constraint (5) states that if trains i and i' have planned arrival times within t^{max} and train i arrives at platform track k before train i' arrives at track k', i.e.

$$x_{ik} = 1, x_{i'k'} = 1, y_{ii'}^{aa} = 1, \tag{23}$$

and trains are on conflicting routes (i.e. $a(I_i, k, I_{i'}, k')$ is *true*), then train i' is allowed to arrive at least $h(i,k,i',k')^{aa}$ minutes later than train i. If at least one of the conditions (23) is not met (e.g. train i is not assigned to track k), then constraint (5) becomes irrelevant as the right-hand side is negative (M is a suitably picked high positive number). If train i' is followed by train i ($y_{ii'}^{aa} = 0$), then i is allowed to arrive at least $h(i',k',i,k)^{aa}$ minutes later than i', which is ensured by constraint (6). Constraints (7) – (12) have a similar meaning for the other combinations of arrival – departure.

Constraints (13) – (14) ensure that a train will not be dispatched to an occupied track. If train i' is followed by train i ($y_{ii'}^{aa} = 0$) and both trains arrive at the same track k, then i is allowed to arrive at least $h(i',k,i,k)^{da}$ minutes after train i' leaves track k, which is expressed by constraint (13). Constraint (14) holds for the reverse order of trains i, i'.

Constraint (15) states that y_{ij}^{aa} is 1 if train i is followed by train j at the arrival and both trains travel on the same in-line.

Constraint (16) ensures that each train is always dispatched to exactly one platform track.

The remaining obligatory constraints (17) – (22) specify the definition domains of the variables.

This multiple-criteria optimisation problem was solved using the lexicographic approach, where the objective functions are ranked according to their importance. In the problem at hand, the first objective function (i.e. to meet the timetable) is more important that the second one (i.e. to respect track preferences). This ordering reflects how decisions are currently made in practice. The solution technique consists of two steps. In the first step the problem (1), (3) – (22) is solved giving the best value of the weighted sum of deviations f_1^{best}. Then the constraint

$$c_i \sum_{i \in U} (u_i + v_i) \leq f_1^{best} \tag{24}$$

is added and the model (2) – (22), (24) is solved. Because both MIP problems are hard and the optimal solutions cannot be found within a reasonable time limit, we decided to implement Local Branching heuristic [6] using the general optimisation software *Xpress* [5].

3 Case study

The model was verified by using the real data of Prague main station and the timetable valid for the years 2004/2005. Prague main station is a large station that at the given time had 7 platforms, 17 platform tracks and 5 arrival/departure line tracks. According to the timetable 2004/2005, the station dealt with 288 regular passenger trains per a weekday.

The results of computational experiments show that the timetable 2004/2005 was not correct with regard to safety requirements. There were some trains travelling on conflicting routes concurrently. That is why their desired arriving or departing times could not be kept. The best solution proposed by the model delays 3 trains at arrival by 5 minutes and 12 trains at departure by 17 minutes in total, and dispatches 32 (11 %) trains to platform tracks different from the planned ones.

A more detailed discussion of computational experiments will be presented at the conference.

4 Conclusions

In the paper, a mixed integer programming model for routing and scheduling trains at a passenger railway station is described. The model gives a solution with regard to particular criteria ranked according to their importance, i.e. the solution with minimal deviation of the arrival and departure times from the timetable that respects the desirability of the platform tracks to be assigned to the trains as much as possible.

Acknowledgements

This research was supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences under project VEGA 1/0296/12 “Public service systems with fair access to service” and by the Slovak Research and Development Agency under project APVV-0760-11 “Designing of Fair Service Systems on Transportation Networks”.

References

[1] Bažant, M., and Kavička, A.: Artificial neural network as a support of platform track assignment within simulation models reflecting passenger railway stations. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* **223** (5), 2009, 505–515.

[2] Billionet, A.: Using integer programming to solve the train-platforming problem. *Transportation Science* **37** (2), 2003, 213–222.

[3] Carey, M., and Carville, S.: Scheduling and platforming trains at busy complex stations. *Transportation Research Part A* **37** (2003), 195–224.

[4] Chakroborty, P., and Vikram, D.: Optimum assignment of trains to platforms under partial schedule compliance. *Transportation Research Part B: Methodological* **42** (2), 2008, 169–184.

[5] *FICO™ Xpress Optimization Suite* [online]. Available from: http://www.fico.com [Accessed 10 October 2011].

[6] Fischetti, M., Lodi, A., Salvagnin, D.: Just MIP it!. In: Maniezzo, V., Stützle, T., Voß, S., eds. *Matheuristics*. Springer, New York, 2009, 39–70.

[7] Zwaneveld, P.J.: *Railway Planning - routing of trains and allocation of passenger lines*. Ph.D. thesis, Erasmus University Rotterdam, Rotterdam, the Netherlands, 1997.

[8] Zwaneveld, P.J., Kroon, L.G., van Hoesel, S.P.M.: Routing trains through a railway station based on a node packing model. *European Journal of Operational Research* **128** (2001), 14–33.

On the constraints in optimal managing of natural resources systems

Jitka Janová [1], Jana Musilová[2]

Abstract. Optimal managing of natural resources systems has been permanently in the focus of interest of worldwide policymakers. Today, more than ever, identifying optimal policies is of strategic importance to address the problems crucial to the economy and the environment. Since the beginning of the mathematical programming science the up-to-date mathematical methods were applied to create decision support for natural resources management. However, where the optimal policies have been determined, these often fail. It was shown that this is due to the shortcomings of the models' structure stemming mainly from neglecting of nonlinearities of real systems and miss identification of constraints among relevant variables.

This contribution serves as an introductory study of constrained systems arising in the natural resources modeling. We focus on two non-equivalent treatments of such constrained dynamical systems (so called vakonomic and nonholonomic approaches), we discuss the possibility to improve the optimal control models for natural resources systems using nonlinear techniques and, as a test case, we formulate a generalized anti-pollution policy problem.

Keywords: natural resources, bioeconomics, constraints, nonholonomic, vakonomic, dynamic optimization

JEL classification: C44
AMS classification: 49N99

1 Introduction

Since the beginning of mathematical programming science the up-to-date mathematical methods were applied to create quality decision support systems for the natural resources management. The most spread approach commonly applied mainly to agriculture and forestry production planning, transportation or scheduling problems is based on linear programming methods. Simultaneously with the development of mathematical methods the models started to use the advanced methods of operations research and mathematical programming. Particularly, this has been the case especially in environmental studies, where such problems as landscape optimization, ecosystem modeling, water resource allocation, forest structure optimization or herd management, etc. were solved using advanced deterministic methods of nonlinear programming. Currently, there is running quite a broad research on stochastic programming models for production and harvesting planning. Let us mention that the methods of stochastic programming (unlike the deterministic methods) are still extensively developed from the mathematical point of view to meet the demand of real optimization problems.

All the above mentioned examples are able to determine static optimal solution, but the natural resources systems are dynamical. It is possible to consider one-year-ahead solution for particular or local needs, but from a long run sustainable perspective these problems must be treated by the dynamical methods. Note, that compared to the number of models developed for static optimization of natural resources systems, the dynamic optimization methods are infrequently used.

Generally, in identifying the optimum long run decision strategies, discrete and continuous approaches are distinguished, the former being connected with sequential decision processes solved via dynamic

[1]Mendel University in Brno/Faculty of Business and Economics, Department of Statistics and Operation Analysis, Zemědělská 1/1665 613 00 Brno, janova@mendelu.cz

[2]Masaryk University/Faculty of Science, Department of Theoretical Physics and Astrophysics, Kotlářská 267/2, 611 37 Brno , janam@physics.muni.cz

programming, the latter with calculus of variations and optimal control theory.

The calculus of variations is inherently connected with the classical mechanics and its fundamental equations of motion - Lagrange equations. Based on the Hamilton approach, an alternative description of a mechanical system was formulated. The Hamilton apparatus was used in formulation of the control theory, where the dynamic optimization problem is viewed as consisting of three (rather than two) types of variables. Aside from the time variable and the state variable, consideration is given to a control variable. The control theory is extensively used and studied in engineering applications (robotics, automation) and also in economic applications covering finance, production and inventory problems, marketing, machine maintenance and replacement. Currently, there exists a considerable amount of literature studying optimal control problems in environmental studies, ecology and natural resources (see [2]). Let us mention only several recent interesting problems studied: e.g. optimal control of invasive species in [5], optimal extraction of natural resources under various assumptions concerning pollution, waste or prices and costs of entering variables in [18], optimal PEST control in agriculture in [4], greenhouse control systems [18] and optimal control of trade offs between water use and fish harvesting [19]. The intense of current research stem from the strong belief that optimal control theory is a powerful tool to identify the long run strategies and to evaluate sustainability implications of alternative policy making practices in ecology and natural resources economics. While it should be useful for managing these systems, the results are still unsatisfactory. Where the optimal control policies have been determined, these often fail. It was shown, that the paradoxical aspects of control theory in natural resources economics result from the shortcomings of the models constructed, stemming mainly from neglecting non-linearities and miss-identification of constraints among the variables (see [15]).

This paper is an introduction to the research of a constraint structure and a solution of the natural resources continuous dynamic systems with constraints (for some notes on the constraint structures in agricultural and forestry see [9] and [12]). First, we introduce the dynamic optimization techniques appropriate for a solution of the natural resources problems. Then, the concept of more general (nonholonomic) constraints in dynamic optimization will be briefly described and the method of incorporating the constraints into the dynamic model will be presented together with examples of constraints in natural resource dynamic problems. As an addition to the explanation of the mathematical background of the constrained system solution, we open the challenging problem of alternative approaches to the constrained dynamic optimization (so called nonholonomic and vakonomic formulation).

2 Methods of continuous dynamic optimization

2.1 The calculus of variations

The fundamental problem of calculus of variations is (see e.g. [3])

$$\max \ (\min) \quad V = \int_{t_0}^{t_1} G(t, q(t), \dot{q}(t))\mathrm{d}t, \tag{1}$$

$$q^\sigma(t_0) = q_0^\sigma, \tag{2}$$

$$q^\sigma(t_1) = q_1^\sigma, 1 \leq \sigma \leq m, \tag{3}$$

where $q(t) = (q^1(t), q^2(t), \ldots, q^m(t))$ and $\dot{q}(t) = (\dot{q}^1(t), \dot{q}^2(t), \ldots, \dot{q}^m(t))$ are state variables and its derivative respectively, and t is time variable [1]. The goal of the fundamental problem (1-3) is to find an optimal trajectory $q^\star(t)$, that yields an extremal (in general stationary) value of the functional $\int_{t_0}^{t_1} G(t, q(t), \dot{q}(t))\mathrm{d}t$. The basic necessary condition in the calculus of variations are the Euler equations:

$$\frac{\partial G(t, q(t), \dot{q}(t))}{\partial q^\sigma} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial G(t, q(t), \dot{q}(t))}{\partial \dot{q}^\sigma} = 0, \ 1 \leq \sigma \leq m. \tag{4}$$

Note, that in particular problems where function G is strictly concave or convex, which is often the case both in economic and physical problems, are the Euler equations the necessary and sufficient conditions for finding the optimal trajectory.

[1] We consider the common assumptions of the calculus of variations: we restrict the set of admissible paths to those continuous curves with continuous derivatives and we assume that the integrand function G is twice differentiable.

Nonholonomic constraints in calculus of variations

Consider a dynamical system described by (1-3) with $G = G(t, q, \dot{q})$ subject to general constraints $f^i(t, q, \dot{q}) = 0, 1 \leq i \leq k$. In physics– the field which the calculus of variations was invented for– we can find two different approaches for incorporating the constraints into the original system. In the first one *the vakonomic approach* we search for the solution of (unconstrained) variational problem associated to the function

$$\bar{G} = G + \sum_{i=1}^{k} \lambda_i f^i,$$

hence the variational problem takes for the constrained system the form

$$\max \ (\min) \quad V = \int_{t_0}^{t_1} \left[G(t, q, \dot{q}) + \sum_{i=1}^{k} \lambda_i f^i(t, q, \dot{q}) \right] \mathrm{d}t, \tag{5}$$

$$q^\sigma(t_0) = q_0^\sigma, \tag{6}$$

$$q^\sigma(t_1) = q_1^\sigma, 1 \leq \sigma \leq m, \tag{7}$$

where multipliers λ_i are considered as additional variables. Namely, the vakonomic solutions can be obtained by the Euler-Lagrange equations

$$\frac{\partial \bar{G}}{\partial q^\sigma} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \bar{G}}{\partial \dot{q}^\sigma} = 0, \tag{8}$$

$$\frac{\partial \bar{G}}{\partial \lambda_i} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \bar{G}}{\partial \dot{\lambda}_i} = 0, \tag{9}$$

which gives

$$\frac{\partial G}{\partial q^\sigma} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial G}{\partial \dot{q}^\sigma} + \sum_{i=1}^{k} \frac{\partial f^i}{\partial \dot{q}^\sigma} \cdot \dot{\lambda}_i - \sum_{i=1}^{k} \left(\frac{\partial f^i}{\partial q^\sigma} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial f^i}{\partial \dot{q}^\sigma} \right) \cdot \lambda_i = 0, \tag{10}$$

$$f^i(t, q, \dot{q}) = 0. \tag{11}$$

The second approach – *nonholonomic* –consists in incorporating the constraint forces into the Euler-Lagrange equations:

$$\frac{\partial G}{\partial q_\sigma} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial G}{\partial \dot{q}_\sigma} = \sum_{i=1}^{k} \frac{\partial f^i}{\partial \dot{q}_\sigma} \cdot \mu_i, \tag{12}$$

$$f^i(t, q, \dot{q}) = 0, \tag{13}$$

where μ are the Lagrange multipliers. The nonholonomic approach is typically used for solving the mechanical systems with nonholonomic constraints in physics (for instructive nonholonomic mechanical problems see e.g. [10], [11]).

Comparing the nonholonomic (12) and vakonomic equations (10) we arrive to the conclusion that the arising systems of differential equations are not equivalent unless

$$\sum_{i=1}^{k} \lambda_i \left(\frac{\partial f^i}{\partial q_\sigma} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial f^i}{\partial \dot{q}_\sigma} \right) = 0. \tag{14}$$

Verifying (14) is complicated task because the multipliers λ_i are a priori unknown. Nevertheless, for special types of constrained systems the equivalence conditions (14) can be simply verified without explicitly solving the constrained dynamics.

It is obvious that for the Lagrangian mechanical system subject to holonomic constraints $u^i(t, q) = 0$, $1 \leq i \leq k$, coincide the nonholonomic (12) and vakonomic equations (10). Now we shall consider a very frequent type of constraint, called *linear integrable*, or *semiholonomic* (see [14]). This constraint has the form $f^i = \dot{q}^{m-k+i} - g^i(t, q^\sigma, \dot{q}^l)$, $1 \leq i \leq k$, $1 \leq k \leq m-1$, $1 \leq l \leq m-k$, where functions g^i obey the

following relations:

$$\varepsilon_l'(g^i) = 0, \quad \varepsilon_l'(g^i) = \left(\frac{\partial g^i}{\partial q^l} + \frac{\partial g^j}{\partial \dot{q}^l}\frac{\partial g^i}{\partial q^{m-k+j}}\right) - \left(\frac{\partial^2 g^i}{\partial t \partial \dot{q}^l} + \dot{q}^s \frac{\partial^2 g^i}{\partial q^s \partial \dot{q}^l} + g^j \frac{\partial^2 g^i}{\partial q^{m-k+j}\partial \dot{q}^l}\right), \quad (15)$$

$$\frac{\partial^2 g^i}{\partial \dot{q}^s \partial \dot{q}^l} = 0. \quad (16)$$

Following proposition concern a special case of semiholonomic constraint, for which functions g^i are independent of variables q^{m-k+j}.

Proposition 1. *For a mechanical system subjected to semiholonomic constraint $f^i = \dot{q}^{m-k+i} - g^i(t, q^l, \dot{q}^l) = 0$, $1 \leq l \leq m-k$, the equivalence condition (14) is satisfied.*

Proof. Let us analyze the expression

$$\frac{\partial f^i}{\partial q^\sigma} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f^i}{\partial \dot{q}^\sigma}$$

included in equivalence conditions (14).

According to (15) and (16) we get for $1 \leq l \leq m-k$

$$\frac{\partial f^i}{\partial q^l} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f^i}{\partial \dot{q}^l} = 0 \quad (17)$$

and

$$\frac{\partial f^i}{\partial q^{m-k+j}} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f^i}{\partial \dot{q}^{m-k+j}} = 0 - \frac{\mathrm{d}}{\mathrm{d}t}\delta^i_j = 0. \quad (18)$$

Hence the equivalence conditions (14) hold independently of values of multipliers λ_i. □

Note that for general semiholonomic constraint $f^i = \dot{q}^{m-k+i} - g^i(t, q^\sigma, \dot{q}^l)$ the conditions (18) do not hold because $\frac{\partial f^i}{\partial q^{m-k+j}} \neq 0$. Nevertheless, the conditions (14) can be satisfied via the appropriate form of Lagrange multipliers. Note that even in technical applications (that are the most elaborated in the calculus of variations) one can find contradictions in the results for semiholonomic constraints. In [8] the coupled rolling motion is considered that is described by Chaplygin system. Since the constraint in the problem obeys the conditions (15-16) it is considered to be semiholonomic. Nevertheless, according to our calculations the vakonomic and nonholonomic systems seems not to be equivalent. This is an issue to be settled by our further research.

2.2 Optimal control theory

The optimal control formulation of a dynamic optimization problem focuses upon control variables, $u(t) = (u^1(t), u^2(t), \ldots, u^p(t))$ that serve as the instrument of optimization. Let us denote the state variables $y(t) = (y^1(t), y^2(t), \ldots, y^r(t))$. Optimal control theory determines the optimal time path for a control variable $u^*(t)$ and we can also find the optimal state path $y^*(t)$. The relation $\dot{y} = f(t, y, u)$ between state variables and control variables must be identified to formulate the optimal control problem. This relation reflects the principle of studied optimization problem. The control variable u is a subject to our discretionary choice through which we influence the state variable y. The criterion of the dynamic optimization problem is reflected by the objective functional

$$V = \int_{t_0}^{t_1} G(t, y, u)\mathrm{d}t. \quad (19)$$

The simplest optimal control problem is

$$\max\ V = \int_{t_0}^{t_1} G(t, y, u)\mathrm{d}t, \quad (20)$$

$$\dot{y} = f(t, y, u), \quad (21)$$

$$y(t_0) = y_0,\ y(t_1)\ \text{free}, u(t) \in U, \quad (22)$$

where (21) refers to as the equation of motion or state equation. While the objective functional (19) reflects the optimization criterion well known to the policy maker and therefore not complicated to compile, the constraint (21), playing a key role in representation of the real world system principles, is in natural resources problems of complex unknown structure. Generally, function f is non-linear, but since the solution of nonlinear systems is very difficult mathematically, the linear relationships are mostly used when formulating the state equations in natural resources systems (for simple nonlinear optimal control problem in the fishery bioeconomics with one state and one control variable see [13]). Note, that linear control system is defined by the state equations

$$\dot{y} = A(t)y + B(t)u. \tag{23}$$

We suggest to adopt nonholonomic optimal control techniques for representing accurately the nonlinear structure of natural resources dynamic systems. Nonholonomic optimal control represents a nonliear approach that enables to consider more complex relations among first derivatives of the state variables. A typical form of nonholonomic control system is given by state equations

$$\dot{y} = F(y)u. \tag{24}$$

We formulate a generalized antipollution policy that is able to reflect nonlinear structure of the real system (see [3] for linear control system). We use the symbol E to represent the energy use, P denotes the stock of pollution with $\dot{P}$ as its flow. The use of energy generates a flow of pollution. Let A stand for the level of antipollution activities that can reduce the pollution stock in a known manner. We assume that the implementation of antipollution activities A in itself requires the use of energy. The criterion of the decision making is maximization of utility $U = U(C(E), P)$ with $C(E)$ being the consumption. The dynamic optimization problem takes the form

$$\begin{aligned}
\max\ V &= \int_0^T U(C(E), P)\mathrm{d}t, && (25)\\
\dot{S} &= -E - A, && (26)\\
\dot{P} &= f_E(S, P)E + f_A(S, P)A, && (27)\\
&P(0) = P_0,\ S(0) = S_0,\ P(T), S(T)\ \text{free}, E \geq 0, 0 \leq A \leq \bar{A}. && (28)
\end{aligned}$$

This is a nonholonomic optimal control problem with state variables S, P and control variables E, A. In our research we focus on adopting the nonholonomic solution techniques from mathematical physics for bioeconomic and natural resources models of type (25-28) to obtain the mathematically correct solutions and considerably improve the applicability of the optimal control results.

3 Conclusion

Incorporating the techniques of variational calculus and optimal control for solving the constrained dynamical system in the area of natural resources systems appears to be promising. Nevertheless, there exist problematic issues in the underlaying mathematical structures. In our contribution we focused on the systems with nonholonomic constraints for which two non equivalent approaches to solution exist. While in economics mainly the vakonomic approach has been adopted for solving the dynamic optimization problems with velocity constraints (see e.g. [6], [17] or [7]), in physics there are strong supporting examples and arguments (e.g. in [20]) that justify the nonholonomic approach as the physically correct one. But in the area of natural resources systems the concept of constrained systems are still to be validated. Particularly, for our further research, we state a hypothesis that in the area of calculus of variations the correctness of vakonomic approach is expected to be shown, while the optimal control applications in natural resources will fit better to the physical approach of nonholonomic constraints.

Acknowledgements

The research is supported by the Czech Science Foundation grants nr. 13-25897S and GA 201/09/0981.

References

[1] Bond, C., A., Farzin, Y. H.: Agroecosystem modeling and optimal economic decisions: Implications for sustainable agriculture *Optimal Control Applications and Methods* **29** (2008), 59–80.

[2] Clark, C., W. : *Mathematical Bioeconomics- The Mathematics of Conservation.* John Wiley & Sons Inc., New Jersey, 2010.

[3] Chiang, A., C.: Elements of dynamic optimization, Waveland Press, Inc., Longe Grove 2000.

[4] Christiaans, T., Eichner, T.,Pethig, R.: Optimal Pest Control in Agriculture *Journal of Economic Dynamics and Control* **31** (2007), 3965–3985.

[5] Epanchin-Niell, R., S., Hastings, A.: Controlling established invaders: integrating economics and spread dynamics to determine optimal management, *Ecology Letters* **13** (2010), 528–541.

[6] Grubbstrom, A.: Application of the Calculus of Variations to Financing Alternatives, *Omega* **19** (1991) 305–316.

[7] Janová, J.: Applications of a constrained mechanics methodology in economics *European Journal of Physics* **32** (2011), 1443–1463.

[8] Janová, J., Musilová, J.: Non-holonomic mechanics: A geometrical treatment of general coupled rolling motion *International Journal of Nonlinear Mechanics* **44** (2009), 98–105.

[9] Janová, J.: On treating the general constraints in agricultural and forestry optimization problems In: *Proceedings of the 27th International Conference on Mathematical Methods in Economics* (H. Brozova, Kvasnicka R. eds.), Czech Univ Life Sci, Kostelec nad Cernymi lesy, 2009, 165–168.

[10] Janová, J., Musilová, J., Bartoš, J.: Coupled rolling motion: a student project in non-holonomic mechanics *European Journal of Physics* **30** (2009), 1257–1269.

[11] Janová, J., Musilová, J.: The streetboard rider: an appealing problem in non-holonomic mechanics *European Journal of Physics* **31** (2010), 333–345.

[12] Janová, J.: Crop planning optimization model: the validation and verification processes *Central European Journal of Operations Research* **20** (2012), 451–462.

[13] Jerry, M., Raissi, N.: The Optimal Strategy for a Bioeconomical Model of a Harvesting Renewable Resource Problem, *Mathematical and Computer Modelling* **36** (2002), 1293–1306.

[14] Krupková, O.: Recent results in the geometry of constrained systems, *Rep. Math. Phys.* **49** (2002), 269–278.

[15] Loehle, C. : Control theory and the management of ecosystems, *Journal of applied ecology* **43** (2006), 957–966.

[16] Maruskin, J., M., Bloch, A., M.: The Boltzmann-Hamel equations for the optimal control of mechanical systems with nonholonomic constraints, *International Journal of Robust and Nonlinear Control* **21** (2010), 373–386.

[17] Molinder, M.: Application of calculus of variations to a continuous time aggregate production model, *International Journal of Production Economics* **41** (1994), 273–280.

[18] Pucheta, J.,et al.: A neuro-dynamic programming based optimal controller for tomato seedling growth in greenhouse systems, *Neural processing letters* **24** (2006), 241–260.

[19] Schulter, M.,et al.: Managing water-use trade-offs in a semi-arid river delta to sustain multiple ecosystem services: A modeling approach, *Ecological research* **24** (2009), 491–514.

[20] Li, S., M.: On the validity of the vakonomic model and the Chetaev model for constraint dynamical systems *Reports on Mathematical Physics* **60** (2007), 107–116.

Examining relations between Slovakia and foreign partners: a cointegrated vector autoregressive approach

Jana Juriová[1]

Abstract. Slovakia as a small open economy is highly dependent on foreign environment. The focus of this study is oriented on Slovakia's relations with its most important foreign trade partners.
Based on relationships derived from economic theory by means of AA-DD model a small open economy structural model is constructed. As most of macroeconomic variables are non-stationary, the suitability of cointegration approach is verified. The study models relations with foreign partners using a Cointegrated Vector Autoregressive approach. The following five long-run equilibrium relations are examined and adjusted for the case of Slovakia: relative purchasing power parity, money supply, output relations, uncovered interest rate parity, and Fisher inflation parity. The foreign partners are characterized by their aggregate domestic demand, consumer prices and interest rate. Also the importance of convergence of Slovakia's economic development to its foreign trade partners is investigated considering its inclusion into the model.
The first version of the estimated model is used to investigate responses of domestic macroeconomic variables to foreign shocks by impulse response functions.

Keywords: small open economy, AA-DD model, cointegrated VAR, impulse response function.

JEL Classification: C51, C52, E17
AMS Classification: 62P20

1 Introduction

A small open economy is an economy that participates in international trade, but is small enough compared to its trading partners and does likely not impact world economy to a great extent. Thus the policy of small open economy does not alter world prices or interest rates, the countries with small open economies are rather price-takers and they are highly dependent on development of foreign environment. Slovakia is a typical example of a small open economy. Moreover, its openness has been increasing over the last decade very rapidly. In 2012 its rate of openness[2] achieved 186.3%, while before 2000 it was still around 120%.

The goal of this paper is to describe and verify macroeconomic relationships between small open economy of Slovakia and its foreign environment. The foreign environment of Slovakia is represented by its most important foreign trading partners as the fluctuations of a small and open economy are mainly driven by its main trading partners [3]. Modelling of small open economy is allowing exogenous assumptions of the conditions in the foreign environment. In this context the paper investigates the long-run structural modelling approach for the Slovak economy. One of the recent attempts to construct a long-run macroeconometric model for Slovakia is e.g. in [2]. Their paper applies a practical approach to incorporate theoretic long-run relationships in a structural vector error correction model of the Slovak and Czech Republic. Their results confirmed the similarity of both economies.

The paper is organized as follows. In the second chapter the theoretical foundations of the model for small open economy are introduced. The third part contains estimated cointegrated VAR model and interpretation of results. The fourth chapter concludes.

2 Long-run equilibrium relations in small open economy (AA-DD model)

As a starting point of structural modelling of small open economy we used the small open economy model developed by Garratt et al. [1] and verified e.g. by Schneider et al. [6]. They derived from the economic theory

[1] VŠB - Technical University of Ostrava, Faculty of Economics, Sokolská třída 33, 701 21 Ostrava, Czech Republic, jana.juriova@vsb.cz.
[2] The openness of economy is defined as the share of total imports and exports of goods and services in gross domestic product at current prices.

based on RBC (real business cycle) models the following five linearized long-run relations: the uncovered interest rate parity, the purchasing power parity, production function, trade balance and real money balance. In this study, these equations were verified and modified by the model AA-DD. The model AA-DD describes equilibrium relations in small open economy. This model represents a synthesis of three market models: the foreign exchange (Forex) market, the money market, and the goods and services market (more detailed description e.g. in [7]. The AA-DD model is described with a diagram consisting of two curves: an AA curve representing asset market equilibriums derived from the money market and foreign exchange markets and a DD curve representing goods market (or demand) equilibriums. One of the basic mechanisms working in these markets is the arbitrage concept and the law of one price. The law of one price states that identical goods sell for the same price worldwide. The existence of arbitrage means that if the prices after exchange-rate adjustment were not equal worldwide, arbitrage of the goods ensures eventually that they will be. The first relationship (1) results from these international arbitrage activities. The relative purchasing power parity theory assumes a relationship between the domestic inflation and foreign inflation over a specified period and the movement in the nominal exchange rate between their two currencies over the same period. This theory states that the exchange rate of one currency against another will adjust to reflect changes in the price levels of the two countries. It means that the currency with the higher inflation rate is expected to depreciate relative to the currency with the lower rate of inflation. However, in practice not all goods and services can be bought in both countries and therefore there remains a certain difference and this leads to introducing a constant term into the equation - the relative purchasing power parity has the form (1). The second relation describes the money market equilibrium when money supply is equal to money demand (2). The real domestic money supply is determined by both domestic output and interest rate. There can be regarded also the trend which can capture the possible effect of the changing nature of financial intermediation. The important relation explaining transmission of business cycle from foreign environment into the domestic economy is output relation. The main channel of business cycle transmission is foreign output and according to Obstfeld and Krugman [4] the current account is a function of exchange rate and output. This implies the inclusion of exchange rate variable into the equation (3). The constant term represents the gap between the domestic and foreign output. The equation (4) describes the interest rate parity which is similar to purchasing power parity, however stands for relation between domestic and foreign interest rates which should determine also the nominal exchange rate. Also this parity relationship is a result of arbitrage activities and is derived from AA curve and based on the international Fisher effect which states that the exchange rate adjusts to the interest rate differential between two countries. It implies that the currency with lower interest rate is expected to appreciate relative to one with a higher rate. The last relation derived from arbitrage conditions is the Fisher inflation parity (5), which describes the relation between domestic interest rates and inflation. It is assumed that higher nominal interest rates reflect inflation.

$$\Delta p_t - \Delta ps_t - e_t = b_{10} + \varepsilon_{1,t+1} \tag{1}$$

$$m_t - \Delta p_t = b_{20} + \beta_{21} y_t + \beta_{22}\left(1 + \frac{r_t}{100}\right) + \varepsilon_{2,t+1} \tag{2}$$

$$y_t = b_{30} + \beta_{31} ys_t + \beta_{32} e_t + \varepsilon_{3,t+1} \tag{3}$$

$$\left(1 + \frac{r_t}{100_t}\right) - \left(1 + \frac{rs_t}{100_t}\right) - e_t = b_{40} + \varepsilon_{4,t+1} \tag{4}$$

$$\left(1 + \frac{r_t}{100}\right) = b_{50} + \beta_{51} \Delta p_t + \varepsilon_{5,t+1} \tag{5}$$

where p_t is domestic price level, ps_t is foreign price level, e_t is nominal exchange rate, m_t is money stock, y_t is domestic output, ys_t is foreign output, r_t is domestic interest rate, rs_t is foreign interest rate, b_{i0}, $i=1,2,\ldots,5$ stands for intercept and $\varepsilon_{i,t+1}$ are long-run stochastic terms.

In all the relations (1) – (5) we can consider also inclusion of trend term, because it can represent e.g. a convergence of Slovak Republic towards its trading partners in equations (1), (3) and (4). However, the inclusion of trend is questionable mainly in current crisis period when all economies are influenced by crisis and thus their economic development is getting closer.

3 Cointegrated VAR model

In this empirical study we apply the modelling strategy to Slovak data to see how the formulated economic theory accounts for the case of the Slovak Republic.

3.1 Input data

The variables for the model are y_t, ys_t, p_t, ps_t, r_t, rs_t, e_t and m_t. A detailed description of these variables is given below. They are all quarterly and seasonally adjusted data and they cover the period 2000Q1-2012Q4. In the model the data are used after logarithm transformation. To measure the domestic inflation we use the consumer price index unlike the paper by Garratt [1] where retail price index is used. To measure the output of foreign trading partners we use their domestic demand instead of GDP. This way we avoid the bias due to their foreign demand if they are also open economies. The most important trading partners of the Slovak Republic are identified on the basis of their shares in the foreign trade of Slovakia.

The list of variables, their description and data source

- y_t – gross domestic product of Slovakia at market prices, millions of national currency, chain-linked volumes with reference year 2005, source: Eurostat;
- ys_t – domestic demand at market prices for the most significant trading partners, millions of national currency, chain-linked volumes with reference year 2005, source: Eurostat;
- p_t – harmonized consumer price index of Slovakia (2005=100), source: Eurostat;
- ps_t – harmonized consumer price index of the most significant trading partners (2005=100), source: Eurostat;
- r_t – short-term interest rate of Slovakia - 3-month rate (%), source: OECD;
- rs_t – short-term interest rate of the most significant trading partners - 3-month rate (%), source: OECD;
- e_t – nominal effective exchange rate (NEER)[3] of Slovakia (27 trading partners), index 2005=100, source: Eurostat;
- m_t – money stock M1 of Slovakia, millions of EUR, source: National bank of Slovakia.

The aggregate variables for the most significant trading partners are composed using weights based on the shares of individual countries in the Slovak foreign trade which are varying in individual years. The most significant trading partners and their average shares in the foreign trade of Slovak Republic are given in the Table 1.

Trading partner	The share in foreign trade (%)
Germany	34.5
Czech Republic	21.1
Italy	9.0
Austria	8.1
Poland	8.0
Hungary	7.9
France	7.2
United Kingdom	4.2

Table 1 Average shares of trading partners in the foreign trade of Slovak Republic for the period 2000-2012

Investigation of the non-stationarity of data

The graphical depiction of variables under consideration (in Annex) suggests that these variables are non-stationary. The Augmented Dickey-Fuller (ADF) test was used to verify the stationarity of original time series and their first differences. The ADF test statistics for the levels and first differences of the original variables in logarithms are reported in Table 2 (the t-statistics are computed using ADF regressions with an intercept and linear time trend for levels and with an intercept for the first differences). The results of the test suggest that it is reasonable to treat all variables in consideration as I(1) variables. For these variables the unit root hypothesis is rejected when applied to their first differences at the significance level of 1%.

[3] The NEER (or, equivalently, the "Trade-weighted currency index") of a country aims to track changes in the value of that country's currency relative to the currencies of its principal trading partners. It is calculated as a weighted geometric average of the bilateral exchange rates against the currencies of competing countries.

Variable	For the levels t-statistic(sign.)	For the first differences t-stat/sign.
y_t	-0.95(0.94)	-7.09(0.00)
ys_t	-1.51(0.81)	-4.63(0.00)
Δp_t	-6.61(0.00)	-6.97(0.00)
Δps_t	-4.900(0.00)	-9.20(0.00)
r_t	-5.49(0.00)	-4.61(0.00)
rs_t	-2.88(0.18)	-4.26(0.00)
e_t	-1.70(0.74)	-5.39(0.00)
m_t	-1.63(0.77)	-6.02(0.00)

Table 2 Augmented Dickey-Fuller unit root test applied to Slovak variables

3.2 Model estimation and verification

In the first stage the order of the unrestricted VAR model is selected. VAR order of one appears to be appropriate when using the AIC and also SIC as the model selection criteria. The next step is to test the cointegration rank. The purpose of the cointegration test is to determine whether a group of non-stationary series are cointegrated or not. We applied the test with the linear trend in the data, as some time series show linear trend, and with an intercept but no trend in the cointegrating equation. The trace test and also the maximum eigenvalue test identify five cointegrating relationships among the eight variables at 5% level (Table 3).

Variable	Trace statistic (critical value)	Max eigen statistic (critical value)
$r \leq 0$	276.3664 (159.5297)	83.26436 (52.36261)
$r \leq 1$	193.1020 (125.6154)	51.65636 (46.23142)
$r \leq 2$	141.4457 (95.75366)	45.35564 (40.07757)
$r \leq 3$	96.09003 (69.81889)	40.85650 (33.87687)
$r \leq 4$	55.23352 (47.85613)	28.88881 (27.58434)
$r \leq 5$	26.34471 (29.79707)	13.06977 (21.13162)

Table 3 Johansen test of cointegration rank

Since five cointegration relations are in line with our theoretical expectations, we proceed estimating vector error correction (VEC) model with five cointegrating relations. The VEC model for the Slovak economy has the following form:

$$\Delta z_t = a_0 + \sum_{i=1}^{p-1} \Gamma_i \Delta z_{t-i} - \alpha \beta' z_{t-1} + u_t \tag{6}$$

where $z_t = (e_t\ r_t\ p_t\ y_t\ ps_t\ m_t\ rs_t\ ys_t)$', a_0 is a vector of intercepts, p is order of underlying VAR model, Γ_i are matrices of short-run coefficients, $\beta' z_{t-1}$ are the error correction terms, α is a matrix of error-correction coefficients and u_t is a vector of disturbances assumed to be white noise.

The matrix β'can be used to impose all the theoretical restrictions necessary for the structural long-run relationships:

$$\beta' = \begin{pmatrix} -1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & \beta_{22} & -1 & \beta_{24} & 0 & 1 & 0 & 0 \\ \beta_{31} & 0 & 0 & 1 & 0 & 0 & 0 & \beta_{38} \\ -1 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & \beta_{53} & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The estimation of a VEC model is carried out in two steps. In the first step, we estimate the cointegrating relations from the Johansen procedure as used in the cointegration test. We then construct the error correction terms from the estimated cointegrating relations and estimate a VAR in first differences including the error cor-

rection terms as regressors. The estimated long-run relationships for the Slovak economy, incorporating all the restrictions (35 restrictions) suggested by the theory are the following (with *t*-statistics in []):

$$\Delta p_t - \Delta ps_t - e_t = 4.678 + \varepsilon_{1,t+1} \tag{7}$$

$$m_t - \Delta p_t = 3.982 - 1.47 y_t + 7.409\left(1 + \frac{r_t}{100}\right) + \varepsilon_{2,t+1} \tag{8}$$

[-20.650] [16.159]

$$y_t = -20.483 + 1.479 ys_t - 2.658 e_t + \varepsilon_{3,t+1} \tag{9}$$

[3.410] [-14.819]

$$\left(1 + \frac{r_t}{100_t}\right) - \left(1 + \frac{rs_t}{100_t}\right) - e_t = 4.678 + \varepsilon_{4,t+1} \tag{10}$$

$$\left(1 + \frac{r_t}{100}\right) = 0.405 - 44.287 \Delta p_t + \varepsilon_{5,t+1} \tag{11}$$

[-9.914]

The first relation confirms the existence of cointegration relationship in purchasing power parity relation, thus empirical evidence on Slovak data supports this relationship. The value of intercept is also very close to that estimated for UK in [1]. The signs of long-run elasticities of the influence of the domestic output and interest rate on the real money supply are the opposite than expected according to the theory. The third long-run output relationship describes the influence of foreign output on the domestic output. The long-run elasticity of the influence of foreign output on domestic output is about 1.5%. The interest rate parity relation includes the intercept, which can be interpreted as the deterministic component of the risk premium, which is approx. 18.7% per annum. However, this value is rather high. The constant in Fisher inflation parity relation implies that the average long-run Slovak interest rate is about 4%. Also the importance of convergence of Slovakia to its foreign trade partners was investigated considering the inclusion of trend into the model. However, the results showed that the estimated parameters for trend terms are very low, though statistically significant and that we can consider the trend terms negligible in the model.

3.3 Impulse response analysis

The short-term dynamic reactions of domestic variables to foreign shock effects are examined by general impulse response functions. Generalized impulses as described by Pesaran and Shin [5] construct an orthogonal set of innovations that does not depend on the VAR ordering. We provide here the examples of the effects of exogenous shocks to a unit (one standard error) unexpected increase on the domestic endogenous variables. The Figure 1 includes the responses of Slovak variables to two shocks from its foreign trading partners: foreign demand shock and foreign monetary shock.

The responses of Slovak GDP to foreign shocks show very high persistence and the Slovak real output reacts positively to both foreign positive shocks in consideration – to the increase in foreign GDP and also in inflation. The temporary increase in GDP of the most important trading partners of Slovakia resulted in longer period of increasing Slovak GDP, short-term growth of Slovak prices a very short period of appreciation of the Slovak effective exchange rate followed by longer period of depreciation. A temporary increase in the foreign inflation causes a short-run rise in domestic inflation, a moderate growth of Slovak GDP and an appreciation of the currency, which is followed by longer period of depreciation. An increase in the foreign price level we can interpret in line with purchasing power parity relation in the following way. Increasing foreign price level causes the rise in domestic price level and the exchange rate depreciates in the long run. In that case it means that the Slovak inflation should be mostly higher than foreign inflation, which is evidenced also by the data used.

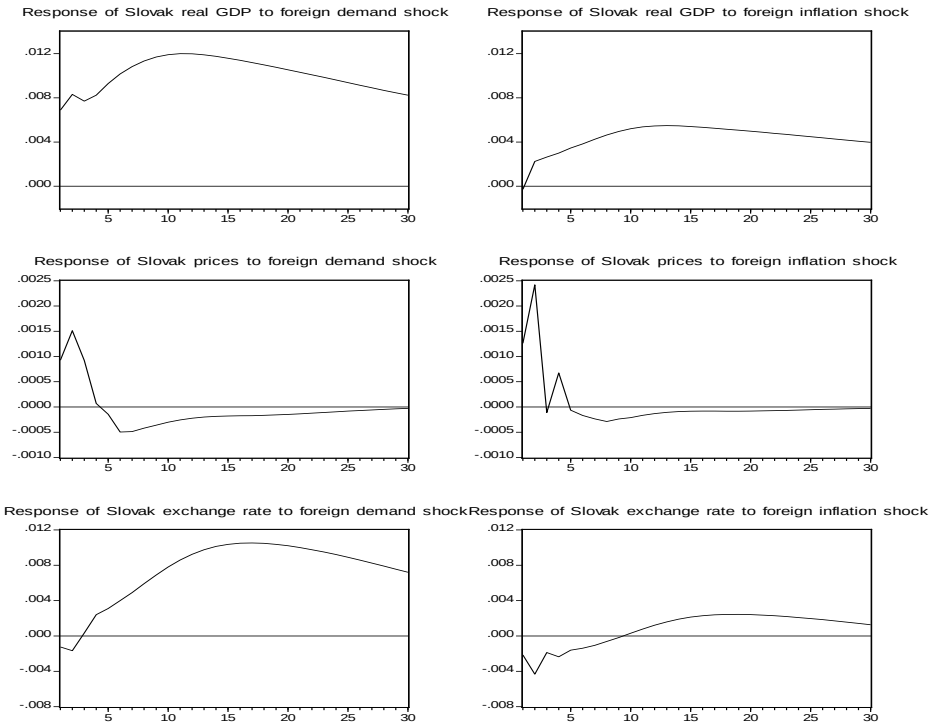


Figure 1 Impulse responses of unit foreign shocks

4 Conclusion

This paper examines long-run cointegration model on the Slovak data. The theoretical model is adjusted according to AA-DD model framework. The proposed relationships were confirmed to be cointegrated and can be used to examine the effects of a one-time rise in the foreign variables on the domestic variables. The results of impulse response analysis confirmed that the demand shock from the most important trading partners of Slovakia has high and persisting influence on the Slovak output. The positive foreign inflation shock is translated into immediate increase of Slovak prices and depreciation of Slovak currency in the long run. In future research the current cointegration model could be extended introducing a break variable.

Acknowledgement

This paper was funded by the Student Grant Competition of VŠB - Technical University of Ostrava within the project SP2013/91 "Modelling cyclical development of small open economy (theoretical concepts)".

References

[1] Garratt, A., Lee, K., Pesaran, M.H., and Shin, Y.: A Long-Run Structural Macroeconometric Model of the UK. *The Economic Journal* **113** (2003), 412-455.

[2] Hančlová, J., Lukáčik, M., Szomolányi, K.: A VEC Model of the Czech and Slovak Economies. In: *Proceedings of 28th International Conference on Mathematical Methods in Economics 2010*. České Budějovice: University of South Bohemia, 2010, 208-213.

[3] Kľúčik, M.: VAR model with current-optimal leading indicators. In: *Proceedings of 30th International Conference on Mathematical Methods in Economics 2012*. Karviná: Silesian University, School of Business Administration, 2012, 408-413.

[4] Obstfeld, M. Krugman, P.: *International Economics: Theory and Policy*. Addison Wesley, 2005.

[5] Pesaran, H., Shin, Y.: Generalized impulse response analysis in linear multivariate models. *Economics Letters* **58** (1998), 17-29.

[6] Schneider, E., Chen, P., Frohn, J.: A Long-Run Structural Macroeconomic Model for Germany: An Empirical Note. *Economics: The Open-Access, Open-Assessment E-Journal* **2** (2008).

[7] Suranovic, S.: *International Economics: Theory and Policy*. Flatworld Knowledge, 2011.

Interactions between sovereign credit default swaps and bonds: The case of EU countries

Veronika Kajurová[1], Jana Hvozdenská[2]

Abstract. Credit default swap markets are considered as a leading indicator of the future development of creditworthiness, which can point out the potential situation in economy. The study examines the relationship between sovereign credit default swap and bond spreads of the EU countries in the period before, during and after financial crisis. The aim of the study is to find out if the price discovery process in the credit default swap and bond markets has been changed by financial crisis and continuing debt crisis in short-run. A vector autoregressive model in the context of Granger causality is employed to capture causal relationships between observed credit default swap and bond markets. Model is employed on daily and weekly data. Results can be beneficial for all participants in the financial markets, especially for regulators and investors as a possible indicator of credit risk. This research showed that the role of both markets has changed. We found out that the number of causal relations grew during the financial crisis period and has decreased during the debt crisis period. The credit default swap markets could be comprehended as a potential indicator of increasing credit risk during the period of financial crisis but they cannot be comprehended anymore in the presence.

Keywords: credit default swap market, bond market, Granger causality

JEL Classification: C32, G15
AMS Classification: 91G70

1 Introduction

Both government bonds and sovereign credit default swaps (CDS) offer investors exposure to the risk and return of sovereign debt. Increased attention to the relation between CDS and bonds started to be devoted after the outbreak of the financial crisis. Lead-lag analysis between CDS markets and bond markets is one of the centers of attention because of exploration the adjustment process between bond and CDS spreads, which can be beneficial for all participants in the financial markets.

First such analysis was published in empirical study by Blanco, Brennan and Marsh [2], which paid attention mainly to the relationship between investment grade bonds and CDS. It was the first study, in which a time series framework was used for credit derivatives. They used data for US and European firms. Besides other findings, they found out the CDS markets "lead" the bond markets. Except this contribution, there are several empirical papers which investigate lead-lag relationship between CDS markets and bond markets, see [5], [9], [7] or [11]. All of them confirm the leading role of CDS with respect to bond markets in majority of cases, which proves the CDS markets move ahead of the bond markets in price adjustment.

There are two main reasons why credit default swap markets are expected to lead the bond markets. 1) The information is reflected earlier in the credit default swap market than in the bond market. 2) Institutional features of the credit default swap markets facilitate a continuous flow of transactions, because short positions in the bond markets are more difficult to establish [9].

Another stream of research is focused on analysis of relationship between sovereign CDS and government bonds. Theoretically, credit default swap markets are considered as a leading indicator of the future development of creditworthiness, which can point out the potential situation in economy. It was confirmed in analysis employed by Palladini and Portes [10]. Contrary, this theoretical knowledge has not been proved in other empirical research, e.g. see [6], who found out that since 2008, in half of the sample countries, price discovery takes place in the bond market and half in the CDS market; or [1], who concluded that the price discovery process is state-dependent and that there are several significant factors in determining which market leads price discovery.

[1] Masaryk University, Faculty of Economics and Administration, Department of Finance, Lipová 41a, 602 00 Brno, Czech Republic, email: vkajurova@mail.muni.cz
[2] Masaryk University, Faculty of Economics and Administration, Department of Finance, Lipová 41a, 602 00 Brno, Czech Republic, email: 174974@mail.muni.cz

In our study, we try to explore interaction between sovereign CDS and government bond spreads of EU countries. The aim of the study is to find out whether the price discovery process in the CDS and bond markets has been changed by financial crisis and continuing debt crisis.

2 Data

In our research, we pay attention to 20 members of the European Union for which data for 5 years sovereign CDS and bonds were available. All data were obtained from Bloomberg on daily and weekly basis.

The total sample period (March 2003 – May 2013) is divided into three sub-periods according to trends in development of the Markit iTraxx SovX Western Europe Index:

- Pre-crisis period (March 2003 – May 2007),
- Financial crisis period (June 2007 – October 2009),
- Debt crisis period (November 2009 – May 2013).

The start of observation differs for each country. It depends on availability of data for CDS.

The crisis period is understood as period of the biggest turmoil in the financial markets. Then the crisis has been transformed into a sovereign debt crisis that has not yet finished, therefore the period after the financial turmoil is denoted as the debt crisis period.

3 Methodology

Granger causality tests are employed to estimate linkages between prices of bond and CDS. It allows us to discover short-run causality. This causality test was developed by Granger [8]. General concept of Granger causality is related to prediction. In case that one time series has an influence on the second time series, the first one should help to improve predictions of the second one [4]. Granger causality means only correlation between present value of one variable and past values of other variables [3].

The standard Granger causality model for two variables can be represented as:

$$Y_t = \sum_{i=1}^{\rho} \alpha y_i \, Y_{t-1} + \sum_{i=1}^{\rho} \alpha x_i \, X_{t-1} + \varepsilon_t \quad (1)$$

$$X_t = \sum_{i=1}^{\rho} \beta x_i \, \Delta X_{t-1} + \sum_{i=1}^{\rho} \beta y_i \, Y_{t-1} + \varepsilon_t \quad (2)$$

where Y_t and X_t are stationary time series, ε_t is uncorrelated white noise, αx_i and βx_i are coefficients chosen to minimize σ^2 (variance), ρ is finite and shorter than the given time series (it can equal infinity but in practice, it is finite due to the length of the available data).

The null hypothesis "X does not Granger cause Y" (equation 1) or "Y does not Granger cause X" (equation 2) is rejected if the coefficients αx_i and βy_i are jointly significant.

In our research, Granger causality test tries to find if credit default swap prices do "Granger-cause" bond prices (past values of credit default swap prices improve the prediction of bond prices), and vice versa - if bond prices do "Granger-cause" credit default swap prices (past values of bond prices improve the prediction of credit default swap prices). We include five lags (in our case 5 days) within lag specification of Granger causality test.

4 Results

The Granger causality tests enabled us to examine two null hypotheses.

1) CDS "does not Granger cause" Bond.
2) Bond "does not Granger cause" CDS.

4.1 Daily data

Table 1 shows summary results of Granger causality tests in all periods employed on daily data.

Full Sample

For full sample data, results indicate that rates of return of credit default swap of six countries of European Union (from total 19 tested) "Granger causes" rates of return of bond. Detailed results are shown in Table 2. Probabilities are lower than 10 % significance level in these cases, therefore the null hypothesis "CDS does not Granger cause Bond" can be rejected in these six cases (Spain, Portugal, Sweden, Hungary, Poland, Czech Re-

public). The second null hypothesis "Bond does not Granger cause CDS" can be rejected for eight tested countries of European Union (Italy, Spain, Portugal, Sweden, Netherlands, Hungary, Poland, Czech Republic).

Period	No. of countries	No. of causal relationships		
		CDS -> bJe ond	Bond -> CDS	Mutual
Full sample	19	6	8	6
Pre-crisis period	12	0	0	0
Financial crisis	16	6	11	5
Debt crisis	19	5	2	1

Table 1 Summary of Granger causality test results on daily data (at 10% significance level)

A mutual relationship from CDS to Bond was proved in six cases (Spain, Portugal, Sweden, Hungary, Poland, Czech Republic). It means CDS is not strongly exogenous in any case. Bond is exogenous only in two cases. Both null hypotheses can be rejected in six cases simultaneously. It shows that there is a backward relation between variables. Full sample results do not indicate the relationships between variables were changing in the time, therefore our attention was paid to the particular periods.

Country	Null hypothesis	Start of the period	Probability			
			Full sample	Pre-crisis	Financial crisis	Debt crisis
United Kingdom	CDS -> Bond	08/11/2008	0,3858	-	0,9516	0,1738
	Bond -> CDS		0,1354	-	0,0397**	0,1340
France	CDS -> Bond	04/07/2003	0,8550	0,9734	0,5946	0,6535
	Bond -> CDS		0,1624	0,5941	0,5143	0,2229
Germany	CDS -> Bond	03/17/2003	0,9949	0,9177	0,7162	0,5547
	Bond -> CDS		0,6762	0,8125	0,8258	0,6722
Italy	CDS -> Bond	03/17/2003	0,1237	0,4727	0,2290	0,0416**
	Bond -> CDS		2*E-09***	0,6100	0,1538	0,4970
Spain	CDS -> Bond	04/12/2004	0,0952*	0,8200	0,1541	0,0092***
	Bond -> CDS		0,0005***	0,9072	0,0862*	0,2058
Portugal	CDS -> Bond	04/07/2003	1*E-06***	0,2564	0,3871	4*E-14***
	Bond -> CDS		2*E-07***	0,9568	0,0858*	0,1786
Sweden	CDS -> Bond	01/15/2007	0,0511*	0,4561	0,0009***	0,6387
	Bond -> CDS		0,0212**	0,4860	0,0024***	0,5381
Netherlands	CDS -> Bond	09/08/2008	0,2067	-	0,6356	0,4426
	Bond -> CDS		0,00933*	-	0,0870*	0,4833
Slovenia	CDS -> Bond	04/12/2010	0,1390	-	-	0,1390
	Bond -> CDS		0,1028	-	-	0,1028
Ireland	CDS -> Bond	03/17/2003	0,9846	-	0,5567	0,2530
	Bond -> CDS		0,5934	-	0,0002***	0,8558
Hungary	CDS -> Bond	03/05/2007	6*E-06***	0,5684	0,0011***	0,0009***
	Bond -> CDS		0,0002***	0,6177	0,0006***	0,2391
Poland	CDS -> Bond	03/05/2007	0,0818*	0,7480	0,0009***	0,6103
	Bond -> CDS		0,0003***	0,3642	0,0035***	0,0342**
Belgium	CDS -> Bond	03/17/2003	0,2628	0,5924	0,1304	0,1258
	Bond -> CDS		0,7940	0,9747	0,9474	0,6122
Romania	CDS -> Bond	04/28/2003	0,7577	-	-	0,7577
	Bond -> CDS		0,5132	-	-	0,5132
Czech Republic	CDS -> Bond	03/05/2007	0,0101**	0,5896	0,0019***	0,4783
	Bond -> CDS		0,0155**	0,8886	0,0005***	0,9162
Bulgaria	CDS -> Bond	06/28/2010	-	-	-	-
	Bond -> CDS		-	-	-	-
Denmark	CDS -> Bond	04/12/2004	0,9824	0,6691	0,0002***	0,4783
	Bond -> CDS		0,9931	0,1674	0,00005***	0,9162

Slovakia	CDS -> Bond	04/12/2010	0,2142	-	-	0,2142
	Bond -> CDS		0,9492	-	-	0,9492
Finland	CDS -> Bond	10/15/2007	0,8059	-	0,5067	0,5418
	Bond -> CDS		0,2063	-	0,0575*	0,3691
Greece	CDS -> Bond	03/05/2007	0,8516	0,4441	0,0023***	0,0471**
	Bond -> CDS		0,2855	0,8847	0,3447	0,0279**

Table 2 Detailed results of pairwise Granger causality tests (daily data)

*denotes significance at 10% level; ** denotes significance at 5% level; *** denotes significance at 1% level

Pre-crisis period

There was not any Granger causality detected in all 12 observed countries during the pre-crisis period. Both null hypotheses "CDS does not Granger cause" and "Bond does not Granger cause" cannot be rejected in any case. Financial markets were stable during this period without doubts about creditworthiness.

Financial crisis period

During the financial crisis period, we observed 16 countries of European Union. Results of the Granger causality test indicate that the null hypothesis "CDS does not Granger cause Bond" can be rejected in six cases (Sweden, Hungary, Poland, Czech Republic, Denmark, Greece). Only in one case (Greece) CDS is strongly exogenous variable.

The null hypothesis "Bond does not Granger cause CDS" can be rejected in 11 cases. Bond is strongly exogenous variable in five cases. Five relationships are mutual. Increased number of relationships can be caused as a consequence of turmoil in the financial markets.

Debt crisis period

The number of relationships has decreased compared to the financial crisis period. Only one relationship is mutual. In this period, we found five causal relationships (Italy, Spain, Portugal, Hungary, Greece) from 19 observed cases in total for the hypothesis "CDS does not Granger cause Bond". CDS is strongly exogenous in four cases. Two relationships were found for the hypothesis "Bond does not Granger cause CDS" (Poland, Greece). Bond is strongly exogenous only in one case.

4.2 Weekly data

Table 3 shows summary results of Granger causality tests in all periods employed on weekly data. Employing Granger causality tests on weekly data showed that they do not confirm results on daily data. The relationship between CDS and bonds is close and because of fast information absorption they do not have such importance in in short-run.

Full Sample

For full sample data, results indicate that rates of return of credit default swap of six countries of European Union (from total 20 tested) "Granger causes" rates of return of bond. Probabilities are lower than 10 % significance level in these cases, therefore the null hypothesis "CDS does not Granger cause Bond" can be rejected in six cases (Sweden, Hungary, Poland, Czech Republic, Slovakia, Greece). The second null hypothesis "Bond does not Granger cause CDS" can be rejected for two observed countries of European Union (Italy and Ireland).

A mutual relationship from CDS to Bond was not proved in any case. It means that CDS is strongly exogenous in six cases and Bond is exogenous in two cases.

Period	**No. of countries**	**No. of causal relationships**		
		CDS -> bond	**Bond -> CDS**	**Mutual**
Full sample	20	6	2	0
Pre-crisis period	12	1	2	0
Financial crisis	16	6	1	1
Debt crisis	20	2	0	0

Table 3 Summary of Granger causality test results on weekly data (at 10% significance level)

Pre-crisis period

Almost any Granger causality was not detected in all 12 observed countries during the pre-crisis period. The null hypothesis "CDS does not Granger cause Bond" can be rejected only in one case (Spain). The hypothesis "Bond does not Granger cause CDS" can be rejected in two cases (Italy and Greece). Results on weekly data are almost same as data on daily basis.

Financial crisis period

During the financial crisis period, we observed 16 countries of European Union. Results of the Granger causality test indicate that the null hypothesis "CDS does not Granger cause Bond" can be rejected in six cases (Spain, Portugal, Ireland, Czech Republic, Denmark, Greece). All cases except of Ireland are strongly exogenous variable. The null hypothesis "Bond does not Granger cause CDS" can be rejected in one case. Bond is not strongly exogenous variable in this case. Only one relationship is mutual.

Debt crisis period

The number of relationships has decreased compared to the financial crisis period. There is no mutual relationship. In this period we found two causal relationships (Czech Republic and Slovakia) from 20 observed cases in total for the hypothesis "CDS does not Granger cause Bond". CDS is strongly exogenous in these cases. No relationship was found for the hypothesis "Bond does not Granger cause CDS" (Poland, Greece). .

5 Conclusions

The aim of the study was to find out if the price discovery process in the CDS and bond markets has been changed by financial crisis and continuing debt crisis in short-run. Granger causality tests were employed to discover relations in the particular periods. Our research showed that the role of both markets has changed. We found out that the number of causal relations grew during the financial crisis period and has decreased during the debt crisis period. According to the obtained results the credit default swap markets could be comprehended as a potential indicator of increasing credit risk during the period of financial crisis but they cannot be comprehended anymore in the presence. The empirical studies showed important role of corporate CDS and corporate bond markets as a leading indicators of the future development of creditworthiness. Our research showed that this rule does not hold when it is examined on sovereign CDS and government bond markets. However, the role of the CDS and bond markets was changed during the financial crisis, the sovereign CDS and government bond markets could be considered as leading indicators, the role of them has changed during the period of debt crisis and the sovereign CDS and government bond markets cannot be considered as leading indicators of the future creditworthiness of the country anymore. These findings can be useful for all participants of the financial markets, policy makers and regulators. Attention in the future research should be paid to long-run causality and other determinants, which can affect prices of sovereign CDS and bonds.

Acknowledgements

Support of Masaryk University within the project MUNI/A/0753/2012 "European financial systems 2013" (Student Project Grant at MU) is gratefully acknowledged.

References

[1] Arce, O., Mayordomo, S. and Peña, J. I.: Credit-risk valuation in the sovereign CDS and bonds markets: Evidence from the euro area crisis. *Working paper.* CNMV, 2012. ISSN: 2172-7147.
[2] Blanco, R., Brennan, S. and Marsh, I. W.: An empirical analysis of the dynamic relationship between investment grade bonds and credit default Swaps. Journal of Finance **5** (2005), 2255-2281.
[3] Brooks, C.: Introductory Econometrics for Finance. Cambridge University Press, Cambridge, 2008.
[4] Cipra, T.: Finanční ekonometrie (Financial econometrics). EKOPRESS, s.r.o., Prague, 2008.
[5] Dötz, N.: Time-varying contributions by the corporate bond and CDS markets to credit risk price discovery. *Discussion Paper Series 2*: Banking and Financial Studies. 2007, 08, Deutsche Bundesbank, Research Centre.
[6] Fontana, A. and Scheicher, M.: An analysis of euro area sovereign CDS and their relation with government bonds. *Working Paper Series 1271*, 2010, European Central Bank.
[7] Forte, S. and Peña, J. I.: Credit spreads: An empirical analysis on the informational content of stocks, bonds, and CDS. *Journal of Banking & Finance* **33** (2009), 2013-2025.
[8] Granger, C. W. J.: Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37** (1969), 424-438.

[9] Norden, L. and Weber, M.: The co-movement of credit default swap, bond and stock markets: an empirical analysis. *European Financial Management* **15** (2009), 529-562.
[10] Palladini, G. and Portes, R.: Sovereign CDS and bond pricing dynamics in the Euro-area. *CEPR Discussion Papers 8651*, 2011, C.E.P.R. Discussion Papers.
[11] Zhu, H.: An empirical comparison of credit spreads between the bond market and the credit default swap market. *Journal of Financial Services Research* **29** (2006), 211-235.

Economic and Financial Problems via Multiobjective Stochastic Optimization

Vlasta Kaňková[1]

Abstract. Multiobjective optimization problems depending on a probability measure correspond to many economic and financial activities. Evidently if the probability measure is completely known, then we can try to influence economic process employing methods of a multiobjective deterministic optimization theory. Since this assumption is fulfilled very seldom we have mostly to analyze the mathematical model and consequently also economic process on the data base. The aim of the talk will be to investigate a relationship between "characteristics" obtained on the base of complete knowledge of the probability measure and them obtained on the above mentioned data base. To this end, the results of the deterministic multiobjective optimization theory and the results obtained for stochastic one objective problems will be employed.

Keywords: Stochastic multiobjective optimization problems, efficient solution, Wasserstein metric, $\mathcal{L}_1$ norm, stability, empirical estimates, Lipschitz property.

JEL classification: C44
AMS classification: 90C15

1 Introduction

To introduce a "rather general" multiobjective stochastic programming problem, let $(\Omega, \mathcal{S}, P)$ be a probability space; $\xi := \xi(\omega) = (\xi_1(\omega), \ldots, \xi_s(\omega))$ s–dimensional random vector defined on $(\Omega, \mathcal{S}, P)$; $F(:= F(z),\, z \in R^s)$, P_F and Z_F denote the distribution function, the probability measure and the support corresponding to ξ. Let, moreover, $g_i := g_i(x,\, z)$, $i = 1, \ldots, l$, $l \geq 1$ be real–valued (say, continuous) functions defined on $R^n \times R^s$; $X_F \subset X \subset R^n$ be a nonempty set generally depending on F, and $X \subset R^n$ be a nonempty deterministic set. If the symbol E_F denotes the operator of mathematical expectation corresponding to F and if for every $x \in X$ there exist finite $\mathsf{E}_F g_i(x, \xi)$, $i = 1, \ldots, l$, then a rather general "multiobjective" one–stage stochastic programming problem can be introduced in the form:

$$\text{Find} \quad \min \mathsf{E}_F g_i(x, \xi),\ i = 1, \ldots, l \quad \text{subject to} \quad x \in X_F. \tag{1}$$

The multiobjective problem (1) corresponds evidently to economic situation in which a "result" of an economic process is simultaneously influenced by a random factor ξ and a decision parameter x, it is reasonable to evaluate this process by a few (say l, $l \geq 1$) objective functions. The decision vector has to be determined without knowledge of the random element realization and it seems to be reasonable to determine "the decision" with respect to the mathematical expectation of the objectives.

It is possible only very seldom to find out simultaneously the solution with respect to all criteria in (1) and moreover, these problems depend on a probability measure P_F that usually has to be estimated on the data base. Consequently, in applications very often the "underlying" probability measure P_F has to be replaced by empirical one. Evidently, then the "solution" and an analysis of the problem have to be done with respect to an empirical problem:

$$\text{Find} \quad \min \mathsf{E}_{F^N} g_i(x, \xi),\ i = 1, \ldots, l \quad \text{subject to} \quad x \in X_{F^N}, \tag{2}$$

[1]Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 4, 182 08 Praha 8, Czech Republic, e-mail: kankova@utia.cas.cz

where F^N denotes an empirical distribution function determined by a random sample $\{\xi^i\}_{i=1}^N$ (not necessary independent) corresponding to the distribution function F.

To analyze the problem (1), first, the results of the multiobjective deterministic problems have to be recalled. Since, it follows from multiobjective theory that the results of uni–objective optimization theory can be useful (under rather general conditions) to investigate the relationship between the results obtained under complete knowledge of P_F and them obtained on the data base, we recall also the results obtained for uni–objective stochastic programming problems. Our aim will be to focus to "underlying" distributions with heavy tails, that correspond just to many economic and financial processes (for more details see e.g. [10] or [12]).

According to the above mentioned facts, the paper is organized as follows. First, we try to recall auxiliary assertions concerning deterministic multiobjective theory (subsection 2.1). Stability and empirical estimates obtained for uni–objective stochastic programming problems are recalled in section 2.2. Section 3 is devoted to the stability analysis of the problem (1). Essential results devoted to the multiobjective problems can be found in section 4.

2 Some Definition and Auxiliary Assertion

2.1 Deterministic Multiobjective Problems

To recall some results of the multiobjective deterministic optimization theory we consider a multiobjective deterministic optimization problem in the following form:

$$\text{Find} \quad \min f_i(x),\ i = 1, \ldots, l^{'} \quad \text{subject to} \quad x \in \mathcal{K}, \tag{3}$$

where $f_i(x)$, $i = 1, \ldots, l^{'}$ are real–valued functions defined on R^n, $\mathcal{K} \subset R^n$ is a nonempty set.

Definition 1. The vector x^* is an efficient solution of the problem (3) if and only if there exists no $x \in \mathcal{K}$ such that $f_i(x) \leq f_i(x^*)$ for $i = 1, \ldots, l^{'}$ and such that for at least one i_0 one has $f_{i_0}(x) < f_{i_0}(x^*)$.

Definition 2. The vector x^* is properly efficient solution of the multiobjective optimization problem (3) if and only if it is efficient and if there exists a scalar $M > 0$ such that for each i and each $x \in \mathcal{K}$ satisfying $f_i(x) < f_i(x^*)$ there exists at least one j such that $f_j(x^*) < f_j(x)$ and

$$\frac{f_i(x^*) - f_i(x)}{f_j(x) - f_j(x^*)} \leq M. \tag{4}$$

Proposition 1. ([4]) *Let $\mathcal{K} \subset R^n$ be a nonempty convex set and let $f_i(x)$, $i = 1, \ldots, r$ be convex functions on $\mathcal{K}$. Then x^0 is a properly efficient solution of the problem (3) if and only if x^0 is optimal in*

$$\min_{x \in \mathcal{K}} \sum_{i=1}^{r} \lambda_i f_i(x) \quad \text{for some} \quad \lambda_1, \ldots, \lambda_r > 0; \quad \sum_{i=1}^{r} \lambda_i = 1.$$

Definition 3. Let $\bar{h}(x)$ be a real–valued function defined on a nonempty convex se $\mathcal{K} \subset R^n$. $\bar{h}(x)$ is a strongly convex function with a parameter $\rho > 0$ if

$$\bar{h}(\lambda x^1 + (1-\lambda)x^2) \leq \lambda \bar{h}(x^1) + (1-\lambda)\bar{h}(x^2) - \lambda(1-\lambda)\rho\|x^1 - x^2\|^2 \quad \text{for every} \quad x^1, x^2 \in \mathcal{K},\ \lambda \in \langle 0, 1\rangle.$$

Proposition 2. ([6]) *Let $\mathcal{K} \subset R^n$ be a nonempty, compact, convex set. Let, moreover, $\bar{h}(x)$ be a strongly convex with a parameter $\rho > 0$, continuous, real–valued function defined on $\mathcal{K}$. If x^0 is defined by the relation $x^0 = \arg\min_{x \in \mathcal{K}} \bar{h}(x)$, then*

$$\|x - x^0\|^2 \leq \frac{2}{\rho}|\bar{h}(x) - \bar{h}(x^0|| \quad \text{for every} \quad x \in \mathcal{K}.$$

2.2 Uni–Objective Stochastic Programming Problems

To recall suitable for us assertions of one criteria stochastic optimization theory we start with the problem:

$$\text{Find} \quad \varphi(F, X_F) = \inf \mathsf{E}_F g_0(x, \xi) \quad \text{subject to} \quad x \in X_F, \tag{5}$$

where $g_0(x,\, z)$ is a real–valued function defined on $R^n \times R^s$.

First, if F and G are two s–dimensional distribution functions for which the Problem (5) is well defined, then we can obtain by the triangular inequality that

$$|\varphi(F,\, X_F) - \varphi(G,\, X_G)| \,\leq\, |\varphi(F,\, X_F) - \varphi(G,\, X_F)| \,+\, \varphi(G,\, X_F) - \varphi(G,\, X_G)|. \tag{6}$$

According to the relation (6) we can study separately stability of the problem (5) with respect to perturbation in the objective function and in constraints set. In this paper we restrict our consideration to the case $X_F = X$ independently of F. To this end we introduce the following assumptions:

A.1
- X is a convex set and there exists $\varepsilon > 0$ such that $g_0(x,\, z)$ is a convex bounded function on $X(\varepsilon)$ ($X(\varepsilon)$ denotes the ε–neighborhood of X),
- $g_0(x, z)$ is a Lipschitz function of $z \in R^s$ with the Lipschitz constant L (corresponding to the $\mathcal{L}_1$ norm) not depending on x.

To introduce the first assertion dealing with the stability of the problem (5) (with $X_F = X$) we denote by $F_i, i = 1, \ldots, s$ one–dimensional marginal distribution functions corresponding to F; $\mathcal{P}(R^s)$ the set of Borel measures on R^s, $\mathcal{M}_1(R^s) = \{P \in \mathcal{P}(R^s) : \int\limits_{R^s} \|z\|_s^1 P(dz) < \infty\}$, $\|\cdot\|_s^1$ denote $\mathcal{L}_1$ norm in R^s.

Proposition 3. ([7]) *Let $P_F, P_G \in \mathcal{M}_1(R^s)$, X be a nonempty set. If A.1 is fulfilled, then*

$$|\mathsf{E}_F g_0(x,\, \xi) - \mathsf{E}_G g_0(x,\, \xi)| \leq L \sum_{i=1}^{s} \int\limits_{-\infty}^{+\infty} |F_i(z_i) - G_i(z_i)| dz_i \quad \text{for every} \quad x \in X.$$

Proposition 3 reduces (from the mathematical point of view) s–dimensional case to one–dimensional. Of course, stochastic dependence between components of the random vector ξ is there neglected. Replacing G by an empirical estimate F^N of F we can employ Proposition 3 to investigate empirical estimates of Problem (5) (with $X_F = X$) and according to Proposition 1 also to analyze relationship between Problems (1) and (2). Evidently, according to Proposition 3 it is reasonable to investigate the behaviour of $\int\limits_{-\infty}^{\infty} |F_i(z_i) - F_i^N(z_i)| dz_i$, $i = 1, \ldots, s$. To this end, we recall the following assumptions:

A.2 $\{\xi^i\}_{i=1}^{\infty}$ is independent random sequence corresponding to F, F^N is an empirical distribution function determined by $\{\xi^i\}_{i=1}^{N}$,

A.3 $P_{F_i}, i = 1, \ldots, s$ are absolutely continuous w. r. t. the Lebesgue measure on R^1.

Proposition 4. ([15]) *Let $s = 1$ and $P_F \in \mathcal{M}_1(R^1)$. Let, moreover A.2 be fulfilled . Then*

$$P\{\omega : \int\limits_{-\infty}^{\infty} |F(z) - F^N(z)| dz \longrightarrow_{N \longrightarrow \infty} 0\} = 1.$$

Proposition 5. [8] *Let $s = 1$, $t > 0$ and Assumptions A.2, A.3 be fulfilled. If there exists $\beta > 0$, $R := R(N) > 0$ defined on $\mathcal{N}$ such that $R(N) \xrightarrow[N\to\infty]{} \infty$ and, moreover,*

$$\begin{aligned}
&N^\beta \int\limits_{-\infty}^{-R(N)} F(z) dz \xrightarrow[N\to\infty]{} 0, \qquad N^\beta \int\limits_{R(N)}^{\infty} [1 - F(z)] dz \xrightarrow[N\to\infty]{} 0,\\
&2NF(-R(N)) \xrightarrow[N\to\infty]{} 0, \qquad 2N[1 - F(R(N))] \xrightarrow[N\to\infty]{} 0,\\
&(\tfrac{12N^\beta R(N)}{t} + 1) \exp\{-2N(\tfrac{t}{12R(N)N^\beta})^2\} \xrightarrow[N\longrightarrow\infty]{} 0,
\end{aligned} \tag{7}$$

then

$$P\{\omega :\, N^\beta \int\limits_{-\infty}^{\infty} |F(z) - F^N(z)| dz > t\} \xrightarrow[N\to\infty]{} 0. \tag{8}$$

($\mathcal{N}$ denotes the set of natural numbers.)

Evidently, it follows from the relations (7), (8) and from the classical result of [2] that the validity of the relation (8) depends on the tails behaviour (for more details see e.g. [9]).

Proposition 6 ([5]). *Let $s = 1$, $t > 0$, $r > 0$, the assumptions A.2, A.3 be fulfilled. Let, moreover, ξ be a random variable such that $\mathsf{E}_F|\xi|^r < \infty$. If constants $\beta, \gamma > 0$ fulfil the inequalities $0 < \beta + \gamma < 1/2, \quad \gamma > 1/r, \quad \beta + (1-r)\gamma < 0$, then*

$$P\{\omega : N^\beta \int_{-\infty}^{\infty} |F(z) - F^N(z)|dz > t\} \longrightarrow_{N \longrightarrow \infty} 0.$$

Analyzing Proposition 6 we can obtain $\beta := \beta(r)$ fulfilling this assertion and simultaneously

$$\beta(r) \longrightarrow_{r \longrightarrow \infty} 1/2, \quad \beta(r) \longrightarrow_{r \longrightarrow 2^+} 0.$$

Proposition 6 covers also some cases of heavy tails distributions. Unfortunately, we cannot obtain by this Proposition any results for the case when there exist only $\mathsf{E}_F|\xi|^r$ for $r < 2$. But just this case corresponds to stable distributions with the tail (shape) parameter $\nu < 2$ (for more details see e.g. [11] or [13]). The shape parameter expresses how "heavy" tails of distribution are. The case $\nu = 2$ corresponds to normal distribution, when the second moment exists. To obtain at least weaker result for the case when the finite moment exists only for $r < 2\,(\nu < 2)$, we recall the results of [1].

Proposition 7 ([1]). *Let $s = 1, \quad \{\xi^i\}_{i=1}^N$, $N = 1, 2, \ldots$ be a sequence of independent random values corresponding to a heavy tailed distribution F with the shape parameter $\nu \in (1, 2)$ and let*

$$\sup_{t>0} t^\nu \, P\{\omega : |\xi| > t\} < \infty, \tag{9}$$

then

$$\lim_{\bar{M} \longrightarrow \infty} \sup_N P\{\omega : \frac{N}{N^{1/\nu}} \int_{-\infty}^{\infty} |F(z) - F^N(z)| > \bar{M}\} = 0. \tag{10}$$

3 Problem Analysis

To analyze the stability of the multiobjective stochastic problem (1) we define the sets $\mathcal{G}(F, X_F)$, $\bar{\mathcal{X}}(F, X_F)$, $\bar{\mathcal{G}}(F, X_F)$ and the function $\bar{g}(x, z, \lambda)$ by the relations

$$\begin{aligned}
\mathcal{G}(F, X_F) &= \{y \in R^l : y_j = \mathsf{E}_F g_j(x, \xi),\ j = 1, \ldots, l \quad \text{for some} \quad x \in X_F;\ y = (y_1, \ldots, y_l)\},\\
\bar{\mathcal{X}}(F, X_F) &= \{x \in X_F : x \quad \text{is a properly efficient point of the problem (1)}\},\\
\bar{\mathcal{G}}(F, X_F) &= \{y \in R^l : y_j = \mathsf{E}_F g_j(x, \xi),\ j = 1, \ldots, l \quad \text{for some} \quad x \in \bar{\mathcal{X}}(F, X_F)\}\\
\bar{g}(x, z, \lambda) &= \sum_{i=1}^{l} \lambda_i g_i(x, z), \quad x \in R^n,\ z \in R^s,\ \lambda = (\lambda_1, \ldots, \lambda_l),\ \lambda_i > 0,\ \sum_{i=1}^{l} \lambda_i = 1.
\end{aligned} \tag{11}$$

Evidently, if the following assumptions are fulfilled

B.1
- X is a convex set and, moreover, there exists $\varepsilon > 0$ such that $g_i(x, z)$, $i = 1, \ldots, s$ are for every $z \in R^s$ a convex functions on $X(\varepsilon)$,
- $g_i(x, z)$, $\imath = 1, \ldots, l$ are Lipschitz functions of $z \in R^s$ with the Lipschitz constant L (corresponding to $\mathcal{L}_1$ norm) not depending on x,

then $\bar{g}(x, z, \lambda)$ is a convex function on $X(\varepsilon)$ and, moreover, it is a Lipschitz function of z with the Lipschitz constant L not depending on x, λ. Consequently, according to to Proposition 3 we can obtain.

Proposition 8. *Let $P_F, P_G \in \mathcal{M}_1(R^s)$, X be a nonempty set. If B.1 is fulfilled, then*

$$|\mathsf{E}_F \bar{g}(x, \xi, \lambda) - \mathsf{E}_G \bar{g}(x, \xi, \lambda)| \le L \sum_{i=1}^{s} \int_{-\infty}^{+\infty} |F_i(z_i) - G_i(z_i)|dz_i, \quad x \in X,\ \lambda_i > 0,\ i = 1, \ldots, s,\ \sum_{i=1}^{s} \lambda_i = 1.$$

Proposition 9. *Let P_F, $P_G \in \mathcal{M}_1(R^s)$, X be a compact set. If B.1 is fulfilled, then*

$$\Delta_n[\mathcal{G}(F,\, X),\, \mathcal{G}(G,\, X)] \leq L \sum_{i=1}^{s} \int_{-\infty}^{+\infty} |F_i(z_i) - G_i(z_i)| dz_i,$$

where the symbol $\Delta n[\cdot,\, \cdot]$ is reserved for the Hausdorff distance of the subsets of R^n (for the definition of the Hausdorff distance see e.g. [14]).

Proposition 10. *Let X be a convex set. If $g_i(x,\, z)$, $i = 1,\, \ldots,\, l$ are strongly convex (with a parameter $\rho > 0$) function on X, then $\mathsf{E}_F \bar{g}(x,\, \xi, \lambda)$ is a strongly convex function on X with the parameter ρ.*

Proof. The assertion of Proposition 10 follows from Definition 3 and Relation (11). □

Employing the assertion of Proposition 1 we can investigate the relationship between the Problems (1), and (2).

4 Empirical Estimates

Theorem 11. *Let Assumptions B.1, A.2, and A.3 be fulfilled, $P_F \in \mathcal{M}_1(R^s)$, X be a compact set. Then*

$$P\{\omega : \Delta_n[\mathcal{G}(F,\, X),\, \mathcal{G}(F^N,\, X)] \longrightarrow_{N \longrightarrow \infty}\} = 1.$$

Proof. The assertion of Theorem 11 follows from Propositions 4, 9 and the relation (11). □

Theorem 12. *Let $t > 0$, $r > 0$, Assumptions B.1, A.2, A.3 be fulfilled. Let, moreover, ξ be a random vector with the components ξ_i, $i = 1,\, \ldots,\, s$ such that $\mathsf{E}_F |\xi_i|^r < \infty$. If constants $\beta,\, \gamma > 0$ fulfil the inequalities $0 < \beta + \gamma < 1/2$, $\gamma > 1/r$, $\beta + (1 - r)\gamma < 0$, then*

$$P\{\omega : N^{\beta}\, \Delta_n[\mathcal{G}(F,\, X),\, \mathcal{G}(F^N,\, X)] > t\} \longrightarrow_{N \longrightarrow \infty}\} = 0.$$

If, moreover, $g_i(x,\, z)$, $i = 1,\, \ldots,\, l$ are strongly convex with a parameter $\rho > 0$ function on X, then also

$$P\{\omega : N^{\beta} \Delta_n[\bar{\mathcal{X}}(F,\, X)),\, \bar{\mathcal{X}}(F^N,\, X)]^2 > t\} \longrightarrow_{N \longrightarrow \infty} 0.$$

Proof. First assertion of Theorem 12 follows from Propositions 6 and 9. The second assertion follows from the first one and from Proposition 2, 10 and the relation 10. □

Theorem 13. *Let Assumptions B.1, A.2 and A.3 be fulfilled, $P_F \in \mathcal{M}_1(R^s)$, $\bar{M} > 0$, X be a compact set. If one–dimensional components ξ_i, $i = 1,\, \ldots,\, s$ of the random vector ξ have distribution functions F_i with tails parameter $\nu_i \in (1,\, 2)$ fulfilling the relations*

$$\sup_{t>0} t^{\nu_i} P_F\{\omega : |\xi_i| > t\} < \infty, \quad i = 1,\, \ldots,\, s,$$

then

$$\lim_{\bar{M} \longrightarrow \infty} \sup_N P\{\omega : \frac{N}{N^{1/\nu}} \Delta_n[\mathcal{G}(F,\, X)),\, \mathcal{G}(F^N,\, X))], > \bar{M}\} = 0 \quad \text{with} \quad \nu = \min(\nu_1,\, \ldots,\, \nu_s\}.$$

If, moreover, $g_i(x,\, z)$, $i = 1,\, \ldots,\, l$ are strongly convex with a parameter $\rho > 0$ function on X, then also

$$\lim_{\bar{M} \longrightarrow \infty} \sup_N \{\omega : \frac{N}{N^{1/\nu}} \Delta_n[\bar{\mathcal{X}}(F,\, X)),\, \bar{\mathcal{X}}(F^N,\, X)]^2 > \bar{M}\} = 0 \quad \text{with} \quad \nu = \min(\nu_1,\, \ldots,\, \nu_s\}.$$

Proof. The assertion of Theorem 13 follows from the assertion of Propositions 2, 7 and 10. □

Remark 1. Let us assume that Assumptions of Theorem 13 are fulfilled and $\beta(\nu) := 1 - 1/\nu$. Then $\beta(\nu)$ is an increasing function of ν and holds up to

$$\lim_{\nu \longrightarrow 1^+} \beta(\nu) = 0, \quad \lim_{\nu \longrightarrow 2^-} \beta(\nu) = \frac{1}{2},$$

The assertions of Theorems 11, 12 and 13 are introduced (under the approach of properly efficient points and their functions mapping), however, since the set of properly efficient points is dense in the set of efficient points our results are not much restricted (for more details see e. g. [3]).

5 Conclusion

The paper deals with multiobjective stochastic programming problems, especially with a relationship between characteristics of these problems corresponding to complete knowledge of the probability measure and them determined on the data base. We have restricted ourselves to investigate the characteristics $\mathcal{G}(F, X)$, $\bar{\mathcal{X}}(F, X))$ and $X_F = X$, generally. Evidently the presented results can be generalized to the characteristic $\bar{\mathcal{G}}(F, X_F)$ and (employing Relation (6)) some type of constraints set depending on the probability measure (see the corresponding results achieved for one objective case [9]). However more detailed investigation in this direction is beyond the scope of this paper.

Acknowledgment This work was supported by the Grant Agency of the Czech Republic under grants 13-14445S and P402/11/0150.

References

[1] Barrio, E., Giné, E., and Matrán, C.: Central limit theorems for the Wassersten distance between the empirical and the true distributions. *Annals of Probability* **27** (1999), 2, 1009–1071.

[2] Dvoretski, A., Kiefer, J. and Wolfowitz, J.: J. and Wolfowitz, J.: Asymptotic minimaxcharacter of esample distribution function and the classical. *Ann. Math. Statist.* **27** (1956),642–669.

[3] M. Ehrgott: Multicriteria Optimization. Springer, Berlin, 2005.

[4] A. M. Geoffrion: Proper efficiency and the theory of vector maximization. J. Math. Annal. Appl. **22** (1968), 3, 618–630.

[5] Houda, M., and Kaňková, V.: Empirical estimates in economic and financial optimization problems. *Bulletin of the Czech Econometric Society* **19** (2012), 29, 50–69.

[6] V. Kaňková: Stability in stochastic programming – the case of unknown location parameter. Kybernetika **29** (1993), 1, 97–112.

[7] Kaňková, V. and Houda M.: Empirical estimates in stochastic programming. In: *Proceedings of Prague Stochastics 2006* (M. Hušková and M. Janžura, eds.), MATFYZPRESS, Prague 2006, 426–436.

[8] V. Kaňková: Empirical estimates in stochastic optimization via distribution tails. *Kybernetika* **46** (2010), 3, 459–471.

[9] Kaňková, V.: Empirical estimates in optimization problems: survey with special regard to heavy tails and dependent samples Bulletin of the Czech Econometric Society **19** (2012), 30, 92–111.

[10] Klebanov, L.B.: *Heavy Tailed Distributions.* MATFYZPRESS, Prague, 2003.

[11] Kozubowski, T.J., Panorska A. K., and Rachev, S. T.: Statistical issues in modelling stable portfolios. In: *Handbook of Heavy Tailed Distributions in Finance* (S.T. Rachev, ed.), Elsevier, Amsterdam 2003, 134–168.

[12] Meerschaert, M.M., and Scheffler, H.P.: Portfolio modelling with heavy tailed random vectors. In: *Handbook of Heavy Tailed Distributions in Finance* (S.T. Rachev, ed.), Elsevier, Amsterdam 2003, 595–640.

[13] Rachev, S.T., and Mitting, S.: *Stable Paretian Models in Finance (Series in Financial Economics and Quantitative Analysis)* John Wiley & Sons, Chichester, 2000.

[14] R. T. Rockafellar, and R. J. B. Wets: Variational Analysis. Springer, Berlin, 1983.

[15] G. R. Shorack and J. A. Wellner: Empirical Processes with Applications to Statistics. John Wiley & Sons, New York, 1986.

Performance of simple heuristic algorithms for the clustering of countries with respect to food supply

Nikola Kaspříková[1]

Abstract. Thanks to increasing availability of the computation power, some computationally intensive heuristic optimization techniques have become popular. These techniques include differential evolution algorithm or threshold accepting method. Results of application of simple general purpose heuristic methods are reported and the impact of values of control parameters on the quality of the solution is discussed. The problem addressed is cluster analysis of countries with respect to the food supply (various items per capita per day). The average silhouette width is used as the objective function for evaluation of the solution and the results are compared with the solution obtained with the commonly used partitioning algorithms. The results suggest that both the threshold accepting method and the differential evolution method are suitable for solving the optimal partitioning problem. These methods easily allow to supply the objective function preferred by the user and reach the solution rather quickly. The values of the control parameters did not exhibit any significant impact on the quality of the solution. The quality of the resulting classification was a little bit higher when using the differential evolution method than when using the threshold accepting method. And both heuristic methods slightly outperformed the standard implementations of the clustering methods (k-means, partitioning around medoids) available in the statistical software packages. The final classification has two groups, where the smaller cluster seems to include the countries usually treated as developing.

Keywords: clustering, silhouette width, threshold accepting, differential evolution.

JEL classification: C61
AMS classification: 90C26

1 Introduction

Many economic optimization problems are difficult to solve analytically (as there may be multiple local optima or there are other undesirable properties) and one has to resort to heuristic techniques, which may reach some suitable solution, even though not necessarily always the optimal one, within reasonable time. Heuristic optimization methods may include trajectory-based methods (there is just one candidate solution, which is modified in each iteration) or population-based methods (such methods work with a population of solutions in every iteration, which is often called a generation), see e.g. [1]. Evolutionary algorithms include the particle swarm optimization or the differential evolution method. Such computationally intensive methods have recently become popular thanks to the rise of the available computation power. Values of the control parameters of these general purpose algorithms often have to be tuned for the particular problem at hand (as there does not seem to be the rule for setting the optimal values of these parameters which would be suitable for all the objective functions). Regarding the differential evolution method, which seems to be a particularly efficient tool for solving difficult optimization tasks, the control parameters of the algorithm include the size of the population, the probability of crossover and the step size. The problem of finding optimal partitioning of multivariate data with respect to particular objective function with the use of heuristic algorithms is addressed in this paper. The average silhouette width is used as the objective function and the results of the application of a couple of simple general purpose heuristic methods are reported.

[1]University of Economics in Prague, Department of Mathematics, Nám. W. Churchilla 4, 130 67 Praha 3, data@tulipany.cz

The organization of the paper is as follows: after a brief introduction to the field of unsupervised classification, the principles of selected heuristic optimization methods are recalled and the description of the data set used in the analysis is provided in the Material and Methods section. Then the results of the analysis are reported.

2 Material and Methods

2.1 Data Description

The dataset used in the analysis is on food supply in particular countries[1]. The supply (in g per capita per day) of 10 food categories (i. e. there are ten variables in the data set) of 175 countries is analyzed with the aim to learn if there is some natural grouping of countries regarding the food supply. Clearly the variables have different scale and it does not make much sense using such variables, so the data preparation step included transforming the original values to ranks.

2.2 Clustering

The cluster analysis, or the unsupervised classification task, aims at the discovering of the best (whatever it means) groupings of the cases included in the analysis in such a way, that the cases in the same cluster are rather similar, whereas the cases from the different clusters are not.

There exist many algorithms for solving this problem, one commonly used clustering of such methods is given in [5]:

- hierarchical methods, which produce a set of clusterings
  - agglomerative hierarchical methods
    - the methods differ mostly in the way of calculating group-group differences from point-point dissimilarities
    - are computationally easy
  - divisive hierarchical methods
    - are computationally difficult
    - available methods are usually monothetic (i. e. the split is just on one variable at each stage)
- optimal partitioning methods
  - the number of clusters is fixed
  - needs some initial partitioning
  - there are various optimization criteria
  - include k-means, partitioning around medoids, CLARA, fuzzy cluster analysis or model-based methods

Some clustering methods have to operate on the original variables values, other methods can build the solution using just the distance matrix. The results of the clustering may be impacted by the choice of the distance measure, we use the usual 2 norm in this analysis.

2.3 The Objective Function

There exist several reasonable objective functions for evaluation of the quality of the clustering model. One of the possibilities is to choose the average silhouette width. We briefly recall how this objective function is defined, for more details see the original sources, such as [3].

[1] The data on food supply can be obtained at http://www.faostat.org.

The average silhouette width may be used for evaluation of the quality of the resulting clustering solution (and it may also help when deciding about the most suitable number of clusters to work with) when the partitioning clustering methods are used.

For every particular case i one defines the average dissimilarity of the case i and all the remaining cases which belong to the same cluster as i (let's denote this cluster as A and the number of cases assigned in this cluster as $|A|$):

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j).$$

Then for every cluster C other than the cluster A the average dissimilarity between the case i and the cluster C is calculated as

$$d(i, C) = \frac{1}{|C|} \sum_{j \in C} d(i, j).$$

Let's denote

$$b(i) = \min_{C \neq A} d(i, C).$$

The *silhouette width* of the case i is

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}.$$

Obviously it is a number from interval $\langle -1, 1 \rangle$ and it can be interpreted as follows:

$s(i) \approx 1 \quad \rightarrow \quad$ the case is well assigned in the cluster considered,
$s(i) \approx 0 \quad \rightarrow \quad$ the case is somewhere between clusters,
$s(i) < 0 \quad \rightarrow \quad$ the case is probably assigned to wrong cluster.

The average silhouette width is then the silhouette width averaged over all the cases in the data set. The values near 0 suggest that the classification structure is weak and may be just artificial, values near 1 suggest the presence of a strong classification structure.

For the presentation of the quality assessment of the resulting clustering, we use the silhouette plot, which is available in the cluster package (see [3]) for the R computing environment (see [2]).

2.4 Threshold Accepting and Differential Evolution Methods

The threshold accepting algorithm is a trajectory-based optimization method, for its principles see [1]. The function for obtaining the neighbour (i. e. next candidate) solution has to be suitable for the particular task being solved and it has to be supplied by the user. The control parameters of the algorithm include the threshold value (usually some sequence with values decreasing with the iteration number).

The differential evolution algorithm, introduced in [4], is a population-based method and it includes several control parameters - the step size F, the probability of the crossover CR, the number of generations and the population size. For details see the original paper [4].

Both algorithms require some suitable choice of the number of iterations (or generations), which may be conveniently checked by observing the plot of the objective function value in response to the iteration number.

We use the implementation of the threshold accepting and the differential evolution methods available in NMOF package (see [1]) in the R computing environment (see [2]). Since the objective function is minimized in this implementation (as is the usual approach), the average silhouette width was always multiplied by -1.

3 Results

The optimal clustering with up to 4 clusters has been searched for. The solutions with just single cluster have been penalized through setting the objective function value to 0 for this case. The use of the heuristic methods allows automatic choice of the most suitable number of groups (2, 3 or 4).

Both the application of the threshold accepting method and the differential evolution method resulted in two clusters solutions. The values of the control parameters for the algorithms did not have any strong

impact on the results. The average silhouette width for the final solution obtained by the threshold accepting with 10000 iterations is 0.3033802 (see Figure 1 which shows how the objective function value changes with the iteration number) and for the differential evolution with F=0.5, CR=0.9, population size 40 and 200 generations it is 0.3033938 (see Figure 2). So the differential evolution did a slightly better job. The partitioning obtained with the differential evolution (let's call it the final clustering) gave one smaller, more clearly defined cluster and one larger cluster which does not seem to have a clear profile - see Figure 3 for the silhouette plot. When inspecting the assignments of particular countries to clusters, the smaller cluster seems to include the countries usually perceived as the developing ones. The prediction model (classification tree) was built to get some insight about the profiles of the clusters and a single predictive variable, which was the supply of Milk, was included in the resulting tree model. The countries with low rank in Milk supply got to the smaller cluster.

The final partitioning is superior to the solution obtained with the standard implementation of the k-means method (in the stats package in [2]), which gives the average silhouette width for two clusters solution 0.3001925, 0.20428 for 3 clusters and 0.1659714 for 4 clusters. Similarly for the partitioning around medoids method (in the cluster package in [2]), which gives 0.3010591, 0.1986875 and 0.2111096 respectively.

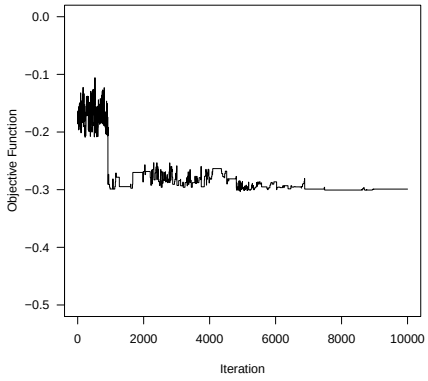


Figure 1 Threshold accepting

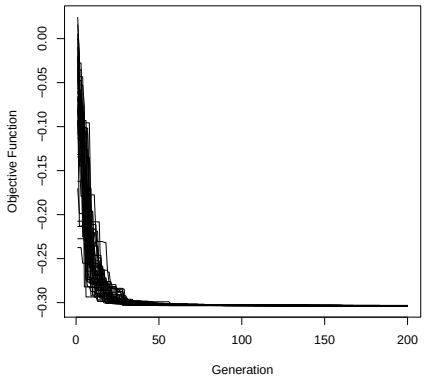


Figure 2 Differential evolution

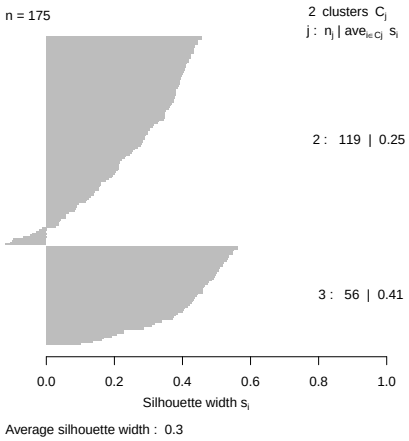


Figure 3 Silhouette plot for the resulting solution

4 Conclusions

It was shown that both the threshold accepting method and the differential evolution method are suitable for solving the optimal partitioning problem. These methods easily allow to supply the objective function preferred by the user and reach the solution rather quickly. The procedures can also automatically determine the best number of clusters, which is also an advantage. The values of the control parameters of the optimization algorithms did not exhibit any great impact on the quality of the solution. The quality of the resulting classification was slightly better when using the differential evolution method than when using the threshold accepting method. And in both cases the result obtained is slightly better than if the standard implementations of the clustering methods available in the software packages are used.

References

[1] Gilli, M., Maringer, D. and Schumann, E.: *Numerical Methods and Optimization in Finance*. Academic Press, 2011.

[2] R Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012.

[3] Rousseeuw, P., Struyf, A. and Hubert, M.: Clustering in an Object-Oriented Environment, *Journal of Statistical Software* **1(4)** (1996), 1–30.

[4] Storn, R. and Price, K.: Differential Evolution A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces, *Journal of Global Optimization* **11(4)** (1997), 341 – 359.

[5] Venables, W. N. and Ripley, B. D.: *Modern Applied Statistics with S*. Springer, 2002.

Consistency versus transitivity in pair-wise comparison matrices

Zuzana Kiszová[1], Jaroslav Ramík[2]

Abstract: Pair-wise comparison matrix is an efficient instrument often used in decision making analysis. Usually, a decision-maker is capable of creating pair-wise comparisons of decision elements forming a pair-wise comparison matrix. Either in multiplicative or in additive approaches there may appear some "incompatibilities" in the pair-wise comparison matrix – inconsistency and/or intransitivity. These properties do not have to occur simultaneously. Inconsistency and/or intransitivity in multiplicative and additive pair-wise comparisons are detected and measured by using consistency index/ratio, transitivity index/ratio, respectively. In this contribution we propose the way how to transform additive (fuzzy) pair-wise comparison matrix into multiplicative one for which the consistency index/ratio and transitivity index/ratio are defined. An example illustrating the application of the above mentioned approaches is supplemented.

Keywords: AHP, decision analysis, fuzzy sets and systems.

JEL Classification: C44

AMS Classification: 90B50, 90C29, 91B06

1 Introduction

Pair-wise comparison matrix is an instrument often used in multcriteria decision making in effort to rank criteria/alternatives or select optimal one. The essence of pair-wise comparisons is to determine which of two objects is more preferred by the decision-maker. There are two approaches to pair-wise comparisons – multiplicative and additive. For both these approaches some properties of pair-wise comparison matrices are investigated in this contribution: reciprocity, consistency and transitivity. If these properties are not satisfied, i.e. the acceptable threshold is exceeded, the decision making process gives inappropriate conclusions. In that case it is necessary to reassess the preferences in pair-wise comparisons and change the pair-wise comparison matrix.

The aim of this contribution is to describe conditions whose satisfaction ensures consistency or transitivity of pair-wise comparison matrices and define how to measure the extent of potential inconsistency/intransitivity. It is necessary to mention that these two incompatibilities may appear independently of each other. This fact is demonstrated by illustrative examples.

2 Pair-wise comparison matrix

Pair-wise comparison matrix is a tool of multicriteria decision making (MCDM). It is an irreducible nonnegative $n \times n$ matrix which entries are results of pair-wise comparisons. To complete such a matrix it is enough to provide $\binom{n}{2} = \frac{n(n-1)}{2}$ pair-wise comparisons, the remaining entries are ensured by *multiplicative reciprocity* property:

$$a_{ij} \cdot a_{ji} = 1, \text{ for all } i, j = 1,2, \dots, n, \tag{1}$$

where a_{ij} are entries of pair-wise comparison matrix $A = \{a_{ij}\}$. A pair-wise comparison matrix $B = \{b_{ij}\}$ is called *additive reciprocal* (*a-reciprocal*), if

$$b_{ij} + b_{ji} = 1, \text{ for all } i, j = 1,2, \dots, n, \tag{2}$$

where b_{ij} are entries of pair-wise comparison matrix $B = \{b_{ij}\}$.

[1] Silesian University in Opava, School of Business Administration in Karvina, Department of Mathematical Methods in Economics, kiszova@opf.slu.cz.

[2] Silesian University in Opava, School of Business Administration in Karvina, Department of Mathematical Methods in Economics, ramik@opf.slu.cz.

There are two properties of pair-wise comparison matrices we are concerned in this contribution. These are consistency and transitivity properties.

3 Multiplicative pair-wise comparisons

In multiplicative approach of pair-wise comparisons there are elements $x_1, x_2, \dots, x_n$ to be compared in pairs to express preference intensity on the scale presented by the closed interval $[1/\sigma\,;\sigma]$, where $\sigma > 1$. In Saaty's approach [3], the scale is $[1/9\,;9]$. The i-th element is compared to the j-th element with respect to a particular property. If element i is more preferred than element j, the result a_{ij} satisfies the inequality $1 < a_{ij} \leq \sigma$. If element i is less preferred than element j, the entry a_{ij} satisfies the condition $1/\sigma \leq a_{ij} < 1$. If these two elements are equally preferred, then $a_{ij} = 1$. The equality $a_{ii} = 1$ for all i is obvious.

Usually, a decision-maker is able to make consistent pair-wise comparisons. *Multiplicative consistency* (*m-consistency*) expresses how much the pair-wise comparisons correspond to each other. A positive $n \times n$ matrix $A = \{a_{ij}\}$ is *m-consistent,* if

$$a_{ij} = a_{ik} \cdot a_{kj} \text{ for all } i, j, k. \tag{3}$$

Multiplicative transitivity (*m*-transitivity) is associated with *m*-consistency as follows. A positive $n \times n$ matrix $A = \{a_{ij}\}$ is *m-transitive,* if

$$\frac{a_{ij}}{a_{ji}} = \frac{a_{ik}}{a_{ki}} \cdot \frac{a_{kj}}{a_{jk}} \text{ for all } i, j, k, \tag{4}$$

where ratio $\frac{a_{ij}}{a_{ji}}$ denotes the relative efficiency of comparison of x_i over x_j.

Notice that if *A* is m-consistent then *A* is m-transitive. Moreover, if $A = \{a_{ij}\}$ is m-reciprocal, then *A* is m-transitive if and only if *A* is m-consistent. Here, we shall investigate m-transitive matrices with elements from the unit interval [0; 1] being not m-reciprocal, however, a-reciprocal.

4 Additive pair-wise comparisons

The multiplicative approach is not always the most acceptable way how to do pair-wise comparisons. Sometimes there is a more natural procedure how to do it – the additive approach. The principle is simple: 100% share of the property in question is divided into two parts, the first part, b_{ij}, is assigned to the i-th object and the second one, b_{ij}, to the j-th object. Moreover, b_{ij} belongs to the closed interval $[0; 1]$ for all i, j.

Let $B = \{b_{ij}\}$ be an $n \times n$ a-reciprocal matrix with $0 \leq b_{ij} \leq 1$ for all i, j, k.

B is called *additive transitive* (*a-transitive*) [4], if

$$b_{ij} - 0.5 = (b_{ik} - 0.5) + (b_{kj} - 0.5) \text{ for all } i, j, k. \tag{5}$$

If *B* is *a*-reciprocal and *m*-transitive, then *B* is called *additive consistent* (*a-consistent*).

There exist some relationships between m-reciprocal and a-reciprocal pair-wise comparison matrices. In order to show them, we define function $\phi(t)$ which transforms a-reciprocal matrix into an m-reciprocal one, see [2]:

$$\phi(t) = \frac{t}{1-t}, \text{ for } t \in (0,1). \tag{6}$$

Proposition 1. *Let $B = \{b_{ij}\}$ be an a-reciprocal $n \times n$ matrix with $0 < b_{ij} < 1$ for all i and j. If $B = \{b_{ij}\}$ is m-transitive then $A = \{a_{ij}\} = \{\phi(b_{ij})\}$ is m-consistent.*

According to Chiclana et al. [1] we shall use transformation function $\varphi_\sigma(t)$ for given $\sigma > 1$ as follows, see [2]:

$$\varphi_\sigma(t) = \sigma^{2t-1}, \text{ for } t \in [0,1]. \tag{7}$$

Proposition 2. *Let $B = \{b_{ij}\}$ be an $n \times n$ matrix with $0 \leq b_{ij} \leq 1$ for all i and j. If $B = \{b_{ij}\}$ is a-transitive then $A = \{a_{ij}\} = \{\varphi_\sigma(b_{ij})\}$ is m-consistent.*

Perfectly consistent or transitive pair-wise comparison matrices occur rarely, therefore, we need some tools for measuring of inconsistency/intransitivity. In practical situations we also need a threshold value such that if the extent of inconsistency/intransitivity of a matrix is smaller than this threshold then the matrix is considered consistent/transitive.

5 Measuring (in)consistency and (in)transitivity

If for some positive $n \times n$ matrix $A = \{a_{ij}\}$ and for some $i, j, k = 1,2, \dots, n$, multiplicative consistency condition (3) does not hold, then A is said to be *multiplicative-inconsistent* (*m-inconsistent*). If for some $n \times n$ matrix $B = \{b_{ij}\}$ with $0 \leq b_{ij} \leq 1$ for all i and j, and for some $i, j, k = 1,2, \dots, n$, (5) does not hold, then B is said to be *additive-intransitive* (*a-intransitive*) [2].

Existence of a positive *spectral radius* $\rho(A)$ and the corresponding positive vector w satisfying the characteristic equation $Aw = \rho(A)w$ follows from Perron-Frobenius theory for any irreducible nonnegative matrix A [2]. Vector $w = (w_1, w_2, \dots, w_n)$, $w_i > 0$ for all $i = 1,2, \dots, n$, is called the *principal eigenvector of* A. When normalized, i.e. $\sum_{i=1}^{n} w_i = 1$, element w_i of vector w represents the relative importance of alternative $x_i, i = 1,2, \dots n$. Then it is possible to rank all alternatives x_i and/or choose the optimal one(s) in this way. Then the spectral radius $\rho(A)$ is called the *principal eigenvalue of matrix A* and the principal eigenvector of A is called the *priority vector of A*.

The extent of m-consistency/m-inconsistency of m-reciprocal $n \times n$ matrix A is measured by the *m-consistency index* $I_{mc}(A)$ [3]:

$$I_{mc}(A) = \frac{\rho(A) - n}{n - 1}. \tag{8}$$

An *m*-reciprocal pair-wise comparison matrix A is *m*-consistent if and only if $I_{mc}(A) = 0$, i.e. $\rho(A) = n$.

By *m-consistency ratio* $CR_{mc}(A)$ we measure the inconsistency independently of the dimension n of A:

$$CR_{mc}(A) = \frac{I_{mc}(A)}{R_{mc}(n)}, \tag{9}$$

where $R_{mc}(n)$ is the *random consistency index* – the mean value of $I_{mc}(A)$ where the elements a_{ij} of A – *m*-reciprocal matrices of dimension n – are uniformly distributed random variables in the interval $[1/9; 9]$. An *m*-reciprocal pair-wise comparison matrix A is considered to be sufficiently *m*-consistent if the *m*-consistency ratio does not exceed the threshold value 0.1. In other words, A is considered to be sufficiently *m*-consistent if the *m*-consistency index $I_{mc}(A)$ is less than one tenth of the random consistency index $R_{mc}(n)$. A more detailed explanation and interpretation of this heuristic threshold can be found e.g. in [3].

By Propositions 1 and 2 we can measure the extent of inconsistency and intransitivity of a-reciprocal matrices.

Let $B = \{b_{ij}\}$ be an a-reciprocal $n \times n$ matrix with $0 < b_{ij} < 1$ for all i and j. The *a-consistency index* $I_{ac}(B)$ of B is defined as:

$$I_{ac}(B) = I_{mc}(A), \text{ where } A = \{\phi(b_{ij})\}. \tag{10}$$

Such matrix B is considered to be *a-consistent* if and only if $I_{ac}(B) = 0$.

Moreover, we define the *a-consistency ratio* $CR_{ac}(B)$ for measuring a-consistency independently on dimension n of the matrix as:

$$CR_{ac}(B) = \frac{I_{ac}(B)}{R_{mc}(n)}. \tag{11}$$

A-reciprocal pair-wise comparison matrix B is regarded to be sufficiently a-consistent if the a-consistency ratio does not exceed the value 0.1.

The corresponding priority vector $w^{ac} = (w_1^{ac}, w_2^{ac}, \dots, w_n^{ac})$ determining the ranking of alternatives $x_i, i = 1,2, \dots n$, is given by the characteristic equation $\phi(B)w^{ac} = \rho(\phi(B))w^{ac}$.

Let $B = \{b_{ij}\}$ be an a-reciprocal $n \times n$ matrix with $0 < b_{ij} < 1$ for all i and j. The *a-transitivity index* $I_{at}^{\sigma}(B)$ of B is defined as:

$$I_{at}^{\sigma}(B) = I_{mc}(A_{\sigma}), \text{ where } A_{\sigma} = \{\varphi_{\sigma}(b_{ij})\}, \tag{12}$$

where $\sigma > 1$ is given value.

Matrix B is *a-transitive* if and only if $I_{at}^{\sigma}(B) = 0$.

Similarly to m-consistency and a-consistency it is possible to measure a-transitivity independently of dimension n of the matrix. The *a-transitivity ratio* $CR_{at}^{\sigma}(B)$ is defined as:

$$CR_{at}^{\sigma}(B) = \frac{I_{at}^{\sigma}(B)}{R_{mc}(n)}. \tag{13}$$

If the a-transitivity ratio does not exceed 0.1, then the matrix B is considered to be sufficiently a-transitive, otherwise, a-intransitive is unacceptable.

The ranking of alternatives $x_i, i = 1,2, \dots n$ is given by the priority vector $w^{at} = (w_1^{at}, w_2^{at}, \dots, w_n^{at})$ which is calculated from the characteristic equation $\varphi_{\sigma}(B)w^{at} = \rho(\varphi_{\sigma}(B))w^{at}$.

6 Example

Let $X = \{x_1, x_2, x_3, x_4\}$ be a set of four alternatives. The alternatives should be ordered from the best to the worst. The preferences on X are given by matrix $B = \{b_{ij}\}$,

$$B = \begin{bmatrix} 0.5 & 0.1 & 0.25 & 0.15 \\ 0.9 & 0.5 & 0.9 & 0.85 \\ 0.75 & 0.1 & 0.5 & 0.15 \\ 0.85 & 0.15 & 0.85 & 0.5 \end{bmatrix}.$$

According to (2) and (3) matrix B is a-reciprocal but m-inconsistent as $b_{13} \cdot b_{32} \cdot b_{21} = 0.02 \neq 0.07 = b_{12} \cdot b_{23} \cdot b_{31}$. Applying (5) we find out that B is a-intransitive: $b_{13} - 0.5 = -0.25 \neq 0 = (b_{12} - 0.5) + (b_{23} - 0.5)$. We have to determine if the extent of inconsistency and intransitivity is unacceptable. Considering $\sigma = 9$ and $R_{mc}(4) = 0{,}89$ [3] we calculate matrices C and D according to (6) and (7):

$$C = \{\phi(b_{ij})\} = \begin{bmatrix} 1 & 0.11 & 0.33 & 0.18 \\ 9 & 1 & 9 & 5.67 \\ 3 & 0.11 & 1 & 0.18 \\ 5.67 & 0.18 & 5.67 & 1 \end{bmatrix},$$

$$D = \{\varphi_9(b_{ij})\} = \begin{bmatrix} 1 & 0.17 & 0.33 & 0.21 \\ 5.80 & 1 & 5.80 & 4.66 \\ 3 & 0.17 & 1 & 0.21 \\ 4.66 & 0.21 & 4.66 & 1 \end{bmatrix}.$$

We compute $\rho(C) = 4.36$, $\rho(D) = 4.38$. According to (8), (11) and (13) we get $CR_{ac}(B) = 0.14 > 0.1$ with the priority vector $w^{ac} = (0.04; 0.67; 0.07; 0.22)$ ranking alternatives $x_2 > x_4 > x_3 > x_1$. Similarly, $CR_{at}^{9}(B) = 0.14 > 0.1$ with the priority vector $w^{at} = (0.06; 0.60; 0.10; 0.25)$, which gives the same ranking of alternatives $x_2 > x_4 > x_3 > x_1$.

Obviously, matrix B is neither sufficiently *a-transitive* nor sufficiently *a-consistent* as the a-transitivity ratio and a-consistency ratio exceeds the value of 0.1.

The decision-maker reconsiders some of his/her preferences in pair-wise comparison matrix B and obtains new pair-wise comparison matrix $E = \{e_{ij}\}$:

$$E = \begin{bmatrix} 0.5 & 0.05 & 0.25 & 0.15 \\ 0.95 & 0.5 & 0.95 & 0.85 \\ 0.75 & 0.05 & 0.5 & 0.15 \\ 0.85 & 0.15 & 0.85 & 0.5 \end{bmatrix}.$$

By (2) and (3) we find out that E is a-reciprocal and m-inconsistent: $e_{14} \cdot e_{42} \cdot e_{21} = 0.02 \neq 0.04 = e_{12} \cdot e_{24} \cdot e_{41}$. From (5) it is obvious that E is also a-intransitive: $e_{14} - 0.5 = -0.35 \neq -0.10 = (e_{12} - 0.5) + (e_{24} - 0.5)$. We have to examine if the extent of inconsistency and intransitivity is unsatisfactory. Considering $\sigma = 9$ and $R_{mc}(4) = 0{,}89$ [3] we compute matrices F and G:

$$F = \{\phi(e_{ij})\} = \begin{bmatrix} 1 & 0.05 & 0.33 & 0.18 \\ 19 & 1 & 19 & 5.67 \\ 3 & 0.05 & 1 & 0.18 \\ 5.67 & 0.18 & 5.67 & 1 \end{bmatrix},$$

$$G = \{\varphi_9(e_{ij})\} = \begin{bmatrix} 1 & 0.14 & 0.33 & 0.21 \\ 7.22 & 1 & 7.22 & 4.66 \\ 3 & 0.14 & 1 & 0.21 \\ 4.66 & 0.21 & 4.66 & 1 \end{bmatrix}.$$

We calculate $\rho(F) = 4.19$ and $\rho(G) = 4.31$. Applying (8), (11) and (13) we obtain $CR_{ac}(E) = 0.07 < 0.1$ with the priority vector $w^{ac} = (0.03;\ 0.75;\ 0.05;\ 0.17)$ giving the ranking of alternatives $x_2 > x_4 > x_3 > x_1$. By a similar way, $CR_{at}^{9}(E) = 0.12 > 0.1$ with the priority vector $w^{at} = (0.05;\ 0.63;\ 0.09;\ 0.23)$ and final ranking of alternatives $x_2 > x_4 > x_3 > x_1$.

Matrix *E* is sufficiently *a-consistent* but unsatisfactory *a-transitive*.

The decision-maker considers more changes in his/her preferences and creates new pair-wise comparison matrix $H = \{h_{ij}\}$:

$$H = \begin{bmatrix} 0.5 & 0.05 & 0.25 & 0.15 \\ 0.95 & 0.5 & 0.6 & 0.7 \\ 0.75 & 0.4 & 0.5 & 0.2 \\ 0.85 & 0.3 & 0.8 & 0.5 \end{bmatrix}.$$

Applying (2) it becomes obvious that *H* is a-reciprocal. By (3) we find out that *H* is m-inconsistent: $h_{13} \cdot h_{32} \cdot h_{21} = 0.10 \neq 0.02 = h_{12} \cdot h_{23} \cdot h_{31}$. According to (5) matrix *H* is a-intransitive: $h_{13} - 0.5 = -0.25 \neq -0.05 = (h_{14} - 0.5) + (h_{43} - 0.5)$. Using (6) and (7) and considering $\sigma = 9$ we are able to find out whether the m-inconsistency and a-intransitivity is sufficient:

$$I = \{\phi(h_{ij})\} = \begin{bmatrix} 1 & 0.05 & 0.33 & 0.18 \\ 19 & 1 & 1.50 & 2.33 \\ 3 & 0.67 & 1 & 0.25 \\ 5.67 & 0.43 & 4 & 1 \end{bmatrix},$$

$$J = \{\varphi_9(h_{ij})\} = \begin{bmatrix} 1 & 0.14 & 0.33 & 0.21 \\ 7.22 & 1 & 1.55 & 2.41 \\ 3 & 0.64 & 1 & 0.27 \\ 4.66 & 0.42 & 3.74 & 1 \end{bmatrix}.$$

We compute $\rho(I) = 4.40$ and $\rho(J) = 4.28$. Using (8), (11) and (13) are got $CR_{ac}(H) = 0.15 > 0.1$ with the priority vector $w^{ac} = (0.04;\ 0.50;\ 0.15;\ 0.31)$ which ranks alternatives $x_2 > x_4 > x_3 > x_1$. Hence, $CR_{at}^{9}(H) = 0.10 \leq 0.1$, with the priority vector $w^{at} = (0.06;\ 0.45;\ 0.17;\ 0.33)$ and final ranking $x_2 > x_4 > x_3 > x_1$.

Matrix *H* is sufficiently *a-transitive* but unsatisfactory *a-consistent*.

The decision-maker once again reconsiders his/her pair-wise comparisons and forms the following pair-wise comparison matrix $K = \{k_{ij}\}$:

$$K = \begin{bmatrix} 0.5 & 0.1 & 0.25 & 0.15 \\ 0.9 & 0.5 & 0.65 & 0.7 \\ 0.75 & 0.35 & 0.5 & 0.2 \\ 0.85 & 0.3 & 0.8 & 0.5 \end{bmatrix}.$$

According to (2) matrix *K* is a-reciprocal. Using (3) it becomes evident matrix *K* is a-inconsistent: $k_{13} \cdot k_{34} \cdot k_{41} = 0.04 \neq 0.09 = k_{14} \cdot k_{43} \cdot k_{31}$. By (5) matrix *K* is a-intransitive: $k_{14} - 0.5 = -0.35 \neq -0.55 = (k_{13} - 0.5) + (k_{34} - 0.5)$. According to (6) and (7) we calculate matrices *L* and *M* considering $\sigma = 9$ and $R_{mc}(4) = 0{,}89$ [3] as

$$L = \{\phi(k_{ij})\} = \begin{bmatrix} 1 & 0.11 & 0.33 & 0.18 \\ 9 & 1 & 1.86 & 2.33 \\ 3 & 0.54 & 1 & 0.25 \\ 5.67 & 0.43 & 4 & 1 \end{bmatrix},$$

$$M = \{\varphi_9(k_{ij})\} = \begin{bmatrix} 1 & 0.17 & 0.33 & 0.21 \\ 5.80 & 1 & 1.93 & 2.41 \\ 3 & 0.52 & 1 & 0.27 \\ 4.66 & 0.42 & 3.74 & 1 \end{bmatrix}.$$

We calculate $\rho(L) = 4.23$ and $\rho(M) = 4.23$. By (8), (11) and (13) we obtain $CR_{ac}(K) = 0.09 < 0.1$ with the priority vector $w^{ac} = (0.05;\ 0.47;\ 0.15;\ 0.33)$, hence the final ranking of alternatives is $x_2 > x_4 > x_3 > x_1$. By the same way, $CR^9_{at}(K) = 0.09 < 0.1$ with the priority vector $w^{at} = (0.06;\ 0.45;\ 0.16;\ 0.33)$ giving again the ranking of alternatives $x_2 > x_4 > x_3 > x_1$.

Consequently, matrix K is sufficiently *a-transitive* and also sufficiently *a-consistent*.

Notice that accidentally the ranking of alternatives given by all four pair-wise comparison matrices *B*, *E*, *H* and *K* is the same.

7 Conclusion

In this contribution we investigated two types of incompatibility of pair-wise comparison matrices called inconsistency and intransitivity. These incompatibilities were measured by consistency index/ratio and transitivity index/ratio and the acceptable extent of inconsistency/intransitivity was defined. By the numerical example it was demonstrated that inconsistency and intransitivity may appear independently of each other, hence measuring inconsistency as well as intransitivity may be important in the real decision making.

Acknowledgements

This paper has been supported by GACR project No. 402090405.

References

[1] Chiclana, F., Herrera, F., Herrera-Viedma, E.: Integrating Multiplicative Preference Relations in a Multipurpose Decision Making Model Based on Fuzzy Preference Relations. *Fuzzy Sets and Systems* **112** (2001) 277–291.

[2] Ramík, J.: Measuring transitivity of fuzzy pairwise comparison matrix. In: *Proceedings of 30th International Conference Mathematical Methods in Economics 2012* (Ramík, J., and Stavárek, D., eds.). Silesian University, School of Business Administration, Karviná, 2012, 751–756.

[3] Saaty, T. L.: *The analytic hierarchy process: Planning, priority setting, resource allocation*. McGraw Hill, New York, 1980.

[4] Tanino, T.: Fuzzy Preference Relations in Group Decision Making. *Fuzzy Sets and Systems* **12** (1984) 117–131.

Some of the Results of the Mathematical Group in the Institute of Automation and Computer Science

Jindřich Klapka[1], Vítězslav Ševčík[2]

Abstract. Research interest within the Mathematics group in the Institute of Automation and Computer Science of the Brno UT, Faculty of Mechanical Engineering is focused on the creation and application of mathematical methods of operational research for project management and production control. Some results achieved by members of the mathematics group and its PhD students are published in Handbook of Optimization, Springer-Verlag, Berlin Heidelberg, 2013.

A heuristic method for the multicriteria selection of the group of project is being developed in the institute. This method enables the objectification of weights of the partial non-linear criteria functions, involves synergistic effects of second- and third- orders in benefit and cost criterion function, and in resource requirements respecting resource sharing and hierarchical contingency relationships among candidate projects, and dialogue of the solution.

In the project scheduling, the two-criteria time-cost minimization problem, solved by fuzzy linear programming, is a basis of the program system created here, as well as a technique for deriving a time-periodical production schedule for a group of products in a group of workplaces having limited capacity, performing cost optimization by lot size alteration. The manager could also observe, monitor or visualise effectively the differences between the scheduled and the actual courses of projects through the SSD Graph.

Keywords: project selection, scheduling, visualising, decision support, multicriterial optimisation, periodic production, lot sizes.

JEL classification: C44
AMS classification: 90B50

1 Decision Support System for Projects Selection

Computer support and mathematical modelling of project management consists (from the mathematical point of view) of the following basic stages (Klapka [6]): prognostic-planning stage, scheduling stage and the stage of analysis and chronicling.

One of the most important problems of prognostic-planning phase is the selection of projects by respecting appropriate criteria and by the limitations of the resources. In industrial enterprises and in the national economies there is often a need for a program system which would make it possible to carry out effectively multicriterial selection of hundreds of projects simultaneously, with tens of criterion functions including nonlinear ones, and tens of resources limitations with respect to the synergistic effects and the hierarchical interdependences between the projects. Since the tasks of this type usually belong to the "ill-defined" ones, it is advisable to use here the interactive dialogue approaches after the projects portfolio has been preliminarily optimized. Most often, the projects selected are the research and development (R&D) projects and the information systems projects. Some conditions for the formulation of the problems of this type are defined e.g. in Eilat, Golany and Shtub [2]. In Santhanam and Kyparisis [14] the problem is solved by means of the nonlinear goal programming. An excellent tool for the solution of the selection problems of great extent of input data is the idea of Stewart [15] who has created the scalarizing function

[1]Brno University of Technology, Faculty of Mechanical Engineering/Institute of Automation and Computer Science Technická 2, 616 69 Brno, Czech Republic, klapka@fme.vutbr.cz

[2]Brno University of Technology, Faculty of Mechanical Engineering/Institute of Automation and Computer Science Technická 2, 616 69 Brno, Czech Republic, sevcik@fme.vutbr.cz

by the modified reference point method, and for its optimization he has used a heuristic method of the effective gradient.

In [9] and [10] we have extended this approach to the possibility of utilizing criterion functions and constraints respecting synergistic effects of the second- and third-orders and hierarchical interdependencies between the projects. This work is based on the optimization by the modified method of effective gradient with the possibility of a dialogue improvement of the solution. The basic idea of the method is presented here.

The application of the presented decision support system proceeds under the cooperation with the Czech industrial enterprises and with other firms. The system is also applied for the teaching process in the Technical University of Brno. Problems of the similar type are still considered by a number of authors, of whom e.g. Litvinchev, López, Alvarez and Fernandez [12] introduced a fuzzyfication of some of the parameters. The lastly mentioned authors successfully managed effective optimization of portfolio of 25 000 projects, while solving the projects selection as two-criteria selection problem, where one criterion is the quality of the portfolio, and the second criterion is a number of projects in the portfolio. In our method, however, we solve a problem with tens of criterion functions with bivalent variables including polynomial ones and quotients of linear functions, tens of resources limitations and hundreds of projects.

1.1 Problem Formulation

The following problem is solved: to choose some of the s projects into the portfolio. Let i be a number of the project ($i = 1, 2, \ldots, s$). The projects belong to different categories (e.g. from the project type and the client type point of view). The categories need not be mutually exclusive. Let $S(k)$ be a set of the projects falling into category k ($k = 1, 2, \ldots, q$). The goal of the solution is to find for all i the values of bivalent variables δ_i for which $\delta_i = 1$ if the project i is selected for the portfolio, and $\delta_i = 0$ in the opposite case. The selection should be made so that all requirements of the solution can be fulfilled, which includes the following:

1. To satisfy the resource constraints

$$\sum_{i=1}^{s} a_{ij}\delta_i - \sum_{i=1}^{s-1}\sum_{k=i+1}^{s} a_{ijk}\delta_i\delta_k + \sum_{i=1}^{s-2}\sum_{k=i+1}^{s-1}\sum_{l=k+1}^{s} a_{ijkl}\delta_i\delta_k\delta_l \leq b_j \tag{1}$$

   where $b_j > 0$ is the total availability of resource j ($j = 1, 2, \ldots, m$), $a_{ij} \geq 0$ is the amount of resource j required by project i, $a_{ijk} \geq 0$ is the amount of resource j shared by projects i and k, a_{ijkl} is the amount of resource j shared by projects i, k, and l. In general it holds that, $a_{ij} \geq a_{ijk}$, $a_{kj} \geq a_{ijk}$, $a_{ijk} \geq a_{ijkl}$, $a_{ijl} \geq a_{ijkl}$, $a_{kjl} \geq a_{ijkl}$ for all i, k, l. In case of the absence of synergistic effect in resource sharing it holds $a_{ijk} = 0$, $a_{ijkl} = 0$.

2. To satisfy contingency constraints

$$\sum_{m \in A_i} \delta_m \geq |A_i|\delta_i \qquad \text{for all } i \in H\ , \tag{2}$$

   where H ($H \subset \{1, 2, \ldots, s\}$ is a set of all projects which are contingent upon the implementation of other projects, A_i ($A_i \subset \{1, 2, \ldots, s\}$ is a set of all projects upon the implementation of which the project i is contingent. $|A_i|$ is the number of elements in the set A_i.

3. To satisfy the directive constraints

$$\delta_i = \begin{cases} 1 & \text{for } i \in B \quad (B \subset \{1, 2, \ldots, s\})\ , \\ 0 & \text{for } i \in D \quad (D \subset \{1, 2, \ldots, s\})\ , \end{cases} \tag{3}$$

   where sets B, D are mandated due to internal and external restrictions.

4. To satisfy the restrictions for mutually exclusive projects: for some i, j ($i, j \in \{1, 2, \ldots, s\}$), can be required: if $\delta_i = 1$, then $\delta_j = 0$, and if $\delta_j = 1$, then $\delta_i = 0$ (e.g. in case when two projects represent alternative levels of activity on the same essential problem).

5. To obtain the highest possible values of criterion functions of gain (benefit)

$$z_j = \sum_{i=1}^{s} c_{ij}\delta_i + \sum_{i=1}^{s-1}\sum_{k=i+1}^{s} c_{ijk}\delta_i\delta_k + \sum_{i=1}^{s-2}\sum_{k=i+1}^{s-1}\sum_{l=k+1}^{s} c_{ijkl}\delta_i\delta_k\delta_l \quad (j = 1, 2, \ldots, p)\,, \tag{4}$$

where $c_{ij} \geq 0$ is the jth benefit derived from implementing project i alone, $c_{ijk} \geq 0$ is the additional jth benefit derived from implementing projects i and k together, and $c_{ijkl} \geq 0$ is the additional jth benefit derived from implementing projects i, k and l together.

Comments:

(a) In a similar way it is possible to formulate a cost-related objective, a negative value of which is maximized by approaching to zero.

(b) The special case when $c_{ijk} = 0$, $c_{ijkl} = 0$ corresponds to the absence of synergistic effect of benefit. Under the simplifying assumptions concerning the additivity of risk, the risk of the set of projects selected may be expressed by the first term in (4) where c_{ij} is now the risk of implementing project i. In this case, we minimize the total risk of portfolio of the selected projects by maximizing

$$\left\{-\sum_{i=1}^{s} c_{ij}\delta_i\right\}.$$

6. To obtain the smallest possible deviation of Stewart function

$$\Phi_k = \frac{\sum_{i\in S(k)} \mu_i\delta_i}{\sum_{i=1}^{s} \mu_i\delta_i} \qquad (k = 1, 2, \ldots, q)\,, \tag{5}$$

from its ideal value π_k (see the above definition of project categories) where μ_i is the cost related to the project i (or e.g. total manpower used by project i). It appears that $\Phi_k \in [0; 1]$, $\pi_k \in [0; 1]$. Let us assume that for at least one i it holds that $\delta_i = 1$.

1.2 Arrangement of the Formulation

Let us choose "asymmetric distance" of Φ_k from π_k, denoted by $\|\Phi_k - \pi_k\|$, in such a way that its value belongs to interval $[0; 1]$ and that for $\Phi_k = 1 \wedge \pi_k \neq 1 \vee \Phi_k = 0 \wedge \pi_k \neq 0$ it holds that $\|\Phi_k - \pi_k\| = 1$ and that for $\Phi_k = \pi_k$ it holds that $\|\Phi_k - \pi_k\| = 0$. For this purpose we define

$$\|\Phi_k - \pi_k\| = \begin{cases} 0 & (\Phi_k = \pi_k)\,, \\ \frac{\Phi_k - \pi_k}{l(\Phi_k - \pi_k) - \pi_k} & (\text{otherwise})\,, \end{cases}$$

where unit-step function $l(x) = 0$ for $x \leq 0$, $l(x) = 1$ for $x > 0$. This means that maximal possible deviations of Φ_k on either side of π_k are equally important.

Then, it is possible to reformulate the problem in the following way:

$$\text{"max"}\, z_j \qquad (j = 1, 2, \ldots, p + q)\,.$$

We solve this problem under constraints (1)–(3), where the criterion functions z_j $(j = 1, 2, \ldots, p)$ are given in (4) with signs of individual terms possibly changed with respect to Comments 5a and 5b. The criterion functions z_{p+k} $(k = 1, 2, \ldots, q)$ are now defined by

$$z_{p+k} = -\|\Phi_k - \pi_k\| = \begin{cases} 0 & (\Phi_k = \pi_k)\,, \\ \frac{\pi_k - \Phi_k}{l(\Phi_k - \pi_k) - \pi_k} & (\text{otherwise})\,, \end{cases} \qquad k = 1, 2, \ldots, q\,. \tag{6}$$

1.3 Optimization and dialogue

In Santhanam and Kyparisis [14] a project selection problem is solved with criterion functions of type (4), resource constraints (1) and contingency constraints (2) for 14 projects ($s = 14$) with accounting for interdependences (type of benefit, cost, resources sharing, and contingency) of up to the third-order

by means of the method based on the goal programming. Decision support system presented here by us enlarges the capabilities of Santhanam's system by the possibilities of solving the problems of more greater extent of input data, utilizing also the balance ratio functions of (5) type, then by the dialogue that makes it possible to solve also the ill-defined problems and by the utilization of the preliminary realistically assessed desired levels of individual criterion functions. With respect to this dialogue, it is not necessary to use accurate methods for optimization, but it is possible to use a heuristic method which makes it possible to enlarge the extent of input data of the solved problem.

For each criterion function z_j we determine the upper bound I_j (in the way of solving an appropriate monocriterial maximization problem with the criterion function z_j), and analogically the lower bound N_j of its optimum value through monocriterial minimization problem. It is very easy to find the bounds for each criterion function. At the same time, the user determines a realistically assessed desired level (reference level) R_j of each z_j. We require

$$N_j \leq R_j \leq I_j \ . \tag{7}$$

In case when the user is not able to order R_j, we can set the initial reference level as follows:

$$R_j = \frac{I_j + N_j}{2} \ .$$

The problem is now transformed to minimizing the scalarizing function

$$\sigma = \sum_{j=1}^{p+q} \left(\frac{I_j - z_j}{I_j - R_j} \right)^h$$

for some $h > 0$ under conditions (1)–(4), (6). In the system presented here, a method of effective gradient (Stewart, [15]), generalized by us for the case of the synergistic effects and hierarchical interdependences of projects, is used for the solution of this minimization problem. The selection $h = 4$ as a compromise between the sensitivity of the method and time and rounding off numerical difficulties proved to be equally right for us and the above-mentioned author. By means of this solution optimal values δ_i for all $i = 1, 2, \ldots, s$ are determined, which define a portfolio of projects selected.

The main idea of our generalized effective gradient method is as follows:

At the beginning of the process of solution, we select $\delta_i = 1$ for all $i = 1, 2, \ldots, s$ (except for a case when some projects must not take place simultaneously, in which case an arbitrary selection can be made between these). Then we calculate

$$u_j = \sum_{i=1}^{s} a_{ij}\delta_i - \sum_{i=1}^{s-1} \sum_{k=i+1}^{s} a_{ijk}\delta_i\delta_k + \sum_{i=1}^{s-2} \sum_{k=i+1}^{s-1} \sum_{l=k+1}^{s} a_{ijkl}\delta_i\delta_k\delta_l$$

for all $j = 1, 2, \ldots, m$, which is the amount of the jth resource requested by those projects that were incorporated in the portfolio. In case that a resource constraint is not satisfied, then for each i for which $\delta_i = 1$, we define $\Delta_i\sigma$ as an increase of function σ caused by dropping project i from the portfolio. For all such i we calculate

$$P_i = \Delta_i\sigma \frac{\sqrt{\sum_j (u_j - b_j)^2}}{\sum_j A_{ij}(u_j - b_j)} \ ,$$

where

$$A_{ij} = a_{ij} - \sum_{k=1,2,\ldots,i-1,i+1,i+2,\ldots,s} a_{ijk}\delta_k + \sum_{k=1,2,\ldots,i-1,i+1,i+2,\ldots,s-1} \ \sum_{l=k+1,k+2,\ldots,i-1,i+1\ldots,s} a_{ijkl}\delta_k\delta_l \tag{8}$$

is the amount of the jth resource consumed due to the implementation of the ith project. P_i is an effective gradient of the scalarizing function σ. The sums in the relation for P_i are realized for all $u_j \geq b_j$ only. The project giving the expression P_i a minimum value, will be dropped from the portfolio.

This step repeats until all resource limitations are fulfilled. Analogically, in a backward course, we introduce into portfolio a maximal number of projects that do not violate any resource constraint.

The dialogue between the user of the system and the person solving the problem, accomplished after the introductory optimization, influences also the value of reference levels R_j in an adaptive way, and

thus also the weights of components of the scalarizing function for the purposes of potential future reoptimization of portfolio.

This dialogue is described in [10]. The dependence of the calculation time for optimization on the number of projects is given in Figure 1.

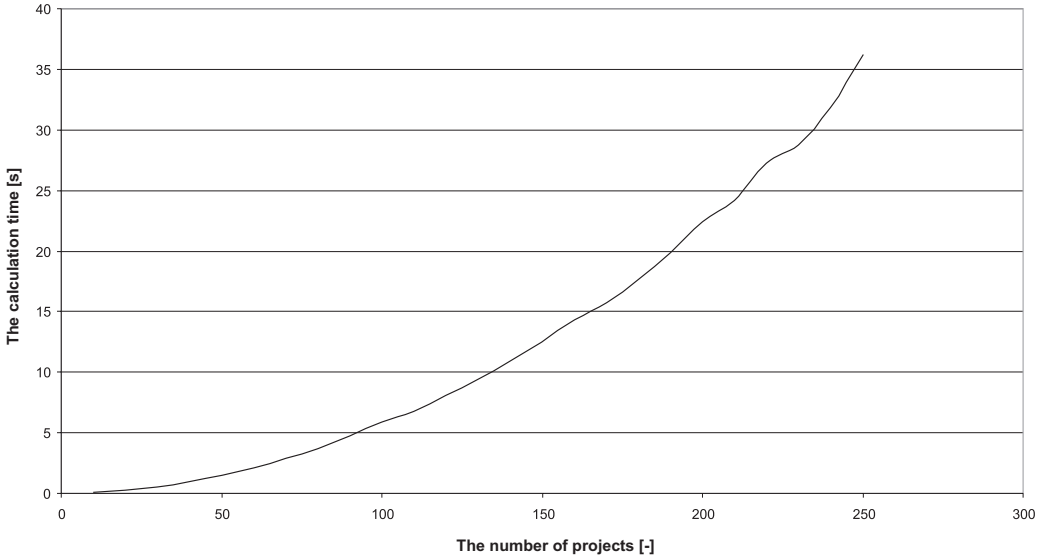


Figure 1 Graph of the dependence of the calculation time on the number of projects ($m = 50$, $p = 20$, $q = 20$).

2 Scheduling of the selected projects

Each project is given by its network graph. For every activity of the project resulting from node i and entering node j the following quantities are given: $y_{ij}^{(0)}$ = normal duration, m_{ij} = maximum crashing, k_{ij} = crashing cost. We search values of the following variables: x_i = time at event i, h_{ij} = crash time for activity (i, j). For each i, j the following conditions holds:

$$\begin{aligned} h_{ij} &\le m_{ij} \,, \\ h_{ij} &\ge 0 \,, \\ x_j - x_i + h_{ij} &\ge y_{ij}^{(0)} \,, \\ x_1 &= T_b && \text{(project start time)} \\ x_i &\ge 0 && (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, n) \end{aligned}$$

Then, partial criterion functions are $Z_1 = \sum k_{ij} h_{ij}$ (crashing cost), $Z_2 = x_n - x_1$ (duration of project). Let us denote $\min Z_1 = U_1$, $\min Z_2 = U_2$. For $Z_1 = U_1$ let us denote $Z_2 = L_2$. For $Z_2 = U_2$ let us denote $Z_1 = L_1$. Then, the two-criterial time-cost minimisation problem is to choose all x_i, h_{ij} which provide $\min Z$ where $Z = \lambda$ under the conditions

$$\lambda \ge \frac{Z_1 - U_1}{L_1 - U_1} \,, \qquad \lambda \ge \frac{Z_2 - U_2}{L_2 - U_2}$$

and the above-mentioned linear condition of the problem. The solution is made by *fuzzy linear programming*. As it is shown in [13] and in the diploma theses carried out at our institute, this method leads to substantially lower total affectable costs of a project than the heuristic "method of cost gradient" known until now.

3 Contemplating the Differences between Scheduled and Actual Course of Projects

For this purpose for each activity a_i of the project and time t we define a scheduled state $\overline{m}_{it}$ such as

$$\overline{m}_{it} = \begin{cases} -1 & (\text{if } a_i \text{ is not begun}) \\ 0 & (\text{if } a_i \text{ is in process}) \\ 1 & (\text{if } a_i \text{ is finished}) \end{cases}$$

and analogically an actual state m_{it}. Deviation d_{it} is defined as the difference between the actual and scheduled states on time t by $d_{it} = m_{it} - \overline{m}_{it}$.

$$d_{it} = \begin{cases} -2 & (\text{if } a_i \text{ is delayed by the second order}) \\ -1 & (\text{if } a_i \text{ is delayed by the first order}) \\ 0 & (\text{if } a_i \text{ is not delayed or in advance}) \\ 1 & (\text{if } a_i \text{ is in advance by the first order}) \\ 2 & (\text{if } a_i \text{ is in advance by the second order}) \end{cases}$$

A set of possible values of deviations d_{it} is $\{-2, -1, 0, 1, 2\}$. It is apparent that such concepts of state and deviations of project activities are very simple and efficient instruments for monitoring and visualising the course of project and the deviation of the actual course from the scheduled one. This non-traditional approach utilising according to [11] the synthesis of Petri nets with Gantt diagram may be combined both with the schedule described in § 2 and also with the commonly available system Microsoft Project. The program systems are prepared for use in Czech enterprises.

4 Time-Periodical Production Scheduling

Periodicity is often an important requirement supposed on the serial production e.g. in mechanical engineering. Scheduling plays an important role in manufacturing and production systems. In the last few decades, many algorithms to solve the production schedule of the group of products in the group of workplaces problem have been developed. Those techniques that are based on exact methods (e.g. [1]) are in most cases suitable for simultaneous production scheduling of a relatively small number of products manufactured on a small number of machines (production stages). For large scale problems, algorithms based on the modern heuristic methods were developed (e.g. [16]); however these approaches provide only an approximate solution and often without error estimation. Exact methods could be advantageously used for such an auxiliary computation as a computation of the lower and upper bound of a demanded criterion function [5, 7] and thus contribute to the efficiency of the heuristic method. The most used criteria are time (total processing time), controllable cost, and idle time. We try to find the optimal solution with respect to all of these measures. The resulting solution consists of the time schedule for all the manufacturing operations and their lot sizes.

In [8] we consider the problem of serial and time-periodical production in the engineering industry where the input parameters are deterministic. We propose a method of improving an existing schedule of the group of products (e.g. in a machinery and equipment manufacturing) with the goal of a cost optimisation through the lot sizes alteration using a multiplication by an appropriately selected constant. The practical benefit of this approach is also evident from the fact that its use can improve the schedule (after a condition of the periodicity of the production is introduced) created by virtually any suitable modern [3, 4, 16] or problem oriented heuristic method, while the mutual shares of lot sizes are preserved.

5 Conclusion

We tried to perfect current situation in using project management in production and manufacturing by making a system of decision support provided with a dialogue regime and by using non-traditional element such as fuzzy linear programming and the approach based upon the synthesis of Petri nets with Gantt diagrams. The newly constructed system makes the integrated application of these elements and functions possible.

Acknowledgements

This paper was supported by research project of BUT IGA No. FSI-S-11-31 "Application of Artificial Intelligence".

References

[1] Balkhi, Z. T. On the optimality of multi item integrated production inventory systems. In: *Proceedings of the 3rd International Conference on Applied Mathematics, Simulation, Modelling, Circuits, Systems and Signals* WSEAS, Wisconsin, 2009, 143–154. ISBN 978-960-474-147-2.

[2] Eilat, H., Golany, B., Shtub A.: Constructing and evaluating balanced portfolios of R&D projects with interactions: A DEA based methodology, *European Journal of Operational Research* **127** 3 (2006), 1018–1039. doi:10.1016/j.ejor.2004.12.001

[3] Jeunet, J., Jonard, N.: Single-point stochastic search algorithms for the multi-level lot-sizing problem, *Computers and Operations Research 32* 4 (2005), 985–1006. doi:10.1016/j.cor.2003.09.009

[4] Kim, Y. W., Suzuki, T., Narikiyo, T.: FMS scheduling based on timed Petri Net model and reactive graph search, *Applied Mathematical Modelling* **31** 6 (2007), 955–970. doi:10.1016/j.apm.2006.10.023

[5] Klapka., J.: A Heuristic Policy of a Multistage Production Process *Ekonomicko-matematický obzor* **16** (1980), 13–22.

[6] Klapka, J.: Contemporary state of mathematical modelling in project management. *Modelling, Measurement and Control D 9* **3** (1994), 43–63.

[7] Klapka. J., Dvorak, J.: A More Efficient Computation of Optimal Lot Sizes in a Multistage Production Process. *Ekonomicko-matematický obzor* **16** (1980), 179–189.

[8] Klapka, J., Matousek, R., Sevcik, V.: Improvement of Time-Periodical Production Schedule of the Group of Products in The Group Of Workplaces Through the Lot Sizes Alteration In: *Mendel 2011 – 17th International Conference on Soft Computing*, Brno University of Technology VUT Press, Brno, 2011, 334–340. ISBN 978-80-214-4302-0.

[9] Klapka, J., Pinos, P.: Decision support system for multicriterial R&D and information systems projects selection. *European Journal of Operational Research* **140** 2 (2002), 434–446. doi:10.1016/S0377-2217(02)00081-4

[10] Klapka, J., Piňos, P., Ševčík, V.: Multicriterial Project Selection. In: *Handbook of Optimization, ISRL 38* (Zelinka et al., eds.), Springer, Berlin Heidelberg, 2013, 245–261. doi:10.1007/978-3-642-30504-7_10

[11] Lee-Kwang, H., Farrel, J.: The SSD graph: a tool for project scheduling and visualization, *IEEE Transactions on Engineering Management* **35** 1, (1988), 25–30. doi:10.1109/17.6001

[12] Litvinchev, I.S., López, F., Alvarez, A., Fernandez, E.: Large-Scale Public R&D Portfolio Selection by Maximizing a Biobjective Impact Measure, *IEEE Transactions on Systems, Man, and Cybernetics* **40** 3 (2010), 572–582. doi:10.1109/TSMCA.2010.2041228

[13] DePorter, E. L., Ellis, K. P.: Optimization of Project Networks with Goal Programming and Fuzzy Linear Programming, *Computers & Industrial Engineering* **19** 1–4, (1990), 500–504. doi:10.1016/0360-8352(90)90168-L.

[14] Santhanam, R., Kyparisis, J.: Multiple criteria decision model for information system project selection. *Computers and Operations Research* **22** 8 (1995), 807–818. doi: 10.1016/0305-0548(94)00069-K

[15] Stewart, T.J.: A multi-criteria decision support system for R&D project selection, *Journal of the Operational Research Society* **42** (1991), 17–26. doi:10.1057/jors.1991.3

[16] Xie, J., Dong, J.: Heuristic genetic algorithms for general capacitated lot-sizing problems, *Computers & Mathematics with Applications.* **44** 1–2 (2002), 263–276. doi:10.1016/S0898-1221(02)00146-3

Multinomial Ordinal Response Models

Martin Klicnar [1] Lenka Činčurová[2]

Abstract. For managing of credit risk in banks, it is fundamental to decide to which customer a loan or service will be offered and to which one it will not. The banks use many methods modeling and forecasting risk – application scorecards, probability of default etc. Inside these methods mainly binary models and logistic regression as technique are applied. The article intends to shortly illustrate and describe extension of these models on situations where response (dependent) variable has more than two categories. The binary models (or dichotomous, i.e. models in which dependent variable has just two categories) are not the only ones, it is possible to use models in which dependent variable has more than two categories – multinomial models.
There are two main groups of such models: with ordered (ordinal) categories and non-ordered (nominal) categories of the dependent variable. The article describes representatives from both groups, but focus on ordinal models, which are likely to be used in the financial institutions than nominal ones. For both groups practical examples are calculated and their results are discussed.

Keywords: Credit scoring, generalized linear regression model, logistic regression, logit, probit, binary, multinomial/multicategorical ordinal and nominal models.

JEL Classification: C35
AMS Classification: 62J12**,** 91G40

1 Introduction

One of the key tool of financial institutions for risk and also cost management is credit scoring and following risk segmentation of customers by the score.

Basic aim of the segmentation is to find good and bad clients, i.e. those who will probably pay, and those who will not. Most of used methods (e.g. binary logistic regression as mostly used) compute probability or number (score) that customer will be good or bad (will pay or not), hence transition from good to bad client is fluent (continuous). However, the way, how to obtain the probability or score, works only with purely good and bad customers, many cases without such easy classification (called indeterminate, intermediate) are excluded and unused. Or it works with all records/customers (each investigated case is decided as good or as bad) , but in such cases some expressive differences in characteristics could be diminished or even disappear. For that reason it is suitable to use multinomial (ordinal) models, in which the intermediate cases create the new (middle) category.

Needs of management or business units could be wider. Main target of firms is to increase sell and profit. But more reasonable is to attract good customers in both terms: of business and of risk. For the former aim as much they will react on offer. For such situations, nominal multinomial models can be used.

This article shortly introduces models and methods which use more than 2 attributes of response variable with focus on ordinal ones and shortly compare them with more used binary methods.

2 Theoretical part

2.1 Binary models

As representative of binary response models, binary logistic regression as most often used is introduced. It supposes that probability of event y_i (labeled as 1) under condition that random explanatory variables have concrete values x_i is derived by logistic distribution, or mathematically – see [1], [2], [5], [6]:

$$P(y_i = 1|x_i) = \pi_i = \frac{1}{1+\exp(-x_i'\beta)} = \frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)},$$

[1] University of South Bohemia in Ceske Budejovice, Faculty of Economics, Studentska 13, 370 05 Ceske Budejovice, martin.klicnar@centrum.cz
[2] University of South Bohemia in Ceske Budejovice, Faculty of Economics, Studentska 13, 370 05 Ceske Budejovice, cincurov@ef.jcu.cz.

where *i* means order of record and β is a vector of unknown parameters. Probability of complementary event is:

$$P(y_i = 0|x_i) = 1 - \pi_i = \frac{1}{1+\exp(x_i'\beta)}.$$

Based on previous relation a *logit* is defined as a logarithm of division of these two probabilities:

$$logit\ \pi_i \equiv \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \ln\left(\frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)} \Big/ \frac{1}{1+\exp(x_i'\beta)}\right) = \ln(\exp(x_i'\beta)) = x_i'\beta .$$

Fraction inside the natural logarithm expresses relative chance that the alternative 1 will happen against that the alternative 0 will happen. The ratio is usually called *odds* or *odds ratio*.

One important comment: it is evident that extreme values for *x* are not problematic. Probability values according to definition must be from interval (0, 1).

Very similar, but not often used, is a *probit* model. For that model normal distribution and probability are used instead of the logistic ones. It could be defined by the following way:

$$P(y_i = 1|x_i) = \pi_i = \Phi(x_i'\beta),$$

where Φ is a standard normal distribution function N(0,1). It is rarely used due to very demanding computation.

While using simple probability linear model, we obtain:
$P(y_i = 1|x_i) = \pi_i = x_i\beta$ and $P(y_i = 0|x_i) = 1 - \pi_i = 1 - x_i\beta$. Here is evident, that it is hardly to ensure basic feature of probability – to be in an interval <0, 1>. Such models are very sensitive on extreme values.

2.2 Multinomial models

Multinomial models are extension of binary models. Models with multinomial dependent variable (i.e. with more than two categories) are divided to 2 basic groups:

- with nominal response variable – there is no ordering among alternatives
- with ordinal response variable – there is ordering among alternatives

Models with nominal response variable

Models with nominal (= non-ordered) response variable could be used when it is necessary to choose from a list of option, for which there is no or very small preference (deciding about mobile phone/operator, brand of car, university, place of holiday etc.).

Typical representative of nominal models is a multinomial logistic model. It can be written for *q* alternatives (categories) this way according to [1], [6]:

$P(y_i = r|x_i) = \pi_{i,r} = \frac{\exp(x_i'\beta_r)}{1+\sum_{j=1}^{q-1}\exp(x_i'\beta_j)}$ for each category $r = 1,\ldots, q\text{-}1$ and

$P(y_i = q|x_i) = \pi_{i,q} = \frac{1}{1+\sum_{j=1}^{q-1}\exp(x_i'\beta_j)}$ for a reference category *q*.

The logit is defined as division of probabilities of the category *r* and of reference category *q:*

$$logit\ \pi_{i,r} \equiv \ln\left(\frac{P(y_i = r|x_i)}{P(y_i = q|x_i)}\right) = x_i'\beta_r .$$

Models with ordinal response variable

There are several possible classifications of multinomial ordinal models and plenty of models. Fahrmeier and Tutz in [1, pp. 81-104] introduce cumulative and sequential models. While Hosmer and Lemeshow, focusing only on logistic regression in [3, pp. 288-291], define three groups of models: the adjacent category, the continuation ratio and proportional odds models. Representatives from all groups will be shortly introduced, but only cumulative models will be followed by practical examples.

Typical example of the cumulative models with ordinal response is the *cumulative logistic model*. It could be described in following way, see e.g. [1], [6]:

$P(y_i \le r|x_i) = \frac{\exp(\theta_r + x_i'\beta)}{1+\exp(\theta_r + x_i'\beta)}$, where $-\infty = \theta_0 < \theta_1 < \ldots < \theta_q = \infty$ and $r = 1, 2, \ldots, q$.

The model is defined through sum of probabilities, when dependent variable will be equal to 1, 2, …, *r* (left-hand side of the equality). Hence, why the term "cumulative" is used.

With respect to $P(y_i > r|x_i) = 1 - P(y_i \le r|x_i) = \frac{1}{1+\exp(\theta_r + x_i'\beta)}$, we can also write and derive:

$$\frac{P(y_i \le r|x_i)}{P(y_i > r|x_i)} = \exp(\theta_r + x_i'\beta) \text{ or } \ln\left(\frac{P(y_i \le r|x_i)}{P(y_i > r|x_i)}\right) = \theta_r + x_i'\beta .$$

For derivation of the model see e.g. Hušek [4, p.195]. By comparing probabilities of the event (one from several possibilities) and its complementary, inserting their ratio into natural logarithm, we again obtain the *logit*.

Other cumulative models are the *probit* model with normal distribution

$$P(y_i \le r|x_i) = \Phi(\theta_r + x_i'\beta), \text{ where } r = 1, 2, \ldots, q.$$

or *grouped Cox model* (or proportional hazards model) with extreme minimal value distribution, see also [1]:

$$P(y_i \le r|x_i) = 1 - \exp\{-\exp(\theta_r + x_i'\beta)\}, \text{ where } r = 1, 2, \ldots, q.$$

Or with *complementary log-log* link: $\log[-\log P(y_i > r|x_i)] = (\theta_r + x_i'\beta)$.

At the end of this part, classification and models derivation from Hosmer and Lemeshow [3] will be presented. The *adjacent-category* logistic model compares each response to the next larger response. Assuming that odds (and logit as well) does not depend on the response and the logit is linear in the coefficients the form of model could be following: $\ln\left(\frac{P(y_i = r|x_i)}{P(y_i = r-1|x_i)}\right) = \theta_r + x_i'\beta$

The second model compares each response to all lower responses. He is called *continuation-ratio* logistic model and could be described such a way: $\ln\left(\frac{P(y_i = r|x_i)}{P(y_i < r-1|x_i)}\right) = \theta_r + x_i'\beta.$

The third and last one compares probability of equal or lower response to probability of larger response and they call it as *proportional odds* model: $\ln\left(\frac{P(y_i \le r|x_i)}{P(y_i > r|x_i)}\right) = \theta_r + x_i'\beta$

This is exactly the same as cumulative logistic model. So, cumulative logistic model and proportional odds model are synonyms (e.g. used in [1]). For more details and important features of each model see [1], [2] and [3].

3 Practical part

Real data from a financial company were used for practical examples of using different methods including several transformations (grouping variables or transferring continuous variable to class one). The data were taken from closed or running contracts, for which one from 3 statuses (*good* - G, *bad* - B, or *intermediate* /or *indeterminate*/ - I) were be defined based on their payment histories. These free statutes have been transformed to statuses 0, 2, 1 (to keep appropriate ordering) for applying of multinomial models and related methods. For all models statistical software SAS were used, concretely procedures Logistic and GENMOD (including consultation on web pages in [7]).

Firstly logistic regression in statistical software SAS (PROC Logistic) was used, firstly for the binary model and status good and bad, then twice for logistic regression for all 3 statuses – concretely general (nominal) logistic regression and cumulative logistic regression. The first two model results (but with another data) were published in [5], all three in [6]. Then procedure GENMOD were used to calculate other cumulative model, namely cumulative probit and again cumulative logit for comparison (both introduced in theoretical section).

In this article, only part of standard SAS outputs is inserted due to limited space - only list of variables selected to the model with relevant statistics, part of parameters estimates, final statistics about model (e.g. several measures of association between estimated model and original data), where it has sense.

As for measures of association – for each model it is different. In the binary one it measures correct assignment only for good and bad, in model with ordinal response measures all three categories (good – indeterminate – bad), in model with nominal (categorical) response these statistics even are missing – in this model it is not stated which status is better and which worse, model only compares one status to reference one.

3.1 Binary model

For easy comparison and not so long outputs parameters alpha, which decide about insertion of concrete variable into a model (maximal probability that the variable is insignificant although I inserted it as significant into the model), was set for all models to 0.01 (1%).

For the binary model (5766 records: 5581 good, 185 bad), logistic regression gives only 5 variables and Somers' D (or GINI) coefficient 0.541 – see the SAS output below.

Parameter	*DF*	*Estimate*	*Standard error*	*Wald Chi-Square*	*Pr > ChiSq*
Intercept	1	0.8192	0.5892	1.9329	0.1644
bapaymth B	1	1.0946	0.2457	19.8493	<.0001
bahomtel Y	1	0.6532	0.2160	9.1491	0.0025
timjob 3	1	0.1116	0.2148	0.2702	0.6032
timjob 5	1	0.3279	0.2226	2.1700	0.1407
timjob 10	1	0.4646	0.2679	3.0072	0.0829
timjob 20	1	1.3118	0.3828	11.7460	0.0006
baaccomc CT	1	0.4551	0.5853	0.6047	0.4368
aaccomc HO	1	1.1697	0.5535	4.4653	0.0346
baaccomc LP	1	0.9226	0.6303	2.1423	0.1433
baaccomc O	1	0.3956	0.5706	0.4808	0.4880
baaccomc TF	1	0.4297	0.5678	0.5728	0.4492
timrel 1	1	3.7522	1.0079	13.8603	0.0002
timrel 5	1	0.8199	0.3427	5.7239	0.0167
timrel 10	1	0.7907	0.2026	15.2278	<.0001
timrel 15	1	0.6162	0.1904	10.4730	0.0012

Association of Predicted Probabilities and Observed Responses: Somers' D 0.541

3.2 Model with three categories – nominal response variable

Multinomial nominal model with 3 statuses (good, i.e. 0 – 5581 records, medium as 1 – 621 records, bad as 2 – 185 records) chose 9 variables – see the results below. But each has double degrees of freedom, because each response (except the reference one), i.e. responses 0 and 1, is compared and estimated against the reference category (response 2).

Effect	*DF*	*Chi-Square*	*Pr > ChiSq*
bac003	2	10.8364	0.0044
bapaymth	2	70.4523	<.0001
baemprol	18	36.6721	0.0058
bahomtel	2	20.5513	<.0001
timjob	8	39.7773	<.0001
baaccomc	10	94.4228	<.0001
appamt	8	40.3286	<.0001
grterm	8	72.8954	<.0001
timrel	8	230.6500	<.0001

Parameter	*target3*	*DF*	*Estimate*	*Standard Error*	*Wald Chi-Square*	*Pr > ChiSq*
Intercept	0	1	-0.5983	1.3571	0.1943	0.6593
Intercept	1	1	2.6207	1.4359	3.3314	0.0680
bac003 Y	0	1	0.3206	0.2156	2.2112	0.1370
bac003 Y	1	1	-0.0469	0.2412	0.0379	0.8457
bapaymth B	0	1	0.9638	0.2585	13.9046	0.0002
bapaymth B	1	1	-0.0510	0.2853	0.0320	0.8581

baemprol	C	0	1	1.6190	1.1940	1.8385	0.1751
baemprol	C	1	1	0.5434	1.2325	0.1944	0.6593
baemprol	MG	0	1	1.0241	1.1643	0.7737	0.3791 ….

The increased number of estimated parameters has impact on SC (Schwartz criterion) measuring model in terms of overfitting (too many variables and estimated parameters). The presented model has SC 5543.066, while the lowest value of SC (5361.163) is obtainable after insertion only 3 (!) variables.

3.3 Model with three categories – ordinal response variable (cumulative model)

Cumulative logistic regression using PROC Logistic

Step	*Effect*	*DF*	*Chi-Square*	*Pr > ChiSq*
1	baaccomc	5	228.1603	<.0001
2	bapaymth	1	95.4431	<.0001
3	timrel	4	80.1009	<.0001
4	grterm	4	58.2071	<.0001
5	appamt	4	46.4525	<.0001
6	timjob	4	34.1833	<.0001
7	bahomtel	1	16.0543	<.0001
8	bac003	1	7.7590	0.0053
9	baemprol	9	25.1272	0.0028

Parameter		*DF*	*Estimate*	*Standard Error*	*Chi-Square*	*Pr > ChiSq*
Intercept	0	1	-2.2369	0.7357	9.2456	0.0024
Intercept	1	1	-0.5547	0.7364	0.5673	0.4513
bac003	Y	1	0.3377	0.1078	9.8179	0.0017
bapaymth	B	1	0.9759	0.1187	67.5670	<.0001
timrel	1	1	-0.4587	0.1218	14.1794	0.0002
timrel	5	1	-0.6670	0.1380	23.3674	<.0001
timrel	10	1	0.4043	0.1216	11.0480	0.0009
timrel	15	1	0.4054	0.1184	11.7125	0.0006

…

Association of Predicted Probabilities and Observed Responses: Somers' D 0.249

The model selected again 9 variables (the same as in the nominal model), more than 5 as in the binary model. But you can see deep fall for Somers' D coefficient (0.541 in the binary model) – it is one of disadvantage of ordered models that the differences between categories are diminishing.

PROC GENMOD and cumulative logit

Procedure GENMOD estimates parameters for all included variables regardless of their significance. Hence, if one need direct comparison with previous models, it is necessary to manually exclude insignificant variables (e.g. exclude those where each parameter has "Pr > ChiSq" > 0.01).

This choice gives very similar result as PROC Logistic. Also it selected 9 variables, the differences are caused by different definition of reference categories of independent variables (PROC Genmod is not so "friendly" for human choice) – see part of the SAS output below.

Parameter		*DF*	*Estimate*	*Error*	*Confidence Interval*		*Chi-Square*	*Pr > ChiSq*
Intercept1		1	1.6893	0.7333	-0.1996	3.5782	5.31	0.0212
Intercept2		1	3.3716	0.7354	1.4772	5.2659	21.02	<.0001
bac003	Y	1	0.3377	0.1078	0.0601	0.6154	9.82	0.0017
bapaymth	A	1	-0.9759	0.1181	-1.2802	-0.6716	68.23	<.0001
timrel	1	1	-0.4588	0.1203	-0.7685	-0.1490	14.55	0.0001
timrel	5	1	-0.6670	0.1362	-1.0179	-0.3161	23.97	<.0001
timrel	10	1	0.4043	0.1228	0.0879	0.7206	10.83	0.0010
timrel	15	1	0.4053	0.1199	0.0965	0.7141	11.43	0.0007

…

Log likelihood -2569.9999

Looking at variable *timrel* we can see almost the same estimates. The estimates for two intercepts are different, but important is their distance measuring distance between "thresholds" for responses 0, 1 and 2 – in former model it is 1.6822, at latter one 1.6823. Almost the same difference.

PROC GENMOD and cumulative probit.

Parameter		*DF*	*Estimate*	*Standard Error*	*Confidence. interval*		*Chi-Square*	*Pr > ChiSq*
Intercept1		1	0.9364	0.4175	-0.1390	2.0118	5.03	0.0249
Intercept2		1	1.7627	0.4181	0.6857	2.8397	17.77	<.0001
bac003	Y	1	0.1711	0.0553	0.0286	0.3136	9.56	0.0020
bapaymth	A	1	-0.5050	0.0587	-0.6562	-0.3537	73.98	<.0001
timrel	1	1	-0.1311	0.0650	-0.2985	0.0364	4.06	0.0439
timrel	5	1	-0.2890	0.0753	-0.4828	-0.0951	14.75	0.0001
timrel	10	1	0.2450	0.0628	0.0834	0.4067	15.25	<.0001
timrel	15	1	0.2387	0.0615	0.0803	0.3972	15.07	0.0001
…								
Log likelihood			-2583.5944					

The result are not same as for cumulative logit, but they are similar as for relations (most of estimates are close to half of logit estimates – except parameter for value "1" of *timrel*, the parameter is insignificant on level 0.01)

4 Comments and conclusion

The presented models lead to very similar results (most significant variables are included in each model), there are some differences in the number of included variables and few included variables differ from model to model.

What is immediately visible, it is difference in structure of result: the nominal multinomial model has two sets of the same parameters (2 categories are compared to reference ones, also variables in the nominal model have double number of degrees of freedom). While the ordinal model has only one set of parameters, but it has 2 constants (thresholds), by which the categories are separated. One constant more is just difference against the basic binary model (except 2 different variables chosen in each model). Inside group of ordinal models, the results are similar, the estimates have different values (levels), but keep relation and basic behavior (significance, importance/weights of concrete estimates). Because the models differ only in used distribution (or link) functions.

The multinomial models give interesting results, they could be viewed as useful extension of the binary models. On possible usage of multinomial models is in identification of most predictive variables (mainly on a border of significance). One advantage is that they use all observations. They are also necessary for cases where the target is more complex (e.g. not only to find best customers in terms of risk, but also to find most active customers). For that purpose the nominal multinomial models are suitable (despite basic requirement of much higher number of estimated parameters). More difficult interpretation and sometimes technical issues for calculation can be chosen as disadvantages for those models.

References

[1] Fahrmeir, L., and Tutz, G.: *Multivariate Statistical Modelling Based on Generalized linear models.* Springer-Verlag New York, Inc. 2001.

[2] Harrel , F. E. Jr.: *Regression Modeling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer-Verlag New York, Inc. 2001.

[3] Hosmer , D. W., and Lemshow, S.: *Applied Logistic Regression*. John Wiley& Sons, Inc., New York, 2000.

[4] Hušek, R.: *Ekonometrická analýza*. Vysoká škola ekonomická v Praze, Nakladatelství Oeconomica. 2007

[5] Klicnar, M.: *Multinomické regresní modely v řízení rizika*. IMPROFORUM 2008, Jihočeská univerzita, Ekonomická fakulta, 2008

[6] Klicnar, M.: *Multicategorial Response Models in Logistic Regression*. MME 2010, Jihočeská univerzita, Ekonomická fakulta, 2010

[7] SAS 9.1 ONLINE DOCUMENTATION: [online] [cit. 29.5.2013] http://support.sas.com/91doc/docMainpage.jsp

A note on quantiles for multidimensional data

Jana Klicnarová[1]

Abstract. In many economic applications, we need to compute multidimensional quantiles or to estimate the center of multidimensional distribution. Our approach to multidimensional quantiles is based on a modified version of interdirections which are given by halving hyperplanes. The methods working with halving hyperplanes do not require any knowledge about the location of the distribution, what is their main advantage in comparison with classical approaches.

Our aim is to study asymptotics for multidimensional quantiles. Under our approach to multidimensional quantiles, the very important tool for studying of these asymptotics are U-statistics. In fact, we need to study asymptotics for U-statistics (especially, weighted U-statistics and incomplete U-statistics).

Keywords: U-statistics, multidimensional quantile, interdirections

JEL classification: C13
AMS classification: 62G05

1 Introduction

In this paper, we focus on the problem of multidimensional quantiles. The theory of multidimensional quantiles is now very developing field. Recently, there were published a lot of interesting papers, see for example [7], [10].

In this paper, we study multidimensional quantiles using of so-called interdirections. Interdirections were presented by Randles in his paper [12]. In short, for two fixed points, the Randles's interdirections are such hyperplanes which go through $k-1$ points of observations and through an origin and which separate these fixed two points. Randles show that the number of such interdirections associated with two points y_1 and y_2 tends to angular distance between these two fixed points.

Through this paper, we suppose observations coming from symmetric distribution (without lost of generality with the center at an origin). Randles ([12]) worked with hyperplanes, which went through $k-1$ data points and the origin. In his paper, he introduced an empirical angular distance between two vectors and proposed a multivariate sign test based on this concept.

Similar concept close to Randles's one was developed also by Oja and Paindaveine in their paper [6]. Oja and Paindaveine obtained similar results to Randles, but they did not use hyperplanes going through $n-1$ observations and origin. Their results are based on all hyperplanes which go through all possible k distinct observation.

Our aim is to show similar results to the above mentioned. Our concept of interdirections is based on hyperplanes, which go through k points of observation and which divide the data set half-to-half. Surely, it is not easy to see, which hyperplanes divide the data set half-to-half, we need to define what it means "to be above" and "to be below" and in case of odd $n-p$ (where n is the number of observations and p is a dimension of our space), which hyperplane is the right one to divide out set half-to-half. The main advantage of our approach is that it leads to methods which do not require any knowledge about the location of the distribution.

[1]University of South Bohemia in České Budějovice, Faculty of Economics, Studentská 13, 370 05 České Budějovice, klicnarova@ef.jcu.cz

2 Notation

Let $X_1, X_2, \ldots, X_n$ be a observation of i.i.d. (independent and identically distributed) random variables in $\mathbf{R}^p$. By I_n, we will denote the subset of p indexes chosen from the set $\{1, 2, \ldots, n\}$, more precisely $I_n = \{i_1, i_2, \ldots, i_p\} \subset \{1, 2, \ldots, n\}$. We will write H^{I_n} for $H_{X_{i_1},\ldots,X_{i_p}} \subset \mathbf{R}^p$, where $H_{X_{i_1},\ldots,X_{i_p}}$ is the hyperplane going through p points – $X_{i_1}, \ldots, X_{i_p}$ – of the observation.

Then we can define the halving hyperplane.

Definition 1. We say, that the hyperplane H^{I_n} is a **halving hyperplane**, if

$$\begin{aligned}\sum_{i=1}^{n} \operatorname{sign}(\det((X_i, X^{I_n}, \mathbf{1}))) &= 0 \quad \text{for } n-p \text{ even},\\ &= 1 \quad \text{for } n-p \text{ odd},\end{aligned}$$

where $(X_i, X^{I_n}, \mathbf{1}) = (X_i, X_{i_1}, \ldots, X_{i_p}, \mathbf{1})$ is a matrix, where the first line is the vector $(X_i, 1)$, the second vector $(X_{i_1}, 1)$ and so on.

Let us denote by $\mathbf{d}(I_n)$ the vector whose components are the cofactors of the first line in the matrix $(x, X^{I_n}, \mathbf{1})$, where x is any p-vector (so, it is normal vector to the hyperplane H^{I_n}). We can see, that x belongs into the hyperplane H^{I_n} if $\mathbf{d}(I_n).(x, 1)$ is equal to zero.

Now, we can compute the distance between any point $x \in \mathbf{R}^p$ and hyperplane H^{I_n} as

$$d(H^{I_n}, x) = \frac{\mathbf{d}(I_n).(x, 1)}{||\mathbf{d}(I_n)||}.$$

(By $||\cdot||$, we denote the L_2-norm.)

Let us define the center given by halving hyperplanes, what is the point which is the nearest one to all halving hyperplanes. More precisely we can define.

Definition 2. The point $\hat{\theta}$ is called the center given by halving hyperplanes, iff

$$\hat{\theta} = \arg\min_{\theta \in \mathbf{R}^p} \left(\sum_{I_n \subset \{1,2,\ldots,n\}} d((\theta, H^{I_n}))\mathbf{I}_{\sum_{j=1}^{n} \det((X_j, X^{I_n}, \mathbf{1}))=0 \vee 1} \right). \tag{1}$$

So, we can see, that in fact, $\hat{\theta}$ is a function of observations. More precisely,

$$\hat{\theta} = \sum_{I_n=(i_1,i_2,\ldots,i_p)\subset\{1,2,\ldots,n\}} f(X^{I_n})\mathbf{I}_{\{H^{I_n} \text{ is halving hyperplane}\}}.$$

So, it is easy to see, that $\hat{\theta}$ is in fact incomplete U-statistic (see for example [8]).

Let us suppose that X_is come from any continuous symmetric distribution with any center $\mathbf{C}$. In the following, we can suppose – without lost of generality – the $\mathbf{C}$ to be equal $\mathbf{0}$.

At the end of this section, let us recall the idea of interdirections. For every two points y_1 and y_2 from $\mathbf{R}^p$, let us denote by $C_{1,2}^{(n)}$ a count of interdirections associated with y_1 and y_2 in the sample $X_1, X_2, \ldots, X_n$. The **interdirection associated with** y_1 **and** y_2 in the sample $X_1, X_2, \ldots, X_n$ is every halving hyperplane H^{I_n} going through p points of the collection $X_1, X_2, \ldots, X_n$ which separates the points y_1 and y_2.

3 Results

At the beginning of this section, let us show, that in case of symmetric distribution, halving hyperplanes tends to the origin (in the sense, that the distance between the origin and the halving hyperplane tends to zero).

Proposition 1. *In case of symmetric distribution with the center in origin and none zero density in a neighbourhood of the origin, we have for any halving hyperplanes H^{I_n}:*

$$d(H^{I_n}, \mathbf{0}) \to 0 \text{ as } n \to +\infty.$$

Proof. Let us suppose that for every n_0 and for every $\varepsilon > 0$ there exists $n > n_0$ such that

$$d(H^{I_n}, \mathbf{0}) > \varepsilon.$$

In such a case, put R^{I_n} for the hyperplane which is parallel to H^{I_n} and goes through the origin. And denote by $R^{I_n}_+$ the space, such that for all $x \in R^{I_n}_+$ there is satisfied: $\det(x, Y^{I_n}, \mathbf{1}) > 0$, where Y^{I_n} is any fixed basis of the hyperplane R^{I_n}. Hence, due to the condition on symmetrization of the distribution, we know, according to Strong Law of Large Numbers (SLLN), that for every choice of I_n

$$\frac{\#(X_i \in R^{I_n}_+)}{n} \to \frac{1}{2},$$

where $\#(\cdot)$ denotes the cardinality of set $(\cdot)$.

Hence, for every $\varepsilon_1 > 0$ we can find n_1 such that for all $n \geq n_1$:

$$\left| \frac{\#(X_i \in R^{I_n}_+)}{n} - \frac{1}{2} \right| \leq \varepsilon_1.$$

So, we obtained, that for every $\varepsilon_1 > 0$ the number of observation between the hyperplanes R^I_n and H^I_n is for n large enough less or equal to $\varepsilon_1 + 1/n$. It is in contradiction with SLLN, requirement of non zero density in the neighbourhood of the origin and condition that the distance between hyperplanes R^{I_n} and H^{I_n} is bigger or equal to fixed ε. □

Proposition 2. *Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. observations with symmetric distribution (with the center at the origin) in $\mathbf{R}^p$, where $p \geq 2$.*

For any fixed $\mathbf{y}_1$ and $\mathbf{y}_2$ in $\mathbf{R}^p$, we denote by

$$\alpha(\mathbf{y}_1, \mathbf{y}_2) := \arccos\left(\frac{\mathbf{y}_1' \mathbf{y}_2}{||\mathbf{y}_1|| \cdot ||\mathbf{y}_2||} \right)$$

the angle between $\mathbf{y}_1$ and $\mathbf{y}_2$, and by $C^{(n)}_{\mathbf{y}_1,\mathbf{y}_2}$ the number of the interdirections associated with $\mathbf{y}_1$ and $\mathbf{y}_2$ in the sample $X_1, X_2, \ldots, X_n$.

Then,

$$\frac{(n-p+1)C^{(n)}_{\mathbf{y}_1,\mathbf{y}_2}}{\binom{n}{p}}$$

converges in quadratic mean to $\frac{1}{\pi}\alpha(\mathbf{y}_1, \mathbf{y}_2)$ as $n \to \infty$.

The proof is similar to a proof of Lemma 1 given by Hallin and Paindaveine in [5].

Proof. For $p = 2$. For every fixed point x_1 of the observation $(X_1, X_2, \ldots, X_n)$, we can find just one other point x_2 from the selection such that the hyperplane going through these two points (x_1, x_2) is a halving hyperplane (in the sense of our definition 1). So, if we have two fixed points y_1 and y_2 which are not contained in the selection, then we can estimate the number of interdirections associated with these two points in this sample as the whole part of half of number of points from the sample between the points y_1 and y_2. By the number of points from the sample between the points y_1 and y_2, we mean the number of observations between two lines – one of them goes through the origin and the point y_1 and second one goes through the origin and the y_2. (This estimation is possible due to the proposition 1).

So, (for n tends to infinity) the halving hyperplane H^{I_n} separates the points y_1 and y_2 in case that the corresponding $\mathbf{d}(I_n) \in (\mathbf{d}(y_1, \mathbf{0}), \mathbf{d}(y_2, \mathbf{0}))$, where $\mathbf{d}(y_1, \mathbf{0})$, resp. $v(y_2, \mathbf{0})$ is a vector whose components are cofactors of the first line in the matrix $(x, y_1, \mathbf{0}, \mathbf{1})$, resp. $(x, y_2, \mathbf{0}, \mathbf{1})$ (normal vectors to hyperplane given by y_1 and the origin, resp. by y_2 and the origin). Because we know, that the distance between halving

hyperplane and the origin tends to zero, it follows from Law of Large Numbers and due to symmetrization of the distribution, that the number of halving hyperplanes for which $\mathbf{d}(I_n) \in (\mathbf{d}(y_1, \mathbf{0}), \mathbf{d}(y_2, \mathbf{0}))$ (in the sense, that $\mathbf{d}(I_n)$ is a convex combination of$\mathbf{d}(y_1, \mathbf{0})$ and $\mathbf{d}(y_2, \mathbf{0})$, more precisely there exists $k \in [0, 1]$ such that $\mathbf{d}(I_n) = k\mathbf{d}(y_1, \mathbf{0}) + (1 - k)\mathbf{d}(y_2, \mathbf{0}))$ tends to $\frac{1}{\pi}\alpha(\mathbf{y}_1, \mathbf{y}_2)$.

For $p \geq 3$. Let us denote:

$$\mathbf{Q} := \{\mathbf{q} = (i_1, i_2, \ldots, i_p) : 1 \leq i_1 < i_2 < \cdots < i_p \leq n\},$$

let H_q be the hyperplane $H_{X_{i_1}, X_{i_2}, \ldots, X_{i_p}}$ such that $q = (i_1, i_2, \ldots, i_p)$ and let $\det(X, H_q, \mathbf{1})$ be the determinant from the matrix with rows $(X, 1), (X_{i_1}, 1), \ldots, (X_{i_p}, 1)$, where $i_j \in q$ for all $j = 1, 2, \ldots p$.

With this notation we get (in case of $n - p$ even):

$$\left(\frac{\binom{n}{p}}{n-p+1}\right)^{-1} C^{(n)}_{\mathbf{y}_1, \mathbf{y}_2} =$$

$$= \left(\frac{\binom{n}{p}}{n-p+1}\right)^{-1} \sum_{\mathbf{q} \in \mathbf{Q}} \frac{1}{2} \mathbf{I}_{\{\sum_{k=1}^{n} \operatorname{sign}(\det(X_k, H_q, \mathbf{1}) = 0)\}} \left(1 - \operatorname{sign}(\det(\mathbf{y}_1, H_q, \mathbf{1})) \cdot \operatorname{sign}(\det(\mathbf{y}_2, H_q, \mathbf{1}))\right)$$

and in case of $n - p$ odd

$$\left(\frac{\binom{n}{p}}{n-p+1}\right)^{-1} C^{(n)}_{\mathbf{y}_1, \mathbf{y}_2} =$$

$$= \left(\frac{\binom{n}{p}}{n-p+1}\right)^{-1} \sum_{\mathbf{q} \in \mathbf{Q}} \frac{1}{2} \mathbf{I}_{\{\sum_{k=1}^{n} \operatorname{sign}(\det(X_k, H_q, \mathbf{1}) = 1)\}} \left(1 - \operatorname{sign}(\det(\mathbf{y}_1, H_q, \mathbf{1})) \cdot \operatorname{sign}(\det(\mathbf{y}_2, H_q, \mathbf{1}))\right).$$

Now, we can see that we obtained an incomplete U-statistic (see [8]) with a symmetric kernel

$$\begin{aligned} g(X_1, X_2, \ldots, X_n) &= \\ &= \frac{1}{2} \left(1 - \operatorname{sign}(\det(\mathbf{y}_1, H_q, \mathbf{1})) \cdot \operatorname{sign}(\det(\mathbf{y}_2, H_q, \mathbf{1}))\right). \end{aligned}$$

And with a finite second moment. It implies that $\left(\frac{\binom{n}{p}}{n-p+1}\right)^{-1} C^{(n)}_{\mathbf{y}_1, \mathbf{y}_2}$ converges in quadratic mean to $\mathrm{E}g(X_1, X_2, \ldots, X_n)$ as n tends to infinity.

To complete the proof we need to show that

$$\mathrm{E}g(X_1, X_2, \ldots, X_n) = \frac{1}{\pi}\alpha(\mathbf{y}_1, \mathbf{y}_2).$$

Let us consider a canonical basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p\}$ in $\mathbf{R}^p$ such that $\mathbf{y}_1 = \mathbf{e}_1$ and $\mathbf{y}_2 := \lambda\mathbf{e}_1 + \mu\mathbf{e}_2$ for some $\lambda, \mu \in \mathbf{R}$. In such a case, we can see that

$$\begin{aligned} \mathrm{E}g(X_1, X_2, \ldots, X_n) &= \\ &= \mathrm{P}\left(\operatorname{sign}(\det(\mathbf{y}_1, H_q, \mathbf{1})) = -\operatorname{sign}(\det(\mathbf{y}_2, H_q, \mathbf{1}))\right). \end{aligned}$$

By A, we denote the subset in the space generated by $\mathbf{e}_1$ and $\mathbf{e}_2$, which is bounded by lines going through origin and projections of y_1 and y_2. Hence, according to lemma 1, we know that for every ε we can find n large enough such that with probability at least $1 - \varepsilon$ the distance of halving hyperplane from the origin is less or equal to ε. It is easily to seen that

$$\left|\mathrm{P}\left(\operatorname{sign}(\det(\mathbf{y}_1, H_q, \mathbf{1})) = -\operatorname{sign}(\det(\mathbf{y}_2, H_q, \mathbf{1}))\right) - \mathrm{P}(((\mathbf{e}_q)_1, (\mathbf{e}_q)_2) \in A\right| \leq$$

$$\leq \varepsilon + \mathrm{P}(((\mathbf{e}_q)_1, (\mathbf{e}_q)_2) \in A^{\varepsilon} \backslash A),$$

where A^ε is a subset of the space generated by $\mathbf{e}_1$ and $\mathbf{e}_2$ such that $x \in A^\varepsilon$ iff $d(x, A) \leq \varepsilon$. So, the right hand side of the inequality could be small enough.

Then, we need to compute

$$\mathrm{P}\left(((\mathbf{e}_q)_1, (\mathbf{e}_q)_2) \in A\right).$$

Using the same arguments as Hallin and Paindaveine we get that

$$\mathrm{P}(((\mathbf{e}_q)_1, (\mathbf{e}_q)_2) \in A) \to \frac{1}{\pi}\alpha(\mathbf{y}_1, \mathbf{y}_2)$$

for $n \to +\infty$.

□

From the proofs of previous propositions and the definition of the center given by halving hyperplanes, we can see, that the key step in proofs of asymptotics theorems for multidimensional quantiles is using of convergence theorem for incomplete U-statistics. The limit theorems for incomplete U-statistics based on i.i.d. data were studied in the seventies of the last century by many researchers. A lot of these results it is possible to find for example in [8].

However, there is a question, if it is possible to prove similar results on multidimensional quantiles in the case of dependent data, more precisely weakly dependent data. The reason for this research is, that in practice usage, we rarely have independent data, more often we handle with nonindependent data set. So, we need the theory of multidimensional quantiles for weakly dependent data, hence we need the asymptotics for U-statistics based on weakly dependent data set.

There are three main ways how to handle with weakly dependent data. The first possibility is to use some mixing conditions. Mixing conditions were introduced by Rosenblatt (see [13]) and are based on relations among σ-fields generated by random sequences. The problem of results based on mixing conditions is that it is quite difficult to verify their conditions, because they are very technical. On the other hand, there is a lot of results on asymptotics for U-statistics derived from mixing sequences, see for example [4], [14].

The second approach uses so-called associated random variables. The theory of associated random variables was introduced in sixties of the last century and it is still being developed by many researchers. Recently, there are published papers on asymptotics for U-statistics based on associated random variables, see for example [1].

The third approach studies the asymptotics of U-statistics which are based on stationary processes, without any requirement of satisfying any technical mixing condition. Such approach we can find in for example [3] and [9]. It seems to be very good approach – the proving techniques are more complicated then in the previous cases – but the conditions on dependent data, which must be verified, are quite easier than in above mentioned approaches.

Acknowledgements

Supported by Czech Science Foundation (project n. P201/11/P164).

The author would like kindly thank to Miroslav Šiman for his ideas, without which the paper can not be written.

References

[1] Beutner, E., Zähle, H.: Deriving the asymptotic distribution of U-and V-statistics of dependent data using weighted empirical processes. *Bernoulli*, **18(3)**, (2012), 803–822.

[2] Bradley, R., C.: *Introduction to strong mixing conditions*. Kendrick Press, Heber City, UT, 2007.

[3] Denker, M., Gordin, M.: Limit theorems for von Mises statistics of a measure preserving transformation. *arXiv preprint arXiv:1109.0635*, (2011).

[4] Dehling, H., Wendler, M.: Central limit theorem and the bootstrap for U-statistics of strongly mixing data. *Journal of Multivariate Analysis*, **101(1)** (2010), 126–137.

[5] Hallin, M. and Paindaveine, D.: Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *The Annals of Statistics*, **30(4)** (2002), 1103–1133.

[6] Oja, H.,Paindaveine, D.: Optimal signed-rank tests based on hyperplanes. *Journal of statistical planning and inference*, **135(2)** (2005), 300–323.

[7] Hallin, M., Paindaveine, D., Šiman, M.: Multivariate quantiles and multiple-output regression quantiles: From L1 optimization to halfspace depth. *The Annals of Statistics*, **38(2)** (2010), 635–669.

[8] Lee, A., J.: *U-statistics. Theory and Practice*. CRC Press, New York, 1990.

[9] Leucht, A., Neumann, M., H.: Degenerate U-and V-statistics under ergodicity: Asymptotics, bootstrap and applications in statistics. *Annals of the Institute of Statistical Mathematics*, **65(2)**, (2013), 349–386.

[10] Paindaveine, D., Šiman, M.: On directional multiple-output quantile regression. *Journal of Multivariate Analysis*, **102(2)**, (2011), 193-212.

[11] Paindaveine, D., Šiman, M.: Computing multiple-output regression quantile regions from projection quantiles. *Computational statistics*, **27(1)**, (2012), 29–49.

[12] Randles, R. H.: A distribution-free multivariate sign test based on interdirections. *Journal of the American Statistical Association*, **408(84)**, (1989), 1045–1050.

[13] Rosenblatt, M.: A central limit theorem and a strong mixing condition, *Proc. Nat. Ac. SC.* **42** (1956), 43-47.

[14] Wendler, M.: U-processes, U-quantile processes and generalized linear statistics of dependent data. *Stochastic Processes and their Applications* **122(3)** (2012), 787–807.

Representative firm's decisions and leading indicators

Miroslav Kľúčik[1]

Abstract. Leading indicators are potential carriers of signals of future changes in aggregate economic activity. They are mostly composed by means of empirical evidence found in development of aggregate time series. There is also a possibility to find their theoretical foundations by examining decisions of a representative firm in short run. The firm aims to minimize the producing costs for a given demand. The decisions involve change of employment, hours worked and stock of inventories according to the firm's expectations. The short run equilibrium is found by means of inter-temporal optimization. The evidence of lead is found by means of shock analysis.
The reduced form equations are utilized for empirical tests using available macroeconomic time series for EU-27 countries. The results show mixed evidence. While for almost every country there is one leading indicator found, the indicators proved the lead in average less than in half cases. The conclusions imply the theory contains evidence of leading indicators, which can be found also in real aggregate macroeconomic time series.

Keywords: leading indicators, economic theory, inter-temporal optimization.

JEL Classification: C62, D21, E32
AMS Classification: 35F20

1 Introduction

The concept of leading indicators arose in the first half of 20th century by empirical research at the National Bureau of Economic Research in the US [1]. It was based on co-movement of macroeconomic aggregate indicators, by definition, the so-called classical business cycles. Mainly, the analysis of business cycles was aimed at short-term economic forecasting, utilizing the lead of some variables against other found in actual data. The economic reasons (rationales) of the lead was systematically classified later by the OECD [2] or De Leeuw [3, 4] roughly into the early stage indicators, rapid response indicators, expectation sensitive indicators and prime movers.

The endeavour for investigation of economic theory behind the leading indicators was not pronounced for decades, probably because of their original empirical nature or, on the other side, their rather poor prognostic abilities [8] resulting into disinterest of most economists. The effort was mostly focused on the interpretation of observed business cycles; the basics of economic theory behind leading indicators are still to be found only in De Leeuw [3, 4]. The principal of profound business cycle research dealt with equilibrium economics based on microeconomic theory, such as real business cycle theory [6] and later New Keynesian economics [7]. De Leeuw used microeconomic equilibrium theory to provide some theoretical evidence of lead/lag behaviour of variables connected to short-term decisions of a representative firm for the three mentioned types of leading indicators - early stage indicators, rapid responsive indicators and expectation sensitive indicators [4]. To make a integration step of classical microeconomic analysis based on agents within general equilibrium economy (households, firms, government, monetary authority) and empirical indicators classified according the OECD a few links can be shed light upon.

Early stage indicators provide technical lead of some production processes against others, such as orders and actual production. This kind of indicators clearly relates only to firms, as shall be the only agent looked upon in this research paper. Rapid responsive indicators evidence the relatively rapid response of some macroeconomic variables to the development of aggregate economy, such as working hours fluctuations or delivery times in case of demand shocks. Also in this case the phenomenon is connected to the decisions of firms, which can alter working hours connected with lower costs in comparison with other decision variables such as employment (given the frictions on the labour market). The third type of leading indicators (expectation sensitive) is connected also with other classical agents, such as households and banks. These can behave differently based on their expectations that lead other major decisions of other agents on the market. Last type of leading indicator is

[1] VSB – Technical University Ostrava, Faculty of Economics, Department of Economics, Sokolská třída 33, 701 21 Ostrava, Czech Republic, miroslav.klucik@vsb.cz.

connected with more profound reasons of business fluctuations such as government and central banks decisions about policies applied on the market (in case of no-anticipation of these steps) or the terms of trade between several countries.

Subsequent analysis will try to build on De Leeuw's approach, clarify the computation steps and enrich it both in theoretical and empirical way. Two types of leading indicators are examined - rapid responsive indicators (hours worked, changes in stock of inventories) and expectation sensitive indicators (expected shipments). Firstly, by deriving of original formulas the short-term behaviour of firm in equilibrium is dubbed, by example in an alternate definition of behaviour. Secondly, an exhaustive way of sensitivity analysis of particular parameters of the reduced form model is provided. Finally, an empirical analysis follows, unlike the De Leeuw's US practical example; it considers EU countries' macroeconomic data.

The following text is organized into the methodology chapter (2nd Chapter) with description of differentiation of short run equilibrium equations for the firm and utilisation of the reduced form equations for simulation of shocks, which shall provide some evidence of lead/lag behaviour of variables. The 3rd Chapter uses the structure of theoretical equations for real-world application of leading indicators for the EU countries. In the conclusions some future possibilities of dealing with this topic is elaborated upon.

2 Theoretical analysis

The original approach of De Leeuw is based on a short-term dynamic firm behaviour according to the work of Holt, Modigliani, Muth and Simon [5]. The aim was to find some theoretical evidence of lead and lag in decisions of a producing firm. The decisions are therefore looked upon in a theoretical state of short-term equilibrium, i.e. the firm decides about input and output in an optimal way.

The basic assumption is that the demand is given, i.e. the buyers can not alter their behaviour in short run, and also the prices are given. The firm decides about the current level of employment, working hours, stock of inventories to minimize the costs of production in order to meet the given demand. De Leeuw gives two alternatives; firstly, the firm produces solely for stock, secondly, solely for orders. For the forthcoming analysis only the stock alternative is taken, basically because no unfilled orders data are available for most EU countries utilized later in empirical analysis. The cost minimization follows:

$$\min_{C_t,r} \sum_{i}^{\infty} \frac{C_{t+i}}{(1+r)^i} \tag{1}$$

where $$C_{t+i} = a_0 + a_1 E_{t+i}\bar{H} + a_2[E_{t+i}(H_{t+i} - \bar{H})]^2 + a_3[\bar{H}(E_{t+i} - E_{t+i-1})]^2 + a_4[(J_{t+i} - c_1 S_{t+i})]^2 \tag{2}$$

$$J_{t+i} = J_{t-1+i} + S_{t+i} - Q_{t+i} \tag{3}$$

and $$Q_{t+i} = c_2 E_{t+i} H_{t+i} \tag{4}$$

C_t are the costs of the producing firm in an infinite time $t+i$, the variable E_t represents the employment costs (wages), H_t the actual working hours, $\bar{H}$ the average hours worked. Parameters indicated $a_0 ... a_4$ and c_1, c_2 are representing the elasticity of changing the costs. The last element in the equation (2) represents the variable stock of inventories (finished goods J_t minus realized supplies S_t). In this case the a_0 parameter can represent some fixed costs, the parameter a_1 indicates elasticity of changing the overall labour input costs, a_2 the costs related to changing of working hours, a_3 costs due to employment changes and a_4 costs associated with decisions about shipments and production time. Last three elements of the equation indicate higher associated costs (squared elements) in the production process. The actual production is restricted by the classical production function, missing capital as input (4). Capital is assumed to be fixed in short time, therefore the firm does not decide about the capital accumulation.

Another variation, still occurring in short run (leading indicators shall provide evidence only short-term lead), might be to maximize the so-called gross operating surplus, i.e. the gap between the actual turnover and producing costs given the same constraint of production function. The decision parameters would be disaggregated into decisions about raw materials, different staff costs etc., but this elements are in the case above (equation 1) included in the stock of inventories variable. This approach would produce the same results.

The minimisation problem has been described above in equations (1) to (4). Basic algebra transformation offers two approaches to solve the problem. First is the use of Lagrangian technique, second, the direct substitution approach. The second approach can be applied in this case.

Looking at the equation (1)-(4) we have the following unknown variables: Q_t, E_t, H_t, S_t, J_t and r. The discount rate is taken as given with no changes in time. The equation (3) can be substituted into equation (2), which leaves out one unknown variable J_t. Also the variable H_t can be easily eliminated by solving equation (4) for H_t and substituted into equation (2). Then there are three variables that affect the level of costs in the short run - expenditures related to employment - number of employees, level of working hours and expenditures related to stock of production, expressed as the shipments S_t, which affect the stocks - J_t.

With some variables having different lag at time *t*, the optimization problem becomes an inter-temporal one. While decision variables can be accounted for infinite time, simplification assumption is made as in [4] that no expectations are created more than 2 periods ahead and the working hours fall back to the average level in two periods (however, these constraint does not shorten the response time of variables in the simulations). Each variable in equation (2) must be dealt as unknown in each period of time *t*, *t+1* etc. Therefore each time shift (with the simplifying assumption now only *t+2*) must be taken into account when computing the equilibrium. The following equations arise taking into account the substitutions made for J_t and H_t:

$$C_t = a_0 + a_1 E_t \bar{H} + a_2 \left[\frac{Q_t}{c_2} - E_t \bar{H}\right]^2 + a_3[\bar{H}E_t - \bar{H}E_{t-1}]^2 + a_4[(J_{t-1} + Q_t - S_t - c_1 S_t)]^2 \tag{5}$$

$$\begin{aligned} C_{t+1} = a_0 + a_1 E_{t+1} \bar{H} + a_2 \left[\frac{Q_{t+1}}{c_2} - E_{t+1} \bar{H}\right]^2 + a_3[\bar{H}E_{t+1} - \bar{H}E_t]^2 \\ + a_4[(J_{t-1} + Q_t - S_t + Q_{t+1} - S_{t+1} - c_1 S_{t+1})]^2 \end{aligned} \tag{6}$$

$$\begin{aligned} C_{t+2} = a_0 + a_1 \frac{c_1 SE_{t+2} - J_{t-1} - Q_t + S_t - Q_{t+1} + SE_{t+1} + SE_{t+2}}{c_2} \\ + a_3 \left[\frac{c_1 SE_{t+2} - J_{t-1} - Q_t + S_t - Q_{t+1} + SE_{t+1} + SE_{t+2}}{c_2} - \bar{H}E_{t+1}\right]^2 \end{aligned} \tag{7}$$

The equations (5) and (6) contain the substituted expression Q_t/c_2 for the expression E_t*H_t. Also the expression after a_4 contains the substitution for J_t according to (3). From the equation (7) follows that the elements after parameters a_2 and a_4 fell out because of mentioned simplification assumptions. Initially, also in the third equation it is needed to get rid of unknowns Q_{t+2}. By using substitution by solving for Q_t in (3), four unknowns remain according to (5)-(7): E_t, E_{t+1}, Q_t, Q_{t+1} and the variables S_{t+1} and S_{t+2}. These are to be understood as expected shipments, i.e. they shall be known in time *t* (this assumption follows [4]).

The minimum costs are then found by differentiation of equation (5) - (7) with respect to E_t, E_{t+1}, Q_t and Q_{t+1}, and by setting the derivatives equal to zero. The solutions are equivalent to De Leeuw [4]. The solved equations for the variable Q_t and E_t represent the actual decision of a firm about production and employment, using the equations (3) and (4) also the current hours worked and level of stocks can be calculated. The solutions for E_{t+1} and Q_{t+1} are representing the planned level of employment and production. The available four equations with for unknowns can be solved either in the matrix form or by substitution analytically. In case of 0 value of discount rate and value of 1 for parameters a_0 to a_4, c_2 and 2 for parameter c_1 (De Leeuw's parameters), the following two equations with two unknowns will emerge[2]:

$$Q_t = 0.077 + 0.115E_{t-1} - 0.654J_{t-1} + 1.577S_t + 0.500SE_{t+1} + 0.115SE_{t+2} \tag{8}$$

$$E_t = -0.231 + 0.404E_{t-1} - 0.288J_{t-1} + 0.519S_t + 0.404SE_{t+1} + 0.250SE_{t+2} \tag{9}$$

The optimal conditions for short-term decisions about costs are now known, the next step is to examine the decisions by using assumptions about parameters and initial levels of employment, working hours, stocks, production and shipments.

2.1 Shock simulations

De Leeuw uses the value 1 for parameters a_0 to a_4 and c_2, and 2 for c_1. The baseline values for simulations are set to 100 for initial value of employment, shipments, production, 200 for initial level of stock of inventories and 1 for average and initial working hours. In the analysis the *r* discount rate is set to 0 for all periods. Then shocks are presented for step changes in demand (shipments - S_t), changes in expectations (expected shipments S_{t+1} and S_{t+2}) and changes of all parameters (including productivity). However, an important fact must be emphasized. De

[2] Both results are similar to De Leuuw's results, which were stated as the baseline parameters.

Leeuw uses the assumption that the shipments in time *t* are not the actual shipments, but expected. In this analysis a more standard case is examined, i.e. the shipments in time *t* are known, which, as it will be seen later, has crucial impact on the results in comparison to De Leeuw's work.

The expectations are modelled similarly as in De Leeuw's paper. The first period expectations are taken as the expectations from previous period plus half of the error of the previous period forecast. The second period expectation is calculated as the forecast for 2 periods ahead from the first period plus half of the difference against the previous period expectations.

Before taking look at the simulations, the expected results regarding the leading indicators approach can be stated. It is expected that the hours worked and also stock of inventories, both as rapid responsive indicators, are going to lead the production and employment in case of their growth and decline turning points (observed high and low point of the time series). The case of expected shipments as leading indicator is not elaborated as it is solely dependent on assumptions.

The level of simulation shocks is assumed in same manner as in De Leeuw's paper, i.e. simulation shock raises the shipments (and expected shipments) in period 6 from 100 to 120 and reduces the shipments back in period 16 to 100 etc.

The Table 1 contains the results of all two shocks on formation and behaviour of each variable regarding the time shift of variables (lagging, coincident, leading) in the turning points (after changes in period 6 and 16).

Type of shock	Turning point 1		Turning point 2	
	H_t	J_t	H_t	J_t
Demand shock	C, L	C, L	C, L	C, L
Expectations shock	-	C, L	C, L	C, L

Table 1 Lead/lag relationship for shocks[3]

The results show that neither working hours nor change in inventories have proven to be leading the production, but on the other side both are leading indicators for the level of employment in all cases. This is a contrary result to the De Leeuw work, which shows that hours worked and stock of inventories are leading indicators in some cases. In the expectations shock the hours worked had shown no conformity with the production and employment, therefore it is omitted.

Sensitivity analysis is performed regarding the values of parameters and shock levels (only demand case shock). The relative stability proved in De Leeuw analysis has been confirmed. The lead/lag behaviour provided in Table 1 is found to be strong and without change in case of parameter changes a_1 to a_3 (tested range 0.5-1.5), and also in the case of c_1 parameter (range 1.0-3.0). Changes in c_1 lead to very different initial state of the system, but in time of the first and second turning point the relationships between the variables is without changes. However, sensitivity of parameter a_4 revealed changes in lead behaviour - at the bottom of tested range (below 0.55) the hours worked variable became leading also to the production (this is the only case where this leading indicator proved his abilities theoretically). The same results showed the sensitivity of c_2 parameter. Below the value of 0.72 the hours worked became leading also the production variable. This implies stronger role of changes in work intensity when sensitivity to changes in stocks is low. Finally, the changes in level of shocks (tested range -50 to 50) and also the changes in discount rate (tested range 0.01 to 0.99) had no impact of lead/lag relationships stated in Table 1. The Figure 1 shows the leading behaviour of hours worked at the case of 0.7 value of c_2 parameter from sensitivity shocks.

3 Empirical Analysis

The forthcoming analysis takes as its basis the trial to prove the leading ability of working hours, stock of inventories and expected shipments in real world data. Empirical analysis shall encompass one solely firm as the representation of an average producing firm in national economy (an agent). Some general assumptions about the market in which the firm exists need to be stated. The labour and the produced good in the model are regarded homogenous. Other assumptions are the same as in the case of theoretical analysis, i.e. in the short term the prices and demand are given, and the firm cannot acquire capital. As no data about shipments are available, the firm is taken to be an exporter, i.e. the export of goods can be regarded as sufficient evidence of a shipment.

[3] C - coincident relationship, L - leading relationship.

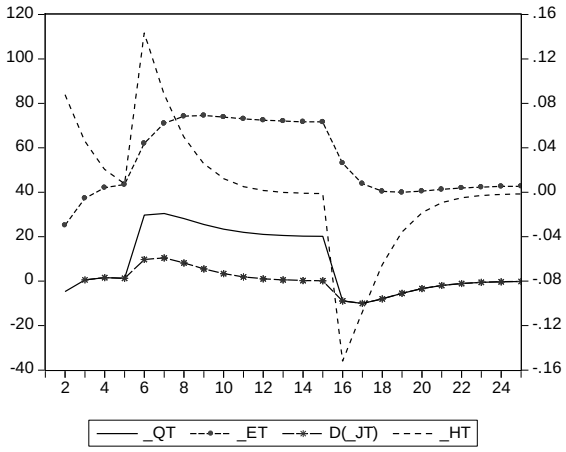


Figure 1 Shock observation and lead of hours worked at observation No. 6 and 16

The production is represented by industrial production index, average employment and working hours are also available for industry. The shipments are introduced through real exports of goods, while expectations can be approximated by expected export orders from business tendency surveys in industry branch. As stock of inventories is available for countries only at quarterly frequency this time series is imitated by estimated stock of inventories in industry, also available from business tendency surveys. All time series are available in monthly frequency. Several steps of transformation are needed for later use in standard regression. Firstly, the time series must be seasonal adjusted (EViews) then their stationarity is ensured through year-over-year transformation. The time series from business tendency surveys are available as balances (qualitative aggregates), and are taken in original form as deviations from trend (the so-called deviation cycle), therefore they do not need to be transformed into year over year growth rates. All time series are smoothed by Henderson Moving Average (X12-ARIMA method) available in EViews, while the smoothing parameter is by default chosen automatically according to the particular time series' variance.

To enable comparison between theoretical and empirical results the original regression is taken as the initial point (see equation 8 and 9), adding hours worked to the equation. Crucial is the observation of sign of each parameter, the lag is taken the same as in the theoretical analysis (1 month, 2 months for expectations). Time series mentioned in previous paragraph are available for 19 European countries from about 2001 until December 2012 (from data sources of Eurostat and National Statistical Institutes).

Country	Qt						Et					
	b_1	b_2	b_3	b_4	b_5	b_6	b_1	b_2	b_3	b_4	b_5	b_6
Bulgaria	-	-	-0.53	0.21	-	1.35	-0.06	0.98	0.06	0.02	-	-
Cyprus	0.56	0.59	-0.34	0.09	0.06	-	0.02	0.97	-0.01	0.01	0.01	
Czech R.	0.71	-0.69	-	0.52	0.22	0.23	0.06	0.85	-	0.03	0.02	0.05
Estonia	-0.13	0.20	-	0.57	-	0.36	0.04	0.95	-0.02	0.03	-	-
Finland	0.38	0.14	-0.09	0.54	0.04	-	0.05	0.92	-0.02	-	-	0.05
France	-	0.37	-0.10	0.72	-	-	0.03	0.93	-0.01	0.02	-	0.03
Germany	0.25	-0.21	-	0.65	-	0.31	0.08	0.91	-0.03	-	-	0.03
Greece	1.20	0.33	-0.62	0.04	-	-	0.08	0.97	-0.05	-	-	-
Hungary	-	-0.23	-	0.73	-	0.49	0.06	0.63	-	0.04	-	0.26
Latvia	0.14	0.25	-	0.59	-	-	-	0.83	-	0.10	-	0.06
Lithuania	0.24	0.25	-	0.23	-	0.30	0.05	0.96	-0.02	0.02	-	-
Netherlands	2.60	-1.62	-0.73	0.06	0.18	0.52	0.03	0.97	-0.01	0.01	-	-
Poland	-0.15	0.73	-	0.43	-	-	-	0.97	-	0.02	-	-
Romania	0.79	-1.72	-	-	0.14	1.81	0.10	1.01	-0.15	0.01	0.03	-
Slovenia	-0.63	-	-	0.43	0.09	1.10	-0.07	0.93	-	0.03	-	0.09
Slovakia	-	0.55	-	0.48	-	-	0.02	0.81	-	0.03	0.01	0.12
Spain	0.95	0.29	-0.54	0.29	-	-	0.11	0.93	-0.07	0.02	-	-
Sweden	0.12	0.13	-	0.74	-	-	0.08	0.89	-0.02	0.05	-	-
UK	1.36	-	-0.45	0.12	-	-	-	1.01	-0.01	-	-	-

Table 2 Regression results for empirical analysis

In the regression runs only variables with significant parameters are retained. For both production equation (8) and employment equation (9) the results are summarized in Table 2. The parameters are marked with b_1 (constant), b_2 (employment), b_3 (stock of inventories), b_4 (exports - shipments), b_5 (expected shipments - exports) and b_6 (hours worked).

No leading indicator is found for production for Latvia, Poland, Slovakia and Sweden. For employment this is the case for Poland. In the theoretical analysis the stock of inventories and hours worked showed to be leading for both production and employment in all cases. The empirical analysis agrees with this fact in 13 cases for stock of inventories and 8 cases for hours worked. The theory in previous chapter implied hours worked as leading indicator of production only in certain range of parameter values, while the same mixed results are evidenced in the empirical analysis. Table 2 shows that stock of inventories, expected shipments and hours worked are leading in less than half cases. All three indicators are significant only for industrial production of Netherlands.

The conclusions stated in Table 2 serve only for comparison between empirical and theoretical expected values. Given the differences between sectorial specialization of each country and different laws, one cannot truly compare the results of countries themselves. However, the market frictions in each country can manifest in different lead.

4 Conclusions

According to the theoretical search for an equilibrium in short run a firm does have an optimum connected to decisions about employment, hours worked and stock of inventories according to expectations, present at the point of minimal costs. The shocks applied to the reduced form equation imply the existence of leading behaviour in case of stock of inventories and hours worked in all cases for employment and in some levels of parameters for hours worked in case of production. The empirical analysis showed that except in four cases out of 40, each dependent variable has found its leading indicator among stock of inventories, expected shipments and hours worked. This proves the theoretical foundation provided, while simultaneously the empirical analysis shows mixed evidence for leading indicators varying among different countries.

The weak spot of the theoretical assumptions is the creation of expectations, which must be stated directly, but despite this fact the leading behaviour emerges as was proved in this paper. However, some new concepts shall be developed regarding the expectation creation, i.e. based on microeconomic theory. Also some cases of widening the period analysed shall be examined.

Acknowledgements

This work was financially supported by the Student Grant Competition of Technical University of Ostrava, Faculty of Economics, within the project SP2013/91 “Modelling the impact of cyclical development on a small open economy (theoretical concepts)”.

References

[1] Burns, A., and Mitchell, W.: *Measuring Business Cycle*. Studies in Business Cycles Vol. 2, National Bureau of Economic Research, New York, 1946.

[2] Gyomai, G., and Guidetti, E.: OECD system of composite leading indicators, OECD, 2012. Retrieved from: http://www.oecd.org/std/41629509.pdf. Accessed 14 December 2012.

[3] De Leeuw, F.: Leading Indicators and the ‘Prime Mover’ View. *Survey of Current Business* **69** (1989) 8, 23–29.

[4] De Leeuw, F.: Towards a theory of leading indicators. In: *Leading Economic Indicators: New Approaches and Forecasting Records* (Lahiri, K., and Moore, G. D., eds). Cambridge University Press, Cambridge, 1991, 15-56.

[5] Holt, C. C., Modigliani, F., Muth, J. F., and Simon, H. A.: *Planning Production, Inventories, and Work Force*, Prentice-Hall, Englewood Cliffs, New Jersey, 1960.

[6] Kydland, F. E., and Prescott, E. C.: Time to Build and Aggregate Fluctuations. *Econometrica* **50** (1982) 6, 1345-1370.

[7] Rotemberg, J. J., and Woodford, M.: Oligopolistic Pricing and the Effects of Aggregate Demand on Economic Activity. *Journal of Political Economy* **100** (1992) 6, 1153-1207.

[8] Stock, J. H., and Watson, M. W.: Introduction to "Business Cycles, Indicators and Forecasting", NBER Chapters. In: *Business Cycles, Indicators and Forecasting*. National Bureau of Economic Research, Inc., Cambridge, MA, 1993, 1-10.

Demographic ageing and its spatial consequences

Renata Klufová[1], Marek Šulista[2]

Abstract. This paper deals with an evaluation of demographic ageing in the Czech Republic and its spatial consequences. A regional analysis of the basic indicators of the economic demography (the age dependency ratio, the economic dependency ratio, the green dependency ratio, and the grey dependency ratio) comprises the main part of this article. The presented analysis can serve as a basic precondition for regional policy recommendations.

Keywords: demographic ageing, factor analysis, spatial autocorrelation, hot spots analysis

JEL classification: C44
AMS classification: 90C15

1 Introduction

Demographic ageing has affected more or less all countries in the world since the second half of the 20th century and has become one of several global problems. According to long-term population projections, the proportion of the elderly is going to increase, and this fact is often mentioned when talking about various social problems, such as pension systems and the increase of social welfare and health care costs.

Ageing, in the demographic sense, relates to the whole population. Only an increase in the proportion of young age groups can make a population that is getting older become younger. Demographic ageing is caused by changes in the character of demographic reproduction – changes in the share of the youthful and post-reproduction parts of the population. These changes could be caused by two factors. The first one is a relative slowdown of the increase in younger age groups, which is mostly a result of a decrease of birth and fertility rates. This type of ageing is called "relative ageing". The second one is an acceleration in the growth of older age groups, which is an outcome of a greater decrease of mortality rates in the higher age ranges. It results in an increase of life expectancy. This type of ageing is known as "absolute ageing". Both types usually occur at the same time.

The Czech Republic is one of the countries experiencing the demographic ageing intensively. According to the up-to-date prognosis [4], the population development of the Czech Republic to 2070 will have the following features:

- The interim period of the population growth by natural increase ends in 10–12 years. This period has been caused by the concurrence of low mortality and the realization of the transformed and postponed fertility. The future will depend on an amount of migration rate.
- The Czech population will further get older. The important factors will also be an expected decrease of fertility and a spin out of the human life length. The Czech population will firstly get older by absolute ageing. The relative ageing will also affect our ageing because changes of lower and higher fertility are expected.
- The expected migration will weaken the influence of anomalies in age structure.
- Expected migration gains will lead to a preservation of economic potential (labour force) in a medium-term perspective. Then we can expect a drop of this potential in a long-term perspective.

Thus, the appraisal of regional differences in the demographic ageing and its consequences is important. The found difference can be used as a recommendation for further regional policies.

[1]University of South Bohemia/Faculty of Economics, Department of Applied Mathematics and Informatics, Studentská 13, 370 05 České Budějovice, e-mail: klufova@ef.jcu.cz
[2]University of South Bohemia/Faculty of Economics, Department of Applied Mathematics and Informatics, Studentská 13, 370 05 České Budějovice, e-mail: sulista@ef.jcu.cz

2 Material and Methods

Three economic generations comprise the age structure of the population:

- The pre-productive age: 0–19 years (Economic generation I),
- The productive age: 20–64 years (Economic generation II),
- The post-productive age: 65 years and older (Economic generation III).

Usually, the age of 15 years is used as the boundary for children in the literature, but many young people are not productive beyond this age, as they continue with their education. Thus, we used the age of 19 years as the lower boundary of the productive age which reflects better the current situation. We tried to include accessible characteristics which influence the demographic development, especially changes in the age structure. With respect to the availability of statistics sources, the basic characteristics of the economic population structure and some other chosen indicators were used for the evaluation of the Czech demographic ageing: the Economic dependency ratio (EDR) – a relation all of economic generations to Economic generation II (it shows how many people have to support one person in the productive age), the Green dependency ratio (GDR) – a relation of Economic generation I to the productive generation, the Grey dependency ratio (GYDR) – a relation of Economic generation II to the productive generation, the Ageing index (AI) – a relation of the post-productive age to children aged 0–15 years, the Mean age (MA), the Infant mortality rate (IMR) – death of a child less than one year of age per 1 000 live births, the Crude nuptiality rate (CNR) – the number of marriages per 1 000 inhabitants, the Crude divorce rate (CDR) – the number of divorces per 1 000 inhabitants, the Crude birth rate (CBR) – the number of births per 1 000 inhabitants, the Crude mortality rate (CMR) – the number of deaths per 1 000 inhabitants, the Incapacity for work (IW) – the notified cases of incapacity for work (cases per 100 sickness insur.), the Crude rate of the net migration (CMS) – migration saldo per 1 000 inhabitants, the cause-specific mortality – the mortality rate from a specified cause for a population per 100 000 inhabitants: cancer (C), cardiovascular disease (CV), breathing system (BS), external causes (EC); the density of population (DP) – the number of inhabitants per km^2, proportion (%) of inhabitants living in towns (IT), the proportion of recipients (OAP) of old-age pensions (in all the pensions), the proportion of foreigners (WF) with working permit (in all foreigners registered by employment bureau), the density of road system (DRS), the completed dwellings (CD) per 1 000 inhabitants, the unemployment rate (RU), the number of registered economic subjects (RES)per 1 000 inhabitants.

The data was obtained from the Czech Statistical Office and the Institute of Health Information and Statistics of the Czech Republic.

The number of characteristics led us to use the factor analysis. It is a method used to describe variability between observed, correlated variables in terms of a potentially lower number of unobserved variables called factors, eg. [5],[3], [10]. There are various rotational strategies that have been proposed. The goal of all of these strategies is to obtain a clear pattern of loadings, that is, the factors that are somehow clearly marked by high loadings for some of the variables and low loadings for the others. This general pattern is also sometimes referred to as the simple structure. We used the software Statistica for this analysis. The analysis was conducted for two years, 2006 and 2011, and then the results were compared in order to describe changes and dynamics of the demographic ageing in the Czech Republic. The main factors were extracted using the method of the principal components and then rotated.

The spatial distribution of demographic development was evaluated by Moran index of spatial autocorrelation I which we applied on the factors strongly correlated with the Economic Dependency Ratio:

$$I = \frac{n}{S_0} \cdot \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \mu)(x_j - \mu)}{\sum_{i=1}^{n} w_{ij}(x_i - \mu)^2}, \tag{1}$$

where w_{ij} is a spatial weight between feature i and j, n is equal to total number of features and S_0 is aggregate of all the spatial weights:

$$S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}. \tag{2}$$

The spatial autocorrelation is characterized by a correlation in a sign among nearby locations in space. According to Fotheringham et. al [9], the spatial autocorrelation is more complex than the one-dimensional autocorrelation because the spatial correlation is multi-dimensional (i.e. 2 or 3 dimensions of space) and multi-directional. The computation of Moran's I is performed by dividing the spatial covariation by the total variation. The resulting values are in the range approximately from -1 to 1. The positive sign represents a positive spatial autocorrelation, while the converse is true for the negative sign. Zero result represents no spatial autocorrelation [2],[13].

In case of the global spatial autocorrelation, the null hypothesis states that "there is no spatial clustering of the values associated with the geographic features in the study area". When the p-value is small and the absolute value of Z score is large enough that it falls outside of the desired confidence level, the null hypothesis can be rejected. If the index value is greater than 0, the set of features exhibits a clustered pattern. If the value is less than 0, the set of features exhibits a dispersed pattern [12],[13].

The global Moran statistic indicates a tendency to clustering but cannot identify types of the clusters (high or low values). It is suitable to use a local statistic to identify clusters of high values (hot spots) or low values (cold spots)[1]. The Hot Spot Analysis calculates the Getis-Ord G_i^* statistic for each feature in a dataset. The resultant Z score indicates where the features with either high or low values cluster spatially. This tool works by looking at each the feature within the context of neighbouring features. A feature with a high value is interesting, but may not be a statistically significant hot spot. To be a statistically significant hot spot, a feature will have a high value and be surrounded by other features with high values as well [13]. The local sum for a feature and its neighbours is compared proportionally to the sum of all features; when the local sum is much different than the expected local sum, and that difference is too large to be the result of random chance, a statistically significant Z score results.

$$G_i^* = \frac{\sum_j w_{ij} d(x_j)}{\sum_j x_j}. \tag{3}$$

Various approaches to conceptualize spatial relationships (neighbourhood of the features used in the calculations). Three of them were tested in our analysis for the hot spots identification of the main factors: the fixed distance, the contiguity and K nearest neighbours. These concepts are described, for example, in [12] or [7].

3 Results and Discussion

The main factors were extracted using the method of the factor analysis. Six factors were chosen with eigenvalues greater than 1 according to the scree plots in both analysed years. The chosen factors explain 75.97% (2006), respectively 71.42% (2011) of the total variance. Tables 1 and 2 show factor loadings. Thus, they represent the most important information on which the interpretation of the factors is based.

We can try to interpret the factors influencing demographic development, with a certain amount of caution, in the following way: the main component "age structure" is the same for both years. The significance of the third factor in 2006, which can be labelled as "urbanization and business [1]", has become stronger in 2011 (as it can be seen in Table 3), whereas the second factor in 2006, "migration", lost its importance in 2011. Mortality played a significant role in demographic development in 2006. Its importance is still decreasing, which corresponds with the main trends discussed in the literature. Many authors use the term "mortality transition" [6],[8]. Immigrants working in the Czech Republic (factor 5 in 2006 and factor 6 in 2011) are also important for the economy in relation to demographic ageing. It also corresponds with various demographic prognoses which consider immigration as one of the future sources of the workforce [15].

The demographic development and its economic consequences are also influenced by particularities of "urban environment" (the population density, a higher crude divorce rate, some causes of mortality connected with the urban environment) which have emerged at the end of the analysed period. The "external causes of mortality" represents another factor of the demographic development playing an important role at the beginning of the period. It corresponds to the above mentioned features of the population development in the Czech Republic till 2070 [4].

[1] EV ... eigenvalue, CPTV ... cumulative percentage of total variation

variable	F1	F2	F3	F4	F5	F6
RU06	-0.49	**0.62**	0.10	0.35	0.03	-0.02
GDR06	**-0.72**	0.13	-0.50	-0.20	-0.02	0.23
GYDR06	**0.92**	0.06	-0.14	0.09	-0.08	0.20
AI06	**0.96**	0.07	0.14	0.12	-0.06	0.03
MA06	**0.94**	-0.12	0.16	0.22	0.01	0.04
IW06	-0.07	0.13	0.24	0.06	**0.72**	-0.01
CNR06	-0.19	-0.43	**0.69**	0.05	0.10	0.09
CDR06	-0.46	-0.19	**0.59**	0.29	0.32	-0.06
CBR06	-0.36	**-0.70**	0.34	-0.06	0.03	0.16
CMR06	0.49	0.06	-0.04	**0.79**	0.13	-0.15
IMR06	-0.29	0.31	0.07	**0.52**	0.02	0.14
CMS06	-0.07	**-0.89**	0.00	0.09	-0.03	-0.03
C06	-0.48	-0.11	0.39	**0.50**	0.09	0.11
CV06	0.31	0.11	-0.30	0.70	-0.25	0.04
BS06	0.49	0.25	0.04	-0.20	0.35	0.47
EC06	-0.09	0.26	-0.07	0.15	0.33	**-0.63**
WF06	0.00	0.00	-0.04	-0.05	**0.86**	-0.07
CD06	0.10	**-0.83**	0.05	-0.17	-0.26	0.10
DRS06	0.10	**-0.53**	-0.09	0.34	0.22	0.51
IT06	-0.02	0.27	**0.82**	0.08	0.24	0.03
PD06	0.34	0.09	**0.78**	-0.02	-0.07	-0.06
RES06	0.36	-0.44	**0.54**	-0.23	0.03	-0.30
OAP06	0.22	-0.16	**0.56**	-0.24	-0.14	0.28

Table 1: Factor loadings 2006

The next step of our analysis was spatial analysis of the main factors. We tested the hypotheses about their (global) spatial autocorrelation and tried to identify hot spots. Table 4 shows correlations of the Economic dependency ratio with the main factors. The values bolded represent significant correlations.

When calculating Moran statistics for highly correlated factors from the Table 4, we found out that all of them show a significant positive spatial autocorrelation, e.g. a tendency for clustering in space. We tried to identify clusters of high/low values (hot spots) in the next step. This part of the analysis was realized with the software GeoDA.

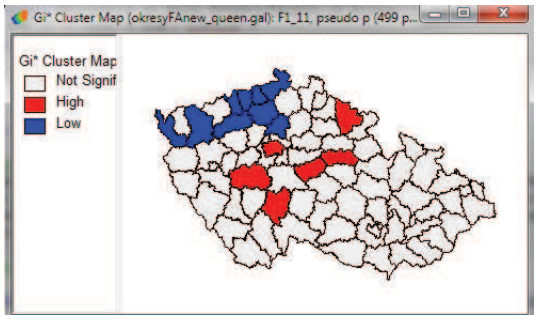


Figure 1: Hot spots - factor 1 - age structure - 2011

Figure 1 shows hot and cold spot of the age structure[2] in 2011 which corresponds with recent differences between districts: the districts in North Bohemia have more favourable age structure, but again there are some districts in the inner parts of the country, mainly along the borders of the administrative regions which have unfavourable age structure. In the Czech Republic, the inner peripheries are usually the

[2]Hot spots found out in 2006 are almost identical.

variable	F1	F2	F3	F4	F5	F6
RU11	-0.13	-0.23	0.31	**-0.57**	-0.52	-0.24
GDR11	**-0.66**	-0.42	-0.21	-0.15	0.37	-0.03
GYDR11	**0.91**	0.12	-0.22	-0.02	0.15	-0.07
AI11	**0.94**	0.23	-0.08	0.02	-0.11	-0.06
MA11	**0.94**	0.17	-0.01	-0.01	-0.18	-0.06
IW11	0.02	-0.17	0.22	**0.76**	0.08	0.08
CNR11	-0.24	0.18	0.01	**0.67**	0.06	0.03
CDR11	-0.31	0.12	**0.67**	0.24	0.07	0.10
CBR11	-0.52	0.30	-0.12	0.16	**0.62**	0.09
CMR11	**0.73**	-0.27	0.45	-0.17	-0.06	-0.08
IMR11	-0.12	0.06	0.49	-0.29	0.12	**-0.50**
CMS11	-0.53	0.12	-0.16	0.04	**0.64**	0.23
C11	0.39	-0.05	**0.69**	0.07	-0.07	0.16
CV11	**0.63**	-0.33	0.08	-0.40	0.02	0.14
BS11	0.32	-0.09	0.10	0.11	0.01	**-0.53**
EC11	0.12	-0.05	-0.07	-0.09	-0.27	-0.45
WF11	0.09	0.17	0.15	0.09	0.05	**0.71**
CD11	-0.39	0.12	-0.32	0.17	**0.71**	0.24
DRS11	0.10	0.00	0.29	0.00	**0.80**	-0.09
IT11	0.15	0.57	**0.61**	0.02	-0.20	-0.12
PD11	0.19	**0.79**	0.10	0.05	0.00	0.03
RES11	0.01	**0.81**	-0.02	0.22	0.16	0.18
OAP11	-0.06	0.66	-0.03	-0.26	0.23	0.18

Table 2: Factor loadings 2011

	2006			2011		
factor	EV	CPTV	interpretation	EV	CPTV	interpretation
F1	5.22	21.76	age structure	6.19	26.89	age structure
F2	4.55	40.73	migration	3.50	42.12	urbanization, business
F3	3.48	55.25	urbanization, business	2.25	51.94	urban environment
F4	2.14	64.17	internal causes of mortality	1.70	59.32	incapacity for work
F5	1.57	70.73	working foreigners	1.55	66.06	migration
F6	1.26	75.97	external causes of mortality	1.18	71.19	working foreigners

Table 3: Main factors of demographic development - comparison 2006 and 2011

year	F1	F2	F3	F4	F5	F6
2006	**0.30**	**-0.67**	0.16	-0.06	-0.12	**0.54**
2011	**0.46**	-0.09	**-0.38**	-0.04	-0.04	**0.45**

Table 4: Correlations of Economic dependency ratio and factors – comparison 2006 and 2011

peripheral zones of metropolitan areas and regional centre areas. In the mid 1990s, the populations stopped declining in some peripheries as a result of the sub-urbanisation processes, but in the other peripheries, the depopulation processes continued. This last category of the inner peripheries can be described as the hard core of Czech peripheral regions and according to Musil and Muller [14] they warrant the development of specific regional policy measures, stressing the creation of new jobs, the improvement of public transport, greater accessibility of service centres, and co-operation among communities.

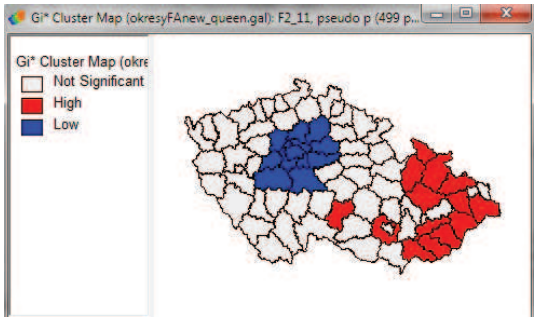


Figure 2: Hot spots - factor 2 - migration - 2006

Figure 2 shows hot and cold spots of migration in the Czech Republic in 2006. The sub-urbanization processes clearly crystallized and after the year 2000, it has become a decisive factor influencing migratory relations within the whole country. Vobecká (2010) stipulates that the sub-urbanization process has been, since 1995, clearly pronounced around primary centres and since 2001, also around smaller, secondary centres. These are so called residential sub-urbanization when people move out of the core cities but their still work there. The improving situation on the Czech housing market speeded up the whole sub-urbanization process. This is the period, in which de-concentration processes (sub-urbanization and partly de-urbanization) started being more important. According to Vobecká (2010), between the years 1995–2006, the most migratory attractive municipalities can be found mainly in close surroundings of big cities (namely Prague, Brno, and Pilsen) and in areas, which are closely tied to advantages offered by the city (within a reasonably short commuting distance) and, at the same time, which have better quality of the environment. On the other hand, there are regions with ecological and economical problems caused by emigration (the north parts of Moravia and areas at the Slovak border).

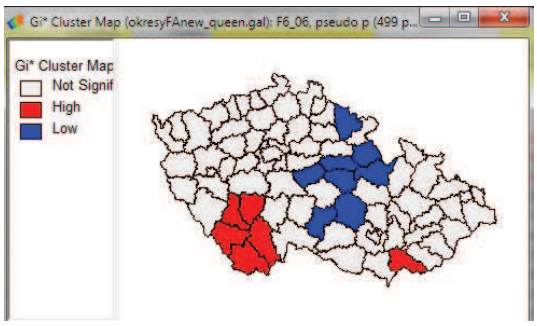


Figure 3: Hot spots - factor 6 - working foreigners - 2011

Figure 3 contains hot spots identified mainly on the basis of foreigners legally working in the Czech Republic (foreigners registered with the Job Center). They provide a picture of the spatial distribution of their activities.

Analysing the differences in the main factors among the regions, we can propose some measures of regional policy aimed at a reduction of the effects of demographic ageing. A detailed spatial analysis of the main factors could serve, for example, as a basis for suggestions on how to improve the Economic Dependency Ratio according to regional conditions.

4 Conclusion

The demographic profile of a region is usually seen as a slowly changing background phenomenon in the analysis of regional competitiveness and regional growth. However, regional demographic change can have a significant impact on regional competitiveness and such a change is often more rapid and profound than those at the national level. In turn, regional population size, growth, composition and distribution are endogenous to regional economic development.

Using spatial statistics, we proceeded to spatial demography. The aim of the paper was to show the usefulness of spatial analysis for regional policy makers. However, the results of any statistical analysis are always limited by the availability of suitable statistical data.

References

[1] Anselin, L.: Local Indicators of Spatial Association - LISA. *Geographical Analysis*, 27 (2),(1995),93–115.

[2] Bailey, T., C. and Gatrell, A., C.: *Interactive Spatial Data Analysis*. Prentice Hall, Harlow, 1995.

[3] Brown, B., L. et al.: *Multivariate Analysis for the Biobehavioral and Social Sciences: A Graphical Approach*. John Wiley & Sons Inc., New York, 2012.

[4] Burcin, B., and Kučera, T.: Prognóza populačního vývoje České republiky na období 2008-2070. In: *Demografická situace České republiky: proměny a kontexty 1993-2008*. (Bartoňová, D., and kol.), Praha, Sociologick nakladatelstv SLON, 2010, 181-212

[5] Clarkson, D., B., and Jennrich, R., I.: Quartic rotation criteria and algorithms, *Psychometrika*, **53** (1988), 251–259.

[6] d'Albis, H. and Lau, S., P. and Snchez-Romero, M.: Mortality transition and differential incentives for early retirement. *Journal of Economic Theory*, **147 (1)**, 2012, 261-283.

[7] Fischer, M., M. and Getis, A.: *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*. Springer, Heidelberg, 2010

[8] Fehr, H. and Jokisch, S. and Kotlikoff, L., J.: Fertility, mortality and the developed worlds demographic transition. *Journal of Policy Modeling*, **(30)**, 2008, 455-473.

[9] Fotheringham, A., S. and Brundson, Ch. and Charlton, M.: *Quantitative Geography: Perspectives on Spatial Data Analysis*. SAGE Publications, London, 2002.

[10] Harris, R., J.: *A Primer of Multivariate Statistics*. Lawrence Erlbaum Associates, Publishers, New Jersey, 2001.

[11] Johnson, R., A., and Wichern, D., W.: *Applied Multivariate Statistical Analysis*. Prentice Hall,New Jersey, 2002.

[12] Lloyd, Ch., D.: *Spatial Data Analysis: an introduction for GIS users*. Oxford University Press, New York, 2010.

[13] Mitchell, A.: *GIS Analysis. Volume 2: Spatial Measurements & Statistics*. The ESRI Press, Redlands, 2005.

[14] Musil, J. and Muller, J.: Vnitřní periferie v České republice jako mechanismus sociální exkluze. *Sociologický časopis/Czech Sociological Review*, **44 (2)**, 2008, 321-348.

[15] Poot, J.: Demographic change and regional competitiveness: the effects of immigration and ageing. *International Journal of Foresight and Innovation Policy*, **4 (1-2)**, 2008, 129–145.

[16] Vobecká, J.: *Spatial dynamics of the population in the Czech Republic*. 1989-2007. Ph.D. thesis plus its synopsis. Charles University and Universit de Bourgogne. Prague and Dijon.

Eliminating the ambiguity of composite indices in continuum economics

Jan Kodera [1], Tran Van Quang [1]

Abstract. The contemporary mainstream macroeconomics works on the assumption of a continuum of producers in which each of them produces just one differentiated good. Under this assumption, producers or goods are numbered by real numbers from interval [0,1] and it seems to be a good approximation of monopolistic competition economic system. But design of composite indices of production or consumption makes problems of ambiguity. As the usual way of aggregation is using Lebesgue integral, it raises questions like whether the integral exists and whether the conditions we impose have any meaningful economic interpretation. It turns out that these problems are difficult to solve and the answers, if found, are unsatisfactory. Even if the integral exists and a composite index for aggregating the size of economic quantities is available, the question of its ambiguity remains unanswered.

Keywords: DSGE models, continuum, aggregation, composite index, ambiguity, Lebesgue's integral

1 Introduction

The notion continuum in Mathematics has penetrated into Economics for more than a half of a century. Its usage is two-fold. First, to put it simply, it is used to indicate the arbitrary divisibility of an economic quantity. For example, to say capital is a continuum means that it can be arbitrarily any real number. In this sense, there is no ambiguity and its correctness is never disputed. But there is another usage of this notion in Economics which is related to the number of varieties of commodities or to the number of agents in an economy. Aumann [1] probably was the first author who uses the notion continuum in this context. In his work, he considers a continuum of agents in market. It is apparent that the number of agents in a market is finite and the idea of an uncountable set of agents does not coincide with reality and therefore it is refuted. But on the other hand, as in natural sciences, the notion a continuum of whatever is a useful modeling tool for analysis.

Dornbusch, Fischer and Samuelson [4] have further developed the original Heckscher-Ohlin $2 \times 2 \times 2$ model [7] in which they use a continuum of types of goods. They have drawn some interesting results which arise from the presence of a continuum of goods. But on the other hand, they also point to possible ambiguity resulting from the continuum assumption. They maintain that the results obtained from a model with a continuum of goods may differ from the ones drawn from a model with a finite number of types of goods. Later, New-Keynesian dynamic stochastic general equilibrium theory, while assuming a finite number of households, also frequently uses continuum of agents (producers of goods and services) and goods (see Gali [5], McCandless [8], Walsh [9]). For New-Keynesian theory the existence of a continuum of agents-producers and a continuum of sorts of goods is essential as important macroeconomic equations are drawn from the micro-economic foundation, i.e. from individual behavior of agents in economy. In order to do so, New Keynesian theorists need to aggregate the behavior of individual agents at the micro-economic level to the macro-economic level by integrating across those continua.

While the authors of DSGE models take the validity of such integration for granted, in our opinion, the validity is not so straightforward. Rather, it can be a stumbling block, if we do not handle properly, it can disprove the validity of the whole model. So far, to our knowledge, no-one has argued on the possible pitfall the aggregation may be exposed to in the literature. The only exception is Jablecki [6]. He rejects the introduction of a continuum of whatever in economics for its unrealistic nature and according to Jablecki, the conclusions on economic theory based on the continuum assumptions are

[1]University of Economics in Prague, Department of Statistics and Probability, nám. W. Churchilla 4, Praha 3, e-mail: {kodera,tran}@vse.cz

incorrect and unapplicable. Unlike Jablecki, whose critical opinion on the continuum assumption may have deep philosophical roots, we accept the fact that when modeling an economy, one has to make some simplifying assumptions. Therefore we do not refuse to accept the usage of continuum approach. Our analysis is hence focused on how to make it applicable for modelling economy, especially in the DSGE model case. As a result, in our paper, we analyze the necessary conditions for the use of continuum approach and point to the possible ambiguity when using some mathematical construction based on the continuum notion. Finally, we would like to show how one can alleviate such ambiguity by using a decimal classification system.

2 Infinity and continuum

As we have already stated in the introduction, the subject of our interest in this paper is the notion continuum in Economics which is used to refer to the number of types of produced commodities, or the number of entities, either consumers or producers, in an economy. This is a very interesting problem whose theoretical foundation is not appropriately discussed in the current leading economic journals. Dynamic stochastic general equilibrium theory which dominates contemporary macroeconomic theory considers a continuum of producers or a continuum of sorts of goods (Gali [5], McCandless [8] and Walsh [9]). A producer and a sort of goods are closely related in the dynamic stochastic general equilibrium theory because the theory assumes that each sort of goods is produced by just one producer. It also assumes that each product has just one price, so the quantities of producers, sorts of goods and prices are the same and therefore, they can be indexed by the same numbers. As a result, it would be enough to deal only with a population of producers as the conclusions drawn from the model for them would be the same as for the whole sorts of goods and prices.

When dealing with a finite population of producers, we use the natural numbers to index them if the number of producers is finite. For example, if we have n agents in an economy, and each of them produces one type of product, we can sufficiently number them by natural numbers from one to n as well as the goods they produce by these numbers. The fact that we have n manufacturers and each of them is assigned to a number from interval of one to n, mathematically we can say that there is a bijection (each agent has just one number and each number corresponds just to one agent) of a set of producers to a set of numbers $\{1, 2, \ldots, n\}$. To put it formally: the cardinality of the set is n. Theoretically, this set can be "extended" to infinity (provided that the assumption that one commodity is produced by one manufacturer is retained). If we index all producers by a whole sequence of natural numbers $\{1, 2, 3, \ldots\}$, we talk about the fact that the set of producers is infinite, but countable (i.e. it is infinite, but it is ordered into a sequence).

If n denotes a natural number, it is possible to use the finite version of Dixit-Stiglitz aggregator [3] to express an aggregate economic quantity. For example, the composite index which aggregates individual production Y_i, $i = 1, \ldots, n$ is expressed by

$$Y = \left[\sum_{1}^{n} Y_i^{1-\frac{1}{\epsilon}} \right]^{\frac{\epsilon}{\epsilon-1}}, \ \epsilon > 1.$$

If we express the aggregate production this way, it is absolutely right and does require any other additional conditions. We can also release the assumption of finite n. Then we have an infinitely countable set of agents, one can used the infinite version of Dixit-Stiglitz aggregator:

$$Y = \left[\sum_{1}^{\infty} Y_i^{1-\frac{1}{\epsilon}} \right]^{\frac{\epsilon}{\epsilon-1}}, \ \epsilon > 1,$$

where Y_i, $i = 1, 2, \ldots$ is a infinite sequence of production indexed by natural numbers. The above composite index of production has an interpretation if the series of production converges. But the assumption of convergence is a little artificial in economic environment i.e. it is not implied by economic requirements.

Now, instead of using a sequence of natural numbers to index the whole population of producers, we use a continuum to describe them. Imagine an economy, in which we have a number of producers that are indexed by all real numbers from interval [0, 1]. In this case, we talk about a continuum of agents in an economy. It means that we assign each agent or each sort of commodity to a real number from

interval [0, 1] and each number from this interval corresponds to any agent. One then may ask what is the cardinality of a continuum in this case. The answer is the cardinality of a continuum is also infinity, but somehow "denser". Mathematically, we say that we map the continuum of producers (or equivalently the set of producers) on interval [0, 1].

But such mapping between continuum of producers and digits in interval $[0, 1]$ may bring up many problems. Let's assume that the production of an individual producer is $Y(i)$, $i \in [0, 1]$. The composite index of production Y is given by Dixit-Stiglitz aggregator for continuum economics, i.e. by integral:

$$Y = \left[\int_0^1 Y^{1-\frac{1}{\epsilon}}(i)\mathrm{d}i\right]^{\frac{\epsilon}{\epsilon-1}}, \quad \epsilon > 1.$$

Function $Y(i)$ on domain $[0, 1]$ can be arbitrary so it may not meet the integrability conditions in the sense of Lebesgue theory of integral. It is a very serious problem. For the aggregation purpose, New Keynesian theory uses mathematical structures (integrals), whose existence is not warranted. To solve this problem, it is necessary to find sufficient conditions for the integrability of function $Y(i)$ by giving them some economic interpretation.

For this purpose, it is necessary to begin with the discussion about the domain of function $Y(i)$. As the integral is used to obtain aggregate economic quantity Y, we have to study the measure on their domain $[0, 1]$. The most common approach is to use Borel measure defined on σ-algebra of Borel sets, which is generated by a system of all open subsets (subintervals (α, β), $\alpha, \beta \in [0, 1]$). In Economics, economic interpretation of each open subintervals (α, β) is very important because we solve the question if the agents or products indexed by close numbers, have also similar features. This question remains unanswered by New Keynesian Economics theorists.

3 The notion of continuum in economics

In the previous section, we have shown that we can index a producer in a population of them either by a natural number from a sequence of natural numbers or by a real number from interval [0,1]. One then may ask what makes economists-theorists to use the continuum of commodities, consumers, producers etc to cope with the whole population of agents in an economy. The introduction of a continuum of producers into economic modelling is clearly inspired by its use in Physics. In Physics, this approach is supported by human intuition.

Consider a meter-long metal stick with a varying density. The density differs from layer to layer, but it changes continuously which results from the physical properties of the rod. This variation in density causes no problem for calculating the weight of the rod as it is composed of a continuum of infinitely thin cross section slices putting side by side consecutively to make the final density. The total weight of the rod is the integral on the interval [0, 1] of the mass density function. In this case, each cross-section layer has its own size which corresponds to a real number from the interval [0, 1]. The density of infinitely thin slice is a function defined on interval [0, 1]. The rod weight is the integral of the density function from 0 to 1, which is the sum of infinite number of layers. As density function is continuous due to its physical nature, the weight must be the integral from zero to one of this density function.

Unlike in Physics, there is a danger that using a continuum of products or manufacturers to describe a population of agents in an economy when modeling may lead to other conclusions than those when the countable or infinitely countable sets are considered as Aumann [1] has already pointed to. The reason might be as follows. Let us recall how the concept of continuum is made. In an economy, each commodity is assigned to an index, which is a real number from interval $[0, 1]$ and vice versa. In this case, we talk about a continuum of types of products. Indexing commodity in this way can raise a serious problems when we introduce a function on domain $[0, 1]$ which for example captures the size of production of individual goods. It can be very obscure because without any mathematical restricting assumptions, and thus may not be integrable. If we want to explain economic meaning of function $Y(i)$, $i \in [0, 1]$, we must also explain economic characteristics of open subinterval (α, β). These subintervals should contain indices of individual product of similar feature. The structure of New Keynesian model and equilibrium conditions imply the fact that products with similar characteristics are produced at a similar amount. This implicitly guarantees the continuity of $Y(i)$ in a sufficiently small neighbourhood which means that integrability is secured.

The fact that theoreticians of New Keynesian Economics have no economic interpretation of open sets of $[0,1]$ matters more than we would expect. Individual real numbers thus indicate only indices of commodities without any ties to the surrounding numbers, which are, however, other commodities indices. In this case, when there is no other structure on the interval, it is possible to arbitrarily renumber types of products. Renumbering the types of products may cause the ambiguity of aggregate quantities displayed by composite indices as we will show in the following section.

4 Ambiguity of aggregation in continuum economics

In this section, we assume that the production of individual agents is given by a Lebesgue integrable function. Even after accepting this assumption which is not explained by New Keynesian Economics, there still remains the danger of ambiguity of aggregation. This indeterminacy results from the possibility to renumber the agents in the environment of a continuum of agents. We will show that one cannot obtain uniquely determined results for an aggregate quantity in the environment of a continuum-of-agents economics because macroeconomic aggregates depends on how the agents are indexed.

Let's denote the production of individual agents function as $Y(t)$. The domain of this function is interval $[0,1]$ and co-domain $R^+ \equiv x \in \Re : x \geq 0$. Let us assume (it is not too realistic assumption in this field) that $Y(i)$ is a Lebesgue integrable function. To express the aggregate of production Y with Dixit-Stiglitz aggregator

$$Y = \left[\int_0^1 Y^{1-\frac{1}{\epsilon}}(i)\mathrm{d}i\right]^{\frac{\epsilon}{\epsilon-1}} \tag{1}$$

where elasticity $\epsilon > 1$ is used. Now, we will argue by an example that the aggregator does not express the production of arbitrary economy unambiguously. Let the production of different goods numbered by real numbers $[0, 0.25)$ has the same size 4. Goods are indexed by numbers in interval $[0.25, 1]$ and they are produced in the same quantity 9. The graphical illustration is on Fig 1. Let us assume that elasticity of substitution for the economy $\epsilon = 2$. Integral (1) gets a form as follows:

$$Y = \left[\int_0^1 Y^{\frac{1}{2}}(i)\mathrm{d}i\right]^2. \tag{2}$$

Function Y in our example is a piece-wise constant function, so after replacing from our example we get:

$$Y = \left[\int_0^{0.25} Y^{\frac{1}{2}}(i)\mathrm{d}i + \int_{0.25}^1 Y^{\frac{1}{2}}(i)\mathrm{d}i\right]^2 = \left[0.25 \times 4^{\frac{1}{2}} + 0.75 \times 9^{\frac{1}{2}}\right]^2 = \left[\frac{1}{2} + \frac{9}{4}\right]^2 = \frac{121}{16}.$$

If we had a finite number of agents and goods and renumbered them, nothing would change as we know that addition is commutative. Renumbering in the case of continuum agents and goods corresponds to bijection from the interval $[0,1]$ on $[0,1]$. Let us consider function $j = \sqrt{i}$ which is a bijection from $[0,1]$ on $[0,1]$. Graphical illustration of function $Y(j)$ after renumbering is displayed on Fig 2. When apply aggregation with help of Dixit-Stiglitz aggregator and using data from our example we obtain:

$$Y = \left[\int_0^{0.5} Y^{\frac{1}{2}}(j)\mathrm{d}j + \int_{0.25}^1 Y^{\frac{1}{2}}(j)\mathrm{d}j\right]^2 = \left[0.5 \times 4^{\frac{1}{2}} + 0.5 \times 9^{\frac{1}{2}}\right]^2 = \left[1 + \frac{3}{2}\right]^2 = \frac{25}{4}.$$

The obtained result shows that renumbering goods has impact on the value of production index, which is a serious problem if we want to aggregate all individual productions by this approach. The total production of economy depends on a purely administrative act which is the mapping of a continuum of goods on interval [0, 1].

5 Decimal classification and following problems

As a solution to the problem we have described above we suggest a decimal classification of products. We label all products in consideration by real numbers from interval $[0,1]$. It means all numbers begin by digit 0 except the last one 1 which can be left out without losing any generality. The continuum of goods would be divided into ten categories $0.0, 0.1, ..., 0.9$ according to their qualitative features. Each category has its own ten subcategories and so on. By doing so, we obtain a topology on the set of all types of

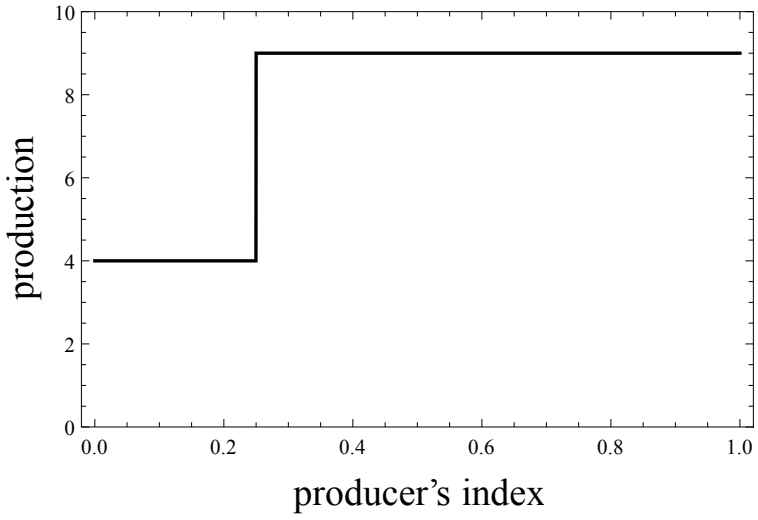


Figure 1 The aggregation of production with its original indices

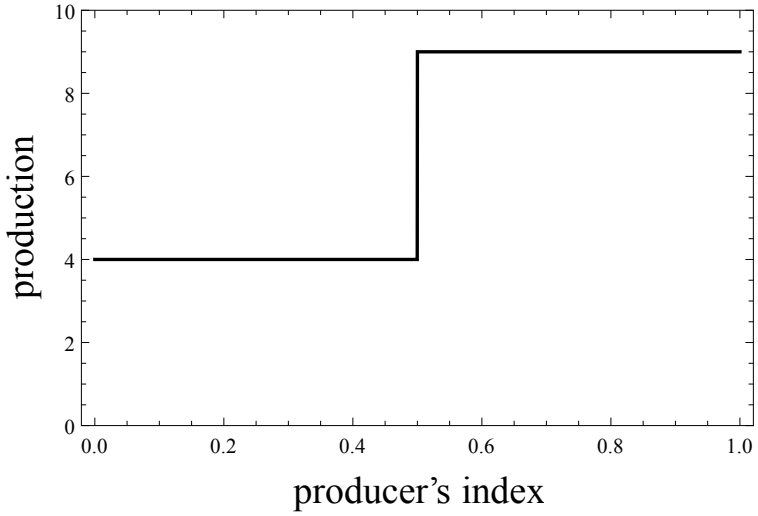


Figure 2 The aggregation of production after renumbering its indices

products, commodities close to each other according to their qualitative characteristics. The closeness of their corresponding indices assigned by a real number from interval (0, 1) in terms of their distance in this interval will depend only on their qualitative characteristics.

Assuming that the re-optimisation of prices (Gali [5]) is applied for goods of close characteristics, we can expect close goods will have a similar size of demand and in equilibrium, they would have an approximately equal size of production. If the size of production of somehow closely related products are not very different, then function of $Y(i)$ could be considered as a continuous one. Then the aggregation across all goods by integration would be feasible. In order to make the case more general, we may allow the existence of a finite number of points of discontinuity, or even an infinite but countable number of points of discontinuity. Such sets have a zero measure, i.e. the integral remains the same.

This idea of decimal classification we have suggested above should not be understood as an implementable proposal. It should be taken as a theoretical concept for how to solve the problem of ambiguity of aggregation across a continuum of agents or goods. As one may have noticed that decimal classification of product cannot entirely eliminate the problem of ambiguity, but it could bring some order into the indexation of different products by real numbers. It is obvious that the values of aggregates will depend on how decimal classification will be applied, i.e how categories subcategories and so on will be defined. But decimal classification introduces a systemic way of ordering of products of different sorts on the interval $[0, 1]$.

This process of sorting cannot continue to infinity, and it must be stopped after a finite number of

steps. The result of the process is finite system of disjunct semi-closed intervals which cover the interval $[0, 1]$. Elements of this intervals belong to agents or types of products with similar characteristics. This method, in our opinion, is suitable for indexing agents and goods and it also allows for aggregating across a whole population to get macro-economics aggregates for the need in New Keynesian Economics.

The other way to solve this problem is to abandon the assumption of a continuum of goods and agents. Instead of it, agents or goods are indexed by natural number and their numbers are finite. In this case, the act of indexing each good is a very simple operation and renumbering is possible without any danger as we have already shown when aggregating an economic quantity, one can use the finite Dixit-Stiglitz aggregator. However, it is necessary to find new methods and limit their use to finite versions to obtain the needed aggregates.

6 Conclusion

Economists have borrowed the notion continuum from Mathematics to capture a whole population of agents in an economy. Each agent in the continuum is arbitrarily indexed by a real number taken from a closed interval $[0, 1]$. To obtain an aggregate quantity, one just needs to integrate a corresponding quantity of an agent across the whole population. Though we do not a priori reject such kind of simplifying assumption for its unrealistic nature as others may, in our work we try to point to some problems this kind of simplification may get into.

In our opinion, the use of notion continuum in economics should be supported by a clear economic interpretation. It is important because it can give us a clue on the integrability of the quantity to be aggregated, no matter whether it is the output, the prices, the consumption and the individual labour force of different skills in a continuum. We have to say that modern New-Keynesian theorists does not bother to tackle this question.

Further, even if the existence of the integral is secured, the ambiguity of the aggregation across a whole continuum would still remain as it has been shown in the article. To solve this problem, we propose a solution which lies in the creation of relatively fixed structure. In this structure each individual quantity is numbered by a decimal classification as it has been shown. Instead of a continuum, there is an idea of finite number of agents, goods or whatever. To our knowledge, this approach not only eliminates the ambiguity connected with the aggregation across a continuum, but it also softens the unrealistic nature of the notion continuum in Economics.

Acknowledgements

The financial support from the Czech Science Foundation under Grant P402/12/G097 is gratefully acknowledged.

References

[1] Auman, R.: Markets with a Continuum of Traders. *Econometrica* **32** (1964), 39–50.

[2] Cantor, G.:. Ein Beitrag zur Mannigfaltigkeitslehre. *Journal fuer die Reine und Angewandte Mathematik* **84** (1878), 242-258.

[3] Dixit, A., Stiglitz, J.: Monopolistic Competition and Optimum Product Diversity. *American Economic Review* **67** (1977), 297308.

[4] Dornbusch, R., Fischer, S. and Samuelson, P. A.: Heckscher- Ohlin Trade Theory with a Continuum of Goods. *The Quarterly Journal of Economics* **95** (1980), 203–24.

[5] Gali, J.: *The ABCs of RBCs: An Introduction to Dynamic Macroeconomic Models*. Princeton University Press, Princeton, 2008.

[6] Jablecki, J.: How Many Traders Can You Fit into a Model? Retrieved from http://mises.org/daily/2646, 2007.

[7] Leamer, E. E.: The Heckscher-Ohlin Model in Theory and Practice. Princeton Studies in International Finance 77. Princeton, NJ: Princeton University Press, 1995.

[8] McCandless, G.: *Monetary Policy, Inflation and Business Cycle*. Harvard University Press, Cambridge, MA, 2008.

[9] Walsh, C. E.: *Monetary Theory and Policy*. The MIT Press, Cambridge, Massachusetts, 2010.

Impact of changes in fares and demand to the tariff zones design

Michal Kohání[1]

Abstract. When designing a tariff system, there are several approaches how to design it. One of them is to divide the region into tariff zones where the price of traveling depends on the number of traveled zones. There are many important factors with major influence on the solution of the problem such as the number of created zones or the price of traveling - fares. Changing of fares also affect the demand for traveling. Higher fare often occurs a drop in demand and number of carried passengers and it also affect the total income from passengers. We will introduce a mathematical model of the tariff zones design problem based on counting zones that contains parameters modeling the impact of fare changes on the demand for traveling. We will solve this model using a universal optimization tool Xpress. We will make a computational study on test data from selected region and compare the impact of obtained solutions to the tariff zones design and its parameters.

Keywords: tariff planning, tariff zones design, IP solver, demand, fares

JEL Classification: C44, C61
AMS Classification: 90C08

1 Introduction

One of important tasks in the designing of integrated transport system is the design of the tariff subsystem. There are several approaches of designing the tariff and one of them is dividing the region into the tariff zones. The price of traveling in such system is determined by the number of traveled tariff zones by the passenger [1][8].

There are many important factors with major influence on the solution of the problem such as the number of created zones or the price of traveling - fares. Changing of fares can change the demand for traveling. Higher fare often occur a drop in demand and number of carried passengers and it also affect the total income from passengers. However, if the price will be lower, it can motivate travelers to use public transport and may lead to the increasing of the number of transported passengers. Commonly used models don't take into account these factors.

In this article we analyze the problem of the tariff system design. In the third chapter we introduce a mathematical model of the tariff zones design problem based on counting zones. In the next chapter we introduce the adapted mathematical model that contains parameters modeling the impact of price changes on the demand for traveling. We make a computational study on test data from selected region and compare the impact of obtained solutions to the tariff zones design and its parameters. We perform numerical experiments using a universal optimization tool Xpress.

2 Tariff zones design problem

In the field of public transport there are several ways of designing the fare and prices for travelling. As was mentioned in [2], [3] and [8], frequently used way is a *distance tariff* system. In this system the price for a trip is calculated according to the length of the trip. We need to have the distance between origin and destination station to calculate the price for the trip. This type of tariff is often used in regional transportation and intercity lines. The *unit tariff* is the opposite of the distance tariff. In this system the prices for all trips are equal and are independent on the distance. The unit tariff is frequently used in city public transport, but it is not very suitable for regional public transportation, especially in large regions, as was mentioned in [8].

Third type of tariff system is a *zone tariff* system. This system combines advantages of both unit and distance tariff. In this type of tariff system the whole region is divided into smaller sub-regions (*the tariff zones*). The price for a trip depends on the starting and the ending zone of the trip. Inside one zone the unit tariff applies. For a trip that involves passing several zones, there are two ways of determining the price, as in [2] and [8].

[1] University of Zilina, Faculty of Management Science and Informatics, Department of Transportation Networks, Univerzitna 1, 01026, Zilina, Slovakia, Michal.Kohani@fri.uniza.sk

If the price is given arbitrarily for each pair of zones, we call it *a zone tariff with arbitrary prices*. An example for this tariff system is the Zilina Regional Integrated Transport System in Slovak Republic, see the tariff prices matrix in the Table 1 and Figure 1. Prices here depend on the pair of origin and destination zones and the number of travelled zones is not important, because prices are given for all pairs of zones separately.

Prices of basic fare single tickets in ŽRIDS (In Euro)

	To zone 1	**To zone 2**	**To zone 3**	**To zone 4**	**To zone 5**	**To zone 6**	**To zone 7**
From zone 1	x	x	x	1.00	1.00	1.30	1.50
From zone 2	x	x	x	0.90	0.90	1.20	1.40
From zone 3	x	x	x	0.45	0.45	0.75	0.95
From zone 4	1.00	0.90	0.45	0.45	0.45	0.70	0.90
From zone 5	1.00	0.90	0.45	0.45	0.45	0.50	0.70
From zone 6	1.30	1.20	0.75	0.70	0.50	0.45	0.45
From zone 7	1.50	1.40	0.95	0.90	0.70	0.45	0.45

Table 1 Price matrix in Zilina Regional Integrated Transport System *(www.dpmz.sk)*

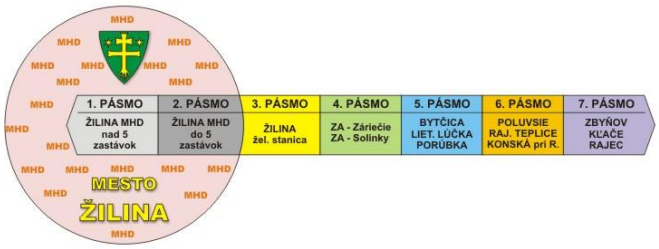


Figure 1 Zone tariff system in Zilina region *(www.dpmz.sk)*

The counting zone tariff system is the second variant of a zone tariff system. The price of trip in this system is calculated according to the number of crossed zones. The prices depend on the origin and the destination zone of the trip, but trips passing the same number of zones must have the same price. The example of a counting zone tariff system in Bratislava region in Slovakia is in the Table 2 and Figure 2.

Number of zones	**Price**
1 regional zones	14,90 €
2 regional zones	25,90 €
3 regional zones	31,90 €
4 regional zones	42,90 €
5 regional zones	52,90 €
6 regional zones	63,90 €
7 regional zones	73,90 €
8 regional zones	84,90 €
9 regional zones	95,90 €
10 regional zones	105,90 €
All zones	114,90 €

Table 2 Prices for travelling for 30 days period – Integrated system of Bratislava region *(www.bid.sk)*

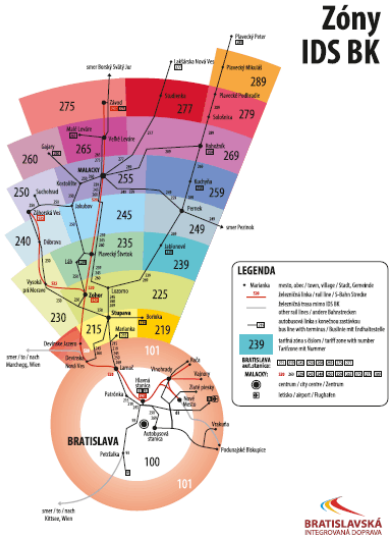


Figure 2 Zone tariff system in Bratislava region *(www.bid.sk)*

According to [2] and [8], zone tariff system is gaining popularity in integrated transport system across Europe due to its simplicity. When a public transportation company or the regional transportation office wants to change its tariff system to a zone tariff, it has to design the zones and to fix the new fares. The goal is to design in the way that resulting system must be accepted by the customers and does not decrease the income of the company. The goal then is to design the zones in that the new and the old price for most of the trips are as close as possible. This means that neither the public transportation company nor the customers will have major disadvantages when changing the current tariff system to a zone tariff. Another goal can be to design fair zones. This approach was described in [6].

3 Mathematical model of the tariff zones design problem

Let all stations in the network of public transport constitute the set I. The station i and j from set I are connected by the edge $(i,j) \in V$, if there is direct connection by public transport line between these two stations. Symbol V denotes the set of edges. The distance between stations i and j is denoted as d_{ij}. For each pair of stations i and j is c_{ij} the current price of travelling between these two stations. The number of passengers between stations i and j is b_{ij} (OD matrix).

If we want to calculate new price of the trip between nodes i and j in the counting zones tariff system, we need to calculate the number of zones crossed on this trip. The calculation of the number of crossed zones can be easily replaced by the calculation of crossed zone borders as was used in [2]. We assume that the node can be assigned only to one zone and then the border between zones is on the node. We will introduce the binary variable w_{rs} for each existing edge $(r, s) \in V$, which is equal to 1 if stations r and s are in different zones and is equal to 0 otherwise. For calculation of the number of crossed borders we need to determine the used path for traveling between stations i and j. We introduce a_{ij}^{rs}, where the used paths will be observed. a_{ij}^{rs} is equal to 1 if the edge (r,s) will be used for travelling between i and j and 0 otherwise.

As was mentioned in [3] and [4], there are two possibilities, how to set a new price for travelling. In the first case *(Price_uni)*, the unit price f for travelling in one zone will be established. This means that for the calculation of the new price it will be necessary to count the number of travelled zones and multiplied it by the unit price per one zone. In the second case *(Price_dif)* two different unit prices will be determined – price f_1 per travelling in the first zone and unit price f_2 for travelling in each additional zone. The final price will be calculated as a sum of the basic price for the first zone and number of other travelled zones multiplied by the unit price for additional zones. According to the numerical experiments in [4], we will use the *Price_dif* setting of prices for travelling. The current or fair price between stations i and j is denoted by c_{ij}. New price determined by the number of crossed zones will be calculated according to *Price_dif* setting as follows (1):

$$n_{ij} = f_1 + \sum_{(r,s)\in V} f_2 a_{ij}^{rs} w_{rs} \tag{1}$$

We introduce binary variables y_i, which represent the "fictional" centre of the zone. Variable y_i is equal to 1 if there is a centre of the zone in node *i* and 0 otherwise. For each pair of stations *i* and *j* we introduce variables z_{ij}. Variable z_{ij} is equal to 1 if the station *j* is assigned to the zone with centre in the node *i* and 0 otherwise. We expect to create at most *p* tariff zones.

According to [3] and [8] there are two different objective functions. First one will be the maximal deviation between the current or fair price and new price determined by the number of crossed zones for all passengers between *i* and *j*. Second one will be the average deviation between current and new price for all passengers. According to the previous research in [3] and advices of experts in [8], in this article we will use the average deviation between current and new price as a criterion in objective function.

The mathematical model *(Mod_simple)* can be written in the form:

$$Minimize\ dev_{avg} = \frac{\sum_{i\in I}\sum_{j\in J}\left|c_{ij}-n_{ij}\right|b_{ij}}{\sum_{i\in I}\sum_{j\in J}b_{ij}} \tag{2}$$

$$subject\ to \sum_{i\in I} z_{ij} = 1,\ for\ j \in I \tag{3}$$

$$z_{ij} \le y_i,\ for\ i, j \in I \tag{4}$$

$$z_{ij} - z_{ik} \le w_{jk},\ for\ i \in I, (j,k) \in V \tag{5}$$

$$\sum_{i\in I} y_i \le p \tag{6}$$

$$z_{ij} \in \{0,1\},\ for\ i,\ j \in I \tag{7}$$

$$y_i \in \{0,1\},\ for\ i \in I \tag{8}$$

$$w_{ij} \in \{0,1\},\ for\ (i,\ j) \in V \tag{9}$$

Conditions (3) ensure that each station will be assigned exactly to one zone. Conditions (4) ensure that the station *j* will be assigned only to the existing centre of the zone. Conditions (5) are coupling between variables for allocation of the station to the zone and the variables for determining the zone border on the edge *(j,k).* Condition (6) ensures that we will create the most *p* tariff zones.

This model will be solved using IP solver with exact methods, so we will obtain exact solution of the problem. Because the objective function (2) in this model is not a linear function, we need to modify this objective function to linear form. This adjustment was shown in [4].

4 Changes in fares and demand

According to [5] and [7], important factors which has major influence on demand in public transport are price, speed of travel, information about transport, safety, comfort and availability. In our model we can influence only the price for travelling. As was mentioned earlier, the goal is to design the zones in that the new and the old price for most of the trips are as close as possible. This means that neither the public transportation company nor the customers will have major disadvantages when changing the current tariff system to a zone tariff. The increase in prices can cause reducing the number of transported passengers. On the contrary, a price reduction may affect the attractiveness of transport and hence the slight increase in the demand and the number of transported passengers. Nevertheless, objective function (2) does not include this possibility so we will have to reformulate it.

We introduce new variables u_{ij}, $v_{ij.}$ Variables u_{ij} represent the calculated prices for travelling in case that new price is lower than current and variables v_{ij} represent the calculated prices for travelling in opposite case. We can also introduce the coefficients *d* and *e*. Coefficient *d* represents the percentage increase in number of passengers in the case of lover new prices, coefficient *e* represents the percentage decrease in number of passengers in the case of higher new prices.

Then we can reformulate mathematical model *(Mod_demand)* to the form:

$$Minimize\ dev_{\max} = \frac{\sum_{i\in I}\sum_{j\in J}(1+d).u_{ij}b_{ij} + \sum_{i\in I}\sum_{j\in J}(1-e).v_{ij}b_{ij}}{\sum_{i\in I}\sum_{j\in J}b_{ij}} \tag{10}$$

$$subject\ to\ (3)-(9)$$

$$c_{ij}-n_{ij} = u_{ij} - v_{ij}, for\ i,j \in I \tag{11}$$

$$u_{ij} \geq 0,\ for\ i,\ j \in I \tag{12}$$

$$v_{ij} \geq 0,\ for\ i,\ j \in I \tag{13}$$

5 Numerical experiments

The goal of numerical experiments was to compare the assignment of stations to the created zones and the average deviation in prices between both models. Numerical experiments were performed on the data of the Zvolen County in Slovak Republic. The stations in the network are represented by the 51 municipalities. Current prices were calculated according to real prices depending on the distance for travelling by regional buses. The OD matrix was estimated using the gravity model as in [4], where the number of passengers between nodes *i* and *j* is calculated as follows:

$$\frac{b_i b_j}{d_{ij}}$$

Parameter b_i represents the number of inhabitants in the node *i*. We assume that the decrease in number of passengers in the case of higher new prices will be higher than increase of passengers with lower new prices. Parameter *d* was set to 0.02 and the parameter *e* had value 0.05. Parameter f_1 was set to 0.7 and parameter f_2 to the value 0.4 and according to the best values obtained in [3]. We perform experiments for 8 different values of parameter *p*, denoted as p_{max} in Tables. To perform the computation we used the general optimization software tool FICO XPRESS 7.3 [9]. The experiments were performed on a personal computer equipped with Intel Core 2 Duo E6850 with parameters 3 GHz and 3.5 GB RAM. In the Table 3 and Table 4 there are results for both models *(Mod_simple)* and *(Mod_demand)* used in this study. In the Table 3 there are compared differences in assignments of stations (column *Difference*) and the number of created zones (column *Zones*). Total deviations between current prices and new prices for all passengers are calculated in the Table 4.

p_{max}	*Zones*		*Difference*
	Mod_simple	*Mod_demand*	
4	4	4	0
6	6	6	0
8	8	8	0
10	10	10	1
13	12	13	3
16	16	16	4
20	20	20	6
25	24	24	4

Table 3 Number of created zones and assignments

p_{max}	*Mod_simple*	*Mod_demand*
4	9780.9	9924.67
6	9655.7	9756.47
8	9628.7	9690.89
10	9230.2	9128.16
13	8455.2	8267.24
16	8266.1	8244.51
20	7859.5	7922.73
25	7691	7746.72

Table 4 Total deviations between current prices and new prices in the system

6 Conclusion

The results of numerical experiments show that the differences between two models (*Mod_simple, Mod_demand*) are very small. From the Table 3 we can see that the number of created zones is equal in almost all cases. There are some differences in assignment of stations to tariff zones but these differences in assignment are in the stations with smaller population, as was found by comparing the results. In the Table 4 we can see that the results obtained by *Mod_simple* model are better than in the model which includes the influence of demand on the number of transported passengers. To proof the above results, we want perform more experiments with real data of selected region in the future.

Acknowledgements

This work was supported by the research grants VEGA 1/0339/13 "Advanced microscopic modelling and complex data sources for designing spatially large public service systems", APVV-0760-11 "Designing of Fair Service Systems on Transportation Networks" and the institutional grant of Faculty of Management Science and Informatics (FRI ZU). We would like to thank to "*Centre of excellence for systems and services of intelligent transport*" (ITMS 26220120050) for built-up infrastructure, which was used.

References

[1] Hejhalová, B.: Kedy sa stanú integrované dopravné systémy osobnej dopravy na Slovensku realitou? *Doprava a spoje : internetový časopis* **1** (2011), 29-36.

[2] Hamacher, H. W., and Schöbel, A.: Design of Zone Tariff Systems in Public Transportation. *Operations Research* **52** (2004), 897-908.

[3] Kohání, M.: Designing of zone tariff in integrated transport systems, In: *Communications : scientific letters of the University of Žilina* **15** (2013), 29-33.

[4] Kohání, M.: Exact approach to the tariff zones design problem in public transport, In: *Mathematical methods in economics : proceedings of the 30th international conference*, Silesian University in Opava, Karviná, 2012, 426-431.

[5] Konečný, V.: Príjmová elasticita dopytu po osobnej doprave v SR, In: *Doprava a spoje - internetový časopis* **1** (2007), 30-37.

[6] Palúch, S.: On a fair fare rating on a bus line. In: *Communications : scientific letters of the University of Žilina* **15** (2013), 25-28.

[7] Poliak, M., and Semanová, Š.: Obstarávanie dopravnej obslužnosti, In: *Perner´s Contacts* **29** (2013), 162-167.

[8] Schöbel, A.: *Optimization in Public Transportation: Stop Location, Delay Management and Tariff Zone Design in a Public Transportation Network*, Springer, 2006.

[9] *XPRESS-Mosel User guide*. Fair Isaac Corporation, Birmingham, 2012.

Portfolio efficiency with respect to higher order stochastic dominance criteria

Miloš Kopa[1]

Abstract. This paper deals with portfolio efficiency testing with respect to higher order stochastic dominance criteria. Firstly, we test the second-order stochastic dominance (SSD) portfolio efficiency, that allows for risk averse decision makers, for portfolios from a regular grid. The grid is created from one-month US Treasury bill (a riskless asset) and 10 representative value-weighted active benchmark stock portfolios. They are formed, and annually rebalanced, based on individual stocks' market capitalization of equity, each representing a decile of the cross-section of stocks in a given year. Secondly, we limit our attention to SSD efficient portfolios and we test their efficiency with respect to higher order stochastic dominance. Finally, we compare all results with mean-variance efficiency frontier. In all our models we assume discrete distribution of monthly returns and no short sales.

Keywords: stochastic dominance, portfolio efficiency, mean-variance efficiency

JEL classification: D81, G11
AMS classification: 91B16, 91B30

1 Introduction

The theory of decision making under risk is one of the most appealing issues within financial mathematics. It is based on the basic economical principles, however it also exploits optimization techniques and statistical tools. Mathematical formulations of decision making problems under risk lead to stochastic programming models which are searching for the optimal solution with respect to a chosen objective (criterion) and feasibility constraints. In financial applications, they turn out to so called portfolio selection problems. These problems basically capture two fundamental principles: non-satiation and risk attitude. While the non-satiation axiom is easy to implement and generally accepted in economics and finance, the risk attitude can be understood or expressed in various ways.

The first portfolio selection problem was introduced by Markowitz (1952). The model jointly focuses on maximizing expected return and minimizing variance of the portfolio, where variance serves as a measure of risk. Consider N risky assets with returns modeled by random vector $\boldsymbol{\varrho}$. Let $\boldsymbol{\lambda} \in \mathbb{R}^M$ be a vector of weights determining the way how the initial wealth is invested. Following Markowitz (1952), short sales are allowed, that is, the set of all feasible portfolios Λ is given by condition: $\mathbf{1}'\boldsymbol{\lambda} = 1$. Consequently, the return and variance of portfolio (with weights) $\boldsymbol{\lambda}$ are $E(\boldsymbol{\varrho}'\boldsymbol{\lambda})$ and $\mathrm{var}(\boldsymbol{\varrho}'\boldsymbol{\lambda})$. Hence, Markowitz portfolio selection problem may be formulated as follows:

$$\max_{\boldsymbol{\lambda}\in\Lambda} E(\boldsymbol{\varrho}'\boldsymbol{\lambda}) - \vartheta \mathrm{var}(\boldsymbol{\varrho}'\boldsymbol{\lambda})$$

where $\vartheta \geq 0$ is a risk aversion parameter.

Alternatively, one may model the risk attitude using utility functions introduced in von Neumann and Morgenstern (1944). Contrary to the bi-objective mean-variance models, application of utility function leads to classical nonlinear programming problem - to maximize expected utility of the final wealth. Unfortunately, identifying the particular utility function of the decision maker is usually very difficult. When the utility function is not precisely known, one can consider at least some suitable, economically meaningful classes of utility functions, for example the class of all nondecreasing and concave utility

[1]Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Prague 8, Czech Republic, kopa@karlin.mff.cuni.cz

functions U_2. In this case, the optimal solution of the maximizing expected utility problem can not be exactly found. However, one can at least compare two portfolios. If one of them gives higher or equal expected utility of final wealth (or returns) than the other one for all considered utility functions then a relation of stochastic dominance (SD) between them exists. The notion of stochastic dominance was introduced in statistics more than 50 years ago and it was firstly applied to economics and finance in Quirk and Saposnik (1962), Hadar and Russell (1969) and Hanoch and Levy (1969). Moreover, considering the particular class of utility functions one can test whether a given portfolio is efficient or not. Until recently, practical applications of stochastic dominance relations in portfolio analysis were hampered by the absence of tractable algorithms to deal with diversification across multiple assets. The pairwise comparisons criteria have only a limited use when looking for efficient portfolios with respect to SD relations. In the last decade, under assumptions of discrete distribution of returns, Post (2003), Kuosmanen (2004) and Kopa and Chovanec (2008) developed linear programming tests to analyze if a portfolio is efficient with respect to the second-order stochastic dominance (nondecreasing concave utility functions) relative to all possible portfolios formed from the considered assets. Later on, Kopa and Post (2013) introduced a general SSD efficiency test which is a generalization of all previous SSD efficiency tests what makes the theory of SSD portfolio efficiency more compact. Moreover, they presented also a new SSD portfolio efficiency test that is much less computationally demanding and applicable even for thousands of scenarios. Although the test does not identify a dominating portfolio, the computational tractability allows for further sensitivity, robustness or bootstrap techniques. The tests are formulated for the case when no short sales are allowed, that is :

$$\Lambda = \{\boldsymbol{\lambda} \in R^M | \mathbf{1}'\boldsymbol{\lambda} = 1, \ \lambda_m \geq 0, \ m = 1, 2, \ldots, M\} \tag{1}$$

however, one can easily modify the results for any nonempty bounded polytope set Λ. If only non-satiation of decision maker is assumed, that is, the set of all non-decreasing utility functions (U_1) is considered, one must distinguish between two concepts of efficiency: " FSD admissibility" and " FSD optimality". A mixed-integer programming test for FSD admissibility was proposed by Kuosmanen (2004) while the FSD optimality test was derived in Kopa and Post (2009). Recently, Post and Kopa (2013) introduced a general linear programming portfolio efficiency test for any Nth-order of stochastic dominance when $N \geq 2$. The notion of portfolio efficiency with respect to stochastic dominance criteria is related also to the other efficiency approaches, mainly, it is related to DEA models and also to mean-risk models, see Branda and Kopa (2012, 2013) for more details.

In this paper we test the efficiency of billions of portfolios using the Post and Kopa (2013) NSD portfolio efficiency test for $N = 2, 3, 4$. Contrary to Post and Kopa (2013) we test the efficiency of all portfolios from the regular grid created from 11 base assets. Moreover, we consider the shorter time period for our data than Post and Kopa (2013) in order to increase the importance of during-crises observations (scenarios). We compare the sets of efficient portfolios for $N = 2, 3, 4$ and mean-variance efficiency frontier among each others.

The rest of this paper is structured as follows. Section 2 introduces the basic notation and formulates definitions of stochastic dominance criteria. Moreover, it recalls the NSD portfolio efficiency test of Post and Kopa (2013). Section 3 presents our empirical study. It introduces the considered data from Kenneth French library, summarizes its basic descriptive statistics and presents the results. Finally, Section 4 concludes the paper with several ideas for future theoretical improvements based on bootstrap and robustness techniques.

2 Nth-order stochastic dominance

Until very recently, stochastic dominance portfolio efficiency tests (Post 2003, Kuosmanen 2004, Kopa and Chovanec 2008, Lizyayev 2012, Dupačová and Kopa 2012) were derived only with respect to class U_1 or U_2. However, in general, one can consider any subset of U_1. Let $U_N \subset U_1$ be the set of N times differentiable utility functions such that: $(-1)^k u^{(k)} \leq 0$ for all $k = 1, 2, ..., N$. As the limiting case, $U_\infty \subset U_1$ is the set of infinitely differentiable utility functions with alternating signs of the derivatives. These functions are also called completely monotonic utility functions. More details about completely monotonic utility functions can be found in Whitmore (1989) and references therein.

For each set $U_N \subset U_1$, $N = 1, 2, ...$ we define weak Nth-order stochastic dominance relation between portfolios $\boldsymbol{\lambda}$ and $\boldsymbol{\tau}$ as follows: $\boldsymbol{\lambda} \succeq_{NSD} \boldsymbol{\tau}$ if $Eu(\boldsymbol{\varrho}'\boldsymbol{\lambda}) - Eu(\boldsymbol{\varrho}'\boldsymbol{\tau}) \geq 0$ for every utility function $u \in U_N$.

If strict inequality holds true for at least one $u \in U_N$ then the relation is called Nth-order stochastic dominance.

Let $F^{(1)}_{\boldsymbol{\varrho}'\boldsymbol{\lambda}}$ be the cumulative distribution function of returns of portfolio $\boldsymbol{\lambda}$. Considering random variables with bounded support (a, b), Levy (2006) presents a necessary and sufficient condition for weak Nth-order stochastic dominance relation in terms of N times cumulated distribution functions

$$F^{(N)}_{\boldsymbol{\varrho}'\boldsymbol{\lambda}}(t) = \int_{-\infty}^{t} F^{(N-1)}_{\boldsymbol{\varrho}'\boldsymbol{\lambda}}(x)dx$$

as follows: $\boldsymbol{\lambda} \succeq_{NSD} \boldsymbol{\tau}$ if and only if $F^{(k)}_{\boldsymbol{\varrho}'\boldsymbol{\lambda}}(b) \leq F^{(k)}_{\boldsymbol{\varrho}'\boldsymbol{\tau}}(b)$, for all $k = 2, 3, ..., N-1$ and $F^{(N)}_{\boldsymbol{\varrho}'\boldsymbol{\lambda}}(x) \leq F^{(N)}_{\boldsymbol{\varrho}'\boldsymbol{\tau}}(x)$ for all $x \in \langle a, b\rangle$. Again, in the case of NSD relation, at least one strict inequality is required.

The infinite-order stochastic dominance relation (ISD) is defined as a limit case of NSD when $N \to \infty$: $\boldsymbol{\lambda} \succeq_{ISD} \boldsymbol{\tau}$ if $Eu(\boldsymbol{\varrho}'\boldsymbol{\lambda}) - Eu(\boldsymbol{\varrho}'\boldsymbol{\tau}) \geq 0$ for every utility function $u \in U_\infty$ and $\boldsymbol{\lambda} \succ_{ISD} \boldsymbol{\tau}$ if $Eu(\boldsymbol{\varrho}'\boldsymbol{\lambda}) - Eu(\boldsymbol{\varrho}'\boldsymbol{\tau}) \geq 0$ for every utility function $u \in U_\infty$ and strict inequality holds for at least one $u \in U_\infty$.

For infinite-order of stochastic dominance, Whitmore (1989) derived a necessary and sufficient condition based on the Bernstein Theorem: $\boldsymbol{\lambda} \succ_{ISD} \boldsymbol{\tau}$ if and only if

$$E\left(e^{-a\boldsymbol{\varrho}'\boldsymbol{\lambda}} - e^{-a\boldsymbol{\varrho}'\boldsymbol{\tau}}\right) \leq 0 \quad \text{for all} \quad a \geq 0.$$

The recursive nature of the definition of the N-times cumulated distribution function makes it clear that (weak) Nth-order stochastic dominance implies (weak) $\overline{N}$th-order stochastic dominance for $\overline{N} > N$. Moreover, any Nth-order stochastic dominance implies infinite-order stochastic dominance. Thistle (1993) also proved that if $\boldsymbol{\lambda} \succ_{ISD} \boldsymbol{\tau}$ then $\boldsymbol{\lambda} \succ_{NSD} \boldsymbol{\tau}$ for some finite N.

The general definition of NSD efficiency for $N \geq 2$ can be seen as an extension of SSD efficiency and, following Post and Kopa (2013), we formulate it in the "NSD optimality" form.

Definition 1. A given portfolio $\boldsymbol{\tau}$ is NSD efficient ($N \geq 2$), if there exists at least one utility function $u \in U_N$ such that $Eu(\boldsymbol{\varrho}'\boldsymbol{\tau}) - Eu(\boldsymbol{\varrho}'\boldsymbol{\lambda}) \geq 0$ for all $\boldsymbol{\lambda} \in \Lambda$ with strict inequality for at least one $\boldsymbol{\lambda} \in \Lambda$.

In the rest of the paper we will assume a discrete probability distribution of returns $\boldsymbol{\varrho}$ that is given by scenarios $\mathbf{x}^t$, $t = 1, 2, ..., T$ that occurs with probability p_t. The scenarios are collected in the following matrix:

$$X = \begin{pmatrix} \mathbf{x}^1 \\ \mathbf{x}^2 \\ \vdots \\ \mathbf{x}^T \end{pmatrix}$$

where $\mathbf{x}^t = (x^t_1, x^t_2, \ldots, x^t_M)$ is the t-th row of matrix X.

Moreover, let $(X\boldsymbol{\tau})^{[k]}$ be the k-th smallest element among $(\mathbf{x}^1\boldsymbol{\tau}), (\mathbf{x}^2\boldsymbol{\tau}), ..., (\mathbf{x}^T\boldsymbol{\tau})$, that is, $(X\boldsymbol{\tau})^{[1]} \leq (X\boldsymbol{\tau})^{[2]} \leq \ldots \leq (X\boldsymbol{\tau})^{[T]}$.

Post and Kopa (2013) consider the following reformulation of Definition 1.

Proposition 1. *An evaluated portfolio $\boldsymbol{\tau} \in \Lambda$ is efficient in terms of N-th order stochastic dominance, $N \geq 2$ relative to all feasible portfolios $\boldsymbol{\lambda} \in \Lambda$ if it is an optimal solution of the maximizing expected utility problem for some admissible utility function $u \in U_N$, that is:*

$$\sum_{t=1}^{T} p_t u(\mathbf{x}^t\boldsymbol{\tau}) \geq \sum_{t=1}^{T} p_t u(\mathbf{x}^t\boldsymbol{\lambda}) \quad \forall \boldsymbol{\lambda} \in \Lambda \Longleftrightarrow$$

$$\sum_{t=1}^{T} p_t u'(\mathbf{x}^t\boldsymbol{\tau})(\mathbf{x}^t\boldsymbol{\tau} - x^t_j) \geq 0, \quad j = 1, 2, ..., M.$$

The proposition follows from the Karush-Kuhn-Tucker first-order condition for selecting the optimal combination of assets: $\max_{\boldsymbol{\lambda}\in\Lambda} \sum_{t=1}^{T} p_t u\left(\mathbf{x}^t\boldsymbol{\lambda}\right)$. An admissible function $u \in U_N$ is now chosen as any function from U_N which is not constant on the relevant interval $\left\langle (X\boldsymbol{\tau})^{[1]}, (X\boldsymbol{\tau})^{[T]} \right\rangle$ in order to avoid some

trivial cases. This reformulation was first introduced by Post (2003) for SSD ($N = 2$) and applies also for higher-order criteria ($N > 2$), but it does not apply for FSD ($N = 1$). Kopa and Post (2009) presented a different utility-based formulation for this case. Finally, Post and Kopa (2013) derived the following necessary and sufficient condition for NSD efficiency of a given portfolio.

Theorem 2. *Let*

$$\begin{aligned}
\theta^*(\boldsymbol{\tau}) = \min_{\beta_n, \gamma_k, \theta} \ & \theta \qquad (2)\\
\text{s.t.} \quad & \left(\mathbf{x}^t\boldsymbol{\tau} - x_j^t\right)p_r \sum_{t=1}^{T}\left(\sum_{n=1}^{N-2} n\beta_n\left(\mathbf{x}^t\boldsymbol{\tau} - \mathbf{x}^T\boldsymbol{\tau}\right)^{n-1} + \right.\\
& \left.(N-1)\sum_{k=t}^{T}\gamma_k\left(\mathbf{x}^t\boldsymbol{\tau} - \mathbf{x}^k\boldsymbol{\tau}\right)^{N-2}\right) + \theta \geq 0,\ j = 1, \cdots, M\\
& (-1)^n\beta_n \leq 0,\ n = 1, \ldots, N-2\\
& (-1)^{N-1}\gamma_k \leq 0,\ k = 1, 2, \ldots, T\\
& \sum_{t=1}^{T}\left(\sum_{n=1}^{N-2} n\beta_n\left(\mathbf{x}^t\boldsymbol{\tau} - \mathbf{x}^T\boldsymbol{\tau}\right)^{n-1} + (N-1)\sum_{k=t}^{T}\gamma_k\left(\mathbf{x}^t\boldsymbol{\tau} - \mathbf{x}^k\boldsymbol{\tau}\right)^{N-2}\right)p_r = 1.
\end{aligned}$$

A portfolio $\boldsymbol{\tau}$ is NSD efficient if and only if $\theta^*(\boldsymbol{\tau})$ given by (2) is equal to zero.

Solving problem (2) one tries to find an "optimal" utility function, that is, a function from U_N that maximizes the expected utility of returns in $\boldsymbol{\tau}$. If $\theta^*(\boldsymbol{\tau}) = 0$ then this optimal function can be reconstructed from the optimal solution of (2), see Post and Kopa (2013) for more details.

3 Empirical application

We follow Post and Kopa (2013) in taking data from the Kenneth French library. We consider a standard set of 10 active benchmark stock portfolios as the base assets. They are formed, and annually rebalanced, based on individual stocks market capitalization of equity, each representing a decile of the cross-section of stocks in a given year. Furthermore, we include the 1-month US Treasury bill as a riskless asset. We use data on monthly returns from January 2002 to December 2011 (120 months). We assume that all scenarios are equiprobable, that is, $p_t = 1/120$, $t = 1, ..., 120$.

Portfolios	Mean	St. deviation	Min	Max	Skewness	Kurtosis
1st decile	0.70	6.28	-20.46	15.05	-0.45	0.51
2nd decile	0.47	6.71	-20.10	17.61	-0.28	0.30
3rd decile	0.68	6.45	-22.07	16.21	-0.41	0.72
4th decile	0.61	6.18	-20.62	15.83	-0.39	0.80
5th decile	0.63	6.13	-20.37	17.52	-0.27	0.67
6th decile	0.74	5.50	-19.42	15.41	-0.40	0.98
7th decile	0.60	5.53	-22.45	14.32	-0.64	1.88
8th decile	0.53	5.45	-20.86	13.69	-0.59	1.42
9th decile	0.48	5.02	-21.44	14.67	-0.75	2.73
10th decile	0.12	4.38	-14.94	10.18	-0.48	0.88

Table 1 Basic data descriptive statistics

First, we construct a regular grid $\{0, 0.01, 0.02..., 0.99, 1\}^{11} \cap \Lambda$ from the base assets. Then we test SSD efficiency of each portfolio from the grid using (2) for $N = 2$. We identify 748 SSD efficient portfolios, all of them are composed from the 1st decile portfolio, the 6th decile portfolio and the riskless asset. Finally, we test NSD efficiency for $N = 3, 4$ and mean-variance efficiency of all these 748 portfolios. We find 101 of these portfolios, that are NSD efficient for $N = 3$. Perhaps surprisingly, these 101 portfolios are NSD efficient for $N = 4$ and mean-variance efficient as well. These portfolios are linear convex combinations of the 6th decile portfolio (the asset with the highest mean return) and the riskless asset. The results are summarized at Figure 1.

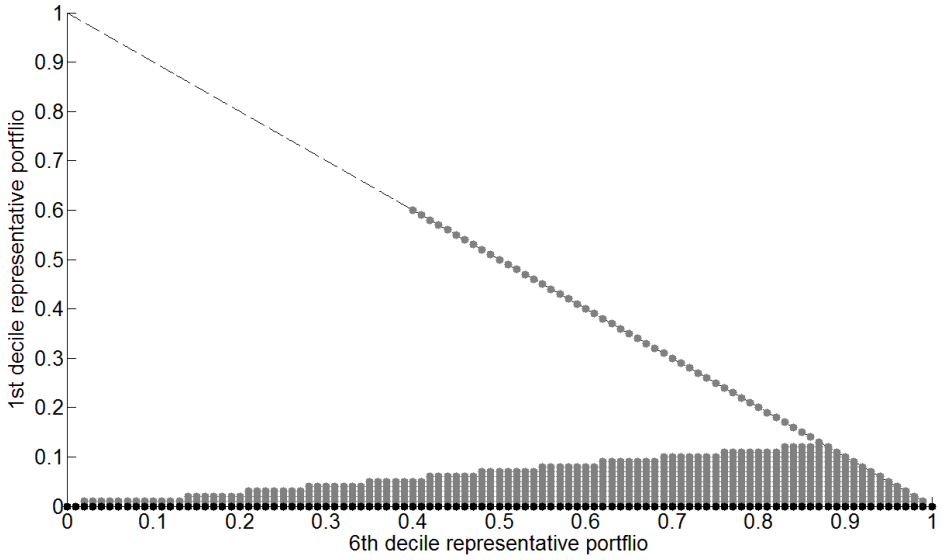


Figure 1 The figure captures a comparison of various efficiency criteria. The grey area and black dots express the set of SSD efficient portfolios. The black dots correspond to portfolios that are NSD efficient for $N = 3, 4$ and mean-variance efficient, too. Each of these portfolios consists only of the 6th decile portfolio and the riskless asset.

4 Conclusions

In this paper a portfolio efficiency with respect to various criteria was analyzed. Considering 11 base assets a regular grid was constructed. Using the general NSD efficiency test of Post and Kopa (2013), the second-order, third-order and fourth-order stochastic dominance efficiency of portfolios from the grid was tested. Then it was compared to the mean-variance efficiency frontier. It was found that 748 portfolios satisfy the conditions of SSD efficiency, but only 101 of them were also identified as third-order stochastic dominance efficient. Moreover, these 101 portfolios were classified as the fourth-order efficient and mean-variance efficient as well.

For future research, this study can be improved in various ways. For example, longer historical data can be used. In addition, one can consider the portfolio efficiency in a more robust way as it was done in Kopa (2010, 2012) or Dupačová and Kopa (2012, 2013) for the first and the second-order stochastic dominance, using contamination techniques and the worst-case approach. Alternatively, one can compare the results also with the first-order stochastic dominance efficiency set using the Kopa and Post (2009) test. Unfortunately, all these improvements would lead to more computationally demanding efficiency tests what requires much better hardware equipment than is currently available.

Acknowledgements

The research was partly supported by the Czech Science Foundation (grant P402/12/0558).

References

[1] M. Branda, M. Kopa: DEA-risk Efficiency and Stochastic Dominance Efficiency of Stock Indices, *Czech Journal of Economics and Finance*, **62**, 2 (2012), 106 – 124.

[2] M. Branda, M. Kopa: On relations between DEA-risk models and stochastic dominance efficiency tests, *Central European Journal of Operations Research*, DOI 10.1007/s10100-012-0283-2, Online first 2013.

[3] J. Dupačová, M. Kopa: Robustness in stochastic programs with risk constraints, *Annals of Operations Research*, **200**, 1 (2012), 55 – 74.

[4] J. Dupačová, M. Kopa: Robustness of optimal portfolios under risk and stochastic dominance constraints, submitted in *European Journal of Operational Research*.

[5] J. Hadar and W. R. Russell: Rules for Ordering Uncertain Prospects, *American Economic Review*, **59**, 1 (1969), 25 – 34.

[6] G. Hanoch, H. Levy: The efficiency analysis of choices involving risk, *Review of Economic Studies* **36** (1969), 335 – 346.

[7] M. Kopa: Measuring of second-order stochastic dominance portfolio efficiency, *Kybernetika*, **46**, 3 (2010), 488 – 500.

[8] M. Kopa: Robustness in SSD portfolio efficiency testing, *Proceedings of International Conference on Operations Research*, pp. 527 – 532 Springer, Berlin, 2012.

[9] M. Kopa, P. Chovanec: A second-order stochastic dominance portfolio efficiency measure, Kybernetika **44**, 2 (2008), 243 - 258.

[10] M. Kopa, T. Post: A portfolio optimality test based on the first-order stochastic dominance criterion, *Journal of Financial and Quantitative Analysis* **44**, 5 (2009), 1103 - 1124.

[11] T. Kuosmanen: Efficient diversification according to stochastic dominance criteria, *Management Science* **50**, 10 (2004), 1390 – 1406.

[12] A. Lizyayev: Stochastic dominance efficiency analysis of diversified portfolios: classification, comparison and refinements. *Annals of Operations Research*, **196**, 1 (2012), 391 - 410.

[13] H. Levy: *Stochastic Dominance: Investment Decision Making Under Uncertainty*. Second edition. Springer Science, New York 2006.

[14] H. M. Markowitz: Portfolio Selection, *Journal of Finance*, **7**, 1 (1952), 77 – 91.

[15] J. von Neumann, O. Morgenstern: *Theory of Games and Economic Behavior*, Princeton University Press, 1944.

[16] T. Post: Empirical tests for stochastic dominance efficiency, *Journal of Finance*, **58** (2003), 1905 – 1932.

[17] T. Post, M. Kopa: General Linear Formulations of Stochastic Dominance Criteria, *European Journal of Operational Research*, Online first 2013, http://dx.doi.org/10.1016/j.ejor.2013.04.015.

[18] J. P. Quirk, R. Saposnik: Admissibility and measurable utility functions, *Review of Economic Studies*, **29** (1962), 140 - 146.

[19] G. A. Whitmore: Stochastic dominance for the class of completely monotonic utility functions, *Studies in the Economics of Uncertainty* (Thomas B. Fomby and Tae Kun Seo eds.), Springer Verlag, New York, (1989), 77 – 88.

Verification of the Linder Hypothesis with using the Gravity Model of International Trade

Jana Kovářová[1]

Abstract. New role of the demand in the trade theories is associated with the theory of overlapping demand by Swedish economist, Staffan Linder. The Linder hypothesis postulates that countries with similar level of income will trade more than countries with dissimilar level of income. On the contrary, the supply side Heckscher-Ohlin hypothesis implies that countries with dissimilar level of income will trade more than countries with similar level of income. The Linder hypothesis is associated with high level of the intra-industry trade (similarity of demands) and the Heckscher-Ohlin theory with high level of the inter-industry trade (different factor endowments). The subject of this paper is an empirical investigation of the Linder hypothesis with using the gravity model of international trade in manufactured goods in the case of the Czech economy with other EU member countries. Motivations for this research are two-fold. First, new information is provided on verification of the Linder hypothesis of international trade in manufactured goods between the Czech Republic and other EU member countries. Second is the application of the gravity model of international trade with using panel data in the static and dynamic form.

Keywords: Linder hypothesis, Heckscher-Ohlin hypothesis, Czech Republic, gravity model of international trade, panel data.

JEL Classification: C23 ; C50 ; F12 ; F14 ; O52.
AMS Classification: 62G05.

1 Introduction

Many trade theories emphasized the supply side in explaining trade direction. For example the Heckscher-Ohlin (HO) model which was developed by two Swedish economists, Eli Heckscher and Bertil Ohlin. In this model comparative advantage is determined by differences in endowments of factors across countries. Other supply oriented model of international trade is the Ricardian model which postulates that comparative advantage is determined by differences in technology. In 1961, Steffan Linder presented alternative view in explaining the international trade direction which is based on the similarity of the demands. Linder hypothesis of international trade departs from the neoclassical theories where supply conditions are the most important factor of trade [4]. Linder proposed that trade patterns in manufacturing are dependent on the similarity of preference among nations. He argued that countries have similar demands for manufacturing with others that have similar per capita income levels and got favorable empirical results [12].

Since Linder did not present a formal model to test his hypothesis[2], many others have tried different approaches to test it. The primary tool for testing Linder hypothesis was rank correlation analysis used for example by Sailors, et al. [22] or Greytak and McHugh [11]. These studies were heavily criticized, however, for their failure to employ regression analysis, a technique that could have controlled for the effects of distance on trade intensities [6]. However, the empirical evidence based upon the regression analysis of the Linder hypothesis was rather mixed. For example, Kennedy and McHugh [14, 15] and Qureshi et al. [20] tested the theory in terms of changes in propensities to trade against changes in income differences between two points in time to control for distance and found no support for the Linder model. Bergstrand [3] made the theoretical link between the Linder model and the gravity model specification, but found little or no evidence to support Linder's hypothesis. On the other hand, Thursby and Thursby [23] tested the Linder hypothesis using gravity model specification and they

[1] VŠB-TU Ostrava, Faculty of Economics, Department of Economics, Sokolská Třída 33, Ostrava 1, 701 21, jana.kovarova.st2@vsb.cz.

[2] Linder proposed a tabular form of test comparing the import propensities between similar-income countries with those for dissimilar countries to test his hypothesis. He arrayed the countries vertically and horizontally in order of per capita incomes and tried to find a tendency toward bunching of the high bilateral import propensities around the main diagonal, which would be in favor of the thesis [17].

found overwhelming support for this hypothesis. Arnon and Weinblatt [1] use a simple gravity equation model to analyze the nature of bilateral trade flows among and between the 35 developed and less developed countries. Their study provides empirical evidence supporting the comprehensive validity of the Linder's hypothesis for all country groups with both high and low income. Chow et al. [13] tested the Linder hypothesis for trade between the four original tiger economies (Hong Kong, Singapore, South Korea and Taiwan) and their major OECD markets during the 1965-1990 period. They conclude that the Linder hypothesis may provide a relatively good explanation of trade for countries above some per capita income threshold and for trade in differentiated products. McPherson et al. [18] support the Linder hypothesis for five East African developing countries (Ethiopia, Kenya, Rwanda, Sudan and Uganda). Their study is based on panel data analysis in the 1984-1992 periods. The results imply that these countries trade more intensively with others who have similar per capita income levels, as predicted Linder. Fillat-Castejon and Serrano-sanz [9] tested the verification of the Linder hypothesis in Spain using data for the 1959-1986 period find international demand to be an important determinant of trade and suggest that foreign markets can be considered an extension of the domestic market. Leitao and Faustino [16] examine the features and determinants of Portuguese intra-industry trade in the years 1995-2003 and find that differences in income levels have a positive impact on intra-industry trade. Bohman and Nilsson [5] suggested a new approach to assess the Linder hypothesis, incorporating the distribution of income within a country. They develop two different variables to capture the similarity in demand structures between two trading partners and the size of the market, constituting a market overlap. The results imply that similarity in structure of demand acts as a catalyst of trade flows between countries. This similarity is more important for the differentiated goods than homogenous goods.

The subject of this paper is an empirical investigation of the Linder hypothesis with using the gravity model of international trade in goods in the case of the Czech economy with other EU member countries. Motivations for this research are two-fold. First, new information is provided on verification of the Linder hypothesis of international trade in goods between the Czech Republic and other EU member countries. Second is the application of the gravity model of international trade with using panel data in the static and dynamic form.

2 Czech trade with EU member states

The Czech economy has changed in terms of the extent and the share of the economy openness during the transformation process. The changes in the territorial structure of the Czech foreign trade were clearly associated with the changes in the commodity structure. Requirements to succeed in the markets of industrialized countries were restructuring of the industrial production, investments in new equipment, purchases of new technologies and human resources. At the beginning of the 90th 20th century, Czech Republic exported mainly raw materials and supplies (e.g. metallurgical and steel products). The second half of the 90th years means positive turn. In the Czech exports begin dominate products with higher added value, i.e. mainly machinery and transport equipment [10]. Table 1 shows the Czech commodity structure with the EU in 2011.

Commodity group	Export		Import	
	Value (mill. CZK)	Share (%)	Value (mill. CZK)	Share (%)
Food and live animals	83 988	3.5	103 486	6.1
Beverages and tobacco	15 875	0.7	13 756	0.8
Crude materials, inedible, except fuels	72 697	3.1	39 273	2.3
Mineral fuels, lubricants and related materials	100 046	4.2	103 124	6.0
Animal and vegetable oils, fats and waxes	4 812	0.2	5 482	0.3
Chemicals and related products, n.e.s.	140 130	5.9	246 854	14.5
Manufactured goods classified chiefly by material	425 381	17.9	397 496	23.3
Machinery and transport equipment	1 271 143	53.4	636 475	37.3
Miscellaneous manufactured articles	262 958	11.0	158 742	9.3
Commodities and transactions not classified elsewhere in the SITC	3 102	0.1	2 923	0.2

Table 1 Czech commodity structure of the foreign trade with the European Union in 2011. [7]

In 2011, more than half of exports (53.4 %) to the EU countries were machinery and transport equipment (SITC 7), which represented the absolute level 1 271 143 mill. CZK. The most exported items were following: the road vehicles (412 437 million CZK, 17.33 % share of total SITC 7 exports), the electrical machinery, apparatus and appliances (225 290 million CZK, 9.47 %), the office machines and automatic data processing (199 552 million CZK, 8.38 %), the telecommunications and sound recording and reproducing (148 277 million CZK, 6.23 %) and the machinery and equipment generally used in the industry (144 488 million CZK, 6.07 %). The aggregate share of these five items in the total Czech exports to the European Union was 47.48 % [7].

Also in imports were the largest commodity group machinery and transport equipment with 37.3 % share of the total Czech imports (636 475 million CZK). The biggest amount of imported manufactured goods was in these categories: the electrical machinery, apparatus and appliances (171 758 million CZK, 10.06 %), the road vehicles (168 058 million CZK, 9.84 %) and the equipment generally used in industry (99 622 million CZK, 5.83 %) [7].

The Czech commodity structure with EU member countries points to a very high level of the intra-industry trade. The Linder hypothesis, which is based on a higher level of intra-industry trade[3], is assumed.

3 Empirical model

The gravity model represents a simple empirical tool for analyzing bilateral trade flows between geographical entities. This model originates from the Newtonian physics notion. The gravity model for trade is analogous: the trade flow between two countries (integration units, trading block etc.) is proportional to the product of each country's "economic mass", generally measured by GDP (national income) and inversely proportional to the distance between the countries respective "economic centers of gravity", generally their capitals [21]. The original gravity equation proposed by Tinbergen is as follows [2]:

$$Trade_{ij} = \alpha \cdot \frac{GDP_i \cdot GDP_j}{Dis_{ij}}, \tag{1}$$

where $Trade_{ij}$ represents the value of the bilateral trade exchange between country i and j, GDP_i and GDP_j are country i and j's respective national incomes, Dis_{ij} is the geographical distance between the countries and α is an intercept. The original equation (1) is transformed into the linear form using natural logarithm:

$$\ln(Trade_{ij}) = \alpha + \beta_1 \ln(GDP_i \cdot GDP_j) - \beta_2 \ln(Dis_{ij}) + \varepsilon_{ij}, \tag{2}$$

where $Trade_{ij}$ is the value of the bilateral trade exchange between country i and j, GDP_i and GDP_j are country i and j's respective national incomes, Dis_{ij} is the geographical distance between the countries, α is an intercept, β_1 and β_2 are the regression coefficients, ε_{ij} is the random component.

The gravity model in the form of the equation (2) is typically used to analyze cross-sectional data. Often other variables are included such as population size to reflect the possible economies of scale or set of dummy variables incorporating institutional characteristic such as economic integration arrangements, infrastructure or cultural and linguistic proximity, historical links, exchange rate, foreign direct investments, common language, common border, inland location, cultural differences and various barriers to trade.

In this study a modified gravity model with panel data structure is used. For verification the Linder hypothesis is included a variable which measures the degree of similarity between the per capita income levels of the selected countries. The estimated equation is in form:

$$\ln(Trade_{ijt}) = \alpha + \beta_1 \ln(GDP_{it} \cdot GDP_{jt}) - \beta_2 \ln(Dis_{ijt}) + \beta_3 Linder_{ijt} + \varepsilon_{ij}, \tag{3}$$

where $Trade_{ijt}$ is the value of the bilateral trade exchange between country i and j in time t, GDP_{it} and GDP_{jt} are country i and j's respective national incomes in time t, Dis_{itj} is the geographical distance between the countries in time t, $Linder_{ijt}$ is the natural logarithm of the absolute value of per capita GDP differential between country i and j in time t, α is an intercept, $\beta_1, \beta_2, \beta_3$ are the regression coefficients, ε_{ij} is the random component.

4 Data and methodology

The Linder hypothesis is tested in the case of Czech Republic (CR) international trade in manufactured goods with other member states of the European Union (EU) in the 1995-2011 periods. The dependent variable $Trade_{ijt}$

[3] The Czech share of the manufacturing intra-industry trade in the total manufacturing was 66.3 % between the 1992-1995 periods and increased by 11.1 % in the 1996-2000 periods [19].

expresses the sum of the export and import of manufactured goods of the Czech Republic with other member states of the European Union. The independent variable $GDP_{it} \cdot GDP_{jt}$ is the economic mass of the Czech Republic and one of the EU member state. According to Linder's theory, GDP of exporting country express the supply and GDP of importing country reflects the demand. It is expected a positive estimated coefficient β_1 because the bigger GDP implies the bigger trade between countries. The independent variable Dis_{itj} is the physical distance between the capital cities of the Czech Republic and one of the EU member state[4]. Longer distance means higher transportation cots thus decreasing the potential for bilateral trade flows. For this reason, it is expected a negative estimated coefficient β_2. The independent variable $Linder_{ijt}$ is the most important variable in this model. It measures the difference between per capita GDP of the Czech Republic and one of the EU member state. According to Linder hypothesis, countries with similar level of income will trade more than countries with dissimilar level of income. Therefore when two countries GDP per capita are close to each other, this income similarity variable would tend to be smaller and the dependent variable $Trade_{ijt}$ would tend to be bigger. It is expected a negative estimated coefficient β_3 to confirm the Linder hypothesis. On the contrary, the positive sign of this estimated coefficient implies that country traded in accordance of the Heckscher-Ohlin trade theory. All data were obtained from the United Nation Conference on Trade and Development (UNCTAD) in current USD. Table 2 presents the summary statistics for the variables included in the gravity model.

Variable	Observation	Mean	Std. Dev.	Min	Max	Skewness	Kurtosis
$ln(Trade_{ijt})$	442	20.41568	1.987548	15.36927	25.11724	-0.225306	2.606046
$ln(GDP_{it} \cdot GDP_{jt})$	442	50.94896	1.971154	46.78412	55.05977	-0.096112	2.259857
$ln(Dis_{itj})$	442	6.757133	0.595602	5.540871	7.721792	-0.430465	2.457037
$Linder_{ijt}$	442	9.186802	1.003350	6.567194	11.48195	-0.340264	2.396873

Table 2 Summary statistics. Self-elaboration with using EViews 7.

The gravity model is estimated with panel data structure in the static and dynamic form. The Least Squares Method (LMS) with fixed time effects is used for static estimate. The Generalized Method of Moments (GMM) is used for dynamic estimate. The lagged values of the dependent variable $Trade_{ijt}$ were chosen as instruments. The advantage of the dynamic estimate is that it takes into account the existence of the long-term contracts.

5 Empirical results

The estimation results for static form are presented in table 3.

Variable	Observation	Coefficient	t-Statistic	Probability
$ln(GDP_{it} \cdot GDP_{jt})$	442	0.847261	51.31611	0.0000
$ln(Dis_{itj})$	442	-1.563049	-38.71572	0.0000
$Linder_{ijt}$	442	-0.287804	-9.978777	0.0000

Table 3 Static results. Self-elaboration with using EViews 7.

The overall performance of this model is very good, with typically high R-squared (0.94) [5]. All explanatory variables are significant, indicating that the gravity model is appropriate in explaining CR international trade in manufactured goods with other EU member states. Presented results confirm that the bilateral trade in manufactured goods of the Czech Republic with other EU member states positively depends on their economic mass measured by GDP and negatively depends on the distances measured by distances between their capitals. The estimated coefficient can be interpreted as follows. When the variable $GDP_{it} \cdot GDP_{jt}$ increased by 1 %, the bilateral trade exchange of the manufactured goods increased by 0.84 % in the 1995-2011 period. On the contrary, when the distance increased by 1 %, the bilateral trade exchange of the manufactured goods decreased by 1.56 % in the same period. The estimated coefficient β_3 has a negative sign. This result confirms the validity of the Linder hypothesis which implies that the Czech Republic traded in manufactured goods with other EU member states in accordance with the postulates of the Linder theory. The estimated coefficient β_3 can be interpreted as follows, when the difference of the income measured by absolute difference in GDP per capita increased by 1 %, the bilateral trade exchange of the manufactured goods decreased by 0,28 %.

The estimation results for dynamic form are presented in table 4.

[4] The distance between capital cities is measured using the great circle formula, which takes into account the longitude and latitude of the capitals.

[5] The adjusted R-squared is 0.93 for static form.

Variable	Observation	Coefficient	t-Statistic	Probability
$ln(GDP_{it} \cdot GDP_{jt})$	442	0.830990	13.87685	0.0000
$ln(Dis_{itj})$	442	-2.180653	-21.07554	0.0000
$Linder_{ijt}$	442	-0.268055	-5.064925	0.0000

Table 4 Dynamic results. Self-elaboration with using EViews 7.

The dynamic results are very similar to the static results. The overall performance of this model is very good too, with high R-squared (0.91) [6]. The estimated coefficients β_1 and β_3 have almost identical value in the dynamic form as well as in the static form. The higher value of the estimated coefficient β_2 indicates greater importance of the transportation costs in the dynamic form.

Conclusion

Linder's trade theory for the first time suggested that in some areas of foreign trade, concretely trade in manufactured goods, it is not only supply but also demand that should be taken into account when explaining bilateral trade flows between countries. Linder suggested that his theory was applicable only to trade in differentiated manufactured goods in which consumer tastes and scale economies were deemed to be particularly important. Linder was pursued that trade in raw material or agricultural products can be adequately explained by the traditional theory with its emphasis on the supply of productive factors, including climate and natural resources. Linder's trade theory also provides an explanation for intra-industry trade, an important phenomenon in international trade, which depends on economies of scale and implies imperfectly competitive markets [8].

The validity of the Linder hypothesis, in the case of the Czech economy with other EU member countries in the 1995-2001 periods, was confirmed using the gravity model of international trade. The model is estimated with panel data structure in the static and dynamic form. The Least Squares Method with fixed time effects is used for static estimate. The Generalized Method of Moments is used for dynamic estimate. The lagged values of the dependent variable $Trade_{ijt}$ were chosen as instruments. The advantage of the dynamic estimate is that it takes into account the existence of the long-term contracts. The results of both estimations are similar with exception of the transaction costs which depends on trade distance. The sensitivity of the dependent variable $Trade_{ijt}$ on the distance is higher in the dynamic specification of the model.

Acknowledgements

This paper was created under SGS project (SP2013/122) of Faculty of Economics, VSB - Technical University of Ostrava.

References

[1] Arnon, A., and Weinblatt, J.: Linder's Hypothesis Revisited: Income Similarity Effects for Low Income Countries. *Applied Economics Letters* **5** (1998), 605-611.
[2] Batra, A.: India's Global Trade Potential: The Gravity Model Approach. *ICRIER Working Paper No. 151* (2004).
[3] Bergstrand, J. H.: The generalized Gravity equation, monopolistic competition, and the factor-proportions theory in international trade. *Review of Economics and Statistics* **71** (1989), 143-153.
[4] Bohman, H., and Nilsson, D.: Introducing income distribution to the Linder hypothesis. *ETSG 2006 Vienna Eighth Annual Conference*, 7-9 September 2006, WIIW and Universität Wien.
[5] Bohman, H., and Nilsson, D.: Market Overlap and the Direction of Exports – A New Approach of Assessing the Linder Hypothesis. *CESIS Electronic Working Paper Series*. Paper No. 86 (2007).
[6] Bukhari, S., et al.: An Empirical Analysis of the Linder Theory of International Trade for South Asian Countries. *The Pakistan Development Review* **44** (2005), 307-320.
[7] CzechTrade: Zahraniční obchod České Republiky v roce 2011, 2012 [online]. Available:http://www.czechtrade.cz/d/documents/01/7-infoservis/analyzy-zo-cr/2012/zahr_obchod_za_rok_2011.pdf.
[8] Dakal, D., et al.: Another Empirical Look at the Theory of Overlapping Demands. The University of New Haven Department of Economics and Finance, *Working Paper No. 0901*, February 11, 2009.

[6] The adjusted R-squared is 0.91 for dynamic form.

[9] Fillat-Castejon, C., and Serrano-sanz, J.: Linder Revisited: Trade and Development in the Spanish Economy. *International Review of Applied Economics* **13** (2004) 323-348.
[10] Fojtíková, L.: *Zahraničně obchodní politika ČR: historie a současnost (1945-2008).* C H Beck, Praha, 2009.
[11] Greytak, D., and McHugh, R.: Linder's trade thesis: an empirical examination. *Southern Economic Journal* **43** (1977), 1386-1389.
[12] Choi, C.: Linder hypothesis revisited. *Applied Economics Letters* **9** (2002), 601-605.
[13] Chow, P. et al.: A Test of the Linder Hypothesis in Pacific NIC Trade 1965-1990. *Applied Economics* **31** (1999), 175-182,
[14] Kennedy, T. E., and McHugh, R.: An intertemporal test and rejection of the Linder hypothesis. *Southern Economic Journal* **46** (1980), 898-903.
[15] Kennedy, T. E., and McHugh, R.: Taste similarity and trade intensity: a test of the Linder hypothesis for U.S. exports. *Weltwirtschaftliches Archiv* **119** (1983), 84-96.
[16] Leitao, N., and Faustino, H.: Country-Specific Determinants of Horizontal and Vertical Intra-Industry Trade in Portugal. *Fifth GEP Postgraduate Conference*, University of Nottingham, 2006.
[17] Linder, S. B.: *An Essay on Trade and Transformation*. John Wiley & Sons Inc., New York, 1961.
[18] McPherson, M. A., et al.: International Trade and Developing Countries: An Empirical Investigation of the Linder Hypothesis. *Applied Economics* **33** (2001) 649-657.
[19] OECD: Intra-Industry and Intra-Firm Trade and the Internationalization of Production, 2002 [online]. Available: http://www.oecd.org/eco/outlook/2752923.pdf.
[20] Qureshi, U. A., et al.: Linder's trade thesis: a further examination. *Southern Economic Journal* **46** (1980), 933-936.
[21] Rahman, M. M.: Australia's Global Trade Potential: Evidence from the Gravity Model Analysis. *Oxford Business and Economics Conference* (2009).
[22] Sailors, J. W., et al.: Empirical verification of Linder's trade thesis. *Southern Economic Journal* **40** (1973), 262-268.
[23] Thursby, J. G., and Thursby, M. C.: Bilateral trade flows, the Linder hypothesis, and exchange risk. *The Review of Economics and Statistics* **69** (1987), 488-495.

Application of structural models in Credit Risk

Petr Královič[1]

Abstract. In last years (especially after the financial crisis in 2008), there is a great emphasis on the correct state of the credit risk. The most commonly used method is probably the portfolio model CreditMetrics. Very sophisticated methods, based on the options theory, especially on Merton's model (1974), are structural models.
The paper focuses on the application of the structural models in credit risk. Specially, a KMV model will be used at a few Czech firms.
First there will be described the main approach of Merton models with details of KMV model. In the application part there is calculated the default probability, using this model. In conclusion, the results are evaluated.

Keywords: KMV model, credit risk, options, probability of default.

JEL Classification: G33
AMS Classification: 91G40

1 Introduction

Credit risk is the risk that the borrower doesn't pay his liabilities at the time. The credit risk of the firm is based on the computing the probability of the firm's solvency, or insolvency, i.e. default.

There are many rating agencies, such as Standard & Poor's, Moody's, etc., which focus of evaluating and computing this credit risk. One way, how to predict the corporate defaults, is using the structural models. These models apply the Merton's work [4]. In the late 1980's, KMV Corporation (founded by Oldrich Vasicek and Stephen Kealhofer) developed the KMV model, where the distance-to-default was used for the first time.

In present, many of practitioners use KMV model. For all, can be named Moody's, who bought the KMV model and Crosbie and Bohn [2] made some modifications. In academic research, for example, Bharath and Shumway [1] and Crouchy [3] are interested in Merton models.

In this paper there will be used one of the structural models. It is the most commercial successful model, called KMV model. This model determines the default probability, i.e. probability, which the firm is bankrupt.

The aim of this paper is to apply the KMV on real data from Czech economy. The first part of the paper will be devoted to description of the principles and the derivation of default probability in the KMV model. Real data from Prague stock exchange and the accounting data from firms will be applied in the second part. The last part of the paper will be the assessment of the results.

2 The Merton Model

The Merton Model was published in year 1974 for the first time [4]. This model is based on two important assumptions. First is about the debt structure. The firm has just issued a single homogenous class of bond, which has maturity at one time. So, the value of the firm at any time t ($V_{A,t}$) is the sum of equity E_t and debt D_t.

$$V_{A,t} = E_t + D_t \,. \tag{1}$$

The second basic assumption is about dynamic of the firm's assets. It is assumed to follow the geometric Brownian motion,

$$dV_A = \mu V_A dt + \sigma_A V_A dz \,, \tag{2}$$

[1] Ing. Petr Královič, VŠB – Technical University of Ostrava, Faculty of Economics, Department of Finance, Sokolska tr. 33, 701 21 Ostrava 1; e-mail: petr.kralovic@vsb.cz.

where V_A is the total value of the firm, μ is the rate of return on the assets, σ_A is the asset volatility and *dz* is a standard Wiener process.

It also assumes that coupon, and dividend payments and taxes are ignored.

If at time *T* the value of the firm is over debt, the bondholders will receive their debt in full and equity holders will get the rest. If the value of the firm is lower than the debt, the equity holders will receive nothing. Value of the equity can be written as the call option with the strike price equal the face value of the debt.

$$V_E = \max(V_A - D;0). \tag{3}$$

The Black-Scholes model shows that the value of the equity in the case of previous assumptions is:

$$V_E = V_A N(d_1) - e^{rT} DN(d_2), \tag{4}$$

where V_E, V_A is the value of the equity, resp. firm, *r* is the risk-free rate, *D* is the face value of the debt and $N(\cdot)$ is the cumulative standard normal distribution function, d_1 is

$$d_1 = \frac{\ln\left(\frac{V_A}{D}\right) + \left(r + \frac{\sigma_A^2}{2}\right)T}{\sigma\sqrt{T}} \tag{5}$$

and d_2 is

$$d_2 = d_1 - \sigma_A\sqrt{T}\,. \tag{6}$$

2.1 Merton's KMV model

This model has two important equations. The first one is the Black-Scholes equation (4), which shows that the value of the equity is a function of the value of firm's assets. The second one makes the relationship between the volatility of the equity and the volatility of the firm's assets. Using Ito's lemma, it can be shown that

$$\sigma_E = \left(\frac{V_A}{V_E}\right)\frac{\partial V_E}{\partial V_A}\sigma_A. \tag{7}$$

Because $\frac{\partial V_E}{\partial V_A} = N(d_1)$ (4), the volatilities of the equity and firm's assets are related by

$$\sigma_E = \left(\frac{V_A}{V_E}\right)N(d_1)\sigma_A. \tag{8}$$

In practice, the value of the equity can be observed from a stock exchange market. The volatility of the equity can be obtained by estimating the implied volatility from the option price or by using a historical approach. If we have the volatility of firm's assets and the value of the firm, the number of standard deviations that the firm is away from default. The KMV calls this as *distance-to-default (DD)*. With Black-Scholes formula, the distance-to-default is given by:

$$DD = \frac{\ln\left(\frac{V_A}{D}\right) + \left(\mu - \frac{\sigma_A^2}{2}\right)t}{\sigma_A\sqrt{t}}, \tag{9}$$

where μ is expected annual return of the firm's assets.

The probability of default, sometimes in KMV model called *expected default frequency (EDF)* is

$$P = N(-DD). \tag{10}$$

3 Application of KMV model

To apply the KMV model, three Czech firms were chosen. Chosen firms were AAA Auto, TATRA, a.s. and Unipetrol. All of these firms were in trouble in the past and that was the next reason, why there were only three firms.

Data were taken from the stock market, especially Burza cenných papírů Praha, or RM-System, which is over the counter market. For all firms data should be between years 2002 and 2011, but in case of AAA Auto, there are only data since September 2007, because it is the moment, when the company enters the Prague stock exchange. Accounting data were used too. From accounting data (balance sheet) was important face value of debt. The risk-free rate was used by Česká národní banka [7]. Time to maturity is provided as one year.

3.1 Computing model and results

Default probability of KMV model is computed by equation (10). Data were computed for every quarter of examined years. First step was calculating the volatility of the equity. It was calculated by the historical approach. The value of equity is market capitalization of the firm. With equations (4) and (8) were computed volatility and value of firm's assets. As written in the previous text, risk-free rate was used from database of Czech national bank.

A kind of trouble may be the book value of liabilities determining. Because in reality there isn't only one zero-coupon debt, but much more, KMV model is based on the assumption [2], [3], that debt is equivalent as the face value of all short-term liabilities and a half of the face value of all long-term liabilities.

If the volatilities and values of equities and firm's assets and values of liabilities are known, and with the expected time to maturity in one year, it is possible to compute the distance-to-default (9). Expected default frequency (10), which is the probability of default, is now simple to calculate. The graphic results are in the Figures 1-3.

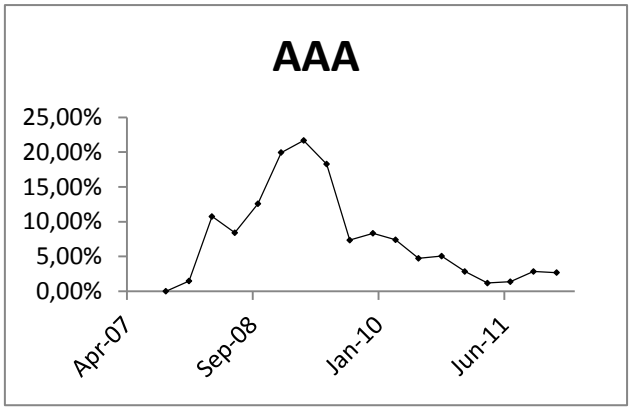


Figure 1 AAA Auto default probability

In Figure 1 is illustrated the default probability of the firm AAA Auto. It can be seen, that firm was in a bad situation between years 2008 and 2009. At this time the EDF was almost 20%. The main reason was the heavy decline of share prices. From about 50 CZK/share to about 10 CZK/share. After that, the share prices rise again, but not so much. At the present, firm announced the plan to retire the Prague stock exchange.

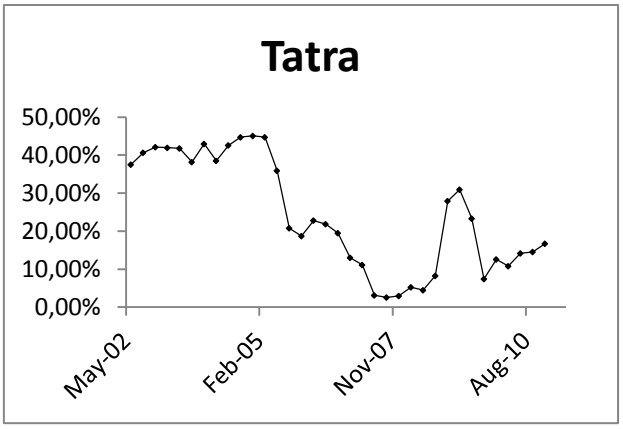


Figure 2 TATRA, a.s. default probability

The Figure 2 illustrated the evolution of the firm TATRA, a.s. In first part (until year 2005) is default probability about 40%. It is huge number. It was due to decreased production and the decline of share prices. For a long time, TATRA, a.s. was under the threat of bankruptcy. But all the time it was resisting. The results are only until year 2010, because later data were unavailable. In last 2 years the share prices still declines and in present are at historical minimum. The last information about this firm is, that it was sold to new subject, which promises a growth of the firm.

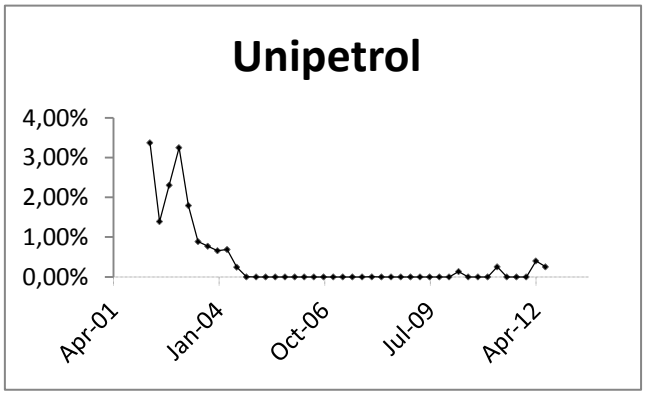


Figure 3 Unipetrol default probability

The Figure 3 illustrates the evolution of default probability of the firm Unipetrol. Until year 2005 was the probability level was between 1 – 4 %. As compared with two other firms, it is very good number. But at this time, Unipetrol was on the brink of bankruptcy. At this time a new Polish company PKN Orlen bought Unipetrol, and the growth started. Between years 2005 and 2010 was default probability basically 0%. A little bit of fluctuations were caused by economy crisis until year 2008.

4 Conclusion

The paper describes the basic idea of the Merton model, especially the Merton's KMV (or Moody's KMV). This model was applied to the real data of three Czech firms. The results, that threat of bankruptcy, which was given by poor economic or market situation, was shown at figures.

The advantage of this model is that it is based on market data, especially share prices. If data are good implemented, this model can be used for prediction the default probability.

There are two disadvantages of the model. First is the assumption of only one simple debt structure. But in real economy it isn't observed. So the results may be distorted, which is probably also the case of this paper. The second disadvantage is that the private companies have difficult valuable equity.

Last, the KMV model, how was described and with public available data, can be used also for prediction the default probability of the Czech public firms.

Acknowledgements

The paper is based on research activities sponsored through SGS research project No. SP2013/59. The support is greatly acknowledged.

References

[1] Bharath, S. T., and Shumway, T.: Forecasting default with the KMV-Merton model.
[2] Crosbie, P. J., and Bohn, J. R.: *Modeling Default Risk*, Research paper, Moody's KMV, 2002.
[3] Crouchy, M., et al.: A comparative analysis of current credit risk models. *Journal of Banking and Finance* **24** (2000), 57-117.
[4] Merton, R. C.: On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance* **29** (1974), 449-470.
[5] Hull, J.: *Options, Futures and Other Derivatives, 8th edition.* Prantice Hall, New Jersey, 2011
[6] Burza cenných papírů Praha (Prague stock exchange), online. www.pse.cz
[7] ČNB database ARAD, online. http://www.cnb.cz/docs/ARADY/HTML/index.htm

Analysis of competitiveness of selected countries using AHP models

Jana Kramulová[1], Josef Jablonský[2]

Abstract. A great effort to measure and analyse competitiveness of countries and regions can be seen in recent years. If we skip the discussion about whether a country or region can even be competitive or whether these issues should be left just for companies, several competitiveness indices can be found. Different authorities and researchers use different methods as well as indicators and perform different rankings. There is no single procedure that should be applied always when comparing countries or regions from the point of view of competitiveness. It can be hardly stated, which of the measurement approaches is the most proper.
The IMD World Competitiveness Online Database works with more than 300 indicators aggregating them subsequently in the Overall Competitiveness Ranking that evaluates 59 world countries in the year 2012. We compare in the paper these results with results from AHP methods using the same subgroups of indicators. The overall competitiveness is measured on the basis of AHP model with absolute measurement and application of weight expert estimation. The differences in both methods are being discussed and differences in results are analysed.

Keywords: competitiveness, AHP model with absolute measurement, countries, weight expert estimate.

JEL Classification: C44
AMS Classification: 90B50

1 Introduction

Scientists as well as politicians often compare various indicators of selected countries and/or regions. They compare e.g. economic performance, level of environmental degradation or unemployment. It is in this sense often spoken about competitiveness. Competitiveness is mainly connected with firms and industries and famous Porter's competitiveness diamond (Porter [12]). Successful firms usually concentrate not far from universities and other firms (suppliers etc.) to gain comparative advantages. This also contributes to faster diffusion of innovations. Important is also rivalry within the countries among their regions.

The process of considering company's competitiveness led step by step to assessments of regional competitiveness. Barkley [2] points out the quantity issue of different national and regional indices. Various approaches usually differ in number of indicators as well as in number of compared regions/nations. Barkley [2] also discusses the usefulness of such indices and rankings. Also Czesaný [4] considers the disadvantages of using composite indicators, especially the possibility of conclusions' simplification when inadequate composite indicator construction or problematic selection of partial indicators and determination of their weights. On the other hand Czesaný [4] sees also advantages, e.g. ability to summarize multiple perspectives, which can be among the main reasons, why competitiveness composite indicators are so popular.

Further problem connected with regional competitiveness, apart from composite indicators difficulties, can be seen straight in the regional competitiveness definition. Kitson et al. [11] describe regional competitiveness as an elusive concept which is very complicated to measure. Gardiner et al. [5] align regional competitiveness with success on export market. Their main focus is then on regional productivity. Boschma [3] is unlike Viturka [15] convinced that region can be assessed as a sum of companies located in its area. Therefore according to Boschma [3] "*successful region is just lucky to host more successful firms on average, and it does not have to be assumed that regions (like firms) act for this to happen*". On the other hand Viturka [15] states that costs and revenues (and subsequently competitiveness) of firms are (significantly) influenced by quality of local public administration and its decisions.

[1] University of Economics, Prague, Faculty of Economics, Department of Regional Studies, W. Churchill sq. 4, 130 67 Prague 3, Czech Republic, jana.kramulova@vse.cz.
[2] University of Economics, Prague, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill sq. 4, 130 67 Prague 3, Czech Republic, Jablon@vse.cz.

Despite aforementioned obstacles many composite indicators evaluating competitiveness of regions at almost any level (nations, regions, cities etc.) exist. We decided to compare in the paper rankings of IMD World Competitiveness Online Database with results from AHP method while using the same subgroups of indicators. The overall competitiveness is measured on the basis of AHP model with absolute measurement and application of weight expert estimation. The differences in both methods are being discussed as well as differences in results.

The paper is divided into several parts. After the theoretical background and overview of data set the computations and results are presented and discussed. Finally the conclusions are drawn.

2 Theoretical background

2.1 Data

As it was already stated, many competitiveness composite indicators exist. The main ones are according to [15] Global Competitiveness Index [14] published yearly by World Economic Forum (hereinafter WEF) in Global Competitiveness Report or data from World Competitiveness Yearbook (WCY) published yearly by Institute for Management and Development (hereinafter IMD). Among others are e.g. Regional Competitiveness Index calculated for EU (Annoni and Kozovska [1]) or countries (Czech Republic has RCICZ, methodology can be found in [6]), or UK Competitiveness Index (Huggins and Thompson [8], Huggins [7]) and many others. We compared results of both WEF and EU indices with IMD approach resulting in the fact, that all of them give very similar rankings (correlation coefficients were very high). Due to limited space we chose just one of them, i.e. IMD approach. However, the methodology introduced below could be applied to the other indices as well.

We decided to attract our attention to national level and cover into analysis indicators chosen by IMD World Competitiveness Online. We did not want to build up a totally new indicator composed from different "new" indicators but on the contrary we tried to compare results obtained from the database with results computed by using other methods while using the same data. The aim of this paper is to assess if the method selection has an impact on ranking of countries in relation to competitiveness issue.

IMD World Competitiveness Online database [9] contains information on 59 world countries (in the year 2012). According to methodology the overall competitiveness is divided into four groups and these are further again divided as shown in Figure 1. 329 indicators are covered in analysis, as Economic Performance contains 78 indicators, Government Efficiency 70, Business Efficiency 67 and Infrastructure 114.

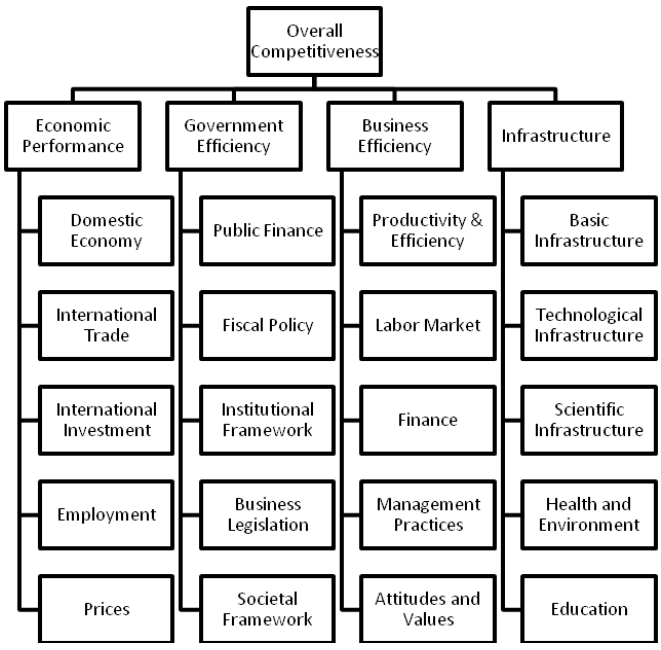


Figure 1 Hierarchy of groups and sub-criteria for evaluation of overall competitiveness according to IMD

2.2 Analytic Hierarchy Process

Competitiveness forms a typical example (another one can be sustainability, see [10]) of multiple criteria decision making (MCDM) problem. An ideal tool for analysis of such MCDM problem is the Analytic Hierarchy

Process (AHP) developed by Saaty [13]. The decision problem is divided into several partly independent hierarchical levels that can be analysed separately. AHP hierarchy usually consists of the following levels:

- overall objective of the decision problem – in our case the overall competitiveness of selected countries,
- criteria used for the evaluation – in our case 4 groups, each of them further divided into 5 sub-criteria (i.e. two hierarchical levels as shown in Figure 1),
- alternatives are usually at the lowest level of the hierarchy – in our case 59 world countries.

The principle of AHP method consists in the division of overall priorities from the topmost level of the hierarchy into the underling levels according to the decision maker's preferences. The decision maker expresses his preferences by comparing the importance of the elements at the given level with respect to an element of the preceding level. In standard AHP models the decision maker's judgments on elements at a given level with respect to an element at the level above are organized into paired comparison matrices. The judgments are estimates of the preference between two elements of the lower level with respect to the element at the level above. Let us denote the paired comparison matrix (1)

$$\mathbf{A} = \left\{ a_{ij} \middle| a_{ji} = \frac{1}{a_{ij}}, a_{ij} > 0, i, j = 1,2,\ldots,k \right\}, \tag{1}$$

where k is the number of elements in the particular comparison set of the lower level. Saaty [13] proposes to use a_{ij} integers in the range 1 through 9 to express preference, where 1 means that the i-th and the j-th element are equally important and 9 means that the i-th element is absolutely more important than the j-th element. The local priorities are derived by solving the following eigenvector problem (2):

$$\begin{aligned} &\mathbf{A}.\mathbf{v} = \lambda_{max}\mathbf{v}, \\ &\sum_{i=1}^{k} v_i = 1, \end{aligned} \tag{2}$$

where λ_{max} is the largest eigenvalue of **A** and **v** is the normalised right eigenvector belonging to λ_{max}. The eigenvector problem (2) is not always easily solvable. That is why several approximation methods can be used. Among them logarithmic least square method is the most popular. In this case the local priorities v_i are derived as the geometric average of all elements in the i-th row of the matrix **A** and then they are normalised:

$$\begin{aligned} v_i' &= \left(\prod_{j=1}^{k} a_{ij} \right)^{\frac{1}{k}}, i = 1,2,\ldots,k, \\ v_i &= \frac{v_i'}{\sum_{i=1}^{k} v_i'}, i = 1,2,\ldots,k. \end{aligned} \tag{3}$$

3 Computations and results

3.1 Our AHP model for competitiveness evaluation

Our AHP model for evaluation of competitiveness of selected countries is presented in Figure 2. As shown below the models consist of 5 levels. The first defines the main objective of the decision problem – evaluation of overall competitiveness. The next level contains four groups of criteria overall competitiveness consists of – we suppose that all of them have the same importance. That is why pairwise comparisons are not used for deriving priorities at this level and all the priorities of four main groups are $p_1 = p_2 = p_3 = p_4 = 0.25$. The priorities of these four main groups are further divided into underling level containing directly sub-criteria (each group is divided into 5 sub-criteria). The priorities of these sub-criteria (i.e. final criteria) are denoted as v_1, v_2, …, v_{20}, where index corresponds to each sub-criteria listed in Figure 1. As in case of the main groups the priorities of sub-criteria are derived by the pairwise comparisons according to their importance with respect to the main groups (preceding level). Of course the condition that the sum of priorities of the elements of an underling level equals to the priority of the element of the preceding level must hold.

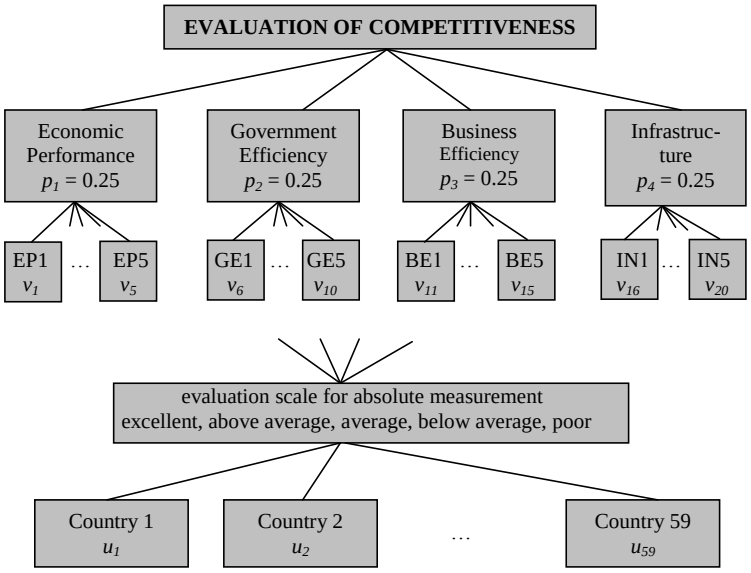


Figure 2 AHP model for evaluation of overall competitiveness

AHP offers two ways of measurements and deriving priorities – relative and absolute. The first uses standard pairwise comparisons. In case of a high number of alternatives or other elements at the level of the hierarchy the relative measurement cannot be used. This is the case of our model, in which the number of alternatives is very high (59 world countries). Subsequently, the number of pairwise comparisons would be vast and it would be technically almost impossible to do them. Therefore absolute measurement was used. It consists in evaluation of elements of the bottom level of the hierarchy (usually alternatives) by their assignment to one of the elements of the evaluation scale. Each of the elements of this scale has its numerical judgment given either directly by decision maker or by pairwise comparisons of elements of the evaluation scale. For simplicity let us suppose that there is the same five-elements evaluation scale with the identical numerical judgment q_i, i = 1,2,…,5 for all criteria of the model. Table 1 presents how the judgments q_i are derived by pairwise comparisons. Consistency index (hereinafter C.I.) should be in all pairwise comparison matrices lower than 0.1, which is fulfilled (C.I. equals 0.059).

Absolute measurement	E	AA	A	BA	P	q_i	C.I.
Excellent	1	3	5	7	9	0.5128	0.059
Above average	1/3	1	3	5	7	0.2615	
Average	1/5	1/3	1	3	5	0.1290	
Below average	1/7	1/5	1/3	1	3	0.0634	
Poor	1/9	1/7	1/5	1/3	1	0.0333	

Table 1 Judgment of elements of the scale with absolute measurement

Let us finally denote the value of the i-th alternative by criterion j as x_{ij}, i=1, 2, …, 59, j =1, 2, …, 20. According to criterion values the alternatives are evaluated by one of the elements of the evaluation scale and the values x_{ij} are replaced by one of the numerical judgment of the evaluation scale – they can be denoted as y_{ij}. Final utility of evaluated alternatives (countries) is simple weighted average of y_{ij} values:

$$u_i = \sum_{j}^{20} v_j y_{ij}, i = 1,2,\ldots,59. \tag{4}$$

The alternatives can be ranked by their utilities u_i as shown in equation (4).

3.2 Computational procedure and results

In the first step of application of the AHP model the weights of all 20 sub-criteria are derived. Table 2 contains pairwise comparison matrix for sub-criteria weights deriving of Economic Performance group. Consistency

index of this matrix equals 0.03, again fulfilling the condition mentioned above. The same procedure (i.e. pair-wise comparison matrix) is applied for remaining three groups – Government Efficiency, Business Efficiency and Infrastructure. All comparisons were object of expert estimation. Due to lack of space Table 3 shows just the resulting weights, not the whole matrices for the other groups. Corresponding consistency indices are again lower than 0.1, namely 0.053 for Government Efficiency, 0.042 for Business Efficiency and 0.026 for Infrastructure.

Economic Performance	DE	IT	II	E	P	v_i	C.I.
Domestic Economy	1	1/3	5	3	1/2	0.1677	0.03
International Trade	3	1	8	6	2	0.4378	
International Investment	1/5	1/8	1	1/3	1/7	0.0372	
Employment	1/3	1/6	3	1	1/5	0.0722	
Prices	2	1/2	7	5	1	0.2852	

Table 2 Weights of sub-criteria of Economic Performance

Government Efficiency	v_i	Business Efficiency	v_i	Infrastructure	v_i
Public Finance	0.0598	**Productivity & Efficiency**	0.4460	**Basic Infrastructure**	0.5098
Fiscal Policy	0.1276	**Labour Market**	0.0816	**Technological Infrastructure**	0.0521
Institutional Framework	0.5079	**Finance**	0.0561	**Scientific Infrastructure**	0.2055
Business Legislation	0.2623	**Management Practices**	0.1690	**Health and Environment**	0.1163
Societal Framework	0.0423	**Attitudes and Values**	0.2472	**Education**	0.1163

Table 3 Weights and consistency indices of sub-criteria of remaining three groups

As all the priorities of four main groups are the same ($p_1 = p_2 = p_3 = p_4 = 0.25$), we decided not to normalise the given weights v_i according to the weight of the preceding node now, but at the very end of computational procedure.

Subsequently, after obtaining weights for all 20 sub-criteria four quintiles were computed to allow matching of the scale with absolute measurement. Afterwards values of four main groups were computed, equation (5) serves as an example for group Economic Performance (5):

$$u_i^{EP} = \sum_{j}^{5} v_j y_{ij}, i = 1,2,\ldots,59. \tag{5}$$

Country	IMD value	Rank IMD	AHP u_i^{EP}	Rank AHP	Rank difference
Germany	77,51	5	0,4028	3	2
Greece	10,79	58	0,0551	59	1
Hong Kong	79,16	4	0,3761	8	4
Hungary	47,31	35	0,1850	33	2

Table 4 Four countries example of rank comparison

As is shown in Table 4, we were able to compare in each main group results of original IMD values and u_i^{EP} values computed by AHP as well as to assess rank differences among both methods in each country. Computing of utility value of four main groups made us possible to finalise the procedure and compute the overall competitiveness ranking of countries u_i by simple average of the four components.

4 Discussion and conclusion

Due to lack of space we will just discuss part of the results obtained. Table 5 shows first 5 and last 5 countries according to both methodologies. We computed the overall competitiveness of all 59 countries, ranked them and then compared with results of IMD methodology. There were slight differences, but the average rank difference equals 3.02, what we see as a very good result. The greatest differences were observed in cases of Bulgaria (the greatest positive shift: +13 positions from 54th to 41st position) and Portugal and USA (the greatest negative shift: –8 positions from 41st to 49th and from 2nd to 10th position). While in case of Bulgaria the main discrepancy between both approaches was in group of Economic Performance (+35), in case of Portugal and USA it was in

group of Infrastructure (–9 and –10). Generally the biggest differences were inside group of Economic Performance, on the other hand the lowest were in case of Infrastructure group. It would be useful to discuss why the results of AHP and IMD approaches vary in different intensities for diverse indicators. Deeper analysis of these relations will be included in the follow-up publications, as here is a limited space for the presentation of comprehensive rankings.

First 5	IMD	AHP	Last 5	IMD	AHP
1	Hong Kong	Hong Kong	**55**	Argentina	Colombia
2	USA	Singapore	**56**	Ukraine	Romania
3	Switzerland	Sweden	**57**	Croatia	Ukraine
4	Singapore	Switzerland	**58**	Greece	Greece
5	Sweden	Germany	**59**	Venezuela	Venezuela

Table 5 First 5 and last 5 countries in both methodologies (IMD and AHP) rankings

As a conclusion, our case study demonstrated that AHP method with absolute measurement represents a simple, fast and useful tool even for a big amount of alternatives. Presented case study also shows that results of this procedure correspond with ranking determined by more sophisticated method.

Acknowledgements

The research is supported by the Czech Science Foundation, project no. P403/12/1387 and University of Economics, Prague, project no. IGA F4/11/2013.

References

[1] Annoni, P. and Kozovska, K.: *EU Regional Competitiveness Index RCI 2010*. Publications Office of the European Union, Luxembourg, 2010.
[2] Barkley, D.L.: Evaluations of Regional Competitiveness: Making a Case for Case Studies. *Review of Regional Studies* **38** (2008), 121-143.
[3] Boschma, R.A.: Competitiveness of Regions from an Evolutionary Perspective. *Regional Studies*. Special Issue **38** (2004), 1001-1014.
[4] Czesaný, S.: Indikátory udržitelného rozvoje [Indicators of sustainable development]. *Statistika* **43** (2006), 431-434.
[5] Gardiner, B. and Martin, R. and Tyler, P.: Competitiveness, Productivity and Economic Growth across the European Regions. *Regional Studies*. Special Issue **38** (2004), 1045-1067.
[6] GaREP and RegioPartner: Hodnocení regionální konkurenceschopnosti krajů a správních oblastí obcí s rozšířenou působností. Metodika [Evaluation of regional competitiveness of regions and municipalities with extended competence. Methodology]. 2011. Available at: http://www.regiopartner.cz/cz/download/1404041849/?at=1. Last accessed [2013-05-14].
[7] Huggins, R.: Creating a UK Competitiveness index: Regional and local benchmarking. *Regional Studies* **37** (2003), 89-96.
[8] Huggins, R. and Thompson, P.: *UK Competitiveness Index 2010*. Centre for International Competitiveness, Cardiff School of Management, University of Wales Institute, Cardiff, 2010.
[9] IMD World Competitiveness Online. Available at: https://www.worldcompetitiveness.com/online. Last accessed [2013-01-24].
[10] Jablonský, J. and Kramulová, J.: AHP model for evaluation of sustainable development in EU-27 countries. In: *Sustainable Development: Industrial Practice, Education & Research* (Digiesi, S., Mossa, G., Mummolo, G. And Ranieri, L., eds.). DIMEG Edizioni, Politecnico di Bari, Bari, 2010, 26–31.
[11] Kitson, M. and Martin, R. and Tyler, P.: Regional Competitiveness: An Elusive yet Key Concept? *Regional Studies*. Special Issue **38** (2004), 991-999.
[12] Porter, M.: *The Competitive Advantage of Nations*. The Free Press, New York, 1998.
[13] Saaty, T.L.: *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. RWS Publications, Pittsburgh, 1990.
[14] Schwab, K. (ed.) and Sala-i-Martín, X.: *The Global Competitiveness Report 2012–20013: Full Data Edition*. World Economic Forum, Geneva, 2012.
[15] Viturka, M. *et al.*: *Kvalita podnikatelského prostředí, regionální konkurenceschopnost a strategie regionálního rozvoje České republiky [Business environment quality, regional competitiveness and regional development strategy of the Czech Republic]*. Grada Publishing, Praha, 2010.

Explaining the differences in on-trade prices of beer in Europe

Ondřej Krčál [1]

Abstract. The goal of this paper is to explain the differences in prices of beer in the hospitality sector in Europe. Using cross-section data for 27 European countries, the paper finds that a rise in labor costs and a reduction in on-trade per capita consumption and number of beers served per employee in the hospitality sector increases the price of beer and the value added in the hospitality sector. Moreover, these variables explain a large proportion of the variance in prices and the value added in the European data. The empirical results are consistent with the predictions of an extension of the Salop circular city model, which is used as a model of beer consumption in the hospitality sector.

Keywords: price, draft beer, circular city, Salop, Europe.

JEL classification: L10, L81, R30

1 Introduction

On-trade prices of beer differ widely in Europe. For 10 euro, consumers are able to buy approximately 1.5 beers in a pub in Norway, 3 beers in Germany, 5 beers in Slovenia, and more than 10 beers in the Czech Republic. On the other hand, prices of other products, such as groceries, garments or electronics, are relatively similar in all European countries. Even prices of other products sold in the hospitality sector, like coffee or accommodation, tend to differ less.

The aim of this paper is to explain differences in on-trade prices of beer in Europe. Although the question is of interest even for people outside of the economic profession, there is no academic work, to my knowledge, explaining the differences in on-trade prices of beer. Most of the papers studying beer market follow different goals, see e.g. [8] for a study of divestiture of brewer-owned public houses in the UK, or [5] for a survey of literature studying the demand for beer. For an application of this literature in competition economics, see [6]. In the study with a goal similar to this paper, Culbertson & Bradford [1] explore the factors responsible for the inter-state differences in off-trade (retail) prices of beer in the US. They find that "a substantial proportion of the variance can be explained by demand pressure, transportation costs, and the price of substitutes, as well as taxes and the presence of legally-mandated exclusive territories governing the wholesale distribution of beer." ([1], p. 275)

This paper finds that on-trade prices of beer in Europe are increasing in labor costs and decreasing in on-trade consumption and beer turnover per employee in the hospitality sector and that a large proportion of the variance in on-trade prices of beer can be attributed to these factors. The effects of these factors are apparent in the pattern of beer prices across Europe. For instance, beer in Ireland or France is more expensive than beer in the Czech Republic or Estonia probably because of differences in labor costs. While an hour of labor in accommodation and food services in France and Ireland costs 21.7 and 15.76 euro respectively, the cost of labor in the Czech Republic or Estonia is approximately 5.3 euro. However, beer in Estonia is still more than twice as expensive as in the Czech Republic. Here, higher Estonian prices can be explained by lower yearly per-capita consumption and beer turnover in the hospitality sector (8 liters per capita, 7.050 liters of beer per employee) compared to the Czech Republic (80 liters per capita, 25.882 liters of beer per employee).

The rest of the paper is organized as follows. Section 2 presents an extension of the Salop circular city model. Section 3 introduces the data, discusses the empirical strategy used, and presents the results of the empirical analysis. Finally, Section 4 concludes.

[1] Masaryk University/FEA, Economics department, Lipová 41a, 602 00 Brno, krcalo@mail.muni.cz

2 A model of beer consumption in the hospitality sector

In this section, I present an extension of the Salop model with endogenous number of firms ([7]). The Salop model seems to be an appropriate model of beer consumption in the hospitality sector which is characterized by spacial product differentiation and low entry costs.

In the circular city model, N pubs serve identical product with quasi-fixed costs F and constant marginal costs c. The profit of pub i is

$$\pi_i(q_i) = \begin{cases} (p_i - c)q_i - F & \text{for } q_i > 0 \\ 0 & \text{for } q_i = 0, \end{cases}$$

where p_i is the price of beer in pub i and q_i is the quantity of beer sold by pub i. The city is inhabited by consumers who are uniformly distributed on a circle with unit circumference. Each consumer visits a pub v times per year, buys a total number of b liters of beer, $b_v = b/v$ liters on each visit. Consumers go to the pub that minimizes their total spending on beer plus total transportation costs $bp_i + v\tau x_i$, where τ represents transportation costs per unit of distance and x_i is the distance to pub i.

The model is a two stage game. In the first stage, N pubs enter the market. In the second stage, they choose profit-maximizing prices. I solve the model using backward induction.

In a symmetric equilibrium in the second stage, the distance between N pubs will be $1/N$. Assuming that pubs $i+1$ and $i-1$ charge the same equilibrium price p^*, the consumer that is indifferent as to whether she buys from pub i or $i+1$ lives at a distance $\hat{x}_i$ from pub i. The distance $\hat{x}_i$ is determined by $bp_i + v\tau\hat{x}_i = bp^* + v\tau(1/N - \hat{x}_i)$. Substituting $v = b/b_v$, the demand function of pub i is

$$q_i(p_i, p^*) = 2b\hat{x}_i = \frac{bb_v(p^* - p_i)}{\tau} + \frac{b}{N}.$$

Pub i chooses the price p_i for which

$$\frac{\partial \pi_i(p_i, p^*)}{\partial p_i} = \frac{bb_v(p^* - 2p_i + c)}{\tau} + \frac{b}{N} = 0.$$

In a symmetric equilibrium in the second stage, the price of firm i is

$$p_i(N) = p^*(N) = c + \frac{\tau}{b_v N}.$$

In the first stage, firms know profit-maximizing prices for different numbers of firms N. Additional firms are willing to enter the market until

$$\pi_i(p^*) = (p^*(N^*) - c)\frac{b}{N^*} - F = \frac{b\tau}{b_v N^{*2}} - F = 0.$$

The equilibrium number of firms in the industry is $N^* = \sqrt{\frac{b\tau}{b_v F}}$. The equilibrium price is

$$p^*(N^*) = c + \frac{\tau}{b_v N^*} = c + \sqrt{\frac{F\tau}{b_v b}}. \tag{1}$$

The model predicts the following effects of individual variables on the equilibrium price of beer:

- Marginal costs c enter the price of beer directly because pubs earn zero profits and a rise in marginal costs has no effect on the equilibrium number of pubs in the market.
- Quasi-fixed costs F increase the price of beer as prices charged by zero-profit pubs have to reflect costs. However, prices increase at a diminishing rate because quasi-fixed costs F reduce the equilibrium number of pubs N^* in the market, so that each pub serves a larger market.
- Transportation costs per unit of distance τ increase the price of beer because they reduce competition in the market. They increase prices at a diminishing rate because they also raise the equilibrium number of pubs N^*. The number of beers per visit b_v has the opposite effect since it reduces total transportation costs of consumers.
- The number beers b reduces the price of beer at a diminishing rate because it increases the equilibrium number of pubs N^* (and consequently competition in the market) at a diminishing rate.

3 Empirical Analysis

This section provides some evidence on the determinants of the price of beer using cross-section data from 27 European countries. First, I introduce the data and empirical methodology. Then I present the results of the empirical analysis and discuss the main findings.

3.1 Data and methodology

The data is compiled from two main sources: most variables are calculated from the 2008 data published in the Ernst & Young study "The Contribution Made by Beer to the European Economy", edition 2009 (see [2]), and the 2008 labor cost data taken from the Eurostat pocketbook "Labour Market Statistics" (see [4], p. 90). The data in both sources are available for the following 27 European countries: Austria, Bulgaria, Cyprus, Czech Republic, Denmark, France, Estonia, Finland, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and United Kingdom.

In the empirical analysis I use two dependent variables:

- Price of beer in the hospitality sector net of taxes (euro/liter) is calculated as

$$\frac{\text{on-trade price (1 liter)}}{1 + \text{VAT rate}} - \text{excise tax},$$

  where the excise tax is calculated as total excise revenues on beer divided by total consumption in liters. All values are taken from [2] except for the 2008 VAT rates of Italy (20%) and United Kingdom (17.5%)[1] which are not listed in [2] and were therefore found on the internet.

- Value added per liter of beer in the hospitality sector (euro) is calculated as value added due to sale of beer in the hospitality industry divided by total consumption in hospitality (total consumption in 2008 times the share of consumption in hospitality). All values are taken from [2].

The dependent variables are regressed against the following independent variables:

- Labor cost (euro) is the 2008 hourly labor cost in accommodation and food services activities taken from [4].

- On-trade consumption per capita (liters) is calculated as total consumption in hospitality (total consumption in 2008 times the share of consumption in hospitality) taken from [2] divided by 2008 population found on the Eurostat webpage.[2]

- Beer turnover per employee and year in the hospitality sector (liters) calculated as total consumption in hospitality (2008 total consumption in liters times the share of consumption in hospitality) divided by the employment in the hospitality sector due to beer sales. All values are taken from [2].

I made an additional adjustment in the data. Since the price of a liter of beer in Estonia (4.5 euro) in [2] is lower than the calculated value added in the hospitality sector (4.7 euro), I replaced the data on the share of on-trade consumption and employment in the hospitality sector for Estonia by the data from 2011 edition of the Ernst & Young study [3], which results in a more realistic value added in the hospitality sector (2.6 Euro). I also tried to omit Estonia from the data. The results without Estonia are similar to the results presented in the following section. Table 1 presents summary statistics for the adjusted data.

The regression equations presented in the following section should reflect the predictions of the model presented in Section 2.

- Labor cost is likely to enter both marginal and quasi-fixed costs. A rise in marginal costs increases the price of beer linearly and a rise in quasi-fixed costs at a diminishing rate. The parameter should be positive and, if part of labor cost is quasi-fixed, a rise in labor cost should increase the price of beer or value added in the hospitality sector at a diminishing rate.

[1] The VAT rate was 17.5% till 1 December 2008, when it was reduced to 15%.

[2] http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&language=en&pcode=tps00001&tableSelection=1&footnotes=yes&labeling=labels&plugin=1

- Beer turnover per employee and year in the hospitality sector reduces labor cost per liter of beer. Hence, a rise in turnover should reduce the price of beer at a diminishing rate. Beer turnover may be also a proxy for the number of beers per visit b_v. The intuition is as follows. The higher the consumption per visit of a pub, the higher is likely to be consumption in a given pub, and the higher tends to be the number of beers sold per employee, at least in small pubs with a low number of staff. For instance, productivity of an employee who is alone in a pub is determined by the consumption of beer in the pub, because there has to be at least one person present all the time. Hence, a rise in turnover reduces the price of beer at a diminishing rate, too.
- On-trade consumption per capita corresponds to the theoretical parameter of beer consumption b. The parameter should be negative and a rise in on-trade consumption per capita should reduce the price of beer or value added in the hospitality sector at a diminishing rate.

Variable	Mean	Median	Minimum	Maximum
On-trade price of beer net of taxes (euro/liter)	4.50355	4.72189	0.884490	9.27491
Value added in the hospitality sector (euro/liter)	2.12149	2.46733	0.342291	4.68165
Labor cost (euro/hour)	11.7811	11.3900	1.55000	25.8000
On-trade consumption per capita (liters/year)	30.2714	26.6636	5.46151	81.4110
Beer turnover per employee (liters/year)	9726.59	8838.12	5442.57	25882.5

Table 1 Summary Statistics

3.2 Results

Table 2 presents OLS regressions with the price of beer net of taxes as a dependent variable. Model I does not include beer turnover per employee. Model II with log of beer turnover has lower Schwarz, Akaike, and Hannan–Quinn information criteria than the same model with turnover and turnover squared. Both models with labor cost and labor cost squared have lower information criteria than the models with square root of labor cost, and the models with the log of labor cost are misspecified according to the RESET test (squares and cubes). Similarly, both models with the log of on-trade consumption have lower information criteria than the models with on-trade consumption and on-trade consumption squared.

	I	II
constant	2.168^{**}	14.73^{**}
	(0.9039)	(4.468)
labor cost	0.6475^{**}	0.6258^{**}
	(0.1146)	(0.1003)
labor cost squared	-0.01315^{**}	-0.01278^{**}
	(0.004359)	(0.003809)
log of on-trade consumption	-0.8894^{**}	-0.5092^{*}
	(0.2859)	(0.2830)
log of beer turnover per employee		-1.490^{**}
		(0.5217)
n	27	27
$\bar{R}^2$	0.8616	0.8944
ℓ	-34.06	-29.81

Standard errors in parentheses
* indicates significance at the 10 percent level
** indicates significance at the 5 percent level

Table 2 The OLS estimates for the dependent variable on-trade price of beer net of taxes

The estimates in Table 2 appear to be consistent with the predictions of the theoretical model. In both models, the labor cost increases the price of beer at a diminishing rate, and on-trade consumption per capita reduces the price of beer at a diminishing rate too. In model II, turnover reduces the price of beer at a diminishing rate. Beer turnover per employee in model II reduces significance of the log of on-trade consumption because correlation between beer turnover and on-trade consumption is relatively high (the correlation coefficient is 0.578). Additionally, all specifications explain more than 86% of the variation in the price of beer ($\bar{R}^2 > 0.86$).

Table 3 studies the effect of different variables on the value added per liter of beer in hospitality. Again, Model I does not include turnover. However instead of the labor cost and labor cost squared it uses the square root of labor cost. The model with the labor cost and labor cost squared gives also reasonable results but has higher information criteria. Model II adds the log of beer turnover per employee (the same equation with turnover and turnover squared has higher information criteria). The parameter of the log of on-trade consumption in model II is negative but not statistically significant. Model III omits the log of on-trade consumption. Finally, model IV uses the labor cost and labor cost squared instead of the square root of labor cost. In this specification, both the log of on-trade consumption per capita and beer turnover per employee are significantly lower than zero.

	I	II	III	IV
constant	0.1317	6.444**	7.733**	1.295**
	(0.4900)	(2.226)	(2.129)	(0.4161)
log of on-trade consumption	−0.3870**	−0.1980		−0.2627*
	(0.1242)	(0.1269)		(0.1514)
square root of labor cost	0.9881**	0.9499**	0.9413**	
	(0.08210)	(0.07304)	(0.07498)	
log of beer turnover per employee		−0.7450**	−0.9524**	
		(0.2578)	(0.2274)	
labor cost				0.2557**
				(0.05336)
labor cost squared				−0.004373**
				(0.002012)
beer turnover per employee				−5.457e-05**
				(2.374e-05)
n	27	27	27	27
$\bar{R}^2$	0.8557	0.8895	0.8829	0.8772
ℓ	−15.63	−11.45	−12.81	−12.27

Standard errors in parentheses
* indicates significance at the 10 percent level
** indicates significance at the 5 percent level

Table 3 The OLS estimates for the dependent variable value added per liter of beer

Again, the findings in Table 3 seem to support predictions of the theoretical model. Labor cost increases the value added per employee at a diminishing rate in all four models. Moreover, the data support the square root specification predicted directly by the theoretical model. Better support for the square root specification compared to the regressions presented in Table 2 might be due to the fact that the value added per employee does not include part the cost of beer production. Hence marginal costs c in equation (1) are lower, which makes the relationship between the labor cost and value added more square-root shaped. The on-trade consumption per capita reduces the value added in models I, II and III at a diminishing rate. However, it is less or not at all significant in models that include the beer turnover per employee. Similarly, the turnover per employee reduces the value added at a diminishing rate in models II and III and at a constant rate in model IV.

I also test the classical assumptions of the OLS regression. In all models, I do not reject normality of residuals using Doornik-Hansen test, Shapiro-Wilk W-test, Lilliefors test, and Jarque-Bera test at the 5% level. Similarly, neither Breusch-Pagan test nor White's test reject homoskedasticity at the 5% level. Furthermore, the RESET test (squares and cubes) in both models does not indicate incorrect specification at the 5% level. Finally, there may be endogeneity in both models since the price of beer or value added per liter of beer may affect the on-trade consumption and turnover per employee. However, the cross-section differences in these variables seem to be mainly due to different cultural norms and drinking habits in these countries. This view is supported by the fact that there is no statistically significant relationship between off-trade (retail) prices of beer and off-trade or on-trade consumption of beer per capita or turnover per employee. Similarly, there is no statistically significant relationship between the share of on-trade consumption and the difference between on-trade and off-trade prices or the ratio of off-trade to on-trade prices (the data on off-trade prices and consumption are taken from [2]).

4 Conclusion

The goal of the paper is to explain the differences in on-trade prices of beer in Europe. Using cross-section data from 27 European countries, the paper finds that more than 86% of the variance in the price of beer net of taxes is due to differences labor costs, on-trade consumption per-capita and beer turnover per employee in the hospitality sector. Specifically, it finds that a rise in labor costs and a reduction in consumption per capita and beer turnover increase the price of beer. Moreover, it finds that all the variables affect prices at a diminishing rate. The findings hold also if value added in the hospitality sector is used instead of prices. These findings are consistent with the predictions of the extended Salop model of circular city which seems to be appropriate for studying the determinants of beer prices in the hospitality sector.

Acknowledgements

This paper was created as a part of the specific research project no. MUNI/A/0797/2012 at Masaryk University.

References

[1] Culbertson, W. P., and Bradford, D.: The Price of Beer: Some Evidence from Interstate Comparisons, *International Journal of Industrial Organization* **9** (1991), 275–289.

[2] Ernst & Young: The Contribution Made by Beer to the European Economy, Full Report, European & Country Chapters, 2009, available at `http://www.brewersofeurope.org/docs/flipping_books/ey_full_2009/index.html`

[3] Ernst & Young: The Contribution Made by Beer to the European Economy, Full Report, European & Country Chapters, 2011, available at `http://www.brewersofeurope.org/docs/flipping_books/contribution_report_2011/index.html`

[4] Eurostat: Labour Market Statistics, 2009, available at `http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-32-11-798/EN/KS-32-11-798-EN.PDF`

[5] Fogarty, J.: The Demand for Beer, Wine and Spirits: A Survey of the Literature, *Journal of Economic Surveys* **24** (2010), 428–478.

[6] Houška, T., and Bil, J.: Definition of Relevant Market in Beer Industry: Application of LA-AIDS Model. In: *Proceedings of the 30th International Conference Mathematical Methods in Economics 2012* (J. Ramík, D. Stavárek, eds.), Karviná, 2012, 320–325.

[7] Salop, S. C.: Monopolistic Competition with Outside Goods, *The Bell Journal of Economics* **10** (1979), 141–156.

[8] Slade, M. E.: Beer and the Tie: Did Divestiture of Brewer-Owned Public Houses Lead to Higher Beer Prices? *The Economic Journal* **108** (1998), pp. 565–602.

A proper fuzzification of Saaty's scale and an improved method for computing fuzzy weights in fuzzy AHP

Jana Krejčí[1], Jana Talašová[2]

Abstract. A modification to the fuzzified Saaty's scale for designing a multiplicative fuzzy pairwise comparison matrix and an improved method for computing fuzzy weights of elements from a multiplicative fuzzy pairwise comparison matrix will be proposed. For simplicity of explanation triangular fuzzy numbers will be used in this paper. In many cases of the fuzzified AHP method that can be found in the literature, the fuzzification of the Saaty's scale is not appropriate. Therefore, a new fuzzification of the Saaty's scale will be suggested. Afterwards, an improvement of the formulas proposed by Buckley [4] and Pan and Yuan [9] will be suggested for computing fuzzy weights from a multiplicative fuzzy pairwise comparison matrix. In the new algorithm the reciprocity condition for the pairwise comparison matrix is taken into account in each step of the computation. The fuzzy weights resulting from the proposed fuzzification of the Saaty's AHP are less uncertain than those calculated by means of Buckley's algorithms.

Keywords: fuzzy pairwise comparison matrix, fuzzified Saaty's scale, triangular fuzzy number, fuzzy AHP

JEL classification: C44
AMS classification: 90C15

1 Introduction

The Analytic hierarchy process (AHP) is a methodology developed by T. L. Saaty in 1970s [10] for coping with unstructured complex problems using a hierarchical structure and pairwise comparisons. It is based on a construction of pairwise comparison matrices of alternatives against each criterion and a pairwise comparison matrix of criteria against the overall goal of the decision making problem. In both cases, the Saaty's 5-point scale or extended 9-point scale are used for making pairwise comparisons. Afterwards, weights of criteria and alternatives are computed from the pairwise comparison matrices and finally global priority weights of all alternatives are obtained by agregation.

After the AHP was introduced in 1970s many ways of fuzzification of the method were proposed. Van Laarhoven and Pedrycz [12] were first who applied fuzzy theory to the AHP. Buckley [4] initiated trapezoidal fuzzy numbers to express the decision maker's preferences and fuzzified geometric mean method for obtaining weights from the Saaty's pairwise comparison matrix. After then many other methods for obtaining fuzzy weights from fuzzy pairwise comparison matrices were introduced. However, as was shown in [3] the geometric mean is the only solution to the problem satisfying logical conditions. That is why we focus on this method and its fuzzification. For correct computation of fuzzy weights of elements also properly fuzzified Saaty's scale is essential. Therefore, we suggest also a modification to the fuzzification of the Saaty's scale.

This paper is organized as follows: In Section 2, the fuzzification of the Saaty's scale will be discussed. Section 3 will be focused on the fuzzification of the geometric mean method for obtaining weights from multiplicative pairwise comparison matrices. Numerical example of the computation of fuzzy weights from fuzzy pairwise comparison matrix will be given afterwards. Finally, the conclusion we will made in

[1]Palacký University Olomouc, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 12, 771 46 Olomouc, Czech Republic, jana.krejci01@upol.cz
[2]Palacký University Olomouc, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 12, 771 46 Olomouc, Czech Republic, talasova@inf.upol.cz

Section 4.

2 Fuzzification of the Saaty's scale

In this section the Saaty's scale used in the AHP will be remined. Then, fuzzifications of the Saaty's scale proposed by various authors will be described, lack of these fuzzifications will be shown and then more appropriate fuzzification of the Saaty's scale will be proposed.

Saaty [10] proposed the fundamental 5-point and extended 9-point scales of intensities of preferences to compare two elements in one level of the hierarchy with respect to an element from the upper level. Making comparisons of every two elements in one level of hierarchy with respect to an element from the upper level of the hierarchy a pairwise comparison matrix is constructed. Let us denote such a pairwise comparison matrix $A = \{a_{ij}\}_{i,j=1}^{n}$, where n is a number of elements $x_1, x_2, \ldots, x_n$ in one level of hierarchy which are compared and a_{ij} expresses how many times more important is the element x_i over the element x_j. When the element x_i is $a_{ij}-$times more important than the element x_j then the element x_j takes only $\frac{1}{a_{ij}}-$th of the importance of the element x_i. Therefore, the pairwise comparison matrix has to be reciprocal, i.e. $a_{ij} = \frac{1}{a_{ji}}$, where $i, j = 1, \ldots, n$.

To capture the vagueness in the decision maker's preferences the Saaty's scales are fuzzified. For the fuzzification triangular fuzzy numbers are usually used. Therefore we will use them in this article too.

A triangular fuzzy number $\tilde{c}$ is a fuzzy number whose membership function is uniquely determined by a triplet of significant values $c_1 \leq c_2 \leq c_3$ in the following way:

$$\tilde{c}\,(x) = \begin{cases} \frac{x-c_l}{c_2-c_1}, & c_1 \leq x \leq c_2, \\ \frac{c_3-x}{c_3-c_2}, & c_2 < x \leq c_3, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

For a triangular fuzzy number $\tilde{c}$ whose membership function is given by (1), the notation $\tilde{c} = (c_1, c_2, c_3)$ will be used hereafter.The real numbers c_1, c_2, and c_3 will be called the lower, the middle, and the upper significant values of the triangular fuzzy number $\tilde{c}$, respectively.

Some examples of the fuzzification of the Saaty's scale proposed in the literature are shown in Table 1 and Table 2. Using the fuzzified Saaty's scale fuzzy pairwise comparison matrices are constructed. In the following, a Saaty's fuzzy pairwise comparison matrix is a matrix $\tilde{A} = \{\tilde{a}_{ij}\}_{i,j=1}^{n}$ whose elements $\tilde{a}_{ij}, i \neq j$, are triangular fuzzy numbers from the fuzzified Saaty's scale in case that the element x_i is more important than the element x_j or their reciprocal in case that the element x_j is more important than the element x_i. On the main diagonal of the Saaty's fuzzy pairwise comparison matrix there are always real number 1 as we compare one element with itself. Therefore, there is no fuzziness in this comparison.

Let us denote that the reciprocal of the triangular fuzzy number is not a triangular fuzzy number. However, for the sake of simplicity it is usually aproximated by a triangular fuzzy number. In the following, the reciprocal of a triangular fuzzy number $\tilde{c} = (c_1, c_2, c_3)$ will be defined as a triangular fuzzy number $\frac{1}{\tilde{c}} = \left(\frac{1}{c_3}, \frac{1}{c_2}, \frac{1}{c_1}\right)$.

Fuzzy	Membership functions used in:			Linguistic
number	[2]	[5]	[7]	term
$\widetilde{1}$	$(1,1,3)$	$(1,1,3)$	$(1,1,2)$	equal importance
$\widetilde{3}$	$(1,3,5)$	$(1,3,5)$	$(2,3,4)$	moderate importance
$\widetilde{5}$	$(3,5,7)$	$(3,5,7)$	$(4,5,6)$	strong importance
$\widetilde{7}$	$(5,7,9)$	$(5,7,9)$	$(6,7,8)$	very strong importance
$\widetilde{9}$	$(7,9,11)$	$(7,9,9)$	$(8,9,9)$	extreme importance

Table 1 Examples of the fuzzified 5-point scale applied in the literature.

The fuzzifications of the Saaty's scale mentioned above are not suitable for constructing fuzzy pairwise comparison matrices. The problem is caused by the fuzzy number $\tilde{1}$ which means that two elements are

Fuzzy	Membership functions used in:			Linguistic
number	[1]	[6]	[11]	term
$\widetilde{1}$	$(1,1,1)$	$(1,1,1)$	$(1,1,1)$	equal importance
$\widetilde{2}$	$(1,2,3)$	$(1,2,3)$	$(2-x^*,2,2+x)$	
$\widetilde{3}$	$(2,3,4)$	$(2,3,4)$	$(3-x,3,3+x)$	moderate importance
$\widetilde{4}$	$(3,4,5)$	$(3,4,5)$	$(4-x,4,4+x)$	
$\widetilde{5}$	$(4,5,6)$	$(4,5,6)$	$(5-x,5,5+x)$	strong importance
$\widetilde{6}$	$(5,6,7)$	$(5,6,7)$	$(6-x,6,6+x)$	
$\widetilde{7}$	$(6,7,8)$	$(6,7,8)$	$(7-x,7,7+x)$	very strong importance
$\widetilde{8}$	$(7,8,9)$	$(7,8,9)$	$(8-x,8,8+x)$	
$\widetilde{9}$	$(9,9,9)$	$(8,9,10)$	$(8,9,9)$	extreme importance

$^*x$ is a fuzzification factor

Table 2 Examples of the fuzzified 9-point scale applied in the literature.

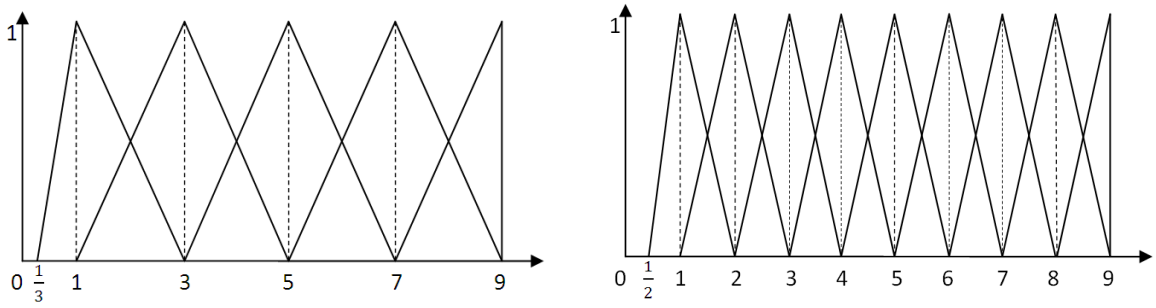


Figure 1 Suitable fuzzified 5-point end extended 9-point scales.

of the same importance. In Table 2, the fuzzy number $\tilde{1}$ is defined as a real number $(1,1,1)$. However, $\tilde{1}$ should contain some fuzziness as we want to deal with vagueness in meanings. In Table 1, on the other hand, the fuzzy number $\tilde{1}$ is defined as a triangular fuzzy number $(1,1,c)$, where $c=2$, resp. $c=3$. Let us assume an example in which two criteria, let say $\mathcal{K}_1$ and $\mathcal{K}_2$, are of the same importance. Then in the fuzzy pairwise comparison matrix of criteria there will be $\tilde{a}_{12}=(1,1,c)$. And because the fuzzy pairwise comparison matrix has to be reciprocal, it follows that $\tilde{a}_{21}=\frac{1}{\tilde{a}_{12}}=\frac{1}{(1,1,c)}=\left(\frac{1}{c},1,1\right)$. However, when an expert says that the criterion $\mathcal{K}_1$ is of the same importance as the criterion $\mathcal{K}_2$, then also the criterion $\mathcal{K}_2$ should be of the same importance as the criterion $\mathcal{K}_1$. Therefore, it should hold that $\tilde{a}_{12}=\tilde{a}_{21}$. According to this consideration we should define the fuzzy number $\tilde{1}$ in such a way that the equation $\tilde{1}=\frac{1}{\tilde{1}}$ holds. Therefore, it follows that it should be defined as $\tilde{1}=\left(\frac{1}{c},1,c\right)$. The vagueness of meaning of other linguistic terms from the fuzzified Saaty's scale is understood so that the meaning of the linguistic term ranges between two neigbouring values of the scale and the typical numerical value corresponding to the linguistic term is the value in the core of the corresponding fuzzy number. The properly fuzzified 5-point and extended 9-point scales are shown in Figure 1.

The elements of the fuzzified Saaty's scales shown in Figure 1 together with their reciprocal form a special structure called *fuzzy scale* on the interval $\left[\frac{1}{9},9\right]$. Generally, the fuzzy numbers $\tilde{c}_1,\ldots,\tilde{c}_n$ defined on the interval $[a,b]$, $a,b\in\mathbb{R}$ form a fuzzy scale if $\sum_{i=1}^{n}\tilde{c}_i(x)=1,\forall x\in[a,b]$, and if the fuzzy numbers are numbered in conformity with their linear ordering, i.e. $\tilde{c}_1<\cdots<\tilde{c}_n$.

3 Fuzzification of the geometric mean method

In this section the fuzzification of the geometric mean method for obtaining fuzzy weights of elements from the Saaty's fuzzy pairwise comparison matrix will be discussed. The approaches of Buckley, Pan and Yuan will be described, their lacks will be pointed out and then improved formulas for computing

fuzzy weights will be proposed.

As was already mentioned in Introduction, the geometric mean method was fuzzified by Buckley in [4]. The formulas for computing fuzzy weight $\tilde{w}_i = (w_{i1}, w_{i2}, w_{i3})$ of the $i-$th element, where $i = 1, \ldots, n$, from the fuzzy pairwise comparison matrix $\tilde{A} = \{\tilde{a}_{ij}\}_{i,j=1}^{n}$ proposed by Buckley are as follow:

$$w_{i1} = \frac{\sqrt[n]{\prod_{j=1}^{n} a_{ij1}}}{\sum_{k=1}^{n} \sqrt[n]{\prod_{j=1}^{n} a_{kj3}}}, \tag{2}$$

$$w_{i2} = \frac{\sqrt[n]{\prod_{j=1}^{n} a_{ij2}}}{\sum_{k=1}^{n} \sqrt[n]{\prod_{j=1}^{n} a_{kj2}}}, \tag{3}$$

$$w_{i3} = \frac{\sqrt[n]{\prod_{j=1}^{n} a_{ij3}}}{\sum_{k=1}^{n} \sqrt[n]{\prod_{j=1}^{n} a_{kj1}}}, \tag{4}$$

where $\tilde{a}_{ij} = (a_{ij1}, a_{ij2}, a_{ij3})$. In general, these fuzzy weights are not constrained to the interval $[0, 1]$. In that case, Buckley [4] suggests to multiply the fuzzy weights by a normalizing constant so that the supports of all fuzzy weights are in the interval $[0, 1]$. In that case the sum of the middle values of the fuzzy weights is not equal to one. However, Buckey did not proceeded correctly during the construction of the formulas (2), (3) and (4). At first, he computed geometric means of the rows of the Saaty's fuzzy pairwise comparison matrix and then computed their sum. Thus he obtained two triangular fuzzy numbers which he then devided. However, he did not take into account that for the normalization the extention principle has to by applied within the whole formula at one time.

Unlike Buckley, Pan and Yuan [9] realized the problem and applied the extention principle on the whole formula for computing normalized fuzzy weights:

$$w_{i1} = \frac{\sqrt[n]{\prod_{j=1}^{n} a_{ij1}}}{\sqrt[n]{\prod_{j=1}^{n} a_{ij1}} + \sum_{k=1, k\neq i}^{n} \sqrt[n]{\prod_{j=1}^{n} a_{kj3}}}, \tag{5}$$

$$w_{i2} = \frac{\sqrt[n]{\prod_{j=1}^{n} a_{ij2}}}{\sum_{k=1}^{n} \sqrt[n]{\prod_{j=1}^{n} a_{kj2}}}, \tag{6}$$

$$w_{i3} = \frac{\sqrt[n]{\prod_{j=1}^{n} a_{ij3}}}{\sqrt[n]{\prod_{j=1}^{n} a_{ij3}} + \sum_{k=1, k\neq i}^{n} \sqrt[n]{\prod_{j=1}^{n} a_{kj1}}}. \tag{7}$$

Fuzzy weights obtained from the Saaty's fuzzy pairwise comparison matrix using the formulas (5), (6) and (7) are constrained to the interval $[0, 1]$ and the sum of the middle significant values of the fuzzy weights is equal to one. However, in these formulas the condition of reciprocity of the fuzzy pairwise comparison matrix $\tilde{A}$ is still not taken into account. Therefore, with respect to the reciprocity the significant values shoud be computed according to these formulas:

$$w_{i1} = \min \left\{ \frac{\sqrt[n]{\prod_{j=1}^{n} a_{ij}}}{\sum_{k=1}^{n} \sqrt[n]{\prod_{j=1}^{n} a_{kj}}} \; ; \; a_{kj} \in [a_{kj1}, a_{kj3}], \; k, j = 1, \ldots, n, \; a_{kj} = \frac{1}{a_{jk}} \right\}, \tag{8}$$

$$w_{i2} = \frac{\sqrt[n]{\prod_{j=1}^{n} a_{ij2}}}{\sum_{k=1}^{n} \sqrt[n]{\prod_{j=1}^{n} a_{kj2}}}, \tag{9}$$

$$w_{i3} = \max \left\{ \frac{\sqrt[n]{\prod_{j=1}^{n} a_{ij}}}{\sum_{k=1}^{n} \sqrt[n]{\prod_{j=1}^{n} a_{kj}}} \; ; \; a_{kj} \in [a_{kj1}, a_{kj3}], \; k, j = 1, \ldots, n, \; a_{kj} = \frac{1}{a_{jk}} \right\}. \tag{10}$$

Let us remark that the formulas (2), (6) and (9) for computing middle significant values of the fuzzy weights are the same.

The algorithm for the compution of fuzzy weights according to the formulas (8) and (10) is described in [8]. The supports of the fuzzy weights computed using the formulas (8) and (10) are in the interval $[0,1]$ and the sum of the middle significant values is equal to one. Moreover, the supports of the fuzzy weights are narrower than the supports of the fuzzy weights computed using the formulas (2), (3) and (4), resp. (5), (6) and (7) because we took into account the condition of the reciprocity of the matrices used in the formulas.

Let us remark that the fuzzy weights obtained from the Saaty's fuzzy pairwise comparison matrix are not triangular fuzzy numbers. We work only with their $3-$point representation. However, the representation of real fuzzy weights by triangular fuzzy numbers is sufficient in real applications.

In the following example we will compute fuzzy weights of the criteria using the formulas proposed by Buckley, Pan and Yuan and by us to compare the results.

Example 1. Let us assume the fuzzy pairwise comparison matrix of the 6 criteria from [8]:

$$\begin{array}{c} \\ \mathcal{K}_1 \\ \mathcal{K}_2 \\ \mathcal{K}_3 \\ \mathcal{K}_4 \\ \mathcal{K}_5 \\ \mathcal{K}_6 \end{array} \begin{array}{c} \begin{array}{cccccc} \mathcal{K}_1 & \mathcal{K}_2 & \mathcal{K}_3 & \mathcal{K}_4 & \mathcal{K}_5 & \mathcal{K}_6 \end{array} \\ \begin{pmatrix} 1 & (1,3,5) & (3,5,7) & (3,5,7) & (3,5,7) & (5,7,9) \\ \left(\frac{1}{5},\frac{1}{3},1\right) & 1 & (3,5,7) & (3,5,7) & (3,5,7) & (5,7,9) \\ \left(\frac{1}{7},\frac{1}{5},\frac{1}{3}\right) & \left(\frac{1}{7},\frac{1}{5},\frac{1}{3}\right) & 1 & \left(\frac{1}{3},1,3\right) & \left(\frac{1}{3},1,3\right) & (3,5,7) \\ \left(\frac{1}{7},\frac{1}{5},\frac{1}{3}\right) & \left(\frac{1}{7},\frac{1}{5},\frac{1}{3}\right) & \left(\frac{1}{3},1,3\right) & 1 & \left(\frac{1}{3},1,3\right) & (1,3,5) \\ \left(\frac{1}{7},\frac{1}{5},\frac{1}{3}\right) & \left(\frac{1}{7},\frac{1}{5},\frac{1}{3}\right) & \left(\frac{1}{3},1,3\right) & \left(\frac{1}{3},1,3\right) & 1 & (1,3,5) \\ \left(\frac{1}{9},\frac{1}{7},\frac{1}{5}\right) & \left(\frac{1}{9},\frac{1}{7},\frac{1}{5}\right) & \left(\frac{1}{7},\frac{1}{5},\frac{1}{3}\right) & \left(\frac{1}{5},\frac{1}{3},1\right) & \left(\frac{1}{5},\frac{1}{3},1\right) & 1 \end{pmatrix} \end{array}. \tag{11}$$

The fuzzy weights of the criteria computed using the formulas proposed by Buckley, Pan and Yuan and by us are given in Table 3. As we can see from the table, the fuzzy weights computed using the formulas (8) and (10) have the smallest vagueness. This was reached by the fact that we applied the extention principle correctly and took into account the condition of the reciprocity of the matrices used in the formulas.

	Fuzzy weights of the criteria using the formulas		
	(2), (3), (4),	(5), (6), (7)	(8), (6), (10)
$\widetilde{w}_1$	$(0.1704, 0.4252, 0.9309)$	$(0.2144, 0.4252, 0.6171)$	$(0.2578, 0.4252, 0.5489)$
$\widetilde{w}_2$	$(0.1303, 0.2948, 0.7118)$	$(0.1545, 0.2948, 0.5126)$	$(0.1837, 0.2948, 0.4516)$
$\widetilde{w}_3$	$(0.0328, 0.0875, 0.2580)$	$(0.0353, 0.0875, 0.2193)$	$(0.0425, 0.0875, 0.1795)$
$\widetilde{w}_4$	$(0.0273, 0.0804, 0.2439)$	$(0.0294, 0.0804, 0.2074)$	$(0.0353, 0.0804, 0.1748)$
$\widetilde{w}_5$	$(0.0273, 0.0804, 0.2439)$	$(0.0294, 0.0804, 0.2074)$	$(0.0353, 0.0804, 0.1748)$
$\widetilde{w}_6$	$(0.0153, 0.0317, 0.0908)$	$(0.0156, 0.0317, 0.0863)$	$(0.0197, 0.0317, 0.0673)$

Table 3 Fuzzy weights of the criteria.

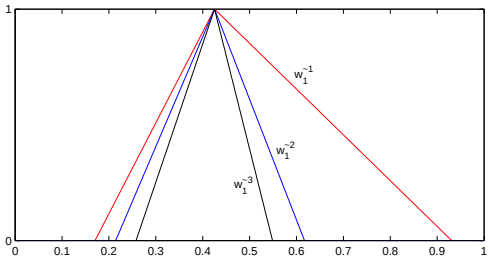


Figure 2 The fuzzy weight of the criterion $\mathcal{K}_1$.

For illustration, fuzzy weights of the criterion $\mathcal{K}_1$ computed using the different formulas are given in Figure 2. The fuzzy weight of the criterion obtained from the formulas (2), (3), (4), resp. (5), (7), resp. (8), (10) is represented by the fuzzy number $\widetilde{w}_1^1$, resp. $\widetilde{w}_1^2$, resp. $\widetilde{w}_1^3$.

4 Conclusion

A proper fuzzification of the Saaty's scale and the fuzzification of the geometric mean method for computing fuzzy weights from a multiplicative fuzzy pairwise comparison matrix were introduced. Incorrect definitions of the fuzzy number $\tilde{1}$ were rectified and a clear explanation of the other elements of the fuzzified Saaty's scale was given. The Saaty's fuzzy pairwise comparison matrix constructed from the fuzzified Saaty's scale was defined then.

Afterwards the fuzzification of the geometric mean method proposed by Buckley [4] and an improvement of the formulas proposed by Pan and Yuan [9] were described. The lacks of these formulas were mentioned and new more adequate formulas (8) and (10) for computing fuzzy weights from a multiplicative fuzzy pairwise comparison matrix were introduced. These formulas are based on the extention principle and take into account the condition of the reciprocity of pairwise comparison matrices. The resulting fuzzy weights computed from the formulas (8), (9) and (10) are then less vague than the fuzzy weights computed from the formulas (2), (3), (4), resp. (5), (6) and (7).

5 Acknowledgement

This research was supported by the grant PrF 2013 013 Mathematical models of the Internal Grant Agency of Palacky University in Olomouc.

References

[1] Alesheikh A. A., Alimohammadi A., Vahidnia M. H. : Hospital site selection using fuzzy AHP and its derivatives, *Journal of Environmental Management* **90** (2009), 3048-3056

[2] Bai H.,Kwong C. K.: A fuzzy AHP approach to the determination of importance weights of customer requirements in quality function deployment, *Journal of Intelligent Manufacturing* **13** (2002), 367-377

[3] Barzilai J., Cook W. D., Golany B.: Consistent weights for judgements matrices of the relative importance of alternatives, *Operations Research Letters* **6** (1987), 131-134

[4] Buckley J.: Fuzzy hierarchical analysis, *Fuzzy Sets and Systems* **17** (1985), 233-247

[5] Chen Y.-H., Wang T.-Ch.: Applying fuzzy linguistic preference relations to the improvement of consistency of fuzzy AHP, *Information Sciences* **178** (2008), 3755-3765

[6] Chou Y.-C., Sun C.-C., Yen H.-Y.: Evaluating the criteria for human resoure for science and technoogy (HRST) based on an integrated fuzzy AHP and fuzzy DEMATEL approach, *Applied Soft Computing* **12** (2012), 64-71

[7] Fan K., Wang J., Wang W.: Integration of fuzzy AHP and FPP with TOPSIS methodology for aeroengine health assessment, *Expert Systems with Applications* **37** (2010), 8516-8526

[8] Krejčí J.: Fuzzy rozšíření Saatyho AHP. [Fuzzy extension of Saaty's AHP] (in Czech), unpublished masters thesis, Palacky University in Olomouc, Olomouc, 2012.

[9] Pan Y., Yuan B.: Bayesian Inference of Fuzzy Probabilities, *Int. J. General Systems* **26** (1997), 73-90

[10] Saaty T. L.: The Analytic Hierarchy Process, New York, McGraw Hill, 1980

[11] Sadiq R., Tesfamariam S. : Risk-based environmental decision-making fuzzy analytic hierarchy process (F-AHP), *Stoch Environ Res Risk Assess* **21** (2006), 35-50

[12] Van Laarhoven P.J.M., Pedrycz W.: A fuzzy extension of Saaty's priority theory, *Fuzzy Sets and Systems* **11** (1983), 253-264

VaR backtesting results under different volatility models

Aleš Kresta[1]

Abstract. Correct risk estimation is nowadays very important and no less challenging part of financial management. It is particularly crucial for financial institutions and their regulatory authorities. The simplest case is the risk estimation of investment into stock market index. Different models can be applied for risk estimation in terms of Value at Risk. These models differ mainly in the stochastic process, which price (respectively returns) time series are assumed to follow. For all the models the accurate prediction of returns' volatility is fundamental. Although it was empirically shown that the returns of financial assets are heteroscedastic, for simple models the homoscedasticity is usually assumed. More precise models take the heteroscedasticy into account. The paper examines the effect of different volatility models application on accuracy of VaR estimation for investment into stock market index. The accuracy is back-tested and the backtesting results are compared. The comparison of the results is made by means of Kupiec's unconditional test and Christoffersen conditional test, i.e. the number of exceptions and their independency in time are tested.

Keywords: Backtesting, Value at Risk, risk management.

JEL Classification: G17
AMS Classification: 91B84

1 Introduction

In order to estimate the risk of financial investments, it is necessary to model the future evolution of returns. This is clearly difficult and no less interesting part of financial management. For calculation of risk (in terms of Value at Risk) we can recognize generally three groups of methods: variance-covariance method, (filtered) historical simulation and Monte Carlo simulation. While it is crucial for all the models to estimate the future variance correctly, we can distinguish also the methods which assume the volatility to be constant over time and methods modeling variance over time. The second group of methods should give more accurate results. Both type models were tested for example by Alexander and Sheedy [1].

Majority of the recent papers published on risk estimation topic is also focused on the dependence modeling (i.e. the authors assume the portfolio composed of more than one asset). We can mention for instance papers published by Huang et al. [8], Ignatieva and Platen [11] or Kresta and Tichý [13]. In these papers the accuracy of models was measured by the backtesting procedure (i.e. the quantity of cases in which the observed loss exceeded the estimated Value at Risk was observed). Statistical tests due to Kupiec [14] and Christoffersen [10] were utilized mostly in these papers, but there were not given attention to the length of period utilized for parameters estimation.

It is obvious that, if the modeling of volatility is inaccurate for a single asset (one-dimensional problem), increasing the dimensionality of the problem (incorporating the dependence to it) would not improve the results (accuracy of the joint model). Hence, we focus solely on the volatility modeling and neglect the dependence issues. Thus, we assume simple one dimensional time series (i.e. the investment into only one asset), which allow us to work with the longer time series with still reasonable requirements on computational time. In this paper we examine the impact of the length of period utilized for parameters estimation on the volatility modeling accuracy and thus backtesting results.

The goal of the paper is to back-test the different volatility models, concretely GARCH model and GJR model with Gaussian and Student innovations, on the chosen time series. Utilized time series is more than 50 years long time series of American stock index S&P 500.

The paper is organized as follows. Applied volatility models are defined in the next section. Then, Value at Risk and method of its backtesting are described. In the last section, utilized dataset is described and backtesting results are presented.

[1] VŠB – Technical University of Ostrava/Faculty of Economics, Department of Finance, Sokolská třída 33, Ostrava, ales.kresta@vsb.cz.

2 Volatility models

Volatility models have become important tool in time series analysis, particularly in financial applications. Engle [6] observed that, although the future value of many financial time series is unpredictable, there is a clustering in volatility. He proposed autoregressive conditional heteroskedasticity (ARCH) process, which has been later expanded to generalized autoregressive conditional heteroskedasticity (GARCH) model by Bollerslev [4]. Further extension assumed in this paper is asymmetric GJR model proposed by Glosten, Jagannathan and Runkle [7]. There are also other volatility models such as IGARCH, FIGARCH, GARCH-M, EGARCH, etc. For their description see e.g. [2].

For all models the conditional mean will also be assumed. Thus the models of time series $\{x_t\}_{t=1}^{N}$ will be of general form as follows,

$$x_t = \mu_0 + \sum_{i=1}^{R} \mu_i \cdot x_{t-i} + \sigma_t \cdot \tilde{\varepsilon}_t , \tag{1}$$

$$\tilde{\varepsilon}_t \sim N(0,1) \quad or \quad \tilde{\varepsilon}_t \sim t_\nu(0,1), \tag{2}$$

where μ_0 is unconditional mean of the series, μ_i are autocorrelation coefficients for lag 1 up to R, σ_t is modeled standard deviation (volatility) and $\tilde{\varepsilon}_t$ is a random number from chosen probability distribution. In the paper two probability distributions are assumed: Gaussian distribution (normal distribution, henceforth n) and Student distribution (henceforth t).

2.1 GARCH model

The GARCH model [4] was proposed as the extension of ARCH model in order to avoid problematic parameters estimation, when there are many of them. The model takes the following form,

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{P} \alpha_i \cdot \sigma_{t-i}^2 + \sum_{j=1}^{Q} \beta_j \cdot \varepsilon_{t-j}^2 , \tag{3}$$

where α_0, α_i and β_j are parameters needed to be estimated. The positive variance is assured if $\alpha_0 > 0$, $\alpha_i \geq 0 \; \forall i$ and $\beta_j \geq 0 \; \forall j$. The model is stationary if $\sum_{i=1}^{P} \alpha_i + \sum_{j=1}^{Q} \beta_j < 1$.

2.2 GJR model

It was shown (firstly by Black [3]) that there is usually different impact of the positive and negative shocks on the volatility. GJR model, proposed by by Glosten, Jagannathan and Runkle, takes this into account. It is similar to the GARCH model (3), but if the previous innovation was negative (dummy variable i_{t-j}) the impact on volatility is bigger (by the parameter γ_j). The model takes the following form,

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{P} \alpha_i \cdot \sigma_{t-i}^2 + \sum_{j=1}^{Q} \beta_j \cdot \varepsilon_{t-j}^2 + \sum_{j=1}^{Q} \gamma_j \cdot i_{t-j} \cdot \varepsilon_{t-j}^2 , \tag{4}$$

where i_{t-j} is a dummy variable which equals to one when innovation ε_{t-j} is negative and null otherwise. Variables α_0, α_i, β_j and γ_j are parameters needed to be estimated. The positive variance is assured if $\alpha_0 > 0$, $\alpha_i \geq 0 \; \forall i$, $\beta_j \geq 0 \; \forall j$ and $\beta_j + \gamma_j \geq 0 \; \forall j$ and model is stationary if $\sum_{i=1}^{P} \alpha_i + \sum_{j=1}^{Q} \beta_j + \frac{1}{2} \sum_{j=1}^{Q} \gamma_i < 1$.

3 VaR and backtesting procedure

Value at Risk (VaR) is nowadays commonly accepted measure of the risk. If we assume a random variable X – the profit from asset / portfolio with the (un)known distribution function F_X , VaR at a given probability level α is the maximum loss which will occur in $1-\alpha$ cases (confidence level),

$$VaR_\alpha(X) = \sup\{x \in R : F_X(-x) \geq \alpha\}. \tag{5}$$

VaR is usually estimated for one day ahead period and then (if needed) recalculated for longer periods. Mostly utilized values of α are 15%, 5%, 1% and 0.5%. For further explanation of VaR concept and methods (or models) utilized for its estimation see e.g. [5, 12].

There are three basic approaches to VaR estimation – (i) analytical formula utilizing parametrical probability distribution function, (ii) stochastic (Monte Carlo) simulation that estimates the quantile of a given distribution numerically, and (iii) (filtered) historical simulation that relates VaR estimation to the quantile obtained from historical observations. For all the models the accurate prediction of returns' volatility is fundamental. In this paper we assume that the financial returns can be modeled by the AR-GARCH and AR-GJR processes, which were described in the previous section. When parameters of these processes are estimated, VaR can be calculated as follows,

$$VaR_{\alpha,t+1} := VaR_{\alpha}(x_{t+1}) = \hat{\mu}_0 + \sum_{i=1}^{R} \hat{\mu}_i \cdot x_{t-i+1} + \hat{\sigma}_{t+1} \cdot q_{\alpha}, \tag{6}$$

where $\hat{\mu}_0$ and $\hat{\mu}_i$ are estimated coefficients of conditional mean, $\hat{\sigma}_{t+1}$ is estimated standard deviation by one of the models defined in previous section and q_α is an appropriate quantile of the innovations distribution (assumed are Gaussian and Student distributions).

By means of backtesting procedure the model is verified. This procedure is based on the comparison of the risk estimated at time t for time $t+1$ with the true loss observed at time $t+1$. Within the backtesting procedure on a given time series the following two situations can arise – the loss is higher or lower than its estimation,

$$I_t = \begin{cases} 1 & if\ r_t < -VaR_t \\ 0 & if\ r_t \geq -VaR_t \end{cases}. \tag{7}$$

While the former case is denoted by 1 as an exception, the latter one is denoted by zero. If the model is accurate, than roughly $(1-\alpha)\cdot n$ exceptions (where n is the length of the data set utilized for backtesting) should be experienced. Bigger quantity of exceptions means, that the model underestimates the risk and vice versa. For further details see [9, 15].

Mostly applied statistical tests are due to Kupiec [14] and Christoffersen [10]. Kupiec's test (henceforth K-test) is derived from a relative amount of exceptions, i.e. whether their quantity is from the statistical point of view different from the assumption. The null hypothesis is that the observer probability of exception occurring is equal to the assumed. A given likelihood ratio on the basis of χ^2 probability distribution with one degree of freedom is formulated as follows:

$$LR = -2\ln\left[\frac{\pi_{ex}^{n_1}\left(1-\pi_{ex}\right)^{n_0}}{\pi_{obs}^{n_1}\left(1-\pi_{obs}\right)^{n_0}}\right], \tag{8}$$

where π_{ex} is expected probability of exception occurring, π_{obs} is observed probability of exception occurring, n_0 is the number of zeros and n_1 is the number of ones (exceptions). The Kupiec's test takes into account only the quantity of exceptions.

By contrast, in order to assess whether the exceptions are distributed equally in time, i.e. without any dependence (autocorrelation), we should define the time lag first. Therefore, we replace the original sequence by a new one, where 01, 00, 11 or 10 are recorded. The null hypothesis is that the probability of exception occurring is independent on the information whether the exception has occurred also previous day. Then we have the likelihood ratio as follows (C-test):

$$LR = -2\ln\left[\frac{\pi_{obs}^{n_1}\left(1-\pi_{obs}\right)^{n_0}}{\pi_{01}^{n_{01}}\left(1-\pi_{01}\right)^{n_{00}}\pi_{11}^{n_{11}}\left(1-\pi_{11}\right)^{n_{10}}}\right], \tag{9}$$

where $\pi_{ij} = \Pr\left(I_t = j \middle| I_{t-1} = i\right)$ and $\pi_{obs} = \dfrac{n_{01}+n_{11}}{n_{00}+n_{01}+n_{10}+n_{11}}$. This test statistic has χ^2 probability distribution with one degree of freedom.

4 Results

Particular models described in the previous section will be applied on data series of log-returns calculated from adjusted closing prices of S&P 500 index over the period January 3, 1950 to December 31, 2012.[2] The length of utilized series of log-returns is 15,850 observations. The distribution of log-returns in time and evolution of the adjusted closing price is depicted in Figure 1.

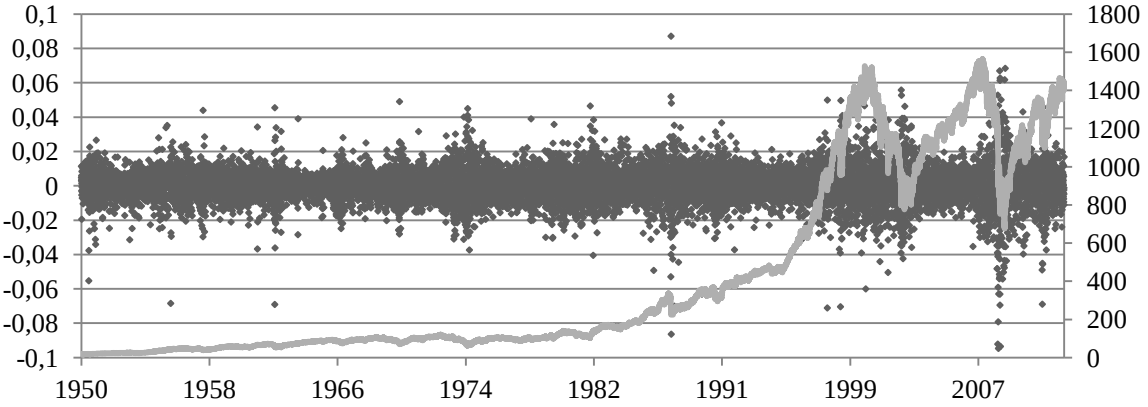


Figure 1 Distribution of log-returns (black, left axis) and evolution of the adjusted closing prices (gray, right axis) of S&P 500 index

It can be observed that although the most of the returns are in interval <-2%, 2%>, the log-returns around ±4% are not exceptional. There is even a quantity of losses around 7%. This corresponds with the high value of the kurtosis (30.683) of the log-returns. It is also interesting that there are more profits than losses (the median is 0.046%), but the average loss is bigger than the average profit (the mean is 0.028%, skewness is -1.301). Moreover, from time to time, there can be observed periods with significantly higher volatility than in the rest of the series. The periods with increased volatility is mostly characterized by the decline of the index value (e.g. 1962, 1987-1988, 1998, 200-2003, 2007-2009). From this observation we can make an assumption that GJR model should better fit the data and thus provide better VaR estimation.

Parameters of all four models were firstly estimated on the whole dataset. For all four assumed models, the first two autoregressive coefficients and only first order of coefficients for volatility equation were found statistically significant (thus we further assume only the 2-1-1 models). According to both the values of log-likelihood function (LLF) as well as Bayesian information criteria (BIC)[3] the models with the Student distribution are superior to those with Gaussian distribution and GJR models are superior to GARCH models (also GARCH-t is superior to GJR-n). While the GJR model is the generalization of GARCH model, the unrestricted GJR model can be tested against restricted GARCH model by means of likelihood ratio test. Both GJR-n and GJR-t are statistically accepted compared to GARCH-n and GARCH-t respectively (p-values of both tests are close to 0). This suggests that GJR model should perform better.

The models (each with $R=2$, $P=1$, $Q=1$) were back-tested with different periods utilized for parameters estimation and the ratio of observed quantity of exceptions to the expected quantity were recorded. Different probability levels are assumed: 0.5% and 15% (required by legal regulations for insurance companies); 1% (required by legal regulations for banks); 5% (original methodology proposed by JP Morgan); 20% and 25% for better clarification. The results are depicted on Figure 2. The ratio should be ideally equal to one or at least close to one (as we want the quantity of observed exceptions to be approximately equal to the expected quantity). As can be seen, for GARCH model (both with Gaussian and Student innovations) the period utilized for parameters estimation is not important and model is applicable only for α bigger or equal to 5% (for $\alpha = 1\%$ the ratio is oscillating around 1.25 and for $\alpha = 0.5\%$ it is oscillating around 1.9, i.e. the observed quantities of exceptions are 1.25 and 1.9 times higher). Model GJR-n provides even worse results than GARCH models: for $\alpha = 1\%$ the ratio converges to two; for $\alpha = 0.5\%$ the ratio converges to 2.5; for $\alpha = 5\%$ the ratio is clearly bigger than 1. The best results are obtained by application of GJR-t model: for $\alpha = 1\%$ and $\alpha = 0.5\%$ the ratios converge to

[2] The dataset was obtained from finance.yahoo.com webserver.

[3] Bayesian information criteria (BIC) is computed as $\log(n) \cdot k - 2 \cdot LLF$, where n is the number of observations, k is the quantity of parameters to be estimated and LLF is the value of log-likelihood function.

1; however for α bigger or equal to 5% the ratios are slightly lower than 1 (it is in interval 0.9-1.0); the curves are smoother than for other models, i.e. the results are not sensitive to small changes in the length of estimation period. From the graph we can see that it is better to utilize longer period for parameters estimation (in our case we further assume 250 days).

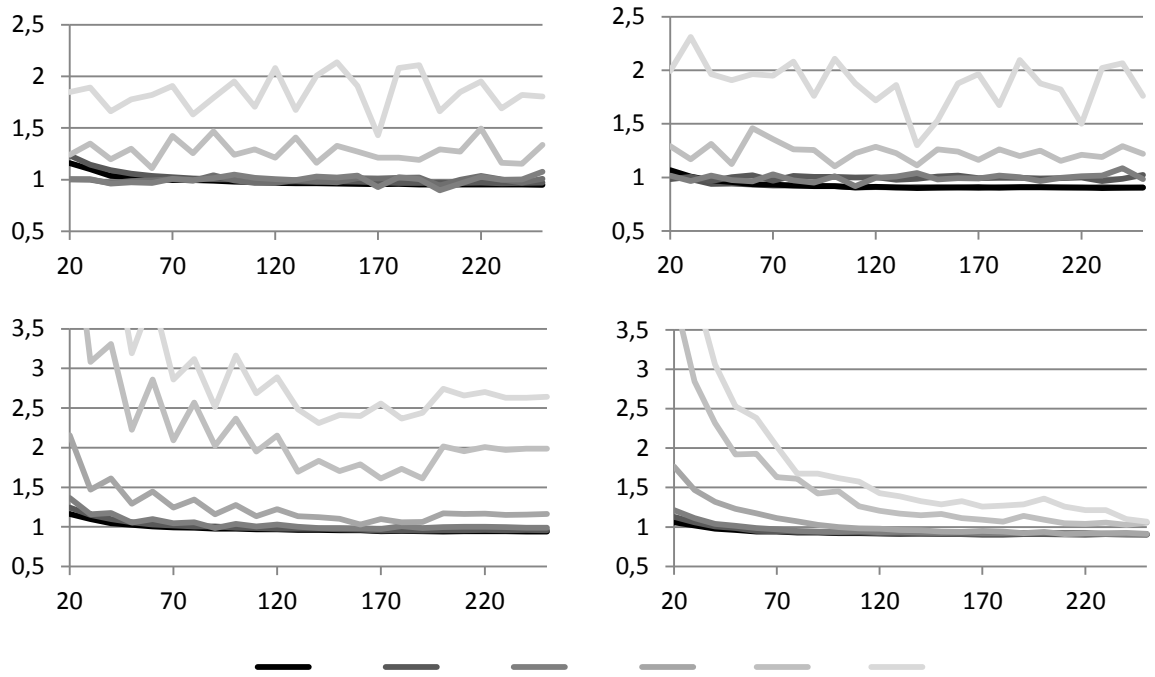


Figure 2 The ratios of observed and expected quantities of exceptions for different periods utilized for parameters estimation (from 20 days till 250 days with the step of 10 days) and different models (from top GARCH-n, GARCH-t, GJR-n, GJR-t)

From the Figure 1 we can see that for $\alpha = 15\%$ and $\alpha = 5\%$ the best model is GARCH-n model. There are not big changes in results for different periods utilized for parameter estimation. The backtesting results of the model for 30 days period are depicted in Table 1. There are shown assumed and observed number of exceptions as well as the p-values of statistical tests defined in previous section. The p-value is a measure of probability of exceptions (to be more specific: the quantity of exceptions in K-test; their distribution over time for C-test) occurring by the chance, assuming the null hypothesis of the test to be true. In statistics, the values 5% or 1% are usually assumed as the critical levels. As can be seen for $\alpha = 15\%$ and $\alpha = 5\%$ the model perform well – the quantity of exceptions is really close to the expected quantity (p-values are even higher than 90%). Also the bunching of exceptions can be rejected (C test). On the other hand, the model performs poorly (the quantity of exceptions is too high) for $\alpha = 1\%$ and $\alpha = 0.5\%$.

	15%	5%	1%	0.5%
Assumed quantity of exceptions	2077.35	692.45	138.49	69.25
Quantity of observed exceptions	2076.00	695.00	187.00	131.00
P-value of K-test	97.44%	92.08%	0.01%	0.00%
P-value of C-test	43.41%	98.30%	76.85%	52.60%

Table 1 Backtesting results for AR(2)-GARCH(1,1)-n model estimated from 30 days

In Table 2 the backtesting results of AR-GJR-t model are shown (period utilized for parameters estimation is 250 days). This model is acceptable for $\alpha = 1\%$ and $\alpha = 0.5\%$ – the quantities of exceptions are close to the assumed and bunching of exceptions can be rejected. On the other hand, the risk for $\alpha = 15\%$ and $\alpha = 5\%$ is overestimated. For both models we can reject the dependence of exceptions for all α.

	15%	5%	1%	0.5%
Assumed quantity of exceptions	2077.35	692.45	138.49	69.25
Quantity of observed exceptions	1873.00	633.00	146.00	74.00
P-value of K-test	0.00%	1.87%	52.49%	57.10%
P-value of C-test	81.61%	83.65%	29.16%	6.74%

Table 2 Backtesting results for AR(2)-GJR(1,1)-t model estimated from 250 days

5 Conclusion

Unexpectedly high decreases in the prices of financial assets and swift changes in volatility are challenging task for any risk model. In this article we utilized the parametric model for Value at Risk estimation based on Gaussian and Student distributions with conditional mean and variance. For modeling the conditional variance we assumed GARCH and GJR models and we examined the influence of the periods utilized for parameters estimation on backtesting results. It was found out that for Value at Risk estimation on probability levels $\alpha = 15\%$ and $\alpha = 5\%$ for S&P 500 index the best model is GARCH model with Gaussian innovations. For Value at Risk estimation on probability levels $\alpha = 1\%$ or $\alpha = 0.5\%$ and for investment into S&P 500 index the most accurate model was GJR model with Student innovations. For this model, in this particular case longer periods for parameters estimation should be applied.

Acknowledgements

The research was supported by the European Social Fund under the Opportunity for young researchers project (CZ.1.07/2.3.00/30.0016) as well as by GAČR (Czech Science Foundation – Grantová Agentura České Republiky) under the project no. 13-18300P.

References

[1] Alexander, C., and Sheedy, E.: Developing a stress testing framework based on market risk models. *Journal of Banking and Finance* **32** (2008), 2220-2236.
[2] Arlt, J. and Arltová, M.: *Ekonomické časové řady*. Grada, Praha, 2007.
[3] Black, F.: The pricing of Commodity Contracts. *Journal of Financial Economics* **3** (1976), 167-179.
[4] Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31** (1986), 307-327.
[5] Dowd, K.: *Beyond Value-at-Risk: The New Science of Risk Management*. John Wiley & Sons Inc., London, 1998.
[6] Engle, R. F.: Auto-regressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50** (1982), 987-1007.
[7] Glosten, L. R., Jagannathan, R., and Runkle, D. E.: On the relation between the expected value and the volatility on the nominal excess returns on stocks. *Journal of Finance* **48** (1993), 1779-1801.
[8] Huang, J. J., Lee, K. J., Liang, H., and Lin, W.F.: Estimating value at risk of portfolio by conditional copula-GARCH method. *Insurance: Mathematics and Economics* **45** (2009), 315–324.
[9] Hull, J.: *Risk Management and Financial Institutions*. Prentice Hall, Upper Saddle River, 2006.
[10] Christoffersen, P.F.: Evaluating Interval Forecasts. *International Economic Review* **39** (1998), 841–862.
[11] Ignatieva, K., and Platen, E.: Modelling Co-movements and Tail Dependency in the International Stock Market via Copulae. *Asia-Pacific Financial Markets* **17** (2010), 261–302.
[12] Jorion, P.: *Value at Risk: The New Benchmark for Managing Financial Risk*. 3rd ed. McGraw-Hill, New York, 2006.
[13] Kresta, A., and Tichý, T.: International Equity Portfolio Risk Modeling: The case of NIG model and ordinary copula functions. *Finance a úvěr – Czech Journal of Economics and Finance* **61** (2012), 141–151.
[14] Kupiec, P.: Techniques for verifying the accuracy of risk measurement models. *Journal of Derivative* **3** (1995), 73-84.
[15] Resti, A., and Sironi, A.: *Risk management and shareholders´ Value in banking: from risk measurement models to capital allocation policies*. John Wiley & Sons Inc., New York, 2007.

Mixed-correlated ARFIMA processes for power-law cross-correlations

Ladislav Krištoufek [1]

Abstract. We introduce a general framework of the Mixed-correlated ARFIMA (MC-ARFIMA) processes which allows for various specifications of univariate and bivariate long-term memory. Apart from a standard case when $H_{xy} = \frac{1}{2}(H_x + H_y)$, MC-ARFIMA also allows for processes with $H_{xy} < \frac{1}{2}(H_x + H_y)$ but also for long-range correlated processes which are either short-range cross-correlated or simply correlated. The major contribution of MC-ARFIMA lays in the fact that the processes have well-defined asymptotic properties for H_x, H_y and H_{xy}, which are derived in the paper, so that the processes can be used in simulation studies comparing various estimators of the bivariate Hurst exponent H_{xy}. Moreover, the framework allows for modeling of processes which are found to have $H_{xy} < \frac{1}{2}(H_x + H_y)$.

Keywords: power-law cross-correlations, long-term memory, econophysics

1 Introduction

Studying long-range correlations has become a stable part of financial econometrics and econophysics in recent years. Long-term memory has been studied for stock indices, bonds, exchange rates, commodities, interest rates and others [2, 8, 6]. Most recently, a focus has been put on analysis of long-range (power-law) cross-correlations as an addition to more traditional research of power-law auto-correlations. Similarly to the power-law autocorrelation case, we assume that a pair of time series is long-range cross-correlated if their cross-correlation function $\rho_{xy}(k)$ follows an asymptotic power law so that $\rho_{xy}(k) \propto k^{2H_{xy}-2}$ as $k \to +\infty$ where H_{xy} is the bivariate Hurst exponent. For $H_{xy} > 0.5$, we have cross-persistent processes which possess non-zero cross-correlations even for very long lags. Contrary to the univariate case, we can have both positively and negatively cross-persistent process with $H_{xy} > 0.5$ due to different properties of the auto- and cross-correlation functions.

As the financial time series provide sufficient number of observations and they have been shown to possess long-range auto-correlations (volatility, traded volume, signs of changes and absolute returns to name the most frequently studied ones), these also create an appropriate setting for the power-law cross-correlations analysis [17, 23, 9, 13]. For estimation of the bivariate Hurst exponent H_{xy}, several estimators have been proposed. Podobnik & Stanley [19] propose the detrended cross-correlation analysis (DCCA) as a bivariate generalization of the detrended fluctuation analysis (DFA) [16]. Zhou [25] then generalizes the method to the multifractal setting with the multifractal detrended cross-correlations analysis (MF-DXA). Kristoufek [11] generalizes the height-height correlation analysis [4, 5] and the generalized Hurst exponent approach [7] into the multifractal height cross-correlation analysis (MF-HXA). The detrending moving average (DMA) [1] is generalized by He & Chen [10] forming the detrended moving-average cross-correlation analysis (DMCA). The most recent estimator from the time domain is proposed by Wang *et al.* [24] – multifractal cross-correlation analysis based in statistical moments (MFSMXA). However, relatively little attention has been given to processes which can be characterized as cross-persistent with a specific bivariate Hurst exponent.

Several processes that possess such long-term correlations have been proposed in the literature. The most frequently discussed and applied ones are the multivariate generalizations of the well-established fractionally integrated ARMA processes (usually labeled as FARIMA and ARFIMA) – VARFIMA or MVARFIMA processes [14] – and fractional Gaussian noise processes or fractional Brownian motions,

[1]Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Opletalova 26, CZ-110 00, Prague, Czech Republic; Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, CZ-182 08, Prague, Czech Republic, e-mail: kristouf@utia.cas.cz

which are their integrated version (these are labeled as fGn and fBm in the literature, respectively). The construction of the multivariate ARFIMA process implies that the bivariate Hurst exponent is the average of the separate Hurst exponents [14]. The same property holds for the fractional Brownian motion [3]. The long-range cross-correlations thus simply arise from the specification of these processes.

Lobato [12] and then in some detail Sela & Hurvich [21] discuss two types of fractionally integrated models – VARFI and FIVAR. VARFI is a vector autoregressive model with fractionally integrated innovations (or error terms), and FIVAR consists of fractionally integrated processes with innovations that come from a VAR model. Nielsen [15] discusses the case of the fractional cointegration in the bivariate long-term memory setting and shows that the coherence of the processes is equal to unity which implies that the bivariate Hurst exponent is the same as the separate Hurst exponents and so again is equal to their average.

All of the above mentioned processes possibly yield cross-persistence but only with the bivariate Hurst exponent equal to the average of the Hurst exponents of the separate processes. There are only two studies which propose models with H_{xy} different from $\frac{1}{2}(H_x + H_y)$. Sela & Hurvich [22] propose an anti-cointegration model, which is in fact a linear combination of ARFIMA processes with a subset of innovations (but not all pairs) being identical across the two processes. The model allows to control the separate Hurst exponents as well as the bivariate Hurst exponent as long as it is lower or equal to the average of the separate parameters. Podobnik *et al.* [18] introduce two-component ARFIMA processes, which are based on ARFIMA-like mixing of two processes. Unfortunately, the authors neither provide any clue how to control the bivariate parameter H_{xy} nor is it evident whether the processes are even stationary.

In this paper, we introduce a new kind of a bivariate process which we call the mixed-correlated ARFIMA process. The process allows to control for the separate and bivariate Hurst exponents as long as the bivariate one is not higher than the average of the separate ones, and additionally allows for short-range dependence as well.

2 Mixed-correlated ARFIMA framework

We start with a general framework of a bivariate series where each of the series consists of a linear combination of two ARFIMA(0,d,0) processes so that

$$\begin{aligned} x_t &= \alpha \sum_{n=0}^{+\infty} a_n(d_1)\varepsilon_{1,t-n} + \beta \sum_{n=0}^{+\infty} a_n(d_2)\varepsilon_{2,t-n} \\ y_t &= \gamma \sum_{n=0}^{+\infty} a_n(d_3)\varepsilon_{3,t-n} + \delta \sum_{n=0}^{+\infty} a_n(d_4)\varepsilon_{4,t-n}. \end{aligned} \tag{1}$$

Innovations are characterized by

$$\begin{aligned} \langle \varepsilon_{i,t} \rangle &= 0 \text{ for } i = 1,2,3,4 \\ \langle \varepsilon_{i,t}^2 \rangle &= \sigma_{\varepsilon_i}^2 \text{ for } i = 1,2,3,4 \\ \langle \varepsilon_{i,t}\varepsilon_{j,t-n} \rangle &= 0 \text{ for } n \neq 0 \text{ and } i,j = 1,2,3,4 \\ \langle \varepsilon_{i,t}\varepsilon_{j,t} \rangle &= \sigma_{ij} \text{ for } i,j = 1,2,3,4 \text{ and } i \neq j. \end{aligned} \tag{2}$$

In words, we have two processes and each one is a linear combination of two long-range correlated processes with possibly correlated innovations. Note that the separate long-term memory parameters d_1, d_2, d_3, d_4 can vary or be the same. We call the set of processes $\{x_t\}$ and $\{y_t\}$ as the mixed-correlated ARFIMA processes (MC-ARFIMA). As MC-ARFIMA is a new kind of process not discussed in the literature, even though these can be seen as a generalization of the anti-cointegration model of Sela & Hurvich [22], we shortly discuss its stationarity. For the wide-sense stationarity, it suffices to state that both $\{x_t\}$ and $\{y_t\}$ are linear combinations of two ARFIMA(0,d,0) processes with correlated innovations which are wide-sense stationary so that MC-ARFIMA processes are stationary as long as $0 \leq d_1, d_2, d_3, d_4 < 0.5$ [20]. Evidently, we have $\langle x_t \rangle = \langle y_t \rangle = 0$ and both processes have finite variance, i.e. $\langle x_t^2 \rangle \equiv \sigma_x^2 < +\infty$ and $\langle y_t^2 \rangle \equiv \sigma_y^2 < +\infty$ since the separate ARFIMA(0,d,0) processes have zero means and finite variances.

As ARFIMA(0,d,0) processes are long-range correlated, their linear combination is also long-range correlated. The higher d will dominate in the linear combination so that process $\{x_t\}$ is integrated of order $\max(d_1, d_2)$ and $\{y_t\}$ of order $\max(d_3, d_4)$. The separate processes are thus wide-sense stationary.

To show that $\{x_t\}$ and $\{y_t\}$ are also jointly wide-sense stationary, we need to show that $\rho_{xy}(k)$ does not depend on t. It can be easily shown that the cross-correlation function is dependent only on the parameters d_1, d_2, d_3, d_4, α, β, γ, δ and σ_{ij} (with $i, j = 1, 2, 3, 4$) and the processes $\{x_t\}$ and $\{y_t\}$ are thus also jointly wide-sense stationary. Based on the cross-correlation structure, the cross-power spectrum can be written as

$$\begin{aligned} f_{xy}(\lambda) = \frac{\alpha\gamma\sigma_{13}}{2\pi}\sum_{k=0}^{+\infty}\sum_{l=0}^{+\infty} a_k(d_1)a_l(d_3)\exp(i(k-l)\lambda) + \frac{\alpha\delta\sigma_{14}}{2\pi}\sum_{k=0}^{+\infty}\sum_{l=0}^{+\infty} a_k(d_1)a_l(d_4)\exp(i(k-l)\lambda) + \\ \frac{\beta\gamma\sigma_{23}}{2\pi}\sum_{k=0}^{+\infty}\sum_{l=0}^{+\infty} a_k(d_2)a_l(d_3)\exp(i(k-l)\lambda) + \frac{\beta\delta\sigma_{24}}{2\pi}\sum_{k=0}^{+\infty}\sum_{l=0}^{+\infty} a_k(d_2)a_l(d_4)\exp(i(k-l)\lambda) = \\ \frac{1}{2\pi}\Big[\alpha\gamma\sigma_{13}\left(1-\exp(i\lambda)\right)^{-d_1}\left(1-\exp(-i\lambda)\right)^{-d_3} + \alpha\delta\sigma_{14}\left(1-\exp(i\lambda)\right)^{-d_1}\left(1-\exp(-i\lambda)\right)^{-d_4} + \\ \beta\gamma\sigma_{23}\left(1-\exp(i\lambda)\right)^{-d_2}\left(1-\exp(-i\lambda)\right)^{-d_3} + \beta\delta\sigma_{24}\left(1-\exp(i\lambda)\right)^{-d_2}\left(1-\exp(-i\lambda)\right)^{-d_4}\Big]. \end{aligned} \quad (3)$$

Using the inverse Fourier transform and the Dirac delta function, we get

$$\begin{aligned} \rho_{xy}(n) = \frac{\alpha\gamma\sigma_{13}}{\sigma_x\sigma_y}\sum_{k=0}^{+\infty}\sum_{l=0}^{+\infty} a_k(d_1)a_l(d_3)\delta(n+k-l) + \frac{\alpha\delta\sigma_{14}}{\sigma_x\sigma_y}\sum_{k=0}^{+\infty}\sum_{l=0}^{+\infty} a_k(d_1)a_l(d_4)\delta(n+k-l) + \\ \frac{\beta\gamma\sigma_{23}}{\sigma_x\sigma_y}\sum_{k=0}^{+\infty}\sum_{l=0}^{+\infty} a_k(d_2)a_l(d_3)\delta(n+k-l) + \frac{\beta\delta\sigma_{24}}{\sigma_x\sigma_y}\sum_{k=0}^{+\infty}\sum_{l=0}^{+\infty} a_k(d_2)a_l(d_4)\delta(n+k-l) = \\ \frac{\alpha\gamma\sigma_{13}}{\sigma_x\sigma_y}\underbrace{\sum_{k=0}^{+\infty} a_k(d_1)a_{n+k}(d_3)}_{\approx\int_0^{+\infty} k^{d_1-1}(n+k)^{d_3-1}dk\propto n^{d_1+d_3-1}} + \frac{\alpha\delta\sigma_{14}}{\sigma_x\sigma_y}\underbrace{\sum_{k=0}^{+\infty} a_k(d_1)a_{n+k}(d_4)}_{\approx\int_0^{+\infty} k^{d_1-1}(n+k)^{d_4-1}dk\propto n^{d_1+d_4-1}} + \\ \frac{\beta\gamma\sigma_{23}}{\sigma_x\sigma_y}\underbrace{\sum_{k=0}^{+\infty} a_k(d_2)a_{n+k}(d_3)}_{\approx\int_0^{+\infty} k^{d_2-1}(n+k)^{d_3-1}dk\propto n^{d_2+d_3-1}} + \frac{\beta\delta\sigma_{24}}{\sigma_x\sigma_y}\underbrace{\sum_{k=0}^{+\infty} a_k(d_2)a_{n+k}(d_4)}_{\approx\int_0^{+\infty} k^{d_2-1}(n+k)^{d_4-1}dk\propto n^{d_2+d_4-1}}. \end{aligned} \quad (4)$$

The results are obtained by using the Stirling's approximation and by approximating the infinite sum by the definite integrals. As we are interested in the asymptotic case $n \to +\infty$, the scaling of $\rho_{xy}(n)$ will be dominated by the highest exponent. This leads us to several interesting settings.

Firstly, let's have $\alpha, \beta, \gamma, \delta \neq 0$ and $\sigma_{ij} \neq 0$ for all $i, j = 1, 2, 3, 4$. Labeling $H_i = d_i + 0.5$ for $i = 1, 2, 3, 4$, we have

$$\begin{aligned} H_x &= \max(H_1, H_2) \\ H_y &= \max(H_3, H_4) \end{aligned} \quad (5)$$

$$H_{xy} = \frac{\max(H_1+H_2, H_1+H_4, H_2+H_3, H_2+H_4)}{2} = \frac{\max(H_1, H_2)+\max(H_3, H_4)}{2} = \frac{H_x+H_y}{2}. \quad (6)$$

Therefore, if the innovations are correlated without restriction, we arrive at $H_{xy} = \frac{1}{2}(H_x + H_y)$.

Secondly, let's again have $\alpha, \beta, \gamma, \delta \neq 0$ and without loss on generality, let's have $\max(H_1, H_2) = H_1$ and $\max(H_3, H_4) = H_4$ so that $\{x_t\}$ is integrated of order d_1 (with $H_x = 0.5 + d_1$) and $\{y_t\}$ of order d_4 (with $H_y = 0.5 + d_4$). Moreover, assume that $\sigma_{23} = \sigma_{32} \neq 0$ and all the other covariances are equal to zero. From Eq. 4, this implies

$$H_{xy} = \frac{H_2+H_3}{2} \leq \frac{H_x+H_y}{2} = \frac{H_1+H_4}{2} = \frac{\max(H_1, H_2)+\max(H_3, H_4)}{2}. \quad (7)$$

The equality holds only if $H_1 = H_2$ and $H_3 = H_4$. For the other cases, it implies that the bivariate Hurst exponent H_{xy} is not equal to the average of the univariate Hurst exponents H_x and H_y while still

showing long-range cross-correlations, i.e. without $H_{xy} = 0.5$. Processes $\{x_t\}$ and $\{y_t\}$ are thus long-range cross-correlated but possess the power-law coherency in a similar manner as the anti-cointegration model of Sela & Hurvich [22].

Apart from the framework introduced above, we can slightly adjust the setting in the following way:

$$x_t = \alpha \sum_{n=0}^{+\infty} a_n(d_1)\varepsilon_{1,t-n} + \beta\varepsilon_{2,t}$$
$$y_t = \gamma\varepsilon_{3,t} + \delta \sum_{n=0}^{+\infty} a_n(d_4)\varepsilon_{4,t-n}. \tag{8}$$

We thus again have $\{x_t\}$ with the long-term memory parameter d_1 (with $H_x = 0.5 + d_1$) and $\{y_t\}$ with d_4 (with $H_y = 0.5 + d_4$) but $H_{xy} = 0.5$ since it can be easily shown that $\rho_{xy}(0) = \frac{\sigma_{23}}{\sigma_x\sigma_y}$ and $\rho_{xy}(k) = 0$ for $k \neq 0$. Therefore, we have two long-range dependent processes which are correlated but not cross-correlated.

In a similar manner, let's have the same assumptions about correlations and parameters $\alpha, \beta, \gamma, \delta$ as in the previous two cases (only innovations ε_2 and ε_3 are correlated) but let's adjust the model specification to

$$x_t = \alpha \sum_{n=0}^{+\infty} a_n(d_1)\varepsilon_{1,t-n} + \beta \sum_{n=0}^{+\infty} \theta_2^n\varepsilon_{2,t-n}$$
$$y_t = \gamma \sum_{n=0}^{+\infty} \theta_3^n\varepsilon_{3,t-n} + \delta \sum_{n=0}^{+\infty} a_n(d_4)\varepsilon_{4,t-n}. \tag{9}$$

Processes $\{x_t\}$ and $\{y_t\}$ are thus linear combinations of ARFIMA(0,d,0) and AR(1) processes. In this case, we again have $\{x_t\}$ with memory d_1 (and $H_x = 0.5 + d_1$) and $\{y_t\}$ with memory d_4 (and $H_y = 0.5 + d_4$). And as the only non-zero correlation between innovations is $\sigma_{23} = \sigma_{32}$, we have $H_{xy} = 0.5$ as the cross-correlations quickly vanish to zero. We thus have two long-range correlated processes $\{x_t\}$ and $\{y_t\}$, which are only short-range cross-correlated.

The MC-ARFIMA framework thus provides quite a wide range of possible model specifications yielding long-range cross-correlated processes which can be either long-range cross-correlated (with or without power law coherency), short-range cross-correlated, pairwise correlated or uncorrelated. Generally, the framework allows for even more possible specifications. In the next section, we present several simulated processes[1] based on specifications developed in this section.

3 Illustrative examples

As Model 1, we use the specification of Eq. 1 with $\alpha = \delta = 0.2$, $\beta = \gamma = 1$, $d_1 = d_4 = 0.4$, $d_2 = d_3 = 0.3$, $\sigma_i^2 = 1$ for $i = 1, 2, 3, 4$ and $\sigma_{23} = 0.9$ with $T = 10000$. We thus obtain two long-range correlated processes with $H_x = H_y = 0.9$ which are also long-range cross-correlated with $H_{xy} = 0.8$. The values of parameters $\alpha, \beta, \gamma, \delta$ are selected to highlight the cross-persistence. The cross-correlation function between simulated processes is shown in Fig. 1. The cross-correlations evidently follow a very slow decay.

As Model 2, we use the specification given in Eq. 9 with $\alpha = \beta = \gamma = \delta = 1$, $d_1 = d_4 = 0.4$, $\theta_2 = \theta_3 = 0.8$, $\sigma_i^2 = 1$ for $i = 1, 2, 3, 4$ and $\sigma_{23} = 0.9$ with $T = 10000$. Model 2 thus represents two processes which are long-range correlated but only short-range cross-correlated. A rapid (exponential) decay of cross-correlation function is presented in Fig. 1. Note that even for a rather strong short-term memory parameter $\theta = 0.8$, the cross-correlations vanish quickly and after lag 15, the cross-correlations are close to zero. The exponential decay to insignificant cross-correlation values is more evident from the cross-correlation function.

As Model 3, we use the specification given in Eq. 8 with $\alpha = \beta = \gamma = \delta = 1$, $d_1 = d_4 = 0.4$, $\sigma_i^2 = 1$ for $i = 1, 2, 3, 4$ and $\sigma_{23} = 0.9$ with $T = 10000$. Model 3 thus represents two processes which are long-range correlated but only correlated (not cross-correlated). Form of the cross-correlation function is presented

[1] R-project codes for MC-ARFIMA are available at http://staff.utia.cas.cz/kristoufek/Ladislav_Kristoufek/Codes.html.

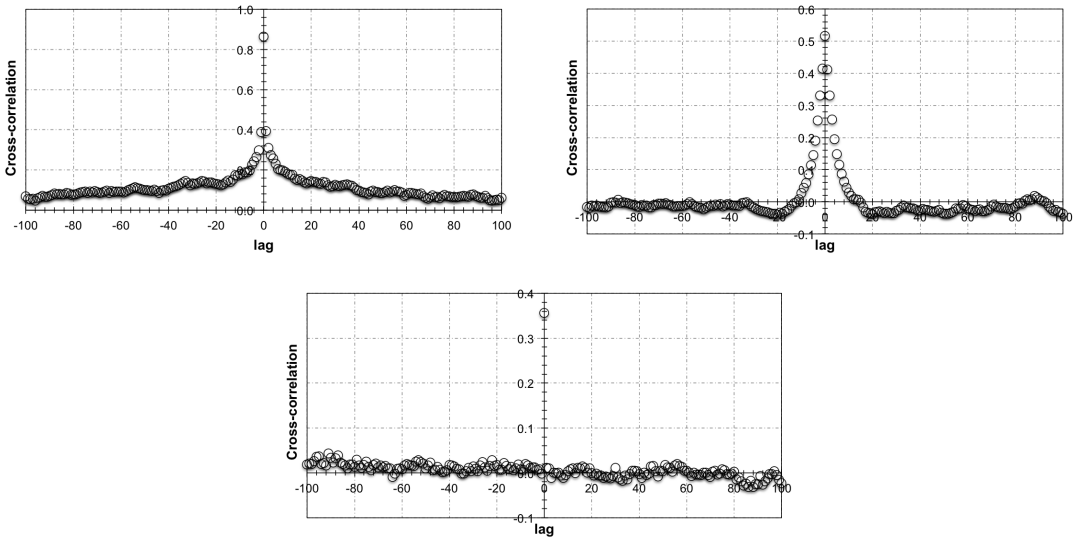


Figure 1: **Cross-correlation functions of Model 1, Model 2 and Model 3.** Cross-correlation function shows a very slow decay for both negative and positive lags for Model 1 (top left), exponential decay for Model 2 (top right) and correlation only at lag 0 for Model 3 (bottom).

in Fig. 1. We observe that indeed the cross-correlation function shows non-zero values only for the zeroth lag which is evident also from the scatter plots for specific lags.

4 Conclusions

Analysis of slowly decaying cross-correlations has recently become a widely and frequently discussed topic in the econophysics literature. However, only little attention has been put on the actual processes that can produce the bivariate Hurst exponent which is different from the average of the separate (univariate) Hurst exponents of the analyzed processes. In this paper, we have introduced a rather general framework of the Mixed-correlated ARFIMA (MC-ARFIMA) processes which allows for various specifications. Apart from a standard case when $H_{xy} = \frac{1}{2}(H_x + H_y)$, it also allows for processes with $H_{xy} < \frac{1}{2}(H_x + H_y)$ but also for long-range correlated processes which are either short-range cross-correlated or simply correlated. The major contribution of MC-ARFIMA lays in the fact that the processes have well-defined asymptotic properties for H_x, H_y and H_{xy} so that the processes can be used in simulation studies comparing various estimators of the bivariate Hurst exponent H_{xy}. Moreover, the framework allows for modeling of processes which are found to have $H_{xy} < \frac{1}{2}(H_x + H_y)$.

Acknowledgements

The support from the Grant Agency of Charles University (GAUK) under project 1110213 and Grant Agency of the Czech Republic (GACR) under project 402/09/0965 are gratefully acknowledged.

References

[1] E. Alessio, A. Carbone, G. Castelli, and V. Frappietro. Second-order moving average and scaling of stochastic time series. *European Physical Journal B*, 27:197–200, 2002.

[2] J. Alvarez-Ramirez and R. Escarela-Perez. Time-dependent correlations in electricity markets. *Energy Economics*, 32(2):269–277, 2010.

[3] P.-O. Amblard and J.-F. Coeurjolly. Identification of the Multivariate Fractional Brownian Motion. *IEEE Transactions on Signal Processing*, 59(11):5152–5168, 2011.

[4] A.-L. Barabasi, P. Szepfalusy, and T. Vicsek. Multifractal spectra of multi-affine functions. *Physica A*, 178:17–28, 1991.

[5] A. L. Barabasi and T. Vicsek. Multifractality of self-affine fractals. *Physical Review A*, 44:2730–2733, 1991.

[6] J. Barunik and L. Kristoufek. On Hurst exponent estimation under heavy-tailed distributions. *Physica A*, 389(18):3844–3855, 2010.

[7] T. Di Matteo, T. Aste, and M. Dacorogna. Scaling behaviors in differently developed markets. *Physica A*, 324:183–188, 2003.

[8] T. Di Matteo, T. Aste, and M. Dacorogna. Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development. *Journal of Banking & Finance*, 29:827–851, 2005.

[9] L.-Y. He and S.-P. Chen. Multifractal detrended cross-correlation analysis of agricultural futures markets. *Chaos, Solitons and Fractals*, 44:355–361, 2011.

[10] L.-Y. He and S.-P. Chen. A new approach to quantify power-law cross-correlation and its application to commodity markets. *Physica A*, 390:3806–3814, 2011.

[11] L. Kristoufek. Multifractal height cross-correlation analysis: A new method for analyzing long-range cross-correlations. *EPL*, 95:68001, 2011.

[12] I. Lobato. Consistency of the average cross-periodigram in long memory time series. *Journal of Time Series Analysis*, 18:137–155, 1997.

[13] F. Ma, Y. Wei, and D. Huang. Multifractal detrended cross-correlation analysis between the Chinese stock market and surrounding stock markets. *Physica A*, 392:1659–1670, 2013.

[14] F. Nielsen. Local Whittle estimation of multi-variate fractionally integrated processes. *Journal of Time Series Analysis*, 32(3):317335, 2011.

[15] M.-O. Nielsen. Spectral analysis of fractional cointegrated systems. *Economics Letters*, 83:225–231, 2004.

[16] C. Peng, S. Buldyrev, S. Havlin, M. Simons, H. Stanley, and A Goldberger. Mosaic organization of DNA nucleotides. *Physical Review E*, 49(2):1685–1689, 1994.

[17] B. Podobnik, I. Grosse, D. Horvatic, S. Ilic, P. Ch. Ivanov, and H. E. Stanley. Quantifying cross-correlations using local and global detrending approaches. *European Physical Journal B*, 71:243–250, 2009.

[18] B. Podobnik, D. Horvatic, A. Lam Ng, H. Stanley, and P. Ivanov. Modeling long-range cross-correlations in two-component arfima and fiarch processes. *Physica A*, 387:3954–3959, 2008.

[19] B. Podobnik and H.E. Stanley. Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series. *Physical Review Letters*, 100:084102, 2008.

[20] Gennady Samorodnitsky. Long range dependence. *Foundation and Trends® in Stochastic Systems*, 1(3):163–257, 2006.

[21] R. Sela and C. Hurvich. Computationally efficient methods for two multivariate fractionally integrated models. *Journal of Time Series Analysis*, 30(6):631–651, 2009.

[22] R. Sela and C. Hurvich. The average periodogram estimator for a power law in coherency. *Journal of Time Series Analysis*, 33:340–363, 2012.

[23] E. L. Siqueira Jr., T. Stošić, L. Bejan, and B. Stošić. Correlations and cross-correlations in the Brazilian agrarian commodities and stocks. *Phy*, 389:2739–2743, 2010.

[24] J. Wang, P. Shang, and W. Ge. Multifractal cross-correlation analysis based on statistical moments. *Fractals*, 20(3):271–279, 2012.

[25] W.-X. Zhou. Multifractal detrended cross-correlation analysis for two nonstationary signals. *Physical Review E*, 77:066211, 2008.

Measuring capital market efficiency: Long-term memory, fractal dimension and approximate entropy

Ladislav Krištoufek Miloslav Vošvrda[1]

Abstract. We utilize long-term memory, fractal dimension and approximate entropy as input variables for the Efficiency Index [Kristoufek & Vosvrda (2013), Physica A 392]. This way, we are able to comment on stock market efficiency after controlling for different types of inefficiencies. Applying the methodology on 38 stock market indices across the world, we find that the most efficient markets are situated in the Eurozone (the Netherlands, France and Germany) and the least efficient ones in the Latin America (Venezuela and Chile).

Keywords: capital market efficiency, long-range dependence, fractal dimension, approximate entropy

1 Introduction

Efficient markets hypothesis (EMH) is one of the cornerstones of the modern financial economics. Since its introduction in 1960s [10, 24, 11], EMH has been a controversial topic. Nonetheless, the theory still remains a stable part of the classical financial economics. Regardless of its definition via a random walk [10] or a martingale [24], the main idea of EMH is that risk-adjusted returns cannot be systematically predicted and there can be no long-term profits above the market profits assuming the same risk. The EMH definition is also tightly connected with a notion of rational homogenous agents and Gaussian distribution of returns. Both these assumptions have been widely disregarded in the literature [6].

There are several papers ranking various financial markets with respect to their efficiency. Research group around Di Matteo [8, 9, 7] show that the correlations structure of various assets (stocks, exchange rates and interest rates) is connected to the development of the specific countries and stock markets. In the series of papers, Cajueiro & Tabak [3, 4, 2, 5] study the relationship between the long-term memory parameter H and development stages of the countries' economy. Both groups find interesting results connecting persistent (long-term correlated) behavior to the least developed markets but also anti-persistent behavior for the most developed ones. Lim [20] investigates how the ranking of stock markets based on Hurst exponent evolves in time and reports that the behavior can be quite erratic. Zunino *et al.* [28] utilize entropy to rank stock markets to show that the emergent/developing markets are indeed less efficient than the developed ones. Even though the ranking is provided in these studies, the type of memory taken into consideration (either long-term memory or entropy/complexity) is limited and treated separately. In this paper, we utilize the Efficiency Index proposed by Kristoufek & Vosvrda [19] incorporating long-term memory, fractal dimension and entropy to control for various types of correlations and complexity using a single measure. Basing the definition of the market efficiency simply on no correlation structure, we can state the expected values of long-term memory, fractal dimension and entropy to construct an efficiency measure based on a distance from the efficient market state. The procedure is then applied on 38 stock indices from different parts of the world and we show that the most efficient markets are indeed the most developed ones – the Western European markets and the US markets – and the least efficient ones are situated in the Latin America and South-East Asia.

The paper is structured as follows. In Section 2, we provide brief description of used methodology focusing on long-term memory, fractal dimension, entropy and efficiency measure. Section 3 introduces the dataset and describes the results. Section 4 concludes.

[1]Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, CZ-182 08, Prague, Czech Republic; Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Opletalova 26, CZ-110 00, Prague, Czech Republic, e-mail: kristouf@utia.cas.cz, vosvrda@utia.cas.cz

2 Methodology

2.1 Long-term memory

Long-term memory (long-range dependence) is usually characterized in time domain by a power-law decay of autocorrelation function and in frequency domain by a power-law divergence of spectrum close to the origin. More specifically, the autocorrelation function $\rho(k)$ with lag k of a long-range correlated process decays as $\rho(k) \propto k^{2H-2}$ for $k \to +\infty$, and the spectrum $f(\lambda)$ with frequency λ of a long-range correlated process diverges as $f(\lambda) \propto \lambda^{1-2H}$ for $\lambda \to 0+$. The characteristic parameter of the long-term memory Hurst exponent H ranges between $0 \leq H < 1$ for stationary processes. The breaking value of 0.5 indicates no long-term memory so that the autocorrelations decay rapidly (exponentially). For $H > 0.5$, the series is persistent with strong positive correlations characteristic by a trend-like behavior while still remaining stationary. For $H < 0.5$, the series is anti-persistent and it switches the direction of increments more frequently than a random process does. As the inputs to the Efficiency Index, we utilize two estimators from the frequency domain – the local Whittle [23] and GPH estimators [13] – which are more appropriate for rather short financial series with a possible weak short-term memory [25, 26], which can easily bias the time domain estimators [27, 1, 17]. Moreover, the frequency domain estimators have well-defined asymptotic properties and the selected two are consistent and asymptotically normal estimators.

2.2 Fractal dimension

Fractal dimension D is a measure of roughness of the series and can be taken as a measure of local memory of the series [19]. For a univariate series, it holds that $1 < D \leq 2$. For self-similar processes, the fractal dimension is connected to the long-term memory of the series so that $D + H = 2$. This can be attributed to a perfect reflection of a local behavior (fractal dimension) to a global behavior (long-term memory). However, the relation usually does not hold perfectly for the financial series so that both D and H give different insights on the dynamics of the series. In general, $D = 1.5$ holds for a random series with no local trending or no local anti-correlations. For a low fractal dimension $D < 1.5$, the series is locally less rough and thus resembles a local persistence. Reversely, a high fractal dimension $D > 1.5$ is characteristic for rougher series with local anti-persistence. For purposes of the Efficiency Index, we utilize Hall-Wood and Genton estimators [14, 15, 16, 12].

2.3 Approximate entropy

Entropy can be taken as a measure of complexity of the system. The systems with high entropy can be characterized by no information and are thus random and reversely, the series with low entropy can be seen as deterministic [22]. The efficient market can be then seen as the one with maximum entropy and the lower the entropy, the less efficient the market is. For purposes of the Efficiency Index, we need an entropy measure which is bounded. Therefore, we utilize the approximate entropy introduced by Pincus [21] which is bounded between 0 (completely deterministic behavior) and 1 (completely random behavior).

2.4 Capital market efficiency measure

According to Kristoufek & Vosvrda [18, 19], the Efficiency Index (EI) is defined as

$$EI = \sqrt{\sum_{i=1}^{n} \left(\frac{\widehat{M_i} - M_i^*}{R_i} \right)^2},$$

where M_i is the ith measure of efficiency, $\widehat{M_i}$ is an estimate of the ith measure, M_i^* is an expected value of the ith measure for the efficient market and R_i is a range of the ith measure. In words, the efficiency measure is simply defined as a distance from the efficient market specification based on various measures of the market efficiency. In our case, we consider three measures of market efficiency – Hurst exponent H with an expected value of 0.5 for the efficient market ($M_H^* = 0.5$), fractal dimension D with an expected

value of 1.5 ($M_D^* = 1.5$) and the approximate entropy with an expected value of 1 ($M_{AE}^* = 1$). The estimate of Hurst exponent is taken as an average of estimates based on GPH and the local Whittle estimators. The estimate of the fractal dimension is again taken as an average of the estimates based on the Hall-Wood and Genton methods. For the approximate entropy, we utilize the estimate described in the corresponding section. However, as the approximate entropy ranges between 0 (for completely deterministic market) and 1 (for random series), we need to rescale its effect, i.e. we use $R_{AE} = 2$ for the approximate entropy and $R_H = R_D = 1$ for the other two measures so that the maximum deviation from the efficient market value is the same for all measures.

3 Application and discussion

We analyze 38 stock indices from various locations – the complete list is given in Tab. 1 – between January 2000 and August 2011. Various phases of the market behavior – DotCom bubble, bursting of the bubble, stable growth of 2003-2007 and the current financial crisis – are thus covered in the analyzed period. The indices cover stock markets in both North and Latin Americas, Western and Eastern Europe, Asia and Oceania so that markets at various levels of development are included in the study. The logarithmic returns are asymptotically stationary (according to the KPSS test), leptokurtic and returns of majority of the indices are negatively skewed (the results are available upon request).

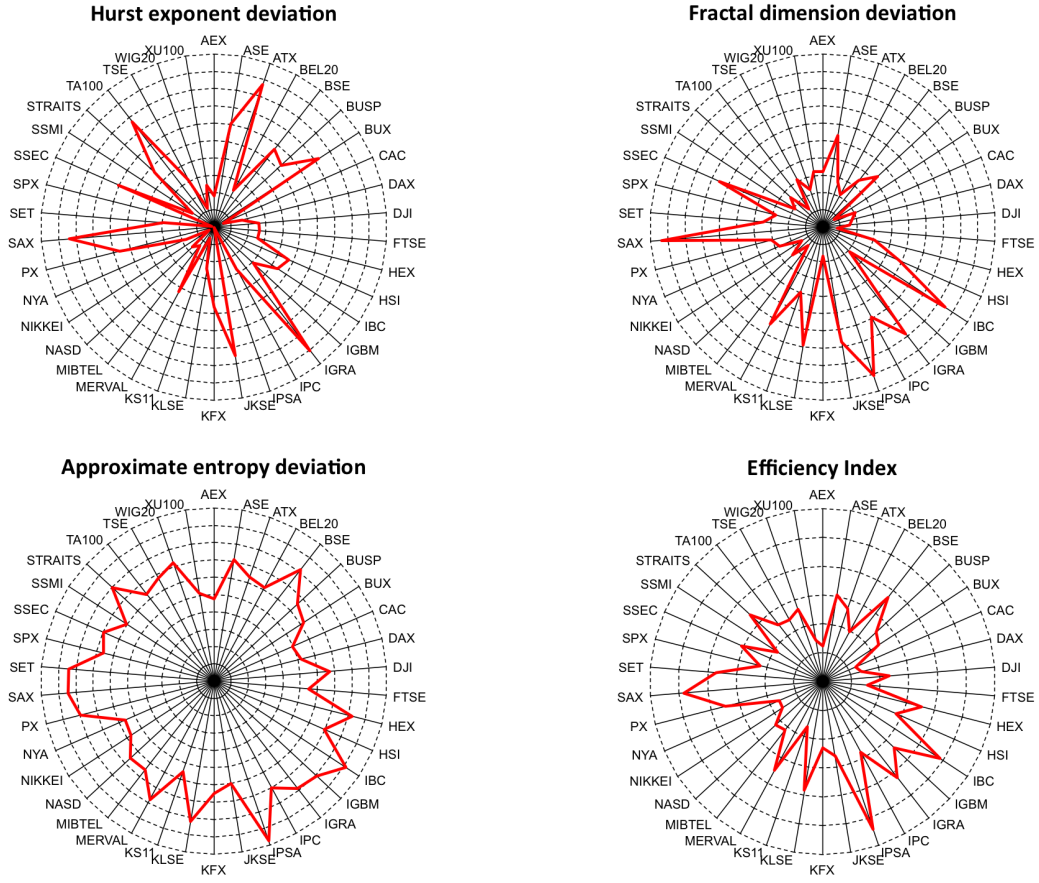


Figure 1: *Hurst exponent, fractal dimension, approximate entropy and efficiency index for analyzed indices.*

Let us now turn to the results. In Fig. 1, all the results are summarized graphically. For the utilized three measures – Hurst exponent, fractal dimension and approximate entropy – we present the absolute deviations from the expected values of the efficient market for comparison. For the Hurst exponent estimates, we observe huge diversity – between practically zero (for IPSA of Chile) and 0.18 (for Peruvian IGRA). Interestingly, for some of the most developed markets, we observe Hurst exponents well below 0.5 (Tab. 1 gives the specific estimates) which is, however, in hand with results of other authors [9, 7]. The results for the fractal dimension again vary strongly across the stock indices. The highest deviation

is observed for the Slovakian SAX (0.19) and the lowest for the FTSE of the UK (0.02). In Tab. 1, we observe that apart from FTSE, all the other stock indices possess the fractal dimension below 1.5 which indicates that the indices are locally persistent, i.e. in some periods, the indices experience significant positively autocorrelated behavior which is well in hand with expectations about the herding behavior during critical events. The approximate entropy estimates are more stable across indices compared to the previous two cases. The highest deviation from the expected value for the efficient market is observed for the Chilean IPSA (0.98) and the lowest for the Dutch AEX (0.48). Evidently, all the analyzed stock indices are highly complex as the approximate entropy is far from the ideal (efficient market) value of 1 and such complexity is not sufficiently covered by the other two applied measures. The inclusion of the approximate entropy into the Efficiency Index thus proves its worth.

Table 1: Ranked stock indices according to the Efficiency Index

Index	Country	Hurst exponent	Fractal dimension	Approximate entropy	Efficiency index
AEX	Netherlands	0.5358	1.4356	0.5246	0.0619
CAC	France	0.5118	1.4592	0.5059	0.0628
DAX	Germany	0.5334	1.4646	0.4807	0.0698
XU100	Turkey	0.5493	1.4350	0.4870	0.0724
FTSE	UK	0.4470	1.5171	0.4500	0.0787
NYA	USA	0.5348	1.4457	0.4418	0.0821
NIKKEI	Japan	0.5063	1.4716	0.4285	0.0825
KS11	South Korea	0.5137	1.4204	0.4473	0.0829
SSMI	Switzerland	0.5297	1.4617	0.3983	0.0929
BEL20	Belgium	0.5481	1.4574	0.3869	0.0981
MIBTEL	Italy	0.5267	1.4728	0.3525	0.1063
NASD	USA	0.5340	1.4526	0.3428	0.1114
SPX	USA	0.5026	1.4437	0.3405	0.1119
KFX	Denmark	0.5927	1.4665	0.3516	0.1148
DJI	USA	0.4477	1.4685	0.3284	0.1165
BUX	Hungary	0.6448	1.4844	0.3811	0.1170
TSE	Canada	0.5626	1.4375	0.3272	0.1210
TA100	Israel	0.6536	1.4739	0.3648	0.1251
BUSP	Brazil	0.6055	1.4142	0.3435	0.1262
JKSE	Indonesia	0.6505	1.3657	0.3986	0.1311
WIG20	Poland	0.5232	1.4545	0.2790	0.1326
ATX	Austria	0.6744	1.4455	0.3669	0.1336
HSI	Hong-Kong	0.5945	1.4033	0.3033	0.1396
IPC	Mexico	0.5550	1.3817	0.2991	0.1398
ASE	Greece	0.6210	1.3926	0.2911	0.1518
SSEC	China	0.6205	1.3698	0.3019	0.1533
IGBM	Spain	0.5615	1.4581	0.1912	0.1691
STRAITS	Singapore	0.5937	1.4500	0.2027	0.1702
PX	Czech Rep	0.6124	1.4386	0.2053	0.1743
MERVAL	Argentina	0.5850	1.3729	0.2225	0.1745
HEX	Finland	0.5524	1.4385	0.1747	0.1768
BSE	India	0.6139	1.4313	0.1842	0.1841
SET	Thailand	0.5591	1.4311	0.1590	0.1851
KLSE	Malaysia	0.5489	1.3620	0.1773	0.1906
IGRA	Peru	0.6806	1.3435	0.2160	0.2108
SAX	Slovakia	0.6673	1.3132	0.1534	0.2421
IBC	Venezuela	0.5881	1.3308	0.0890	0.2439
IPSA	Chile	0.4997	1.3187	0.0239	0.2711

Putting the estimates of the three measures together, we get the Efficiency Index which is also graphically presented in 1. For the ranking of indices according to their efficiency, we present Tab. 1. The most efficient stock market turns out to be the Dutch AEX closely followed by the French CAC and the German DAX. We can observe that the most efficient markets are usually the EU (or rather Eurozone) countries followed by the US markets and other developed markets from the rest of the world – Japanese NIKKEI, Korean KS11, Swiss SSMI. The least efficient part of the ranking is dominated by the Asian

and the Latin American countries. At the very end, we have the Slovakian SAX, Venezuelan IBC and Chilean IPSA. The efficiency of the stock markets is thus strongly geographically determined which is connected to the stage of development of the specific markets.

4 Conclusions

We have utilized long-term memory, fractal dimension and approximate entropy as input variables for the Efficiency Index [19]. This way, we are able to comment on stock market efficiency after controlling for different types of inefficiencies. Applying the methodology on 38 stock market indices across the world, we find that the most efficient markets are situated in the Eurozone (the Netherlands, France and Germany) and the least efficient ones in the Latin America (Venezuela and Chile). The Efficiency Index thus well corresponds to the expectation that the stock market efficiency is connected to the development of the market.

Acknowledgements

The support from the Grant Agency of Charles University (GAUK) under project 1110213, Grant Agency of the Czech Republic (GACR) under project 402/09/0965 and project SVV 265 504 are gratefully acknowledged.

References

[1] J. Barunik and L. Kristoufek. On Hurst exponent estimation under heavy-tailed distributions. *Physica A*, 389(18):3844–3855, 2010.

[2] D. Cajueiro and B. Tabak. Evidence of long range dependence in Asian equity markets: the role of liquidity and market restrictions. *Physica A*, 342:656–664, 2004.

[3] D. Cajueiro and B. Tabak. The Hurst exponent over time: testing the assertion that emerging markets are becoming more efficient. *Physica A*, 336:521–537, 2004.

[4] D. Cajueiro and B. Tabak. Ranking efficiency for emerging markets. *Chaos, Solitons and Fractals*, 22:349–352, 2004.

[5] D. Cajueiro and B. Tabak. Ranking efficiency for emerging equity markets II. *Chaos, Solitons and Fractals*, 23:671–675, 2005.

[6] R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223 – 236, 2001.

[7] T. Di Matteo. Multi-scaling in finance. *Quantitatice Finance*, 7(1):21–36, 2007.

[8] T. Di Matteo, T. Aste, and M. Dacorogna. Scaling behaviors in differently developed markets. *Physica A*, 324:183–188, 2003.

[9] T. Di Matteo, T. Aste, and M. Dacorogna. Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development. *Journal of Banking & Finance*, 29:827–851, 2005.

[10] E. Fama. The behavior of stock market prices. *Journal of Business*, 38:34–105, 1965.

[11] E. Fama. Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance*, 25:383–417, 1970.

[12] M.G. Genton. Highly robust variogram estimation. *Mathematical Geology*, 30:213–221, 1998.

[13] J. Geweke and S. Porter-Hudak. The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4(4):221–238, 1983.

[14] T. Gneiting and M. Schlather. Stochastic Models That Separate Fractal Dimension and the Hurst Effect. *SIAM Review*, 46(2):269–282, 2004.

[15] T. Gneiting, H. Sevcikova, and D.B. Percival. Estimators of fractal dimension: Assessing the roughness of time series and spatial data. Technical report, Department of Statistics, University of Washington, 2010.

[16] P. Hall and A. Wood. On the performane of box-counting estimators of fractal dimension. *Biometrika*, 80(1):246–252, 1993.

[17] L. Kristoufek. How are rescaled range analyses affected by different memory and distributional properties? A Monte Carlo study. *Physica A*, 391:4252–4260, 2012.

[18] L. Kristoufek and M. Vosvrda. Capital markets efficiency: Fractal dimension, Hurst exponent and entropy *(in Czech)*. *Politická ekonomie*, 16(2):208–221, 2012.

[19] L. Kristoufek and M. Vosvrda. Measuring capital market efficiency: Global and local correlations structure. *Physica A*, 392:184–193, 2013.

[20] K.-P. Lim. Ranking market efficiency for stock markets: A nonlinear perspective. *Physica A*, 376:445–454, 2007.

[21] S. Pincus. Approximate entropy as a measure of system complexity. *PNAS*, 88:2297–2301, 1991.

[22] S. Pincus and R.E. Kalman. Irregularity, volatility, risk, and financial market time series. *PNAS*, 101(38):13709–13714, 2004.

[23] P.M. Robinson. Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, 23(5):1630–1661, 1995.

[24] P. Samuelson. Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, 6:41–49, 1965.

[25] M. Taqqu, W. Teverosky, and W. Willinger. Estimators for long-range dependence: an empirical study. *Fractals*, 3(4):785–798, 1995.

[26] M. Taqqu and V. Teverovsky. On Estimating the Intensity of Long-Range Dependence in Finite and Infinite Variance Time Series. In *A Practical Guide To Heavy Tails: Statistical Techniques and Applications*, 1996.

[27] V. Teverovsky, M. Taqqu, and W. Willinger. A critical look at lo's modified r/s statistic. *Journal of Statistical Planning and Inference*, 80(1-2):211–227, 1999.

[28] L. Zunino, M. Zanin, B. Tabak, D. Pérez, and O. Rosso. Complexity-entropy causality plane: A useful approach to quantify the stock market inefficiency. *Physica A*, 389:1891–1901, 2010.

Portfolio competitions and rationality

Aleš Antonín Kuběna[1], Martin Šmíd[2]

Abstract. We study investment competitions in which the players with highest achieved returns are rewarded by fixed prizes. We show that, under realistic assumptions, a game the participants play lacks a pure equilibrium and that the "max-min" solution of the game lies in one of the extremal points of the feasible set, namely in the one having maximal probability that the portfolio return falls into its normal cone. We analyse empirically a portfolio competition held recently by the Czech portal "lidovky.cz"; we find that the majority of people do not behave according to the game-theoretic conclusions. Consequently, searching for factors influencing a choice of particular stocks, we find that that the only significant determinant of the choice is a size of the stock's issuer.

Keywords: portfolio competition, game theory, behavioural finance

JEL classification: C7, D03
AMS classification: 91B99

1 Introduction

With the public availability of the Internet, various investment competitions started to be held, usually with the following rules: each player obtains a virtual sum of money which he has to divide into several (real-life) financial assets. After a pre-determined time, gains of the players are evaluated (according to the real-life prices) and a selected number of the best players are rewarded by monetary prizes. If several participants achieve the same evaluation, the prize(s) divide(s) equally.

Whether the organizers realize it or not, those games are far from being a simulation of a real-life investment; the main reason for this is the fact that the objectives of "players" in real life differ from those in the game. In particular, while the actual return is simultaneously the gain in real life, which forces a risk averse individual to diversify (see [1]), only the best returns bring positive gains in the competition which, as shown in Section 2 of the present paper, makes even a risk-averse participant to take positions which are the most risky ones from the point of view of portfolio selection theory. In particular, the only portfolios getting a positive max-min gain are those lying in extremal points of the feasible set.

Analysing an actual portfolio game held by Czech internet portal "lidovky.cz", however, we found that people do not behave according to such a conclusion. As shown in Section 3, only 16.8% of participants chose portfolios lying in extremal points.

Section 4 tries to give alternative explanations of the player's behaviour. It is shown that, out of several fundamental and technical-analysis indicators, the only significant factor of a stock's selection is the size of the stock's issuer.

Even if all the game theoretic results and the method of our subsequent analysis are rather straightforward, we regard our work as original because, to our best knowledge, there is no other paper analysing this type of competition.

The paper is concluded by Section 5.

[1]Institute of Information Theory and Automation AS CR, Dpt. of Econometrics, e-mail: kubena@utia.cas.cz
[2]Institute of Information Theory and Automation AS CR, Dpt. of Econometrics, e-mail: smid@utia.cas.cz

2 Game Theoretic Approach

Denote $R \in \mathbb{R}^n$ a random vector of asset returns, possibly discounted by a deterministic risk free rate r_0, having an absolutely continuous joint distribution such that

$$\operatorname{supp}(R) = (-1, \infty)^n.$$

We assume that the set of feasible actions of the players is defined as

$$S = \{\pi \in \mathbb{R}^n : \gamma \leq 1'\pi \leq 1, 0 \leq \pi_i \leq \alpha, 1 \leq i \leq n\}$$

where α and γ are some constants. In the definition above, π stands for a vector fractions of the initial sum invested into the individual assets.

We assume the competitors to be risk averse, the i-th one having a strictly increasing utility function u_i. For simplicity, we assume that (the participants act as if) there is only single prize which implies that the motility of the i-th player is

$$v_i = \mathbb{E}(u_i(Z_i))$$

where Z_i is a gain of the player given by

$$Z_i \quad = Z_i(\pi_1, \ldots, \pi_m) = \quad \begin{cases} \frac{1}{k_i} & \text{if } R \in \Gamma_i \\ 0 & \text{otherwise} \end{cases}$$

Here

- $\Gamma_i = \Gamma_i(\pi_1, \ldots, \pi_m) := \{r : \pi_i' r > \pi_j' r, j \notin K_i\}$
- $K_i = \{1 \leq j \leq m : \pi_j' R = \pi_i' R\}$,
- $k_i = |K_i|$
- $\pi_1, \pi_2, \ldots, \pi_m$ are the strategies (portfolios) of individual players.

Remark 1. The vector $(Z_1, \ldots, Z_m)$ is uniquely defined by $\rho = \frac{R}{|R|}$ a.s. where the support of ρ is the unit sphere.

In the present paper, assume all the strategies be the pure ones, i.e. deterministic. Then

Proposition 1. *The set*

$$K_i = \{1 \leq j \leq m : \pi_j' R = \pi_i' R\}$$

is a.s. deterministic.

Proof. Because $\mathbb{P}[w'S = 0] = 0$ for any absolutely continuous random vector S, any deterministic vector $w \neq 0$ and a constant c, we have

$$e_{i,j} := 1\{\pi_j' R = \pi_i' R\} = 1\{(\pi_j - \pi_i)' R = 0\} \overset{a.s.}{=} 1\{\pi_j = \pi_i\}$$

i.e., is deterministic, so it has to be

$$K_i = \{j : e_{i,j} = 1\}.$$

□

Corollary 1. $v_i = u_i(1/k_i)p_i$ where $p_i = \mathbb{P}(R \in \Gamma_i)$.

The following Proposition shows that some portfolios would never win regardless of the distribution of R

Proposition 2. *If π_i is not an extremal point of $C := \operatorname{conv}(\pi_1, \pi_2, \ldots, \pi_{\mathrm{m}})$ then $Z_i \equiv 0$ a.s.*

Proof. Assume WLOG that the first k strategies are extremal points of C. From the basic convex analysis (see [2]) we have that $\pi_i = \sum_{j=1}^{k} \lambda_j \pi_j$ where $\lambda_j \geq 0$. For Z_i to be positive, there should exists at least one possible return value r such that $\lambda_j r'(\pi_i - \pi_j) > 0$ for all $j \leq k$ giving $r'0 > 0$ by summing over all j. □

The following result says that the best max-min strategy is to take the most "advantegeous" corner of S; however, no equilibrium in pure strategies exists whenever there do not exist a group of stocks strongly outperforming the rest.

Theorem 3. *Denote $E = (e_1, \ldots, e_r)$ the set of extremal points of S and put*

$$\sigma_i = \mathbb{P}(\rho \in N_S(e_i))$$

where

$$N_S(e) = \{r : r'(\pi - e) \leq 0 \text{ for all } \pi \in S\}$$

is a normal cone.
(i) If $m \geq n + 2$ then

$$\max_{\pi_i} \min_{\pi_j, j \neq i} v_i = 0$$

whenever $\pi_i \notin E$.
(ii)

$$\max_{\pi_i} \min_{\pi_j, j \neq i} v_i \geq u_i(\frac{1}{m})\sigma_i$$

whenever $\pi_i \in E$.
(iii) Denote $I_j = \left\lfloor \frac{1}{\alpha} \right\rfloor$. If there is a player, say the i-th one, such for each $j \geq 1$ there exist $j_1, j_2, \ldots, j_{I_i+1}$, differing from j fulfilling

$$\mathbb{P}(R_j \geq R_{j_k}) > \frac{u_i(\frac{1}{m})}{u_i(1)}, \qquad 1 \leq k \leq I_j \tag{1}$$

then there exists no symmetric equilibrium in pure strategies.

Before proving the Theorem note that the RHS of (1) goes to zero with the growing number of participants.

Proof. (i) Assume $\pi_1 \notin E$. Then, by basic convex analysis, there exist $\pi_2, \ldots, \pi_{n+2} \in E$ such that π_1 lies in their convex hull so (ii) is implied by Proposition 2.
(ii) If $\pi_1 \in E$, then clearly $N_S(\pi_1) \subseteq \Gamma_1$ for any $\pi_2, \ldots, \pi_m$ so

$$v_i = u_i(1/k_i)p_i \geq u_i(1/m)\sigma_i.$$

(iii) Let $\pi \in S$ be an equilibrium. Let j be one of its non-zero components and let j_k fulfil (1) so that $\pi_{j_k} < \alpha$ (such j_k has to exist because at the weight of at most I_j components may equal to α). Consider portfolio $\tilde{\pi} = (\pi_1, \pi_2, \ldots \pi_j - s, \ldots \pi_{j_k} + s, \ldots, \pi_n)$ where s is small enough for $\tilde{\pi}$ to be feasible. However, if the i-th player holds portfolio $\tilde{\pi}$ and the rest of players hold π then it will be

$$p_i = \mathbb{P}(\tilde{\pi}R > \pi R) = \mathbb{P}(s(R_i - R_{j_k}) > 0) > \frac{u_i(\frac{1}{m})}{u_i(1)}$$

hence the expected utility given $\tilde{\pi}$ (equal to $p_i u_i(1)$) would be greater than that given the equilibrium (being $u_i(1/m)$).

□

Summarizing: if one wants to be sure with a positive expected gain, he has to choose one of the extremal points as his strategy. If, in addition, there are no significant leaders among stocks and/or there is a large number of participants, then no such or another point is a pure equilibrium, i.e., possible common strategy.

3 Empirical Evidence

In order to verify whether actual people behave according to the game theoretic conclusions, we analysed a portfolio competition held by Czech news internet portal "lidovky.cz" this year. The competition started in April and is supposed to end in July. According to the rules, its participants could split a virtual million Czech crowns among 27 stocks listed in Table 3, and a (fictitious) bank account yielding 0.4% p.a. The three participants with the highest value of their virtual portfolios, measured on July 9, are

promised to obtain 30.000, 20.000, and 10.000 Czech crowns, respectively. If there were more participants with the highest value of their portfolios then the prize would be divided equally.[1] The upper limit α of an investment asset is 40% for stocks, 50% for the bank account, respectively. The rules also say that at least 10% could be invested into a single stock if it is invested into it which, however, was violated by 6 portfolios for unknown reasons.[2]

The data we used come from the internet site of the competition `http://portfolio.lidovky.cz` and a subsequent preprocessing by a special software written in `C++` by us and by a free OCR program `gocr`. As the text recognition appeared to be inaccurate, several consistency checks were performed and, subsequently, manual correction were made; nevertheless, it is still possible that there are minor errors left in data caused by an inaccurate OCR recognition, which may be, however, regarded as noise if the data is analysed statistically .

There was as much as 2699 portfolios competing in the game. Even if it is highly probable that some players created multiple identities to increase their chances, we neglect this suspicion as we have no means to identify those cases.

There is 9828 extremal points of a feasible set in total,[3] 345 of which were occupied by portfolios of 453 (16.8%) participants (the most popular being portfolio CETV 40%, NWR 40%, ORCO 20% which was used 8 times). In other words, only 16.8% of players behaved "rationally" in the sense of Theorem 3. Out of remaining (non-extremal) portfolios, 975 was dominated in the sense of Proposition 2, having no chance for the first prize given the configuration portfolios of the other players. We used Iredundancy problem algorithm to determine which portfolios were dominated (see [3], Chp. 19 for details).

As the participants could optionally publish their gender and age, which was actually done by 2163 of them, 1559 of them, respectively, we tested for a correlation of a type of strategy chosen (possible types being an extremal point, a non-extremal not dominated point and a non-extremal dominated point) with these values. However, no significant results have been found here.

These facts lead us to a conclusion that people did not behave according to game theory given that only pure strategies are assumed.

4 Alternative Explanation.

Opposed to a rational approach, a hypothesis of purely random choice of portfolio, i.e., that the portfolios are chosen from the uniform distribution on the feasible set, suggests itself. This hypothesis, however, is falsified by the fact that SCHHV was never chosen because the probability that some stock has zero weight in all the 2699 portfolios is less than 0.02.

There could be many potential factors possibly influencing the choice of the stocks. In the present introductory paper, we restricted ourselves to considering selected data concerning the individual stocks published by the Prague stock exchange on their website, in particular to

- price-earning ratio (P/E),
- market capitalization (the monetary value of the issued shares), measuring the size of the firm,
- long-time trend (the ratio of the price of the stock at the time of the game's start and the average of the highest and the lowest price form the last year) and
- the short-time trend (the ratio of an OLS trend, computed from observations of the price from the five weeks preceding the competition, and the current price of the stock).

Note that, while the first two factors belong to fundamental analysis, the latter are more technical-analysis ones. In order to discover the dependency of a particular stock's choice on those factors, we run the logistic regression with the relative frequency as a dependent variable and the four mentioned factors

[1] It is, however, not said what would happen in case of equality on the second and/or the third place.

[2] We neglect these lower bounds in our theoretical analysis in Section 2 as they bring non-convexity of the feasible set which consequently complicates the treatment.

[3] Note that this number depends only on the number of stocks

Code	Name	p	a	MC	P/E	long	short
AAA	AAA Auto Group N.V.	0.17	3.0	1.565	5.98	0.043	0.002
CETV	CE Media Enterprises Ltd.	0.15	3.2	5.797	0	8.753	−0.037
ČEZ	ČEZ, a.s.	0.50	12.2	304.502	7.44	−0.163	−0.011
EFORU	E4U a.s.	0.04	0.7	0.167	10.56	0.022	−0.001
ENCHE	ENERGOCHEMICA SE	0.06	0.9	3.810	0	0.002	0.000
ENRGA	Energoaqua, a.s.	0.08	1.3	1.185	8.92	0.056	0.000
ERSTE	Erste Group Bank AG	0.42	8.5	223.957	0	0.108	−0.021
FOREG	Fortuna Entertainment Group N.V.	0.37	7.5	5.096	15.21	0.105	0.020
JIP	VET ASSETS a.s.	0.04	0.7	0.010	0	−0.052	0.000
KB	Komerční banka, a.s.	0.43	8.3	144.095	15.53	0.038	0.008
LAZJA	Jáchymov Property Management, a.s.	0.03	0.4	0.477	89.91	−0.025	0.000
NWR	New World Resources Plc	0.22	4.7	17.877	0	−0.318	−0.019
OCELH	OCEL HOLDING SE	0.09	1.5	3.798	0	0.000	0.000
ORCO	Orco Property Group S.A.	0.18	3.7	5.899	0.12	−0.148	−0.014
PEGAS	PEGAS NONWOVENS SA	0.26	5.2	4.661	12.92	0.064	−0.009
PM ČR	Philip Morris ČR a.s.	0.43	9.2	22.302	12.74	0.051	0.001
PRSLU	Pražské služby, a.s.	0.05	0.9	0.795	18.09	−0.066	0.000
PVT	RMS Mezzanine, a.s.	0.03	0.6	1.225	0	0.036	0.032
SCHHV	SPOLEK PRO CHEM.A HUT.VÝR.,a.s	0.00	0.0	775.763	0	0.000	0.000
SMPLY	Severomoravská plynárenská, a.s.	0.12	2.0	13.251	17.05	0.007	0.000
TEL. O2	Telefónica Czech Republic, a.s.	0.35	6.9	93.245	10.67	−0.181	−0.019
TMR	Tatry mountain resort, a.s.	0.16	3.3	7.814	0	0.029	0.004
TOMA	TOMA, a.s.	0.08	1.2	1.006	7.8	0.010	−0.001
UNI	UNIPETROL, a.s.	0.26	4.7	31.190	0	0.003	0.000
VCPLY	Východočeská plynárenská,a.s.	0.09	1.6	6.657	13.94	−0.037	−0.018
VGP	VGP NV	0.02	0.4	6.504	19.35	0.000	0.000
VIG	VIENNA INSURANCE GROUP	0.23	4.1	123.110	12.93	0.104	0.000

Table 1: Menu of stocks: p - frequency of choice, a - average weight (in %), MC - market capitalization, P/E - price to earning ration, *long* - long term trend, *short* - short term trend.

	Coefficient	Std. Error	t-ratio	p-value
const	−2.02345	0.238733	−8.4758	0.0000***
MC	8.09090e–06	2.55727e–06	3.1639	0.0047***
P/E	−0.0145792	0.0103617	−1.4070	0.1740
long	−0.0203306	0.125300	−0.1623	0.8727
short	−12.5787	16.7717	−0.7500	0.4616

Sum squared resid	16.92489	S.E. of regression	0.897746
R^2	0.462091	Adjusted R^2	0.359632
$F(4, 21)$	4.510016	P-value(F)	0.008704
Log-likelihood	−31.31136	Akaike criterion	72.62272
Schwarz criterion	78.91320	Hannan–Quinn	74.43415

Table 2: Result of logistic regression $p = 1/(1 + e^{-(1,MC,P/E,long,short)'\beta})$

as independent ones.[4] The results, shown by table 3, clearly show that only the market capitalization comes out as significant.[5]

5 Conclusion

We analysed a rather general case of a portfolio competition. As the behaviour of players in an actual game of this type appeared to be inconsistent with the from the game-theoretical point of view, we tried to give a simple behavioural explanation: in particular, we found that - out of four factors - players take only the size of the stock's issuer into account when constructing their portfolios.

Acknowledgements

Support from the Czech Science Foundation under grant 402/09/0965 is gratefully acknowledged.

References

[1] Cuthbertson, K., and Nitzsche, D.: *Quantitative financial economics: stocks, bonds and foreign exchange*. Wiley, 2005.

[2] Ekeland, I., and Témam. I: *Convex Analysis and Variational Problems*. Siam, 1976.

[3] Goodman, J.E., and Rourke J.O. eds. *Handbook of Discrete & Computational Geometry*. Chapman and Hall/CRC, 2010.

[4] Even if there is a dependency between choices of the particular stocks by the individuals, we omit it relying on the law of large numbers promising a closeness of the relative frequencies with the probabilities.

[5] We excluded SCHHV from the regression as it is clear from its data that the stocks is not traded.

Empirical copula for missing observations

Adam Kubetta [1]

Abstract. Empirical copula (EC) is a rank-based method for estimating the dependence structure of a random vector, which is convenient when someone wants to estimate non-parametrically the whole distribution of the random vector in two separated stages: first the margins and then the dependence structure fully represented by a copula. The discrete nature of an EC can be successively smoothed using Bernstein polynomials approximation to obtain differentiable estimate of underlying copula or its density. In practice, however, some components of some observations are often missing. In this case the marginal distributions of individual vector components can be estimated without any changes using all available information, whilst the second step is not so straightforward. One can employ only complete observations to capture the dependence but then the mapping of estimated copula on individual marginal quantiles is not reflected appropriately. In the article the author suggests to generalize the classical EC to be applicable also on incomplete observations and further shows that Bernstein approximation based on this generalized EC is only slightly modified and all its important attributes remain unchanged.

Keywords: Bernstein approximation, empirical copula, missing observations

JEL classification: C44
AMS classification: 90C15

1 Introduction

A motivation for this study was the task to estimate the measure called CoVaR [1] for Czech banks. As the classical measure VaR describes the critical value of one specific subject's future loss, the CoVaR measures the same value under some conditions on other subject's losses. When some heavy-tailed losses of certain subjects are considered, the unconditional VaR can significantly change for subjects with losses, which are correlated. Therefore CoVaR can be used to compare the sensitivity of individual subjects to a financial contagion.

Although originally the method was considered to be applied on quite dense public market data, here only a sparse data (monthly reports to the regulator) are considered, because the majority of the Czech banking sector is not publicly traded. That is why the fully empirical estimate can not be used. Moreover, all subjects have history of different lengths so that the periods of their coexistence are even shorter. The reports in times when all subjects submitted the data are called complete observations in this text. For complete datasets the method based on quantile regression, for example, mentioned in [1] seems reasonable. Anyway, traditional approaches usually do not consider partly incomplete datasets and therefore can be applied indeed only on those complete observations.

Due to the mentioned data characteristics the choice of an approach via copulas seems natural as it can divide the task on separable steps that can be easily adjusted. In the first step the distributions of marginal losses for individual subjects is estimated beforehand using all information in the data. In the second step, which is studied in detail in this text, the dependence structure of margins is captured and as explained further, this step can be also adjusted on incomplete data. Finally, the CoVaR can be, more or less, directly calculated from the full distribution estimate obtained from the previous steps.

[1] Charles University in Prague, Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics
Sokolovská 83, 186 75 Praha 8, kubetta@karlin.mff.cuni.cz

2 Notation

In the text $d \in \mathcal{N}$ always denotes the dimension of the underlying random vector $\mathbf{X} = \{\mathrm{X}_1, ..., \mathrm{X}_d\}$ with random components $\mathrm{X}_i \in \mathcal{R}$, $i \in \mathbb{i} = \{1, ..., d\}$. All references to d-dimensional objects are abbreviated to d-objects, for example d-vector $\mathbf{X}$, etc. The observed sample of length n from $\mathbf{X}$, denoted $\mathbb{x} = \{\mathbf{x}_1, ..., \mathbf{x}_n\}$, determines the series of its d-ranks $\mathbb{r} = \{\mathbf{r}_1, ..., \mathbf{r}_n\}$. Values $\mathbf{r}_k \in \mathcal{N}^d$, for $k \in \mathbb{k} = \{1, ..., n\}$, say that the k^{th} observation $\mathbf{x}_k \in \mathcal{R}^d$ is $(\mathbf{r}_k)_i^{\text{th}}$ (according to the increasing magnitude) when comparing i^{th} coordinates of all observations. For simplicity, ties are not considered here or decided randomly. On the other hand, any i^{th} coordinate of any k^{th} observation $\mathbf{x}_k$ might be missing, which is denoted $(\mathbf{x}_k)_i = x_{k,i} = \square$. Technically, whole vectors can be unavailable, $\{\square, ..., \square\}$.

Any arithmetic operation a applied on $\square$ returns $a(\square) = \square$ so that a rank corresponding to incomplete observation is also incomplete. Missing components are matched with ranks with the highest orders and n_i denotes the number of those observations having the i^{th} component available. The vector of n_i is denoted $\mathbf{n} = \{n_1, ..., n_d\}$.

Vectors taking values only from d-dimensional unit interval $\mathcal{I}^d = [0, 1] \times ... \times [0, 1]$ are distinguished from those from $\mathcal{R}^d$ by using letters $\mathbf{u}$ or $\mathbf{w}$ instead of $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$. The binary operators of product and division applied on vectors are considered component-wise, the product is then denoted $*$ to avoid misinterpreting as the scalar (dot) product.

Complete operator $\mathcal{C}$ applied on observations (ranks) returns the set of only those observations (ranks), which are complete, and similarly $\mathcal{C}$ applied on indices $\mathcal{C}(1, ..., n)$ returns only those referring to complete observations (ranks). For example, indices of incomplete observations can be expressed as $\mathbb{k} \backslash \mathcal{C}(\mathbb{k})$.

Vector inequalities $\mathbf{x} \leq \mathbf{y}$ means that $x_i \leq y_i$ for each $i \in \mathbb{i}$ and λ is the Lebesgue measure on $\mathcal{R}^d$.

3 Brief overview of copula

This section sums up the necessary essentials regarding copulas, which can be all found in [3].

Definition 1 (Copula)**.** For any $d \in \mathcal{N}$, d-copula is a function $C : \mathcal{I}^d \to \mathcal{I}$ that satisfies the following conditions:
(C1) $C(u_1, ..., u_d) = 0$ whenever $u_i = 0$ for at least one index $i \in \mathbb{i}$
(C2) $C(u_1, ..., u_d) = u_i$ whenever $u_j = 1$ for each index $j \neq i$
(C3) C is d-increasing, i.e. the volume of an arbitrary d-interval $[\mathbf{x}, \mathbf{y}] = [x_1, y_1] \times ... \times [x_d, y_d] \subset \mathcal{I}^d$ measured by C is non-negative, or symbolically

$$\int_{\mathcal{I}^d} \mathbb{1}(\mathbf{u} \in [\mathbf{x}, \mathbf{y}]) \mathrm{d}C(\mathbf{u}) = \sum_{\mathbf{z} \in \{x_1, y_1\} \times ... \times \{x_d, y_d\}} (-1)^{\mathbb{1}(x_1 = z_1) + ... + \mathbb{1}(x_d = z_d)} C(\mathbf{z}) \geq 0$$

Copulas are particularly important in the probability theory due to the following result by Sklar.

Theorem 1 (Sklar)**.** *For each random d-vector $\mathbf{X}$ with d-CDF F and marginal CDFs $F_1, ..., F_d$ there is a d-copula $C_{\mathbf{X}}$ such that $F(\mathbf{x}) = C_{\mathbf{X}}(F_1(x_1), ..., F_d(x_d))$ and this copula is unique on the domain specified by the Cartesian product of ranges $F_1(\mathcal{R}) \times ... \times F_d(\mathcal{R})$.*

Remark 1. Note that the range of a *continuous* random variable CDF is always the whole unit interval $\mathcal{I}$, thus the d-copula associated with a continuous random d-vector is always unique on the whole $\mathcal{I}^d$. Moreover, copula can then be interpreted as d-CDF of $\mathbf{U} = \{F_1(\mathrm{X}_1), ..., F_d(\mathrm{X}_d)\} \in \mathcal{I}^d$, i.e. the random d-vector $\mathbf{X}$ transformed by its margins. Since the transformation of a continuous random variable specified by CDF (of the variable) is always uniformly distributed on $\mathcal{I}$, the random d-vector $\mathbf{U}$ has indeed uniformly distributed margins.

4 Empirical Copula

The definition of classical EC (see definition 3) can be also found in [3]. However, the author finds useful to clarify its meaning beforehand, so that a reader can get an insight into the way how this definition is further generalized. Other definitions in this section are proposed by the author and relate only to a notation.

CDF of a random d-vector $\mathbf{X}$ is defined as the following function $F_{\mathbf{X}} : \mathcal{R}^d \to \mathcal{I}$.

$$F(\mathbf{x}) = F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}(\mathbf{X} \leq \mathbf{x}) = \mathbb{P}(\{\mathrm{X}_1, ..., \mathrm{X}_d\} \leq \mathbf{x})$$

This function can be estimated based on complete observations $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ of the d-vector $\mathbf{X}$ by the *empirical* d-CDF, which is the following function $\hat{F}_{\{\mathbf{x}_1,...,\mathbf{x}_n\}} : \mathcal{R}^d \to \mathcal{I}$.

$$\hat{F}(\mathbf{x}) = \hat{F}_{\{\mathbf{x}_1,...,\mathbf{x}_n\}}(\mathbf{x}) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}(\mathbf{x}_k \leq \mathbf{x}) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}(\{x_{k,1}, ..., x_{k,d}\} \leq \mathbf{x})$$

The d-copula corresponding to a random d-vector $\mathbf{X}$ having marginal CDFs $F_1, ..., F_d$ can be also written as the function $C_{\mathbf{X}} : \mathcal{I}^d \to \mathcal{I}$.

$$C(\mathbf{u}) = C_{\mathbf{X}}(\mathbf{u}) = \mathbb{P}(\{F_1(\mathrm{X}_1), ..., F_d(\mathrm{X}_d)\} \leq \mathbf{u})$$

Analogically, one can estimate this function based on a complete sample $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ as

$$\hat{C}(\mathbf{u}) = \hat{C}_{\{\mathbf{x}_1,...,\mathbf{x}_n\}}(\mathbf{u}) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}(\{F_1(x_{k,1}), ..., F_d(x_{k,d})\} \leq \mathbf{u})$$

However, this estimate depends on marginal CDFs $F_1, ..., F_d$, which are not observed. Plugging empirical CDFs $\hat{F}_1, ..., \hat{F}_d$ to the previous formula (in the place of $F_1, ..., F_d$ respectively) yields the estimate called empirical copula. After substituting, the expression can be further simplified, because $n\hat{F}_i((\mathbf{x}_k)_i) = n\hat{F}_i(x_{k,i})$ is nothing else than the rank $(\mathbf{r}_k)_i$ for each $k \in \mathbb{k}$ and $i \in \mathbb{i}$.

Definition 2 ($\mathbf{n}$-lattice, elementary interval)**.** For a given $\mathbf{n} \in \mathcal{N}^d$ the $\mathbf{n}$-lattice $\mathcal{L}(\mathbf{n})$ is a collection of coordinates from $\mathcal{I}^d$ created as the following Cartesian product

$$\mathcal{L}(\mathbf{n}) = \left\{\frac{1}{n_1}, ..., \frac{n_1}{n_1}\right\} \times ... \times \left\{\frac{1}{n_d}, ..., \frac{n_d}{n_d}\right\}$$

Each point from $\mathcal{L}(\mathbf{n})$ can be addressed by a d-index $\mathbf{i} \in \{1, ..., n_1\} \times ... \times \{1, ..., n_d\}$ as $\mathbf{i}/\mathbf{n}$ as well as its neighbourhood $\mathbf{i}^{\text{th}}$ elementary interval

$$\Xi_{\mathbf{i}/\mathbf{n}} = \Xi_{\mathbf{i}/\mathbf{n},\mathbf{n}} = \left[\frac{(\mathbf{i}-1)}{\mathbf{n}}, \frac{\mathbf{i}}{\mathbf{n}}\right] = \left[(\mathbf{i}-1) * \left\{\frac{1}{n_1}, ..., \frac{1}{n_d}\right\}, \mathbf{i} * \left\{\frac{1}{n_1}, ..., \frac{1}{n_d}\right\}\right]$$

Definition 3 (Classical empirical copula)**.** Empirical d-copula of a random d-vector $\mathbf{X}$ with marginal CDFs $F_1, ..., F_d$ based on a complete random sample $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ is defined for $\mathbf{u} \in \mathcal{L}(\{n, ..., n\})$ as

$$\mathrm{EC}(\mathbf{u}) = \mathrm{EC}_{\{\mathbf{x}_1,...,\mathbf{x}_n\}}(\mathbf{u}) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}(\mathbf{r}_k \leq n\mathbf{u})$$

The domain of EC is restricted on the lattice $\mathcal{L}(\{n, ..., n\})$ only in order to avoid technical difficulties, because the traditional definition requires copula to have continuous margins. There are some approaches that extends the EC domain to $\mathcal{I}^d$ in such a way that EC complies with the definition of copula. They differ in allocating the probability mass over each $\mathbf{i}^{\text{th}}$ elementary interval $\Xi_{\mathbf{i}/\mathbf{n}}$.

Remark 2 (EC approximation)**.** The uniform allocation over elementary intervals gives the so called *checkerboard approximation*, others are e.g. check-min or shuffle-of-min approximations, see [2, 4]. However, further study is focused mainly on capturing the copula on the lattice $\mathcal{L}(\mathbf{n})$.

Definition 4 ($\mathbf{x}$-manifold)**.** For given $\mathbf{x} \in \mathcal{R}^d$ that may contain missing components the $\mathbf{x}$-manifold $\mathcal{M}(\mathbf{x})$ is defined as

$$\mathcal{M}(\mathbf{x}) = \left\{\mathbf{y} \in \mathcal{R}^d : y_i = x_i \text{ for all such } i \text{ that } x_i \neq \square\right\}$$

For any complete observation $\mathbf{x}$ holds $\mathcal{M}(\mathbf{x}) = \mathbf{x}$. When $\mathbf{x}$ contains one missing component at i^{th} position, $\mathcal{M}(\mathbf{x})$ is a line parallel to i^{th} axis, when $\mathbf{x}$ contains two missing components at positions i and j, $\mathcal{M}(\mathbf{x})$ is a plane parallel to both i^{th} and j^{th} axis, etc.

Definition 5 (**u**-space)**.** For given $\mathbf{n} \in \mathcal{N}^d$ and $\mathbf{u} \in \mathcal{L}(\mathbf{n})$ that may contain missing components the **u**-space $\Xi_{\mathbf{n}}(\mathbf{u}) \subset \mathcal{I}^d$ is defined as

$$\Xi(\mathbf{u}) = \Xi_{\mathbf{n}}(\mathbf{u}) = \bigcup_{\mathbf{w} \in \mathcal{M}(\mathbf{u}) \cap \mathcal{L}(\mathbf{n})} \Xi_{\mathbf{w},\mathbf{n}}$$

For any complete vector $\mathbf{u} \in \mathcal{L}(\mathbf{n})$ the $\Xi_{\mathbf{n}}(\mathbf{u})$ coincides with the elementary interval $\Xi_{\mathbf{u}/\mathbf{n}}$.

Proposal 1 (Extension of empirical copula)**.** For given $\mathbf{n} \in \mathcal{N}^d$ the function $\mathrm{EC} : \mathcal{L}(\mathbf{n}) \to \mathcal{I}$ is called empirical copula if EC complies with the following requirements.
(EC1) EC is a d-CDF on the lattice $\mathcal{L}(\mathbf{n})$
(EC2) For $\mathbf{U} \sim \mathrm{EC}$ holds $\mathbb{P}(\mathbf{U} \in \cup_{\mathbf{r} \in \mathbb{r}} \Xi_{\mathbf{n}}(\mathbf{r}/\mathbf{n})) = 1$
(EC3) For $\mathbf{U} \sim \mathrm{EC}$ holds $\mathbb{P}(\mathbf{U} \in \Xi_{\mathbf{n}}(\mathbf{r}_k/\mathbf{n}) | \mathbf{U} \in \cup_{\mathbf{r} \in \mathcal{C}(\mathbb{r})} \Xi_{\mathbf{n}}(\mathbf{r}/\mathbf{n}) = 1/n$ for all $k \in \mathcal{C}(\mathbb{k})$

The first condition is clear. The second says that the probability mass described by an empirical copula can not be allocated outside the collection of **u**-spaces associated with rescaled ranks of observations. The third condition means that all complete observations (as well as their rescaled ranks) are weighted equally and proportionally to incomplete ones.

Remark 3. Note that in higher dimensions ($d \geq 3$) there are more possibilities how to construct an EC.

The series of pictures in Figure 1 shows the process of creating an empirical copula (here for $d = 3$) by the author's algorithm, which can be described here only roughly due to limited scope of the article.

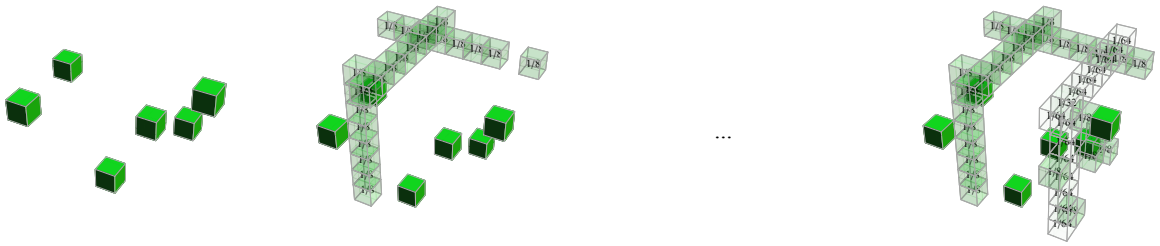


Figure 1 Finding an EC. The final density-plot of an EC for $d = 3$ can be seen on the right.

In the beginning the probability mass corresponding to all complete observations is located to adequate elementary intervals in advance. This is the first and also the last step for complete datasets and can be seen on the left. Then a certain ratio is associated with each incomplete observation based on a mutual spatial relation of observations. After that follows a multi-level process of projecting the mass already located onto elementary intervals that has not been matched with any amount yet. The probability mass once matched to an elementary interval is never changed, which requires to consider the order of projecting carefully. The final result can be seen as the checkerboard empirical copula density after multiplying each amount by $n_1 ... n_d$. Each result can be easily checked, whether it has uniform margins.

Although the purpose of this text is not to study the asymptotics, the following theorem states some trivial sufficient conditions for convergence of EC to C.

Theorem 2 (Empirical copula convergence)**.** *Let $m \in \mathcal{N}$ be fixed number. If a random sample $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ taken from a random vector $\mathbf{X}$ with a copula $C_{\mathbf{X}}$ that have a density with respect to Lebesgue measure λ on $\mathcal{I}^d$ contains for arbitrary $n \in \mathcal{N}$ no more than m incomplete observations, then $\mathrm{EC}_{\{\mathbf{x}_1,...,\mathbf{x}_n\}}$ converges uniformly to $C_{\mathbf{x}}$ on $\mathcal{I}^d$ when $n \to \infty$.*

Proof. The result follows from the facts that $\mathrm{EC} \to C$ for complete observations (uniform convergence of CDF) and that the probability that $\mathbf{X} \in \cup_{k \in \mathbb{k} \setminus \mathcal{C}(\mathbb{k})} \Xi_{\mathbf{n}}(\mathbf{r}_k/\mathbf{n})$ converges to 0, because

$$\lambda \left(\cup_{k \in \mathbb{k} \setminus \mathcal{C}(\mathbb{k})} \Xi_{\mathbf{n}}(\mathbf{r}_k/\mathbf{n}) \right) \to 0$$

as $n \to \infty$ and $C_{\mathbf{X}}$-measure is absolutely continuous with respect to λ. □

5 Bernstein approximation of a copula

An option, how to extend a copula specified only on $\mathcal{L}(\mathbf{n})$ to the whole $\mathcal{I}^d$, is to use some approximation (see Remark 2). The approximation described in this section, studied e.g. in [4] (only for complete datasets), is based on polynomials and thus returns a smooth function on $\mathcal{I}^d$.

For $n \in \mathcal{N}$, $k = 0, ..., n$ and $u \in \mathcal{I}$ a *Bernstein polynomial* $b_{n,k}(u)$ (or k^{th} Bernstein basis function of degree n at u) is defined as the following expression

$$b_{n,k}(u) = \binom{n}{k} u^k (1-u)^{n-k}$$

In order to simplify further notation in this text, let $b_{n,k}(u) = 0$ for other indices then those specified above and let $\mathbf{b}_{\mathbf{n},\mathbf{k}}(\mathbf{u}) = b_{n_1,k_1}(u_1)...b_{n_d,k_d}(u_d)$ for $\mathbf{n} \in \mathcal{N}^d$, $\mathbf{k} \in \{0, ..., n_1\} \times ... \times \{0, ..., n_d\}$ and $\mathbf{u} \in \mathcal{I}^d$ denote the object that might be called *the product Bernstein polynomial*.

Proposal 2 (Bernstein approximation of copula)**.** Bernstein approximation of a copula C can be expressed as

$$B_{\mathbf{n}}(C)(\mathbf{u}) = \sum_{\mathbf{w} \in \mathcal{L}(\mathbf{n})} C(\mathbf{w}) \mathbf{b}_{\mathbf{n},\mathbf{n}*\mathbf{w}}(\mathbf{u})$$

Remark 4. Note that the classical Bernstein approximation of a copula, described e.g. in [4], is a special case of the proposal above, where $\mathbf{n} = \{n, ..., n\}$. Be aware that the Bernstein approximation is not an interpolation, i.e. $\mathbf{u} \in \mathcal{L}(\mathbf{n}) \Rightarrow B_{\mathbf{n}}(C)(\mathbf{u}) = C(\mathbf{u})$ generally does not hold.

Theorem 3 (Bernstein approximation attributes)**.** *For an arbitrary continuous copula C the following statements are true.*
(BA1) $B_{\mathbf{n}}(C) \to C$ *uniformly on* $\mathcal{I}^d$ *when* $\mathbf{n} \to \{\infty, ..., \infty\}$
(BA2) $B_{\mathbf{n}}(C)$ *is a copula for any* $\mathbf{n} \in \mathcal{N}^d$

Proof. (BA1) is a special case of a known result from analysis [5]. (BA2) can be verified analogically to the classical case, see sketch of proof in [4], by showing that $B_{\mathbf{n}}(C)$ complies with (C1), (C2) and (C3).

The Bernstein polynomial $b_{n,k}(0) = 0$ for arbitrary indices $0 < k \leq n$ and also $\mathbf{n} * \mathbf{w} \geq \mathbf{0}$ for any $\mathbf{n} \in \mathcal{L}(\mathbf{n})$, hence $\mathbf{b}_{\mathbf{n},\mathbf{n}*\mathbf{w}}(\mathbf{u}) = 0$ for all $\mathbf{w} \in \mathcal{L}(\mathbf{n})$ whenever $u_k = 0$ for at least one index k, which implies that $B_{\mathbf{n}}(C)$ satisfies (C1). Similarly, $b_{n,k}(1) = 0$ for indices $0 < k < n$ and $b_{n,k}(1) = 1$ for $0 < k = n$ and generally also holds $\Sigma_{k=1,...,n}\ k\ b_{n,k}(u)/n = u$ from which follows that $B_{\mathbf{n}}(C)$ satisfies also (C2), because

$$B_{\mathbf{n}}(C)(\{1, ..., 1, u_i, 1, ..., 1\}) = \sum_{\mathbf{w} \in \mathcal{L}(\mathbf{n}) \cap \mathcal{M}(\{1,...,1,\square,1,...,1\})} C(\{1, ..., 1, w_i, 1, ..., 1\}) b_{n_i,n_i w_i}(u_i) = u_i$$

The d-derivative of $B_{\mathbf{n}}(C)$ can be simplified considering the rule $\partial_u b_{n,k}(u) = n\,(b_{n-1,k-1}(u) - b_{n-1,k}(u))$ for $n \in \mathcal{N}$ and $k = 0, ..., n$ to the following form

$$D(B_{\mathbf{n}}(C))(\mathbf{u}) = n_1...n_d \sum_{\mathbf{w} \in \mathcal{L}(\mathbf{n})} \left(C(\mathbf{w}) - C\left(\mathbf{w} - \tfrac{1}{\mathbf{n}}\right)\right) \mathbf{b}_{\mathbf{n}-1,\mathbf{n}*\mathbf{w}-1}(\mathbf{n})$$

The expression is obviously non-negative and can be seen as the density of $B_{\mathbf{n}}(C)$-measure on $\mathcal{I}^d$ so the condition (C3) on $B_{\mathbf{n}}(C)$ is verified. □

The estimate $B_{\mathbf{n}}(C)$ takes values of C only at the lattice $\mathcal{L}(\mathbf{n})$, and therefore C can be successively replaced by an empirical copula EC when estimating it. Therefore one can define the Bernstein Copula estimate as an object depending only on a sample $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ as

$$\mathrm{BC}_{\mathbf{n}}(\mathbf{x}) = B_{\mathbf{n}}\left(\mathrm{EC}_{\{\mathbf{x}_1,...,\mathbf{x}_n\}}\right)(\mathbf{x}) = \sum_{\mathbf{u} \in \mathcal{L}(\mathbf{n})} \mathrm{EC}_{\{\mathbf{x}_1,...,\mathbf{x}_n\}}(\mathbf{u}) \mathbf{b}_{\mathbf{n},\mathbf{n}*\mathbf{u}}(\mathbf{x})$$

Theorem 4 (Bernstein Copula estimate convergence)**.** *For $n \to \infty$ the Bernstein Copula estimate calculated based on a complete random sample $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ drawn from the underlying random vector $\mathbf{X}$ having a continuous copula $C_{\mathbf{X}}$ converges to this copula $C_{\mathbf{X}}$ uniformly on $\mathcal{I}^d$.*

If the sample is incomplete but meets all the requirements from Theorem 2 on empirical copula convergence, then the statement is still valid. The result follows immediately from composing two uniform convergences in Theorems 2 and 3.

6 Comparison

In figure 2 some different approaches to a 2-vector density estimation are compared on an artificial data. All use the same kernel estimates of margins. The first approach constructs a MLE of the vector density expressed through the AMH parametric copula, where of course only complete data can be used in the maximization. The second approach uses the classical EC constructed again only from complete observations. The last approach employing the proposed extended EC can process all available data.

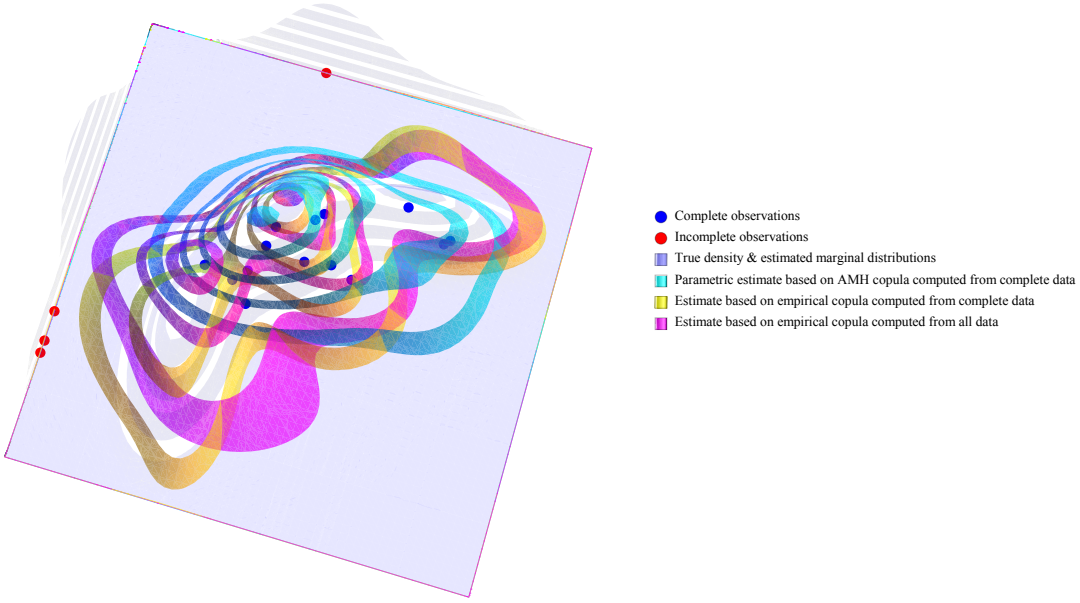


Figure 2 Comparison of random vector estimates.

7 Conclusion

The proposed extension of the empirical copula introduces the opportunity to estimate the dependence structure of a random vector based on observations, which do not have to be necessarily complete. When all observed data are complete, the extended algorithm coincides with the classical one. The empirical approach leaves open the choice of method for estimating the random vector margins, because the empirical copula is a rank based function of a random sample.

Acknowledgements

This work was supported by the grant SVV-2013- 267 315 and the Czech Science Foundation project No. 13-01930S.

References

[1] Adrian, T., and Brunnermeier, M. K.: *CoVaR*. National Bureau of Economic Research (Working Paper 17454), Cambridge, 2011.

[2] Kulpa, T.: On Approximation of Copulas, *International Journal of Mathematics and Mathematical Sciences* **22** (1999), 259–269.

[3] Mai, J. F., and Scherer, M.: *Simulating Copulas*. Imperial College Press, London, 2012.

[4] Mikusiński, P., and Taylor, M. D.: Some Approximations of n-Copulas, *Metrika* **72** (2010), 385–414.

[5] Pallini, A.: Bernstein-type Approximation of Smooth Functions, *Statistica* **65**, No 2 (2005), 169–191.

The Synchronization of Financial and Trade Integration of the EU10 Countries by Using the Method of Classical and Dynamic Correlation

Zuzana Kučerová[1], Jitka Poměnková[2]

Abstract. Financial and trade integration are interconnected processes and should not be assessed separately. Using a data sample of EU10 member countries over the period 1993-2012 we assess the relationship between financial and trade integration in the EU10 member countries. As the basic measurement of the dependence between indicators we use the method classical correlation. Because the sample size is small and with respect to the annual character of the data, we additionally calculate the moving correlation to evaluate behaving dependence in the time. Consequently we apply the method of dynamic correlation which provides a detailed view to the structure of dependency also from the frequency point of view. The result of the dynamic correlation is used for validation of the results obtained from the previous methodological steps. The empirical evidence of the dependence is validated across all three mentioned approaches. We conclude that there is a strong linear association between financial and trade integration in the pre-crisis period (1993-2007) in the EU10 countries. However, this relationship was seriously weakened by the financial crisis (2008-2010). We also found out, that the progress in financial integration was smaller compared to the progress of trade integration in the EU10 countries.

Keywords: financial integration, foreign trade, international investment position, correlation analysis.

JEL Classification: C23, C36, E44, F36, F42
AMS Classification: 62P20, 62M10, 91B84

1 Introduction

The integration of financial markets contributes to the overall integration and economic growth by removing the exchange rate risk and the barriers and frictions in cross-border capital movement. This allows the capital to be allocated more efficiently (Baele *et al*. [3]). Financial integration is an important factor in increasing the efficiency of a financial system and lowering the costs for business as well as for consumers. However, the process of financial integration of the past decade was associated with an unprecedented accumulation of risks and the national and supranational financial regulation and supervisory practices lagged behind the highly integrated, fast expanding and sophisticated financial sector (European Commission [5]).

Foreign trade (i.e. trade integration) is an important factor influencing financial integration. Lane and Milesi-Ferreti [14], [15] mention several important linkages between foreign trade and trade with foreign assets and liabilities. Firstly, the high volume of trade with goods and services evokes the corresponding financial transactions. Foreign direct investments had a great impact on the external balance of the "new" EU member countries from Central and Eastern Europe; large trade deficits originating from the transformation process were compensated by investment inflows (i.e. by increasing financial integration). International trade and international financial flows are thus able to equilibrate the balance of payment. Secondly, a high share of bilateral trade linkages between countries leads to the willingness of economic agents to increase the number of financial transactions with these countries. Investors have a better knowledge of foreign companies from these countries and are thus more prone to buying the shares of these companies (the "familiarity effect"). Thirdly, a high degree of trade openness of a country reflects the liberal approach of macroeconomic policy authorities not only in the area of foreign trade, but also in the area of cross-border capital flows.

The aim of the paper is to assess the relationship between financial and trade integration in the EU10 member countries over the period 1993-2012. We use quantity-based measures of financial integration derived from

[1] Faculty of Economics, VSB-Technical University of Ostrava, Dept. of Economic Policy, Sokolská tř. 33, Ostrava 1, 701 21, zuzana.kucerova@vsb.cz.
[2] Faculty of Economics, VSB-Technical University of Ostrava, Dept. of Economic Policy, Sokolská tř. 33, Ostrava 1, 701 21, jitka.pomenkova@vsb.cz.

the countries' international investment positions with a view to the foreign trade and the method of correlation analysis (including the moving correlation method). The text is structured as follows.

1.1 Previous empirical research

Lane and Milesi-Ferretti [13] created a methodology to produce a unique data set containing an estimation of foreign assets and liabilities for a large set of industrial and developing countries. This is one of the first attempts to study the foreign assets and liabilities. In another paper, Lane and Milesi-Ferretti [15] examine the cross-country and time-series variation in the size of international balance sheets. They study the relation between foreign assets and liabilities on one side and a set of various regressors (GDP per capita, trade openness, external liberalisation, financial depth, stock market capitalisation, privatisation revenues etc.) on the other side. They find that international trade and stock market capitalisation are the two most important variables influencing international balance sheets. This study was updated in Lane and Milesi-Ferretti [17].

Kose *et al*. [10] focus on cross-country trade and financial linkages and produce a comprehensive analysis of the roles of both trade and financial integration in driving the growth-volatility relationship. They conclude that both trade and financial integration significantly weaken the negative association between output volatility and growth. Kose *et al*. [11] analyse the impact of selected macroeconomic variables (the depth of financial markets, trade openness, real GDP per capita, macroeconomic policies stability, institutional quality, and the regulation of an economy) on a country's financial openness (the sum of financial assets and liabilities relative to nominal GDP). They conclude that foreign direct investments and cross-border flows of equity securities are safer for the economy than cross-border flows of debt securities especially in the case of a low level of a country's financial openness and quality of institutions.

According to Rusek [18] and Spiegel [20], a common currency fosters the foreign trade of the euro area countries (the "euro effect"). Whereas foreign trade requires external financing, trade integration intensifies financial integration. Sebnem *et al*. [19] investigate the underlying channels of the "euro effect" on financial integration, i.e. the elimination of the currency risk among euro area countries, various financial sector legislative-regulatory reforms or increased goods trade. They find that the impact of this effect on financial integration is primarily driven by eliminating the currency risk. While financial and trade integration are highly correlated processes, trade in goods does not play a role in explaining the positive effect of euro on financial integration.

Aviat and Coeurdacier [2] explore the complementarity between bilateral trade in goods and bilateral asset holdings in a simultaneous gravity equations framework. According to results, trade in goods and trade in assets are closely related. They find a very robust and significant effect of trade on financial asset holdings and this causality runs in both ways; however, the impact of asset holdings on trade in goods is smaller. Kucerova [12] confirms the same results by using the simultaneous equations model. Aizenman and Noy [1] also study the endogenous determination of financial and trade openness. They construct a theoretical framework leading to two-way feedbacks between financial and trade openness and then identify these feedbacks empirically. They find that countries cannot choose the degree of financial openness independently of their degree of trade openness.

1.2 Methods

In order to quantify co-movements between time series, we use the method of *dynamic correlation* according to Croux et al. [4]. It measures the similarity of the frequency components of two time series *y* and *z* and can be defined as:

$$\rho_{yz}(\omega) = \frac{C_{yz}(\omega)}{\sqrt{S_z(\omega) S_y(\omega)}}, \tag{1}$$

where *Cyz* is a co-spectrum (the real part of the cross-spectrum) and *Sy*, *Sz* are the individual spectra of time series *y* and *z* for frequencies ω. The dynamic correlation values belong to the interval from -1 to +1. Integrating the equation (1) in the frequency band from ω_1 to ω_2 a dynamic correlation coefficient arises:

$$\rho_{yz}(\omega_1, \omega_2) = \frac{\int_{\omega_1}^{\omega_2} C_{yz}(\omega) d\omega}{\sqrt{\int_{\omega_1}^{\omega_2} S_z(\omega) d\omega \int_{\omega_1}^{\omega_2} S_y(\omega) d\omega}} \tag{2}$$

which evaluates the common behaviour of two time series in the given band of frequencies. For ω_1=0, ω_2 =π the integration is done over the whole defined frequency range and thus the dynamic correlation coefficient corresponds to the classical correlation coefficient (Fidrmuc et al. [6]).

1.3 Data

The yearly data 1993-2012 used to calculate the measures of financial integration are from the International Monetary Fund International Financial Statistics online database, specifically a category called the international investment position (IMF [8]). Incomplete data for some countries and some years have been completed from the on-line database External Wealth of Nations Mark II (Lane and Milesi-Ferretti [16]). Data concerning nominal exports and imports (in USD) are also extracted from the on-line database IMF IFS. Data concerning nominal GDP (in USD) are extracted from the on-line database IMF World Economic Outlook (WEO) Database, version April 2013 (IMF [9]). Each variable defined below (IFI, GI, GEQ, TRADE, etc.) was calculated as the average of corresponding values for 10 representative countries of EU, namely Bulgaria, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia, Slovenia.

Foreign assets include several categories: foreign assets, foreign direct investment abroad, portfolio investment equity securities, portfolio investment debt securities, financial derivatives, other investment/sectors, reserve assets. Foreign liabilities assets include these categories: foreign direct investment in the economy, portfolio investment equity securities, portfolio investment debt securities, financial derivatives, other investment/sectors.

The variable IFI_{it} is an average indicator of financial integration. It is *a quantity-based measure* of financial integration. This indicator is constructed as follows:

$$IFI_{it} = \frac{\left(FA_{it} + FL_{it}\right)}{GDP_{it}}, \tag{3}$$

where FA_{it} is the stock of total foreign assets of country i in time t, FL_{it} is the stock of total financial liabilities of country i in time t and GDP_{it} is the nominal GDP of country i in time t. Absolute levels of any variable do not reflect the size of the economy properly, can be misleading and are thus not convenient for direct comparisons of different countries. Therefore, it is better to adjust the *IFI* indicator by including the nominal GDP of countries in order to take into account the size of the economy and reveal the true differences in the level of financial integration. In our empirical analysis we use also separated indicators *FA* and *FL*.

The second average measure is *the investment-based measure* of financial integration (*GI*): it contains only foreign direct investments and portfolio investments (equity and debt securities). The other categories were dropped from this measure because they are either volatile (other investments) or time series are not long enough (financial derivatives). The construction of this adjusted measure is as follows:

$$GI_{it} = \frac{\left(FDIA_{it} + FDIL_{it} + PEQA_{it} + PEQL_{it} + PDEA_{it} + PDEL_{it}\right)}{GDP_{it}} \tag{4}$$

where $FDIA_{it}$ is the stock of foreign direct investment assets of country i abroad, $FDIL_{it}$ is the stock of foreign direct investment liabilities of the rest of the world in country i, $PEQA_{it}$ is the stock of portfolio equity assets of country i abroad, $PEQL_{it}$ the stock of portfolio equity liabilities in country i, $PDEA_{it}$ the stock of portfolio debt assets of country i abroad, and $PDEL_{it}$ is the stock of portfolio debt liabilities in a country i.

A third possible average measure of financial market integration – *the equity-based measure* of financial integration (*GEQ*) – is based solely on the equity cross-holdings – that is, flows of portfolio equity and foreign direct investments. As international trade in debt instruments can be sometimes influenced by special factors, it was omitted in this indicator:

$$GEQ_{it} = \frac{\left(FDIA_{it} + FDIL_{it} + PEQA_{it} + PEQL_{it}\right)}{GDP_{it}} \tag{5}$$

Trade openness is expressed by using the *TRADE* indicator and this indicator of trade integration is constructed as follows:

$$TRADE_{it} = \frac{\left(EX_{it} + IM_{it}\right)}{GDP_{it}}, \tag{6}$$

where EX_{it} is the total sum of exports of country i in time t, IM_{it} is the total sum of imports of country i in time t and GDP_{it} is the nominal GDP of country i in time t. The higher the value of this indicator, the higher the country's trade openness is.

1.4 Correlation Analysis

In the first step of empirical analysis we calculate classical (static) correlation coefficients for the annual data 1993-2012. In the second step the dynamic correlations were calculated according eq. (2) for ω_1=0, ω_2 =π. Correlated variables are denoted as bold in the tab. 1. Tab. 1 (lower left triangle) shows a significant (p-value=0,000 for all coefficients) positive dependence between trade and other measures of financial integration measured by static correlation. When calculating the dynamic correlation over the whole frequency range (Tab. 1, upper right triangle), we should get the same results as in the case of the classical correlation. Minor changes are due to the computational precision of the software Matlab R2011b. of calculation.

	TRADE	**IFI**	**GI**	**GEQ**		**TRADE**	**FDI**	**PEQ**	**PDE**
TRADE	1	0,8287	0,8252	0,8168	**TRADE**	1	0,8034	0,8369	0,8715
IFI	0,8270	1	0,9739	0,9802	**FDI**	0,8013	1	0,8183	0,9412
GI	0,8233	0,9736	1	0,9975	**PEQ**	0,8350	0,8160	1	0,8830
GEQ	0,8148	0,9799	0,9975	1	**PDE**	0,8696	0,9410	0,8815	1

Table 1 Correlation coefficients: static and dynamic correlation

Source: own calculation
Note: Lower left triangle contains correlation coefficient; upper right triangle presents dynamic correlation.

Unfortunately, classical correlation does not provide a detailed view of the structure of dependence. Therefore, we proceed with the calculation of the moving correlation (Tab. 2). Moving correlation is calculated as classical static correlation, but on the moving time window. We establish the moving part of the size of 10 observations, we start with time window 1993-2002 and we move per one observation, i.e. 1994-2003, 1995-2005 etc. Denote the sample size is n=20.

TRADE	**1993-2002**	**1994-2003**	**1995-2004**	**1996-2005**	**1997-2006**	**1998-2007**	**1999-2008**	**2000-2009**	**2001-2010**	**2002-2011**	**2003-2012**
	1	2	3	4	5	6	7	8	9	10	11
						Correlation coefficient					
IFI	0,8786	0,8997	0,9228	0,9102	0,9468	0,9253	0,9163	0,3730	0,3805	0,3254	0,3847
GI	0,6314	0,8836	0,9018	0,8811	0,9355	0,9269	0,9086	0,3739	0,3808	0,3324	0,4008
GEQ	0,6598	0,8845	0,9052	0,8992	0,9402	0,9196	0,8913	0,3887	0,3931	0,3453	0,3941
						p-value					
IFI	0,0008	0,0004	0,0001	0,0003	0,0000	0,0001	0,0002	0,2884	0,2780	0,3588	0,2723
GI	0,0502	0,0007	0,0004	0,0008	0,0001	0,0001	0,0003	0,2872	0,2776	0,3481	0,2510
GEQ	0,0379	0,0007	0,0003	0,0004	0,0001	0,0002	0,0005	0,2669	0,2611	0,3285	0,2598
						Correlation coefficient					
FDI	0,6409	0,8839	0,9039	0,8983	0,9363	0,9182	0,8709	0,3523	0,3596	0,3140	0,3729
PEQ	0,6635	0,6769	0,7596	0,8060	0,8726	0,9001	0,7835	0,6569	0,6532	0,5835	0,4866
PDE	0,6430	0,8648	0,8921	0,9015	0,9241	0,9375	0,9299	0,5494	0,5307	0,4666	0,4948
						p-value					
FDI	0,0459	0,0007	0,0003	0,0004	0,0001	0,0002	0,0010	0,3180	0,3075	0,3769	0,2886
PEQ	0,0365	0,0316	0,0108	0,0049	0,0010	0,0004	0,0073	0,0391	0,0406	0,0766	0,1538
PDE	0,0449	0,0012	0,0005	0,0004	0,0001	0,0001	0,0001	0,1000	0,1145	0,1740	0,1460

Table 2 Correlation coefficients: moving version of static correlation

Source: own calculation

In the Tab. 2, we can see the statistically significant dependence till the year 2008. After that, a moving part containing values after 2008 indicates a substantial decrease in the level of correlation caused by the financial crisis, which began in 2007 and has turned into economic and debt crisis. There is only one indicator (*PEQ*, containing shares, stocks and depository receipts) showing a significant dependence with trade after year 2008. In the case of *PDE*, the drop in correlation coefficients was not so large, but the results are not as significant as in the case of the *PEQ*. FDI experienced the worst drop in correlation.

Additional detailed view of the structure of correlation provides the dynamic correlation (equation (1)). We present the development of dynamic correlation between *TRADE* and the selected indicators in relation to different frequencies in Fig. 1a (*FDI*, *PEQ*, *PDE*) and Fig. 2a (*IFI*, *GI*, *GEQ*). For a better illustration, we present moving correlation figures too; in order to compare the results of these two methods, see Fig. 1b and Fig. 2b.

According to the results, there is the same tendency in the correlation in the both groups of indicators (*FDI*, *PEQ*, *PDE* and *IFI*, *GI*, *GEQ*) within the period 2000-2009.

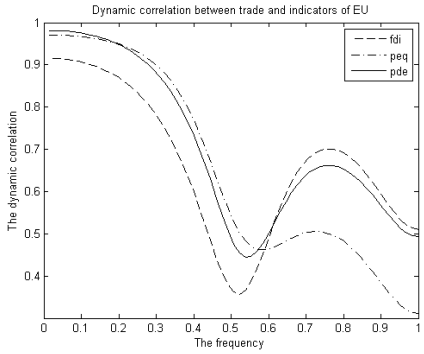


Figure 1a Dynamic correlation curves

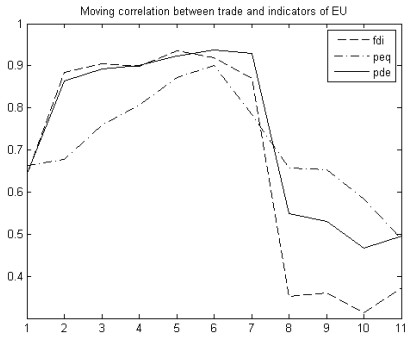


Figure 1b Moving correlation curves

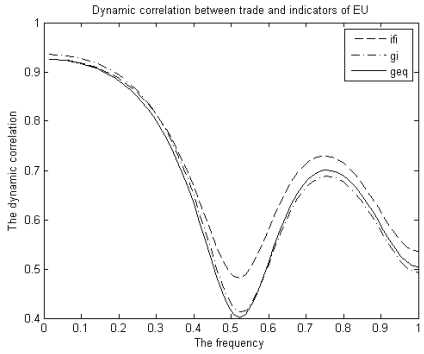


Figure 2a Dynamic correlation curves

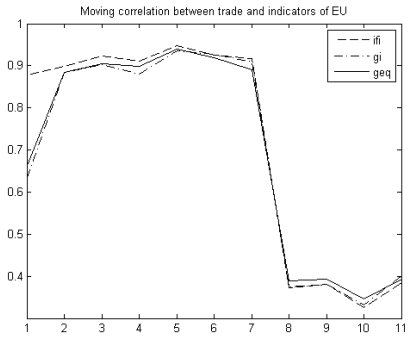


Figure 2b Moving correlation curves

In Tab. 1, we present the dynamic correlation over the whole frequency range (equation (2)). The variability of all dynamic correlation curves is presented in Tab. 3. In the case of dependencies between *TRADE* and the selected financial indicators (Fig. 1a), we see the volatility of the dynamic correlation curve, which varies for different lengths of cyclic components (frequencies). For all indicators, there is a high correlation with the indicator of trade in long cycles and economic cycles (low frequency value on the *x*-axis, i.e. the frequency value on the *x*-axis ranges from 0.001-0.33), the correlation is close to 0.8. Dynamic correlations corresponding to short cycles also achieve high levels in the case of the *FDI* and *PDE* indicators (higher frequencies, i.e. 0.65-0.85). This similar tendency is apparent in the Fig. 2a for the *IFI*, *GI* and *GEQ* indicators. Contrary to the Fig. 1a, all the indicators have the same development tendency.

	TRADE	**IFI**	**GI**	**GEQ**		**TRADE**	**FDI**	**PEQ**	**PDE**
TRADE	0%	2,008%	2,804%	2,671%	**TRADE**	0%	2,867%	5,507%	3,423%
IFI	2,008%	0%	0,169%	0,157%	**FDI**	2,867%	0%	3,279%	0,0745%
GI	2,804%	0,169%	0%	0,0002%	**PEQ**	5,507%	3,279%	0%	1,735%
GEQ	2,671%	0,157%	0,0002%	0%	**PDE**	3,423%	0,075%	1,735%	0%

Table 3 Variability of dynamic correlation

Source: own calculation

1.5 Conclusion

The process of monetary integration in Europe has to a great extent influenced the European financial markets because a single currency requires well-integrated financial markets. Financial integration is a substantial condition for introducing a single currency. However, foreign trade (i.e. trade integration) is an important factor influencing financial integration. This paper analysed the relationship between financial and trade integration in the EU10 member countries over the period 1993-2012. We used quantity-based measures of financial integration derived from the countries' international investment positions with a view to foreign trade by using various the methods of classical and dynamic correlation.

We concluded that financial and trade integration has been deepening since 1993. Both integration processes were broken by the world financial crises (the process of financial integration in 2008 and the process of trade integration in 2009). The deepening integration trend was restored one year later in both cases. We also conclude that there is a linear association between financial and trade integration in the pre-crisis period (1993–2008) in the EU10 countries. However, this relationship was seriously weakened by the financial crisis. These integration processes are interconnected, i.e. the more the countries trade the more financially integrated they are. It confirms a strong relationship between the two main balance-of-payment components: the current account and the financial account.

Acknowledgements

The research behind this paper is supported by the Grant Agency of the Czech Republic within Project No. 403/11/2073.

References

[1] Aizenman, J., and Noy, I.: Endogenous Financial and Trade Openness. *Review of Development Economics* **2** (2009), 175–189.
[2] Aviat, A., and Coeurdacier, N.: The Geography of Trade in Goods and Asset Holdings. *Journal of International Economics* **1** (2007), 22–51.
[3] Baele, L., Ferrando, A., Hördahl, P., Krylova, E., and Monnet, C.: Measuring Financial Integration in the Euro Area. *European Central Bank Occasional Paper* **14** (2004).
[4] Croux, C., Forni, M., and Reichlin, L.: A Measure of Comovement for Economic Variables: Theory and Empirics. *The Review of Economics and Statistics* **2** (2001), 232–241.
[5] European Commission: European Financial Stability and Integration Report 2010. *Commission Staff Working Paper* **SEC(2011) 489** (2011).
[6] Fidrmuc, J., Ikeda, T., and Iwatsubo, K.: International transmission of business cycles: Evidence from dynamic correlations. *Economics Letters* **114** (2012), 252-255.
[7] Green, W.H.: *Econometric Analysis*. Prentice Hall, New Jersey, 2012
[8] International Monetary Fund: *International Monetary Fund International Financial Statistics* [online]. 2013 [cit. 2013-04-22]. Available from Internet: http://www.imf.org/external/data.htm.
[9] International Monetary Fund: *World Economic Outlook* [online]. 2013 [cit. 2013-04-25]. Available from Internet: http://www.imf.org/external/data.htm.
[10] Kose, M. A., Prasad, E. S., and TERRONES, M. E.: How Do Trade and Financial Integration Affect the Relationship between Growth and Volatility? *Journal of International Economics* **1** (2006), 176–202.
[11] Kose, M. A., Prasad, E. S., and Taylor, A. D.: Thresholds in the Process of International Financial Integration. *Journal of International Money and Finance* **1** (2011), 147-179.
[12] Kucerova, Z. Measuring Financial Integration in Central Europe through International Investment Positions. *Eastern European Economics* **4** (2009), 25-41.
[13] Lane, P. R., and Milesi-Ferretti, G. M.: The External Wealth of Nations: Measures of Foreign Assets and Liabilities for Industrial and Developing Countries. *Centre for Economic Policy Research Discussion Paper* **2231**(1999).
[14] Lane, P. R., and Milesi-Ferretti, G. M.: External Capital Structure: Theory and Evidence [online]. *International Monetary Fund Working Paper* **WP/00/152** (2000).
[15] Lane, P. R., and Milesi-Ferretti, G. M.: International Financial Integration. *International Monetary Fund Working Paper* **WP/03/86** (2003).
[16] Lane, P. R., and Milesi-Ferretti, G. M.: The External Wealth of Nations Mark II. *Journal of International Economics* **73** (2007), 223-250.
[17] Lane, P.R., and Milesi-Ferretti, G. M.: The Drivers of Financial Globalization. *American Economic Review* **98** (2008), 327-32.
[18] Rusek, A..: Financial Integration and the New EU Member Countries: Challenges and Dilemmas. *Prague Economic Papers* **1** (2005), 17–32.
[19] Sebnem, K. O., Papaioannou, E., and Peydro, J. L.: What Lies Beneath the Euro's Effect on Financial Integration? Currency Risk, Legal Harmonization, or Trade? *Journal of International Economics* **1** (2010), 75–88.
[20] Spiegel, M. M.: Monetary and Financial Integration in the EMU: Push or Pull? *Review of International Economics* **17** (2009), 751–776.

Pay-to-Bid Auctions: Evidence from Czech Data

Vojtěch Kuna[1], Ondřej Krčál[2]

Abstract. This paper examines properties of pay-to-bid auctions. Pay-to-bid auctions represent a relatively new type of all-pay auctions in which participants have to pay a fee in order to increase the purchase price by a small fixed bidding increment. The player who submits the last bid wins the right to buy the auctioned item at the current purchase price, which is normally substantially lower than the value of the item. While most of the existing studies examine the data on the international server Swoopo.com, we present a unique dataset retrieved from the Czech auction site Bonus.cz. In the paper, we introduce econometric models that explain the total number of bids and seller's revenue using data from the period between July 2009 and April 2010 where the basic properties of the auctions correspond to the stylized facts found in the US data. We compare the results with predictions of the model of pay-to-bid auctions introduced by [3]. We find that the empirical evidence is consistent with theoretical predictions and suggests risk-seeking preferences of bidders.

Keywords: internet auctions, bidding fees, risk-seeking preferences.

JEL Classification: D44

AMS Classification: 62J05, 91A80

1 Introduction

Pay-to-bid auctions have been introduced as a new type of internet auctions relatively recently. Former leading provider of these auctions Swoopo.com, operating in Germany, Austria, the US, the UK, Canada and Spain, entered the market only in 2005. Pay-to-bid auctions quickly managed to raise interest of shoppers, media and academics with the number of websites soaring over one hundred in the US alone and monthly number of unique visitors reaching 16% of the traffic held by eBay, the undisputed leader in online auctions [3].

While the pay-to-bid auctions offered on the internet may differ in details, the basic concept remains the same across the websites. The auction site (or seller) offers a single item in an auction, for example a new model of mobile phone. The initial price is zero and there is a timer counting down from 30 seconds. If a player places a bid, the price of the item increases by a fixed amount and the timer is set at 30 seconds again. Players pay a bidding fee for each bid they make, usually in the form of pre-paid tokens. If nobody bids before the timer reaches zero, the person who made the last bid wins the right to buy the item at the current purchase price.

Using the data from the US version of Swoopo, the existing studies provide interesting stylized facts about pay-to-bid auctions. For instance, [1] reports that while most auctions end with a loss for the seller, the rest more than compensates for the losses. In fact, the average revenue from an auction reaches 150% of the retail price of the item on the auction. Moreover, the main part of seller's revenue comes from the bidding fees, not from the final purchase prices winners pay for the items. There are several game-theoretical models explaining the stylized facts ([1], [2], [3]). All papers note that the basic model of fully rational, risk-neutral players in an environment with full information is unable to explain the empirical data. However, they differ in the assumptions they use to improve the fit of their models. For example, [1] introduces the sunk-cost fallacy, and [2] assumes asymmetric information among players. In this paper, we compare the empirical evidence from the Czech auction server Bonus.cz with predictions of the model presented in [3], which is able to explain the profit of the auction sites using the basic environment of utility-maximizing players by adding risk-seeking preferences.

In the rest of the paper, we proceed as follows. In Section 2, we describe our dataset retrieved from Bonus.cz and provide basic descriptive statistics. In Section 3, we present the intuition behind the theoretical explanation in [3] and summarize predictions of the model. In Section 4, we introduce the econometric models, present the results and compare the empirical findings with theoretical predictions. Finally, we conclude in Section 5.

[1] Masaryk University, Department of Economics, Lipová 41a, Brno, 602 00, Czech Republic, v.kuna@mail.muni.cz.

[2] Masaryk University, Department of Economics, Lipová 41a, Brno, 602 00, Czech Republic, krcalo@mail.muni.cz.

2 Data

The full dataset contains information about more than 55 000 auctions from the auction site Bonus.cz completed in the period from July 2009 to April 2012. Our software robot collected all information about the completed auctions provided on the site – the name of the item, retail price, purchase price, bidding increment, number of winner's bids, and special rules used in the auction. Bonus offered ten different special rules, such as shorter countdown times, limited number of bids by a single player, the right to sell the item back to the site for 60% of its retail price or the right to sell the item for the full price in bidding tokens. There were auctions with zero, one, or more than one special rules applied.

Using the collected data, we calculated two additional variables: the total number of bids equal to the final purchase price divided by the bidding increment and the seller's revenue equal to the total number of bids times the sum of bidding increment and price of a token. The price of a token itself had to be calculated from the available data. Although the basic price of the token on the site was set to 8 CZK, the actual average price was somewhat lower due to the fact that tokens could also be bought in packs with free tokens included. Unfortunately, the data on the number of paid and free tokens were available only for the winners of the auctions. Therefore, we calculated the average price of a token (approx. equal to 6.1 CZK) using the percentage of free bids which was calculated from the data on free and paid tokens used by the winners.

Furthermore, we divided all auctioned items into the following 13 categories covering the basic range of consumer products and services:

- culture – books, movies, magazines, etc.,
- free time activities – vouchers (e.g. for horseback riding, sauna, parachuting) and tickets,
- gold – actual gold or virtual gold, which is equivalent to a monetary prize,
- electronics – computers, tablets, televisions, phones, cameras, etc.,
- food and drinks – mostly luxury items such as chocolate pralines, expensive coffee or bottles of alcohol, often in the form of gift baskets,
- appliances – washing machines, coffee machines, vacuum cleaners, microwave ovens, etc.,
- toys and sport equipment – Lego, Barbie dolls, sports equipment, slot car tracks, etc.,
- clothing and accessories – mostly clothing, bags and handbags,
- video games – mostly game consoles (Xbox, Playstation, Nintendo) and games for all platforms,
- cosmetics – mostly gift packages including creams, gels, shampoos,
- phone credit – phone credit of different nominal values,
- bidding tokens,
- and others.

Moreover, we divided the data in two periods according to the size of the bidding increment. The bidding increments were exclusively 1 CZK or 0.1 CZK in the first period (from July 2009 to April 2010) and mostly 0.01 CZK in the second period (from April 2010 to April 2012). The periods also differed in bidding behavior of buyers and in profitability of the auction site. While the average profitability of the auctions in the first period was positive and relatively high, which is consistent with the evidence from the US data (see [1] or [3]), it was negative in the second period. The reasons for this change remain a matter of speculation and are beyond the scope of this study. However, the decreasing trend of auction profitability is noticeable already during the first period. This suggests that changes in the parameters of the auctions, most importantly the introduction of auctions with the bidding increment of 0.01 CZK, were a reaction to the fall in profitability, and not its cause. Because of the abnormal results of the auctions in second period, we restrict our study only to the first period. The total number of observations in the restricted dataset is 13 817.

The stylized facts about the pay-to-bid auctions in the first period are as follows. The average purchase price is 258 CZK, which is only 6.3% of the average retail price. Hence, the purchase price paid by the winner covers only a small part of the retail price. Adding the costs of bidding tokens increases the average amount paid by the winners to 899 CZK which amounts to 21.8% of the average retail price. Moreover, while most auctions resulted in net loss for the seller (the median profit is –354 CZK), the average profit was positive (1 035 CZK, markup of 27.2%). The median and average number of bids in an auction was 268 and 802, respectively, showing a skewed distribution with a lot of auctions ending relatively early and a long tail of auctions with high number of bids.

3 The Platt et al.'s model of pay-to-bid auctions

The game-theoretical model introduced in [3] is ideal for our analysis because it offers clearly defined predictions that can be tested on the aggregated auction data. Furthermore, its theoretical parameters correspond to the variables in our dataset. The model assumes a common valuation of the auctioned item, known to all potential bidders. This allows us to isolate an important aspect of the game - the players are willing to bid (and pay the bidding fee) while knowing that the other players are willing to pay for the item more than the current purchase price and can overbid them in the following period. Their bid is therefore similar to a bet that no other player will bid after them. Because the auctioned items are new and normally available from traditional retailers, the

common valuation seems a reasonable assumption. In the rest of the section, we will present a basic setup of the game, the intuition behind the equilibrium outcome and predictions that will be tested on our data.

The game contains a constant number of players watching the auction. The auction is divided into periods which end after the timer runs out or after someone places a bid. In each period, all players who are not currently winning simultaneously decide whether to place a bid. If only one player decides to place a bid, she pays the bidding fee and enters the next period as the current winner and the purchase price increases by the bidding increment. If there are more players placing a bid, the player who pays the bidding fee and becomes the current winner is randomly selected out of these players. If no player places a bid, the auction ends. The current winner pays the purchase price and receives the item.

In this game, the so-called *indifference condition* is used to guarantee a reasonable equilibrium outcome. It means that in all periods, players have the same expected utility from bidding as well as from abstaining from bidding. They use individual mixed strategies with certain probabilities of bidding that guarantee aggregate probabilities of another bid being placed in the current period, such that players in the previous period are exactly indifferent about placing their bids. While there are other symmetric equilibria, they produce "degenerate" outcomes, meaning that auctions end without any bids or exactly after one bid. If players decided to use mixed strategies with higher probabilities of bidding in period q, they would also cease to be indifferent about their bids in the previous period $q - 1$. Due to a higher probability of being outbid in the next period, abstaining from bidding in period $q - 1$ would become a strictly dominant strategy. Consequently, placing a bid would become a strictly dominant strategy in period $q - 2$ as the winner in that period could not be outbid. This effect would propagate back to the first period, causing the "degenerate" outcome. Similar outcomes arise also for other deviations from the strategy leading to indifference between placing and not placing the bid.

Hence we consider only individual strategies that are consistent with the indifference condition. These strategies result in unique aggregate probability of the q^{th} bid being placed, conditional on the $q - 1^{\text{st}}$ bid having already been placed. We will also assume that players follow the same strategy in the first period even though they would not be forced to do so by the indifference condition because there are no previous periods.

The authors find symmetric subgame perfect equilibria for two cases of players' utility functions - risk-neutral bidders and bidders with risk preferences. Let's consider the situation with risk-neutral players first. The players have the utility function $u(w) = w$, where w is the wealth of the player which depends on the initial endowment of the player, on the bidding decision of the player and on the outcome of the auction. Using the indifference condition, the probability density function for unconditional probability of an auction ending exactly after q bids is given by

$$f(q) = \left(1 - \mu_{q+1}\right) \prod_{j=1}^{q} \mu_j = \frac{b}{v - s \cdot q} \prod_{j=1}^{q} \left(1 - \frac{b}{v - s \cdot (j-1)}\right) \tag{1}$$

where q is the number of bids, b is the bidding fee, v is the value of the item, s is the bidding increment and μ_q is the aggregate probability of the q^{th} bid occurring. Because of the form of the utility function, the probability density function $f(q)$ is not a function of the unobservable endowment of players. The unconditional probability of the auction ending $f(q)$ is decreasing in the number of bids q. The expected number of bids is increasing in the value of the item v and decreasing in the bidding fee b and bidding increment s. The authors also show that expected revenue from this auction is equal to the item's valuation v and is independent from the bidding fee b and bidding increment s.

Next, we describe the more general situation in which preferences of players are represented by the CARA utility function $u(w) = (1/\alpha)(1 - \mathrm{e})^{-\alpha w}$, where α is the absolute risk aversion parameter. Positive values of α indicate risk-averse preferences and negative values indicate risk-seeking preferences. The advantage of the functional form used is that the unobservable initial endowment is not present in the final probability density function, which is given by

$$f(q) \equiv \left(1 - \mu_{q+1}\right) \prod_{j=1}^{q} \mu_j = \frac{e^{\alpha b} - 1}{e^{\alpha b} - e^{\alpha(b + sq - v)}} \prod_{j=1}^{q} \left(\frac{1 - e^{\alpha\left[b + s(j-1) - v\right]}}{e^{\alpha b} - e^{\alpha\left[b + s(j-1) - v\right]}}\right). \tag{2}$$

As for risk neutral players, the function is decreasing in number of bids q. The expected number of bids is increasing in the value of the item v and decreasing in the bidding fee b, bidding increment s and risk-aversion parameter α. A reduction in α leading to more risk-seeking preferences increases the expected number of bids, as players are willing to accept lower monetary payoffs and more adverse probabilities of winnings. The expected revenue from the auction cannot be analytically derived in this case. However, the effect of changes in parameters can be found using numerical simulations. The authors in [3] show that for risk-seeking players, the expected revenue increases in bidding fees b and decreases in bidding increment s and in risk-aversion parameter α. If players are more risk-seeking, they have additional utility from the risk which the seller is able to extract as revenue thanks to the indifference condition. A rise in the bidding fee b or a reduction in the bidding increment s cause higher variance in outcomes which contributes to higher utility of risk-seeking players. Conversely, a rise

in bidding fee b and a reduction in the bidding increment s reduces the expected revenue for risk-averse players. Furthermore, our simulations showed that the expected revenue is increasing in the value of the item v. Predictions of the model are summarized in Table 1.

	Players' risk preferences		
	risk neutral	**risk-seeking**	**risk-averse**
Expected number of bids E($f(q)$)	*increasing* in v *decreasing* in b and s	*increasing* in v *decreasing* in b, s, and α	*increasing* in v *decreasing* in b, s, and α
Expected revenue E(rev)	*increasing* in v (E(rev) = v)	*increasing* in v and b *decreasing* in s and α	*increasing* in v *and* s *decreasing* in b and α

Table 1 Predictions of the model

4 Empirical analysis

In this section, we present econometric models with two explained variables: number of bids (in log) and seller's revenue (in log of CZK). Because there is no theoretical foundation concerning the exact form of the estimated equations, we use several specifications and examine the estimated effects of the variables with respect to their stability across all specifications. Both explained variables were estimated as depending on the auctioned item's retail price, dummy variables for the value of bidding increment, dummy variables for special auction types and dummy variables for item categories. Some specifications also include the item's price to the power of 2 and to the power of 3, log of the price or interaction of the price with the dummy for the bidding increment value. The model specifications are described in Table 2. We estimated the models on the data from the first period (with the bidding increments of 1 and 0.1 CZK) using the ordinary least squares method with heteroskedasticity-consistent standard errors. Variables with high p-values were sequentially eliminated until the resulting model consisted only of the variables significant at least at the 0.10 level.

specification	**retail price**	**interaction**	**dummies**
base	basic	-	types, categories
log	log of	-	types, categories
exponential	second power, third power	-	types, categories
interaction	basic	crown * price	types, categories
combined	second power, third power	crown * price	types, categories

Table 2 Estimated specifications

Predictions of the theoretical model [3] are summarized in Table 2. Specifically, the model predicts that the expected number of bids increases in retail price, which measures the value of the item v, and decreases in bidding increment s. Hence the effect of retail the price on the number of bids should be positive and the effect of the dummy variable for the increment 1 CZK should be negative. These effects hold regardless of risk preferences of bidders. Furthermore, the model predicts a positive effect of retail price on seller's revenue. The bidding increment s has no effect on expected revenue for risk-neutral preferences, positive effect for risk-averse preferences and negative effect for risk-seeking preferences.

Table 3 presents summary of the results. For each variable and model, we show a total number of specifications for which the estimated parameter is positive, negative and not statistically significant at the 10% level [positive / negative / not significant][3]. If the results are consistent at least across three different specifications and there is no major conflict, we conclude that a variable has either positive (+) or negative impact (–). Otherwise, results remain inconclusive (0).

[3] For the sake of easy orientation, we present results of various parameterizations using just two variables - *retail price* and *increment 1 CZK*. In some of the estimations, the influence of price is not strictly positive throughout all possible values. However, the price range where the influence is negative is limited to very low or very high prices and only a handful of items fall into this category.

	model	
variable	number of bids	seller's revenue
retail price	+ [5/0/0]	+ [5/0/0]
increment 1 CZK	- [0/5/0]	- [0/5/0]
some tokens returned after the end of auction	+ [5/0/0]	+ [5/0/0]
right to purchase the item with a price cut	+ [5/0/0]	+ [5/0/0]
limited number of bids by a single player	- [0/5/0]	- [0/5/0]
progressively shortening timer	- [0/4/1]	0 [0/2/3]
shorter timer	0 [1/0/4]	0 [1/0/4]
right to sell the item back	0 [0/0/5]	0 [0/0/5]
appliances	+ [5/0/0]	+ [5/0/0]
toys and sport equipment	+ [5/0/0]	+ [5/0/0]
bidding tokens	+ [5/0/0]	+ [5/0/0]
electronics	+ [5/0/0]	+ [3/0/2]
culture	- [0/5/0]	- [0/5/0]
free time activities	- [0/5/0]	- [0/5/0]
video games	- [0/3/2]	- [0/3/2]
gold	0 [2/1/2]	0 [2/3/0]
food and drinks	0 [2/0/3]	0 [2/0/3]
clothing and accessories	0 [0/1/4]	0 [0/1/4]
cosmetics	0 [2/0/3]	0 [2/0/3]
phone credit	0 [2/0/3]	0 [2/0/3]
average adjusted R^2	0.35	0.35

Table 3 Estimation results

The results show the dominant influence of a variable: + (positive), – (negative), 0 (inconclusive); total number of specifications where variable parameter estimation is [positive / negative / not significant].

The results presented in Table 3 show a positive effect of retail price on the number of bids and seller's revenue and a negative effect of bidding increment on the number of bids. This is consistent with the model predictions regardless of players' risk preferences. On the other hand, the predicted effect of the bidding increment depends on risk preferences of players. A negative effect of bidding increment is consistent with predictions of the model with risk-seeking players. Hence our findings suggest that risk-seeking bidders were dominant in the online auctions on Bonus. This result is supported by Platt at el. [3] who find that players participating in US version of Swoopo had predominantly risk-seeking preferences as well.

Moreover, Table 3 provides interesting findings concerning the effects of special auction types. If some tokens are returned after the end of auction or the winner can purchase the item with a price cut, players tend to bid more, which in turn generates higher revenue for the seller. Similarly, if the number of bids by a single player is limited, the average number of bids and the revenue of the seller are lower. Hence, the results for special auction types seem to be consistent with the assumption of rational players. Furthermore, there seem to be categories of auctioned items which are generally more popular and able to attract more bids and the revenue the seller earned on these items was higher compared to the category others. This is especially true for appliances, toys and sport equipment, bidding tokens and electronics. On the other hand, the items in categories free time activities, culture and video games attracted rather less bids and generated relatively low seller's revenues.

5 Conclusion

In this paper, we analyze a unique dataset recovered from the Czech auction site Bonus.cz. In the period from July 2009 to April 2010, the data reveals similar pattern of profitability from individual auctions to that reported by other studies using US data from Swoopo.com (see [1] and [3]). It shows that a majority of the auctions ended with a low number of bids and a loss for the auction site. However, a relatively low number of auctions attracted a large number of bids, so that the overall average profit for the seller was positive. Using the data for the period from July 2009 to April 2010, we estimate the effect of retail price, bidding increment and dummy variables for special types of the auction and categories of the items on number of bids and revenue of the seller. We find that a rise in retail price and a reduction bidding increment increases the number of bids. Furthermore, we find the same effect of both variables on revenue of the auction site. We compare the results with the predictions of the

model with rational players under perfect information in [3]. We find that the basic version of model with risk-neutral players is incapable of explaining our data. However, after incorporating risk-seeking preferences into the model, the predictions correspond perfectly to the empirical results. Hence, our analysis suggests that the bidding pattern on Bonus.cz is consistent with the optimal choices of utility-maximizing consumer under perfect information and risk-seeking preferences.

Acknowledgements

This paper was created as a part of the specific research project no. MUNI/A/0797/2012 at Masaryk University.

References

[1] Augenblick, N: *Consumer and Producer Behavior in the Market for Penny Auctions: A Theoretical and Empirical Analysis.* (2009), available at http://www.stanford.edu/~ned789/Augenblick_JMP_Penny_Auction.pdf.

[2] Byers, J. W., Mitzenmacher, M., and Zervas, G.: *Information Asymmetries in Pay-Per-Bid Auctions: How Swoopo Makes Bank.* (2010), available at http://arxiv.org/pdf/1001.0592v3.pdf.

[3] Platt, B. C., Price, J., and Tappen, H.: The Role of Risk Preferences in Pay-to-Bid Auctions. (2012), available at https://economics.byu.edu/SiteAssets/Pages/Faculty/Brennan%20Platt/PayToBid.pdf.

Economic Analysis of Czech Regions by Econometric Models, MCDM and DEA

Martina Kuncová[1], Jana Sekničková[2]

Abstract. The Czech statistical office regularly publishes economic data about Czech regions. Based on these data we would like to analyze selected economic indices, such as income of regions, amount of grants and subsidies transferred to other subjects etc., and to model them using by econometric models. We try to use equilibrium models and to interpret obtained results. With respect to econometric results we would like to compare Czech regions with each other. The economics of regions can be evaluated also by multicriteria decision making (MCDM) analysis or using data envelopment analysis (DEA) models and we would like to compare econometric results also with MCDM and DEA ones.

Keywords: Czech Regions, economic indicators, budgets, DEA, multicriteria, regression

JEL Classification: C01, C30, C44, C67
AMS Classification: 62J05, 62J12, 62P20, 90B50, 90C29, 91B06

1 Introduction

Since 2000, the Czech Republic is divided into thirteen regions – Central Bohemian Region, South Bohemian Region, Plzeň Region, Karlovy Vary Region, Ústí nad Labem Region, Liberec Region, Hradec Králové Region, Pardubice Region, Vysočina Region, South Moravian Region, Olomouc Region, Moravian-Silesian Region, Zlín Region – and Prague, the Capital City (see figure 1). Each region has its own elected Regional Assembly and hetman. In Prague, their powers are executed by the city council and the mayor.

Figure 1 Czech regions [4]

In this paper we would like to analyze selected data from regional budgets and also from the Czech statistical office and compare all fourteen Czech regions with each other. For our analysis we used official budgets of regions from years 2005 – 2012 available from official web pages of Czech regions. Relevant data from these budgets are income of region (in thousands of CZK), common expenditures (mainly transferred grants and subsidies, in thousands of CZK), capital expenditures (mainly investments, in thousands of CZK) and expenditures of regional assembly (in thousands of CZK). We also use data taken from the web pages of Czech Statistical Office [4], especially unemployment rate (in percentage), ratio of economic activity (in percentage), average wage (in CZK), average age (in years) and the number of free workplaces.

[1] University of Economics Prague, Department of Econometrics, W.Churchill Sq.4, 13067 Prague 3, Czech Republic, kuncovam@vse.cz
[2]University of Economics Prague, Department of Econometrics, W.Churchill Sq.4, 13067 Prague 3, Czech Republic, jana.seknickova@vse.cz

We assume that region with higher income is more effective than the one with lower budget. Higher income indicates higher amount of grants and subsidies transferred to the economic subjects. Also the region with higher amount of investments is evaluated better as well as region with lower unemployment rate and higher economic activity. Of course we would like region with young people so the lower average age is desirable and higher average wage is also wanted.

For comparison of regions it is possible to use various kinds of techniques and methods such as statistical methods, econometric models, data envelopment analysis (DEA) models, or multicriteria decision making (MCDM) models, especially multicriteria evaluation of alternatives (MCEA). The aim of this article is to compare the Czech regions from the economic point of view via econometric model (chapter 2), MCEA (chapter 4) and DEA models (chapter 3) and also compare the results of this methods (chapter 5).

2 Econometric model

For analysis of relationships between described economic indicators we used the simple econometric model of simultaneous equations. This model consists of two equations in following form:

$$C_t = C_0 + cY_t + u_t, \tag{1}$$

$$Y_t = C_t + I_t + G_t, \tag{2}$$

where C_t denotes consumption (in the case of this paper amount of expenditures) in the time t, Y_t denotes income of region, I_t denotes investments and G_t denotes expenditures of regional assembly. Symbols C_0 and c are parameters of model, C_0 denotes autonomous consumption and c denotes marginal propensity to consume. It is obvious that $c = \frac{\partial C_t}{\partial Y_t}$ and so this parameter can be interpreted as the increase in consumption with respect to increase of one unit of income.

The previous model can be transformed into two independent equations in the form:

$$C_t = \frac{C_0}{1-c} + \frac{c}{1-c} I_t + \frac{c}{1-c} G_t + v_t, \tag{3}$$

$$Y_t = \frac{C_0}{1-c} + \frac{1}{1-c} I_t + \frac{1}{1-c} G_t + v_t, \tag{4}$$

and the parameters of these equations can be simply estimated by ordinary least squares (or generalized least squares) [7]. We applied this theory on regional data and we obtained data in table 1.

Region		*c*	*c(std err)*	*p-value*	***Rank***
1.	Prague, the Capital City	0.668544	0.028304	6.19E-08	**10**
2.	Central Bohemian	0.628461	0.035494	4.52E-07	**11**
3.	South Bohemian	0.829656	0.018657	7.60E-10	**5**
4.	Plzeň	0.498379	0.037294	3.08E-06	**14**
5.	Karlovy Vary	0.887146	0.016535	2.05E-10	**1**
6.	Ústí nad Labem	0.829405	0.015272	1.88E-10	**6**
7.	Liberec	0.853120	0.018474	5.84E-10	**3**
8.	Hradec Králové	0.740908	0.016342	6.64E-10	**7**
9.	Pardubice	0.577770	0.026701	1.14E-07	**13**
10.	Vysočina	0.668813	0.084357	4.20E-03	**9**
11.	South Moravian	0.729998	0.024013	1.07E-08	**8**
12.	Olomouc	0.850651	0.024865	4.73E-09	**4**
13.	Moravian-Silesian	0.613880	0.018846	6.65E-09	**12**
14.	Zlín	0.864907	0.008427	2.20E-12	**2**

Table 1 Marginal propensity to consume

Note that for all regions parameter c is statistically significant at 1% significance level and parameter C_0 is nonsignificant at the same level. As consumption in our model includes mainly grants and subsidies intended for transfer to other subjects, marginal propensity to consume denotes how many percentage of regional income is

transferred in the form of grants and subsidies. It is clear that higher value of parameter *c* is better for people and the rank with respect to this parameter is in the last column of table 1.

We can see that the highest ratio of grants and subsidies in budget has Karlovy Vary Region, the second one is Zlín Region and the third is Liberec Region. Region Prague, the Capital City placed tenth place. It is given by large expenditures of town council and huge investments. The last region is Plzeň with really low consumption.

For information we analyzed also dependency of regional incomes on other economic factors introduced in introduction (average of years 2005 – 2012). For 10% level of significance regional income (Y_t) is dependent on constant, consumption (C_t), investments (I_t) and economic activity (A_t) in the form:

$$Y_t = -54.48 + 0.66\,C_t + 0.96\,I_t + 0.97A_t + e_t \tag{5}$$

and for 5% and 1% level of significance the model has this form:

$$Y_t = 1.30 + 0.83\,C_t + 0.98\,I_t + e_t. \tag{6}$$

Both models are statistically significant at 1% level and remaining variables are insignificant in these models.

3 Data envelopment analysis model

In the second step of our analysis we used the data envelopment analysis (DEA) not only to compare the regions but also to find which one is efficient. DEA belongs to the operational research methods, especially to the linear programming models, that have been used many times in private or public sector to evaluate the performances of many different kinds of entities (countries, regions, enterprises, schools, hospitals, insurance companies, military units etc.) engaged in many different kinds of activities in many different contexts [2]. These entities must have identical inputs and outputs to measure the efficiency from the same parameters.

DEA models are widely used in comparison of countries, regions or districts from various points of view. Melecky and Stanickova [9] compared evaluated the performance of the four Visegrad countries and their NUTS 2 regions. Friebel and Friebelova [5] measured life quality in 14 Southwest Czech districts (LAU 1) by DEA using 4 inputs and 1 output. These articles are aimed at the measurement of competitiveness by economic and social characteristics or measuring of the life quality.

The basic idea of DEA models consists in estimation of an efficient frontier that defines production possibility set of the problem. Based on the set of available decision making units (DMUs) DEA estimates so-called efficient frontier, and projects all DMUs onto this frontier. If a DMU lies on the frontier, it is referred to as an efficient unit, otherwise inefficient. DEA models can be oriented to inputs or outputs. In the case of input oriented models we assume fixed level of outputs (CCR-I), the output oriented model assumes fixed level of inputs and maximize level of outputs with respect to given inputs (CCR-O) [1]. These models are used if we assume constant return to scale. In the case of variable return to scale we work with BCC (Banker, Charnes, Cooper) models. The review and detailed information about DEA models can be found in [2]. The basic idea for the efficiency calculation is to maximize the rate of weighted sum of outputs divided by weighted sum of inputs. For example the model transformed (Charnes-Cooper transformation) into the linear programming form can be defined as follows (CCR-I):

$$\begin{aligned}
&\text{Maximize } z = \sum_{i=1}^{r} u_i y_{iq} \\
&\text{Subject to: } \sum_{i=1}^{r} u_i y_{ik} \le \sum_{j=1}^{m} v_j x_{jk}\ ,\ k = 1, 2, \ldots, n \\
&\sum_{j=1}^{m} v_j x_{jq} = 1 \\
&u_i \ge 0\ ,\ i=1,2,\ldots,m \\
&v_j \ge 0\ ,\ j=1,2,\ldots,r,
\end{aligned} \tag{7}$$

where *q* represents the evaluated DMU, y_{rj} are known outputs, x_{ij} are known inputs of the *j*th DMU, u_r and v_j are the variable weights to be determined by the solution of this problem. The efficient unit U_q lies on the efficient frontier in case that the optimal efficiency (calculated by the model) $z = 1$. The inefficient units have *z* lower than 1 (in CCR-I model) [1]. In some situations the outputs cannot be influenced or changed (like average age) – so the model is transformed into the nondiscreditory model where the weighted output is decrease by these uncontrollable outputs (models UCCR-I and UBCC-I).

In our case we compared 14 regions with respect to available data. The models had five inputs (unemployment rate, economic activity, average wage, average age and free workplaces per capita) and four outputs (marginal propensity to consume, income per capita, consumption per capita and investments per capita). The DEA model assumes inputs are minimized and outputs are maximized for efficient units. In this case we had to transform input data to minimize them. The results (objective values) obtained after transformation from DEA models are in table 2. Note that the last column (*AAD*) displays average absolute deviation from efficient frontier. Zero values denote regions those are efficient in all six models.

From the data in table 2 we can easily conclude that five regions are efficient. There are Prague, the Capital City, South Bohemian, Karlovy Vary, Liberec and Zlín Region – these are efficient in the case of all six DEA models. The absolutely inefficient is Moravian-Silesian Region with average absolute deviation more than 20%.

Region	*CCR-I*	*CCR-O*	*BCC-I*	*BCC-O*	*UCCR-I*	*UBCC-I*	***AAD***
Prague, the Capital City	1	1	1	1	1	1	***0***
Central Bohemian	0.8144	1.2279	1	1	0.7843	1	***0.1049***
South Bohemian	1	1	1	1	1	1	***0***
Plzeň	0.7047	1.419	1	1	0.6695	1	***0.1741***
Karlovy Vary	1	1	1	1	1	1	***0***
Ústí nad Labem	0.9778	1.0226	1	1	0.9154	1	***0.0216***
Liberec	1	1	1	1	1	1	***0***
Hradec Králové	0.8816	1.1343	0.982	1.1245	0.8684	0.9145	***0.1021***
Pardubice	0.7687	1.301	1	1	0.7132	1	***0.1365***
Vysočina	0.8128	1.2303	0.9965	1.1520	0.7512	0.9313	***0.1484***
South Moravian	0.8530	1.1724	0.9793	1.1612	0.8049	0.9063	***0.1317***
Olomouc	0.9588	1.0430	0.9840	1.0334	0.8977	0.8993	***0.0561***
Moravian-Silesian	0.7633	1.3101	0.9892	1.3073	0.6955	0.9085	***0.2101***
Zlín	1	1	1	1	1	1	***0***

Table 2 DEA results

4 Multicriteria evaluation of alternative model

In the third step of our analysis we used the multicriteria evaluation of alternative (MCEA). MCEA belongs to the category of discrete multi-criteria decision making models where all the alternatives (a_1, a_2, …, a_p) and criteria (f_1, f_2, …, f_k) are known. To solve this kind of model it is necessary to know the preferences of the decision maker. These preferences can be described by aspiration levels (or requirements), criteria order or by the weights of the criteria. For our analysis we used three methods – WSA, TOPSIS and ELECTREE III. that need only decision matrix **Y** and weights of criteria as inputs [5]. WSA (Weighted Sum Approach) sorts the alternatives based on the values of their utility functions which in this case are assumed to be linear. It requires the information about the weights of the criteria. Higher value of utility means better alternative.

The basic concept of the TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) method is that the best alternative should have the shortest distance from the ideal alternative and the farthest from the basal alternative. The method is also able to rank the alternatives using the relative index of distance of the alternatives from the basal alternative. Higher relative index of distance means better alternative. The user must supply only the information about the weights of criteria

ELECTRE III. method uses pairwise comparison of the alternatives and summarizes the weights of the criteria where the alternative is better than the other. According to the strength of the preference the final indifference classes are made. It is necessary that all the alternatives are nondominated (so there is no alternative that is better in at least one criterion and no worse in any other) [5].

For analysis we use the same criteria as in DEA model and weight of each criterion was equal (v_j = 1/9, j = 1, …, 9). We obtained the results displayed in table 3 (note that all alternatives are nondominated).

Column *Utility* displays weighted sum gained by WSA and *Rank WSA* presents order with respect to maximal utility. Similarly, *R.R.D.* column displays the relative ratio of distance to the basal alternative and *Rank TOPSIS* presents order with respect to maximal relative ratio of this distance. Column *Rank ELECTRE III.* presents order with respect to ELECTRE classes and *Average rank* is the mean of three previous ranks. *Total rank* sorts regions with respect to average rank.

MCEA with equal weights places South Bohemian Region at the first place before Prague, the capital city and Ústí nad Labem Region. Central Bohemian Region is the fourth and Karlovy Vary Region has the fifth place. The last ones are Pardubice Region together with South Moravian Region.

Region	*Utility*	*Rank WSA*	*R. R. D.*	*Rank TOPSIS*	*Rank ELECTRE III.*	*Average rank*	***Total rank***
Prague, the Capital City	0.71530	1	0.71081	1	5	2.3	**2**
Central Bohemian	0.42405	5	0.34276	4	3	4.0	**4**
South Bohemian	0.49468	3	0.40463	2	1	2.0	**1**
Plzeň	0.25370	13	0.24751	13	8	11.3	**12**
Karlovy Vary	0.50629	2	0.34048	5	6	4.3	**5**
Ústí nad Labem	0.44655	4	0.36677	3	2	3.0	**3**
Liberec	0.37764	7	0.29446	8	10	8.3	**8**
Hradec Králové	0.32561	10	0.27495	10	10	10.0	**10**
Pardubice	0.24963	14	0.25215	12	12	12.7	**13**
Vysočina	0.38033	6	0.32939	7	4	5.7	**6**
South Moravian	0.29332	12	0.24053	14	12	12.7	**13**
Olomouc	0.33361	9	0.27070	11	12	10.7	**11**
Moravian-Silesian	0.32052	11	0.33479	6	7	8.0	**7**
Zlín	0.35227	8	0.27497	9	9	8.7	**9**

Table 3 Multicriteria evaluation of alternatives – results

5 Results

Now we can compare results of all three parts. Table 4 summarizes the previous ranks.

Region		*Econometric rank*	*DEA rank*	*MCEA rank*	***Total rank***
1.	Prague, the Capital City	10	1	2	**5**
2.	Central Bohemian	11	9	4	**8**
3.	South Bohemian	5	1	1	**1**
4.	Plzeň	14	13	12	**14**
5.	Karlovy Vary	1	1	5	**1**
6.	Ústí nad Labem	6	6	3	**6**
7.	Liberec	3	1	8	**3**
8.	Hradec Králové	7	8	10	**9**
9.	Pardubice	13	11	13	**13**
10.	Vysočina	9	12	6	**10**
11.	South Moravian	8	10	13	**11**
12.	Olomouc	4	7	11	**7**
13.	Moravian-Silesian	12	14	7	**12**
14.	Zlín	2	1	9	**3**

Table 4 Recapitulation of results

Econometric and DEA models provide similar results expected Region Prague, the Capital City. With reference to marginal propensity to consume Prague, the Capital City places the tenth rank however DEA and MCEA place it at first two positions. This fact is clear from budget point of view. Prague, the Capital City has large income that divides between grants, investments and also expenditures of regional assembly. These expenditures

are much higher than in the case of others regions and so ratio of grants is smaller. In DEA and also MCEA models the other economic indicators play the role and so marginal propensity to consume is not so important with respect to income, investments and grants per capita where Prague, the Capital City places the first rank, as well as unemployment rate, economic activity, average wage and free workplaces.

The last column of table 4 denotes the rank of region with respect to average rank of all three approaches. Note that regions efficient according to DEA models are placed at first five places. First two places can be divided between South Bohemian and Karlovy Vary Regions, next two places are allocated to Liberec Region and Zlín Regions. Surprisingly Prague, the Capital City is placed at the fifth position. There follow inefficient regions (according to DEA model) and the last one is Plzeň Region.

6 Conclusion

The aim of this contribution was to analyze the economic position of the fourteen regions of the Czech Republic by different methods to set the order of the regions and also to compare the results of the methods. At first econometric model of simultaneous equations was used, next DEA models were calculated and finally mucltricriteria evaluation of alternatives was applied. It is clear that DEA models are not able always to create the order of the alternatives but the results are close to the econometric model (except of Prague but this region is always completely different than the others because of the capital city). The MCEA methods do not cover the relations among the criteria and so these results are more nuanced. As for the regions we can see that Prague do not need to be the best region, as Karlovy Vary starts to be better together with South Bohemian region, Liberec and Zlín.

Acknowledgements

The research was supported partly by the Grant Agency of the Czech Republic GAČR No. P403/12/1387 and also by the Internal Grant Agency of the University of Economics in Prague IGA F4/16/2013 and IGA F4/11/2013.

References

[1] Charnes A., Cooper W.W., and Rhodes, E.: Measuring the efficiency of decision making units. *European Journal of Operational Research* **2** (1978), 429–444.
[2] Cooper, W. W., Lawrence, M.S., and Zhu, J.: *Handbook on Data Envelopment Analysis*. Kluwer Academic Publishers, Norwell, 2004.
[3] Cooper, W.W., Seiford, L.M., and Tone, K.: *Introduction to Data Envelopment Analysis and Its Uses*. Springer Science + Business Media Inc., New York, 2006.
[4] Czech Statistical Office: Regions and districts. [online], available at http://www.czso.cz/x/krajedata.nsf/krajenejnovejsi/xj [cit. 2013-05-18]
[5] Figueira, J., Greco, S., and Ehrgott M.: *Multiple Criteria Decision Analysis – State of the Art Surveys*. New York : Springer Science + Business Media Inc., 2005.
[6] Friebel, L., and Friebelova, J.: Quantitative evaluation of life quality of Czech districts. *Proceedings of 30th* International Conference Mathematical Methods in Economics, Karvina, 2012, 190-195.
[7] Greene, W.H.: *Econometric Analysis*, 7th edition, Prentice-Hall, 2011.
[8] Jablonsky, J.: Software Support for Multiple Criteria Decision Making Problems. *Management Information Systems* Vol. **4** (2009), 29-34.
[9] Melecky, L., and Stanickova, M.: Assessment of efficiency in Visegrad countries and regions using DEA models. *Central European Review of Economic Issues* **15** (2012), 145-156.
[10] *Wikipedie*. The Czech Republic [online], available at http://en.wikipedia.org/wiki/Czech_Republic#Administrative_divisions [cit. 2013-05-22]

Comparison of selected municipalities of Vysocina region from the economic activity and traffic accessibility point of view using DEA model

Martina Kuncová[1], Jana Sekničková[2], Libuše Měrtlová[3], Hana Vojáčková[4]

Abstract. Traffic accessibility can be the key factor for economic activity, employment and unemployment for each municipality, district region. It is influenced by various factors like number of kilometers of roads, highways and railroads and also by travel times and number of public transportation lines or distance from the district center or from the center of the region. This article describes the analyses of selected municipalities from the economic activity and traffic accessibility point of view using DEA models. DEA (Data Envelopment Analysis) models are usually used to find the relative efficiency among homogenous units according to selected criteria (inputs and outputs). In this article we try to compare municipalities to find the problematic places.

Keywords: DEA models, Vysocina Region, traffic accessibility, economic activity

JEL Classification: C44, C67
AMS Classification: 90B50, 90B90

1 Introduction

Vysocina region lies in the center of the Czech Republic between Bohemia and Moravia. This region belongs to the agricultural and industrial ones. The unemployment rate is usually below the republic average but it differs from the districts inside the region. On the other hand the average wages are lower than in other regions. The regional disparities can be seen especially in the socio-economic indicators. If we take into account the GDP per capita and unemployment rate we can see that there are also big differences between the centers of the region and in the municipalities. Those which lie far from the industrial centers or from the highways have lower GDP and also higher unemployment. These facts cause the higher demand on social benefits, the decrease of inhabitants in those areas and the rising pressure for people to travel to the far-away centers. For travelling the traffic accessibility is crucial. The measurement of traffic accessibility is not easy because it is influenced by many factors such as number of kilometers of roads, highways and railroads and also by travel times and number of public transportation lines or distance from the district center or from the center of the region. In this paper we would like to compare the selected areas in Vysocina region according to the traffic accessibility conditions and economic activity represented by unemployment and number of businesses (companies and entrepreneurs).

For comparison of regions or municipalities it is possible to use various kinds of techniques and methods such as multi-criteria evaluation of alternatives, statistical methods, econometric models or DEA models. In this first phase we have decided for the data envelopment analysis (DEA) not only to compare the selected areas but also to find which one is efficient. DEA belongs to the operational research methods, especially to the linear programming models, that have been used many times in private or public sector to evaluate the performances of many different kinds of entities (countries, regions, enterprises, schools, hospitals, insurance companies, military units etc.) engaged in many different kinds of activities in many different contexts [2]. These entities must have identical inputs and outputs to measure the efficiency from the same parameters.

DEA models are widely used in comparison of countries, regions or districts from various points of view. Melecky and Stanickova [9] compared evaluated the performance of the four Visegrad countries and their NUTS

[1] College of Polytechnics Jihlava, Department of Economic Studies, Tolstého 16, 58601 Jihlava, Czech Republic, kuncova@vspj.cz
[2]University of Economics Prague, Department of Econometrics, W.Churchill Sq.4, 13067 Prague 3, Czech Republic, jana.seknickova@vse.cz
[3] College of Polytechnics Jihlava, Department of Economic Studies, Tolstého 16, 58601 Jihlava, Czech Republic, mertlova@vspj.cz
[4] College of Polytechnics Jihlava, Department of Electrotechnics and Informatics, Tolstého 16, 58601 Jihlava, Czech Republic, vojackova@vspj.cz

2 regions. Friebel and Friebelova [5] measured life quality in 14 Southwest Czech districts (LAU 1) by DEA using 4 inputs and 1 output. These articles are aimed at the measurement of competitiveness by economic and social characteristics or measuring of the life quality. If we talk about the efficiency of traffic and transportation system it is harder to find a comparison of regions or municipalities because of the lack of data. The studies of traffic accessibility concentrate usually on the small area and they are created under the support of grants as the process of data collection is long and expensive. A big study of the accessibility of the employers in the district Bruntal (Czech Republic) was made under the grant of the Grant Agency of the Czech Republic [6]. Authors analyzed the situation of 302 municipalities on the basis of the public and private transport accessibility, time and kilometers distance from the main employers. For the comparison they used network and statistical analysis. One study using DEA model for evaluation of traffic and transportation system was made in China [11]. This study uses fixed assets investment and quantity of employee as inputs and passengers volume, passengers' transportation volume, freights volume and freight turnover volume as outputs to evaluate the development of traffic and transportation in China's 31 province.

In the beginning of our analysis of the traffic accessibility in the Vysocina region we decided to compare selected areas from the traffic accessibility and unemployment point of view. As there are only 5 districts in Vysocina region it is not enough for the DEA model. But it is not possible to obtain the relevant data for all 704 municipalities and so we analyze the municipalities with extended competence called in Czech ORP (15 in Vysocina region). These municipalities lie between the NUTS IV (LAU 1) – districts (5 in Vysocina region), and NUTS V (LAU 2) – municipalities (704 in Vysocina region). Figure 1 shows the Vysocina region and the areas that belong to each municipality with extended competence (so all ORP's).

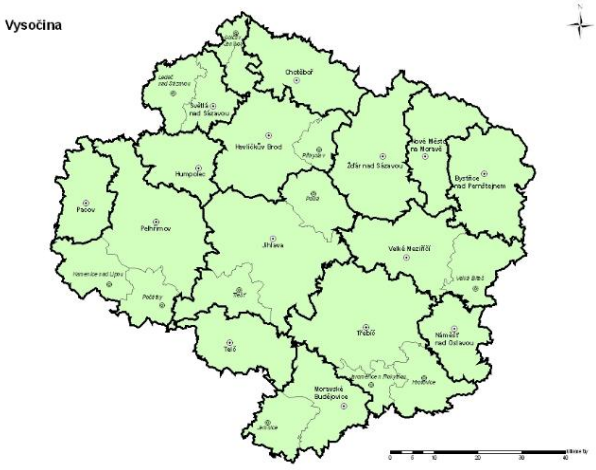


Figure 1 Vysocina region and 15 areas with municipalities with extended competences (ORP = obec s rozšířenou působností) [4]

2 Data and methods

The basic idea of DEA models consists in estimation of an efficient frontier that defines production possibility set of the problem. Based on the set of available decision making units (DMUs) DEA estimates so-called efficient frontier, and projects all DMUs onto this frontier. If a DMU lies on the frontier, it is referred to as an efficient unit, otherwise inefficient. DEA also provides efficiency scores and virtual units for inefficient DMUs. Reference units are hypothetical units on the efficient frontier, which can be regarded as target units for inefficient units. DEA models can be oriented to inputs or outputs. In the case of input oriented models we assume fixed level of outputs (CCR-I), the output oriented model assumes fixed level of inputs and maximize level of outputs with respect to given inputs (CCR-O) [1]. These models are used if we assume constant return to scale. In the case of variable return to scale we work with BCC (Banker, Charnes, Cooper) models. The review and detailed information about DEA models can be found in [2] and [3]. The basic idea for the efficiency calculation is to maximize the rate of weighted sum of outputs divided by weighted sum of inputs. For example the model transformed (Charnes-Cooper transformation) into the linear programming form can be defined as follows (CCR-I):

$$\text{Maximize } z = \sum_{i=1}^{r} u_i y_{iq}$$

$$\text{Subject to: } \sum_{i=1}^{r} u_i y_{ik} \le \sum_{j=1}^{m} v_j x_{jk} \text{ , } k = 1, 2, \ldots, n \qquad (1)$$

$$\sum_{j=1}^{m} v_j x_{jq} = 1$$

$u_i \geq 0$, i=1,2,…,m

$v_j \geq 0$, j=1,2,…,r,

where q represents the evaluated DMU, y_{rj} are known outputs, x_{ij} are known inputs of the jth DMU, u_r and v_j are the variable weights to be determined by the solution of this problem. The efficient unit U_q lies on the efficient frontier in case that the optimal efficiency (calculated by the model) $z = 1$. The inefficient units have z lower than 1 (in CCR-I model) [1].

The aim of DEA is to separate the DMUs into efficient and inefficient ones according to the defined inputs and outputs. DEA models are based on the fact that it is possible to find the efficient frontier formed by combinations of inputs and outputs of some DMUs. The units lying on the frontier are considered as efficient and the remaining ones as inefficient. Their efficiency score is measured as a distance from the efficient frontier [2]. The number of DMUs should be high enough because if we have few units and a lot of inputs and outputs, all units are considered to be efficient. It is not necessary to have a lot of criteria especially when we would like to describe the situation in a graph and show the efficient frontier.

Typical CCR and BCC models suppose it is possible to change inputs and outputs so as the inefficient DMU could be efficient. Sometimes it is necessary to use inputs or outputs that are given and cannot be change. In this situation the models with uncontrollable or nondiscretionary inputs or outputs have to be applied. The model stays the same as in (1) only in the objective function the weighted sum of outputs is decreased by the weighted sum of nondiscretionary inputs [2].

In our case we compare 15 municipalities with extended competence (Figure 1). According to the available data taken from the web pages of Czech Statistical Office [4] and web pages with regional characteristics [10], [8] we have decided to analyze the relationship between the traffic accessibility and the economic activity. The traffic accessibility can be described by many different factors. We have chosen the distance from the town Jihlava (the county seat) from the ORP, the average distance of the municipalities from its ORP (calculated from all the municipalities that belongs to the ORP) and the possibility to reach one of these places. The distances are measured in kilometers or in minutes (as a time to reach the town), the possibility to reach is expressed by the number (or average number) of lines (bus or train together) – it means how many times it is possible to go from a municipality to the ORP or Jihlava (or vice versa) daily. The outputs should characterize the economic activity, so the first one is unemployment and the second one number of all types of businesses in the ORP area. As in the usual DEA model the outputs are supposed to be maximized (and inputs minimized) we have transferred the unemployment percentage into the employment (100 – unemployment percentage). Also the number of lines can be view to be maximized or minimized. Firstly we take into account the regional authority point of view and so it is better to minimize it (because of the cost connected with the public transportation). Afterwards we change this criterion to be output just to maximize it (people point of view). We have thought also to include the number of economically active persons but this criterion is highly correlated (correlation coefficient 0.997) with the number of businesses and so it in not good to have them both in a model. In the last step we have added also the number of applicants for a job as an input. The first model uses these inputs and outputs:

Inputs:

- Distance from Jihlava in kilometers
- Time to reach Jihlava in minutes
- Number of lines to Jihlava
- Average distance to the ORP in kilometers
- Average time to reach ORP in minutes
- Average number of lines to ORP

Outputs:

- Unemployment (change into Employment)
- Number of businesses

As you can see we have 4 inputs that can be characterized as nondiscretionary – these are the distances in kms or time in minutes (because it is hard or impossible to reduce the distance) so we also try the models with nondiscretionary inputs.

3 Results

When we would like to create a graph with the efficient frontier we can use only 1 input and 2 outputs or 2 inputs and 1 output. When we consider both outputs and number of lines to ORP as an input we can see that there are only 2 efficient ORPs – Jihlava and Humpolec. It can be also influenced by the fact that these two ORPs are towns that are very close to the biggest motorway D1 and that is why a lot of buses goes there. On the other hand Velke Mezirici is also close to D1 but it is very far from the efficient frontier. It is because in Jihlava and Humpolec the unemployment rate is smaller (they are industrial centers) than in Velke Mezirici and Bystrice nad Pernstejnem or Trebic where the unemployment rate is higher than 10 %.

As Jihlava is the biggest town of the Vysocina region a lot of people work there. If we compare the relationship between the employment as an output and the distance of the ORP from Jihlava and number of lines to Jihlava as 2 inputs (Figure 2) Jihlava must be efficient but if we erase it then Havlickuv Brod would lie on the efficient frontier with Pelhrimov, Svetla nad Sazavou and Pacov. First two units are district towns and so the number of lines to Jihlava is large, next two units have very low unemployment. Inefficient ORP Telc has higher unemployment rate but also very high number of lines to Jihlava in comparison with the distance. It is because this town is registered on the UNESCO's List of World Cultural Heritage sites. As we cannot change the distance, the only possibility how can Telc became efficient is to increase employment (or decrease unemployment). It is clear that this recommendation is easy to say but hard to put into practice. The result is not influenced by other criteria that are important (such as number of businesses, number of free working places, wages, .. – but some of them are unavailable for the municipalities). The next unefficient Namest nad Oslavou has the highest unemployment rate in Vysocina region (14,1 %).

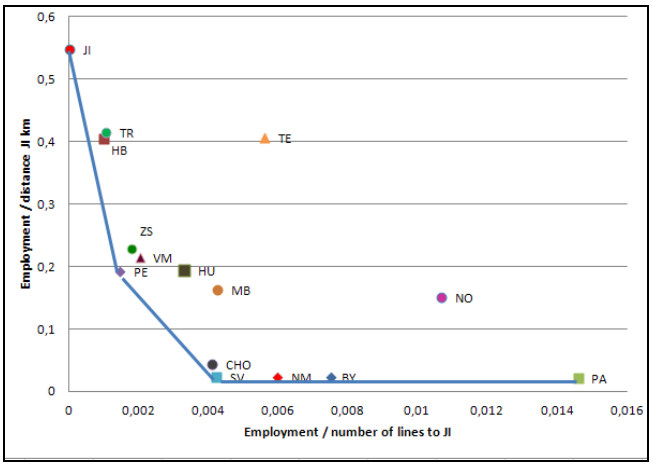


Figure 2 Efficient frontier for 2 inputs and 1 output

For the complete results we have used DEA Excel Solver [7]. We have mentioned above that 4 of our 6 inputs are nondiscretionary and that is why we have calculated not only CCR and BCC models but also its alternative with this specifications (nondiscretionary CCR model = ND-CCR, nondiscretionary BCC model = ND-BCC). The first two models were calculated as output oriented because outputs can be better changed than inputs, the second two models are input oriented. All the results are in the Table 1. From the 15 ORPs nearly half of them is efficient in all models – Humpolec, Chotebor, Jihlava, Pacov, Pelhrimov, Svetla nad Sazavou and Velke Mezirici. In all these municipality areas the unemployment rate is lower than 9.9 % whereas from the inefficient ORPs only Havlickuv Brod and Zdar nad Sazavou has the unemployment rate lower than 10 %.

These results take into account the regional authority point of view (with minimization number of lines). As we have calculated output oriented model the Table 2 shows possible changes of 2 outputs for the inefficient DMUs to be efficient. It has been mentioned that in reality it is not easy to decrease the unemployment rate but we can see that in the areas of Moravske Budejovice, Namest nad Oslavou and Trebic the unemployment should decrease for more than 4 % which is today unreal. The extreme change of number of businesses should happen in Namest nad Oslavou to be efficient (or rapidly decrease the number of lines to Jihlava) – so we see that this ORP is the most inefficient.

The number of lines can be seen also from the usual people's point of view – so we have tried to change the model so as the number of lines has started to be two new outputs. In this case BCC output oriented model has figured out only 4 inefficient units: Bystrice nad Pernstejnem, Chotebor, Nove Mesto na Morave and Zdar nad Sazavou. CCR output oriented model adds Havlickuv Brod, Pelhrimov and Trebic. Table 3 describes the possi-

ble increase of the number of lines for the DMUs to be efficient. It is clear that for the places with low number of lines to Jihlava the increase should be very large but it is because of the other criteria not only because of the low number (we put into the table also Pacov which is efficient with low number of lines). This analysis tells us that there are some localities far from Jihlava (Bystrice nad Pernstejnem, Nove Mesto na Morave, Chotebor) where it is better to increase the number of lines to its ORP than to Jihlava (as it is more expensive) or try to increase number of businesses (and number of entrepreneurs) to be comparable with other localities in Vysocina region. As a last step we have added the number of applicants for a job as a new input but the results has stayed nearly the same as in the previous analysis.

DMUs (ORP) / Model	CCR-O	BCC-O	ND-CCR	ND-BCC
Bystrice nad Pernstejnem (BY)	0.96906	0.96906	0.96906	1
Havlickuv Brod (HB)	0.98481	0.99148	0.98481	0.98634
Humpolec (HU)	1	1	1	1
Chotebor (CHO)	1	1	1	1
Jihlava (JI)	1	1	1	1
Moravske Budejovice (MB)	0.89364	0.92677	0.64587	0.69496
Namest nad Oslavou (NO)	0.95292	0.95292	0.85447	0.90909
Nove Mesto na Morave (NM)	0.98320	0.98320	0.98320	1
Pacov (PA)	1	1	1	1
Pelhrimov (PE)	1	1	1	1
Svetla nad Sazavou (SV)	1	1	1	1
Telc (TE)	0,99648	1	0.99352	1
Trebic (TR)	0.94032	0.94536	0.92310	0.92364
Velke Mezirici (VM)	1	1	1	1
Zdar nad Sazavou (ZS)	0.9623	0.99084	0.9623	0.97344

Table 1 Results of the DEA models using 6 inputs and 2 outputs.

Outputs / DMUs	BY	HB	MB	NO	NM	TE	TR	ZS
Real unemployment	12.24	8.57	13.88	14.1	10.92	11.27	13.1	8.39
Virtual unemployment	9.44	7.79	7.08	9.85	9.39	11.27	8.06	7.55
% change	2.8	0.78	6.8	4.25	1.53	0	5.04	0.84
Real Number of businesses	3730	10339	4423	2500	3764	2759	15448	9295
Virtual number of businesses	3849	10428	4773	4084	3828	2759	16341	9381
% change	3.19	0.86	7.9	63.36	1.71	0	5.78	0.92

Table 2 Possible changes of outputs according to CCR-O model

Outputs / DMUs	BY	CHO	NM	PA	ZS
Real number of lines to Jihlava	2	4	2	2	21
Virtual number of lines to Jihlava	16.9	15.8	7.8	2	21.4
% change	745	295	290	0	1.9
Real number of lines to ORP	14	10	11	12	12
Virtual number of lines to ORP	14,2	10,2	12.4	12	13
% change	1.4	2	12.7	0	8.33

Table 3 Possible changes of outputs according to CCR-O model with number of lines as output

This analysis can be influence by the fact that we have a lot of inputs and some of them can be correlated. The correlation matrix showed us that the highest correlation (90%) is between the distance to Jihlava and time to reach Jihlava (which can be expected) but the high correlation is also between the number of lines to ORP and time to reach ORP (86%) which is interesting. So we have tested some other models with lower number of inputs and outputs. If we take into account the municipality point of view that would like to minimize number of

lines (to Jihlava and ORP, 2 inputs) and maximize employment and number of businesses (2 outputs) the results are nearly the same like in first model only Velke Mezirici starts to be inefficient. So we can say that in this model the results and number of efficient units are not influenced so much by correlated inputs. When we change the point of view to respect the inhabitants requirements (so the number of lines starts to be 2 outputs that should be maximize and we have taken 3 inputs: unemployment, time to reach Jihlava and time to reach ORP that should be minimize), we have different results in CCR and BCC models. CCR (input and output) models are more strict and showed only 4 efficient areas – Jihlava, Moravske Budejovice, Pacov, Pelhrimov, whilst the BCC models has had 8 efficient – it has added 4 other areas (Humpolec, Namest nad Oslavou, Svetla nad Sazavou, Telc). When we compare results of all models we can see that 4 areas are always inefficient: Bystrice nad Pernstejnem, Nove Mesto na Morave, Trebic, Zdar nad Sazavou. From the inhabitants point of view also Chotebor belongs to inefficient areas whereas from the municipality perspective Havlickuv Brod, Moravske Budejovice, Namest nad Oslavou and Telc do not belong to efficient units.

4 Conclusion

The aim of this contribution was to compare the relative relationship between the factors connected with the traffic accessibility of the selected municipalities and its economic activity. According to the results from DEA models we can say that in Vysocina region there exist efficient municipalities with extended competences (especially those near to D1 highway) and inefficient ones where the situation could be better by changing some parameters. One possibility is to decrease the number of lines to Jihlava in the far-away ORPs and increase the number of lines to the center of its ORP to increase the business activity and decrease the unemployment rate. The other part shows that for some areas it would be better to increase the number of lines to Jihlava and ORP for the same effect. We can also see that there are some municipalities which are always inefficient no matter what point of view is taken into account. The results can be used by the Regional authority in the process of employment subsidy allocation and for the support of expansion by new businesses in inefficient areas. The research will continue when more information will be available.

Acknowledgements

The research was supported partly by the Grant Agency of the Czech Republic GAČR No. P403/12/1387, by the Internal Grant Agency of the University of Economics IGA F4/11/2013 "Optimalizační modely v ekonomickém rozhodování" and also by the Internal Grant Agency of the College of Polytechnics Jihlava – grant no. 2013005 "Vliv dopravní infrastruktury a vzdálenosti od centra kraje na ekonomickou a sociální úroveň okresů a ORP Kraje Vysočina."

References

[1] Charnes A., Cooper W.W., Rhodes E.: Measuring the efficiency of decision making units. *European Journal of Operational Research* **2** (1978), 429–444.
[2] Cooper, W. W., Lawrence, M.S., Zhu, J.: *Handbook on Data Envelopment Analysis*. Kluwer Academic Publishers, Norwell, 2004
[3] Cooper, W.W., Seiford, L.M., Tone, K.: *Introduction to Data Envelopment Analysis and Its Uses*. Springer Science + Business Media Inc., New York, 2006.
[4] Czech Statistical Office: Regions and districts. [online], available at http://www.czso.cz/x/krajedata.nsf/krajenejnovejsi/xj [cit. 2013-05-18]
[5] Friebel, L., Friebelova, J.: Quantitative evaluation of life quality of Czech districts. *Proceedings of 30th* International Conference Mathematical Methods in Economics, Karvina (2012), 190-195.
[6] Horak, J. et al: Dopravní dostupnost významných zaměstnavatelů v okrese Bruntál. [online], available at *gis*.vsb.cz/gacr_mtp/Clanky/**dostupnost**BR_X2.pdf [cit. 2013-05-20]
[7] Jablonsky, J.: Software Support for Multiple Criteria Decision Making Problems. *Management Information Systems* Vol. 4 (2009), 29-34
[8] Kraj Vysocina: Dopravni plan. [online], available at http://www.kr-vysocina.cz/dopravni-plan/d-4040364/p1=7232 [cit. 2013-05-20]
[9] Melecky, L., Stanickova, M.: Assessment of efficiency in Visegrad countries and regions using DEA models. *Central European Review of Economic Issues* **15** (2012), 145-156
[10] Regionální informační servis: *Správní obvody obcí s rozšířenou působností*. [online], available at http://www.risy.cz/cs/krajske-ris/kraj-vysocina/obce-s-rozsir-pusobnosti/ [cit. 2013-05-18]
[11] Yu, W., Chen, S., Yang, Y.: Evaluation on Overall Relative Efficiency of China's Traffic and Transportation System Based on Data Envelopment Analysis. IEEE International Conference on Green Systems and Intelligent Services, Nanjing (2007), 303-307.

Markets, social networks, endogenous preferences, and opinion leaders[1]

Michal Kvasničkа[2]

Abstract. This paper studies the impact of opinion leaders ("stars") and their fans on equilibrium market prices within the Bell's model (JEBO 2002). The simulation shows that 1) the model may not converge when the opinion leader consumes the good that is extremely scarce—it can create infinite cycles in her fans' preferences; 2) the preferences may not be completely polarized in the same situation—the agents with non-polarized preferences prevent the cycles; 3) while the agents in the Bell's model consume only the more abundant good when the other good is extremely scarce, the presence of the opinion leader eliminates this when she consumes the scarce good, and 4) the presence of the opinion leader and her fans can sometimes surprisingly lower the price of the good that the opinion leader consumes.

Keywords: endogenous preferences, market, social network, opinion leaders, agent-based simulation

JEL classification: D83, D51, D85
AMS classification: 68U20, 91D30, 91B69

1 Introduction

The seminal agent-based study on endogenous preferences and how they are influenced by interactions within a social network is A. M. Bell's paper "Locally interdependent preferences in a general equilibrium environment" [1] which explored the endogenous preference adaptation for a grid social network. The robustness of her conclusions was later tested by Kvasnička [2] who showed that some of Bell's original conclusions do not apply for other symmetric network structures. The present paper enhances the Bell's model in another way: it studies the impact of an opinion leader ("a star") on human preferences within the framework of Bells model. It introduces a new kind of agent ("an opinion leader" or "a star") and a second social network, the asymmetric star network of her fans. By means of agent-based simulations it investigates how the properties of the fans network (number of agents in the network, their concentration, and the strength of the opinion leader's impact on one fan) affect the structure of the model equilibrium and the relative price of the two goods. It is the first step toward a quantitative modeling of an important marketing problem how opinion-leaders (e.g. pop stars) can be used to enhance the demand for one of competing products in environments where fashion matters.

2 Model

The model enhances Bell's model of "exchange economy", see [1]. There are two kinds of agents: "ordinary people" and one "opinion leader". All ordinary agents consume two comparable kinds of goods (e.g. white and black t-shirts). The opinion leader consumes only good 1 (the white t-shirts). In every period, all agents get an initial endowment of each good which they exchange with each other in the centralized market at the market clearing price. The ordinary agents' preferences evolve over time: each ordinary agent increases her preference for the good that has been recently more popular (i.e. more consumed) in her neighborhood. Her neighborhood consists of three parts: 1) the agent herself, 2) her eight closest agents, and 3) the opinion leader, if the agent is her fan. The opinion leader's preferences do not change in time.

More formally, there are N agents: one opinion leader indexed $i = 1$ and $N - 1$ ordinary agents indexed $i = 2, \ldots, N$. In every period, each agent gets the same endowment: e_1 units of good 1 and e_2 units of good 2 ($e_1, e_2 > 0$,

[1]This paper has been created as a part of project of specific research no. MUNI/A/0797/2012 at the Masaryk University.
[2]Masaryk University, Faculty of Economics, Department of Economics, Lipová 41a, Brno, 602 00, the Czech Republic, michal.kvasnicka@econ.muni.cz

and $e_1 + e_2 = 100$). Agent i then demands x_{i1} units of good 1 and x_{i2} units of good 2 to maximize her one-period Cobb-Douglas utility function subject to the constraint given by her endowment, i.e.

$$\max_{x_{i1},x_{i2}} x_{i1}^{a_{it}} x_{i2}^{1-a_{it}} \qquad \text{s.t.} \quad p_1 x_{i1} + p_2 x_{i2} = p_1 e_1 + p_2 e_2, \tag{1}$$

where p_1 and p_2 are the prices of good 1 and good 2 respectively, and a_{it} is agent i's relative preference for good 1 at time t. The initial value of the preference parameter a_{i0} is drawn independently for each ordinary agent $i > 1$ from the continuous uniform distribution $U(0, 1)$; the opinion leader's preference $a_{1t} = 1$ for each t.

Agent i's demand for the two goods is then

$$x_{i1}(p_1, p_2) = a_{it}\left(e_1 + \frac{p_2}{p_1}e_2\right), \qquad x_{i2}(p_1, p_2) = (1 - a_{it})\frac{p_1}{p_2}\left(e_1 + \frac{p_2}{p_1}e_2\right). \tag{2}$$

Since the total endowment is given, the market clearing relative price of good 1 in terms of good 2 is

$$\frac{p_1^*}{p_2^*} = \frac{e_2 \sum_{\forall j} a_{jt}}{e_1 \sum_{\forall j}(1 - a_{jt})}. \tag{3}$$

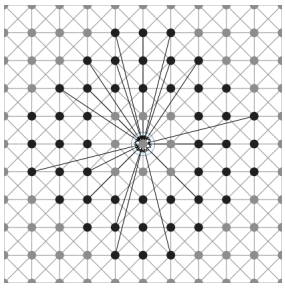
The relationships among the agents are defined through two kinds of social networks: one network defining friends, and the other defining fans of the opinion leader. Each social network is represented by a graph, in which agents are vertices and their relationships are edges (connections). Agent i has relationship to agent j in graph G if i is connected with an edge to j; we then write $i \sim j \in G$. Both social networks are created independently for each simulation and are fixed within the simulation. For an example see Figure 1.

The first social network portrays friendship, and is equivalent to the Bell's original social network, see [1]. It is represented by an undirected graph G_1, which is a grid network on a toroid; see [6] for its definition. (Since the graph G_1 is undirected, the friendship relationship is symmetric, i.e. $i \sim j \in G_1 \Leftrightarrow j \sim i \in G_1$.) One can see the network as a cellular automaton on a lattice where each agent is represented by a cell, the set of the agent's friends consists of the cell's Moore neighborhood, and the lattice edges are connected to each other in such a way that an agent at the very top of the lattice is at the same time located also at the very bottom of the lattice, and so on. Let us define index function $n(i) = \{j : j = i \vee j \sim i \in G_1\}$ as the set of indices of agent i's friends and the index of the agent i herself.

The second social network portrays the relationships of the fans to the opinion leader. This network is represented by a directed star graph G_2; see [6] for its definition. In this network, the fans are connected to the opinion leader in such a way that each fan cares about the consumption of the opinion leader but not about the consumption of the other fans; the opinion leader does not care about the consumption of her fans. (The relation in directed networks is neither reflexive, nor symmetric, i.e. $1 \sim 1 \notin G_2$, and $i \sim j \in G_2$ does not imply $j \sim i \in G_2$.) The fans are selected from among a subset A of the ordinary agents that are not the opinion leader's friends. Let $\rho = |A|/(N - 9)$ denote the share of the agents that can be selected as opinion leader's fans on the total population of all ordinary agents except the opinion leader's friends. Let $\pi \leq \rho$ denote the share of the actual fans on the total population of the ordinary agents that are not the opinion leader's friends. The network G_2 is constructed this way: 1) construct the set A that contains indexes of all ordinary agents that are not the opinion leader's friends that are located within the smallest circle around the opinion leader that includes at least $\rho(N - 9) + 9$ agents (I assume that the agents are distributed equidistantly in the friendship space, as on the lattice); 2) construct the set B as $\pi(N - 9)$ randomly chosen indexes from the set A, and 3) connect the agents in B to the opinion leader in G_2, i.e. set $j \sim 1 \in G_2$ for $\forall j \in B$. Notice that the parameter ρ together with π determines how concentrated within the whole population the fans are. For instance, $\pi = 0.1$ and $\rho = 1$ means that there are few fans and they are sparsely scattered within the whole population of agents. On the other hand, $\pi = 0.3$ and $\rho = 0.3$ means that all the fans are concentrated around the opinion leader—the 30 % of opinion leader's closest non-friend agents are her fan, and there are no fans farther from the opinion leader. Let us define Boolean function f such that $f(i) = 1$ if $i \sim 1 \in G_2$, and $f(i) = 0$ otherwise.

Agent $i > 1$'s neighborhood thus consists of the agent herself, her friends (i.e. the agents with indexes in $n(i)$), and the opinion leader if the agent is her friend (i.e. $f(i) = 1$). After observing the consumption in her neighborhood, each ordinary agent $i > 1$ adjusts her preferences in such a way that she increases the preference for the good that is consumed more in her neighborhood. Specifically, agent $i > 1$ sets the future value of her preference parameter $a_{i,t+1}$ at

$$a_{i,t+1} = a_{it} + r\left(\frac{\sum_{j\in n(i)} x_{j1} + sf(i)x_{11}}{\sum_{j\in n(i)} x_{j1} + sf(i)x_{11} + \sum_{j\in n(i)} x_{j2} + sf(i)x_{12}} - 0.5\right) \quad \text{for } i > 1, \tag{4}$$



The dots denote the agents, the lines denote the edges of the social networks. The agents on the very top wrap over to the very bottom and the agents at the right edge wrap over to the left edge.
The encircled dot in the middle of the figure is the opinion leader.
The black dots are the agents that can be selected as fans, i.e. agents with indexes $i \in A$.
The short gray vertical, horizontal, and diagonal lines are the vertices of the friends network G_1.
The long black lines are the vertices of the fans network G_2.

Figure 1 An example of the social networks for $N = 100$, $\pi = 0.2$, and $\rho = 0.5$.

where the adjustment parameter $r \in (0, 1)$ regulates the speed of the preference adjustment and s is the strength of impact of the opinion leader's consumption on a fan. Setting $s > 1$ allows us to study the cases where the fans value the consumption of the opinion leader higher then the consumption of their friends.

The evolution of agents' preferences and consumption is simulated in the agent-based computational fashion (for introduction to it, see e.g. [4]). First, the model is initialized: N agents are created, assigned their random initial preference $a_{i,0}$ drawn from $U(0, 1)$, and organized in the social networks G_1; then one agent is selected as the opinion leader and her preference is changed to $a_{1,0} = 1$; then the social network G_2 is created. The simulation than proceeds in discrete steps repeated until the model converges (i.e. the agents' preferences change no more), or the maximal amount of steps is reached. In each step, 1) the market clearing relative price p_1^*/p_2^* is calculated for the agents' current preferences (equation 3), 2) each agent's equilibrium consumption is calculated (equation 2), and 3) each agent $i > 1$'s preference is adjusted (equation 4).

3 Results of simulations

The model has been simulated for $N = 100, 900$, and 2 500 agents, the concentration of fans $\rho = 0.1, 0.2, \ldots, 1$, the share of fans $\pi = 0.1, 0.2, \ldots, 0.7$ (only combinations with $\pi \leq \rho$ were simulated), and the strength of opinion leader's impact $s = 1, 2, 3$. The adjustment constant r was set to 0.5 as in [1]. Three variants of the model were simulated for each random seed: 1) the "grid" model, i.e. Bell's original model without the opinion leader and her fans where agent 1 was treated as an ordinary agent, 2) the "stubborn" model, i.e. an intermediate model where agent 1 had the preference parameter $a_{1t} = 1$ at each t but had no fans (i.e. G_2 was not present), and 3) the "leader" model, i.e. the full model described above. The three variants can be straightforwardly compared because only the initialization is stochastic (the rest of the simulation is deterministic), and the order of steps in the initialization secures that the common part of each two variants is the same for each initial random seed. The model was simulated one hundred times for each feasible combination of parameters and each variant. The total number of the simulation runs was 573 300 triples. The maximal amount of simulation steps was set to 5 000. The model was simulated in NetLogo 5.0.4 [5] and the results were analyzed in R [3]. The web interface of the model is available at http://www.econ.muni.cz/~qasar/english/models.html.

3.1 Simulation outcomes

Bell [1] discusses two kinds of simulation outcomes on the grid friends network G_1. In both, the system converges, each good is consumed by some agents, and fewer agents consume the scarcer good than the more abundant one. The difference between the outcomes lies in the agents' equilibrium preferences. In the first case, each agent specializes in consumption of only one kind of good (i.e. $a_{it} \in \{0, 1\}$, which Bell calls "polarized" preferences), while in the second case, at least some agents consume both goods (i.e. have $a_{it} \in (0, 1)$). Bell claims that the second outcome is unstable, and hence cannot occur in a simulation [1, p. 321]. In the polarized state, the agents with the same preference are clustered together (for an example of the clusters, see Figure 3, panel a). The clusters arise because no ordinary agent can keep its polarized preference unless she is surrounded by a sufficient number of other agents with the same preference. The reason why more agents specialize in consumption of the more abundant good is the negative feedback provided by the market: when the number of agents consuming a good is out of proportion to the good endowment, the relative price of the goods is unequal; the consumers of the cheaper good consume more, which motivates some consumers of the other good to change their preferences. In the limit, this market feedback presses the relative price of the goods to unity. However, there is also the "bandwagon" effect: the higher the proportion of consumers of one good, the higher the probability that an agent is surrounded

by the consumers of this good, and hence that she switches her preference to this good. The bandwagon effect rises demand for the more abundant good, and hence presses the relative price of the goods away from unity. See [2] for a more detailed analysis.

Kvasnička [2] showed that there may be a richer set of outcomes with a general social network: 1) the simulation may not converge, 2) it may converge but some agents can remain non-polarized, 3) the simulation can converge, all agents can be polarized but all of them consume only the more abundant good, and 4) the simulation converges, all agents are polarized and both kinds of good are consumed by some agents, which is the Bell's standard outcome. All these four states occurred also in the present simulation, though the first three cases occurred only with extreme proportions of the endowment: either $e_1 \geq 95$ and $e_2 \leq 5$, or $e_1 \leq 5$ and $e_2 \geq 95$. The standard logistic regression was not suitable here for the detection of the combinations of treatments for which these outcomes were realized because the model behavior is highly non-linear; decision trees constructed by R packages rpart, rpart.plot and Rattle [7] were used instead.

The state 1, the non-convergence, is caused by fans network G_2—the grid and stubborn models always converge. The model with the opinion leader did not converge in 22 737 out of 573 300 runs (i.e. about 3.97 % of runs). Most non-convergent runs happened when $e_2 = 99$ (i.e. $e_1 = 1$), $N = 900$ or 2 500, and $\pi \geq 30$ %. For details, see the decision tree in Figure 2. When the model does not converge, it is because it cycles. Let us show the mechanism on an example. Assume the model with $N = 2500$ agents, endowments $e_1 = 1$ and $e_2 = 99$, $\pi = 0.2$, $\rho = 0.3$ and $s = 2$. Because of the bandwagon effect, the model soon reaches the state where only the opinion leader consumes the scarce good 1 and all other agents consume only the abundant good 2. The opinion leader then consumes 2 500 units of good 1, while each other agent consumes only about 99 units of good 2. Thus according to equation (4), all fans increase their taste for good 1 in the next step. This rises the demand for good 1, which rises its relative price, and the consumption of all agents who consume good 1 declines. In the following step, the fans' preferences for the good drop according to equation (4) again. The opinion leader is again the only agent consuming the scarce good 1, and the new cycle begins.

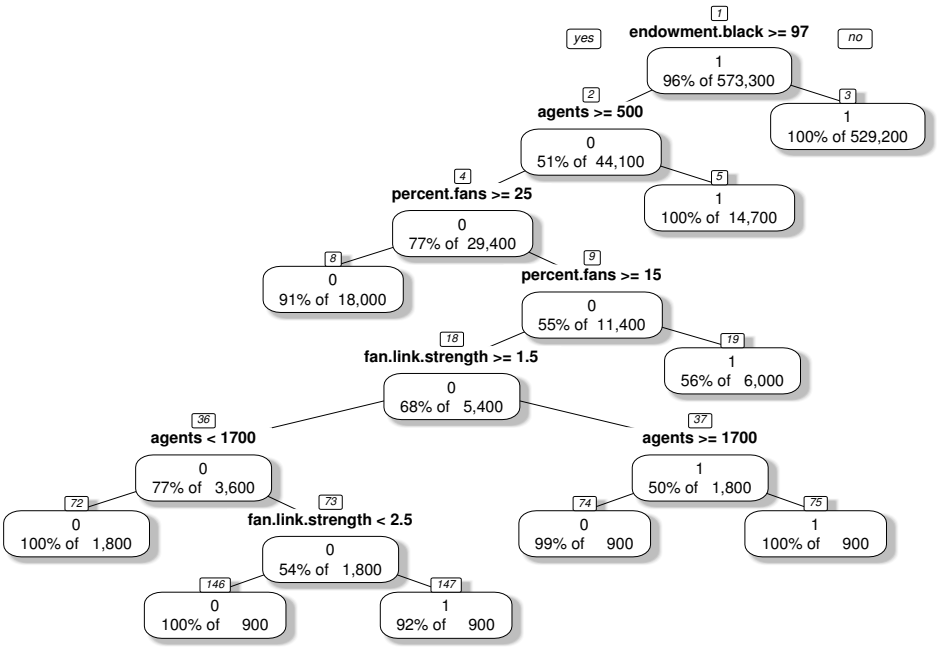


Figure 2 Decision tree showing when the leader model converged. Convergence is denoted by 1, N is denoted as agents, e_2 as endowment.black, π as percent.fans, and s as fan.link.strength.

The state 2, where the model converges but some agents have non-polarized preferences, occurs in 4 930 runs out of the 550 563 converged runs (i.e. 0.9 % of runs). It only occurs in the leader model when $N = 100$, $e_1 = 5$ and $e_2 = 95$, and $s \geq 2$. Usually, there were one to four non-polarized agents (the fans) that advanced the relative price of the scare good 1 in such a way that the opinion leader's consumption of good 1 was low enough, so that the other fans were not tempted to change their preferences. This balancing problem obviously becomes more delicate when there are more agents, and the state was not observed in data for higher N.

The state 3, where the model converges, all agents' preference are polarized but only the more abundant good is consumed because of the bandwagon effect, occurred in the grid model when the endowment was extremely unequal. It always occurred in the grid model when $e_j = 99$ (and $e_{-j} = 1$), and often also when $e_j = 95$ and $N = 100$. Good 1 was solely consumed in 31 650 and good 2 in 58 653 runs out of 545 633 converged polarized runs totally (together 16.6 % of the converged polarized runs). The presence of a stubborn consumer of good 1 (whether she had a fans network or not) eliminated this state when the scarce good was good 1—it was at least consumed by agent 1 herself. When the scarce good was good 2, then the presence of a stubborn consumer of good 1 slightly rose the number of runs when this happened: the state occurred in 58 946 runs in the stubborn model and in 59 015 runs in the leader model.

3.2 Opinion leader's impact on relative price

To study the impact of the opinion leader' presence on the relative price of good 1, we will confine to the converged polarized runs when both goods were consumed in positive quantities. It may seem intuitive that the presence of the opinion leader that always consumes good 1 and her fans that can follow her in her consumption should rise the demand for good 1, and hence its price. However, it is not this simple. In fact, the inclusion of the opinion leader and her fans increased the relative price of good 1 in 360 910 runs out of 454 703 runs in view (i.e. in 79.4 % runs), did not change it in 35 973 (7.9 %) runs, and decreased it in 57 820 (12.7 %) runs. It is instructive to decompose the total effect to the impact of the opinion leader herself (the change from the grid to the corresponding stubborn model) and the impact of the fans network (the change from the stubborn model to the corresponding leader model). The detailed decomposition is presented in Table 1.

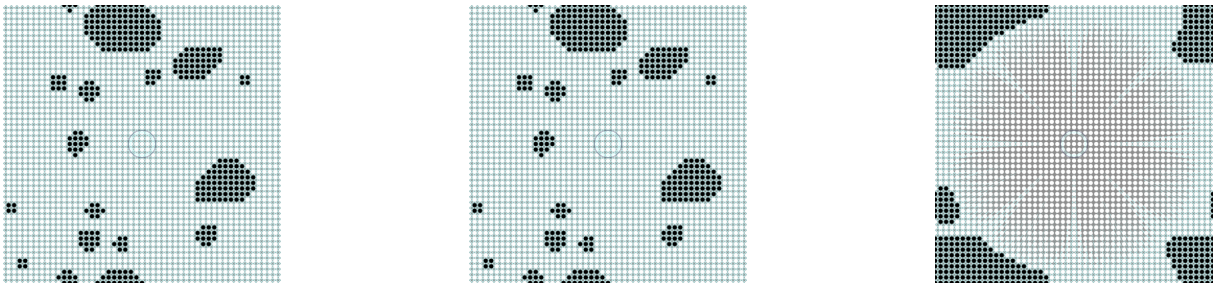
	total rise			total no change			total decline		
	decline	no change	rise	decline	no change	rise	decline	no change	rise
decline	0	0	94 343	0	0	4 988	7 842	3 298	6 189
no change	0	0	123 588	0	27 790	0	31 465	0	0
rise	10 477	8 885	123 617	3 195	0	0	9 026	0	0

Table 1 Number of converged polarized runs where both goods were consumed decomposed. Each block corresponds to a total change in the relative price; e.g. the block denoted "total rise" includes the runs where the relative price of good 1 rose with transition from the grid model to the leader model. Within the blocks, the rows correspond to a change of the relative price when one agent was assigned the stubborn preference for good 1 (the transition from the grid to the stubborn model). The columns correspond a change of the relative price when the stubborn agent was given the fans network G_2 (the transition from the stubborn to the leader model). Each number is the number of the corresponding runs; e.g. 4 988 is the number of runs where the relative price declined when the stubborn consumer of good 1 was introduced, it rose again when the stubborn agent was given a network of fans, and the two effects had precisely the same magnitude with the opposite sign, so that the relative price was the same in the leader model as in the corresponding grid model. (Note that the relative price can take only values from a discrete set because of equation (3), polarized preferences, and discrete number of agents.

When one agent in the grid network was assigned the stubborn preference for good 1 (i.e. the grid model was changed into the stubborn model), the relative price of good 1 rose in 34 % runs, did not change in 40 % runs, and decreased in 26 % runs. The reason is that when one agent in the grid network is given a stubborn preference for good 1, the clusters of consumers may evolve differently, and the number of the consumers of each good can randomly rise, decline, or stay the same (as the transition from panel a to panel b in Figure 3). The relative price of good 1 declines more often when the number of agents is high and the endowment e_2 is high. For instance, the relative price of good 1 declines in 53 % runs when one stubborn consumer of good 1 is added to the grid network of $N = 2500$ agents with endowment $e_2 = 90$, but it declines in only 9 % runs with $N = 100$ and $e_2 = 10$.

When the fans network G_2 was added (i.e. the stubborn model was changed into the leader model), the relative price of good 2 rose in 77 % runs, did not change in 9 % runs, and declined in 14 % runs. The relative price declined more often when the endowment of good 1 was higher then the endowment of good 2 (72 % of this relative price declines happened when $e_2 \leq 40$), and when the number of agents was high (80 % of the relative price declines happened when $N \geq 900$). The reason is again that the clustering evolves differently when the fans network G_2 is added. As the clusters evolve differently, the number of the consumers of each good can randomly rise, decline, or stay the same, and the relative price changes accordingly.

The most interesting case of the decline of the relative price of good 1 caused by introduction of the fans



(a) grid model (b) stubborn model (c) leader model

Figure 3 An example of the outcome of the simulation where the relative price of good 1 consumed by the opinion leader decreased when because of the network G_2 of her fans. Here $N = 2500$, $e_2 = 20$, $\pi = \rho = 0.7$, and $s = 3$. The opinion leader or corresponding agent has been moved to the center and denoted by the circle. The black dots correspond to the agents with preference $a_{it} = 0$, the white dots to the agents with preference $a_{it} = 1$.

network G_2 happens when the opinion leader's fans are sufficiently concentrated, i.e. π/ρ is close to unity. An example can be seen in Figure 3. The introduction of the fans network G_2 (the transition from the panel b to the panel c) cleanses the surrounding of the opinion leader, and creates a great cluster of consumers of good 1 which rises the demand for good 1, and hence its relative price. The side effect is that consumers of good 2 often create a smaller number of bigger clusters, which allows more agents to keep the preference for good 2. This second effect lowers the relative price of goods 1. Quite often the second effect is stronger than the former one, and the relative price of good 1 declines.

4 Conclusions

The simulation shows that the presence of an opinion leader and a network of her fans within Bell's model modifies Bell's results: 1) The model may not converge because the presence of the opinion leader that stubbornly consumes a good that is extremely scarce can create infinite cycles in her fans' preferences. 2) The preferences need not be completely polarized in the same situation—the agents with non-polarized preferences can eliminate the cycles. 3) While the agents in the Bell's model consume only the more abundant good when the other good is extremely scarce, the presence of the opinion leader can eliminate this effect when she consumes the scarce good, or slightly enhance it when she consumes the abundant good. 4) The most surprising effect is that the presence of the opinion leader and her fans can sometimes lower the demand, and hence the price of the good that the opinion leader consumes. This suggests that demand manipulation through opinion-leaders may be risky even when the leader succeeds in rising demand of her fans.

References

[1] Bell, A. M.: Locally interdependent preferences in a general equilibrium environment, *Journal of Economic Behavior & Organization* **47** (2002), 309–333.

[2] Kvasnička, M.: Markets, Social Networks, and Endogenous Preferences. In: *Proceedings of 30th International Conference Mathematical Methods in Economics*, vol. 2, (Ramk J. and Stavrek, D., eds.), Silesian University, 2012.

[3] R Development Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Wien, 2011, http://www.R-project.org/.

[4] Tesfatsion, L.: Agent-Based Computational Economics: A Constructive Approach to Economic Theory. In: *Handbook of Computational Economics*, vol. 2, (L. Tesfatsion and K. L. Judd, eds.), Elsevier, 2006, 831–880.

[5] Wilensky, U. (1999). *NetLogo*. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL, http://ccl.northwestern.edu/netlogo/.

[6] Wilhite, A.: Economic activity on fixed networks. In: *Handbook of Computational Economics*, vol. 2, (L. Tesfatsion and K. L. Judd, eds.), Elsevier, 2006, 1013–1045.

[7] Williams, G.: *Data Mining with Rattle and R*, Springer, 2011.

Payoff size variation problem in simple reinforcement learning algorithms[1]

Michal Kvasničk[2]

Abstract. This paper shows that the speed of the reinforcement learning depends on the size of the payoffs, at least when all payoffs are positive. When the speed of learning is too fast, the agents tend to learn to play the actions which they randomly chosen in the first rounds of the learning process. The compositions of the agents' strategies then on the aggregate level resembles the initial individual agent's mixed strategy. This may create artificial effects in the simulations where the size of payoffs depend on the model treatments because the speed of learning cannot be tuned in.

Keywords: reinforcement learning, agent-based simulation, economic experiments, voluntary provision of public goods

JEL classification: C92, D83, H41
AMS classification: 68U, 68W, 91B, 91E

1 Introduction

Standard economic models based on optimizing omniscient agents and instantly attained equilibria are not able to explain many observed phenomena. For instance, they often fail to explain the outcomes of laboratory experiments with human subjects. That is why there are attempts to create alternative formal explanatory models. One approach to this goal is Agent-based computational economics (for a general overview see [14]). The ACE models populated with boundedly rational agents are able to predict not only the eventual equilibrium but also the adjusting process leading to it. Their structure makes them especially useful for modeling the behavior observed in the experiments, see [6]. Since the agents in these models are only boundedly rational they usually have to learn how to act from the feedback provided by their model environment. Thus the learning process constitutes an important part of the models (for a general overview of the learning algorithms used in the ACE model see [4]).

In this paper, we will explore one overlooked property of one of the most often used learning algorithms: the simple reinforcement learning (we use the adjective "simple" to distinguish the reinforcement learning algorithms used in ACE from the more complex algorithms used in the field of artificial intelligence, see [12].) We will claim that the speed of the simple reinforcement learning depends on the size of the payoffs, at least when all payoffs are positive. When the speed of learning is too fast, the agents tend to learn to play the actions which they randomly chosen in the first rounds of the learning process. The compositions of the agents' strategies then on the aggregate level resembles the initial individual agent's mixed strategy. This is no problem in most simulations because the modeler has tools to tune in the speed of learning. However, this may create strange effects in the simulations where the size of payoffs depends on the model treatments—then the speed of learning varies with the treatment, and hence cannot be fine-tuned.

Several parts of the claim has been previously known. Sutton and Barto claim without any proof that the proper setting of the algorithm depends on the size of the payoffs [12, p. 31] and that the eventual mixed strategy may be biased by its prior [12, p. 35]. Fudenberg and Levine claim (also without any proof or elaboration) that there is a positive probability that the algorithm converges to a state "where an inferior choice is played with probability 1" [7, p. 73]. The closest to this paper came Bell [3] who inquired the influence of the initial prior on the algorithm's outcomes with a constant payoff. As far as we know, the influence of the variable payoff size on the outcomes of the algorithm has never been explored.

The rest of the paper proceeds like this: The second section describes the reinforcement learning algorithm and shows the claim analytically in the simplest possible case of one agent playing against a deterministic automaton.

[1]This paper has been created as a part of project of specific research no. MUNI/A/0797/2012 at the Masaryk University.
[2]Masaryk University, Faculty of Economics, Department of Economics, Lipová 41a, Brno, 602 00, the Czech Republic, michal.kvasnicka@econ.muni.cz

The third section provides a more complex example: a simulated version of the provision of public goods experiment where the treatment is the number of agents. It shows that the agents learn to contribute more when the number of agents in the game is higher. The fourth section compares the reinforcement learning to the replicator dynamics and then discusses various ways to solve the problem. It discusses also one more possibility, namely that the discussed effect is not only a computational artifact but a true property of human learning.

2 Speed of reinforcement learning: an analytical example

The basic idea behind the reinforcement learning is that the rewarded behavior is strengthened and the punished behavior is weakened. More specifically, it is assumed that an agent (human or animal) can choose an action from a known discrete set of actions. She has a mixed strategy, i.e. chooses each action with some probability. After choosing an action, the agent earns a payoff that depends on the chosen action and the state of the environment (which can include actions chosen by other agents). After observing the payoff associated with the chosen action, the agent updates her mixed strategy. The probability that the chosen action would be played in the future is increased in proportion to the payoff (it is decreased if the payoff was negative). The future probabilities of all other actions are changed accordingly so that the probabilities of all action sum to unity. The reinforcement learning is classified as "non-conscious" learning (see [4]) because the algorithm does not explicitly model the agent's cognitive reflection. Specifically, it models neither the agent's belief about the state of her environment, nor her belief about the strategies of other agents. Nonetheless, the reinforcement learning was successfully used in modeling the situations where the agents learned consciously (such as in laboratory experiments with human subjects) and where the situation was strategic (e.g. in games), see [4, p. 939]. Very often it produced better predictions of human subjects' behavior in standard games then game theory, see [6, p. 1003].

There are many variants of the simple reinforcement learning algorithm but the differences between them are minor: alternative variants differ only in the speed of learning (it may be constant or slowing down in time) and in the inertia after a change of the environment, see [4, p. 905]. For our exposition, we use the variant of the algorithm taken from [1]. It works like this: The learning proceeds in discrete rounds. In each round t, agent i chooses an action j with probability p_{ij}^t which is calculated from her propensities R_{ij}^t to play the action j as

$$p_{ij}^t = \frac{e^{\lambda R_{ij}^t}}{\sum\limits_{\forall k} e^{\lambda R_{ik}^t}}. \tag{1}$$

After observing the payoff, the propensity to play each action j is updated as

$$R_{ij}^{t+1} = qR_{ij}^t + I_{ij}^t \pi_i^t, \tag{2}$$

where π_i^t is the payoff agent i gained in round t and $I_{ij}^t = 1$ if the action j was chosen in the round t, and $I_{ij}^t = 0$ otherwise. There are three parameters: the initial propensities R_{ij}^1, the "forgetting" parameter $q \in (0, 1]$, and the "focus" parameter $\lambda \geq 0$. The forgetting parameter q allows the agent to change her mind if her environment (and hence payoffs) changed; it also secures the numerical stability of the algorithm. The initial propensities R_{ij}^1 determine the agent's initial mixed strategy; e.g. $R_{i1}^1 = R_{i2}^1 = \ldots = R_{im}^t$ means that in the first round agent i chooses each of her m actions with probability $1/m$. The focus parameter λ is said to determine "the extent to which the agent focuses on choices with higher values of R_{ij}^t" [1, p. 211]. If $\lambda = 0$, then each action is chosen with the same probability. As λ rises, the higher and higher probability is attached to the action with the highest propensity; it is chosen with probability equal to 1 in the limit.

The three parameters together with the size of the payoffs determine the speed of an agent's learning. Intuitively, the speed of learning means how fast the agent's mixed strategy degenerates to the choice of one pure action. This can be measured as the speed with which the entropy of the agent's mixed strategy decreases in time. Entropy E_i^t of agent i's mixed strategy in round t is defined as

$$E_i^t = -\sum_{k=1}^{m} p_{ik}^t \log_m p_{ik}^t \tag{3}$$

where the base of the logarithm is the number m of agent i's actions. Notice that $E_i^t \in [0, 1]$ reaches its maximum when agent i plays each action with the same probability $1/m$, and its minimum when she plays one action with probability 1 and the other actions with probability 0. It has an intermediate value for other mixed strategies and decreases as the agent focuses. We claim that 1) the speed of learning depends beside others also on the payoff

size, and 2) if the speed of learning is too high, the chosen pure action need not to be the right one, i.e. the action with the highest payoff. From some speed of learning, the ex ante probability that the right action would be chosen decreases in the speed of learning and converges to its probability in the initial mixed strategy.

We will show my claim first in the simplest possible environment. Since even this case is not fully analytically tractable, we will discuss analytically only the speed of learning between the first and second round, and the corresponding ex ante probability that the right action is reinforced. The setting is like this: Let us assume a simulation where one agent plays against a deterministic automaton. The agent has two actions, a_1 and a_2. The payoffs of these actions are deterministic, $s\pi_1 > 0$ and $s\pi_2 > s\pi_1$ respectively where s is a "size" of the payoffs (let us say π_1 is normalized to 1). Let us further assume that the agent's initial propensities to play a_1 and a_2 are $R^1_{11} = R^1_{12} = r$ respectively, i.e. she plays each action with probability $1/2$ in the first round.

We have to explore two probabilities: the unconditional probability Ep^2_{12} that the right action a_2 is chosen in the second round and the probability p^- that the action randomly chosen in the first round is chosen also in the second round. Let us start with Ep^t_{12}. If the action a_1 is chosen in the first round, the agent sets the future propensities $R^2_{11} = qr + s\pi_1$ and $R^2_{12} = qr$. Hence, the conditional probability that she chooses action a_2 in the second round when she has chosen a_1 in the first round is then $p^2_{12}|a_1 = e^{\lambda qr}/(e^{\lambda qr+\lambda s\pi_1} + e^{\lambda qr}) = 1/(1 + e^{\lambda s\pi_1})$. On the other hand, if the action a_2 is chosen in the first round, the agent sets $R^2_{11} = qr$ and $R^2_{12} = qr + s\pi_2$. Then the conditional probability that she chooses action a_2 in the second round when she has chosen a_2 in the first round is $p^2_{12}|a_2 = e^{\lambda qr+\lambda s\pi_2}/(e^{\lambda qr} + e^{\lambda qr+\lambda s\pi_2}) = e^{\lambda s\pi_2}/(1 + e^{\lambda s\pi_2})$. Since each action is chosen with probability equal to $1/2$ in the first round, the unconditional probability Ep^2_{12} is

$$Ep^2_{12} = 1/2p^2_{12}|a_1 + 1/2p^2_{12}|a_2 = \left(\frac{e^{\lambda s\pi_2}}{1 + e^{\lambda s\pi_2}} + \frac{1}{1 + e^{\lambda s\pi_1}}\right)/2. \tag{4}$$

The probability that the same action is chosen in both rounds is

$$p^- = 1/2(1 - p^2_{12}|a_1) + 1/2p^2_{12}|a_2 = \left(\frac{e^{\lambda s\pi_1}}{1 + e^{\lambda s\pi_1}} + \frac{e^{\lambda s\pi_2}}{1 + e^{\lambda s\pi_2}}\right)/2. \tag{5}$$

The inspection of the equations (4) and (5) shows the following properties: First, from the equation (5) we can see that the probability p^- that the same action is played in the second round as in the first round is rising in the focus parameter λ, in the average size of the payoffs s, and in their product λs. If $\lambda s = 0$, then $p^- = 1/2$. As the product λs rises, the probability that the action chosen randomly in the first round is played again in the second round monotonically rises and converges to unity, i.e. $\lim_{\lambda s\to\infty} p^- = 1$. Notice that in our case, the agent's mixed strategy entropy E^t_i decreases as the probability that any action is chosen rises above $1/2$ (and the probability that the other action is chosen decreases below $1/2$). That means that the speed of learning between the first and the second round increases in λs.

Second, from equation (4) we can see that the unconditional probability Ep^2_{12} that the right action a_2 is chosen in the second round is increasing in π_2 and decreasing in π_1. In other words, the higher the difference in payoffs between the actions, the higher ex ante probability that the agent will choose the right action in the second round.

Third, from the same equation we can also see that the unconditional probability $Ep^2_{12} = p^1_{12} = 1/2$ in two cases: 1) when $\lambda s = 0$, i.e. when there is no learning at all and the agent chooses her actions independently in each round, each action with the same probability; 2) in the limit when λs is high ($\lim_{\lambda s\to\infty} Ep^2_{12} = 1/2$). In this case, the agent chooses in the second round the same action as in the first round (see the first point above), i.e. there is no further learning since the agent is locked in the previously randomly chosen action. The $Ep^2_{12} = 1/2$ because each action is chosen with this probability in the first round. Notice also that the agent's (with probability $1/2$ inefficient) action is locked, and the agent cannot change her mind: she chooses the previously chosen action in each round, i.e. the other action is never tried and the propensity to play it decreases to zero while the action randomly chosen in the first round is reinforced forever.

Fourth, $dEp^2_{12}/d(\lambda s) > 0$ for $\lambda s = 0$. This together with $Ep^2_{12} = p^1_{12} = 1/2$ for $\lambda s = 0$ and $\lim_{\lambda s\to\infty} Ep^2_{12} = 1/2$ implies that there is a level l such that $dEp^2_{12}/d(\lambda s) < 0$ for any $\lambda s > l$, i.e. the probability that the dominant action a_2 is chosen in the second round decreases in λs, i.e. it decreases with the speed of learning.

We can summarize it this way: If λs is too small, the agents learn too slowly. If the product is too high, the agents learns so fast that he may learn to play a dominated action. Obviously, this is no problem in most simulations—the modeler simply has to set the focus parameter λ properly: the higher s, the lower λ. This way the learning speed may be calibrated for instance to fit the convergence speed observed in an experiment. (However, it might then be difficult to interpret the parameters q and λ as behavioral.) However, there is one overlooked instance

where this is indeed a problem: in simulations in which the size of payoffs depend on the treatment, i.e. it changes within the simulation. Then the speed of the learning can change within the simulation and unexpected things can happen. We will provide an example of such a simulation in the following section.

3 Case study: voluntary provision of public goods

In this section, we will describe an agent-based computational model of the simplest version of the voluntary provision of public goods experiment. The experiment consists of T discrete rounds. There are $N > 2$ players, the same in all rounds. In each round, each player is given some endowment w. She can contribute part of the endowment to a public good and save the rest for herself. Agent i's payoff is $\pi_i^t = (w - c_i^t) + M \sum_{\forall k} c_k^t$ in the round t, where c_i^t is agent i's contribution to the public good in round t and $M \in (0, 1)$ is the payoff of the public good. Notice that the only dominant (and hence equilibrium) action of every agent (both in each round and in the whole finitely repeated game) is $c_i^t = 0$. However, in the typical situation $M > 1/N$, the socially optimal action is $c_i^t = w$.

The stylized facts on results of the experiment can be found in [2, 8, 9]. They show that most people follow neither their dominant, nor their socially optimal strategy, but contribute somewhere in-between. The typical average contribution is about one half in the first round and it decreases in the time, however not to the zero. Since the game theory is not predictive here and since we know that most agent change their actions in time, it seems plausible that they learn how to play the game. Hence the experiment is a natural candidate for an agent-based computational simulation trying to explain the agents' behavior.

We will use the version of the experiment taken from [13] for the simulation. Here, the return on the public good $M = 0.5$, the endowment $w = 40$ and the set of actions is limited to three actions: $a_1 : c_i^t = 40$, $a_2 : c_i^t = 20$, and $a_3 : c_i^t = 0$. The only treatment in the simulation (not in the experiment [13]) is the number of the agents, $N = 4, 7, 10, 20, 40$, and 100. Notice, that the structure of the game implies that an increase in the number of agents rises the size of payoffs for $c_i^t > 0$. The parameters of the reinforcement learning algorithm were casually calibrated in such a way that they follow the stylized fact with the typical number of agents, $N = 7$. The forgetting parameter $q = 0.9$, the focus parameter $\lambda = 0.005$, and initial propensities $R_{ij}^1 = 0$ for each agent i's each strategy j. The model was simulated for 21 rounds, which is more than enough for a comparison with data from any experiment. Each simulation was repeated one thousand times. The model was simulated and the resulting data were analyzed in R [11].

The results of the simulations are summarized in Table 1. In both its panels, the columns denote the number of agents in the game and the rows denote rounds. The right panel shows that the average entropy of agents' mixed strategies (averaged over the agents and the simulations). It can be clearly seen that the agents learn much faster when there are more agents, and hence the size of payoffs is higher. For instance, when $N = 4$, there is some entropy in the agents' mixed strategies even after twenty rounds, i.e. the agents are still learning even after twenty rounds. On the other hand, when $N = 20$, the learning process has stopped after 13 round, and when $N = 100$, the learning has stopped after only two rounds.

The left panel of Table 1 shows the average contributions to the public good in percents of the endowment w (averaged over the agents and the simulations). This can be seen a measure of how fast (and if at all) the agents learn to play their dominant free riding strategy a_3. The more agents learned to play the dominant strategy, the less they contribute, and hence the lower is the average contribution. In the first round, the expected value of the average percentage contribution is $1/2$, which is given by the initial mixed strategy $(1/3, 1/3, 1/3)$. The average contribution in later rounds is rising in N for $N \geq 7$. This is clearly an artifact of the learning algorithm since there is nothing in the agents' preferences or in the structure of the game that could cause it. The reason is that the learning is faster when the number N of agents in the game is higher. The faster the learning, the higher probability that the agents learn a wrong action, and hence contribute more. The extreme case happens when $N = 100$. Then the agents learns their actions almost instantly—they learn to play the action they randomly chose in the first round. The reason is simple. The expected total contribution of 100 agents is $100/3 \times 0 + 100/3 \times 20 + 100/3 \times 40 = 2000$; an agent's payoff is then about 1000 (we can neglect the agent's individual private savings here). The agent then sets the propensity of the randomly chosen action to $R_{ij}^2 = qR_{ij}^1 + \lambda\pi_i^1 \doteq 0.005 \times 1000 = 5$. The probability that the same action would be chosen in the second round is then about $e^5/(2 + e^5) \doteq 99$ %. The eventual expected value of the average contribution is then the same as the initial one: $1/2$. It is because each agent learned to play always the action she randomly chose in the first round, and each agent chose each action initially with probability equal to $1/3$. Now, one third agents play the pure action a_1, one third the pure action a_2, and one third the pure action a_3.

When the number of agents in the game is lower, the learning is slower, and the entropy of the agents' mixed strategies drops to zero later. The longer learning allows the agents to learn better action. However, not all agents

(a) average percentage contribution to the public good

$t \setminus N$	4	7	10	20	40	100
1	0.509	0.499	0.495	0.499	0.501	0.502
2	0.499	0.493	0.490	0.492	0.498	0.501
3	0.483	0.492	0.492	0.488	0.495	0.502
4	0.490	0.479	0.482	0.488	0.495	0.502
5	0.490	0.478	0.473	0.484	0.494	0.502
6	0.461	0.474	0.475	0.481	0.494	0.502
7	0.456	0.458	0.466	0.479	0.494	0.502
8	0.459	0.453	0.458	0.478	0.494	0.502
9	0.460	0.450	0.450	0.476	0.494	0.502
10	0.442	0.451	0.452	0.479	0.494	0.502
11	0.438	0.437	0.443	0.477	0.494	0.502
12	0.435	0.431	0.438	0.475	0.494	0.502
13	0.428	0.428	0.441	0.475	0.494	0.502
14	0.422	0.429	0.438	0.476	0.494	0.502
15	0.416	0.425	0.434	0.476	0.494	0.502
16	0.403	0.415	0.430	0.477	0.494	0.502
17	0.427	0.408	0.433	0.476	0.494	0.502
18	0.397	0.406	0.427	0.476	0.494	0.502
19	0.400	0.403	0.426	0.477	0.494	0.502
20	0.397	0.396	0.421	0.476	0.494	0.502
21	0.404	0.392	0.424	0.476	0.494	0.502

(b) average entropy of agents' mixed strategies

$t \setminus N$	4	7	10	20	40	100
1	1.000	1.000	1.000	1.000	1.000	1.000
2	0.964	0.924	0.865	0.610	0.199	0.002
3	0.926	0.838	0.728	0.344	0.045	0.000
4	0.887	0.752	0.598	0.185	0.009	0.000
5	0.842	0.665	0.476	0.096	0.002	0.000
6	0.802	0.582	0.383	0.050	0.001	0.000
7	0.757	0.511	0.306	0.028	0.000	0.000
8	0.712	0.448	0.249	0.016	0.000	0.000
9	0.671	0.388	0.200	0.008	0.000	0.000
10	0.633	0.337	0.160	0.005	0.000	0.000
11	0.597	0.296	0.127	0.003	0.000	0.000
12	0.563	0.268	0.102	0.001	0.000	0.000
13	0.533	0.240	0.083	0.001	0.000	0.000
14	0.508	0.218	0.069	0.000	0.000	0.000
15	0.485	0.192	0.057	0.000	0.000	0.000
16	0.461	0.174	0.049	0.000	0.000	0.000
17	0.435	0.157	0.041	0.000	0.000	0.000
18	0.419	0.145	0.035	0.000	0.000	0.000
19	0.403	0.135	0.030	0.000	0.000	0.000
20	0.388	0.125	0.027	0.000	0.000	0.000
21	0.374	0.116	0.025	0.000	0.000	0.000

Table 1 The left table shows the average contribution to the public goods in percents for a given number of agents (columns) in a given round (rows). The right table shows the average entropy of the agents' mixed strategies for a given number of agents (columns) in a given round (rows). Each value is averaged over one thousand simulations.

need to learn their dominant action. For instance, Table 1 shows that with $N = 20$, the agents' actions are locked after the 13th round but the agents still contribute in average 47.6 % of their endowment (instead of zero, as predicted by their dominant strategy).

4 Discussion

We have shown that the speed of reinforcement learning depends beside others on the size of payoffs. The higher payoffs, the faster learning. If the learning is too fast, the agents can learn to play the action they have randomly chosen in the first round. In such a case, the final distribution of the agents' pure strategies resembles the initial one agent's mixed strategy. When the modeler cannot control for this effect, unexpected results can happen. In the contribution to the public good simulation in the previous section, the result was that the average contribution rose with the number of agents in the game.

It might be interesting to compare the result with the replicator dynamics. Miller [10] showed for a similar version of the experiment that the more agents in the game, the slower the agents learn their dominant free riding actions through the replicator dynamics. He also showed that it is because the more agents imply the higher payoffs. However, there are two key differences between the replicator dynamics and reinforcement learning. First, in the replicator dynamics, the slow learning to play the right action is the result of the slow learning while in the reinforcement learning it is the result of too fast learning. Second, in the replicator dynamics, the agents eventually learn to play their dominant actions while in the reinforcement learning they can get stuck with suboptimal actions.

Although the problem discussed in this paper has been overlooked so far, there exists enhancements of the algorithm that can overcome it. The lock-in effect can possibly be alleviated by introduction of experimentation (I^t_{ij} in equation (2) is redefined so that $I^t_{ij} = 1$ when action j has been chosen, or $I^t_{ij} = \delta$ otherwise, where $\delta \in [0, 1]$) or by introduction of fictive payoffs to the non-chosen actions as in EWA (the term $I^t_{ij}\pi^t_i$ in equation (2) is redefined so that I^t_{ij} is as in experimentation and π^t_i is replaced by the payoff that the agent would obtain if she has chosen action j and the other agents have not changed their actions). However, neither of these techniques can solve the

problem completely if the payoffs are high enough since δ should be typically small.

The influence of the payoffs' size can be completely eliminated by substituting the gross payoffs determined by the game with some form of net payoffs, i.e. a transformation of the gross payoffs. There are many ways how to calculate the net payoffs. For instance, one can subtract an aspiration level from the gross payoffs, or calculate the fictive payoffs $\tilde{\pi}_{ij}^t$ and then subtract the $\max(0, \min_{\forall j}(\tilde{\pi}_{ij}^t)$ from the gross payoffs. Both these techniques can eliminate the variation in the speed of leaning if used properly. However, some modelers can object to it because it may be incompatible with learning from the feedback. For instance, in his seminal paper, Cross [5, p. 247] claims: "It should be stressed that the concept of opportunity cost has no place in this analysis either. The theory of opportunity cost is derived from the maximization hypothesis, and its introduction would be inconsistent with the action-taking orientation of this paper."

Finally, it should be mentioned that there is one more possibility, namely that the variation of the speed of learning with the payoff size and the resulting effects are not just an computational artifact but a property of human (and animal) learning. Indeed, one of the stylized facts of the contribution to public goods experiments is that the average contribution does rise with the number of human subjects in the game, see [2, 8, 9]. When the agent-based model of learning uses net payoffs (as e.g. in [2]), the fact must be explained by other-caring preferences. The variation in the speed of reinforcement learning then offers an alternative (and perhaps easier) explanation. Clearly, more research on how people truly learn is needed.

References

[1] Arifovic, J.—Ledyard, J.: Scaling Up Learning Models in Public Good Games, *Journal of Public Economic Theory* **6** (2004), 203–238.

[2] Arifovic, J.—Ledyard, J.: Individual evolutionary learning, other-regarding preferences, and the voluntary contributions mechanism, *Journal of Public Economic Theory* **96** (2012), 808–823.

[3] Bell, A. M.: Reinforcement Learning Rules in a Repeated Game, *Computational Economics* **18** (2001), 89–111.

[4] Brenner, T.: Agent Learning Representation: Advice on Modeling Economic Learning. In: *Handbook of Computational Economics*, vol. 2, (L. Tesfatsion and K. L. Judd, eds.), Elsevier, 2006, 831–880.

[5] Cross, J. G.: A stochastic learning model of economic behavior, *Quarterly Journal of Economics* **87** (1973), 239–266.

[6] Duffy, J.: Agent-Based Models and Human Subject Experiments. In: *Handbook of Computational Economics*, vol. 2, (L. Tesfatsion and K. L. Judd, eds.), Elsevier, 2006, 831–880.

[7] Fudenberg, D.—Levine, D. K.: *The Theory of Learning in Games*, MIT Press, 1998.

[8] Holt, Ch. A.: *Markets, Games, & Strategic Behavior*, Pearson Education, 2007.

[9] Ledyard, J. O.: Public Goods: A Survey of Experimental Research. In: *Handbook of Experimental Economics*, (J. H. Kagel and A. Roth, eds.), Princeton University Press, 1995.

[10] Miller, J. H.: Can evolutionary dynamics explain free riding in experiments?, *Economic Letters* **36** (1991), 9–15.

[11] R Development Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Wien, 2011, http://www.R-project.org/.

[12] Sutton, R. S. and Barto, A. G.: *Reinforcement Learning: An Introduction*, 2nd ed. in progress, available at http://webdocs.cs.ualberta.ca/˜sutton/book/the-book.html, 2012.

[13] Šeneklová J.—Špalek, J.: Jsou ekonomové jiní? Ekonomický model versus realita, *Politická ekonomie* 1 (2009), 21–44.

[14] Tesfatsion, L.: Agent-Based Computational Economics: A Constructive Approach to Economic Theory. In: *Handbook of Computational Economics*, vol. 2, (L. Tesfatsion and K. L. Judd, eds.), Elsevier, 2006, 831–880.

Time series representation and appropriate estimation

Petr Lachout[1]

Abstract. We present a discussion on time series models, particularly, on ARMA and ARIMA model. Considering non-uniqueness of the model representation, we propose to select an appropriate model optimizing simultaneously two criteria: "fit" and "model complexity". These criteria are discordant. Therefore, we have to consider their suitable combination; e.g. AIC or BIC criterion. Our contribution is in larger set of competitive models than monographs suggest.

Keywords: ARMA process, ARIMA process, AIC criterion, algebra of polynomials.

JEL classification: C22
AMS classification: 91B84; 62M10

1 Introduction

Time series analysis is developed under background of the Hilbert space L_2 and the calculus of polynomials in the backward shift operator, see e.g. [1], [2], [3], [4], [5]. This theoretical description and consequent results possess relevant impact to practice. Modern analysis of financial and economic data can be hardly made without this fruitful theory.

Our realization is focused to time series models based on equations expressed by means of polynomials in the backward shift operator. Let us mention such models like AR, MA, ARMA, ARIMA, ARCH, GARCH, etc. Any such equation can be divided by a polynomial with non-zero roots. Obtained equation is equivalent with former one because the division is reversible. Multiplication with the same polynomial is giving the former equation. Thus, the theoretical model can be written in several equivalent forms. Unfortunately, the forms are theoretically equivalent, but, different in practice. Each of them can result in a different estimation of the model.

In the paper we presents the realization for the model ARMA and consequently for the model ARIMA. For them, we suggest family of competitive models which is larger than model family proposed in literature; see e.g. [1], [2], [3], [4], [5]. We select among models using AIC criterion. The proposed procedure is demonstrated on a data set.

2 Description of the idea for ARMA process

Let us recall a definition of the ARMA process.

Definition 1. A random process $(\mathsf{Y}_t, t \in \mathbb{N})$ is called $\mathsf{ARMA}(\mathsf{p},\mathsf{q})$ if it fulfills:

1. It is stationary.
2. It is causal.
3. It fulfills an equation

$$\forall t \in \mathbb{N} \qquad \Phi(\mathsf{B})\,\mathsf{Y}_t = \Theta(\mathsf{B})\,\mathsf{Z}_t, \tag{1}$$

[1]Dept. Probability and Statistics, Charles University in Praha, Faculty of Mathematics and Physics, Petr.Lachout@mff.cuni.cz

where

$$\Phi(z) = 1 + \phi_1 z + \phi_2 z^2 + \cdots + \phi_{\mathsf{p}} z^{\mathsf{p}},$$
$$\Theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_{\mathsf{q}} z^{\mathsf{q}}$$

are polynomials, $\mathsf{p}, \mathsf{q} \in \mathbb{N}_0$, B is the backward shift operator acting from $\mathbb{R}^{\mathbb{N}}$ to $\mathbb{R}^{\mathbb{N}}$, and, Z is a White Noise with zero mean and positive variance σ_{Z}^2.

To guarantee uniqueness of polynomials, it is assumed that the polynomials Φ, Θ are possessing no common root.

Let us recall that autoregression, moving averages and uncorrelated sequences (i.i.d. is a particular uncorrelated sequence) belong among ARMA processes. Particularly, $\mathsf{AR}(\mathsf{p}) = \mathsf{ARMA}(\mathsf{p}, 0)$, $\mathsf{MA}(\mathsf{q}) = \mathsf{ARMA}(0, \mathsf{q})$, and, $\mathsf{ARMA}(0, 0)$ means uncorrelated sequence.

Because, the ARMA process is stationary and causal, all roots of Φ must be outside of the unit circle; i.e. their norm is larger then 1. Moreover, the process can be written as an infinite sum

$$\mathsf{Y}_t = \frac{\Theta}{\Phi}(\mathsf{B})\mathsf{Z}_t = \Gamma(\mathsf{B})\,\mathsf{Z}_t = \sum_{\tau=0}^{+\infty} \gamma_\tau \mathsf{Z}_{t-\tau}, \tag{2}$$

where

$$\Gamma(z) = \frac{\Theta}{\Phi}(z) = \sum_{\tau=0}^{+\infty} \gamma_\tau z^\tau,\ \gamma_0 = 1$$

and $\frac{\Theta}{\Phi}$ denotes division of the polynomial Θ by the polynomial Φ.

The equality (1) can be rewritten in many equivalent forms. It is because, the formula can be divided by any polynomial with roots outside of the unit circle without any loss of information.

Let us consider a polynomial

$$\mathsf{Q}(z) = q_0 + q_1 z + q_2 z^2 + \cdots + q_\kappa z^\kappa$$

with roots outside of the unit circle. For simplicity we also require $q_0 = 1$.

Dividing the equality (1) by Q we are receiving its equivalent form

$$\forall t \in \mathbb{N} \qquad \widetilde{\Phi}(\mathsf{B}|\mathsf{Q}, +\infty)\,\mathsf{Y}_t = \widetilde{\Theta}(\mathsf{B}|\mathsf{Q}, +\infty)\,\mathsf{Z}_t, \tag{3}$$

where

$$\widetilde{\Phi}(z|\mathsf{Q}, +\infty) = \frac{\Phi}{\mathsf{Q}}(z) = \sum_{\tau=0}^{+\infty} \widetilde{\phi}_\tau z^\tau,\ \widetilde{\phi}_0 = 1,$$
$$\widetilde{\Theta}(z|\mathsf{Q}, +\infty) = \frac{\Theta}{\mathsf{Q}}(z) = \sum_{\tau=0}^{+\infty} \widetilde{\theta}_\tau z^\tau,\ \widetilde{\theta}_0 = 1.$$

Seeking for a process $\mathsf{ARMA}(\widetilde{\mathsf{p}}, \widetilde{\mathsf{q}})$, we actually allow an inaccuracy in the equation

$$\forall t \in \mathbb{N} \qquad \widetilde{\Phi}(\mathsf{B}|\mathsf{Q}, \widetilde{\mathsf{p}})\,\mathsf{Y}_t = \widetilde{\Theta}(\mathsf{B}|\mathsf{Q}, \widetilde{\mathsf{q}})\,\mathsf{Z}_t + \varepsilon_t, \tag{4}$$

where

$$\widetilde{\Phi}(z|\mathsf{Q}, \widetilde{\mathsf{p}}) = \sum_{\tau=0}^{\widetilde{\mathsf{p}}} \widetilde{\phi}_\tau z^\tau,\ \widetilde{\Theta}(z|\mathsf{Q}, \widetilde{\mathsf{q}}) = \sum_{\tau=0}^{\widetilde{\mathsf{q}}} \widetilde{\theta}_\tau z^\tau,$$
$$\varepsilon_t = \sum_{\tau=\widetilde{\mathsf{q}}+1}^{+\infty} \widetilde{\theta}_\tau \mathsf{Z}_{t-\tau} - \sum_{\tau=\widetilde{\mathsf{p}}+1}^{+\infty} \widetilde{\phi}_\tau \mathsf{Y}_{t-\tau}.$$

Taking $\widetilde{\mathsf{p}}, \widetilde{\mathsf{q}} \in \mathbb{N}_0$ large enough, the variance of the inaccuracy ε_t becomes to be smaller than a prescribed level. Hence, we would need enormous number of observation to be able to distinguish between models

$\mathsf{ARMA}(\mathsf{p},\mathsf{q})$ and $\mathsf{ARMA}(\widetilde{\mathsf{p}},\widetilde{\mathsf{q}})$. Thus, we receive several competitive models for an observed time series. Fortunately, these model will differ in the number of active coefficients. Thus, we can select appropriate model using simultaneously two criteria: "fit" and "model complexity". To be able to decide, we have to accept a convenient combination of these two criteria. Several such combinations are used in practice, e.g. criterion AIC or criterion BIC. The approach is suggested in monographs; e.g. [1], [2], [3], [4], [5].

Standard suggestion is to consider a set of competitive models $\mathsf{ARMA}(\widetilde{\mathsf{p}},\widetilde{\mathsf{q}})$, $0 \le \widetilde{\mathsf{p}} \le \widehat{\mathsf{p}}$, $0 \le \widetilde{\mathsf{q}} \le \widehat{\mathsf{q}}$, or, $\mathsf{ARMA}(\widetilde{\mathsf{p}},\widetilde{\mathsf{q}})$, $0 \le \widetilde{\mathsf{p}}$, $0 \le \widetilde{\mathsf{q}}$, $\widetilde{\mathsf{p}} + \widetilde{\mathsf{q}} \le \Delta$, where $\widehat{\mathsf{p}},\widehat{\mathsf{q}} \in \mathbb{N}$ (resp. $\Delta \in \mathbb{N}$) are properly chosen.

We intend to employ more of equivalent models. For that, we suggest to incorporate also the observation that some coefficients in competitive models can vanish. For such a model, number of coefficients decrease and the criterion can decrease, also. Our suggestion is to consider a set of competitive models $\mathsf{ARMA}(\widehat{\mathsf{p}},\widehat{\mathsf{q}})$, $\phi_i = 0$, $i \in I$, $\theta_j = 0$, $j \in J$ for each selection $I \subset \{1,2,\ldots,\widehat{\mathsf{p}}\}$, $J \subset \{1,2,\ldots,\widehat{\mathsf{q}}\}$.

Similar idea can be implemented for estimation of ARIMA model. Since, a random process $(\mathsf{Y}_t, t \in \mathbb{N})$ is $\mathsf{ARIMA}(\mathsf{p},\mathsf{d},\mathsf{q})$ if and only if random process $(\Delta^{\mathsf{d}}\mathsf{Y}_t, t \in \mathbb{N})$ fulfills $\mathsf{ARMA}(\mathsf{p},\mathsf{q})$. Recall that Δ^{d} denotes the d-difference, i.e. $\Delta^0\mathsf{Y}_t = \mathsf{Y}_t$, $\Delta^{\mathsf{d}+1}\mathsf{Y}_t = \Delta^{\mathsf{d}}\mathsf{Y}_{t+1} - \Delta^{\mathsf{d}}\mathsf{Y}_t$.

Hence for ARIMA process, we suggest to consider a set of competitive models $\mathsf{ARIMA}(\widehat{\mathsf{p}},\mathsf{d},\widehat{\mathsf{q}})$, $\phi_i = 0$, $i \in I$, $\theta_j = 0$, $j \in J$ for each selection $I \subset \{1,2,\ldots,\widehat{\mathsf{p}}\}$, $J \subset \{1,2,\ldots,\widehat{\mathsf{q}}\}$.

Moreover, if we hesitate between ARMA and ARIMA we can join both cases and, for example, consider competitive models $\mathsf{ARIMA}(\widehat{\mathsf{p}},\mathsf{d},\widehat{\mathsf{q}})$, $\phi_i = 0$, $i \in I$, $\theta_j = 0$, $j \in J$ for each selection $I \subset \{1,2,\ldots,\widehat{\mathsf{p}}\}$, $J \subset \{1,2,\ldots,\widehat{\mathsf{q}}\}$, $\mathsf{d} \in \{0,1\}$.

3 Illustrative example

Let us consider daily closing prices of major European stock indices, 1991-1998. Particularly, we concentrate to the stock index CAC (France). The data are freely available, e.g. in the software package R under the name "EuStockMarkets".

In the sequel, we will analyze CAC index in years 1991-1998. Numerical computations are done using the software package R.

Extracting a trend and a seasonal part we receive a time series possessing stationary features. This time series we try to fit with an ARMA model. We consider a set of competitive models $\mathsf{ARMA}(3,3)$, $\phi_i = 0$, $i \in I$, $\theta_j = 0$, $j \in J$ for each selection $I \subset \{1,2,3\}$, $J \subset \{1,2,3\}$. Thus, we receive $2^6 = 64$ competitive models. Now, we use AIC criterion to order them. Ordered values of the AIC criterion look like:

13376.15	13377.40	13380.20	13380.70	13380.87	13381.39	13381.44	13381.61
13381.63	13381.73	13381.95	13382.27	13382.41	13382.65	13382.66	13382.69
13382.72	13382.72	13382.80	13382.84	13382.84	13382.85	13383.02	13383.21
13383.23	13383.26	13383.31	13383.31	13383.33	13383.34	13383.38	13383.49
13383.62	13384.23	13384.23	13384.27	13384.61	13384.79	13384.99	13385.11
13385.18	13385.24	14251.00	14386.76	14437.97	14439.51	14439.85	14440.96
14507.44	14683.33	14806.87	14984.58	15024.94	15026.94	15925.37	15964.66
16024.95	16153.41	16300.86	17532.96	NA	NA	NA	NA

The best model according to AIC criterion differs to the second one at 1.25. Then the difference between the second and the third one is 2.80. Hence AIC criterion is slowly increasing to the 42^{th} model. Then AIC criterion jumps to the 43^{th} model for 865.76. After that AIC criterion is visibly increasing and the last four models are inconvenient because they are giving estimate forming a polynomial with a root in the unit circle.

Let us introduce the first twenty models:

	AIC		choice phi		choice theta
1	13376.147	...	• • •	...	• • 0
2	13377.403	...	• • •	...	• • •
3	13380.199	...	0 • •	...	• • 0
4	13380.695	...	• 0 0	...	• 0 •
5	13380.868	...	• • 0	...	0 0 •
6	13381.392	...	• • 0	...	0 0 0
7	13381.442	...	• 0 0	...	• 0 0
8	13381.609	...	• • 0	...	• 0 •
9	13381.631	...	• 0 0	...	0 0 •
10	13381.727	...	0 • •	...	• • •
11	13381.950	...	• 0 •	...	0 0 •
12	13382.266	...	• 0 0	...	0 0 0
13	13382.413	...	• 0 •	...	0 0 0
14	13382.653	...	• 0 0	...	• • •
15	13382.662	...	• 0 •	...	• 0 •
16	13382.689	...	0 • •	...	• 0 0
17	13382.716	...	• • 0	...	• • •
18	13382.724	...	• 0 •	...	0 • •
19	13382.803	...	• • •	...	• 0 •
20	13382.839	...	• • 0	...	• • 0

Symbol '•' means no restriction on the coefficient and '0' means that the coefficient is fixed as zero.

The models leads to estimation of their coefficients. These estimates are:

		$\widehat{\phi}_1$	$\widehat{\phi}_2$	$\widehat{\phi}_3$		$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_3$
1	...	0.53562	-0.54279	-0.88112	...	1.53368	0.95138	0.00000
2	...	0.54808	-0.54175	-0.88902	...	1.56477	0.99573	0.02450
3	...	-0.00000	-0.57452	-0.34447	...	1.02026	0.42784	0.00000
4	...	-0.96819	-0.00000	-0.00000	...	0.045494	0.000000	-0.044267
5	...	-1.012549	0.042983	-0.000000	...	0.000000	0.000000	-0.042714
6	...	-1.010695	0.043764	-0.000000	...	0	0	0
7	...	-0.96555	-0.00000	-0.00000	...	0.044348	0.000000	0.000000
8	...	-0.52697	-0.42874	-0.00000	...	0.477771	0.000000	-0.058177
9	...	-0.97089	-0.00000	-0.00000	...	0.000000	0.000000	-0.043754
10	...	-0.00000	-0.56611	-0.35537	...	1.013480	0.418849	-0.018818
11	...	-0.992952	-0.000000	0.023576	...	0.00000	0.00000	-0.04228
12	...	-0.96833	-0.00000	-0.00000	...	0	0	0
13	...	-0.991445	-0.000000	0.024749	...	0	0	0
14	...	-0.96778	-0.00000	-0.00000	...	0.0461018	0.0053509	-0.0434824
15	...	-0.9726982	-0.0000000	0.0045586	...	0.041231	0.000000	-0.043731
16	...	-0.000000	-0.963770	0.030197	...	1.0005	0.0000	0.0000
17	...	-0.44877	-0.50116	-0.00000	...	0.564738	0.030783	-0.044593
18	...	-1.01341	-0.00000	0.04262	...	0.000000	-0.039433	-0.045252
19	...	-0.492517	-0.486769	0.026708	...	0.520576	0.000000	-0.047138
20	...	-0.26757	-0.67149	-0.00000	...	0.74677	0.06496	0.00000

The coefficients estimates are similar for the first and the second model. The consequent models are giving estimates which are very different.

4 Conclusion

Statistical features of an ARMA process are determined by coefficients $\phi_1, \phi_2, \ldots, \phi_{\mathsf{p}}$, $\theta_1, \theta_2, \ldots, \theta_{\mathsf{q}}$, p, q. Allowing inaccuracy, we immediately receive several different models which are describing the same process up to a given uncertainty.

Presented numerical example is showing this phenomenon on a real data set. Let us list some models together with their order in ordering given by AIC criterion:

- $\mathsf{ARMA}(3,2)$ is the best model for the data set,
- $\mathsf{ARMA}(3,3)$ is the second best,
- $\mathsf{ARMA}(3,2)$ with $\phi_1 = 0$ is the third best,
- $\mathsf{ARMA}(1,3)$ with $\theta_2 = 0$ is the 4^{th},
- $\mathsf{ARMA}(2,3)$ with $\theta_1 = \theta_2 = 0$ is the 5^{th},
- $\mathsf{AR}(2)$ is the 6^{th},
- $\mathsf{ARMA}(1,1)$ is the 7^{th},
- $\mathsf{ARMA}(3,3)$ with $\phi_1 = 0$ is the 10^{th},
- $\mathsf{AR}(1)$ is the 12^{th},
- $\mathsf{ARMA}(1,3)$ is the 14^{th}.

Comparing estimated acf (autocorrelation function) and pacf (partial autocorrelation function), one can use a rough rule suggested in monographs, see e.g. [1], [2], [3], [4], [5]. Since estimate of pacf possesses only two significant values for lag 1 and 2, the rule leads to the $\mathsf{AR}(2)$. In the ordering given by AIC criterion, this model is the 6^{th} best model for our data.

Acknowledgements

The research was supported by the Czech Science Foundation (grant P402/12/G097).

References

[1] Box, G.E.P.; Jenkins, G.M.: *Time Series Analysis Forecasting and Control*. Holden-Day, San Francisco, 1976.

[2] Brockwell, P.J.; Davis, R.A.: *Time Series: Theory and Methods* Springer-Verlag, New York, 1987.

[3] Cipra, T.: *Analýza časových řad s aplikacemi v ekonomii*. SNTL/ALFA, Praha, 1986.

[4] Durbin, J.; Koopman, S.J.: *Time Series Analysis by State Space Models*. Oxford University Press, Oxford, 2001.

[5] Wei, W.W.S.: *Time Series Analysis*. Addison-Wesley Publishing Company, Inc., Redwood City, 1994.

Similarity measures for response patterns on dichotomously scored items

Tommaso Lando[1], Lucio Bertoli-Barsotti[2]

Abstract. Response patterns from test, questionnaires, or other kind of evaluation instruments can give important information about latent variables which cannot be directly observed, such as the "ability" of the respondent individuals. The study of similarity between response patterns leads to a classification of the respondents that can explain the latent variable under a new perspective. There are several known similarity measures, mostly developed in fields such as the bibliometric and scientometric research, which are based on co-occurrences between "objects", and satisfy certain properties which unfortunately don't suit to our problem. Taking inspiration from these measures, we propose a class of similarity measures for response patterns on dichotomously scored items. This study is meant as an example of applying these measures within the item response theory framework, and as a critical assessment of the usefulness of this approach. Our method focuses on the position of respondents in the multidimensional space spanned by the aggregated subject-subject co-occurrences. A number of choices/parameters can be left to the user, but we provide default options reflecting our preferences.

Keywords: similarity, map, Rasch model

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

In the context of the Item Response Theory (IRT), a probabilistic "response model" refers to the issue of the measurement of a quantitative latent (unobservable) trait θ, given a set of k items used to measure that latent trait. In particular, in IRT, dichotomous response models define the probability of a score $x_{vi} = 1$ (to be interpreted as a correct answer), or $x_{vi} = 0$ (to be interpreted as a wrong answer), where v designates the individual, with $v = 1,2,...,n$, and i the item, with $i = 1, 2, ...,k$. The mathematical form of the general item response model is $p(x_{vi}) = P(x_{vi} \mid \theta_v, \beta_i)$, where the probability function p depends on a person parameter θ and on an item parameter β. Person and/or item parameters are possibly vector valued. Nevertheless, in their most simple form, they are scalar valued – the parameter β_i representing the item "difficulty", and the parameter θ_v representing the "ability" of respondent. The Rasch model (RM) [6] is a well-known instance of IRT response model, with scalar valued person and item parameters. The RM can be written:

$$p(x_{vi}) = P(X_{vi} = x_{vi} \mid \theta_v, \beta_i) = \frac{exp\{x_{vi}(\theta_v - \beta_i)\}}{1 + exp(\theta_v - \beta_i)}$$

It is worth noting that the difficulty parameter is on the same metric as the latent trait. As can be seen, the response model is based on a two dimensional data matrix (x_{vi}). For each respondent, there will be a row vector $x_v = (x_{v1},...,x_{vk})$ of item responses of length k. Let us refer to the vector x_v as a "response pattern". Individual v also corresponds to a total score (so called raw score) r_v, which is the sum of the elements of x_v. Similarly, we will denote s_i the sum $\sum_v x_{vi}$ (so called item marginal sum). In the sequel, we will assume, without loss of generality, that $s_1 \geq s_2 \geq \ldots \geq s_k$. A remarkable property of the RM is that the maximum likelihood estimates of person and item parameters are co-monotone with the corresponding marginal sums, respectively, r_v and s_i [2]. This corresponds to the intuitive notion that the ability of person v increases when his/her raw score increases. In particular, under the RM, all the persons obtaining the same raw score will receive the same estimated ability. Otherwise said, for a given a raw score the estimated abilities of a RM are not dependent on the response pat-

[1] VŠB TU Ostrava, Department of Finance 70121, Sokolská 33 Ostrava (Czech Republic), tommaso.lando@unicatt.it
[2] University of Bergamo, Via dei Caniana 2, 24127 Bergamo (Italy), lucio.bertoli-barsotti@unibg.it

terns. For a fixed raw score r_v, $0 \le r_v \le k$, the most probable response pattern is given by $x_{vi} = 1$ for every $i \le r_v$ and 0 otherwise (which is called *perfect Guttman* pattern). Correspondingly, the most improbable pattern is given by $x_{vi} = 0$ for every $i \le k - r_v$ and 1 otherwise (most *aberrant* response pattern).

The aim of this paper is to find an appropriate index to measure the "similarity" between individuals, based on the "similarity" between their corresponding response patterns. This could be useful to get an interpretation of the data (maybe) different from the traditional one, only based on the parameter estimates and the total scores.

In scientometric research, similarity measures are known, which are based on co-occurrence data. For example, co-citation data can be used to study the relations among authors or journals, co-authorship data can be used to study scientific cooperation. Unfortunately none of the known similarity measures [7] can be applied to response patterns, since they seem to be unsuitable for our particular situation (see section 3). Our idea is to integrate information from co-occurrence with information from the item marginal sums. For each pattern we define a particular corresponding distribution, based on the item marginal sums s_i, and we call it "conformity" distribution. Then, we propose a general class of similarity measures for response patterns, based on co-occurrences but also on the "dissimilarity" between these distributions.

Since some of the results we base on come from majorization theory, in section 2 we provide a small summary of the main concepts that may be useful to the reader.

2 Majorization ordering and generalizations

Majorization [5] is a pre-order on vectors used in mathematics to decide whether the components of a vector are more (or less) "spread out" or more (or less) "equal" than the components of another vector. Functions of the vectors that are consistent with this pre-order are called Schur-convex. Many statistical "dissimilarity" or "divergence" measures belongs to this class of functions. Majorization and Schur-convexity can be generalized from vectors to measurable functions [1].

Definition 1. *Let f be a μ-measurable function in a set I (where μ is a positive and finite measure on I). We define the increasing re-ordering of f with respect to μ as:*

$$f_{\uparrow}^{(\mu)}(t) = inf\{z: \nu_z \ge t\}, \quad t \in [0, \mu(I)],$$

where $\nu_z = \mu(B_z)$, $B_z = \{y \in I: f(y) \le z\}$.

Definition 2. *Let f, g be μ-integrable and μ-measurable functions on I ($f, g \in L^1(I,\mu)$). We say that g (μ)-majorizes f and write $f \prec_{(\mu)} g$ if:*

1. $\int_0^z f_{\uparrow}^{(\mu)}(t)dt \ge \int_0^z g_{\uparrow}^{(\mu)}(t)dt$, $\forall z \in [0, \mu(I)]$.
2. $\int_I f d\mu = \int_I g d\mu$.

When conditions for majorization are not completely satisfied, there are also weak definitions of majorization. In particular, we will need the following definition.

Definition 3. *Let $f, g \in L^1(I,\mu)$. We say that g weakly (μ)-majorizes f from above and write $f \prec_{(\mu)}^{w} g$ when only condition* 1) *of "Definition 2" is satisfied.*

A generalization of Karamata's theorem [4] allows us to characterize a class of functionals which are consistent with strong and weak (μ)-majorization. This result can be summarized by the following theorem.

Theorem 1. (μ)*-majorization theorem*
Let $f, g \in L^1(I,\mu)$.

1. $f \prec_{(\mu)} g$ *if and only if:*

$$\int_I \varphi(f)\, d\mu \le \int_I \varphi(g)\, d\mu \text{, for all convex functions } \varphi.$$

2. $f \prec_{(\mu)}^{w} g$ *if and only if:*

$$\int_I \varphi(f)\, d\mu \le \int_I \varphi(g)\, d\mu \text{, for all decreasing convex functions } \varphi .$$

(Proof: see [1]).

The theorem says that the functional $\int_I \varphi(f)\, d\mu$ is consistent with strong (μ)-majorization (φ convex) or weak (μ)-majorization from above (φ convex and decreasing).

3 Similarity between response patterns

As a general rule, we assume that similarity between patterns is a function of the number of co-occurences in the patterns. Co-occurrence, in our case, should be seen as the number of times when individual *v* and individual *w* give the same response, right or wrong, to the items. In order to define a similarity measure, say $S(x_v, x_w)$, between two response patterns, x_v and x_w, we assume that information from co-occurrence data should be "weighted" by information from the item marginal sums.

Let $\bar{x}_v$, with elements $\bar{x}_{vi} = 1 - x_{vi}$, $v = 1,2,...,n$, the reverse-scored pattern version (or "dual") of x_v. Then, we can obtain the number of co-occurrences c_{vw} between individual *v* and individual *w* by:

$$c_{vw} = x_v' \cdot x_w + \bar{x}_v' \cdot \bar{x}_w .$$

Anyway, the number of co-occurrences alone is a too rough measure, and is not enough to decide whether a pair of response patterns are similar or not. Consider, for example, the following response patterns (from a $n \times k$ data matrix, with $n > 5$):

1) x_1 = (1,0,0,…,0,0,0)
2) x_2 = (0,1,0,…,0,0,0)
3) x_3 = (1,1,0,…,0,0,0)
4) x_4 = (0,0,0,…,0,0,1)
5) x_5 = (0,0,0,…,0,1,1).

First, let's consider x_1, x_2 and x_4. These three patterns present the same number of co-occurrences between them, that is $k-2$. Anyway x_4 has an unexpected (aberrant) structure with respect to the others (remind that the item marginal sums are ordered in decreasing order, $s_1 \ge s_2 \ge \ldots \ge s_k$). Then, although it has the same raw score (1) and the same number of co-occurrences, its similarity with patterns x_1 and x_2 should be lower than the similarity between those two (which are not aberrant). So we want $S(x_v, x_w)$ to be such that $S(x_1, x_2) \ge S(x_1, x_4)$.

Further, consider the couples x_1, x_3 and x_4, x_5: depending on the scores s_1, s_2, s_{k-1}, s_k we could desire $S(x_1, x_3) = S(x_4, x_5)$ or even $S(x_1, x_3) < S(x_4, x_5)$, giving more weight to similarity between "aberrant" patterns, which are, by definition, less frequent.

So, similarity index proposed will hopefully follow these properties, it will not be based only on the "rough" co-occurrence that should be weighted according to the item difficulty. In particular, our solution is to weight each outcome in a pattern with the corresponding item marginal sum.

From pattern x_v we create a vector that we call *conformity distribution* $f_v = (f_{v1},...,f_{vi},...,f_{vk})$, where:

$$f_{vi} = x_{vi} s_i + \bar{x}_{vi} (n - s_i),$$

According to the total scores s_i we can also define the pattern which follows the most logical structure (score 1 for the "easy" items and 0 for the "difficult" ones). Let x^* be any pattern with conformity distribution f^* such that:

$$f_i^* = \max\{ s_i , n - s_i \}$$

for $i = 1,2,...,k$ (notice that x^* could not belong to the dataset and, moreover, could not be univoquely identified - but this is not essential for our purposes).

First of all, we want to find an index to measure the "distance" between x_v and x^*. An effective way to measure distance between distributions should be based on the ratios f_{vi} / f_i^*. Intuitively, we could say that f_v is "close" to f^* if their ratio is (on average) close to 1. Suppose that we want to find which distribution, between f_v and f_w is

closer to f^*. Since $\sum_i f_{vi} \le \sum_i f_i^*$ and $\sum_i f_{wi} \le \sum_i f_i^*$, we can compare the ratio vectors f_v / f^* and f_w / f^* using weak majorization (form above) with respect to the measure defined by f^* (called μ^*). Notice that vectors $f_v / f^* = (f_{v1} / f_1^*, \ldots, f_{vk} / f_k^*)$ and $f_w / f^* = (f_{w1} / f_1^*, \ldots, f_{wk} / f_k^*)$ are μ^*-measurable functions.
Further, the reason why majorization has to be studied with respect to μ^* is that, since we divide f_v and f_w by f^*, the increasing re-ordering of these functions must consider the different "weights" given by μ^*.
So $\frac{f_v}{f^*} \prec^{w}_{(\mu^*)} \frac{f_w}{f^*}$ if and only if:

$$\int_0^x \left(\frac{f_v}{f^*}\right)_{\uparrow}^{(\mu^*)} dt \ge \int_0^x \left(\frac{f_w}{f^*}\right)_{\uparrow}^{(\mu^*)} dt, \quad \forall x \in [0, \sum_i f_i^*],$$

which means that f^* is "closer" to f_v than to f_w. Then, for the (μ)-majorization theorem we have:

$$\Phi(f_v, f^*) = \sum_i \varphi\left(\frac{f_{vi}}{f_i^*}\right) f_i^* \le \sum_i \varphi\left(\frac{f_{wi}}{f_i^*}\right) f_i^* = \Phi(f_w, f^*)$$

for any decreasing convex function φ.
This suggests to use, as *dissimilarity measure* between two conformity distributions f_v and f_w, the function:

$$d(f_v, f_w) = \left|\Phi(f_v, f^*) - \Phi(f_w, f^*)\right|.$$

From this result, we could define a similarity index $S(x_v, x_w)$ by the ratio between similarity (based on "rough" co-occurrences) and dissimilarity (based on distributions), that is:

$$S(x_v, x_w) = \frac{c_{vw}}{1 + [d(f_v, f_w)]^{\alpha}},$$

with $\alpha \ge 0$.
This defines a general class of similarity measures, depending on the choice of φ in $\Phi(f_v, f^*)$ and α, which can vary according on the weight we want to give to co-occurrence rather than dissimilarity. For simplicity, we can take φ = -ln since it is convex, decreasing and takes value 0 when $f_v = f^*$.
The following theorems show two fundamental properties of S that can be useful to understand its results in many particular situations.

Theorem 2
Let x_u, x_v, x_w, x_z be response patterns such that $c_{uv}=c_{wz}=k-1$. Suppose that $f_{vi} = f_i^* > f_{ui}$, for a fixed index i between 1 and k, and $f_{zj} = f_j^* > f_{wj}$, for a fixed index j between 1 and k. Then:

1. *$f_i^* > f_j^*$ if and only if $S(x_u, x_v) < S(x_w, x_z)$;*
2. *$f_i^* < f_j^*$ if and only if $S(x_u, x_v) > S(x_w, x_z)$;*
3. *$f_i^* = f_j^*$ if and only if $S(x_u, x_v) = S(x_w, x_z)$.*

Proof:
It's easy to see that

$$d(f_u, f_v) = \varphi\left(\frac{1 - f_i^*}{f_i^*}\right) f_i^*, \quad d(f_w, f_z) = \varphi\left(\frac{1 - f_j^*}{f_j^*}\right) f_j^* \quad .$$

$f_i^* > f_j^*$ if and only if $\frac{1 - f_i^*}{f_i^*} < \frac{1 - f_j^*}{f_j^*}$. Since φ is decreasing and $c_{uv}=c_{wz}$ we get $d(f_u, f_v) \ge d(f_u, f_w)$ and then $S(x_u, x_v) < S(x_w, x_z)$. Similarly we can prove 2) and 3).

Theorem 3
Let x_u, x_v, x_w be response patterns such that $c_{uv}=c_{uw}= k-2$. Suppose that $f_{ui} \ne f_{vi} = f_{wi}$, for a fixed index i between 1 and k, $f_{vj} \ne f_{uj} = f_{wj}$, for a fixed index j between 1 and k, and $f_{wl} \ne f_{ul} = f_{vl}$, for a fixed index l between 1 and k. Then:

1. *$s_i \ge s_j \ge s_l$ implies $S(x_u, x_v) \ge S(x_u, x_w)$;*
2. *$s_j \ge s_l \ge s_i$ implies $S(x_u, x_v) \ge S(x_u, x_w)$.*

Proof:
From the assumptions, we know that *u* and *v* present the same values in any cell except from cell *i* and cell *j*. We could say that the two patterns present two "non-co-occurrences". Similarly, also *u* and *w* present two non-co-occurences, in cell *i* and *l*. The proof will be divided analyzing four different stuations.
1) $s_i \geq s_j \geq s_l \geq n/2$.
Note that $s_i \geq s_j$ implies $1\text{-}s_i \leq 1\text{-}s_j$, so $\varphi\left(\frac{1-s_i/n}{s_i/n}\right)s_i/n \geq \varphi\left(\frac{1-s_j/n}{s_j/n}\right)s_j/n$ (since φ is decreasing). So we get:

$$d(f_u, f_v) = \left|\varphi\left(\frac{1-s_i/n}{s_i/n}\right)s_i/n - \varphi\left(\frac{1-s_j/n}{s_j/n}\right)s_j/n\right| = \varphi\left(\frac{1-s_i/n}{s_i/n}\right)s_i/n - \varphi\left(\frac{1-s_j/n}{s_j/n}\right)s_j/n .$$

Similarly, since $s_i \geq s_l$, we get:

$$d(f_u, f_w) = \varphi\left(\frac{1-s_i/n}{s_i/n}\right)s_i/n - \varphi\left(\frac{1-s_l/n}{s_l/n}\right)s_l/n .$$

The two dissimilarities differ only by the second addend. $s_j \geq s_l$ implies $1\text{-}s_j \leq 1\text{-}s_l$, so:

$$\varphi\left(\frac{1-s_j/n}{s_j/n}\right)s_j/n \geq \varphi\left(\frac{1-s_l/n}{s_l/n}\right)s_l/n$$

(since φ is decreasing), and then $d(f_u, f_v) \leq d(f_u, f_w)$. Finally, considering that $c_{uv}=c_{uw}$, we get $S(x_u, x_v) \geq S(x_u, x_w)$.
2) $s_i \geq s_j \geq n/2 \geq s_l$.
In this situation, $d(f_u, f_v)$ remains the same as before (point 1)), but for $d(f_u, f_w)$ we get

$$d(f_u, f_w) = \varphi\left(\frac{1-s_i/n}{s_i/n}\right)s_i/n + \varphi\left(\frac{s_l/n}{1-s_l/n}\right)(1-s_l/n)$$

(since $f_{ui}= f_i^*=s_i$, $f_{wi}=1\text{-}\ s_i < f_i^*$ and $f_{ul}= f_l^*=1\text{-}s_l$, $f_{wl}=s_l < f_l^*$). Then we surely have $d(f_u, f_v) \leq d(f_u, f_w)$ and so $S(x_u, x_v) \geq S(x_u, x_w)$.
3) $s_i \geq n/2 \geq s_j \geq s_l$.
In this situation, $d(f_u, f_w)$ is the same as before (point 2)) and similarly we get:

$$d(f_u, f_v) = \varphi\left(\frac{1-s_i/n}{s_i/n}\right)s_i/n + \varphi\left(\frac{s_j/n}{1-s_j/n}\right)(1-s_j/n)$$

The two dissimilarities differ only by the second addend. $s_j \geq s_l$ implies $1\text{-}s_j \leq 1\text{-}s_l$, so:

$$\varphi\left(\frac{s_j/n}{1-s_j/n}\right)(1-s_j/n) \leq \varphi\left(\frac{s_l/n}{1-s_l/n}\right)(1-s_l/n)$$

(since φ is decreasing), and then $d(f_u, f_v) \leq d(f_u, f_w)$ which implies $S(x_u, x_v) \geq S(x_u, x_w)$.
4) $n/2 \geq s_i \geq s_j \geq s_l$.
This situation is very similar to point 1), we get:

$$d(f_u, f_v) = \left|\varphi\left(\frac{s_i/n}{1-s_i/n}\right)(1-s_i/n) - \varphi\left(\frac{s_j/n}{1-s_j/n}\right)(1-s_j/n)\right| = \varphi\left(\frac{s_j/n}{1-s_j/n}\right)(1-s_j/n) - \varphi\left(\frac{s_i/n}{1-s_i/n}\right)(1-s_i/n) ,$$

$$d(f_u, f_w) = \varphi\left(\frac{1-s_l/n}{s_l/n}\right)s_l/n - \varphi\left(\frac{1-s_i/n}{s_i/n}\right)s_i/n .$$

Since we already know (from point 3)) that $\varphi\left(\frac{s_j/n}{1-s_j/n}\right)(1-s_j/n) \leq \varphi\left(\frac{s_l/n}{1-s_l/n}\right)(1-s_l/n)$, then we get $d(f_u, f_v) \leq d(f_u, f_w)$ which implies $S(x_u, x_v) \geq S(x_u, x_w)$.

As a particular consequence of these theorems, take simple patterns x_1, x_2, x_3 and x_4 considered previously.

First, since $s_1 \geq s_2 \geq \ldots \geq s_k$, from theorem 3) follows $S(x_1, x_2) \geq S(x_1, x_4)$ as we wished (take $x_1=x_u$, $x_2=x_v$, $x_4=x_w$; $i=1$, $j=2$, $l=k$). Further, regarding similarity between x_1, x_3 and x_4, x_5, from theorem 2) follows that, depending on the scores s_2 and s_{k-1} we may get different results. For example, take $s_2=n-s_{k-1}$ (which means $f_2^*= f_{k-1}^*$): then we get $f\, S(x_1, x_3)= S(x_4, x_5)$. Take now $s_2<n-s_{k-1}$: then we get $S(x_1, x_3)< S(x_4, x_5)$, otherwise, if $s_2>n-s_{k-1}$ (it may happen even if not very common in a dataset) we get $S(x_1, x_3)> S(x_4, x_5)$. Intuitively, the heavier the weight (given by f^*) in the "non-co-occurrent" data, the lower the similarity.
By the last intuitive explanation of S's property, we can also wish that the heavier the weights in the "co-occurrent" data, the higher the similarity. Here, the weights in the co-occurrent data set are given by f_v (or identically f_w). This suggests to simply consider the sum of the weighted co-occurrences between patterns x_v, x_w, so:

$$c^*_{vw} = \sum_C f_{vi} = \sum_C f_{wi},$$

where $C = \{i: f_{vi} = f_{wi}\}$. c^*_{vw} alone could be taken as a similarity measure but, since its value is influenced by the weights given by the scores, c^*_{vw} could get higher if the two patterns are close to x^*, and a lower if they are not (for example if they are aberrant). In the previous situation, we surely get $c^*_{13} \geq c^*_{45}$ only because $s_1 \geq s_k$. To avoid this situation, the second proposed similarity measure S' consists in a normalization of c^*_{vw}. In particular, we can divide c^*_{vw} by $m^*_{vw}= \max\{c^*_{vv}, c^*_{ww}\}$. This value decreases as the patterns x_v and x_w tend to be aberrant. The normalization leads to the following formula:

$$S'(x_v, x_w) = \frac{c^*_{vw}}{m^*_{vw}}.$$

4 Conclusion

Both indexes proposed, S and S', seem to meet the requirements which should be necessary to a similarity measure between response patterns. They are not only based on co-occurrence data but also on information provided by the item marginal sum, giving a different weight to each item.
The study of similarities in response patterns could be useful in IRT, in particular in comparison with Rasch measurement. With the RM, latent variable, such as individual ability, can be measured with maximum likelihood estimation. In particular, it is known that these estimates are strictly related to marginal sums r_v [2]. Otherwise said, within the Rasch measurement framework classification of individuals is only based on their raw scores [3]. Then, as a limit case, a couple of individuals that give opposite response to each item could, for example, get the same raw score r_v: nevertheless, it is hard to think that their abilities is completely equal and matched.
In this paper we presented a deterministic approach (as opposed to the probabilistic approach of Rasch) to the problem of measurement. Similarity measures could provide a deeper analysis of individuals and their abilities, not only related to the scores but also to the corresponding item difficulties. This study provided a new perspective in latent variable analysis, where also information from response patterns could play a key role and can be useful to data interpretation.

References

[1] Bertoli-Barsotti, L.: Maggiorazione e Schur-convessità: generalizzazioni ed applicazioni statistiche. *Dipartimento di Statistica e Matematica applicata alle Scienze Umane "Diego de Castro" - Università degli Studi Torino* **7** (2001).

[2] Bertoli-Barsotti, L.: An order-preserving property of the maximum likelihood estimates for the Rasch model, *Statistics and Probability Letters* **61** (2003), 91-96.

[3] Del Pino, G., San Martin, E., Gonzales, J. and De Boeck, P.: On the relationships between sum score based estimation and joint maximum likelihood estimation. *Psychometrika* **73** (2008), 145-151.

[4] Karamata, J.: Sur une inegalité relative aux functions covexes. *Publ. Math. Univ. Belgrade* **1** (1932), 145-148.

[5] Marshall, A., and Olkin, I.: *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York, 1979.

[6] Rasch, G., *Probabilistic models for some intelligence and attainment tests*. MESA Press, Chicago, 1993.

[7] Van Eck, N.J., and Waltman, L.: How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology* **60(8)** (2009), 1635-1651.

Income Inequality and Gross Domestic Private Product Growth in Different Groups of Countries

Hana Lipovská[1], Jakub Fischer[2]

Abstract.

The aim of this paper is to analyse the relationship between economic growth, government expenditure and income inequality. Most studies suggest that there is a weak or non-existent relationship between growth and income inequality. We use Gross domestic private product growth instead of GDP growth. The analysis was performed by employing an OLS model for 73 countries. The clustering was used to classify data set into 5 subsets. A positive relationship between the GDPP growth and income inequality was found. This relationship is strong for developing and transitional economies and weak for the developed countries including the G7.

Keywords: economic growth, Gini index, government consumption

JEL Classification: H5, I3, O4
AMS Classification: 91B62

1. Introduction

The relationship between economic growth and income inequality measured by the Gini index is a frequently discussed macroeconomic topic addressed by numerous economic policies. Empirical studies discovered either no association or a negative relationship between economic growth and inequality in society (e.g.: Alesina, Rodrick [1]). Barro [2] proves that while the income inequality diminishes economic growth in poor states, it improves the growth rate in developed states, with GDP per capita equal to $2000 being the benchmark for sorting poor and developed states. Kuznets [5] suggested a widening inequality in the early phases of economic growth (e.g. developing countries) and a narrowing inequality in the later phases of industrialization and urbanization. There is no doubt, that the relationship is highly influenced by the way the GDP growth and income inequality is measured. Higgs [4] (based on Kuznets´ findings) pointed out that government product might be excluded from the GDP. Based on those studies, we developed adjusted macroeconomic indicator – Gross Domestic Private Product Growth (GDPP growth). The aim of this paper is to show that income inequality might contribute to the more rapid economic growth measured as Gross Domestic Private Product growth instead of Gross Domestic Product growth. The rest of the paper is organized as follows: we introduce briefly the theory of economic growth, show the importance of using GDPP instead of GDP and explain why income inequality accelerates the GDPP growth, while income equality contributes to the stagnation. We introduce the data as well as the statistical and econometrical methods we used. We then present results of our analysis. Firstly the relations between GDP growth rate and the Gini index and between GDPP growth rate and the Gini index are discussed. Secondly the impact of the Gini index, GDPP per capita and Gross Domestic Savings on GDPP growth is examined for 5 different groups of countries. Finally we analyze how the income inequality measured by the Gini index is influenced by GDPP growth, Gross Domestic Savings and Research and Development expenditures.

2. GDP and GDPP growth, Gini index

GDP and GDPP growth

Economic growth as the improving of economic production over time is rooted in several sources such as human capital, expected longevity, lower birthrate, stable formal institutions (e.g. legal or political), modest inflation and lower governmental consumption to GDP. There are three standard methods for estimating GDP – the output method, the income method and the expenditure method. For the aim of our contribution, the expenditure

[1] Masaryk University, Faculty of Economics and Administration, Lipová 41a, 602 00, Brno, lipovska@mail.muni.cz.
[2] University of Economics, Prague, Faculty of Informatics and Statistics, Department of Economic Statistics, W. Churchill Sq. 4, 13067, Praha 3, fischerj@vse.cz.

method is appropriate because it allowed us to eliminate government expenditures from the GDP to calculate GDPP. GDP growth is than calculated as the percentage change in GDP

$$GDP = C + I + NX + G, \qquad (1)$$

C being households consumption, I investments, NX net exports and G part of government expenditures. The government expenditures are the key part of this equity: government expenditures grew from the end of the 19th century to the 80´s in the 20th century. Panel data published by the World Bank suggests lower government expenditures, because the variable General government final consumption does not include the defense expenditures. The average government consumption for all economies in the dataset was 26%, although in developed and transitional countries it reached 32% and in African and Asian states it was just 17%. Government final consumption according to the World Bank includes cash payment for operating activities of the government in providing goods and services, compensation of employees, interest, subsidies, grants, social benefits or rents [9], but they do not include transfer payments. Barro pointed out, that government expenditures are one of the factors causing diminished economic growth [2]. For finding a relation between economic growth and income inequality, there are several reasons why government expenditures should not be part of the growth. According to Higgs [4] governments expenditures improve intermediate, rather than final production. Furthermore, it is difficult to calculate the market price of the goods and services purchased by government and finally the government expenditures are not influenced by the economic motives, but by political forces. This can be exemplified by the boom in government spending pulled by the need for supporting economic growth. Such spending is financed by tax-payers or by debt. In both cases, this artificial economic growth does not respond to the growth potential of such economies. We can illustrate it by a simple example: in the first year $GDP_1 = 3\,800$ and $G_1 = 760$, so $GDPP_1 = 3040$. Next year GDP_2 grew to 3857, so the GDP growth was 1.5%. However, this growth was influenced by the higher $G_2 = 877$, which means that $GDPP_2 = 2980$ and the GDPP growth was actually -1.9 %. In this case the GDPP growth explains the growth that is not influenced by governmental expenditures. Despite the questionable nature of such a claim, we believe, that GDP growth does not show the real potential growth of the national economy. We suggest using GDPP growth as the change of GDP adjusted by the government expenditure for calculating the relation between the income inequality and the tendency of the economies.

Gini index

The income inequality is often approximated by the Gini index measuring the spread between perfect equality and the real income (or wealth) distribution. Gini index is calculated as the ratio between the line of equality that forms a 45 degree angle with the x-axis and the Lorenz curve $L(x)$. The Gini index can be expressed as

$$GC = 1 - 2\int_0^1 L(x)dx, \qquad (2)$$

reaching the highest values around 65 (e.g.: Seychelles) and the lowest around 23 – 24 (Scandinavian economies).

This study and our findings could be improved with new proxy variable used instead of the Gini index. As Kuznets noticed, the income inequality should not be measured as the state in one moment, because of a migration among income groups. For this paper, however, the standard Gini index was used as we have not sufficient dataset available.

3. Data and models

A frequent problem when working with macroeconomic data is that only a sparse matrix is available, which means having to begin the analysis with as large a dataset as possible to get appropriately large dataset for the final modeling. We have used the World Bank data that is valuable for its high quality. Initially we used a data set of 214 cases; nevertheless only 73 cases could be used for our model. In spite of the fact that this article analyses cross-sectional data, for the primary data set data for the first decade of the 21st century (2001 – 2010) was used. As the benchmark year 2006 was chosen – this year was far away from the dot.com boom and following crises from the beginning of the millennium, and it was not heavily influenced by 2007 bubble and recent recession. That is the reason why data on GDP, GDP per capita, economic growth and inflation are referred to 2006. For the developed economies annual data is available, but this is not the case for developing or

underdeveloped countries, which provides this type of data sporadically. For all other variables excluding those above and GDPP growth the arithmetical average of available data from the period 2001 – 2010 was calculated.

After diminishing GDP by the government expenditures, the growth in GDPP was calculated. For each of the 73 states the average GDPP growth was calculated as the geometrical mean:

$$GDPP_{growth_i} = \sqrt[T]{\prod_{t=1}^{T} \frac{(1-0{,}01G_t) \times GDP_t - (1-0{,}01G_{t-1}) \times GDP_{t-1}}{(1-0{,}01G_t) \times GDP_{t-1}}},$$

$$i = 1, \dots 73, \min T = 1, \max T = 9. \qquad (3)$$

The Gini index was available for just several years of the discussed period moreover the World Bank data does not contain a Gini index for most of the member states of the Organization for Economic Co-operation and Development. If the World dataBank dataset was used, our analysis would miss data for e.g.: USA, Japan or Czech Republic. Aware of the risks of using in one analysis two types of data (and possibly two different methodologies), we have completed the data set with data from the OECD database. The arithmetical average of the Gini index enters our final model.

The final data set was created by excluding all economies, for which key macroeconomic data was not available. Furthermore, Bosnia and Herzegovina as well as Lesotho were excluded as outliers for a model using gross domestic savings to GDP. The final data set covers 75 % of world estimated GDP.

The OLS estimator was used to estimate a model explaining sources of income inequality and a model explaining sources of the GDPP growth. For macroeconomic data the risk of violation of homoscedasticity exists, therefore the White HCE estimator was used. The normality of residuals was not considered to be a problem, because the size of the data set allows us to assume asymptotical normality. World Bank as well as OECD data are widely valued for high quality. Nevertheless, there can be significant differences in methodology in space and time. The least developed countries as well as dictatorships are typical examples of this problem. The deliberate overestimation or underestimation of data can result in a significantly distorted model. On the other hand, excluding those economies from our model would lead to lack of spatial variety. For these reason, all the available data were finally used.

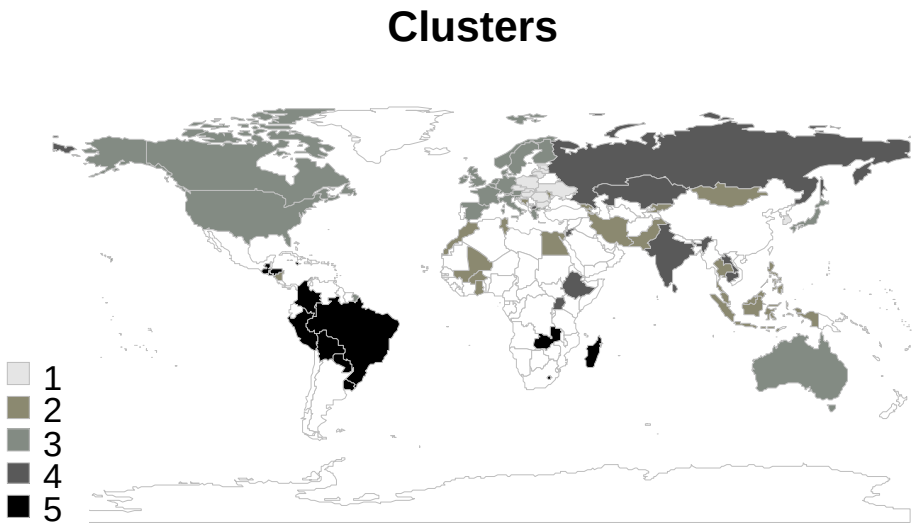


Figure 1 The spatial distribution of characteristic groups of states. (Source: author´s calculation and graphics)

Based on Kuznets [5] we expected differences between Gini index, GDPP growth and other variables among typical groups of states (according to the phase of their economies), so the cluster analysis was used. Five-means clustering was used on variables GDP per capita (PPP, 2006), GDPP growth and Gini index. The mean values for those variables and for all five clusters are depicted in Figure 2. The first cluster consists of transitional economies including the Czech Republic, as well as Israel and South Korea (as the only Asian state). The third cluster of Latin America states and four African states can be easily described. The remaining two clusters

cannot be clearly described in a geographical way, however all the states are similar in economic features: the second cluster includes several African and Asian states as well as Bosnia and Herzegovina, the forth cluster consists of African and Asian states, Russia and India (fig. 1). The forth cluster is the most important for our analysis, as it contains rapid GDPP growth and low Gini index (fig. 2). The impact of belonging to the clusters on the intercept is expressed by four dummies $D_i, i = 1, \dots 4$, the fifth cluster responds to the intercept of estimated models.

We used 22 variables included dummies. The first group of variables consists of variables connected to the gross domestic product: GDP and GDPP (measured in PPP), GDP per capita and GDPP per capita and GDP growth as well as GDPP growth. The second group of variables consists of $trade$ (export to GDP), income inequality proxy $Gini$ and gross domestic savings ($Savings$). Other variables describe government interventions, fiscal and monetary policy: $totaltax\_rate$ measures the percentage of aggregate company taxes, tax_rev measures economic freedom, governmental spending $expenses$, governmental spending on defense $military$, proxy of monetary policy $inflation$ and the unemployment rate U. The level of human capital was approximated by the total spending on research and development to GDP ($R\&D$), the demographical variables such as the size of population, population growth and rural population to the total population were used for the first models. However, for the OLS estimator just several of the variables above were statistically significant.

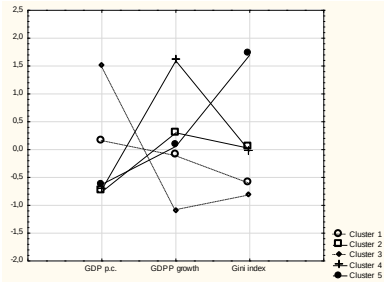


Figure 2 Graph of means (based on clustering analysis results)

4. Estimation results

No significant linear relationship between GDP growth and the Gini index was proved in our dataset. Part of the GDPP growth can be obviously explained by the income inequality measured by the Gini index however, this relation is still not very strong. Though, the relationship is positive: the higher values of Gini index are connected to the higher GDPP growth rate:

$$GDPP_{growth} = 3.06 + 0.078\, Gini. \qquad (4)$$

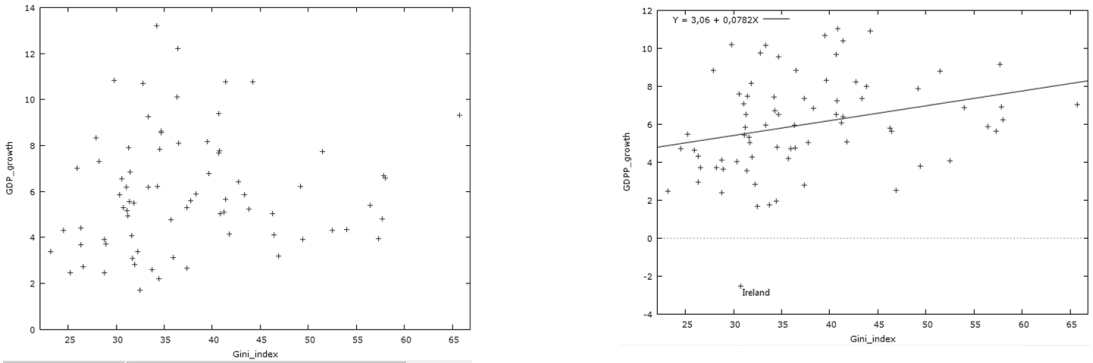


Figure 3 Relation between GDP growth and Gini index (left graph) and relation between GDPP growth and Gini index (right graph)

A correlation between GDP growth and Gini index $\mathrm{corr}(\mathrm{GDP}_{\mathrm{growth}}, \mathrm{Gini}) = 0.106$ is not significant at the p = 0.05 level, on the other hand correlation between GDPP growth and Gini index $\mathrm{corr}(\mathrm{GDP}_{\mathrm{growth}}, \mathrm{Gini}) = 0.282$ is significant on the p = 0.05 level.

	mean	minimum	maximum
Gini index	37.66	23.18	65.77
GDP per capita (PPP)	14813	711,16	53849
GDPP growth (%)	6.005	-2.53	11.04
GDP growth (%)	5.96	1.69	13.19

Table 1 Average, minimum and maximum values of key variables

Table 1 clearly shows the difference between GDP growth and GDPP growth. While there is no significant difference between the mean values, difference between minimal values is more than 4 percentage points and difference between maximum values is nearly 2 percentage points.

4. 1. GDPP growth model

Variable	Parameter estimate	SE	p-value	
$const.$	2,743	1,455	0,0638	*
$Gini\ index$	0,0624	0,0229	0,0084	***
$GDP\ p.c.$	$-7{,}181 \times 10^{-5}$	$2{,}978 \times 10^{-5}$	0,0187	**
$SAVINGS$	0,0479	0,0171	0,0066	***
D_1	1,139	0,475	0,0194	**
D_2	1,2397	0,615	0,0479	**
D_4	4,636	0,616	$1{,}88 \times 10^{-10}$	***
n	73			
R^2	0,71			

Table 2 Estimated OLS model of GDPP growth, with Gini index, GDP per capita, Gross domestic savings and dummies for characteristic groups of states as the explaining variables

The GDPP growth rate is explained by the income inequality, GDP per capita, Gross domestic savings and the clusters. As was expected, the higher Gini index and the Gross domestic savings effect the higher GDPP growth. The low base effect causes that the higher GDPP growth rate is typical for countries with lower GDP per capita. The GDPP growth in the South American and other states from the 5th cluster is *ceteris paribus* 2.74 %, transitive economies, included the Czech Republic, growth by 3.87 % and states from the 2nd cluster by 3.97 %. The highest growth (7.37 %) is in 4th cluster economies, including India and Russia. The average GDPP growth in those states was 10.15 % that is more than 4 percentage points higher in comparison with data set average. The GDPP growth in most developed states is significantly different from South America even if the explaining variables are taken in account: for the following average values $\overline{\mathrm{GINI}_2} = 37.858$, $\overline{\mathrm{GDPpc}_2} = 4433.5$, $\overline{\mathrm{SAVINGS}_2} = 16.206$ is the estimated GDPP growth in developed countries 5.53 %, for $\overline{\mathrm{GINI}_5} = 53.521$, $\overline{\mathrm{GDPpc}_5} = 6245.5$ and $\overline{\mathrm{SAVINGS}_5} = 13.127$ is the estimated GDPP growth in South America and several African states 6.22 %.

4. 2 GINI Index model

Variable	Parameter estimate	SE	p-value	
$const.$	39.309	3.458	3.46×10^{-17}	***
$SAVINGS$	-0.1733	0.089	0.0567	*
$\left(\frac{R\&D}{GDP}\right)$	-3.524	1.665	0.0381	**
$GDPP\ growth$	1.865	0.529	0.0008	***
D_1	-10.062	2.472	0.0001	***
D_2	-9.968	2.456	0.0001	***
D_4	-16.815	3.248	6.42×10^{-6}	***
n	73			
R^2	0.71			

Table 3 Estimated OLS model of income inequality measured by Gini index, with Gross domestic savings, Research and development expenditures, GDPP growth and dummies for characteristic groups of states as the explaining variables

The higher Gross domestic savings tend to lower the income inequality as do the Research and Development expenditures do. *R&D* expenditures are a proxy variable for the human capital and technology development: the high *R&D* rates are typical for Israel and Scandinavian countries. The transitional economies have a lower Gini index in comparison to South America and highly developed countries (the difference being 10 percentage points). The same effect is characteristic for the 2nd cluster economies. If the country is a member of the 4th group of countries (e.g.: India, Russia), the Gini index is the lowest. That is completely in accordance with our assumption, that in the fast developing countries the high GDPP growth rate and high income inequality are closely connected.

	1st cluster	2nd cluster	3th cluster	4. th cluster	5th cluster
$\overline{SAVINGS_\iota}$	21.299	16.206	23.791	15.978	13.127
$\overline{\frac{R\&D}{GDP_\iota}}$	1.2151	0.315	2.0925	0.349	0.216
$\overline{net\_growth_\iota}$	5.6854	6.786	3.168	10.119	6.152
$\overline{\mathrm{GINI_i}}$	31.9	36.8	33.78	37.3	47.7

Table 4 Estimation of Gini index for average values of explaining variables in different clusters

5. Conclusion

In our contribution, we have found a positive relation between income inequality measured by the Gini index and economic growth measured by the Gross domestic private product growth. Income inequality impact on the growth and vice versa is most significant among the different groups of states, which are similar in terms of Gross domestic product per capita, income inequality and growth rate. The countries of Latin America are those with the most pronounced income inequality differences but also with the most rapid growth. On the other hand, developed countries have modest inequality as well as modest growth. Those findings are completely in accordance with Kuznets´s [5] hypothesis.

As we have proven, economic growth (as the change in the status quo) might exist just in those societies, where the *space for growth* exists. Government interventions in form of government expenditures and other ways of narrowing the income gap between different groups of citizens may distort the natural potential for economic growth.

References

[1] Alesina, A., and D. Rodrik. (1994). Distribution Politics and Economic Growth. In: Quarterly Journal of Economics 109, 465–490.
[2] Barro, R. J.: Determinants of Economic Growth: A Cross-Country Empirical Study. NBER Working Paper Series. August 1996.
[3] Barro, R. J.: Inequality and Growth in a Panel of Countries. In: Journal of Economic Growth 5 (March 2000), 5-32.
[4] Higgs, R.: Government Bloat Is Not Growth: Real Gross Domestic Private Product, 2000-2011, www.mises.org.
[5] Kuznets, S. (1955). Economic Growth and Income Inequality. In: American Economic Review 45, 1–28.
[6] OECD, Government at a glance 2009, Paris, 2009.
[7] Rodríguez, F. C. Inequality, Economic Growth and Economic Performance: A Background Note for the World Development Report 2000, World Bank.
[8] Řezanková, H., Húsek, D. and V. Snášel: Shluková analýza dat, Praha, Professional Publishing, 2009.
[9] World Bank: World Development Indicators 2011, April 2011.

Periodic Timetable of Routes with Closed Scheduling Vehicles and Time Coordination in One Point Optimization Model

Filip Lorenz[1], Dušan Teichmann[2], Michal Dorda[3]

Abstract. The aim of this paper is the time coordination of regular bus transport service connecting the city of Ostrava with the Leoš Janáček Airport in Mošnov. The subject of the proposal will be based on a mathematical model that would enable to build up a timetable. The introduction of the paper is dedicated to the importance of the problem, possible use of different types of optimization methods and analysis of available sources of input data. The main part of the paper is focused on the design of an original model for the design of timetable, which allows the direct application of the results of work in practice. Computational experiments are performed in optimization software Xpress – IVE in conditions of upcoming summer season 2013. In the conclusion, the attention is given to evaluation of achieved results.

Keywords: time coordination of connections, linear programming, optimization

JEL Classification: C 61
AMS Classification: 90C05

1 Motivation for solving the task and current state of knowledge

The motivation to solve the problem is to increase the fluency of the transportation process of passenger transport to and from the airport Ostrava by special bus route. Traffic on the route is subject to a special mode - it is required to provide a regular headway between lines in both directions which is equal to 60 min. All lines of the route are served by 2 vehicles. The problem lies in the fact that the current timetable has no dependence on the times of departures and arrivals of individual flights from / at the airport. Passengers who use bus lines are often forced to unnecessarily wait before their departure or after their arrival. Therefore, the bus route is not too attractive to passengers and is little used. The aim is to propose a mathematical model to create a timetable for the bus route connecting the airport without above mentioned delays of passengers. It is expected that properly designed timetable may increase interest in public transport to the airport.

Our problem belongs to the group of co called coordination tasks. The basic problem of time coordination of lines was formulated by authors of the book [1] for the first time. On the basis of their publication the staff of Institute of Transport at VŠB-TU Ostrava has begun with research of time coordination. In the past, there were solved some works relating the problem of time coordination in isolated nodes of transport network [2]. There were also executed some experiments in which public transport was coordinated in several nodes concurrently [3]. From other works, which were oriented on solving this and similar topics, we can mention, for example, coordination tasks for route sections [4] - [7]. Works, which solves coordination in transport nodes, use mathematical models which were assembled by prof. RNDr. Jaroslav Janáček, CSc. from the University of Žilina in the past. The model from 2007 (this model has not been published) was also the inspiration for the solution to our problem. However, the original model has to be modified because the original model did not work with regular headway between the lines and with so called closed operating mode on the route. By closed operating mode on the route we mean that all the lines of the route are served by the same set of vehicles. The consequence of it is that each time shift of a line causes the same time shift of all other lines of the route. Original model worked only with passenger transfers in one direction; our model must be able to work with transfers in both directions.

[1] VŠB – Technical University of Ostrava / Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, Ostrava – Poruba, phone: +420 602 937 767, e-mail: filip.lorenz.st1@vsb.cz.
[2] VŠB – Technical University of Ostrava / Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, Ostrava – Poruba, phone: +420 597 324 575, e-mail: dusan.teichmann@vsb.cz.
[3] VŠB – Technical University of Ostrava / Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, Ostrava – Poruba, phone: +420 597 325 754, e-mail: michal.dorda@vsb.cz.

2 Formulation of the model

Let be a transfer node (the airport Ostrava) given, for which we would like to coordinate arrivals and departures of buses and arrival and departure times of charter flights. Let us consider that following sets are given: the set *I* of bus arrival times at the airport, the set *J* of flight departure times, the set *K* of time slots for departures of buses, and the set *L* of flight arrival times. For each arriving bus line $i \epsilon I$ it is known the earliest possible time t_{1i} of its arrival at the airport and for each departing bus line $k \epsilon K$ the earliest possible time t_{3k} of its departure out of the airport. It is also known the time interval *N* within which it is possible to move with the arrivals and departures of the buses on the route. For each time point, at which a flight departure $j \epsilon J$ occurs, it is known its time position t_{2j} and the number of flight departures B_j for the given period. For each time point of flight arrival $l \epsilon L$ it is also known its time position t_{4l} and the number of flight arrivals A_l for the given period. Further, there are known minimal periods T_{dep} in which it is desirable that the bus lines should arrive before flight departure (these data we consider to be constant for all flight departures) and minimum time T_{arr} that must elapse after the flight arrival to enable easy passengers transfer. It is required that the bus lines should arrive and depart with the regular headway and the time period between the time of bus arrival at the airport and its departure back to Ostrava has to be equal to 20 minutes. The maximal allowable time shift *x* of individual bus lines can be 59 minutes (it is given by defined regular headway between two adjacent lines in one direction).

Now let us introduce variables of the model. The variable *x* is used to model the time shift of the whole system of the bus lines. For each flight departure time $j \epsilon J$ the variable h_j is defined.

This variable will model the time loss of travellers given by the interval between the moment of the bus line arrival before the flight departure $j \epsilon J$ and the beginning of the appropriate time T_{dep} necessary for check-in. For each flight arrival time $l \epsilon L$ the variable q_l is established. This variable will model the time loss which is given by the interval between the time T_{arr} of check-in termination after flight arrival $l \epsilon L$ and the time of the bus line departure. For each pair - the arriving bus line $i \epsilon I$ and the flight departure at the time position $j \epsilon J$ - we introduce a bivalent variable z_{ij}. This variable models the existence of transfer links between the bus line $i \epsilon I$ and the flight departure at the time position $j \epsilon J$. In the case that the passenger transfer between the bus line and the flight will be realized than the variable z_{ij} takes the value equal to 1, in the case that there is no passenger transfer than the variable takes the value equal to 0. Analogously, for each pair - the departing bus line $k \epsilon K$ and the flight arrival at the time position $l \epsilon L$ - a bivalent variable m_{kl} is also established.

This variable models the existence of transfer links between the bus line $k \epsilon K$ and the flight arriving at the time position $l \epsilon L$. If the passengers transfer from the arriving flight to the bus line than the variable m_{kl} is equal to 1, if there is no transfer than the variable m_{kl} takes the value 0.

The mathematical model will be as follows:

$$\min f(x, z, m, h, q) = \sum_{j \epsilon J} B_j \, h_j + \sum_{l \epsilon L} A_l q_l \tag{1}$$

subject to:

$$t_{2j} - (t_{1i} + x) - T_{dep} \geq T(z_{ij} - 1) \qquad for\ i \epsilon I, j \epsilon J \tag{2}$$

$$t_{2j} - (t_{1i} + x) - T_{dep} \leq h_j + T(1 - z_{ij}) \qquad for\ i \epsilon I, j \epsilon J \tag{3}$$

$$(t_{3k} + x) - t_{4l} - T_{arr} \geq T(m_{kl} - 1) \qquad for\ k \epsilon K, l \epsilon L \tag{4}$$

$$(t_{3k} + x) - t_{4l} - T_{arr} \leq q_l + T(1 - m_{kl}) \qquad for\ k \epsilon K, l \epsilon L \tag{5}$$

$$\sum_{i \epsilon I} z_{ij} = 1 \qquad for\ j \epsilon J \tag{6}$$

$$\sum_{k \epsilon K} m_{kl} = 1 \qquad for\ l \epsilon L \tag{7}$$

$$x \leq N \tag{8}$$

$z_{ij} \in \{0,1\}$ $for\ i \in I, j \in J$ (9)

$m_{kl} \in \{0,1\}$ $for\ k \in K, l \in L$ (10)

$x \geq 0$ (11)

$h_j \geq 0$ $for\ j \in J$ (12)

$q_l \geq 0$ $for\ l \in L$ (13)

Function (1) represents the optimization criterion which is the total time loss of passengers who use the airport shuttle service. The group of constraints (2) ensures that if it is not possible to transfer due to the time reasons, transfer link will not be created. The group of constraints (3) ensures linking between the constraints and the first term of the objective function. Group of constraints (4) has the same function as the group of constraints (2) for the relation flight arrival - bus departure. The group of constraints (5) has the same function as the group of constraints (3); however, the link is ensured to the second term of the objective function. The group of constraints (6) ensures that each flight departure is assigned to exactly one arriving bus.

Constraints (7) ensure that each flight arrival is assigned to exactly one departing bus. Constraint (8) ensures that the time shift of the arriving and departing bus links takes the value which will not exceed the permitted limit. Constraints (9), (10), (11), (12) and (13) are obligatory and define domains of definition for individual variables.

3 Computational experiments

In this section we will present the results of executed computational experiments. Computational experiments are carried out to verify the correctness of the proposed mathematical model and its applicability to real-life situations. Bus arrival times at their earliest possible time positions belong to first group of the input data that are necessary for experiments. The route is served by 13 lines in each direction per day. The first bus line arrives at the airport Ostrava at 8:25 (this time is the earliest possible arrival time), followed by other lines arriving with the 60 minutes headway. The last bus line arrives at 20:25. The first bus line departs from the airport at 8:45, followed by other departing also with 60 minutes headway, the earliest possible departure of the last bus line from the airport is at 20:45. Other input data relating to charter flights are summarized in Table 1 (the flight arrivals) and Table 2 (the flight departures).

Time position of flight arrivals and departures are expressed in minutes that elapse from the selected point in time 0 (7:00).

Table 1

Arrival number l	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Time position t_{4l}	135	150	170	180	190	230	235	240	255	270	315	320	345	420	455
Count A_l	18	1	1	15	3	22	9	22	11	68	11	11	16	17	14
Arrival number l	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Time position t_{4l}	465	495	505	515	530	570	595	600	630	650	660	670	705	760	770
Count A_l	14	11	29	5	13	15	10	16	11	24	11	15	21	11	11

Table 2

Departure number j	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Time position t_{2j}	195	230	250	275	280	310	320	325	330	355	380	385	480	510
Count B_j	18	15	3	9	22	10	39	15	11	11	11	16	17	14
Departure number j	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Time position t_{2j}	515	555	565	575	600	660	680	700	710	720	740	750	820	850
Count B_j	15	32	9	5	15	10	11	11	49	34	12	9	11	12

Optimization experiments were carried out using the optimization software Xpress-IVE [8]. The measure of the solution quality is the difference between the value of the objective function for the solution obtained by solving the mathematical model and the value of the same criterion calculated for the current state. The values of the optimization criterion for the current state and for the state obtained by the model are shown in Table 3.

Table 3

The total time loss for the current state [min]	23 065
The total time loss gained as solution of the proposed mathematical model [min]	21 635
Difference [min]	1 430

From the comparison of results it is clear that the total time loss of passengers at the Ostrava airport was decreased by 1 430 minutes.

The result of the optimization calculation is also the time shift x of the whole system of bus lines; it takes the value 25 minutes. This value is added to the earliest possible time positions of all lines. After shifting the first bus line arrives at the airport in Ostrava at 8:50 and other lines will arrive with the regular headway equal to 60 minutes. The last line arrives at the airport in Ostrava at 20:50. The first bus line departs from the airport at 9:10 and others depart with the regular headway equal to 60 minutes. The last line departs from the airport in Ostrava at 21:10. Economic savings of the proposal based on the mathematical modelling result from the savings of time for passengers who use the bus service.

4 Conclusion

The article deals with the time coordination problem. The goal is to coordinate the lines of the bus route providing transportation to the airport with departures and arrivals of charter airlines. The article contains a linear mathematical model which has the optimization criterion based on the total time loss of passengers who use the bus line. The proposed mathematical model was tested in the conditions of the Ostrava Airport. Using the mathematical model it was achieved reduction equal to 1 430 minutes in time losses of passengers using the bus shuttle service. The mathematical model can be also applicable for other routes connecting smaller regional airports, for example, in Brno, Kosice, etc.

References

[1] Černý – Kluvánek: ZMTD

[2] Král, M. Časová koordinace přípojné taktové dopravy v železniční stanici Ostrava – Svinov. Ostrava: VŠB – TU Ostrava, Fakulta strojní, 2010.

[3] Kroužil, K. Časová koordinace osobní železniční dopravy v přestupních stanicích na trati Ostrava-Svinov – Český Těšín. Ostrava: VŠB – TU Ostrava, Fakulta strojní, 2010.

[4] Kozel, P. Časová koordinace spojů veřejné hromadné dopravy v úseku Frýdek-Místek – Dobrá. Ostrava: VŠB – TU Ostrava, Fakulta strojní, 2009.

[5] Kozel, P. Síťová koordinace spojů veřejné hromadné dopravy. Ostrava: VŠB – TU Ostrava, Fakulta strojní, 2012.

[6] Kozel, P.: The time coordination of bus links in both transport directions. Proceedings of the 29th international conference on mathematical methods in economics 2011, Part 1. (Dlouhy, M. and Skocdopolova, V. eds.). Professional Publishing, Praha, 2011, 357-362.

[7] Gabrisova, L. and Jancek, J.: Polygon regular location problem. In: *Proceedings of the 30th International Conference Mathematical Methods in Economics 2012, Part I.* (Ramík, J. and Stavárek, D. eds.). Silesian University in Opava, School of Business Administration in Karviná, Karviná, 2012, 196-200.

Numerical investigation of 3-D nonlinear financial model

Ladislav Lukáš [1]

Abstract. The paper concerns with numerical investigation and some generalization of 3-D nonlinear financial model, which have been reported in the literature recently. The model consists of three simultaneous nonlinear ordinary differential equations consisting of three state variables: the interest rate, the investment demand, and the price index, in particular. We introduce time dependent constitutive coefficient representing rate of investment instead of time invariant one. Further, we admit more complex construction of right hand side terms including time delayed response of state variables. Numerical investigation of model with different forms of investment rate function is reported in detail. We analyze also an influence of different initial functions upon generated trajectories. All computations for solving initial value problems are performed by sw Mathematica.

Keywords: financial model, nonlinear financial model, initial value function, initial value problem, numerical solution, state variables.

JEL Classification: C63, G19
AMS Classification: 37N40, 91G80

1 Introduction

No doubt, dynamic models in finance and their numerical solutions have attracted scientists and financial engineers since computers appeared. At present, we may classify two main streams of research within those challenging topics. The first one is based upon theory of stochastic differential equations, and we refer to [5] just for a brief touch. The second one concerns, it is mainly focused on theory of deterministic nonlinear dynamic systems, which can produce chaotic behavior, strange attractors, etc., and we refer to [1] and [7] for further reading. In 2001, an interesting nonlinear dynamic financial model was published in [3] and [4]. The model was further investigated and developed for delayed feedback in [2], as well. We also refer to [6] and [8] for some general aspects of formulation and handling delays in dynamic financial systems.

The paper concerns with numerical analysis of two effects – i) introduction of time dependent constitutive coefficient representing rate of investment instead of time invariant one, ii) construction of right hand side terms including time delayed response of state variables thus enabling an analysis of different initial functions influence upon generated trajectories. The model contains three state variables denoted x, y, and z, which represent interest rate, investment demand and price exponent, in particular. The first derivatives $\mathrm{d}x/\mathrm{d}t$, $\mathrm{d}y/\mathrm{d}t$, $\mathrm{d}z/\mathrm{d}t$ with respect to time represent the changing rates of the state variables as usual, which govern evolution of the system. Following [3] and [4], a core of the model consists of three phenomenological equations giving components and their influences upon these rates.

The changing rate of interest rate x, i.e. $\mathrm{d}x/\mathrm{d}t$, is composed from two factors coupled in linear form (1), a) the surplus between investment and saving, and b) the structural adjustment induced by goods prices.

$$\mathrm{d}x/\mathrm{d}t = f_1(y - s)\, x + f_2\, z \tag{1}$$

where s is a parameter representing an amount of saving being constant in a certain period of time by assumption, and f_i , i=1,2 are two constitutive constants.

The changing rate of investment demand y, i.e. $\mathrm{d}y/\mathrm{d}t$, is given by equation (2) and is composed from three factors including quadratic influence of x, which maintains proper nonlinearity of the model.

$$\mathrm{d}y/\mathrm{d}t = f_3(e - \alpha y - \beta x^2) \tag{2}$$

where e stands for benefit rate of investment being assumed constant in a certain period of time, too, and f_3, α, β are constitutive constants, where α expresses opposite proportional influence of y upon its rate, and β moderates an influence of x^2 upon that rate, thus expressing a higher order investment propensity caused by interest rate thereon.

[1] Univ. of West Bohemia, Faculty of Econ., Dept. of Econ. and Quantitative Methods, Husova 11, 30614 Plzeň, lukasl@kpm.zcu.cz.

Finally, the changing rate of price exponent z, i.e. $\mathrm{d}z/\mathrm{d}t$, is given by equation (3), and it takes a linear form again. It contains two constitutive constants f_i , i=4,5, so that f_4 passes an influence of market contradiction between supply and demand being expressed by variable z, and f_5 is to express the inflation rate influence.

$$\mathrm{d}z/\mathrm{d}t = -f_4\, z - f_5\, x \tag{3}$$

2 Simplified model and model with delayed feedback

Summarizing the model, we may inspect three state variables x, y, and z, and nine independent constitutive parameters f_i , i=1,.., 5 and s, e, α, β to be adjusted. An additional role of f_i is to maintain correct dimensioning. However following [3], we use a simplified model given by (4) with three non-negative constitutive parameters a, b, and c, only, which gauge saving amount, investment cost, and elasticity of demands of commercials transformed.

$$\begin{aligned} \mathrm{d}x/\mathrm{d}t &= z + (y-a)x, \quad x_0 = x(0) \\ \mathrm{d}y/\mathrm{d}t &= 1 - by - x^2, \quad y_0 = y(0) \\ \mathrm{d}z/\mathrm{d}t &= -x - cz, \quad z_0 = z(0) \end{aligned} \tag{4}$$

The model (4) can be simply extended to the model with delayed feedback, which is given by equations (5).

$$\begin{aligned} \mathrm{d}x/\mathrm{d}t &= z + (y-a)x + k_1(x - x(t-\tau_1)),\ t \geq 0, \quad x(t) = u_0(t),\ t \in [-\tau_1, 0] \\ \mathrm{d}y/\mathrm{d}t &= 1 - by - x^2 + k_2(y - y(t-\tau_2)),\ t \geq 0, \quad y(t) = v_0(t),\ t \in [-\tau_2, 0] \\ \mathrm{d}z/\mathrm{d}t &= -x - cz + k_3(z - z(t-\tau_3))\ t \geq 0, \quad z(t) = w_0(t),\ t \in [-\tau_3, 0] \end{aligned} \tag{5}$$

It contains three feedback intensities $k_j \in \mathrm{R}$, in general, delay times $\tau_j > 0$, j=1, 2, 3, and instead of initial values x_0, y_0, z_0 as in (4), it contains three initial histories $u_0(t)$, $v_0(t)$, $w_0(t)$. For $\tau_j = 0$ the model (5) turns to model (4).

3 Numerical results – case studies

All calculations were performed by Mathematica notebooks we developed for numerical investigation of both models (4) and (5) using extremely powerful command `NDSolve`. For the paper, we select just two topics to present – *i*) numerical investigation of model (4) with function $b(t)$ just replacing the constant b, *ii*) analysis of influence of different initial functions on state-space trajectories in model (5).

Figure 1 show trajectories of model (4) with different $t_{\max}$, $t_{\max}$ = 20 on the left, and 100 on the right, where we can identify also five loops run. Later, we compare the left image with other ones generated with model (5) with different initial history functions.

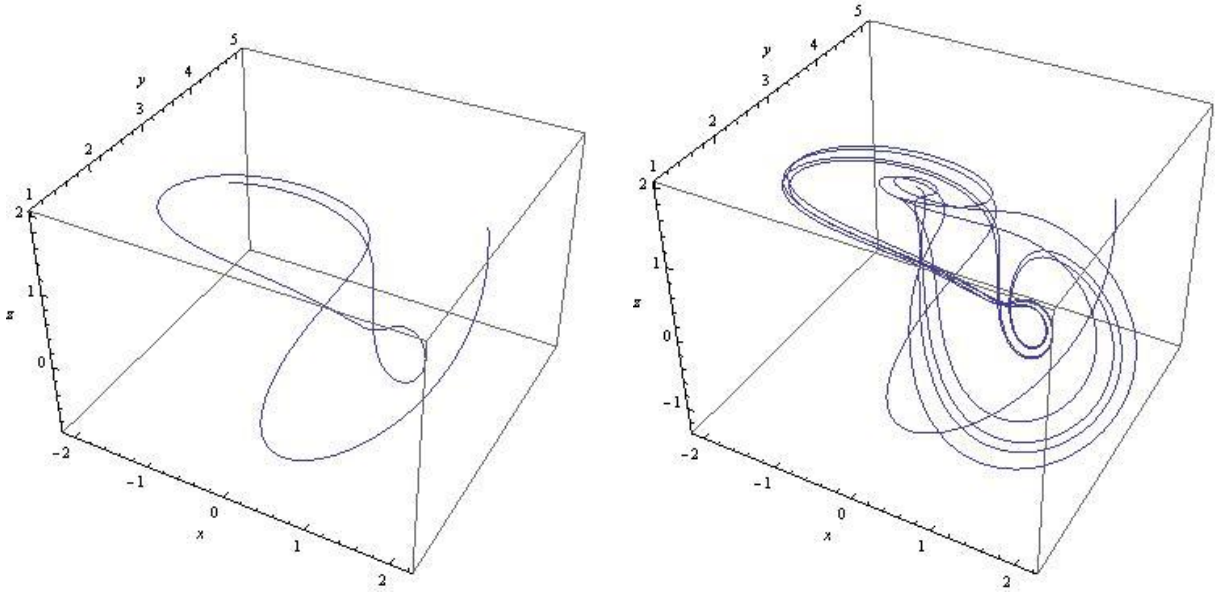
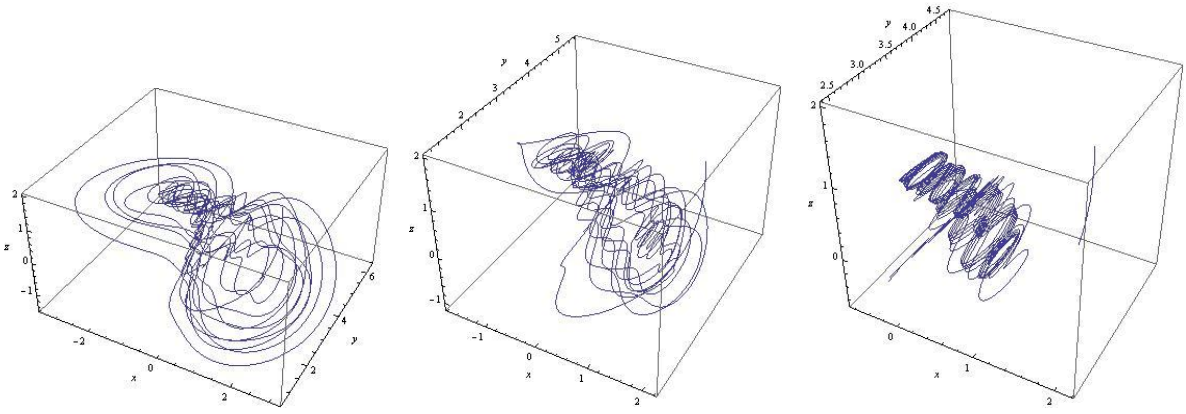


Figure 1 Model (4), a=3, b=0.1, c=1, x_0=2, y_0=3, z_0=2



β_1=0.1, β_2 =5, β_3=3, β_0=0 β_1=0.25, β_2 =10, β_3=3, β_0=0.25 β_1=0.28, β_2 =15, β_3=0, β_0=0.45

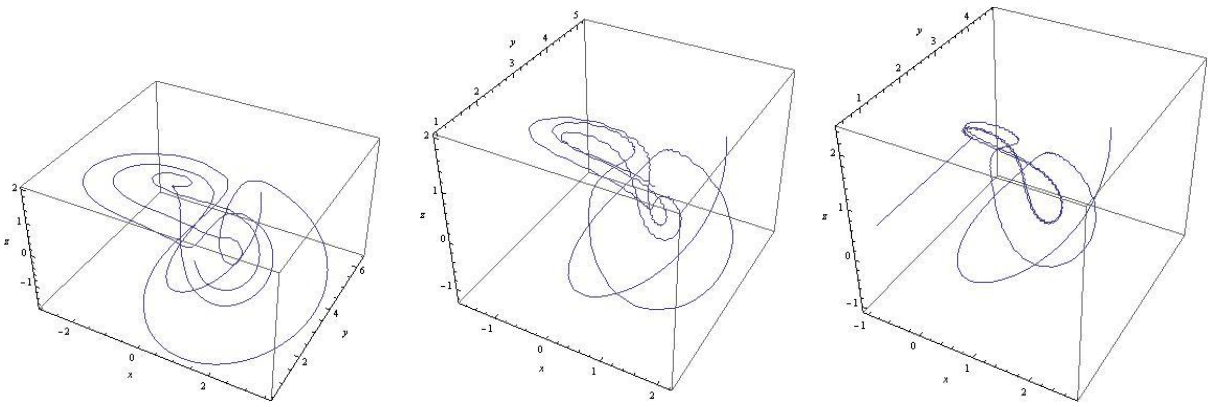
Figure 2 Model (4) with $b_1(t)$, a=3, c=1, x_0=2, y_0=3, z_0=2, t_{max} = 100

Figure 2 show trajectories of modified model (4), when parameter b is replaced by function $b_1(t)$ given by (6), which demonstrates the case that this parameter is not constant within time any more but undergoes scaled periodic forced changes elevated by β_0. The influence of increasing frequency and different elevation is evident.

$$b_1(t) = \beta_1 (\sin(\beta_2 t + \beta_3) + \beta_0) \tag{6}$$

Next, Figure 3 show similar numerical experiments but with shorter time t_{max} and another type of scaled periodic changes given by function (7), which takes non-negative values only.

$$b_2(t) = \beta_1 |\sin(\beta_2 t + \beta_3)| \tag{7}$$



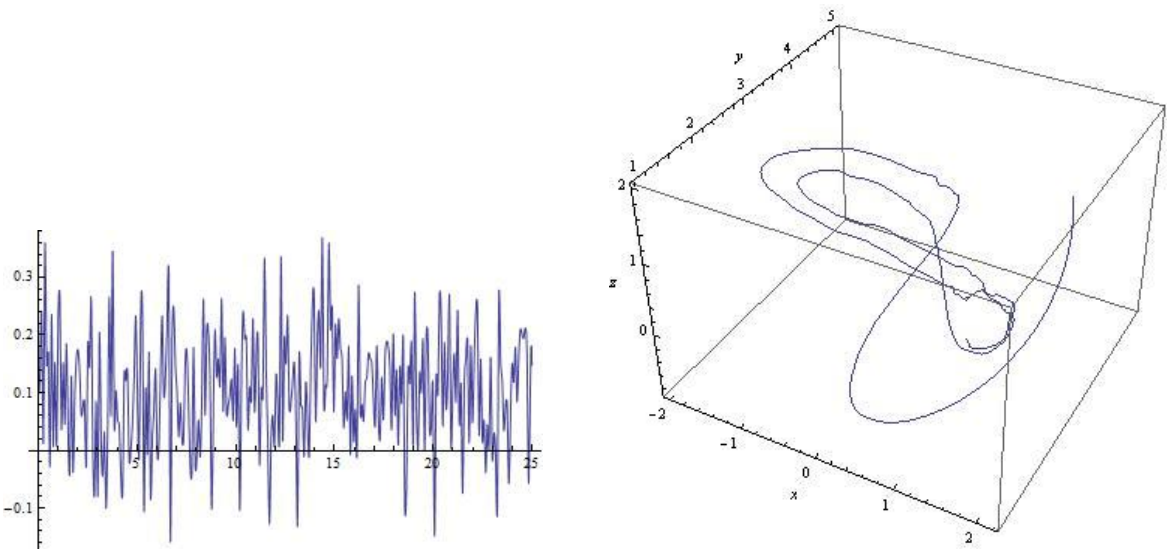
β_1=0.1, β_2 =5, β_3=3, β_0=0 β_1=0.25, β_2 =10, β_3=3, β_0=0.25 β_1=0.28, β_2 =15, β_3=0, β_0=0.45

Figure 3 Model (4) with $b_2(t)$, a=3, c=1, x_0=2, y_0=3, z_0=2, t_{max} = 50

We see different behavior of trajectories, which seems to be nothing else but due to non-negativity of changing values of the parameter b, which is replaced by the function $b_2(t)$, now. Inspecting the right most image of the Figure 3 thoroughly, we may discover rather fine vibrating modes being wound along the middle part of that trajectory, which is quite interesting at all. However, it is also interesting that `NDSolve` has reported exceeding of 10000 steps in numerical integration procedure allowed by default during calculation the given model. Increasing the value of scaling parameter β_1 gets `NDSolve` into serious numerical stability problems, at least when using default options only.

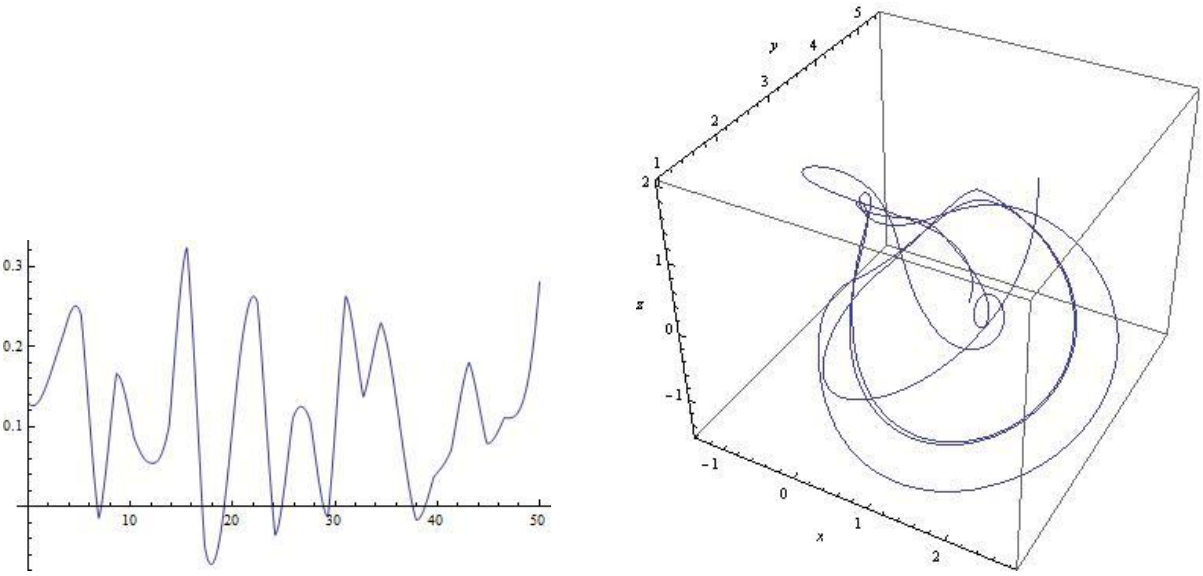
Finally, within this first investigation topic, we have selected still another type of replacement of the parameter b. Instead of deterministic periodic function, we used numerical simulated pseudorandom process built with cubic spline interpolation. Sure, we tried to submit a stochastic process into `NDSolve` directly, but it failed. Hence, we generate a set of n random values B_n sampled from normal distribution with given mean value μ and variance σ^2, first. Then we distribute them uniformly along the given time period [0, t_{max}], and using the Mathematica command `Interpolation` we create a cubic spline interpolant of the generated data. Such function is already acceptable by `NDSolve`. The Figures 4 and 5 show results of numerical experiments with two generated functions of such kind, $b_3(t)$ and $b_4(t)$, respectively. Comparing both results, we might conclude that

much noisy-looked function $b_3(t)$ caused just a low-sized perturbation of the calculated trajectory, whereas $b_4(t)$ produced smooth one.



$b_3(t)$, $t \in [0, t_{max}=25]$, $B_n \in N[\mu=0.1, \sigma^2=0.1]$, $n=300$

Figure 4 Model (4) with $b_3(t)$, $a=3$, $c=1$, $x_0=2$, $y_0=3$, $z_0=2$, $t_{max} = 25$



$b_4(t)$, $t \in [0, t_{max}=50]$, $B_n \in N[\mu=0.1, \sigma^2=0.1]$, $n=30$

Figure 5 Model (4) with $b_4(t)$, $a=3$, $c=1$, $x_0=2$, $y_0=3$, $z_0=2$, $t_{max} = 50$

The second topic we have investigated numerically concerns a short insight upon an influence of initial histories on launching part of trajectories generated by model (5), for $t \in [0,t_p]$, $t_p>0$. Our results are depicted on Figures 6, 7 and 8, respectively. In order to simplify the problem and to enable a comparison with already generated trajectory by model (4), as well, we assume initial histories $u_0(t)$, $v_0(t)$, $w_0(t)$ to take the following form (8).

$$u_0(t) = x_0\eta(t), \quad v_0(t) = y_0\eta(t), \quad w_0(t) = z_0\eta(t), \quad t \in [-\tau,0], \quad \tau_i = \tau, \; i=1,2,3, \; \tau>0, \quad \eta(0)=1 \tag{8}$$

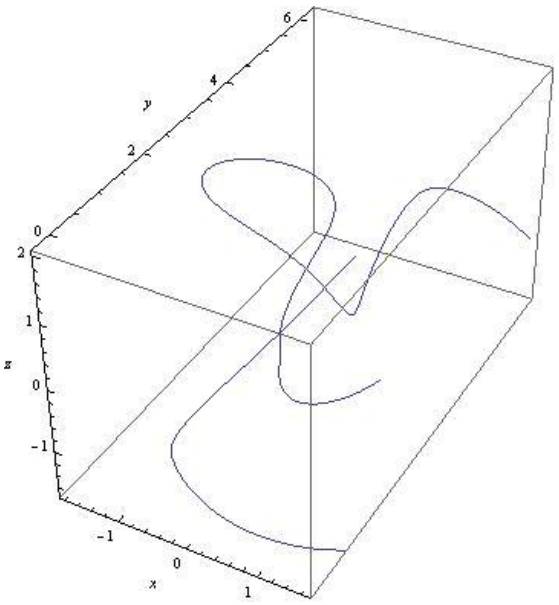
In that case, we select the function $\eta(t)$ only. All computations were performed by `NDSolve` again, which underlines its tremendous versatility. We make choice of five relatively simple classes of functions, which are listed below in formulas (9).

$$\begin{gathered} \eta_1(t)=\exp(\gamma_1 t), \quad \eta_2(t)=\cos(\gamma_2 t), \quad \eta_3(t)=\exp(\gamma_{3,1} t)\cos(\gamma_{3,2} t), \\ \eta_4(t)=1+\gamma_4 t, \quad \eta_5(t)=1+\gamma_{5,1} t + \gamma_{5,2} t^2, \quad t \in [\tau,0], \quad \gamma_i, \gamma_{j,1}, \gamma_{j,2} \in R, \; i=1,2,4, \; j=3,5 \end{gathered} \tag{9}$$

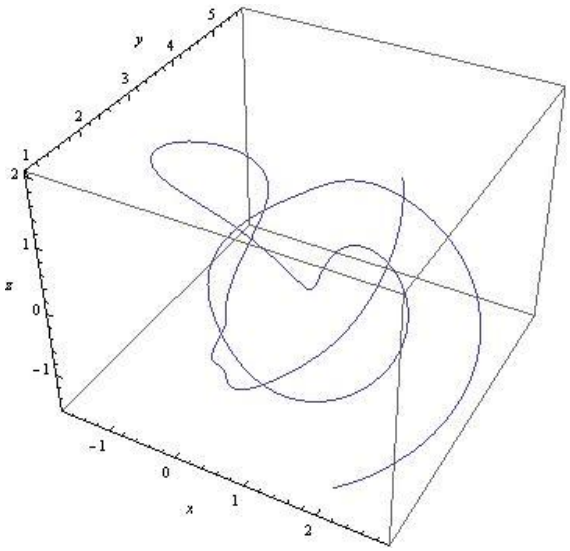
Hence, the initial histories of model (5) are given simply within `NDSolve` by following expressions (10).

$$x[t/;t\leq 0]==x0\eta[t],\ y[t/;t\leq 0]==y0\eta[t],\ z[t/;t\leq 0]==z0\eta[t] \tag{9}$$

where `x0, y0,` and `z0` stand for initial values x_0, y_0, z_0,and function $\eta[t]$ takes one of $\eta_k(t)$, k=1,.., 5. For all computations we set $t_p = 10\tau$, thus defining the period $[0,t_p]$ to be ten times longer than $[-\tau,0]$, the period of initial histories given, with τ =2. The feedback intensities are k_1=0.1, k_2=0.2, and k_3=0.1. Initial values x_0, y_0, and z_0 are set traditionally as follows x_0=2, y_0=3, z_0=2. In a similar way, we take values of constitutive parameters a=3, b=0.1, and c=1, which maintains the model (5) to be compatible with (4) one.

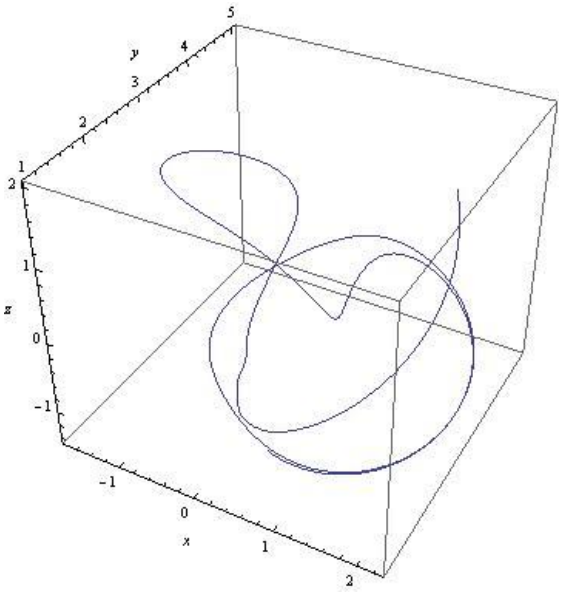


$\eta(t) = \eta_1(t)$, γ_1=5

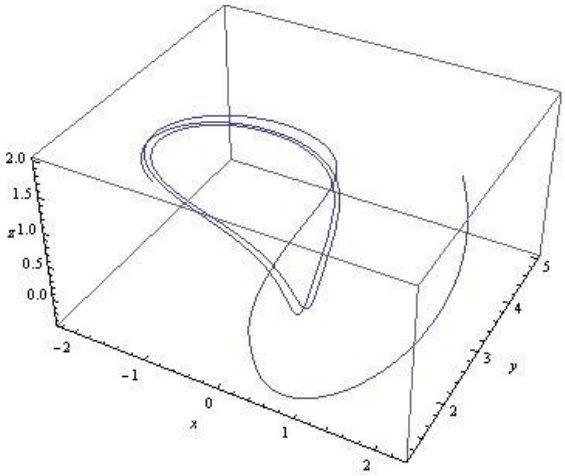


$\eta(t) = \eta_2(t)$, γ_2=10

Figure 6 Model (5)



$\eta(t) = \eta_3(t)$, $\gamma_{3,1}$=2.5, $\gamma_{3,2}$=15

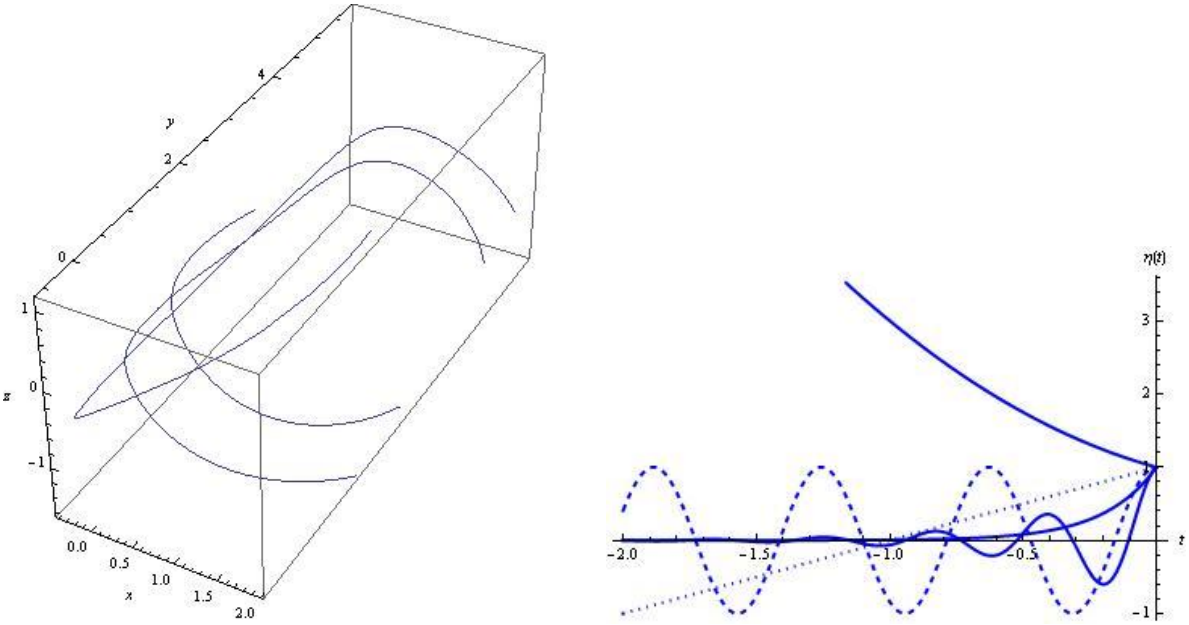


$\eta(t) = \eta_4(t)$, γ_4=1

Figure 7 Model (5)

Inspecting these results, we may just confirm expected dependence of starting parts of trajectories upon initial histories. Any of five trajectory images of model (5) can be directly compared with the trajectory generated by model (4), already presented on the left side of Figure 1, with similar constitutive and initial conditions to see the differences. Finally, we have to point out that presented numerical investigations form just a small fraction of numerical experiments we have already conducted with this dynamic nonlinear financial model in its basic simplified form, as well as with proposed generalizations. The full parametric space of the basic model (4) is R^3xR^3

for three constitutive parameters and three initial conditions. The complexity of model (5) increases dramatically, because of added feedbacks and initial history functions instead of just three numbers x_0, y_0, z_0.



$\eta(t) = \eta_5(t)$, $\gamma_{5,1} = -1$, $\gamma_{5,2} = 1$ functions $\eta_k(t)$, k=1,.., 5

Figure 8 Model (5)

4 Conclusions

We hope the large possibilities of numerical solution of nonlinear dynamic financial system with chaotic behavior using Mathematica command NDSolve were sketched. Proposed and discussed replacement of constitutive parameter b to be time dependent taking form of simple periodic function, pseudorandom process, etc., instead of constant one within the basic 3-D model (4) can be simply extended to other two parameters a and c, too.

Further research on the topic seems to be very challenging and will be focused both on complex program of numerical investigation of large parametric space and advanced theoretical development including analysis of bifurcation, as well as empirical verification of the models.

Acknowledgements

The research project was supported by Grant SGS2012-036 of the Faculty of Economics, University of West Bohemia in Pilsen.

References

[1] Cai, G., Huang, J.: A New Finance Chaotic Attractor. *Int.J. Nonlinear Science*, 3 (2007), 3, 213-220. http://worldacademicunion.com/journal/1749-3889-3897IJNS/IJNSVol3No3Paper08 .

[2] Chen, W.-C.: Nonlinear dynamics and chaos in a fractional-order financial system. Chaos Solitons Fractals, 36 (2008), 1305-1314.

[3] Ma, J.-H., Chen, Y.-S.: Study for the bifurcation topological structure and the global complicated character of a kind of nonlinear finance system (I). *Appl. Math. Mech.-Engl.* 22 (2001), 11, 1240-1251.

[4] Ma, J.-H., Chen, Y.-S.: Study for the bifurcation topological structure and the global complicated character of a kind of nonlinear finance system (II). *Appl. Math. Mech.-Engl.* 22 (2001), 12, 1375-1382.

[5] Sauer, T.: Numerical Solution of Stochastic Differential Equations in Finance. http://math.gmu.edu/~tsauer/pre/sde.pdf .

[6] Son, W.-S., Park, Y.-J.: Delay effect on the dynamical model of financial system. http://www.phys.sinica.edu.tw/~socioecono/econophysics2010/pdfs/SonPaper.pdf .

[7] Yang, M., Cai G.: Chaos Control of a Non-linear Finance System. *JournalUncertain Systems*, 5 (2011), 4, 263-270. http://www.worldacademicunion.com/journal/jus/jusVol05No4paper03.pdf .

[8] Zhang, R.: Bifurcation Analysis for a Kind of Nonlinear Finance System with Delayed Feedback and Its Application to Control of Chaos. Hindawi Publ.Corp., *J.Appl.Math.*, 2012, Article ID 316390, 18 pp.