TECHNICAL UNIVERSITY OF LIBEREC
FACULTY OF ECONOMICS
Studentská 2, 461 17 Liberec

Proceedings of 26th International conference

MATHEMATICAL METHODS IN ECONOMICS 2008

September 17 – 19, 2008

Editors: Ing. Pavla Řehořová, Ph. D.
Ing. Kateřina Maršíková, Ph. D.
Ing. Zbyněk Hubínka (typesetting & conversions to TEX)
Foto: Jan Hastrdlo

The publication doesn't pass the language correction.

Papers are sorted by authors' names in alphabetical order.

ISBN: 978-80-7372-387-3

Contents

Jozef Baruník, Lukáš Vácha

Institute of Information Theory and Automation

Academy of Sciences of the Czech Republic

Pod vodárenskou věží 4, 182 08 Praha 8

e-mail: `barunik@utia.cas.cz, vachal@utia.cas.cz`

Neural Networks with Wavelet Based Denoising Layer: Application to Central European Stock Market Forecasting[1]

Abstract

Traditional prediction methods for time series often restrict on linear regression analysis, exponential smoothing, and ARMA. These methods generally produce reasonable prediction results for stationary random time series of linear systems. In the recent decades, development in econometrics resulted also in methods which are capable of forecasting more complex systems, such as Wavelet decomposition or Neural Networks. These methods proved to better explain the complex stock market behavior. In this paper we apply neural network with wavelet denoising layer method for forecasting of Central European Stock Exchanges, namely Prague, Budapest and Warsaw. Hard threshold denoising with Daubechies 6 wavelet filter and three level decomposition is used to denoise the stock index returns, and two-layer feed-forward neural network with Levenberg-Marquardt learning algorithm is used for modeling. The results show that wavelet network structure is able to approximate the underlying process of considered stock markets better that multi-layered neural network architecture without using wavelets. Further on we discuss the impact of structural changes of the market on forecasting accuracy on the daily stock market data. Stock markets change their structure rapidly with changing agent sentiment structure. These changes then have great impact on the prediction accuracy.

Keywords: *neural networks, hard threshold denoising, time series prediction, wavelets.*

1 Introduction

Traditional prediction methods for time series often restrict on linear regression analysis, exponential smoothing, and ARMA. These methods generally produce reasonable prediction results for stationary random time series of linear systems. In the recent decades, development in econometrics brought also methods which are capable of forecasting more complex systems, such as stock markets. These mainly include Wavelet Decomposition [10], [4], [1] and Neural Networks analysis, [5], [9]. Further on, the idea of combining both methods into single wavelets

[1]A support from the Grant Agency of Charles University (GAUK) under grant 46108 and Czech Science Foundation (GAČR) under grants 402/08/P207 and 402/06/1417

and neural networks has resulted in the formulation of wavelet networks. This research area is new, and there is tremendous potential for its development and application to various fields. This paper is one of the first attempts to fit this exciting method to Central European Stock markets, namely Prague Stock Exchange, Budapest Stock Exchange and Warsaw Stock Exchange. Our main expectation will be that this method will improve the single neural network approach.

2 Denoising with Wavelets

Classical time series denoising approaches are rooted in Fourier analysis where noise is assumed to be represented mainly as high frequency oscillations. The wavelet based denoising, assumes that analysis of time series at different resolutions might improve the separation of the true underlying signal from noise. Let us begin with description of the basics of Wavelet decomposition theory.

Wavelets There are two types of wavelets: father wavelets ϕ and mother wavelets ψ . The father wavelet integrates to unity and the mother wavelet integrates to zero. The father wavelet, also called scaling function, essentially represents the smooth, trend, i.e., the low frequency part of the signal, on the other hand the mother wavelet represents the details, i.e., the high frequency part of the signal. The mother wavelet is compressed or dilated in time domain, to generate cycles to fit the actual time series. The formal definition of the father ϕ and mother ψ wavelet is

$$\phi_{j,\,k} = 2^{-\frac{j}{2}}\phi\left(\frac{t-2^{j}k}{2^{j}}\right), \quad \psi_{j,\,k} = 2^{-\frac{j}{2}}\psi\left(\frac{t-2^{j}k}{2^{j}}\right)$$

where j is the scale (or dilatation) and k is the translation (or shift). Commonly, many types of wavelets can be possibly used, including Haar wavelet, Mexican hat, Morlet wavelet, Daubechies wavelet etc. In the empirical part, we will use Daubechies wavelet db6.

Any time series $x(t)$ can be built up as a sequence of projections onto father and mother wavelets indexed by both j , the scale, and k , the number of translations of the wavelet for any given scale. Usually k is assumed to be dyadic. The wavelet coefficients are approximated by integrals

$$s_{J,\,k} \approx \textstyle\int_{-\infty}^{\infty} x\left(t\right)\,\phi_{J,\,k}\left(t\right)\,dt, \quad d_{j,\,k} \approx \textstyle\int_{-\infty}^{\infty} x\left(t\right)\,\psi_{j,\,k}\left(t\right)\,dt$$

$j = 1,\,2,\ldots,\,J$, where J is the maximum scale. The wavelet representation of the time series $x(t)$ in $L^2(R)$ [2] can be given by

$$x\left(t\right) = \sum_{k} s_{J,\,k}\phi_{J,\,k}\left(t\right) + \sum_{k} d_{J,\,k}\psi_{J,\,k}\left(t\right) + \sum_{k} d_{J-1,\,k}\psi_{J-1,\,k}\left(t\right) + \ldots + \sum_{k} d_{1,\,k}\psi_{1,\,k}\left(t\right)$$

where the basis functions $\phi_{J,\,k}\left(t\right)$ and $\psi_{j,\,k}\left(t\right)$ are assumed to be orthogonal. When the number of observations is dyadic, the number of the wavelet coefficients of each type at the finest scale 2^1 is $N/2$, labeled $d_{1,\,k}$. The next scale 2^2 has $N/2^2$ coefficients, labeled $d_{2,\,k}$. At the coarsest scale 2^J there are $N/2^J$ coefficients $d_{J,\,k}$ and $s_{J,\,k}$.

[2] Square integrable real-valued function, $\int_{-\infty}^{\infty} x^2\left(t\right)\,dt < \infty$.

3 Nonlinear Wavelet Denoising

The simplest method of nonlinear wavelet denoising is via thresholding. The procedure sets all wavelet coefficients that has lower value than some fixed constant to zero. Two thresholding rules were instrumental in the initial development of wavelet denoising, both for their simplicity and performance: hard and soft thresholding [4].

Threshold Selection Optimal thresholding occurs when the threshold is set to the noise level, i.e., $\eta = \sigma_{\in}$. Setting $\eta < \sigma_{\in}$ will allow unwanted noise to enter the estimate while setting $\eta > \sigma_{\in}$ will destroy information that belongs to the underlying signal. Following [3] we can set a universal thresholding as

$$\eta^U = \widehat{\sigma}_{\in}\sqrt{2\log N},$$

where N is the sample size. In practical situations the standard deviations of noise $\sigma_{\in}$ is not known. The most commonly used estimator of $\sigma_{\in}$ is the maximum absolute deviation (MAD) standard deviation [10].

$$\widehat{\sigma}_{MAD} = \frac{median\left(|d_{1,1}|, |d_{1,2}|, ..., |d_{1,N/2-1}|\right)}{0.6745}.$$

The denominator is needed to rescale the numerator so that $\widehat{\sigma}_{MAD}$ is tuned to estimating the standard deviation for Gaussian white noise [4].

Hard Thresholding In our paper we use a hard thresholding. The hard thresholding rule on the wavelet coefficients o_t is given by

$$\delta_\eta^H(o_t) = \{ \begin{array}{c} o_t \text{ if } |o_t| > \eta \\ 0 \text{ otherwise} \end{array},$$

where η is the threshold value. The operation is not a continuous mapping, it only affects input coefficients that are less or equal to the threshold η. After obtaining the thresholded wavelet coefficients using δ_η^H we compose the denoised times series via an inverse wavelet transform (IDWT) so we get $\widehat{x}_{den}(t)$. For a more detailed treatment see [10].

4 Wavelet Network Structure

Wavelet Network is a network combining the ideas of the feed-forward neural networks and the wavelet decomposition. Wavelet networks use simple wavelets and wavelet network learning is performed by the standard type algorithm such as Conjugate-Gradient, or more efficient Levenberg-Marquardt [6], [8]. Neural networks can be viewed as universal approximation tools for fitting linear or nonlinear models, as [5] showed. Limiting space of this paper do not allow us to explore Neural Networks estimation methodology in detail, but reader is advised to follow i.e. [12], or [9] for very good explanation.

There are basically two main approaches to form wavelet networks. In the first approach, the wavelet decomposition is decoupled from the learning component of neural network architecture. In other words, the series are firstly decomposed / denoised using wavelets, and then fed to the neural network. In the second approach, the wavelet theory and neural networks are combined into a single method, where the inputs $x_1, ..., x_k$ with weights $\omega_1, ..., \omega_n$ are combined to estimated output in Multilayer feed-forward network (MPL):

$$\widehat{x}_{DWNN}(t) = \sum_{i=1}^{N} \omega_i f(\gamma_i x(t) + \beta_i),$$

where f is an activation function, γ_i, β_i, ω_i are network weight parameters that are optimized during learning, and N is number of hidden layers. If we feed this network with nonlinear wavelet denoised values $\widehat{x}_{den}(t)$(see section 3) we will get the form of the wavelet neural network (DWNN) used in this paper.

5 Results

In the testing, we focus on sample of 1050 daily returns from 7.1.2004 to 3.4.2008 of value-weighted indices PX-50, BUX and WIG (Prague, Budapest and Warsaw stock exchanges respectively). The dataset was downloaded from the server www.stocktrading.cz. In the prediction task we start with denoising of the 512 data sample with db6 wavelet filter and a 3-level decomposition. Then two-layer neural network with 5 neurons in each layer and Levenberg-Marquardt learning algorithm is used to learn the sample. To avoid over-fitting of the network we use a window of 50 real out-of-sample data on which we test the estimated model on one day predictions. This algorithm is repeated 10 times with moving window of 50, so the final prediction of 500 data is obtained. Finally we compare this method to neural network approach so we can see if the wavelet denoising layer improves the forecasts. Architecture of the network is again two layers with 5 neurons and Levenberg-Marquardt optimization.

As for evaluation, we focus mainly on out-of-sample performance, as it is most important in financial time series forecasting. We consider Root Mean Square Error statistics (RMSE) to see the performance of out-of-sample prediction. Further on, we use statistics proposed by Pesaran Timmerman – SR (PT) [11], which evaluates the correctness of the signs prediction. Such statistics is often used in financial literature as the predicted positive change predicts buy signal, negative change sell signal which allows evaluating a trading strategy. Pesaran Timmerman statistics is based on the null hypothesis that a given model has no economic value in forecasting direction and is approximately normally distributed. In other words, we test the null hypothesis that the signs of the forecasts and the signs of actual variables are independent. If the prediction of signs is statistically dependent, we approached a good forecasting model with economic significance.

Finally to test the performance of the wavelet network and the neural network approach we use the statistic proposed by Clark and McCracken (CM) [2]. They compare out-of-sample accuracy for two models which are nested. The statistics is normally distributed under the null hypothesis of equal predictive ability of the two models.

To compare the results, we can see that Wavelet Neural Networks (DWNN) performed best on BUX returns with lowest RMSE, while RMSE of PX50 was little bit higher, and WIG

	PX50		BUX		WIG	
	MPL	DWNN	MPL	DWNN	MPL	DWNN
RMSE	0.8	0.175	0.08	0.11	0.37	0.41
CM	0.78***		0.25***		-7.6	
SR (PT)	0.52	0.56**	0.51	0.53***	0.47	0.49
*, **, ***, 1%, 5% and 10% significance levels						

Table 1: Prediction results for PX50, BUX and WIG

more than double of PX50 and BUX. This would indicate that the DWNN model will forecast the PX50 and BUX returns better than WIG. This is confirmed by Pesaran Timmerman statistics which is significant on 5% level of significance for PX 50, 10% level of significance of BUX but is not significant at all for WIG. Success rate of correct forecasted direction is 0.56 for PX50, 0.53 for BUX and 0.49 for WIG. As to the Multilayer Feedforward architecture (MPL) without denoising the directional accuracy is not significant at all. But if we compare the DWNN and MPL using Clark and McCracken statistics, we can see that DWNN yields significantly better prediction accuracy for PX50 and BUX series. For WIG the results are not significantly different, which would be expected as WIG does not seem to be significantly predictable even using Pesaran Timmerman.

To sum up the achieved results we can say, that DWNN performs significantly better on the PX50 and BUX returns prediction. Although reader can notice, that the success rate is quite low. This is probably caused by quite large data sample which contains large structural changes as recent stock market crash of February 2008, etc. Even though Wavelet Neural Networks are considered as a universal approximation theorem as mentioned before, reader can see that if we feed it with data which simply cannot be approximated, its performance is poor. As the stock market structure changes in time quite quickly, testing the prediction on such a large data samples does not seem to provide reasonable predictions. Thus our last test is focused on the division of the dataset into 3 month moving windows, where we simply look at how does the Success Rate statistics of Pesaran Timmerman evolve in time.

Figure 1 demonstrates that our assumption was right. For some periods, the statistics for PX50 and BUX returns is 0.6 to 0.7, which means that 60% to 70% of the one day sign change is predicted correctly. For WIG returns the predictability does not excess 55%, which can lead us to suspicion that WIG simply does not contain strong predictable patterns. As to other two tested series, we can see that the pattern strongly evolves over time, and that it would make sense to adjust appropriate method for forecasting each 3-month. This would be done using adjusting wavelets and their levels, number of hidden layer of neural network, etc. which we leave for further research.

Conclusion

Our results indicate that the Wavelet Neural Network might outperform simple Neural Networks while forecasting Central European stock exchanges. More concretely, PX50 and BUX returns were predicted using DWNN structures significantly better than using MPL Network structure. We also conclude that the stock market structural changes affected the final stock

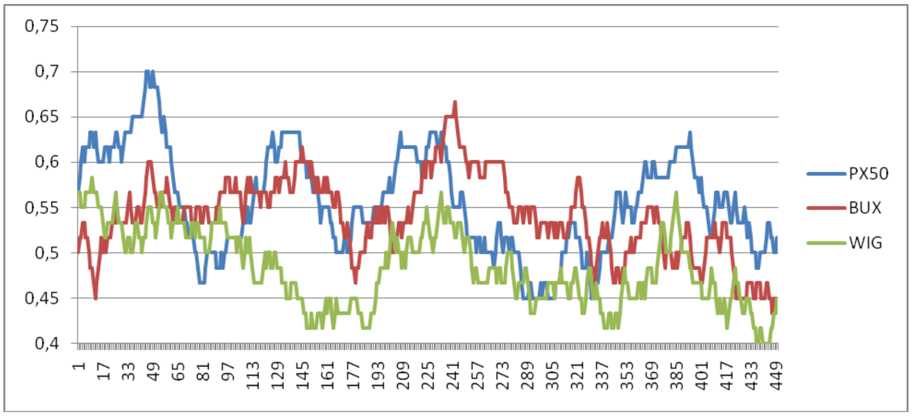


Figure 1: Success Rate statistics of Pesaran Timmerman of PX50, BUX and WIG in time.

market direction prediction greatly. Dataset used for testing was quite large and contained structural changes such as large market crash of February 2008 which leads to significantly lower prediction accuracy of the used methods. For this reason we also used the three month moving window using which we have showed, that for some periods the prediction accuracy reached sustainable 60% to 70%, meaning 60% to 70% future directions of the stock markets were predicted correctly using PX50 and BUX data. On the other hand, the prediction accuracy using WIG data did not improve a lot. Thus we conclude that this market does not simply contain stronger predictable patterns. Further research should concentrate on the exploring of dynamically adjusted wavelet types, number of hidden layers of neural networks or other parameters specific to structural changes in the forecasted underlying series.

References

[1] Abramovich F., Bailey T.C., Sapatinas T. (1999), *Wavelet Analysis and its Statistical Applications*, J. R. Statist. Soc. D, 48.

[2] Clark, Tod E., and Michael W. McCracken (2001): *Tests of Forecast Accuracy and Encompassing for Nested Models*, Journal of Econometrics 105: 85-110.

[3] Donoho, D. L. and Johnstone, I. M. (1994) *Ideal Spatial Adaptation by Wavelet Shrinkage.* Biometrica, 81, 422-455.

[4] Gençay, R., Selçuk, F., Whitcher, B. (2002), *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics.* Academic Press.

[5] Hornik, K., (1989): *Multilayer Feedforward Networks are universal Approximators,* Neural Networks 2, 359-366.

[6] Levenberg, K. (1944): *A Method for the Solution of Certain Problems in Least Squares.* Quart. Appl. Math. 2, 164-168

[7] Mallat S. (1998), *A Wavelet Tour of Signal Processing,* Academic Press

[8] Marquardt, D. (1963): *An Algorithm for Least-Squares Estimation of nonlinear Parameters.* SIAM J.Appl.Math. 11, 431-441.

[9] McNelis, P.D. (2005): *Neural Networks in Finance: Gaining predictive edge in the markets,* Elsevier Academic Press, ISBN 0-12-485967-4

[10] Percival, D. B., Walden, A. T. (2000), *Wavelet Methods for Time series Analysis.* Cambridge University Press.

[11] Pesaran, M.H., and A.Timmermann (1992): *A Simple Nonparametric test of Predictive Performance,* Journal of Business and Economic Statistics 10: pp. 461-465

[12] Poggio, T. and F. Girosi (1990): *Networks for Approximation and Learning.* Proc. of the IEEE, vol. 78, no.9, pp. 1481-1497.

Jozef Baruník, Miloslav Vošvrda

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4, 182 08 Praha 8

Institute of Economic Studies
Faculty of Social Sciences
Charles University Prague
Smetanovo nábř. 6, 110 00 Praha 1

e-mail: vosvrda@utia.cas.cz, barunik@utia.cas.cz

Cusp Catastrophe Theory: Application to U.S. Stock

Abstract

We show that the cusp catastrophe model explains the crash of stock exchanges much better than alternative linear and logistic models. On the data of U.S. stock markets we demonstrate that the crash of October 19, 1987 may be better explained by cusp catastrophe theory, which is not true for the crash of Sept. 11, 2001. With the help of sentiment measures, such as index put/call options ratio and trading volume (the former models the chartists, while the latter the fundamentalists), we have found that the 1987 returns are clearly bimodal and contain bifurcation flags. The cusp catastrophe model fits these data better than alternative models. Therefore we may say that the crash may have been led by internal forces. However, the causes for the crash of 2001 are external, which is also evident in much weaker presence of bifurcations in the data. Thus alternative models may be used for its explanation.

Keywords: *cusp catastrophe, bifurcations, singularity, nonlinear dynamics, stock market crash*

JEL *C01, C53*

1 Introduction

Unexpected stock market crashes has been a nightmare for the financial world ever since the capital market existed. The catastrophe theory attempts to unfold a part of information we might need to understand the crash phenomenon. It describes how small, continuous changes in control parameters, or independent variables influencing the state of the system, can have sudden, discontinuous effects on dependent variables. In the paper, we apply the theory to sudden stock market changes that are known as crashes. Zeeman [7] was the first to qualitatively describe the *"unstable behavior of stock exchanges"* by Thom [5] catastrophe theory. We extend his ideas by incorporating quantitative analysis.

The article is rather empirical as it puts the theory to test on financial data. As only a few papers deal with an empirical analysis of catastrophe theory, this paper may contribute to this research. We build on the Zeeman's qualitative description, and primary aim of the research is to answer the question of whether catastrophe models are capable of indicating the stock market crashes.

What we regard as the most significant aspect is testing on the real-world financial data. Our key assertion is that the cusp catastrophe model is able to fit the data more properly than an alternative linear regression model, and/or nonlinear (logistic) model. We fit the catastrophe model to the data of October 19, 1987 crash, known as Black Monday which was the greatest single-day loss (31%) that Wall Street has ever suffered in continuous trading. As for comparison, we use another large crash, that of September 11, 2001. The final part is devoted to the assumption that while in 1987 the crash was caused by internal forces, in 2001 it was external forces, namely 9/11 terrorist attack. Thus the catastrophe model should fit the data of 1987 well, as the bifurcations leading to instability are present. However, it does not seem to perform better than linear regression on the 2001 data. As the control variables we use the measures of sentiment, precisely OEX[1] Put/Call ratio which appears to be very good measure of the speculative money in the capital market, against trading volume as good proxy for large, fundamental investors.

2 The Cusp Catastrophe Model

Let us assume one dependent variable Y, and a set of n independent variables $\{X_1, X_2, ..., X_n\}$. Then y represents realization of a random variable Y, and x_i represents realizations of X_i. To obtain greater flexibility than using linear regression technique, $2n+2$ additional degrees of freedom are introduced. This could be done by defining control factors $\alpha_x = \alpha_0 + \alpha_1 x_1 + ... + \alpha_n x_n$ and $\beta_x = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n$. These factors determine the predicted values of y given realizations of $\{x_1, x_2, ..., x_n\}$, meaning that for each value x there might be three predicted values of the state variable. The predictions will be roots of the following canonical form

$$0 = \alpha_x + \beta_x (y-\lambda)/\sigma - ((y-\lambda)/\sigma)^3 \,, \tag{1}$$

which describes the cusp catastrophe response surface containing a smooth pleat. λ and σ are location and scale parameters. In the literature on catastrophe theory, α_x and β_x are so called *normal* and *splitting* factors, however, we prefer the notions a*symmetry* and *bifurcation* factors, respectively. Hence the statistical estimation problem is to find the estimates for the $2n+4$ parameters: $\{\lambda, \sigma, \alpha_0, ..., \alpha_n, \beta_0, ..., \beta_n\}$ from n observations of the $n+1$variables $\{Y, X_1, ..., X_n\}$.

2.1 Stochastic Dynamics and Probability Density Function (PDF)

Let y_t be the function of time t for $t \in \langle 0, T \rangle$. From a dynamic system's point of view, Equation (1) can be considered as the surface of the equilibrium points of a dynamic system of the state variable y_t which follows the ordinary differential equation $dy_t = g(x, y_t)dt$, where $g(x, y_t)$ is the right hand side of Equation (1). For real world applications, it is necessary to add a non-deterministic behavior into the system, as the system usually does not determine

[1]OEX are options with the Standard & Poor's 100 Index underlying

its next states entirely. We may obtain a stochastic form by an adding of the Gaussian white noise term[2]. The system is then described by a stochastic differential equation of the form

$$dy_t = \left(\alpha_x + \beta_x \frac{(y_t - \lambda)}{\sigma_{y_t}} - \left(\frac{(y_t - \lambda)}{\sigma_{y_t}} \right)^3 \right) dt + \sigma_{y_t} dW_t, \tag{2}$$

and $\sigma_{y_t}^2$ is an instantaneous variance of the process y_t. The W_t is a standard Wiener process and $dW_t \sim N(0,\, dt)$. Hartelman [4] has established a link between a deterministic function of catastrophe system and a pdf of the corresponding stochastic process. He showed that the pdf $f\,(y_t)$ will converge in time to a pdf $f_S\,(y_\infty)$ corresponding to a limiting stationary stochastic process. This has led to a definition of stochastic equilibrium state and bifurcation which is compatible with their deterministic counterpart. Instead of fitting the deterministic process where the equilibrium points of the system are of a main interest, the attention is drawn to relative extremes of the conditional density function of y. Following Wagenmarkers [6], the pdf of y is:

$$f_S\,(y_\infty|x) = \xi \exp \left[\alpha_x y_\infty + \frac{\beta_x}{2} \left(\frac{y_\infty - \lambda}{\sigma_{y_\infty}} \right)^2 - \frac{1}{4} \left(\frac{y_\infty - \lambda}{\sigma_{y_\infty}} \right)^4 \right] \tag{3}$$

The constant ξ normalizes the pdf so it has unit integral over its range. The modes and antimodes of the cusp catastrophe pdf can be obtained by solving the equation $df_S(.|.)/dy_. = 0$, which will yield exactly implicit cusp surface equation - Equation (1). The parameters will be estimated by method of estimation developed by Hartelman [4], Wagenmarkers [6].

As β_x changes from negative to positive, the pdf $f_S\,(y_\infty|x)$ changes its shape from unimodal to bimodal. It is also the reason why the β_x factor is called *bifurcation* factor. For $\alpha_x = 0$, the pdf is symmetrical, other values control asymmetry, thus α_x is *asymmetry* factor. Thoughtful reader has certainly noted that catastrophe theory models are an extension to traditional models, therefore they have to satisfy the requirement of the empirical testability. It should be remembered, that there is no single statistical test for acceptability of the catastrophe model. Due to the multimodality of cusp catastrophe, traditional measure for goodness of fit cannot be used. Considered residuals can be determined only if the probability density function at time t is one-peaked, and as the model generally offers more than one predicted value, it is difficult to find a tractable definition for a prediction error. In testing we follow Hartelman [4] approach. A comparison of the cusp and a linear regression model is made by means of a likelihood ratio test, which is asymptotically chi-squared distributed with degrees of freedom being equal to the difference in degrees of freedom for two compared models. As it may not be sufficient to reliably distinguish between catastrophe and non-catastrophe models, Hartelman [4] compares catastrophe model also to a nonlinear logistic model. As the cusp catastrophe model and the logistic model are not nested, Akaike information criterion (AIC), and Bayesian information criterion (BIC) statistics are used in a testing routine to compare the models.

[2] Cobb[1], Cobb, Watson[2], Cobb, Zacks[3]

3 Empirical Testing

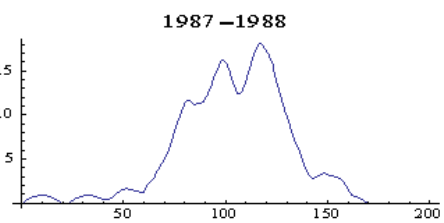
3.1 Data Description

We primarily test the model on the set of daily data which contains most discussed stock market crash of October 19, 1987, known as Black Monday. The crash was the greatest single-day loss that Wall Street had ever suffered in continuous trading, 31%. The reasons for Black Monday have been widely discussed among professional investors and academics. However, not until today is there a consensus on the real cause. For comparison, we use another large crash, that of September 11, 2001. Our assumption is that while in 1987 the crash was caused by internal forces, the 2001 crash happened due to external force, namely the terrorist attack on the twin towers. Therefore the catastrophe model should fit the data of 1987 well as bifurcations leading to instability are present.

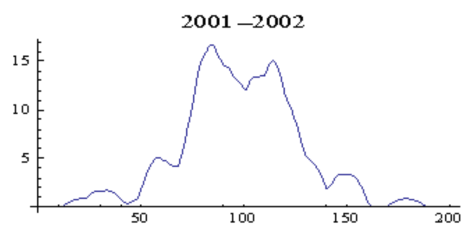
The data represents the daily returns of S&P 500 in the years 1987-1988 and 2001-2002 as the crashes took place inside these intervals. For the asymmetry side, we have chosen the daily change of down volume representing the volume of all declining stocks in the market. The trading volume represents good measure of the fundament, as it correlates with the volatility, and more importantly, good measure of what the large funds, representing fundamental investors, are doing. For bifurcation side OEX Put/Call ratio represents very good measure of speculative money. It is a ratio of daily put volume divided by daily call volume of the options with underlying Standard and Poor's 100 index. As financial options are the most popular vehicle for speculation, it represents the data of speculative money, while extraordinary biased volume or premium suggests excessive fear or greed in the stock market. These should be internal forces which causes the bifurcation.

3.2 Results

All the data are differenced once in order to gain stationarity. It can be seen that the data are leptokurtic, and much more interestingly, multimodal. For illustration of bimodality, we use kernel density estimation - see Graphs 1 and 2 (we use Epanechnikov kernel which is of following form: $K(u) = \frac{3}{4}\left(1-u^2\right)(|u| \leq 1)$ with smoother bandwidth so the bimodality can be seen):



Graph 1



Graph 2

Kernel density of the 2 year returns of 1987 and 1988 shows clear bimodality, and so does the kernel density of the second set of the data, i.e. years 2001 and 2002. The first test we consider is Hartelman's test for multimodality. It is evident from the previous figures that the returns are far from being unimodal. However as noted in Wangenmakers [6], there may occur inconsistencies between the pdf and the invariant function with respect to the number of stable states: examples of which can be found in Wangenmakers [6]. Thus, we make use of the proposed Hartelman's kernel program to test for the multimodality and we have found that there is 75% probability that the 1987-1988 data contains at least one bifurcation point, and 26% probability that the the 2001-2002 data contains at least one bifurcation point. These results are also consistent with our assumption, that the first crisis was drawn by internal market forces (c.f. the presence of the bifurcations in the data), whereas the 2001 crash was caused mainly due to external forces, 9/11 attack.

Encouraged by the knowledge that bifurcations are present in our datasets we can now move to cusp fitting. As has been mentioned before, we use Hartelman's cuspfit software[3] for this purpose. The methodology is simple. First, the linear, nonlinear (logistic) and the cusp catastrophe models have been fitted to the data. Then we have tested whether the cusp catastrophe model fits the data better than the other two models by the procedure described at the beginning of the empirical part of this paper. We have obtained the following results:

model	linear	logistic	**cusp**
R^2	0.1452	0.2558	**0.4025**
log likehood	-6.09×10^3	-5.17×10^2	$\mathbf{-4.95 \times 10^2}$
AIC	1.23×10^3	1.05×10^3	$\mathbf{1.00 \times 10^3}$
BIC	1.24×10^3	1.07×10^3	$\mathbf{1.03 \times 10^3}$
parameters	4	5	6

Table 1: Results of the fits to the 1987 – 1988 data

In Table 1 there are the results of the cusp fit to the data of 1987-1988 which contains the crash of October 19, 1987. We can see that log likelihood is largest for the cusp catastrophe model. Chi-squared test, Akaike and Schwarz-Bayesian information criteria also favor the catastrophe model, and *R2* is again much better for the cusp catastrophe. Thus we can conclude that the cusp catastrophe model offers a more suitable explanation for the 1987 stock market crash. We believe that the quality of the fit arises from the choice of the variables. We have also tried other possible variables in order to explain the bifurcations, but none has proved as successful. The choice of the variables is logical as the tests for the bifurcations in the data confirmed their presence.

Let us have a look at the second set of the data that of years 2001-2002. The results are in Table 2, and we can see, that the catastrophe model in this case is rather superfluous. The log likelihood is greater than in the linear model, but lower than in the logistic model. Also other information criteria favor the logistic model. These results are in fact expected as of our earlier assumption, (i.e. that the 1987 crash was driven by internal forces, and the 2001 crash by external). While the 2001 data does have some bifurcations, the cusp catastrophe model clearly cannot fit the data significantly better than other models. This seems to be true, and for these data the catastrophe model did not perform better. However, for the 1987 crash the model seems to fit the data much better, and that is the sign, that the crash has occurred due to internal market forces.

[3]Applications are available at Han van der Maas's Website (`http://users.fmg.uva.nl/hvandermaas/`)

model	linear	logistic	cusp
R^2	0.1128	0.4682	0.2023
log likehood	-0.61×10^3	0.45×10^3	-0.55×10^3
AIC	0.12×10^4	0.91×10^3	0.11×10^4
BIC	0.12×10^4	0.93×10^3	0.11×10^4
parameters	4	5	6

Table 2: Results of the fits to the 2001 – 2002 data

4 Conclusions

Uncertain behavior of stock markets has always been on the leading edge of the research. Using the Cobb [1], Hartelman [4] and Wagenmarkers [6] results we have managed to test cusp catastrophe theory on the financial data, and we have arrived at very interesting results which may help to move the frontier of understanding the stock market crashes further on. We may thus confirm, that the catastrophe models explains the stock market crash much better then alternative linear regression models, or nonlinear logistic model. We have fitted the data of the two stock market crashes, the first being the crash of October 19, 1987, and the second September 11, 2001. We have used the sentiment measures to model the proportion of technical and fundamental players in the market. OEX put/call ratio is a very good measure of the technical players and represents the speculative money in our model and the trading volume is the measure of fundamental players and represents the excess demand.

We have clearly identified the bimodality of the returns using the test for multimodality which confirms that there is 75% probability that there is at least one bifurcation point in the data. Finally, the cusp catastrophe model fits these data much better than other models that have been used. Hence we conclude that the internal processes of the first dataset led to the crash in 1987. On the other hand, the crash of the September 11, 2001 can be better explained by the alternative logistic model. We have also found only 26% probability that there is at least one bifurcation point in these data, which is also in line with our second assumption: that due to the fact that this crash was caused by external forces the presence of the bifurcations in the data is much weaker.

Our findings may contribute to the frontier of the research, as it is the first attempt to quantitatively explain stock market crashes by cusp catastrophe theory. The testing has been conducted only on the restricted datasets. Thus further work is to test on different data which describes the situations when the changes in speculative money in the stock market lead to a crash. The main significant question, that of whether cusp catastrophe theory may help with an early indication of the stock market crashes still remains to be answered.

Acknowledgement

The authors would like to thank Prof. van der Maas for supplying a copy of Hartelman's dissertation and for publishing invariant stochastic catastrophe theory tools on his web site. A support from the Czech Science Foundation under grants 402/06/0990 and Grant Agency of Charles University (GAUK) under grant 46108 and Centre of research MSMT LC06075 is gratefully acknowledged.

References

[1] COBB, L. *Parameter estimation for the cusp catastrophe model.* Behavioral Science 26, 1981,75–78.

[2] COBB, L., WATSON, B. *Statistical catastrophe theory: an overview.* Math. Model 1, 1980, 311–317.

[3] COBB, L., ZACKS, S. *Applications of Catastrophe Theory for Statistical Modeling in the Biosciences.* Journal of the American Statistical Association 392,1985, 793-802.

[4] HARTELMAN, P. A. *Stochastic Catastrophe Theory.* Dissertatie reeks 1997-2, Faculteit Psychologie, Universiteit van Amsterdam, 1997.

[5] THOM, R. *Structural Stability and Morpohogenesis*, New York: Benjamin,1975.

[6] WAGENMARKERS, E. J., MOLENAAR, P. C. M., HARTELMAN, P. A. I., van der MAAS, H. L. J. *Transformation invariant stochastic catastrophe theory.* Physica D, 211, 263, 2005

[7] ZEEMAN, E. C., *On the unstable behaviour of stock exchanges.* Journal of Mathematical Economics 1(1), 1974, 39-49.

Jaromír Baxa

Institute of Economic Studies, Faculty of Social Sciences
Charles University in Prague
Department of Econometrics, Institute of Information
Theory and Automation of the AS CR

e-mail: jaromir.baxa@centrum.cz

A Simple Real business cycle model— —An application on the Czech economy

Abstract

This paper presents a simple real business cycle model calibrated for the Czech economy. Calibrated model can explain observed fluctuations of the main macroeconomic variables to a large extent, but not perfectly. The deviations between behavior of simulated and observed variables were caused either by theoretical assumptions of the model or they are in accordance with common knowledge of the Czech economy with labor market rigidities and volatile trend growth.

Inadequate theoretical structure of the model is connected with behavior of consumption over the business cycle. The simulated consumption is still correlated with the output gap but its variance is not lower than the variance of output as supposed by the permanent income hypothesis. According to this theory behind the life time utility function household should tend to smooth the consumption over the cycle. Consequently simulated output gaps are caused mostly due to fluctuations of investment with only minor role of consumption.

Problem with higher volatility of investment and output is caused by high estimated variance of Solow residuals which are derived from linear trend assuming constant pace of technological progress over time. This might be true in long-term in developed economies however the trend growth of transition countries at least over the period under investigation is volatile itself containing periods of decreasing growth, too.

Keywords: *Real Business Cycles, Calibration, DYNARE*

1 Introduction

Since early theoretical papers in the field of the real business cycles (further referred as RBC) occurred in the eighties ([17], [22], [24]), it has been shown that many applications of these kind of models are able to replicate most of the stylized facts of economic fluctuations in developed countries[1]. However, applications on countries in transition or developing countries are quite rare[2] despite especially for the Czech economy there's an evidence that its stylized facts are in

[1] See [3], or [10] for references and extensions, application on the U.S. data in [18], sceptical reply in [11].

[2] Application with an RBC-style model with the Czech data can be found in [23], another one with application on emerging economy is study of various RBC models on Chile [1] or an open economy RBC model for Slovakia [26].

accordance with many features generated by standard real business cycle models as reported by Hloušek [12]. In this paper we calibrated the base-line RBC model on the Czech economy with the aim to replicate the most salient features of its economic cycle. It is a great challenge for such a simple model as the Czech economy experienced many structural changes within last decade. Also, the macroeconomic development includes both period with negative GDP growth and period with growth rates exceeding 5% per year.

Our paper is organized as follows. Section 2 provides a description of the model used in this paper, section 3 summarizes important "stylized facts" of the Czech economy. Data description and calibration follow in section 4 and the results are described in section 5. Summary of main findings closes this paper.

2 A Baseline Real Business Cycle Model

Essential features of the model model used here follow basic versions of the real business cycle model described in [3] and [24]. Households solve traditional maximizing problem: they maximize their life time utility U, given as the expected discount flow of utilities from consumption c_t and leisure l_t:

$$\max_{c_t,l_t} U = E_0 \sum_{t=0}^{\infty} \beta^t u(c_t, l_t) \tag{1}$$

The expectation operator E_0 indicates that the expectations are conditional upon information set available at time $0, \beta \in (0;1)$ is time discount factor determining preference for current utility from consumption and leisure according to the future. Instantaneous utility function, $u(.)$, is assumed to be continuous and twice differentiable in both variables. The utility function takes the form of the logarithmic function (2), special case from the constant elasticity class of utility function[3], with parameter $psi > 0$ measuring importance of leisure for instantaneous utility.

$$u(c_t, l_t) = \log c_t + \psi \log l_t. \tag{2}$$

Households get their income from their firms equipped with a production technology producing, for simplicity, a single homogeneous product y_t. The inputs for the production are capital k_t and labour of the members of the households n_t. In this model, the production side of the economy is represented by a standard neoclassical Cobb-Douglass production function

$$y_t = e^{z_t} k_t^{\alpha} n_t^{1-\alpha}, \tag{3}$$

with parameter $\alpha \in (0, 1)$ measuring capital's share on output. The production function is augmented for stochastic productivity shock z_T which follows an AR(1) process (4) with persistence determined by its autocorrelation coefficient $\rho \in (0, 1)$ suggesting stationary nature of these shocks. It is this presence of shocks into technology that cause fluctuations of the

[3]The motivation for using of the logarithmic utility function in the RBC models is that there is a range of microeconomic and asset prices evidence suggesting the coefficient of risk aversion close to 1 ([24]) and estimates suffer with very large standard errors [20].

overall economic activity in this model. Stochastic innovations ε_t are assumed to be normally distributed with zero mean and variance σ.

$$z_t = \rho z_{t-1} + \varepsilon_t. \tag{4}$$

The level of capital stock increases each period by a margin of investment which is not used as a replacement for depreciated capital with depreciation rate $\delta \in (0,1)$:

$$k_{t+1} = i_t + (1-\delta)k_t. \tag{5}$$

The model equations are closed with two identities. The first one says that within each period households divide their one unit of time into labour and leisure, thus $n_t + l_t = 1$. Similarly, output of each period can be used either for consumption and investment $y_t = c_t + i_t$.

The model is constructed as its main variables y, c, i, n, k and l do not grow over time when the model is in its steady state. Consequently these variables represent deviations of trends rather than development of variables in levels.

Solution of the households' problem consisting from the equations (1)-(5) is straightforward using Lagrange multiplier method. The non-linear system consists from the equation representing the intratemporal optimality condition for labour/leisure ratio (8) stating that the marginal product of labour equals the marginal rate of substitution between leisure and utility from consumption. Furthermore the Euler equation (9) describes the intertemporal optimality condition for the consumption stream and finally equations specifying variables follows (10)-(12):

$$\psi \frac{c_t}{l_t} = (1-\alpha) e^{z_t} \left(\frac{k}{n}\right)^{\alpha} \tag{6}$$

$$\frac{1}{c_t} = \beta \frac{1}{c_{t-1}} E_t \left[1 + \alpha e^{z_{t+1}} \left(\frac{n_{t+1}}{k_{t+1}}\right)^{1-\alpha} - \delta \right] \tag{7}$$

$$y_t = e^{z_t} k_t^{\alpha} n_t^{1-\alpha} \tag{8}$$

$$z_t = \rho z_{t-1} + \varepsilon_t \tag{9}$$

$$y_t = c_t + i_t \tag{10}$$

$$n_t + l_t = 1 \tag{11}$$

$$k_{t+1} = i_t + (1-\delta)k_t \tag{12}$$

The system (6)-(12) has been linearised around its steady state and solved using DYNARE toolbox.

3 The Czech Business Cycle

Stylized facts that can characterize the Czech business cycle have been summarized by Hloušek ([12]) recently. He found that components of main macroeconomic aggregates, consumption and investment are both procyclical and also more volatile than the output gap as retrieved by the band-pass filter (consumption has nearly the same volatility, but the permanent income hypothesis would suggest higher smoothing). The same holds for the gap of employment as well. On the other hand government expenditures behave acyclically over the selected years. Regarding overall data we were able to confirm his results on our dataset (1Q 1996 – 3Q 2007) using different filtration method (Hodrick-Prescott filter).

The basic business cycle model has been developed for the description of the closed economy without explicit attempt to model behavior of net exports and of the government sector. Keeping this in mind the output time series has been constructed as pure sum of consumption and investment with government expenditures included in consumption C. As they behave acyclically the behavior of consumption and the sum of consumption and government expenditures is very similar[4]. In accordance to other literature ([1]) the time series of investment has been augmented for investment into housing, because this component of investment is often biased by external factors such as demographical changes or changes in economic policy.

4 Parameterization of the Czech Economy

The model was parameterized using the Czech quarterly data from 1Q 1996 to 3Q 2007. All the variables were represented in 2000 real counterparts and they were first seasonally adjusted using the ARIMA X12 method and then detrended using the Hodrick-Prescott filter. Most of the data were obtained from the system of national accounts provided by the Czech Statistical Office.

Corresponding time discount factor β is related to the subjective discount rate $\xi : \ \beta = \frac{1}{1+\xi}$, which in steady state can be associated with the real interest rate. The value of β was set to 0.99, which is a frequent value used in the literature ([4] and other textbooks, the same value is used for the Czech economy in [25]) as this value corresponds to quarterly interest rate 1%, often seen as a reference value[5]

Parameters α, δ and φ are jointly determined by the steady state of the economy. If we assume that the economy oscillates around its steady state (which might be relatively strong assumption) the values of variables at steady state might be replaced by corresponding means. Steady state ratio of investment and output implies the relation between the values of capital share α and depreciation rate δ according to (13):

$$\alpha = \left(\frac{\delta + \xi}{\delta} \right) \frac{\bar{i}}{\bar{y}}, \tag{13}$$

[4]Standard deviation of cyclical component increased by 0.2% only, the first autocorrelation decreased from 0.79 to 0.64 when government expenditures were included in consumption directly.

[5]Slightly lower value 0.987 corresponding to the quarterly interest rate 1.3% is used in [3], for the U.S. economy the calibration of time discount factor is often even lower as in [22] (0.95) or [24] (0.98) reflecting higher interest rates and real returns on capital in the U.S. during previus periods. For the Czech economy β ranges from 0.985 in [13] to 0.993 [14] based on [2]. Evans and Sezer ([26]) estimated the value of real return on capital across the EU and they found that it ranges from 3% to 5.5%. Minor changes in β within the range from 0.98 to 0.993 didn't affect the results.

where the $\bar{i}/\bar{y}$ is the average investment output ratio, which is approximately 0.26 given specifications of investment and output used here.

This leads to most appropriate choice of $\alpha = 0.43$ and $\delta = 0.0153$, as the value of capital output ratio α is the same as parameter of production function estimated in [5][6]. Also resulting depreciation rate at 1.53% per quarter (implying 6.1% per year) is not far from microeconomic studies that usually estimate it around 1.2% per quarter (see [8], [15], [19] for details).

Remaining parameter ψ was calculated from steady state relations of consumption, output and working hours using expression (14) and estimated share of leisure 61% yielding the value of $\psi = 1.21$. This formulation assured that the resulting labour/leisure ratio at steady state would correspond to the observed long-run characteristic of the Czech economy.

$$\psi = (1-\alpha)\frac{1}{\bar{c}/\bar{y}}\left(\frac{1-n}{n}\right) \tag{14}$$

Finally the structure of productivity shocks was calibrated using procedure described in [16]. The output is given by Cobb-Douglass production function and the series of Solow residuals is revealed. Their deterministic and stochastic components are separated assuming that the deterministic one represents the trend growth of the economy whereas the stochastic part causes business cycles. To allow for such decomposition production function becomes

$$y_t = A_t e^{z_t} k_t^{\alpha} n_t^{1-\alpha} \tag{15}$$

where A_t is the deterministic part attributed to the trend growth. For simplicity assume that in previous discussion about the structure of the model the value of A_t was set to 1 and now we generalize the discussion to a more general case. Nevertheless it is very easy to show that size of A_t changes only the level of consumption, capital, output and investment keeping both dynamics and the great ratios unchanged.

The Solow residual (SR) are now calculated from the time series of output, capital (revealed from perpetual investment method and employment:

$$\log SR_t = \log Y_t - \alpha \log K_t - (1-\alpha)\log N_t. \tag{16}$$

It consists from two parts, thus we can write:

$$\log SR_t = \log A_t + z_t \tag{17}$$

[6]More recently Hájková and Hurník ([9]) allowed for time-varying nature of parameter α and they found that it decreased from 0.47 to 0.4 within 1995-2005 period. On the other hand Hájek [8] in his study on the sources of economic growth in the Czech Republic and decomposition of the factor productivity uses 0.35. Higher value of $\alpha = 0.52$ was used by Marek [21] in his simulation of the Czech pension system and alternative scenarios of its reforms, he derived this parameter from the Czech national accounts. Gollin [7] for example argues that values of α above 0,5 (relatively frequent for emerging markets) are biased because official statistics are often unable to account for compensations of self-employed and in the shadow economy.

The deterministic part is assumed to follow a deterministic linear growth with quarterly growth rate γ

$$\log A_t = \log A_{t-1} + \log \gamma \tag{18}$$

and the stochastic part is the technology shock introduced in section two, which follows the AR(1) process.

$$z_t = \rho z_{t-1} + \varepsilon_t \tag{19}$$

Resulting value of autocorrelation of shocks obtained from this procedure equals to 0.7 for a trend growth estimated to 0.77% per quarter. Despite the estimated growth rate is not robust over the change of the sample, the autocorrelation of stochastic part remains stable. The variance of innovations was set to 0.011.

5 The Results and Evaluation of the Model

Tables 1 and 2 reports descriptive statistics of observed and simulated variables. The variance of output implied by the model is almost two times higher than observed one and several important differences are also in variance ratios. The model predicts lower volatility of investment in comparison to volatility of output, on the other hand consumption is much smoother in the model than in the data. Similarly comparing autocorrelations of these two variables consumption is smoother despite observed data show that consumption is slightly less persistent than investment or output. The same information arises from correlation matrices with predicted output/consumption correlation lower and investment/output correlation higher. These features arise from internal structure of the model which assumes that consumption follows path in line with the permanent income hypothesis.[7] Also high value of autocorrelation of consumption implies that its deviations from equilibrium might be long-term with relatively low tendency to return to steady state. To conclude this calibrated model leads to business cycles generated mostly by shifts in investment which goes against observed and estimated stylized facts.

The variance of employment according to variance of output seems to be very similar both in model and data, but observed employment has slightly higher persistence. Correlation coefficients with output show that correlation between employment and output is much lower than predicted (0.35 to 0.94). It shows that the Czech employment doesn't respond to fluctuations in output perfectly suggesting that observed persistence of employment and its insensitivity on output fluctuations can be caused by labor market rigidities or another forces that limit the effect of adjustment processes.

Figure 1 shows responses of the model to technology shock. We assume that at time 0 a one percent shock in productivity z_t hits the economy which was at steady state before as illustrated on the first plot. According to our specification of shocks the effect slowly decreases and vanishes out after about 15 periods. In response to the shock investment increase by amount

[7] One way how to overcome this counterfactual results suggested by Beorgeing-Soto [1] is to separate consumption into time series of durables and nondurables and including expenditures on durables into investment, the intuition behind is that people buy cars and consumer electronics more often during good times than in bad times.

	mean	st. dev.	var x/var y	1st autocorr	2nd autocorr	corr(x,y)
y	0.0000	0.0117	1.0000	0.7333	0.4941	1.0000
i	0.0000	0.0506	4.3251	0.6681	0.3665	0.8423
c	0.0000	0.0161	1.3783	0.6436	0.4274	0.5602
b	0.0000	0.0060	0.5128	0.8566	0.6324	0.3468

Table 1: Descriptive statistics of observed variables

	mean	st. dev.	var x/var y	1st autocorr	2nd autocorr	corr(x,y)
y	0.0000	0.0229	1.0000	0.7014	0.4895	1.0000
i	0.0000	0.0793	3.4593	0.6574	0.4157	0.9689
c	0.0000	0.0081	0.3543	0.9927	0.9765	0.5025
b	0.0000	0.0123	0.5348	0.6504	0.4036	0.9401

Table 2: Descriptive statistics of simulated variables

of 4.5% and consumption by about 0.5%. Consequently the output in the period after shock is higher for 5%. Increased investment caused increase of capital which in ten periods increases by 15% and then, after positive effect of shock diminishes slowly deteriorates so that after 40 periods corresponding to 10 years it is still 7% above its steady state. Employment increases in response to the shock by 0.4% and decreases after, 3-4 years after shock it is lower than at steady allowing to benefit from accumulated wealth from the good times. Then it returns very slowly to its steady state value despite the deviation is negligible (lower than 0.05%). Good times and high amount of capital turns into slow decrease of consumption to its steady after peak at 5th period.

6 Conclusions

Behavior of simulated variables corresponds to observed fluctuations of the main macroeconomic variables to some extent, but not perfectly. Most of the deviations between behavior of simulated and observed variables is caused either by theoretical assumptions of the model, others are in accordance with common knowledge of the Czech economy with labor market rigidities and volatile trend growth.

The first case – inadequate theoretical structure of the model – is connected with behavior of consumption over the business cycle. The simulated consumption is still correlated with the output gap but its variance is not lower than the variance of output as supposed by the permanent income hypothesis. According to this theory behind the life time utility function household should tend to smooth the consumption over the cycle. Furthermore predicted values of autocorrelation function of consumption are substantially higher than observed ones and this supports the argumentation that the Czech consumption path is influenced by changes in economic conditions much more than it should be if this model fits the data generating process perfectly. Consequently simulated output gaps are caused mostly due to fluctuations

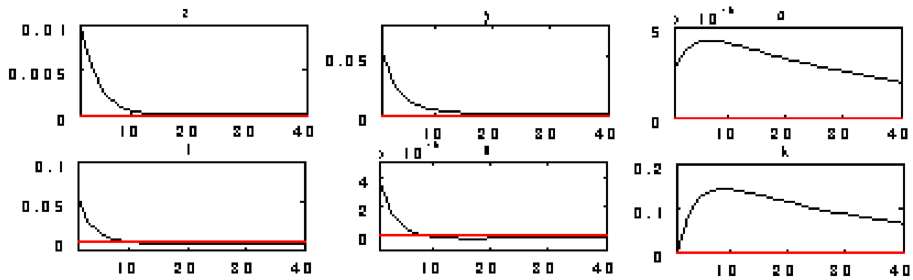


Figure 1: Impulse responses of the technology shock z

of investment with minor role of consumption.

The second case of deviations can be seen from discrepancies between observed and simulated paths of employment. Generally these differences imply that the observed unemployment is much more persistent and insensitive on economic fluctuations suggesting there are more important causes of changes in employment than substitution between time and leisure with respect to observed productivity.

Finally one problem with higher volatility of investment and output, which is caused by very high estimated variance of Solow residuals derived from linear trend assuming constant pace of technological progress over time. This might be true in long-term in developed economies however the trend growth of transition countries at least over the period under investigation is volatile itself.

References

[1] Bergoeging, R., Soto, R., 2002: *Testing Real Business Cycle Models in an Emerging Economy.* The Central Bank of Chile Working Paper No. 159.

[2] Christiano, L.J., Eichenbaum, M.. and Evans C.L. (1999): *Monetary policy shocks: What have we learned and to whatend?* Handbook of Macroeconomics, editors J.B. Taylor, M. Woodford, Elsevier.

[3] Cooley, T. (ed.) 1995: *Frontiers of Business Cycle Research.* Princeton: Princeton University Press.

[4] Dejong, D., Dave, Ch. 2007: *Structural Macroeconometrics.* Princeton: Princeton University Press.

[5] Dybczak, K., Flek, V., Hájková, D., Hurník, J.: *Supply-Side Performance and Structure in the Czech Republic(1995-2005).* CNB Working Paper, 6/2004.

[6] Flek, V. (ed.) 2007: *Anatomy of Czech Labour Market.* Prague: Karolinum.

[7] Gollin, D., 2002: *Getting Income Shares Right.* Journal of Political Economy, vol. 110(2), pp. 458-474.

[8] Hájek, M. 2006 *Zdroje růstu, souhrnná produktivita faktorů a struktura v České republice.* Politická ekonomie, 2006, č. 2, s. 170-189.

[9] Hájková, D., Hurník, J.: *Cobb-Douglas Production Function: The Case of a Converging Economy.* Czech Journal of Economics and Finance (Finance a uver), vol. 57(9-10), pp 465-476.

[10] Hartley, J. et al. 1998: *Real Business Cycles: A Reader.* London and New York: Routledge.

[11] Hartley, J. 1999: *Real Myths and a Monetary Fact.* Applied Economics, vol. 31, pp. 1325 – 1329.

[12] Hloušek, M. 2006: *Stylized Facts of Business Cycle in the Czech Republic.* Conference proceedings, Mathemtical Methods in Economics, 2006, Pilsen.

[13] Ježek, M., Houska, J., Schneider, O. 2004: *Pension Reform: How Macroeconomics May Help Microeconomics–The Czech Case.* Charles University IES Working Paper No. 88/2005.

[14] LAXTON, D. and PESENTI, P. 2003: *Monetary Rules for Small, Open, Emerging Economies.* Journal of Monetary Economics, Elsevier, vol. 50(5), pp. 1109-1146.

[15] Lízal, L.-Svejnar, J. 2002: *Enterprise Investment During the Transition: Evidence from Czech Panel Data.* The Review of Economics and Statistics, vol. 84(2), May 2002, pp. 353-370.

[16] King, R., Rebello, S., 1999: *Resuscitating real business cycles.* Handbook of Macroeconomics, editors J.B. Taylor, M. Woodford, Elsevier.

[17] Kydland, F. E. and Prescott, E. C. 1982: *Time to Build and Aggregate Fluctuations.* Econometrica, vol. 50, no. 6, pp. 1345-1370.

[18] Kydland, F. E. and Prescott, E. C. 1990: *Business Cycles: Real Facts and Monetary Myth.* Federal Reserve Bk of Minneapolis Quarterly Review, vol. 10, no. 4, pp. 9-22.

[19] Kočenda, E., 2001. *Development of Ownership Structure and its Effect on Performance: Czech Firms from Mass Privatization.* CERGE-EI Working Papers wp188.

[20] Kocherlakota, N. 1996: *The Equity premium, it's still a puzzle.* In: Journal of Economic Literature, 34, pp. 42-71.

[21] Marek, D., 2007: *Penzijní reforma v České republice: model důchodového sytému s kombinovaným financováním.* Disertační práce, IES FSV UK, 2007.

[22] Plosser, Ch. I. 1989: *Understanding Real Business Cycles.* Journal of Economic Perspectives, vol. 3, no.3, pp.51-57.

[23] Polanský, J., Vašíček, O., 2006: *Analysis of th eCzech Real Business Cycle Model.* Conference proceedings, Mathemtical Methods in Economics, 2006, Pilsen.

[24] Prescott, E. C. 1986: *Theory ahead of Business Cycle Measurement.* Federal Reserve Bank of Minneapolis Quarterly Review, vol. 14, no. 2, pp. 3-18.

[25] Vašíček, O., Musil, K. 2006: *Behavior of the Czech Economy: New Open Economy Macroeconomics DSGE model.* Research Centre for Competitiveness of Czech Economy Working Paper 23/2006.

[26] ZEMAN, J., SENAJ, M. 2008: *Modeling the Development of the Slovak Economy Using Basic DSGE Model.* Economic Modeling, 16, 3/2008, pp. 2-7.

Štefan Berežný, Ján Plavka

Department of Mathematics
Faculty of Electrical Engineering and Informatics
Technical University of Košice
B. Němcovej 32, 042 00 Košice

e-mail: Stefan.Berezny@tuke.sk, Jan.Plavka@tuke.sk

Štefan Lyócsa

Department of business informatics and mathematics
Faculty of Business Administration in Košice
University of Economics in Bratislava
Tajovského 13, 042 00 Košice
e-mail: Stefan.lyocsa@euke.sk

Optimization under Categorization

Abstract

On the present, businesses are claiming higher demand for more efficient cash transfer, not only in banking industry, but also in other industries, where large amount of financial resources needs to be transferred, either in cash or in electronic form. Aside that transportation of cash is financially demanding, it is also an important issue with regard to security of transportation. Most companies prefer non – cash transfers, or keeping cash transfers as low as possible. This can be addressed using optimization techniques.

In this paper, we describe a process of optimizing cash transfers between affiliates so, that at the same time we want to have an equal and minimum amount of cash, needed for transportation and transfer purposes. We have designed a mathematical model, which describes a simplified situation of discussed reality and by using a polynomial algorithm we have found an optimum of the designed problem. The criteria for finding an optimum were defined in advance, before seeking the solution of the problem.

The model have been developed using simple operators, such as: *Sigma max* and *min*. The basic structure is represented by a weighted graph, and feasible sets are represented by paths or by the spanning tree of the graph.

Keywords: *polynomial algorithm, categorization of edges, path, spannig tree, optimization.*

1 Problem formulation – cash transfer

The term cash transfer we are using in this paper does not refer to the accounting definition of cash flow. For the purpose of this paper, by cash flow we refer to money settlements between institutions in physical (transport of money) and non-physical (account transfers) way. We will describe one of the issues regarding cash flows in companies, which have to transfer larger amount of money within their own capacities, using cash or non – cash transfers. Transportation of larger amount of money is financially demanding, because we need to ensure physical security of money and therefore companies prefer non – cash transfers or minimize of cash transfers, or eventually they prefer to optimize cash transfers within their affiliates, which may consequently lead to cost savings. One of the possibilities to optimize is following:

The process of cash transfer optimization in a company, within its affiliates, is performed so, that the amount of transfers is uniformly distributed for all groups of affiliates. These affiliates belong to the same group, concurrently controlled by an authority. The goal is to minimize the amount of financial resources, within realized transfers. The affiliates are controlled by one, or more authorities (which may be represented by central departments). Affiliates may be represented for example by branch offices.

For every branch office, we have the information about the largest amount of money, which can be transferred and to whom. We also know, between which branch offices cash transfer is not taking place directly. In these cases, financial resources are transferred using other branch offices. The task is to find such a financial path between two branch offices A and B, that all interested parties are loaded, if it‘s possible, with the same number of transfers and at the same time, we want to minimize the amount of money for the most loaded branch offices. We also assume some form of a hierarchy between affiliates. The basic communication is taking place between central authorities and their affiliates (in some cases, for one group of affiliates, more than just one central authority may be assumed) and a higher level of communication is taking place between central authorities (in some exceptional cases a communication may be allowed between office branches, which belong to different groups). This communication hierarchy within a company is important for the construction of a mathematical model and the optimization in this model.

2 Mathematical model

We are constructing a graph $G = (V, E)$, where the set of vertices V represents branch offices, which are forming the financial net, in which financial transfers (cash or non – cash) are taking place. The most challenging problem is concerned with cash transfers; therefore, we focus mostly at the problem of cash transfers. Non – cash transfers will not disrupt this model. We connect two nodes with an edge only, if it is possible to make a direct transfer between two office branches. Generally, we could assume a complete graph, in which edges where financial transfers are not possible, would have been weighted by zero, but as a consequence, such a model would unnecessarily complicate the optimization process. Edges which represent transfers within the same group of affiliates (e.g. branch offices under one central authority) will be regarded as one group. In this manner, we will get categories $S_1, \ldots, S_p$, where $p \leq |V|$ (generally the value of p may not be bounded by the cardinality of the set of vertices of the graph G). No edge, can be at once dedicated to more categories, so all categories of edges are disjunctive to each other. Every edge $e \in E$ has a weight $w(e) \in Z, w(e) \geq 0$. The weight $w(e)$ correspond to the assumed amount of financial resources, which will be transferred between branch offices so, that the transfer will not represent a loss (this lower bound has its meaning for cash transfers). In this paper, we will not focus ourselves to the methodology of setting weights $w(e)$. This would require a separate research. Let A denote a vertex, which represents a branch office, from which a financial transfer is going to be performed, and denote B a node of a graph, to which this transfer needs to be done. With a symbol P, we will denote a path from nodes A to B. Now, let's consider following objective functions:

$$L(P) = \max_{1 \le i \le p} \sum_{e \in S_i \cap P} w(e) - \min_{1 \le i \le p} \sum_{e \in S_i \cap P} w(e),$$

$$B(P) = \max_{1 \le i \le p} \sum_{e \in S_i \cap P} w(e),$$

where $P \in \Delta(G)$, and we assume, that $\max_{e \in \emptyset} w(e) = 0$. The family of feasible sets $\Delta(G)$ represents the set of all paths between nodes A and B, or the set of all spanning trees of graph G. By minimizing the objective function $L(P)$ we achieved, that between two office branches, the amount of transferred money will be approximately the same, or if we would assume to value edges of the graph G. by a unit, then we would acquire the most uniform numbers of transfers between all groups of office branches. By this we achieve, that in the company, the costs regarding transfers will be uniformly distributed. By concurrently minimizing the objective function $B(P)$ we will secure, that transfers will carry as less amount of money as possible and therefore, we also minimize risks concerned with transfers of larger amount of money. If we would minimize the most heavily loaded edge on the path P, this would not necessarily mean minimum amount of transfers between one group of branch offices. Even though, in the next optimization, we will use modified objective function $B(P)$, because some kind of uniformity is generated by minimization of objective function $L(P)$ (providing, that within each category, there are no big differences in the weights of edges). Therefore, in the next section, we will use a modified objective function $B(P)$.

$$B(P) = \max_{e \in S_i \cap P} w(e) \text{ or } B(P) = \sum_{e \in S_i \cap P} w(e).$$

when solving, we will use the modified objective function $B(P)$, even though we won't acquire the desired optimum in absolute numbers, but it will be possible to find the result in polynomial time.

3 Solving the problem

Before we describe the progress of searching for optimum solution of the problem described in the section above, we will shortly resume the common knowledge about similar problems, which we can use when solving our optimization problem. If we have a graph $G = (V, E)$, decomposition of the set of edges E into disjunctive categories $S_1, \ldots, S_p$ and non negative values of edges $w(e)$, while minimization of non modified objective function $B(P)$, than we have a known problem described in [1] (the operator *max* and Σ are used). In this paper, it has been shown, that it is an NP-complete problem, even for two categories. Likewise, in [11] authors have shown, that a similar problem, where instead of the path, we are searching for the spanning tree of the graph, it is also an NP-complete problem. The issues surrounding minimization of objective function $L(P)$ were described in [2]. Similarly as in the cases before, this problem can be described as NP-complete problem, when it regards the path, but also the skeleton, already even for two categories. Despite these negative results, we can use the polynomial algorithm, which was published in [7]. This algorithm instead of searching for path P, is searching for such a spanning tree T of a graph G, for which the value of the objective function $L(T)$ is minimized, providing that all edges of the graph have the same integer weight. Just like when we have used the modified objective function $B(P)$, in which only an operator max or Σ is used, we will acquire known polynomially solvable bottleneck problem of a path, or spanning tree, or a problem of finding the shortest path, or a problem of the cheapest spanning tree of the graph G. The presented objective function may be written as follows:

$$B\left(P\right)=\max_{e\in P}w\left(e\right)\ \text{ or }\ B\left(P\right)=\sum_{e\in P}w\left(e\right).$$

We will describe the process of searching for the best possible solution of our problem in polynomial time by utilizing already known polynomial algorithms.

- In the first step, we will find the cheapest spanning tree T with regard to the objective function $L(T)$, while valuing every edge of the graph G by a unit. For this, we will use an algorithm, which is described in [7]. If the set of the edges would be only in two disjunctive categories, then we could use an improved, faster algorithm, which may also be found in the work of [7] in sections 4 and 5. The stated algorithm makes possible to find all optimum skeletons. We will arrange these spanning trees according to the value of $L(P_i)$, where P is the path between vertices A and B in the i-th spanning tree T_i for $i = 1, \ldots, k$. We will denote these skeletons as $T_1, \ldots, T_k$.
- In the next step, we will find the shortest path with regard to the function $B(P)$ and valuation of edges by $w(e)$. These paths can be found using a known polynomial algorithm. We will denote the found paths as $P_1, \ldots, P_l$.
- We can easily verify, whether one of the found paths $P_1, \ldots, P_l$ is a subset of one of the found skeletons, because in every skeleton of a graph is an expressly given path between two nodes. If one of the paths satisfies the stated condition, than this path will be regarded as optimal for our problem, and will be denoted as P_{opt}. This path is actually only an upper estimate, with regard to the stated objective functions and values of the edges of the graph G. The optimum skeleton will guarantee, that the optimum path will have at least as much transfers, as is the value of $L(T)$. Optimal path will guarantee, that the largest, one time transfer between office branches is the least possible or, that the amount of transferred financial resources on the path from A to B is minimal. The only possible improvement can be done with regard to the equalization of the number of transfers between single groups of branch offices. This improvement cannot be generally done in a polynomial time as stated in [1, 2, and 11].

If at this stage, we haven't found the P_{opt}, it means, that the paths between vertices A and B in optimum spanning trees are different from the optimum paths, which have been found with regard to the second objective function. For the paths, found in spanning trees of a graph G, the following has to be true; the value of the objective function $B(P_i)$ for $i = 1, \ldots, k$ has to be greater than optimum, which was found and is equivalent to the value of $B(P_i)$ for $j = 1, \ldots, l$. When searching for other best solutions, we can choose one from two possibilities: either we will sacrifice the value of the first objective function $L()$ or the value of the second objective function $B()$. Because we cannot ensure an improvement of the value of the function $L()$ in polynomial time for general cases, in next stages, we will sacrifice the value of equalization of the number of transfers in favour of minimal amount of transferred financial resources.

- We will find conjunctions of all optimum paths with the paths from optimum spanning trees. These conjunctions will be denoted as $U_1, \ldots, U_r$, where $r = k \times l$ and at the same time, we will arrange them with regard to their cardinality, from the lowest to the biggest.
- We will take every conjunctions with minimum cardinality and we will arrange them, with regard to the value of the objective function $L(P)$.
- The path with minimum value of the objective function $L()$ will be denoted as P_{opt}. This is the searched path, whiles the value of the equalized number of transfers, could have been worsened only by two times the corresponding cardinality of conjunction U.

With this process we can find the solution in polynomial time, which mostly approaches the real optimum solution, while we have utilized only the known polynomial algorithms, which are tightly connected with the issues of optimization, with regard to the categorization of graph‘s edges.

4 Conclusion

The stated process, is only one of the possibilities, how to handle optimization of cash transfers in a company. When dealing with issues surrounding categorization of edges, it is possible to make use of other objective functions and also polynomial algorithms, which can be used when searching for optimum, which are also described in papers [3, 4, 5, 6, and 9]. Application and modification of problems with categorization of edges, may also be found in [8, 10]. An independent consideration may be given to the methodology of developing weights to the edges in a graph G. One of the possible approaches, which can be utilized when developing weights, is a methodology, which is described in [12]. In the inscribed mathematical model, we could also choose alternative approaches, which would lead to creation of a new, approximative algorithm, or we could establish additional conditions, which would lead to a change from NP-complete problem to a polynomial solvable problem. This would require a new algorithm or the reformulation of the initial problem so, that the new mathematical model would be possible to solve in a polynomial time, while satisfying realistic situations. Considering the limited space in this paper, alternatives sketches in this conclusion will be representing the topics of further research and publication.

References

[1] AVERBAKH, I., BERMAN, O.: Categorized bottleneck-minisum path problems on networks, Operation Research Letters 16 (1994) 291-297.

[2] BEREŽNÝ, Š., CECHLÁROVÁ, K. and LACKO, V.: Optimization problems on graphs with categorization of edges, in: Proc. SOR 2001, eds. V. Rupnik, L. Zadnik-stirn, S. Drobne (Predvor, Slovenia, 2001) 171-176.

[3] BEREŽNÝ, Š., CECHLÁROVÁ, K. and LACKO, V.: A polynomially solvable case of optimization problems on graphs with categorization of edges, in: Proc. of MME'99 (Jindřichův Hradec, 1999) 25-31.

[4] BEREŽNÝ, Š., LACKO, V.: Balanced problems on graphs with categorization of edges, Discussiones Mathematicae Graph Theory 23 (2003) 5-21

[5] BEREŽNÝ, Š., LACKO, V.: Special problems on graphs with categorization, in: Proc. of MME'2000 (Praha, 2000) 7-13.

[6] BEREŽNÝ, Š., LACKO, V.: Easy (polynomial) problems on graphs with categorization, in: Proc. of New trends of aviation development (Air Force Academy of gen. M. R. Štefánik, Košice, 2000) 36-46.

[7] BEREŽNÝ, Š., LACKO, V.: The Color-balanced Spanning Tree Problems, Kybernetika vol. 41 (2005) 539 –546

[8] BEREŽNÝ, Š.: Further options of transport. In: MOSATT 2007: Modern Safety Technologies in Trasportation: Proceedings of the International Scientific Conference: 25th – 27th September 2007, Zlata Idka. Košice: Robert Breda, 2007. s. 19-21. Internet: www.mosatt.org ISBN 978-80-969760-2-7.

[9] BEREŽNÝ, Š.: Overview of problems on graphs with categorization. In: 22nd European Conference on Operational Research: Book of Abstracts: Prague, July 8-11, 2007. Praha: VŠE, 2007. p. 35.

[10] BEREŽNÝ, Š., PLAVKA, J.: Optimalizácia stresu na pracovisku. (submitted)

[11] RICHEY, M. B., PUNNEN, A. P.: Minimum weight perfect bipartite matchings and spanning trees under categorizations, Discrete Appl. Math. 39 (1992) 147-153.

[12] PEKARČÍKOVÁ, M., HAJDUOVÁ, Z., TURISOVÁ, Z.: Implementing TPM system – minimizing downtime breakdown and failures of equipment, Acta Avionica X Nb 15 (2007) ISSN 1335-9479.

Biography

Štefan Berežný graduated in Discrete mathematics from P. J. Šafárik University of Košice. At present he is an assistant professor at the Faculty of Electrical Engineering and Informatics, Technical University of Košice. His scientific research is focusing on combinatorial optimization, algorithms and complexity and management.

Ján Plavka graduated in Discrete mathematics from P. J. Šafárik University of Košice. At present he is an Associate professor at the Faculty of Electrical Engineering and Informatics, Technical University of Košice. His scientific research is focusing on computer science, algorithms and complexity. In addition, he also investigates questions related with the discrete dynamic systems.

Štefan Lyócsa graduated in Business Administration from University of Economics, faculty of Business Administration in Košice. At the present, he is studying as a post graduate student, at the department of business informatics and mathematics, focusing his research at two main topics: real options theory (modelling optimal decision making) and quality improvement (Six Sigma).

Tomasz Błaszczyk, Tomasz Wachowicz

Department of Operations Research
University of Economics in Katowice
ul. 1-Maja 50, 40-287 Katowice

e-mail: blaszczyk@ekonom.ae.katowice.pl,
wachowicz@ekonom.ae.katowice.pl

Application of Even-swap Method in Analyzing New Energy Development Projects

Abstract

The scope of this paper is to show the possibilities of applying the even-swap method in the analysis of the power-plants planning projects. According to the current situation on the Polish energy market and the need of reconstruction of old power sources many investments are required. Investment projects, beside the economical capacity, have to satisfy ecological requirements (e.g. consider CO_2 emission limits) and be acceptable both by authorities and the local community. A clear and understandable scoring system could be a useful tool to support the alternatives research. The authors propose to apply the even-swap method, which allows to build a ranking of the projects. Even-swap method is a multiple objective decision making tool basing on the sole aggregating criterion, but rejects the notions of utility and desirability scores. The approach focuses on estimating the value of each objective in terms of another. It results in a final score of project expressed in terms of one selected objective (e.g. money). Such a score is easy to calculate and interpret, which allows to avoid some tensions and disputes concerning the rules for selecting the winning projects.

Keywords: *even swaps, multi-criteria decision making, pre-feasibility analysis, energy generation sources*

1 Principles of Even Swaps

1.1 Making trade-offs with even swaps

The even-swap method is a simple support tool for multiple criteria decision making. However, the decision problem should be adequately structured before it is analyzed by means of even swaps. The even-swap method was itself proposed as a part of the PrOACT methodology [1], [2]. PrOACT divides any decision situation into the five elements: problem, objectives, alternatives consequences and trade-offs and then reassemble the results of their analysis into the right choice. The trade-offs between objectives can be made by means of even swaps. The procedure focuses on an iterative elimination of the alternatives and the objectives. The procedure employs the basic notion of vector domination, according to which alternative A dominates B if A is better then B in terms of some objectives and no worse then B in terms of all the others. The even-swap method eliminates thus each dominated alternative from the decision process. Furthermore, it basis on the straightforward assumption, which considers that one objective can be ignored in the decision making process if all the alternatives result in equal consequences for this very objective. The even-swap method tries to adjust then the consequences of all considered alternatives to make them equivalent in terms of some objective. It requires increasing the value of an alternative in terms of one objective while decreasing its value by the equivalent amount in terms of another objective. The equivalent amounts are defined subjectively by the decision maker and reflect the structure of their preferences. After such an adjustment the considered objective can be eliminated and domination analysis can be conducted once again with respect to the rest of objectives. The process of the objectives elimination is repeated until the decision maker is able to find the best decision (the one that dominates over all the others).

Let us consider a simple numerical example. We have two alternatives with different consequences for two objectives: price and time (Table 1).

Objectives	**Alternatives**	
	1	**2**
Price (in USD)	400	370
Time (in months)	18	20

Table 1: Table of consequences

Decision maker's preferences for the objectives are to minimize both the price and time. There is unfortunately no alternative that satisfies both the objectives simultaneously, therefore decision maker is unable to consider which of these two alternatives is more preferable (while considered the problem multi-objectively). The even swaps can be used now to evaluate a final score of these two alternatives. According to the idea of even swaps the decision maker needs to make the offers equal for one objective. Let us assume they have chosen the issue of time. The best consequence for this issue is "18 months". Therefore while comparing the consequences decision maker has to decide, what is an equivalent amount (in terms of price) they are going to pay for shortening the time from 20 to 18 months. Let us assume they determine that this 2 months difference is valued as 100 USD of additional costs. This amount may reflect some objective costs that the decision maker has to pay for seeding up the alternative number 2, but all the subjective and "non-measurable" costs may (and should) be taken into consideration as well. Now the equivalent amount should be swapped between the consequences, which means decreasing the value of time while increasing the value of price within the alternative number 2. The results of the even-swap analysis are given in Table 2.

Since both the alternatives result now in equal consequence for criterion of time, it can be eliminated from the decision analysis. After this elimination we can easily find the most preferable alternative, which is the one number 1 (it results in lower costs then alternative number 2). The even-swap analysis could be conducted in terms of criterion of time as well. In such a case the decision maker would have to determine the equivalent amount of months that could be swapped for the 30 USD difference in price.

Objective	Alternatives	
	1	**2**
Price (in USD)	400	~~370~~ 470
Time (in months)	18	~~20~~ 18

Table 2: Making swaps between the consequences

2 Constructing a scoring system with even swaps

We can apply even-swap method to construction of a scoring system, that can be used for ranking and evaluating the alternatives. The problem of construction of such a scoring system for negotiation support has already been described in details, including all required modifications and the system legitimacy [6], [7]. The general idea of a even-swap based scoring system is to find a base objective that will be used as a final score for the offers' evaluation. It can be chosen from the set of all objectives that the alternatives are described with. We recommend to choose the quantitative objective as the base, which will make the interpretation of the final score intuitively easy. Secondly, the iterative elimination of the alternatives and the objectives needs to be modified. While conducting the swaps analysis none of the alternatives must be eliminated. The equivalent amounts need to be accumulated for each alternative to determine its overall score. Third, making all the offers equal for one particular objective requires selecting the base option as a common result for all the alternatives. It is recommended to use the best option (i.e. the consequence that satisfied the very issue the most) as a base. The equivalent amount may be interpreted then as the cost of upgrading the alternative's consequence to the best possible resolution. When the swaps analysis for all the objectives is completed, the overall scores of the offers are obtained as the costs of their upgrading to the ideal (best) one.

Constructing a scoring system for the problem with a multitude of alternatives may be troublesome, because the even-swap analysis requires comparing virtually each pair of alternatives. For such problems it is possible to construct the scoring system by carrying out the within-objectives swap analysis and aggregating then the scores for the complete alternative by simple adding the equivalent amounts of the consequences that comprise this alternative. It is allowed to conduct the within-objective analysis because the even-swap method employs an additive aggregation of the scores, which means it assumes that the score associated with each option of one objective does not depend on the options of the other objectives [3]. Conducting the within-objective analysis of the options makes the process of even swaps easier. It consists then of four simple steps:

1. Setting the base objective.
2. Defining the best options (consequence) for all remaining objectives.
3. Assessing the equivalent amounts for options within objectives.

4. Aggregating the final scores of alternatives by adding the equivalent amounts of the options that comprise them.

3 Power generation project example

Polish energy sector companies are shortly before making decisions about starting investments in power generation sources. As far as the main plants owners are national companies, there possible is a higher social influence on the decisions which must be made. Of course, the techno-economical criteria are still basically considered at the phase of investment pre-feasibility studies, but the social acceptance could also be a factor of the success or defeat. There are well known examples of power plant projects failed in respect of socio-political barriers (even in final phase, like in German Kalkar plant [4]). The democratic societies require the information campaigns explaining the needs, goals, constraints and social costs and benefits. This country-wide explanation must be clear and understandable for most of the miscellaneous citizens. Probably, explanations basing on the comparisons focusing on the per capita costs motivate people to take a view on the trade-offs in terms of their own home budgets. In this paper we chose to employ the even-swap approach to show the idea of the core in information campaign. Fixed data and estimations are taken from Polish energy professionals portal: cire.pl [5]. To make the alternatives more clear and comparable we aggregated the evaluation criteria into five:

- Environmental friendliness, a subjective qualitative criterion, expressing the social acceptance of each alternative,
- "Green" certificate (certificate of origin), factual qualitative criterion, classifying each kind of energy to renewable or non-renewable class due to national and European law. In 2008 in Poland every energy seller must to ensure not less than 5% of volume from renewable sources (with "green" certificate of origin). While having no opportunities of producing this amount they are able to buy required certificates on a free market. At May 30th, 2008 the price of single certificate for 1 megawatt-hour (MWh) in Poland was equal 70,95 EUR.
- Carbon dioxide (CO_2) emission, factual quantities criterion, measuring the average quantity of CO_2 emitted while production of one energy unit (1MWh) with using the standard technology in each alternative. However, now in the whole EU free emission allotments in limited quantity are fixed, but probably, when the planned investments will step into operational phase three will be a must to buy all emission permissions on market. Permission price in our example (25,99 EUR per 1000 kilograms) was actual in Poland at June 3rd, 2008.
- Average investment expenditures per power unit, costs estimated due to actual or actually completed power plants constructions calculated per power unit equal 1000 megawatt (MW) of electrical power installed.
- Average cost of energy production per energy unit, average actual cost netto (including losses and taxes) in Polish energy producers (estimated in the alternative of nuclear energy).

Focusing on criteria explained above in this simplistic example we considered three energy source alternatives:

- Wind energy: energy is generated by wind turbines, constructed on a posts in windy regions. This kind of energy is made from a typical renewable source—it is environmentally friendly because of zero-emissions production. However, there are several cases known, where this kind of investments were blocked by pro-ecological societies in case of being dangerous for migrating birds and landscape changes. Wind energy is also very expensive because of its investment costs per energy unit and the fact, that in actual European regulations it must be bought by power suppliers. The singular wind turbine power is not higher than 3 MW in Polish weather conditions. That means the necessity of buying more than 350 turbines to ensure the required 1000 MW. Unstable wind power and directions demands keeping the power reserve (usually generated by coal-based power plants). Quantity of the energy produced this is not easy to control and regulate due to the market requirements. Average annual production is around 35% of the power installed, depends on the wind exposition and force.
- Nuclear energy: cheapest in production, but the most controversial source. Modern technologies employed in third generation nuclear plants are safe but the remembrances of the Chernobyl disaster are still alive in Poland. In "old EU" countries (e.g. in Germany and Sweden) there were also several pursuits to refuse nuclear power generation in aid of wind, water and geothermal sources. Energy producing in this kind of power plant makes no CO_2 emission, no other air, water nor soil pollution but requires safe disposals to warehouse after-production radioactive wastes. Wastes transportating and warehousing in case of accidents or sabotage acts is strongly dangerous and safety assurance is strict and expensive.
- Coal energy: the traditional source of energy in Polish industry. There are one of the biggest European coal beds laying mainly in the southern (bituminous) and central (lignite) Poland. The technology of fuel preparation and energy generation is well known and continuously improved. The efficiency of nuclear and coal power plants is equal in average, but the coal-based production consequences are the emission of CO_2 with nitrogen- and sulfur oxides, are dangerous poisons as well.

In this analysis we passed over the alternatives of water, geothermal, biomass, and solar energy with malice aforethought. Power generation from this sources could be realized with different technologies and resources. So that, the investment expenditures per unit and production costs are very sensitive and vary due to many specific conditions.

The above problem may be presented in a table of consequences (see Table 3).

Before we start the swaps analysis we need to recalculate some data to make the mutual comparison easier. Three energy types that we consider here differ in terms of average investment expenditures, but the power plants they require have also different time of exploitation (the time the plant can operate with no major rebuilding or maintenance works). We have estimated the average time of exploitation for each plant, that are 40 years for wind and coal energy plants, and 60 years for the nuclear plant. Furthermore, we need to take into account all the technical issues of each plant that will determine its final energy production. Wind

Energy type	environmental friendliness / social acceptance	"green" certificate requirement	CO_2 emission / MWh	average investment expenditures (EUR/1MW)	average cost of energy production (EUR/1MWh, netto)
Wind	yes/no	can sell	0	1 100 000	120
Nuclear	yes/no	must buy	0	3 000 000	25
Coal	no	must buy	1000 kgs	1 500 000	50

Table 3: Table of consequences for original power generation problem

energy plant average production is only 35% of its maximum capacity. Therefore to assure 1000 MW production we need to built a plant, the capacity of which exceeds 2857 MW. The traditional coal energy plant is usually divided into four blocks. While three of them are running the forth one is under maintaining and repairing process and then waits to replace the block that requires reparation the most. Thus to achieve 1000 MW production in coal energy plant we need to build a station of the 1333 MW capacity. The energy required to stop the activity of the nuclear plant is also about of its capacity. Basing on these numbers we can calculate the average investment expenditures per year required to produce 1000 MW of energy. These are consequently:

- Windy energy: 1000 * 1 100 000 * 2,857 / 40 = 78 571 428 EUR,
- Nuclear energy: 1000 * 3 000 000 * 1,333 / 60 = 66 666 666 EUR,
- Coal energy: 1000 * 1 500 000 * 1,333 / 40 = 50 000 000 EUR.

Starting the even-swap procedure we choose the base objective as an average investment expenditure per year. For all other objectives we will need to determine equivalent amounts in terms of an average investment expenditure per year, which we will try to do with some objective data (i.e. official prices, costs, penalties etc.). The final score we obtain then can be easily interpreted as a total cost of operation of each the plant per year.

We start the analysis with the objective of CO_2 emission. The best consequence for this objective is 0 kgs/ MWh and both the nuclear and wind energy result in this consequence. The coal energy produces the emission at the level of 1000 kgs/MWh. We can easily calculate the equivalent amount for reduction of the emission from the level of 1000 to 0 kgs/MWh by determining the total cost of the permission required for such an emission. The cost of 25,99 EUR (for emission of 1000 kgs of CO_2) we need to multiply by the power of the plant (1000 MW) and number of hours it works during a year. We obtain thus the total cost of 227 672 400 EUR.

Analyzing the "green" certificate requirements we find two different consequences for this objective as "can sell" and "must buy". Since the former generate revenues for the plant and the latter additional costs, we will calculate the equivalent amounts for a hypothetical option of “works optimally”. It means the plants with former consequence must buy the “green” certificate (that allows them to operate) in amount of 5% of their production capacity. The total cost we obtain is 70,95 EUR (for 1 MWh) multiplied by 5% of power plant capacity and number of hours it works during a year, which equals 31 076 100 EUR. For the wind energy plant with the initial option can sell“ the new works optimally“ option means it can sell certificate for all of its production, which will result in a revenue of 590 445 900 EUR. These numbers will be taken into account as the equivalent amounts in a trade-off analysis.

Average cost of energy production results in three different consequences for the plants. We calculate the equivalent amounts for modification of these consequences to the best option (hypothetical one) of no costs“. For each type of power plant we need to multiply its individual costs (in EUR/1MWh) by the production capacity (1000 MW) and the number of hours it works during a year. That results in the following equivalent amounts: wind energy—1 051 200 000 EUR, nuclear energy –219 000 000 EUR, coal energy—438 000 000 EUR.

We have made all the alternatives equal for the objectives under consideration and the results of our swaps analysis are presented in Table 4. The last column of the table is a sum of all the equivalent amounts for each objective, which is our final score of each alternative (a total cost of operation of each the plant per year).

Energy type	environ. friendl.	”green” certificate requirement	CO_2 emission / MWh	avg. investment expenditures (EUR/1MW)	avg. cost of energy prod. (EUR/1MWh)	total cost of operation per year (EUR)
Wind	yes/no	~~can sell~~ -590 445 900	~~0~~ 0	~~1 100 000~~ 78 571 428	~~120~~ 1 051 200 000	539 325 528
Nuclear	yes/no	~~must buy~~ 31 076 100	~~0~~ 0	~~3 000 000~~ 66 666 666	~~25~~ 219 000 000	316 742 766
Coal	no	~~must buy~~ 31 076 100	~~1000 kgs~~ 227 672 400	~~1 500 000~~ 50 000 000	~~50~~ 438 000 000	746 748 500

Table 4: Calculation of equivalent amounts and a final score of alternatives

We did not take into consideration the criterion of environmental friendliness yet, but as we see in Table 4 we do not need now the further analysis. The wind and nuclear energy differ in a total costs but result in the same consequence in terms of environmental friendliness. Therefore, no matter how big the costs of the promotional campaign for social acceptance could be, the nuclear energy would remain more preferable (less expensive) then the wind one. The comparison with the coal energy is simple too. Since the consequence for the environmental friendliness is far less preferable then for wind and nuclear energy, the costs of the campaign would be far more higher. That will result in rising the total cost of operation of the coal energy plant. To make the scoring system completed let us assume that the costs of the promotional campaign for improving the social acceptance are 400 000 EUR for wind and nuclear energy and 700 000 EUR for the coal energy. We obtain then the final scores of the alternatives: wind energy—539 725 528 EUR, nuclear energy—317 142 766 EUR, coal energy—747 448 500 EUR. As we see, the cheapest (and most preferable in sense of costs) alternative is a nuclear energy. It is 70% cheaper then the wind energy and more then 2 times cheaper then the coal one. We should obviously take into account now the utility of the scores that would reflect the final decision maker attitude toward the money payoffs, but the differences between the alternatives are so significant that we can assume it would not change the preferences over the alternatives. What is more, presenting the final score in money payoffs allows to justify objectively the choice and make it more clear for the public.

4 Summary

New power plant projects are very difficult and expensive enterprises. They require a detailed analysis since the first, conceptual phases until its final closure. In the matter of fact, those analyses must be done according to multiobjective and multicriterial systematic nature of this kind of project. Usage of different sophisticated examples is described in both scientific and professional resources. This way we already have a choice of good tools helpful in strategical, pre-feasibility, feasibility, and operational planning. The approach we discussed is not expected to replace other methods and techniques where AHP, Electre, Promethee or others have been used. This is a proposal of supplement to the toolbox employed in the energy projects planning.

In this paper we tried to explain how the even-swap approach could help decision-makers to manage the dialogue with people who are not supposed to understand well quantitative methods. The case of energy sources and their development is the nation-wide issue, and most of citizens are involved in its results. In our opinion the approach shown above makes the alternatives easier to convince or to refuse. Showing trade-offs by this way we are making the differences between alternatives clearer because, in most cases, citizens used to feel them by their own home budgets. However, it is not difficult to follow it with the quantitative criteria, but we also have qualitative ones. For example, the "environmental friendliness" is not easy to measure, but the even-swap approach helps us to ask citizens: "how much are you able to pay for it?" or to ask decision-makers "how much are you able to pay for making this criterion less important?".

References:

[1] HAMMOND, J.S.; KEENEY, R.L.; RAIFFA, H., *Smart Choices. A Practical Guide to Making Better Life Decisions.* Broadway Books, 2002, New York.

[2] HAMMOND, J.S.; KEENEY, R.L.; RAIFFA, H., *Even Swaps: A Rational Method for Making Trade-offs.* Harvard Business Review, March-April 1998.

[3] KEENEY, R.L.; RAIFFA, H., *Structuring and Analyzing Values for Multiple-Issue Negotiations*, in: YOUNG, H.P. (ed.): *Negotiation Analysis.* The University of Michigan Press, 1991, 131-152.

[4] MARTH, W., KOEHLER, M., *The german fast breeder program (a historical overview.)* Energy vol. 23, No. 7/8, 1998, pp. 593–608.

[5] The Center of Energy Market Information (PL): http://www.cire.pl

[6] WACHOWICZ, T., *NegoCalc: A Spreadsheet Based Negotiation Support Tool with Even-Swap Analysis*, *Proceedings of the Group Decision and Negotiation Meeting 2008*, 2008, University of Coimbra, Portugal.

[7] WACHOWICZ, T., *Even Swaps Method for Developing Assessment Capabilities of E-Negotiation System*, in: M. Ganzha, M. Paprzycki, T. Pełech-Pilichowski (eds.), *Proceedings of the International Multiconference on Computer Science and Information Technology*, 2007, vol. 2: 597-606.

Andrey Boyarshinov

Perm State University
Economic faculty
Bukierev st., 15, Perm, Russia

e-mail: andrey.boyarshinov@gmail.com

An Approach to Portfolio Management: Multidimensional Optimization

Abstract

During past decades, financial markets have been rapidly developing throughout the world. A large number of new market instruments has appeared. With constantly widening area of existing investment possibilities, it's becoming increasingly more difficult to manage investment portfolios. Old and trusted methods no longer work as well as they used to, thus a need for new methods and models is ever present.

In this paper a problem of optimal portfolio management is approached. A new method of portfolio optimization based on statistic modelling is proposed and an attempt is made to solve the problem with respect to both risk and profitability. The presented method is based upon multidimensional conditional optimization and allows usage of a wide variety of optimization criteria: risk, profit, or more complex functions. Using proposed method, a portfolio structure can be derived which corresponds to a given optimization criteria, or a Pareto optimum curve can be acquired representing various portfolios with best possible risk/profitability figures. It is also worth noting that various functions can be used for risk and profitability calculations, for example VaR for risk and ARIMA model prediction for expected profitability.

A sample problem of portfolio optimization is included in the paper. The problem described is a task of finding three portfolio compositions: a maximum-yield portfolio, a minimum-risk portfolio and an optimal portfolio structure with respect to both criteria. The solution is fully described, and the acquired results are explained.

Keywords: *portfolio, management, risk, optimization*

1 Model description

In the problem, a portfolio management model is approached. The model is defined in a discrete timeline; a day length is accepted as a time unit. The problem is solved within a time period divided into smaller 'investment periods' of the same length, Δ. Managing impulses are applied to a portfolio at the starting point of each investment period.

To describe the model, the following definitions are used:

$x(t) = (x_1(t), x_2(t), ..., x_n(t))$ – A vector with component values representing quantitative volumes of portfolio positions on corresponding securities at time point t (as a result, components of x vector can only be nonnegative at any time point);

n – number of securities allowed for inclusion into the portfolio;

$p(t) = (p_1(t), p_2(t), ..., p_n(t))$ – A vector with component values representing market quotes of corresponding securities at time point t;

$C(t)$ – Remaining cash in the portfolio at time point t;

$\Sigma(t) = S(x, p, C) = S(x(t), p(t), C(t))$ – cash equivalent of portfolio value at time point t;

$u(t+\Delta) = (u_1(t+\Delta), ..., u_n(t+\Delta)) = U(x(t), C(t), p(t+\Delta), Y(t+\Delta), R(t+\Delta))$ – Managements vector;

$Y(t) = Y(x(t-\Delta), C(t-\Delta), p(t-k), ..., p(t))$ – Yield function;

$R(t) = R(x(t-\Delta), C(t-\Delta), p(t-k), ..., p(t))$– Risk function;

k – Depth of historical sample used for calculation of management parameters.

2 Management problem formulation

Component values of p vector are defined by external factors (stock, bond, future and commodities markets, exchanges).

Cash equivalent of portfolio value is calculated as such:

$$\begin{aligned} &\Sigma(t) = \textstyle\sum_{i=1}^{n}[x_i(t)]p_i(t) + C(t); \\ &C(t) = \textstyle\sum_{i=1}^{n}\{x_i(t)\}p_i(t). \end{aligned} \tag{1}$$

$[x_i(t)]$is an integer part of $x_i(t)$ and $\{x_i(t)\}$is its fractional part.

Managing vector $u(t+\Delta)$ influence on portfolio structure can be described in incremental form:

$x(t+\Delta) = x(t) + u(t)$, or

$$\delta x(t+\Delta) = u(t).$$

Where Δ represents length of investment period as described above.

Management vector $u(t)$ has an economical sense: increments of portfolio positions at the beginning of $[t; t+\Delta]$investment period.

The model also incorporates different restrictions which may be applied to calculations of $u(t+1)$:

$$\sum_{i=1}^{N}(x_i(t) + u_i(t+\Delta))p_i(t+\Delta) = \sum_{i=1}^{N}[x_i(t)]p_i(t+\Delta) + C(t), \tag{2}$$

$$\frac{(x_i(t) + u_i(t+\Delta))p_i(t+\Delta)}{S(x(t), p(t+\Delta), C(t))} \leq \alpha, \tag{3}$$

$$\frac{(x_i(t) + u_i(t+\Delta))p_i(t+\Delta)}{S(x(t), p(t+\Delta), C(t))} \geq \beta. \tag{4}$$

Balance restriction (2) shows that portfolio structure that is being set for the investment period must have a market value equal to that of a portfolio value at the point of decision making (including free cash). This restriction is used to verify that money do not appear out of nowhere and do not disappear there.

Restrictions like (3) and (4) are examples the ones usually set by controlling organizations.

Optimization criteria can be either maximum of Yield function $Y(t)$, minimum of Risk function $R(t)$or some kind of more complex criterion.

In this paper, an example using Pareto optimization is described [3].

2.1 Determining Yield function

Yield function $Y(t)$ is determined in this paper as a mathematical expectation [1,2] of positive portfolio yields on a time period equal in length to investment period.

Portfolio yield over an investment period is calculated as follows:

$$y_{\Delta}(t) = \frac{\Sigma(t)}{\Sigma(t-\Delta)}. \tag{5}$$

Yield function can then be derived:

$$Y(t) = \frac{1}{P}\sum_{m\in\Omega} y_{\Delta}(t-m), \tag{6}$$

Where $\Omega = \{m : 0 \le m \le N-1,\ y_{\Delta}(t-m) > 0\}$;

N – size of historical sample of portfolio yields;

P – number of points in Ω.

The Yield function can be chosen with virtually no restrictions, the only requirement being is that the function could be calculated and translated into an algorithmic language. The simple Yield function taken in this paper is only for illustrative purposes.

2.2 Determining Risk function

Risk function $R(t)$ is determined as a Value-at-Risk [5] of the portfolio calculated using variance-covariance method with time horizon being equal to investment period length, historical sample of 250 points and probability $\alpha = 0.95$ [6]:

$$R(t) = q_{1-\alpha}\sqrt{S^T M_c S}, \tag{7}$$

$q_{1-\alpha}$– Percentile of an α-level normal distribution.

M_c – Portfolio securities yields covariation matrix.

S – Vector of elasticities (sensitivities of portfolio yield absolute changes to securities yields changes)

The Risk function can also be chosen with little to no restrictions, if that function can be calculated and translated into an algorithmic language. Risk function taken in this paper for illustrative purposes is based upon observations in earlier works [6].

2.3 Criteria of choosing securities for portfolio

As defined in the problem formulation, management is carried out via 'instant' impulse (which in the perspective of examined problem corresponds to a single trading day/session). Under such circumstances portfolio manager must firstly be absolutely sure to be able to buy or sell required number of securities, and secondly, to be able to make it with minimal transaction costs. Thus, security selection should be based upon at least these two criteria:

Securities are to be of high enough market liquidity to allow for buying and selling them in required amounts;

Securities are to have low bid/ask quote spread to allow for minimization of transaction costs [6].

3 Solution method description

The very first step in solving the described problem is to define the set of securities to be included in investment portfolio. Different principles can be used when selecting securities: the ones described above, for example, or others.

The goal after selecting securities is to find the Pareto set. For instance, two-dimensional (with only two securities) and three-dimensional (correspondingly, with three different securities) problems can be solved very easily via the direct approach lying in building a full Pareto set with given precision. However with increasing the model dimensionality, time consumption by calculations aimed to derive full set increases exponentially. In reality, however, full Pareto set is irrelevant – only a number of points are needed, and those points form the effective boundary of the full set. To find that effective boundary the Nelder-Mead method [4] is used because of its lack of sensitivity to model dimensionality.

Second step is to derive the portfolio with overall minimal risk $R_{\min}$ based upon optimization criteria $R(t) \to \min$. This structure will also determines the minimal acceptable yield $Y_{\min}$– all other structures have higher risk and those of them with lower yield are not acceptable. In the very same way the portfolio with overall maximal yield $Y_{\max}$is found ($-Y(t) \to \min$being the optimization cruteria) and its corresponding maximal acceptable risk $R_{\max}$– if it's exceeded in a portfolio, its yield will still be below $Y_{\max}$and thus such portfolios do not belong to the effective boundary.

Next step is to define required precision of calculating the Pareto boundary. A desired number of points N in the boundary is set, and that number of points is spread over one of the intervals, either $[R_{\min,}R_{\max}]$ or $[Y_{\min}Y_{\max}]$, dividing it into equal fractions. For each of these points a problem of finding portfolio corresponding to the best value of the opposite characteristic is solved. For example, if $[R_{\min,}R_{\max}]$ is broken up by N points, a portfolio is found for each point (each point has its exact risk value) which nets the highest possible yield under that risk.

The problems are solved by minimizing the following functions:

$R(t) + \Phi(Y) \to \min$ to find a portfolio of minimal risk if given exact yield value,

$-Y(t) + \Phi(R) \to \min$ to find a portfolio of maximal yield if given exact risk value, Here $\Phi(\cdot)$ represents penalty function increasing rapidly when risk or yield deviate from given values. This penalty function is used in demonstrational problem:

$$\Phi(Y) = k(Y(t) - Y^*)^2. \tag{8}$$

The last step results in a complete Pareto boundary created with the given precision.

The last step is needed only if the decision is to be made by an expert based upon consideration of all possible portfolios after comparing portfolio figures or their structure. In case of simple criteria like $R \to \min$(fully conservative portfolio with risk lowered as much as possible) or $Y \to \max$(highly aggressive portfolio aimed to achieve the best possible yield) or a kind of complex criterion which can be calculated and minimized only first two steps are needed.

4 Sample problem

As a sample problem, a task of managing a portfolio of three securities is chosen: $x(t) = (x_1(t), x_2(t), x_3(t))$. The problem has three management parameters only two of which, considering (1), are independent. Time interval on which management is required is set as one year. Each investment period is one month long, thus management interval consists of twelve investment periods.

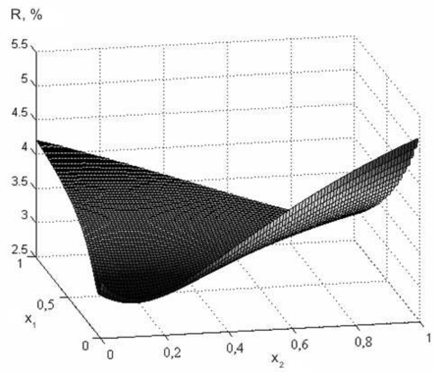
Sample problem includes restrictions applied to management criteria as shown in Table 21.

Three securities were chosen from a wide spectrum of available securities traded on an everyday basis on RTS trading. The main criterion for selection is daily liquidity and supplemental criterion (after initial selection) is a security's share in equal-weight portfolio's risk – marginal and incremental VaRs [5].

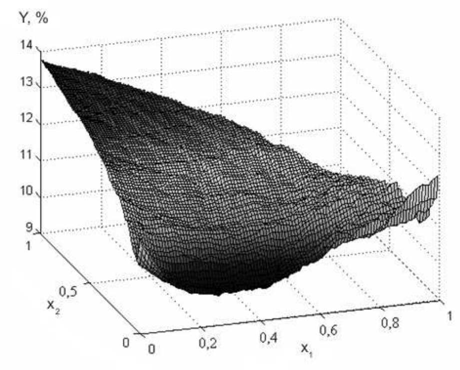
The problem is solved with two different approaches:

4.1 Straightforward approach

With only two independent criteria it is easy to build a set of possible criteria combinations with good precision. Graphs 1 and 2 show respectively profile of risk and profile of yield as 3D plots, in space of two management criteria (x_1, x_2), third criterion x_3 is dependent on first two. Management criteria on these graphs are presented in relative values (0–100% of portfolio) to be more visually representative. Graph 3 shows the Pareto set (acquired using data from Graphs 1 and 2) along with its effective boundary.



Graph 1: Profile of Risk function in respect to management criteria



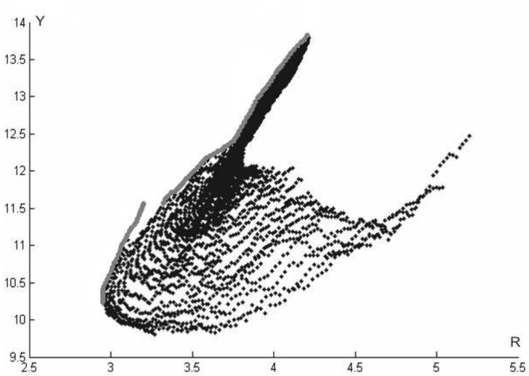
Graph 2: Profile of Yield function in respect to management criteria

4.2 Proposed approach

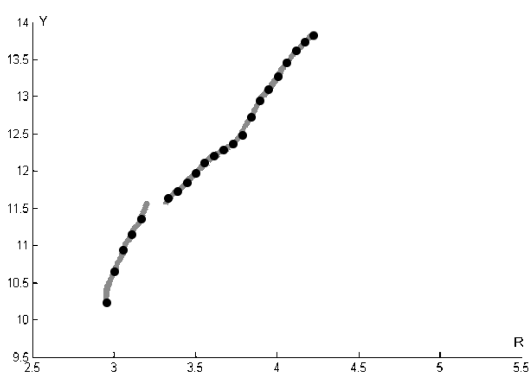
When solving the problem with approach discussed here, there's no need to calculate lots of portfolio variations as done in straightforward approach. On Graph 4 result of calculations is shown. The approximation of Pareto boundary was calculated with low precision specifically for illustrative purposes.

5 Results and explanation

Table 1 contains values of yield acquired in historical simulations while using different optimization criteria and comparison between yields acquired using passive management (portfolio is formed once and is not touched for the whole period of one year) and active management (at start of each investment period portfolio structure is subject to changes). Simulations were performed on real market data and represent the whole year 2007. Starting point of the period is January, 10 2007 and the ending point is January 9, 2008.



Graph 3: Full Pareto set and its effective boundary



Graph 4: Pareto set effective boundary and its approximation

Results in Table 1 show that active management performed using proposed method allows for greater yield than passive management. Values for Risk and Yield restrictions (third and

Criterion	Management principle	Yield, %
$R(t) \to \min$	Active	59,88
	Passive	49,65
$Y(t) \to \max$	Active	17,63
	Passive	4,88
$R(t) \to \min$ while $Y(t) \geq 12\%$	Active	76,87
	Passive	60,15
$Y(t) \to \max$ while $R(t) \leq 4\%$	Active	61,11
	Passive	57,57

Table 1: Results of managing portfolio during 12 investment periods

fourth criteria) are taken from expert analysis of market situation at the beginning of year 2007.

Comparison of the results with change of RTS Index within the management interval (RTSI has increased roughly by 35%) also shows positive results of active management.

References

[1] Aivazyan, S. A., Enjukov, I. S., Meshalkin, L. D. *Applied Statistics.* Finansy i statistica, Moscow, 1998, 1022 p.

[2] Kremer, N. S. *Probability Theory and mathematical statistics* – UNITI-DANA, Moscow, 2002, 543 p.

[3] Makarov, I. M., Vinogradskaya T. M., Rubchinskiy A. A., Sokolov V. V. *Theory of choice and decision-making*, Nauka, Moscow, 1982, 328 p.

[4] Nogin, V. D., Protodjakonov I. O., Evlampiev I. I. *Basics of optimization theory*, Vyshaya Shkola, Moscow, 1986, 384 p.

[5] Lobanov, A. A., Chugunov A. V. *Financial risk-manager* encyclopedia, Alpina-Publisher, Moscow, 2007, 878 p. ISBN 5-9614-0528-6

[6] Boyarshinov, A. M. *Mathematical methods of market risk valuation in application to Russian stock market.* Society for Computational Economics: Economics and Finance 2006, article no. 127. Available at `http://econpapers.repec.org/paper/scescecfa/127.htm`

Hana Brůhová-Foltýnová

Charles University Environment Centre in Prague

Ovocný trh 5, 116 36 Praha 1

e-mail: `bruhova@hotmail.com`

Econometric Estimation of the Value of Time in Urban Areas: The Case Study from Pilsen

Abstract

The aim of this paper is to apply advanced econometric techniques to estimate the monetary value of travel time in a Czech urban area. The model is cast in the discrete-choice framework and is applied on a data-set of actual transport behavior of Pilsen inhabitants. The data has been sampled using a standardized questionnaire survey carried out in June and July 2005.

To estimate the monetary value of the travel time, it is necessary to estimate the probability of the choice of a particular transport mode for a specific trip purpose. I consider the following transport modes: car, city public transport (tram, bus and trolleybus and their combinations), bicycle, and walking. The Cube software provides the trip characteristics (such as the duration and length of the trip based on the starting and destination addresses) for each trip and the used mode and also for all possible (alternative) modes, which could have been used for the trip.

I use a multinomial logit model. The set of explanatory variables include the trip costs, travel time, socio-demographic characteristics (such as the social status), and attitudes towards various transport modes. The choice is estimated for the following categories of the purposes of the reported trips on the reference day: journeys to work, and journeys to home. The value of time estimations seems to be quite robust to the model specifications. As can be expected, the value of the time differs according to the trip purpose. My results suggest that the current practice of the Ministry of Transport significantly overestimates the actual value of time.

Keywords: *Discrete Choice Models, Random Parameter Model, Value of travel time, passenger transport*

1 Introduction

The travel time values or the values of travel time savings are usually used for estimation of effects of changes in traffic velocity on behavior and also for assessment of social benefits of transport infrastructure projects. The Czech transportation policy lacks a rigorous estimation of the value of travel time. The Czech Ministry of Transport regularly publishes guidelines for assessment of economic efficiency of road infrastructure projects in the placecountry-regionCzech Republic using the program HDM-4. These guidelines suggest using the value of travel time which is settled a little above the Czech average wage (164 CZK/hour = about 6.56 EUR/hour in 2007). It is well understood that the subjective value of travel time would be equal to the individual wage rate only under the stringent of conditions, which require that the working time should be fully decided by the individual and that the time assigned to travel by any mode does not enter the individual's utility. Such conditions are rarely satisfied in reality and therefore the current practice of the Ministry should be investigated using independent pieces of research.

Given the expected large amounts of investment in transport infrastructure until 2013[1], co-financed from the EU Structural Funds, the need for an efficiency analysis of transport projects is boosted. This paper is a first attempt to estimate the value of time from revealed preference data in the placecountry-regionCzech Republic and thus it can contribute to assessing of transport infrastructure investments. Our results contribute to understanding whether the use of the travel time value based on the average wage as suggested by the Ministry of Transport is a good or poor approximation to the actual value.

To estimate the value of time, both the revealed preference (RP) and stated preference (SP) data have been used in various studies. Because the SP data usually express hypothetical situations, the RP data are likely to deliver more reliable results. Therefore this paper follows the RP approach.

2 Methodology

What is usually treated in the transport economics as the value of time is the subjective value of travel time (further as SV). It is the amount the individual is willing to pay in order to reduce by one unit his/her travel time [1]. Anyway comparing only travel time and travel costs is usually inadequate because in the subjective value also other characteristics influence the choice of the travel mode (the "quality" of each travel mode and a subjective perception of the travel mode) [2].

During the last thirty years the SV has been estimated using discrete travel-choice models in transport economics. The value is calculated as the rate of substitution between time and money from the estimated utility function. This represents the willingness to pay to increase the quality (or time) in one unit, which is:

$$SV_{kj} = \frac{\partial U/\partial q_{kj}}{\partial U/\partial I}, \tag{1}$$

where U is the indirect utility function, j represents alternatives (possible transport modes), k is the characteristics of the observation, the numerator expresses the increase in utility after an increase in quality q, and the denominator is the increase in utility after an increase in income I [1].

2.1 Econometric model

The revealed preference data are analyzed using a multinomial logit model for each alternative (each possible transport mode). This model allows for unobserved components of the utility. Multinomial logit models (MNL) are used to model relationships between a polytomous response variable and a set of regressor variables. Contrary to the generalized logit model or

[1] The transport infrastructure investments till 2013 represent about 5.77 bill. EUR, i.e. about 21.6% of the whole sum from the EU Structural Funds allocated for the placeCzech Republic, plus 1.01 bill. EUR co-financed from the Czech public funds (see http://www.strukturalni-fondy.cz/op-doprava).

the conditional logit model our model includes both characteristics of the individuals and the alternatives (different travel modes). An important property of MNL models is Independence from Irrelevant Alternatives (IIA), what means that the ration of the choice probabilities for any two alternatives for a particular observation is not influenced systematically by any other alternatives.

The utility functions of the model take the form

$$U_{ij} = \beta C_{ij} + \chi T_{ij} + \sum_{k} \alpha_{jk} x_{ki} + \varepsilon_{ij}, \tag{2}$$

where $i = 1, \ldots, n$ denotes individual, $j = 1, \ldots, m$ denotes choice alternative, x_{ki} is the kth regressor variable, which characterizes the individual i, and ε is the stochastic part of the utility function. C_{ij} is the cost of using the alternative j for the trip of the individual i, and T_{ij} is the time of the alternative j for the trip of the individual i. The Greek letters β, χ, α_j denote unknown parameters which have to be estimated. Note that the coefficients of the regressors β, χ are constant across alternatives, while coefficients not directly related to alternatives (such as socio-demographic characteristics of individuals α_{jk}) should be alternative specific. This means that different socio-demographic groups are allowed to have different preference shifters towards the transport modes. That is why the coefficients α_{jk} take the alternative-specific subscripts, while the coefficients β and χ do not. However, the coefficients associated to the socio-demographic characteristics are not identified. To achieve their identification, the usual choice is to set them equal zero for one of the alternatives [3]. I follow this practice and set to zero the coefficient of the last alternative—cycling.

The parameters are estimated using the maximum likelihood method. The standard errors and the p-values are computed using the non-parametric bootstrap (I choose 1000 bootstrap repetitions).

Taking the derivatives of the utility function (2) with respect to the monetary costs and by time, the following expression for the value of time is reached [1,2]:

$$SV = \frac{\partial U_{ij}/\partial T_{ij}}{\partial U_{ij}/\partial C_{ij}} = \frac{\chi}{\beta}$$

Note that the expression does not depend on i or j, which is a consequence of the multinomial-logit formulation.

Therefore, the mode choice is explained by the monetary costs, travel time, and selected socio-economic characteristics (age, sex, number of the household members, ownership of a driving license), and some attitudes towards transport and the environment. The attitudes are measured on a five-point Likert-type scale and include approaches towards recycling (Question: 'How often do you recycle the household waste (grass, plastic, papers) because of the environment? Answer: Always—never'), the existing behavior saving the environment (Question: 'I do what is good for the environment no matter that is costs more or takes more time. Answer: Definitely disagree—definitely agree') and willingness to pay more for the environment (Question: 'How often are you willing to pay higher prices to save the environment? Answer: Very much unwilling—very much willing'). The choice is estimated for two categories of the purposes of the reported trips on the reference day: journeys to work, and journeys to home.

In the case of journeys to home we distinguish the purpose of the journey (from work, from shopping and from other destinations).

2.2 Data collection

To estimate the value, we use data for passenger transport. The data come from a standardized questionnaire survey carried out in June and July 2005 in the Czech city of CityplacePilsen. The sample population includes adults over 18 years old (N=763) from two residential areas (Slovany and Lochotn). A quota sample (residential area, age, gender, education level) was used. The sample is representative of age (18-84), gender and education structure. There are 47.8% of men and 52.2% of women in the sample. 15.8% of respondents have a university education; about 42% secondary and 8.0% have primary education. The data collection was carried out on working days (Wednesdays, Thursdays and Fridays) in the period of June and July 2005 in the respondents' households.

The questionnaire includes (i) a description of regular week journeys (within the city), including frequency and used modes of transport; (ii) a diary—a description of all journeys made during the previous (reference) day (an ordinary working day in the data collection period); (iii) stated motivation for the use of the particular transport mode; (iv) stated individual attitudes towards various aspects of transport modes (price, time, availability, reliability, safety, convenience) and the environment; and (v) economic and socio-demographic characteristics of the respondents and their families.

All trips from the previous day done by the respondents on the Pilsen city area are observed as an independent chain. The following alternatives for transport mode are included: car (divided on a driver and a car passenger), public transport (tram, bus and trolleybus and their combinations), bicycle, and walking (the minor modes like train or intercity bus are excluded[2]). The trip characteristics (such as the duration and length of the trip and the trip costs) for each trip and for each alternative mode are calculated by the Cube software. (Data on the start and destination addresses of all trips, the transport mode used and the trip purpose come from the questionnaire-diaries.) The Cube software is owned and run by the Správa veřejného statku města Plzně (Authority of the Municipal Goods, allowance organization of the city of Pilsen)[3].

The (marginal) costs for each mode are calculated as following. The cost of the car use is settled as 4 CZK per kilometer driven. If the car is used for business (paid by the employer), the price of the car use is zero. The price of public transport is zero for owners of time tickets (month or year ticket) and 14 CZK per one trip for the others. The costs of walking and cycling are zero.

3 Main findings

I run several models with different regressors to check the robustness of the results. Nevertheless, all models include the time and the monetary costs. For all models, the time and cost

[2] The Cube software is calibrated only for the area of city of CityplacePilsen.

[3] Cube is a transportation planning software system designed for forecasting of passengers and freight movements developed by Citylabs Inc. Cube offers advanced and flexible tools for the generation, distribution, mode split and assignment of personal and freight transport as well as detailed analysis of environmental issues. It consists from 3 components: GIS, Application Manager (flow-chart style interface for designing and developing the model stream), and Scenario Manager (tool kit for developing user interface to the model and to create, manage and run tests, including full batch capabilities).

coefficients are significant and have the expected sign (parameters β and $\chi$from the equation 2). On the contrary, the attitudes and socio-demographic characteristics (parameters $\alpha_j$from the equation 2) are not always significant, but have the expected sign.

Results for six selected models are organized in Tables 1 and 2. Table 1 displays the results for the journeys to work.

Variable		**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Time		−0.09850***	−0.09738***	−0.09582***	−0.10432***	−0.11719***	−0.10556***
Cost		−0.12672***	−0.13362***	−0.14235***	−0.11946***	−0.12456***	−0.12240***
Constant	PT	−0.14575	+0.05955	+0.86881	+0.20234	+1.57927‡	+0.36135
	Car	−20.85303***	−17.72107***	−16.05045***	−18.17954***	−47.55230***	+0.84520
	Walk	−1.12077	+0.71502	−0.21898	−0.21187	+1.64815	+1.20284
Sex	PT	+1.83857**	+1.83563*	+1.81105**	+1.83698**	+2.16494**	+1.82249**
	Car	+0.89477‡	+0.84575‡	+0.72486‡	+0.89506‡	+1.58384*	+0.39183
	Walk	+2.30534*	+2.19295*	+2.25007*	+2.31213**	+4.40912***	+1.93781**
No of Household Members	PT	+0.17971*	+0.15286†	+0.19541*	+0.21894*		+0.24442*
	Car	+0.24036**	+0.21301*	+0.27825**	+0.24568*		+0.21476*
	Walk	−0.59631‡	−0.70902†	−0.57803‡	−0.61207†		−0.53751†
Driving license	PT	+0.36784	+0.34266	+0.10539	+0.29494	+0.23861	
	Car	+19.80793***	+17.92170***	+19.22440***	+18.78119***	+50.13590***	
	Walk	+1.53311†	+1.80638*	+1.18989‡	+1.43768†	+1.87793*	
Age	PT					−5.82946	
	Car					−19.55394*	
	Walk					−57.42747**	
Age^2	PT					+5.91569	
	Car					+27.57920*	
	Walk					+78.14375†	
Recycling	PT	+0.23222	+0.26735	+0.28693			
	Car	+0.99467*	+0.98295**	+0.83475†			
	Walk	+0.39617	+0.39490	+0.48511			
I do well for Environment	PT			−0.21449‡			
	Car			−1.10376**			
	Walk			−0.22332‡			
Willingness to pay more for Env.	PT		−0.03374				
	Car		−0.35491‡				
	Walk		−0.65629*				
Pseudo R2		0.50786	0.52424	0.53141	0.48435	0.50524	0.41669
Log likelihood		−117.3478	−113.4422	−112.381	−124.3825	−119.3436	−140.704
SV (CZK/min)		0.7773	0.7288	0.6731	0.8733	0.9408	0.8624
SV (CZK/hour)		46.6383	43.7270	40.3878	52.3958	56.4499	51.7451

Note: Minimum significance level 0.2‡; 0.1†; 0.05; 0.01**; 0.000***. PT = city public transport.*

Table 1: Model Results for travels to work

The regressors of age, age^2 and attitudes towards recycling and the environment are in general less significant than the regressors of the number of household members, ownership of the driving license or sex. The similar regressors significance was received also for trips to home.

Table 2 summarizes results of models for trips to home. The same six models as for trips to work were modeled. Further the models distinguish the origin point of the trip home: It is specified, if the trip was from work, from shopping, or from other starting points. Thanks to

Variable		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Time		−0.08995***	−0.08996***	−0.09015***	−0.08903***	−0.09505***	−0.08368***
Money		−0.07058***	−0.07190***	−0.07517***	−0.07248***	−0.07612***	−0.07853***
Time Work		−0.00542	−0.00527	−0.00528	−0.00630	−0.00780	−0.02007*
Time Shop		+0.04145**	+0.04164**	+0.04224**	+0.04119*	+0.04425**	+0.03006*
Constant	PT	−0.60406	−0.23468	−0.16822	+0.07635	+1.02932‡	+0.31862
	Car	−0.60406	−5.85878***	−4.01759**	−5.72795***	−4.09121**	−0.67784‡
	Walk	+0.30996	+0.93456‡	+0.89286	+0.23251	+1.07010‡	+0.81441‡
Sex	PT	+2.00381***	+1.99085***	+2.05616***	+1.90500***	+2.13565***	+1.81538***
	Car	+1.76612***	+1.75391**	+1.91009***	+1.69055***	+2.53722***	+1.03893*
	Walk	+1.89903***	+1.85341**	+1.96447***	+1.84959***	+2.17309***	+1.65679***
No of Household Members	PT	−0.10855*	−0.11262†	−0.09542†	−0.09400†		−0.10553†
	Car	+0.04505‡	+0.04756‡	+0.05153‡	+0.04948‡		−0.00975
	Walk	−0.19480**	−0.20790**	−0.18419*	−0.20885**		−0.21632**
Driving license	PT	+0.02507	−0.07459	−0.09932	+0.04064	+0.01597	
	Car	+4.66647***	+4.56165***	+4.19415***	+4.63995***	+4.85589***	
	Walk	+0.48988‡	+0.35182	+0.35771	+0.46084‡	+0.49888‡	
Age	PT					−9.16772†	
	Car					−23.49958**	
	Walk					−13.30908*	
Age^2	PT					+9.56945‡	
	Car					+32.49530**	
	Walk					+16.80771*	
Recycling	PT	+0.31983†	+0.30422†	+0.30011†			
	Car	+0.25560‡	+0.24449‡	+0.06712			
	Walk	−0.11904	−0.12837	−0.15210			
I do well for Environment	PT			−0.10346‡			
	Car			−0.52275***			
	Walk			−0.14791‡			
Willingness to pay more for Env.	PT		−0.08433‡				
	Car		−0.14049‡				
	Walk		−0.15528†				
Pseudo R2		0.3709	0.37413	0.38302	0.36417	0.37503	0.2832
Log likelihood		−276.4617	−275.0439	−271.1349	−279.4189	−274.6452	−314.9998
SV (CZK/min)		1.2744	1.2512	1.1993	1.2283	1.2487	1.0656
SV (CZK/hour)		76.4664	75.0709	71.9569	73.7003	74.9212	63.9348
SV/work (CZK/min)		1.3512	1.3245	1.2695	1.3153	1.3512	1.3212
SV/work (CZK/hour)		81.0740	79.4687	76.1713	78.9156	81.0694	79.2691
SV/shop (CZK/min)		0.6872	0.6720	0.6374	0.6600	0.6674	0.6828
SV/shop (CZK/hour)		41.2298	40.3227	38.2413	39.6026	40.0420	40.9678

Note: Minimum significance level 0.2‡; 0.1†; 0.05; 0.01**; 0.000****

Table 2: Model Results for travels to home

this separation it is possible to receive the subjective value of travel time according to the trip purposes (see Table 2).

Significant regressors are above all sex, number of household members, and the ownership of the driving license. The attitudes are in general less significant regressors, but they influence the SV results—in the models where attitudes are included (Model 1-3) the SV is higher in comparison with models without any attitude variables.

4 Discussion and conclusions

The estimations of the subjective value of travel time are quite robust to the model specifications. Much lower values than recommended by the Czech Ministry of Transport were obtained for all trip purposes. The estimated values vary between 40 and 56 CZK (compared to 164 CZK) for trips to work and between 38 and 74 CZK for trips to home. The SV varies according to the trip purposes. The highest SV was estimated for trips from work to home (from 76 to 81 CZK). On the contrary, the lowest SV was received for trips to home from shopping (from 38 to 41 CZK). The results confirm that the SV of leisure trips (in concordance with results of similar foreign studies[4]) is lower than the SV of trips to or from work.

In the future, I plan to use the data in more depth. First, I plan to include also other trip purposes like leisure. Second, I want to estimate the model also for mode choices on a regular day in addition to the reference day. Finally, I plan to combine these two datasets using a discrete-choice panel data technique. Using an appropriate procedure, it is possible to model a different degree of trust in the two data sources combined in a single econometric model.

References

[1] JARA-DÍAZ, S. *Transport Economic Theory.* Elsevier Ltd., 2007. ISBN 978-0-08-045028-5

[2] HENSHER, D.A., BUTTON, K.J. (Eds.) *Handbook of Transport Modelling.* Elsevier Ltd., 2000. ISBN 0-08-043594-7

[3] TRAIN, K.E. *Discrete Choice Methods with Simulation.* Cambridge University Press, Cambridge, 2003. ISBN 0-521-81696

Acknowledgement

The research was supported by the Grant from the Czech Science Foundation No. 402/07/P484 'Theoretical Inquiries on Transport Regulation'.

[4] Such as results of the European projects UNITE, Deliverable 7 (see `www.its.leeds.ac.uk/projects/unite/`) a MC-ICAM, Deliverable 3 (see `http://www.its.leeds.ac.uk/projects/mcicam/index.html`).

Michal Černý, Irena Šindelářová

Department of Econometrics
University of Economics Prague
W. Churchill sq. 4, 130 67 Prague 3

e-mail: `cernym@vse.cz, isin@vse.cz`

Some Pictures and Scripts for Teaching IPMs

Abstract

In the text we discuss some geometric aspects of linear programming and provide a visualisation tool for a type of a short-step central-path following algorithm. The tool is intended to help in teaching interior point algorithms.

Keywords: *linear programming, central path algorithms, short-step algorithms, Khachiyan's grid, visualisation*

The nice property of linear programming, contrary to many other fields of mathematics, is that its nature is essentially geometric. Therefore, such important terms as polyhedron, analytic center, central path etc. may be quite easily visualised when their definitions are taught. Moreover, one can often visualise also the ideas underlying particular important theorems: for instance, the duality theorem—probably the most important result in theory of linear programming—has an intuitive geometric, sometimes called "billiard", meaning (see [8], p. 93).

A lot of ideas underlying LP algoritms are of quite geometric nature, too. An example of a wonderful and easy geometric idea is the well-known Shor-Yudin-Nemirovskij's ellipsoid method. Even the Khachiyan's algortihm may be visualised: it may be regarded as the ellipsoid method where all "important" points, say vertices of the bounded full-dimensional rational polyhedron (if nonempty) and centres of the ellipsoids, lay on a special grid: a grid of finitely many rational points with bounded bit-size. Rounding of coefficients of matrices describing the ellipsoids may be also geometrically demonstrated: if the ellipsoid is coordinate-aligned and centered at zero, rounding eigenvalues to a restricted bit-size number may be seen as shortening semi-axes so that their lengths (and also squares of their lengths) are points on the Khachyian's grid. Hence, in some sense Khachyian's algorithm may be visualised as a discrete version of the ellipsoid method.

To give one more example: a well-known nice geometric idea is the Karmarkar's affine transformation, nicely described e.g. in [7]. Interesting geometric aspects of the Karmarkar potential are given in [9].

The aim of this text is to provide teachers of linear programming with a simple visualisation tool of a version of a central-path algorithm which will be sketched later; it is a nice representative of the entire family of central-path short-step algorithms and is suitable to be presented in lectures. We choose this algorithm as (i) it is one of the algorithms with so far best-known

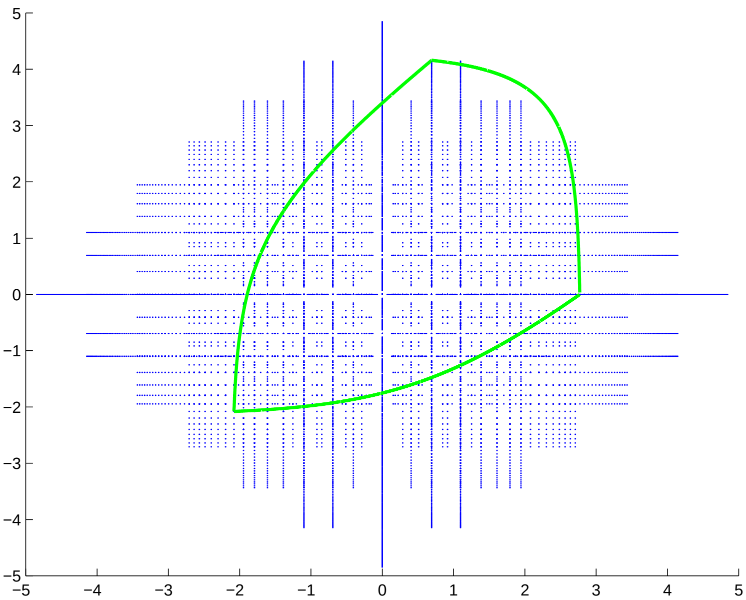


Figure 1: Khachiyan's grid: points $[\frac{a}{b}, \frac{c}{d}] \in \mathbb{R}^2$ *with integral* $a, b, c, d > 0$ *such that sum of bit sizes of* a, b, c *and* d *does not exceed 10 bits (10 bits are taken just as an example) and a polyhedron with vertices on the grid—a triangle in* $\mathbb{R}^2$ *with vertices* $[\frac{1}{8}, \frac{1}{8}], [2, 64]$ *and* $[16, 1]$*. Both axes are log-scaled.*

complexity (its iteration bound is $O(\sqrt{n}L)$, where n denotes dimension and L is the bit-size of the linear program); (ii) its idea is nicely explained in the excellent book [4]. The book does not give full theoretical analysis, however the idea behind the algorithm is shown in a clear way which is suitable for many undergraduate linear-programming courses (nevertheless, a full analysis exceeds the level of such lectures; an almost-full analysis is found in [7]).

The algorithm. Assume we are given a full-rank matrix $\boldsymbol{A}$ and vectors $\boldsymbol{b}, \boldsymbol{c}$ (all vectors are column) and our task is either to find $\boldsymbol{x}^* \in \operatorname{argmax}\{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}, \boldsymbol{x} \ge \boldsymbol{0}\}$ (for our purposes, an approximate optimum is sufficient) if such exists or state that no such exists. (In theory, one usually works with rational arithmetic; here we omit the arithmetic-precision considerations. Our task is just to give a didactic tool to plot some pictures, so we may admit that $10^{-10} = 0$ or so.) Let us be given a small positive number ϵ which will govern the exactness of the algorithm.

For the purpose of visualisation we will assume the dimension $n = 2$ and $\mathcal{F} = \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}, \boldsymbol{x} \ge 0\}$ is nonempty, bounded and has dimension two. (This means the the optimum exists.) However, we describe the algorithm in general. Let m be the number of rows of $\boldsymbol{A}$.

First, we construct a type of a self-dual embedding of the linear program $\operatorname{argmax}\{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}, \boldsymbol{x} \ge \boldsymbol{0}\}$. Let

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{A} & -\boldsymbol{b} \\ -\boldsymbol{A}^{\mathrm{T}} & \boldsymbol{0} & \boldsymbol{c} \\ \boldsymbol{b}^{\mathrm{T}} & -\boldsymbol{c}^{\mathrm{T}} & 0 \end{pmatrix}, \quad \boldsymbol{d} = \boldsymbol{1} + \boldsymbol{D}\boldsymbol{1}, \quad \boldsymbol{E} = \begin{pmatrix} \boldsymbol{D} & -\boldsymbol{d} \\ \boldsymbol{d}^{\mathrm{T}} & 0 \end{pmatrix} \tag{1}$$

Now set

$$\boldsymbol{q} = \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{x} \\ \tau \\ \eta \end{pmatrix}, \quad \boldsymbol{e} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ n+m+2 \end{pmatrix}. \tag{2}$$

Here, $\boldsymbol{q}$ is a vector of variables: $\boldsymbol{x}$ has $n = 2$ components and $\boldsymbol{y}$ has m components. Now the following holds:

(i) the linear program $\operatorname{argmax}\{-\boldsymbol{e}^{\mathrm{T}}\boldsymbol{q} : \boldsymbol{E}\boldsymbol{q} \le \boldsymbol{e}, \boldsymbol{q} \ge \boldsymbol{0}\}$ is bounded and feasible and for every optimal solution $(\boldsymbol{q}^*)^{\mathrm{T}} = ((\boldsymbol{y}^*)^{\mathrm{T}}, (\boldsymbol{x}^*)^{\mathrm{T}}, \tau^*, \eta^*)$, $\boldsymbol{x}^*, \boldsymbol{y}^*$ and τ^* form a solution to the Goldman-Tucker system

$$\boldsymbol{A}\boldsymbol{x} - \tau\boldsymbol{b} \le \boldsymbol{0},\ -\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} + \tau\boldsymbol{c} \le \boldsymbol{0},\ \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \le \boldsymbol{0},\ \boldsymbol{x} \ge \boldsymbol{0},\ \boldsymbol{y} \ge \boldsymbol{0},\ \tau \ge 0;$$

(ii) every strictly complementary optimal solution of $\operatorname{argmax}\{-\boldsymbol{e}^{\mathrm{T}}\boldsymbol{q} : \boldsymbol{E}\boldsymbol{q} \le \boldsymbol{e}, \boldsymbol{q} \ge \boldsymbol{0}\}$ is a solution of the Goldman-Tucker system with either $\tau > 0$ or $\boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} < 0$,

(iii) $\boldsymbol{q} = \boldsymbol{1}$, $\boldsymbol{p} = \boldsymbol{1}$ are feasible for the system $\boldsymbol{E}\boldsymbol{q} + \boldsymbol{p} = \boldsymbol{e}$, $\boldsymbol{q} > \boldsymbol{0}$, $\boldsymbol{p} > \boldsymbol{0}$.

By the Goldman-Tucker theorem, either there exists a solution to the Goldman-Tucker system with $\boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} < 0$, and then the real optimum of the original program $\operatorname{argmax}\{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \boldsymbol{x} \geq \boldsymbol{0}\}$ does not exist, or there exists a solution with $\tau > 0$, and then $\frac{1}{\tau}\boldsymbol{x}$ is optimum of the original linear program (and, moreover, $\frac{1}{\tau}\boldsymbol{y}$ is optimum of its dual program).

So now we have a system $\boldsymbol{E}\boldsymbol{q} + \boldsymbol{p} = \boldsymbol{e}$, $\boldsymbol{q} \geq \boldsymbol{0}$, $\boldsymbol{p} \geq \boldsymbol{0}$ where we know an interior feasible point in advance; moreover it is a "good" starting point for the algorithm.

To get an optimum of the original system (or conclude that it does not exist) we need to find a strictly complementary solution of another system. This is the essence of the algorithm: it gets the linear program

$$\operatorname{argmax}\{-\boldsymbol{e}^{\mathrm{T}}\boldsymbol{q} : \boldsymbol{E}\boldsymbol{q} + \boldsymbol{p} = \boldsymbol{e}, \boldsymbol{q} \geq \boldsymbol{0}, \boldsymbol{p} \geq \boldsymbol{0}\} \tag{3}$$

as its input and will converge to a such a solution.

At the beginning, we set $\mu := 1$ and exponentially fast decreasing μ we will approximately follow the central path $\{\operatorname{argmax}\{(\boldsymbol{q}, \boldsymbol{p}) : \mu \sum_{i=1}^{m+n+2}(\ln \boldsymbol{p}_i + \ln \boldsymbol{q}_i) - \boldsymbol{e}^{\mathrm{T}}\boldsymbol{q}\}; \mu \in (0, 1]\}$. To (approximately) follow the central path (keeping within the quadratic-convergence region), we have to start in a "good" point: indeed, the initial point $\boldsymbol{p} = \boldsymbol{1}$, $\boldsymbol{q} = \boldsymbol{1}$ is a point on the central path with $\mu = 1$.

Given μ, the point on the central path $\operatorname{argmax}\{(\boldsymbol{q}, \boldsymbol{p}) : \mu \sum_{i=1}^{m+n+2}(\ln \boldsymbol{p}_i + \ln \boldsymbol{q}_i) - \boldsymbol{e}^{\mathrm{T}}\boldsymbol{q}\}$ may be written as a solution to a non-linear system $\boldsymbol{E}\boldsymbol{q} + \boldsymbol{p} = \boldsymbol{e}$, $\boldsymbol{q}_i \cdot \boldsymbol{p}_i = \mu$ for all $i = 1, 2, \ldots, n+m+2$, $\boldsymbol{p} \geq \boldsymbol{0}$, $\boldsymbol{q} \geq \boldsymbol{0}$. Given an approximate point near the central path $(\boldsymbol{q}, \boldsymbol{p})$, we want to update $\boldsymbol{q} := \boldsymbol{q} + \boldsymbol{\alpha}$, $\boldsymbol{p} := \boldsymbol{p} + \boldsymbol{\beta}$ by performing a Newton-like step: substituting this into the system, we get $\boldsymbol{E}(\boldsymbol{q} + \boldsymbol{\alpha}) + (\boldsymbol{p} + \boldsymbol{\beta}) = \boldsymbol{e}$, $(\boldsymbol{q}_i + \boldsymbol{\alpha}_i) \cdot (\boldsymbol{p}_i + \boldsymbol{\beta}_i) = \mu$ for all i's, $(\boldsymbol{q} + \boldsymbol{\alpha}) \geq \boldsymbol{0}$, $(\boldsymbol{p} + \boldsymbol{\beta}) \geq \boldsymbol{0}$ and we neglect the second-order terms $\boldsymbol{\beta}_i \cdot \boldsymbol{\alpha}_i$. Hence, we solve the linear system

$$\begin{pmatrix} \boldsymbol{E} & \boldsymbol{J} \\ \operatorname{diag}(\boldsymbol{q}) & \operatorname{diag}(\boldsymbol{p}) \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ \mu \cdot \boldsymbol{1} - \boldsymbol{q} * \boldsymbol{p} \end{pmatrix} \tag{4}$$

($\boldsymbol{J}$ stands for unit matrix, $\operatorname{diag}(\boldsymbol{q})$ for a diagonal matrix with entries of $\boldsymbol{q}$ on the diagonal and $\boldsymbol{q} * \boldsymbol{p} = (\boldsymbol{q}_1 \cdot \boldsymbol{p}_1, \boldsymbol{q}_2 \cdot \boldsymbol{p}_2, \ldots, \boldsymbol{q}_{n+m+2} \cdot \boldsymbol{p}_{n+m+2})^{\mathrm{T}}$) where $\boldsymbol{q}, \boldsymbol{p}$ act as constants and $\boldsymbol{\alpha}, \boldsymbol{\beta}$ as unknowns. It may be shown that such an „approximate" step from $(\boldsymbol{q}, \boldsymbol{p})$ to $(\boldsymbol{q} + \boldsymbol{\alpha}, \boldsymbol{p} + \boldsymbol{\beta})$ fulfills the constraints $\boldsymbol{q} + \boldsymbol{\alpha} > \boldsymbol{0}, \boldsymbol{p} + \boldsymbol{\beta} > \boldsymbol{0}$ and does not deviate from the exact central path too far. (With some further analysis, this implies that the procedure converges.) Now we may summarize the algorithm. Of course, we are not correct... To be fully correct, we would have to compute in rational arithmetic and perform further analysis, when it is appropriate to state that no optimum exists, and also determine the error of the approximate solution (or perform the usual final "2^{-2L}-truncation" step to get an exact optimum). We will not do this here; for purposes of visualisation, we assume the optimum exists, and we do not need to estimate the error in detail.

input: $\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}, \epsilon$

output: approximate $\boldsymbol{x} \in \operatorname{argmax}\{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \boldsymbol{x} \geq \boldsymbol{0}\}$ or report `optimum does not exist`

[1] Construct $\boldsymbol{E}$ as in (1) and set $\mu := 1$, $\boldsymbol{q} := \boldsymbol{1}$, $\boldsymbol{p} := \boldsymbol{1}$;

[2] if $\mu < \epsilon$ then:

look at $\boldsymbol{q}^{\mathrm{T}}$ as $(\boldsymbol{y}^{\mathrm{T}}, \boldsymbol{x}^{\mathrm{T}}, \tau, \eta)$; if $\tau \approx 0$, report `optimum does not exist`; otherwise return $\frac{1}{\tau}\boldsymbol{x}$ as the approximate solution. Stop.

[3] Set $\mu := (1 - \frac{1}{2\sqrt{n}})\mu$;

[4] find $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ by solving the linear system (4) with current $\boldsymbol{q}$ and $\boldsymbol{p}$, update $\boldsymbol{q} := \boldsymbol{q} + \boldsymbol{\alpha}$, $\boldsymbol{p} := \boldsymbol{p} + \boldsymbol{\beta}$ and go to [2].

Now it would be necessary to show that the procedure converges to the strictly complementary solution of (3); moreover, it holds that if an optimum exists, then the algorithm converges to the analytic centre of the optimal face.

We will plot $\mathcal{F} = \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \boldsymbol{x} \geq \boldsymbol{0}\}$ and the trajectory of the algorithm: to plot the trajectory, in each step, we look at current $\boldsymbol{q}^{\mathrm{T}}$ as $(\boldsymbol{y}^{\mathrm{T}}, \boldsymbol{x}^{\mathrm{T}}, \tau, \eta)$ as in step [2] and we plot the point $\frac{1}{\tau}\boldsymbol{x}$.

However, the algorithm in fact does not pass through $\mathcal{F}$ but a more-dimensional polyhedron

$$\{\boldsymbol{q} : \boldsymbol{E}\boldsymbol{q} \leq \boldsymbol{e}, \boldsymbol{q} \geq \boldsymbol{0}\}. \tag{5}$$

So it is instructive, in each step, to plot the projection of this polyhedron into the plane, too.

Given a definition $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \boldsymbol{x} \geq \boldsymbol{0}$ of a nonempty two-dimensional polyhedron, to visualise the analytic centre it suffices to run the algorithm with $\boldsymbol{c} = \boldsymbol{0}$; it is instructive to see how the analytic centre moves if we, for instance, add some redundant inequalities into the defining system (e.g., some inequalities are present more than once).

At the site `http://nb.vse.cz/~cernym/ipm.zip`, some MatLab scripts plotting the pictures, that may help in teaching, are available. The function

`ip(`$\boldsymbol{A}$`,`$\boldsymbol{b}$`,`$\boldsymbol{c}$`,mode)`

gets an $m \times 2$ matrix $\boldsymbol{A}$, a column vector $\boldsymbol{b}$ with m entries and a column vector $\boldsymbol{c}$ with two entries such that the feasible region $\{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \boldsymbol{x} \geq \boldsymbol{0}\}$ is two-dimensional, nonempty and bounded and plots the trajectory of the described algorithm. If `mode` $= 1$, then the projections of (5) into $\mathbb{R}^2$ are in each step plotted too.

The function `getmatrix(`$\boldsymbol{x}$`,`$\boldsymbol{y}$`)`, where $\boldsymbol{x}$ and $\boldsymbol{y}$ are column vectors with l entries, first computes the convex hull $\mathcal{C}$ of the points $[\boldsymbol{x}_1, \boldsymbol{y}_1], \ldots, [\boldsymbol{x}_l, \boldsymbol{y}_l]$ and returns a matrix $(\boldsymbol{A} \quad \boldsymbol{b})$ such that $\mathcal{C} = \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$. For instance, the command

```
Ab =  getmatrix([2 + cos([0:2*pi./7:2*pi])]',
      [3 + sin( [0:2*pi./7:2*pi])]')
```

forms a system of inequalities which define a regular septahedron centered at $[2,3]$. Now calling

```
ip(Ab(:,1:2),Ab(:,3),[0 1]',0)
```

and

```
ip(Ab(:,1:2),Ab(:,3),[0 1]',1)
```

yields the following pictures.

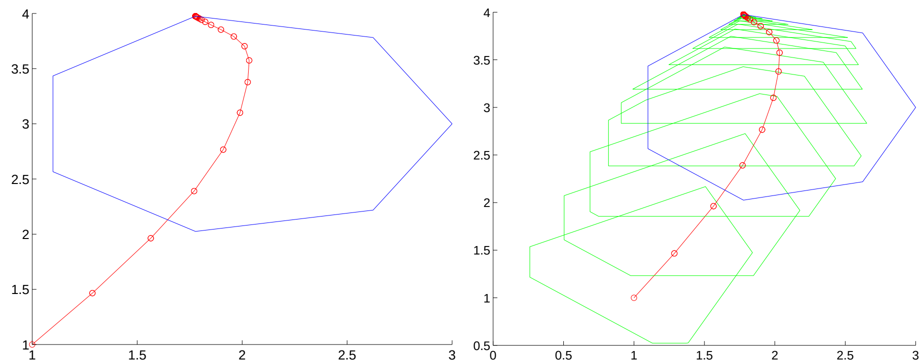


Figure 2: Trajectory of the algorithm

Calling `ip(Ab(:,1:2),Ab(:,3),[-1 0]',` *mode*`)` (again with either *mode*=`0` or *mode*=`1`) shows how the algorithm converges to the analytic centre of the optimal face (which is the vertical edge of the septahedron). Also note how the algorithm follows the central path $\{\text{argmax}\{(\boldsymbol{q},\boldsymbol{p}) : \mu \sum_{i=1}^{m+n+2}(\ln \boldsymbol{p}_i + \ln \boldsymbol{q}_i) - \boldsymbol{e}^{\mathrm{T}}\boldsymbol{q}\}; \mu \in (0,1]\}$: first the direction is governed by the terms $\ln \boldsymbol{p}_i + \ln \boldsymbol{q}_i$ and then by $-\boldsymbol{e}^{\mathrm{T}}\boldsymbol{q}$.

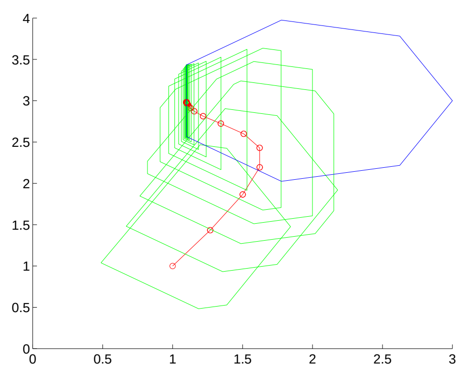


Figure 3: Convergence to the analytic centre of the optimal face

The function `ip2(`$\boldsymbol{A}_1$`,`$\boldsymbol{b}_1$`,`$\boldsymbol{c}_1$`,`$\boldsymbol{A}_2$`,`$\boldsymbol{b}_2$`,`$\boldsymbol{c}_2$`, mode)` plots two trajectories of two linear programs. It is, for instance, instructive to see how the analytic centre moves if we change the definition of the same polyhedron. For example, first generate a triangle

```
Ab =  getmatrix([2 + cos([0:2*pi./3:2*pi])]',
      [3 + sin( [0:2*pi./3:2*pi])]')
```

and call

```
ip2(Ab(:,1:2),Ab(:,3),[0 0]',
   [Ab(:,1:2); Ab(1, 1:2)], [Ab(:,3); Ab(1,3)], [0, 0]', 0)
```

The system $\boldsymbol{A}_1\boldsymbol{x} \leq \boldsymbol{b}_1$ defines the triangle with three inequalities; the system $\boldsymbol{A}_2\boldsymbol{x} \leq \boldsymbol{b}_2$ with four inequalities: the fourth is a redundant copy of the first one. (Note that if $\boldsymbol{c} = \boldsymbol{0}$ then the optimal face is the entire polyhedron and hence the algorithm converges to its analytic centre.)

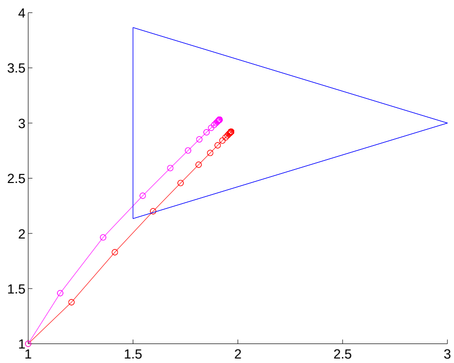


Figure 4: Convergence to analytic centres of the same polyhedron with two different definitions

As the algorithm converges to the analytic centre of the optimal region, it is suitable for L_1-regression. It is well-known that given points $[\boldsymbol{x}_i^{\mathrm{T}}; y_i]$, $i = 1, 2, \ldots, N$, finding $\widehat{\boldsymbol{\beta}}_{L_1} \in \operatorname{argmin}_{\boldsymbol{\beta}}\{\sum_{i=1}^{N} |y_i - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i|\}$ is equivalent to the linear program $\operatorname{argmax}_{(\boldsymbol{\beta},\boldsymbol{r})}\{-\boldsymbol{1}^{\mathrm{T}}\boldsymbol{r} : \boldsymbol{X\beta}+\boldsymbol{r} \geq \boldsymbol{y}, \boldsymbol{X\beta} - \boldsymbol{r} \leq \boldsymbol{y}, \boldsymbol{r} \geq \boldsymbol{0}\}$, where $\boldsymbol{X}$ is a matrix with rows of $\boldsymbol{x}_i^{\mathrm{T}}$'s and $\boldsymbol{y}$ is a vector of y_i's. (Our algorithm requires variables to be nonnegative, hence we have to use the usual trick with $\boldsymbol{\beta} \equiv \boldsymbol{\beta}^+ - \boldsymbol{\beta}^-$; $\boldsymbol{\beta}^+, \boldsymbol{\beta}^- \geq \boldsymbol{0}$ and as we start the algorithm with $\boldsymbol{\beta}^+ = \boldsymbol{\beta}^- = \boldsymbol{1}$, the initial $\boldsymbol{\beta}$ is zero.) The optimum of the linear program need not be unique; then there is a question which solution shall be taken as $\widehat{\boldsymbol{\beta}}_{L_1}$. The analytic centre of the optimal region is a reasonable choice (as it is the "farthest-from-all-borders" point); unlike the simplex method, with our algorithm we get it free of charge.

The script `L1reg(degree,`$\boldsymbol{x}$`,`$\boldsymbol{y}$`)` gets two N-component column vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ and fits the L_1-regression polynomial od degree `degree`. The pictures show how fast the regression function converges from $\boldsymbol{\beta}_{\text{start}} = \boldsymbol{0}$ to the optimum. Just for illustration, the usual $(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}$-estimate is shown with a dashed line. For instance, generate some random data with `x = [0:0.1:3]'`,

```
y =  1 + 1.25.*x - 1.5.*x.^2 + 0.3.*x.^3 +
     random('norm', 0, 0.2, size(x)).
```

Then, calling `L1reg(1,x,y)` and `L1reg(3,x,y)` yields the following pictures.

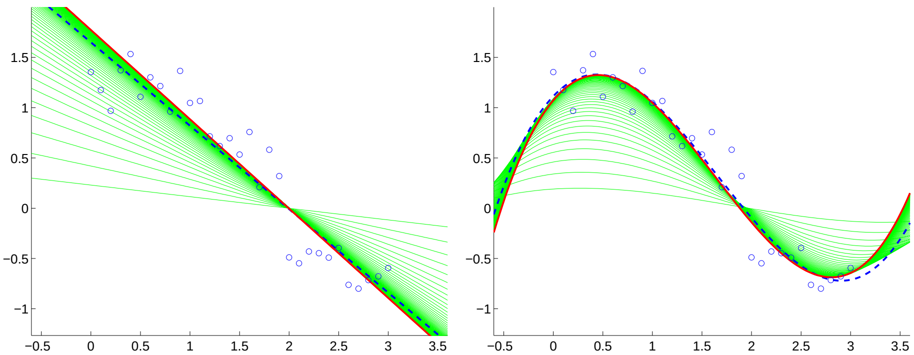


Figure 5: Convergence to L_1-regression polynomials of degree 1 and 3 (compared to L_2-regression (dashed))

The script also produces a picture of evolution of the regression coefficients over time (i.e. values of $\boldsymbol{\beta}$ in k-th iteration, where k is on the x-axis). For instance, if we fit a polynomial of degree 9 to the data described, we get the following picture. (We stop if $\mu < 10^{-9}$.)

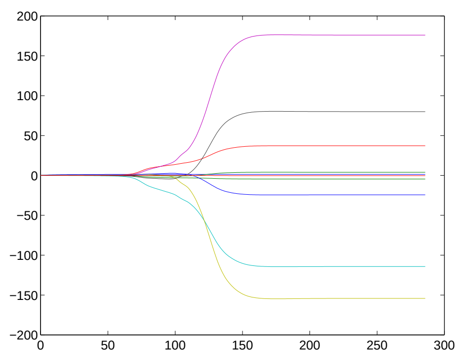


Figure 6: Evolution of the regression coefficients during computation

Now say we are fitting a line, hence $\boldsymbol{\beta}$ has two components. Fix the found (nearly) optimal residuals $\boldsymbol{r}$ and perform a slight δ-relaxation of the polyhedron: we get the system $\boldsymbol{X\beta}+\delta\mathbf{1} \geq \boldsymbol{y}-\boldsymbol{r}, \boldsymbol{X\beta}-\delta\mathbf{1} \leq \boldsymbol{y}+\boldsymbol{r}$ in three variables, $\boldsymbol{\beta}_1$ (intercept), $\boldsymbol{\beta}_2$ (slope) and δ. Now, given δ, we may plot a "δ-confidence polyhedron" which shows how $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ may change such that the sum of residuals does not increase more that δN. This is done by the script `confp(`$\boldsymbol{x}$`,`$\boldsymbol{y}$`,`δ`)`, where $\boldsymbol{x}$ and $\boldsymbol{y}$ are as before. If δ is a column vector, then for each component, one polyhedron is plotted. For our problem, calling `confp(x, y, [0:.01:.1]')` we get the following picture.

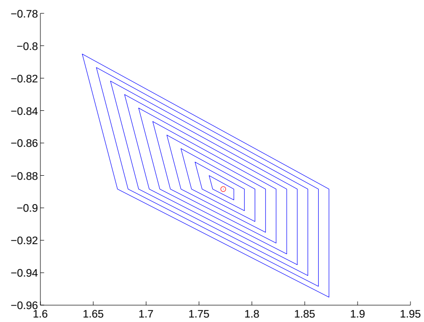


Figure 7: Example of "δ-confidence" polyhedra

There are more examples of linear programming problems that nicely show the convergence of the interior-point algorithm. Our last example concerns the problem of inscribing the largest circle into a polyhedron defined by a set of inequalities $\boldsymbol{Ax} \leq \boldsymbol{b}$. This issue is easily solved by a linear program: maximize r subject to the following constraints: (i) if a line $y = ax + b$ bounds the polyhedron from below, add a constraint $s_y - as_x - \sqrt{a^2+1} \cdot r \geq b$; (ii) if a line $y = ax + b$ bounds the polyhedron from above, add a constraint $s_y - as_x + \sqrt{a^2+1} \cdot r \leq b$; (iii) if a line $x = b$ bounds the polyhedron from left, add a constraint $s_x - r \geq b$, (iv) if a line $x = b$ bounds the polyhedron from right, add a constraint $s_x + r \leq b$, (v) $r \geq 0$. s_x and s_y are variables for the centre of the inscribed circle and r its diameter. (For our algorithm, we need five nonnegative variables: r, s_x^+, s_x^-, s_y^+, s_y^-: $s_x \equiv s_x^+ - s_x^-$, $s_y \equiv s_y^+ - s_y^-$.)

The script `circ(`$\boldsymbol{A}$`,`$\boldsymbol{b}$`)` plots how the algorithm approximates the largest circle inscribed into the nonempty polyhedron $\{\boldsymbol{x} : \boldsymbol{Ax} \leq \boldsymbol{b}\}$. For instance, generating a septahedron with the command

```
Ab =  getmatrix([4 + cos ([0:2*pi./7:2*pi])]',
      [3 + .5.*sin([0:2*pi./7:2*pi])]')
```

and calling `circ(Ab(:,1:2), Ab(:,3))` will yield the last picture of this article.

References

[1] Coppersmith D. — Winograd S. (1990), *Matrix Multiplication via Arithmetic Progressions.* Journal of Symbolic Computation 9, pp. 251–280.

[2] Černý, M., *Linear programming is in* **P** (in Czech, 2007). Internet: `http://nb.vse.cz/~cernym/lpinp.pdf`.

[3] den Hertog, D. (1994), *Interior Point Approach to Linear, Quadratic and Convex Programming.* Mathematics and Its Applications vol. 277, Kluwer Academic Publishers.

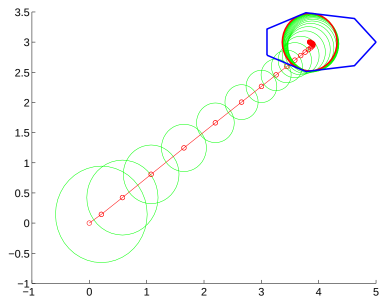


Figure 8: Convergence to the largest circle inscribed into a polyhedron

[4] Matoušek J. — Gärtner B. (2007), *Understanding and Using Linear Programming.* Springer Verlag, Berlin.

[5] Nazareth, J. L. (2003), *Differentiable Optimization and Equation Solving: A Treatise on Algorithmic Science and the Karmarkar Revolution.* Springer Verlag, Berlin.

[6] Regenar, J. (2001), *A Mathematical View of Interior-Point Methods in Convex Optimization.* MPS/SIAM Series on Optimization, Philadelphia.

[7] Roos C. — Terlaky T. — Vial J.-P. (2006), *Interior Point Methods for Linear Optimization.* Springer Verlag, Berlin.

[8] Schrijver A. (1998), *Theory of Linear and Integer Programming.* Wiley and sons, New York.

[9] Ye, Y. (1997), *Interior Point Algorithms: Theory and Analysis.* Wiley and sons, New York.

Miroslav Čulík, Jiří Valecký

Finance department

Faculty of Economics

VŠB – Technical University Ostrava

Sokolská 33, 701 21 Ostrava

e-mail: `miroslav.culik@vsb.cz, jiri.valecky@vsb.cz`

Electricity Price Forecasting and Simulation

Abstract

The business structure of the electricity generating sector is relatively easy to predict, if the environment is regulated. After the successful deregulation of some sectors, there has been movement toward deregulation and liberalization in many European countries. This process has important impact on the most market participants, from the electricity generators, distributing utilities to the final consumers.

As a consequence of power sector liberalization, price modeling and forecasting has emerged as an important input to a wide range of decision problems in the electricity sector. Electricity prices have become extremely volatile compared to prices in other commodity markets and choices of appropriate model for modeling and forecasting have a large effect on the solution to many decision problems. To find the most suitable model including deriving its parameters, the choice of appropriate set of input data that includes relevant information about behavior of electricity prices is important, as well.

The aim of this study is to analyze the application possibilities of selected models for day-ahead electricity prices forecasting at the European electricity markets. For the electricity prices forecasting, diffusion and econometric models will be applied. The study is focused on the daily average electricity prices in Europe over the period 2006 – 2007, the forecasts will be made for the January of 2008. Forecasting results of the used models will be compared and discussed. In the end, possibilities and applicability for modeling and forecasting of applied models is analyzed.

Keywords: *Electricity price, forecasting, confidence interval, Geometric Brownian Motion, mean-reversion process.*

Introduction

Over the last decades, most countries, not only in Europe, but also in overseas, have decided for electricity market liberalization. The prime motivation for liberalization is, in the long run, to increase the competition, stimulate technical innovation, promote efficient investments, etc.

Originally, the power sector was characterized by a highly vertically integrated market structures with existence of no competition. Recently, the monopolistic structure was replaced by deregulated competitive markets, where consumers can choose their energy supplier. Moreover, liberalization of the power system created need for organized markets at the wholesale level. It was necessary to organize trading system, exchanges and pools for energy power.

Deregulation process is usually associated with three types of risk: organizational risk, market risk and regulatory risk.

Organizational risk is associated with the transformation of the companies from traditional monopoly to commercial or market-oriented subject. The main sources of the organizational risk are government authorities, which remove barriers for entering competitions. Competition presence create forces on the efficiency, cost reduction, etc.

Market risk represents another important problem for generator at the deregulated market. Under monopolistic conditions, price formation is relatively easy to understand. Price (tariffs) are negotiated by regulator, information are available and used by centralized planned utilities. Competition introduces the possibility for the final consumer freely decide for the power supplier, increase in the price volatility, asymmetric information, financial uncertainty presence, etc.

Regulatory risk is a risk, when governmental authorities – after successful deregulation – continue to maintain certain degree of regulation. The main purpose is that the regulatory authorities ensure fair implementation of the deregulation and engages anti-competitive practices. The regulator has to decide how to balance controls on prices, investments, anti-competitive behavior and, of course, to protect captured customers. From the point of view of electricity generating utilities, due to the fact, that regulators behavior is less predictable, it is extremely difficult to predict how the regulators will react to any future actions.

As a consequence of power sector liberalization, price modeling and forecasting has emerged as an important input to a wide range of decision problems in the electricity sector. Electricity prices have become extremely volatile compared to prices in other commodity markets and choices of appropriate model for modeling and forecasting have a large effect on the solution to many decision problems. To find the most suitable model including deriving its parameters, the choice of appropriate set of input data that includes relevant information about behavior of electricity prices is important, as well.

The aim of this study is to analyze the application possibilities of selected models for day-ahead electricity prices forecasting at the European electricity market. Daily average electricity prices will be analyzed. Forecasting results of the used models will be compared and discussed. In the end, possibilities and applicability for modeling and forecasting of applied models is analyzed.

1 Models description

In this study, both diffusion and econometrics models will be applied. In the case of diffusion models, discrete and continuous returns will be used. Closer description is the contents of the following chapters 1.1 - 1.3., more on models can be found in Box (1970), Dixit and Pindick (1994), Eydeland and Wolynick (2003), Bunn (2004), Mazer (2007), etc.

1.1 Geometric Brownian Motion

Geometric Brownian motion (GBM) with drift is an important special case of Ito process. Its applicability relies on the assumption, that some economic variables (for instance speculative asset prices) have tendency to wander far from their starting point.

If the generalized Ito process is described by following equation,

$$dx = a\,(x,t)\,dt + b\,(x,t)\ dz, \tag{1}$$

where $a\,(x,t)$ and $b^2\,(x,t)$ are known (nonrandom) function (expected instantaneous drift rate of the Ito process and instantaneous variance rate), dz is increment of a Wiener process and holds that $dz = \tilde{z} \cdot \sqrt{dt}$ (here $\tilde{z}$ denotes random variable from standard normal distribution $N(0,1)$), than in the case of GBM holds, that $a\,(x,t) = \mu x$ and $b\,(x,t) = \sigma x$, where $\mu$and σ are constants and equation (1) can than be rewrite to the following form,

$$dx = \mu \cdot dt + \sigma \cdot x\ \cdot dz \tag{2}$$

If $x\,(t)$ is given by (1), then $F\,(x) = \log x$is GBM with drift,

$$dF = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma\ \cdot dz, \tag{3}$$

where $\left(\mu - \frac{\sigma^2}{2}\right) dt$ is drift of the random process and $\sigma^2 t$is variance. If we denote S_t as the electricity price at time t, equation (3) can be expressed in this way,

$$S_{t+dt} = S_t \cdot \exp\left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma \cdot dz, \tag{4}$$

Here, μ is denotes logarithm change in price (continuous return), i.e.

$$\mu = \log \frac{S_{t+dt}}{S_t}, \tag{5}$$

Otherwise, return can be expressed as discrete,

$$\mu = \frac{S_{t+dt} - S_t}{S_t}, \tag{6}$$

in this case price in the upcoming period, t+dt, is given by,

$$S_{t+dt} = S_t \cdot (1 + \mu \cdot dt + \sigma \cdot dz)\,. \tag{7}$$

For expected value hold $E\,(S_t) = S_0 \cdot \exp^{\mu \cdot t}$, for variance$\sigma^2\,(S_t) = S_0^2 \cdot \exp^{2 \cdot \mu \cdot t} \cdot \left(\exp^{\sigma^2 t} - 1\right)$.

Quantile for the lognormal probability distribution of the random variable is given as follows, see Zmeškal (2004),

$$S_t^{\alpha} = S_0 \cdot \exp\left(\mu \cdot t + \Phi^{-1}\,(\alpha) \cdot \sigma \cdot \sqrt{dt}\right), \tag{8}$$

where Φ^{-1} is inverse function to standard density function and $\alpha$is confidence level.

1.2 Mean-Reverting processes

For commodities, interest rates, exchange rates etc., mean-reversion model has more economic logic than above described geometric Brownian model. In this case, while in the short-run the prices can fluctuate randomly up and down, in the long-run they have the tendency to revert to the long-run equilibrium price.

In the financial-economics literature appears several different ways to model the mean-reversion process.

The most basic mean-reversion model is the (arithmetic) Ornstein-Uhlenbeck model, which is defined as follows (without drift), see Dixit, Pindick (2004),

$$dS = \eta \cdot S \cdot \left(\bar{S} - S\right) \cdot dt + \sigma \cdot S \cdot dz, \tag{9}$$

where η is the speed of reversion, and $\bar{S}$ is the long-run equilibrium level, to which S tends to revert. In this case, the expected change in S depends on the difference between S and $\bar{S}$. If S is above (below) $\bar{S}$, it is more likely to fall (rise) over the next time interval.

If drift it is supposed, equation (9) has this form,

$$dS = \left\{\mu + \left[\eta \cdot \left(\bar{S} \cdot \exp\left(\mu \cdot t\right) - S\right)\right]\right\} \cdot S \cdot dt + \sigma \cdot S \cdot dz. \tag{10}$$

If the current value of S at t_0 is S_0, and S follows equation (9) then the expected value at future time t is,

$$E\left(S_t\right) = \bar{S} + \left(S_0 - \bar{S}\right) \cdot e^{-\eta \cdot t}, \tag{11}$$

and the variance of $\left(S_t - \bar{S}\right)$ is defined in this way,

$$var\left(S_t - \bar{S}\right) = \frac{\sigma^2}{2\eta} \cdot \left(1 - e^{-2\eta \cdot t}\right). \tag{12}$$

It is obvious from these equations, that the expected value of S_t converges to $\bar{S}$ as t becomes large and the variance converges to $\frac{\sigma^2}{2\eta}$.

Quantile for the probability distribution of the random variable is given as follows, see Dixit, Pindick (2004),

$$S_t^{\alpha} = \bar{S} + \left(S_0 - \bar{S}\right) \cdot e^{-\eta \cdot t} + \Phi^{-1}\left(\alpha\right) \cdot \frac{\sigma^2}{2\eta} \cdot \left(1 - e^{-2\eta \cdot t}\right). \tag{13}$$

Figure 1 below presents a combination of Geometric Brownian Motion (exponential drift) with mean-reverting model. This was adapted from Metcalf and Hasset (1995).

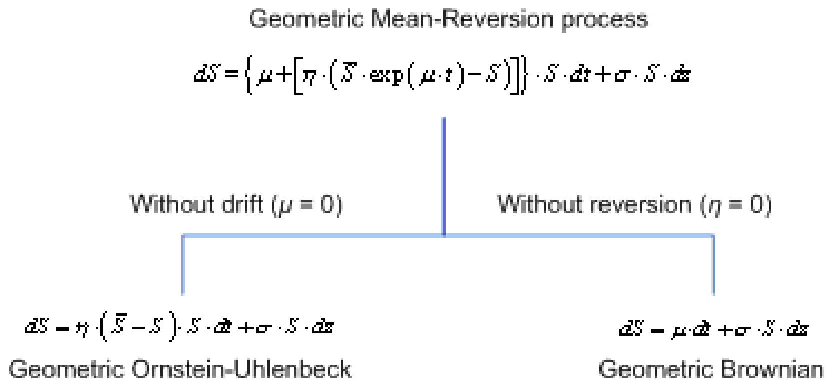


Figure 1: Combination of Geometric Brownian Motion

1.3 Econometric models

ARIMA (Autoregressive Integrated Moving Average) models are suitable for modelling non-stationary time series. ARIMA models are based on the three parts: (1) an autoregressive part, (2), integrated process (3) and a contribution from a moving average.

The autoregressive part (AR) of the model has its origin in the theory that individual values of time series can be described by linear models based on the preceding observations whereas the moving average models (MA models) means that time series values can be expressed as dependent on the preceding estimation errors.

In general, an ARIMA(p,d,q) can be described with backshift operator, B, as follows,

$$k_p\left(B\right)\left(1-B\right)^d X_t=m_q\left(B\right)u_t, \tag{14}$$

where k_p are the parameters of the autoregressive part, m_q are the parameters of the moving average part and u_t is an error term distributed IID. For the backshift operator, it holds,

$$\left(1-B^i\right)X_t=X_t-X-X_{t-i} \tag{15}$$

Equation (14) is a general ARIMA model of an economic time series. In fact and in many cases, the seasonality presence (periodic fluctuation) is detected; therefore seasonal ARIMA models are used to capture this typical feature. The general SARIMA (p,d,q)x(P,D,Q) model can be expressed as follows,

$$k_p\left(B\right)K_p\left(B^S\right)\left(1-B\right)^d\left(1-B^S\right)^D X_t=m_q\left(B\right)M_Q\left(B^S\right)u_t \tag{16}$$

where K_P and M_Q are parameters of seasonal autoregressive and seasonal moving average part, D is seasonal difference and S is seasonal period.

2 Forecast evaluation

There are many models focusing on forecast evaluation, see (Kaminski, 2004; Zmeškal, 2004; Weron 2006; etc). The most common approach is through the comparison of mean squared errors, or similar statistic based on other economic loss measures, for alternative forecasts at several horizons. Only those measures which are applied will be described.

Mean squared error (MSE) of an **estimator** is one of many ways to quantify the amount by which an **estimator** differs from the true value of the quantity being estimated. As a **loss function**, MSE is called squared error loss. MSE measures the **average** of the square of the error and is defined as,

$$MSE\left(\widehat{\theta}\right) = E\left[\left(\widehat{\theta} - \theta\right)^2\right] \tag{17}$$

where $\widehat{\theta}$ is estimated parameter and θ is estimator.

Root mean square deviation (RMSD) (also root mean square error (RMSE)) is a frequently-used measure of the differences between values predicted by a model or an estimator and the values actually observed from the thing being modeled or estimated. These individual differences are also called **residuals**, and the RMSD serves to aggregate them into a single measure of predictive power. The RMSD of an **estimator** with respect to the estimated parameter $\widehat{\theta}$ is defined as the square root of the **mean squared error**,

$$RDSM\left(\widehat{\theta}\right) = \sqrt{MSE\left(\widehat{\theta}\right)} \tag{18}$$

Clemets and Hendry (1993) proposed comparison of forecast performances through the generalized forecast error second moment (GFESM), which is defined as,

$$GFESM\left(h\right) = \left[E\left(E_{T,h}E_{T,h}^{/}\right)\right], \tag{19}$$

where e_{T+h} is the error in an h- steps forecast made at time T and $E_{T,h}^{/} = \left(e_{T+1}, \ldots e_{T+h}\right)$.

3 Application

The aim of this part is to apply above described models for daily electricity prices forecast.

There are two groups of models applied: (a) diffusion models (GBM, mean-reversion) with continuous and discrete returns, and (b) econometric model. For estimation of the models parameters, LSE procedure is used. In the end, forecasting results are compared to the true values, furthermore, models applicability for modeling and forecasting purposes is analyzed.

3.1 Data

The object of analysis is time series of daily electricity average spot prices. Prices are expressed in EUR/MWh, the length of the data series is two years (i.e. daily averages between 1st January 2006 and 31th December 2007)[1], see Figure 2.

[1] Data source: `www.nordpool.com`

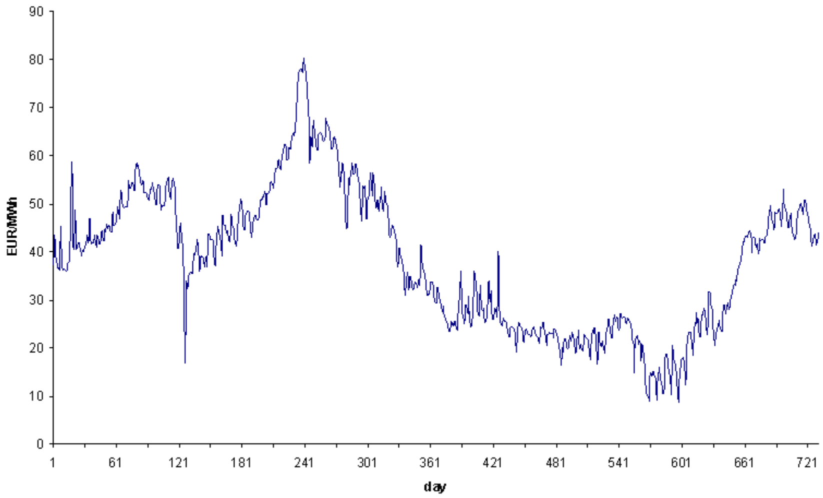


Figure 2: Daily average electricity prices in Europe (EUR/MWh, 2006-2007)

3.2 Diffusion models – derivation and application for simulation and forecast

Calculation procedure was conducted in the following steps:

(a) Estimation of the model parameters by employing LSE procedure. Derived parameters for GBM and mean-reversion model are stated, see Table 1 and 2.

	Geometric Brownian Motion			
parameter		**drift rate** μ(%)	**st. deviation** σ (%)	**dt**
returns	**continuous**	**-2,419**	**36,28**	**0,00274**
	discrete	**0,194**	**40,47**	**0,00274**

Table 1: Estimated parameters of Geometric Brownian Motion

Based on the results in Table 1, GBM model for daily electricity prices forecasting can be written as $S_{t+dt} = S_t \cdot \exp\left(-0,02419 - \frac{0,3628^2}{2}\right) \cdot 0,00274 + 0,3628 \cdot \sqrt{0,00274} \cdot \tilde{z}$ (continuous returns), or $S_{t+dt} = S_t \cdot \left(1 + 0,00197 \cdot 0,00274 + 0,4047 \cdot \sqrt{0,00274} \cdot \tilde{z}\right)$ if we work with discrete returns. Here, $\tilde{z}$ denotes random variable from N(0, 1).

For mean-reversion process, based on the results in Tab.2 general formula (9) can be written as $S_{t+dt} = S_t + 0,208 \cdot (38,65 - S_t) \cdot 0,00274 + 0,3617 \cdot \sqrt{0,00247} \cdot \tilde{z}$ (continuous returns), or $S_{t+dt} = S_t + 0,344 \cdot (44,06 - S_t) \cdot 0,00274 + 0,4011 \cdot \sqrt{0,00247} \cdot \tilde{z}$ if we work with discrete

	Mean – Reversion Model				
parameter		**long-run equilibrium $\bar{S}$**	**speed of reversion η**	**st. deviation σ (%)**	**dt**
returns	**continuous**	**38,65**	**0,208**	**36,17**	**0,00274**
	discrete	**44,06**	**0,344**	**40,11**	**0,00274**

Table 2: Estimated parameters of Mean-Reversion Model

returns. All the parameters are statistical significant at 95 % confidence level (p-value and F-statistic is used).

(b) Calculation of expected electricity prices on the basis of estimated models and estimation of the confidence intervals at 95 % and 99 % confidence levels, see Figures 3.-6.

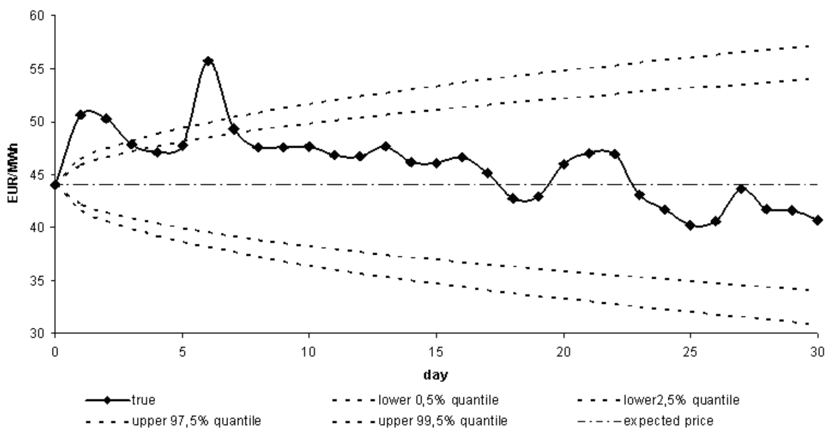


Figure 3: True and expected values incl. confidence intervals (GBM, discrete ret., Jan 2008)

(c) Generation of the random variable from N (0,1) and simulation of daily electricity price on the basis of estimated models for each day of January 2008.

(d) Constructing density functions on the basis of simulations, see Figures 7 – 10.

(d) Calculation of 95% and 99% confidence intervals and comparing to the true value, see Table 3 and Figure 7.

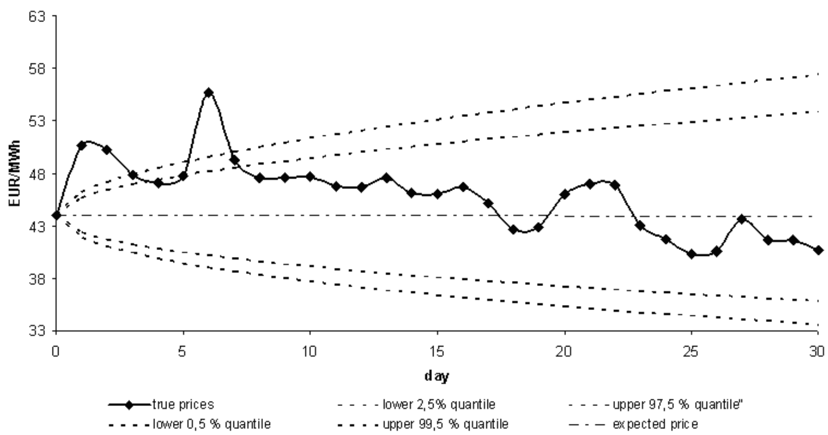


Figure 4: True and expected values incl. confidence intervals (GBM, cont. ret., January 2008)

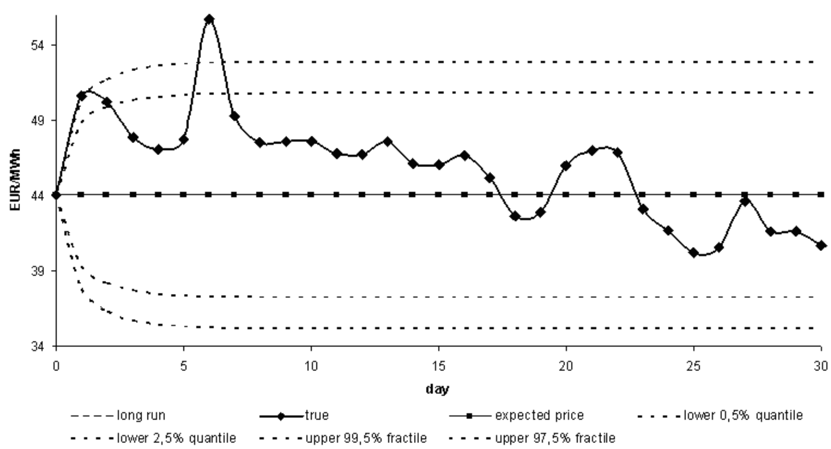


Figure 5: True and expected values incl. confidence intervals (M-R , discrete ret., January 2008)

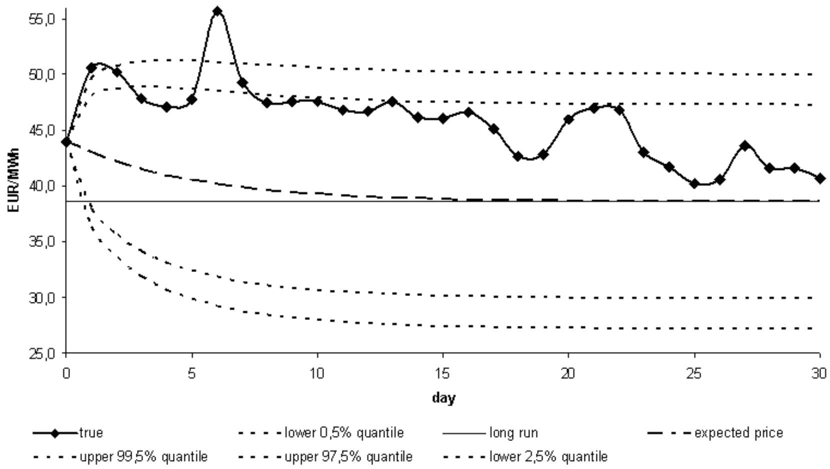


Figure 6: True and expected values incl. confidence intervals (M-R, cont. ret., January 2008)

model		**Expected value**	**95 % confidence level**		**99% confidence level**	
			upper	**lower**	**upper**	**lower**
GBM	**discrete**	44,04	54,05	34,03	57,2	30,88
	continuous	43,94	53,88	35,84	57,44	35,84
Mean Rever-sion	**discrete**	44,06	50,81	37,3	52,93	35,17
	continuous	38,66	47,35	29,97	50,02	27,24

Table 3: Electricity price forecasts for 30th January 2008 (true value 40,06 EUR/MWh)

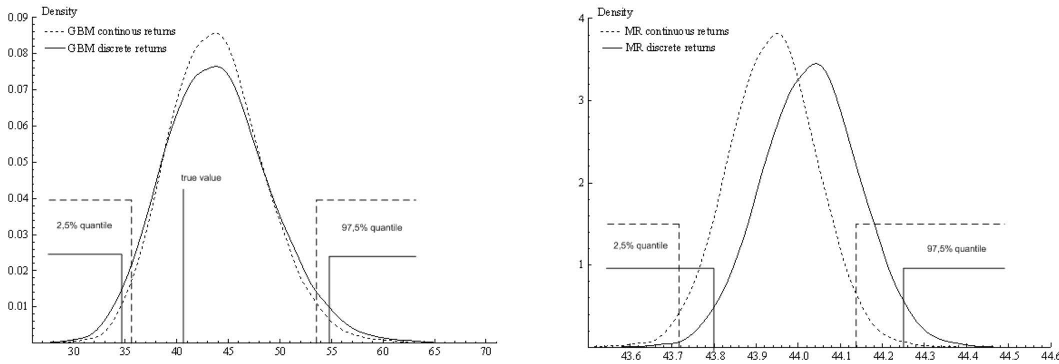


Figure 7: Electricity price density function (left- GBM, right – Mean reversion) on 30th January 2008

3.3 Econometric model – results

First, the seasonal unit root test had to be performed. For this purpose, the HEGY test was employed and the first order of differencing with week (7-days) season period was traced.

Thereafter, the econometric model was estimated in GiveWin by employing maximum likelihood method as SARIMA(0,0,8)x(0,||7||,0) in the following notation,

$$\left(1 - B^7\right) S_t = m_8 \left(B\right) u_t. \tag{20}$$

Following Table 4 includes PcGive module outputs for electricity prices differences and residuals; Figure 8 depicts true price, expected price and 95% confidence interval.

According to the estimation results, the residuals are non-normally distributed (Normality test) and non-correlated (Portmanteau test) and the heteroscedasticite is detected (ARCH effect test). The resulting coefficients of (20) are: m_1 = 0,811549, m_2 = 0,721555, m_3 = 0,729682, m_4 = 0,724344, m_5 = 0,679084, m_6 = 0,670672, m_7 = -0,172740, m_8 = -0,0762864.

4 Conclusion

The aim of this study was to develop proposed models and verify their applicability for forecasting purposes. Attention was focused on the European electricity market and its daily average electricity prices over the horizon analysed; these historical data were used for model estimation and subsequently used for forecast and simulation in January 2008.

Proposed models can be divided into two groups. First, diffusion models (GBM mean-reversion model) and their parameters were on the basis of LSE procedure estimated, second, econometric model on the basis of maximum likelihood method was derived.

In the first part, the intention was to propose model and forecast expected electricity price during January 2008. Graphical presentation of the price expectations including confidence intervals is in the Figures 3 – 6. (for diffusion models) and in Figure 8 (for econometric model). Consequently, the selection of the best model for the price prediction was made. As a selection

```
---- Maximum likelihood estimation of ARFIMA(0,0,8) model ----
The estimation sample is:  1 - 724
The dependent variable is: D_Price_Season7  (Sešit1.xls)

                 Coefficient  Std.Error  t-value  t-prob
MA-1                0.811549    0.03714     21.9   0.000
MA-2                0.721555    0.04767     15.1   0.000
MA-3                0.729682    0.04906     14.9   0.000
MA-4                0.724344    0.05073     14.3   0.000
MA-5                0.679084    0.05049     13.4   0.000
MA-6                0.670672    0.04846     13.8   0.000
MA-7               -0.172740    0.04759    -3.63   0.000
MA-8              -0.0762864    0.03751    -2.03   0.042

log-likelihood    -1630.87615
no. of observations       724  no. of parameters           9
AIC.T               3279.7523  AIC                4.53004461
mean(D_Price_Season7)    0.0461389  var(D_Price_Season7)      22.3091
sigma                 2.01282  sigma^2               4.05143

BFGS using numerical derivatives (eps1=0.0001; eps2=0.005): Strong convergence
Used starting values:
0.75621; -0.56297; -0.34780;0.083108;0.057210; -0.037413; -0.040083; -0.00056801

Descriptive statistics for residuals:
Normality test:   Chi^2(2) =   486.35 [0.0000]**
ARCH 1-1 test:    F(1,714) =   53.431 [0.0000]**
Portmanteau(26):  Chi^2(18)=   19.331 [0.3717]
```

Table 4: Results of PC Give module

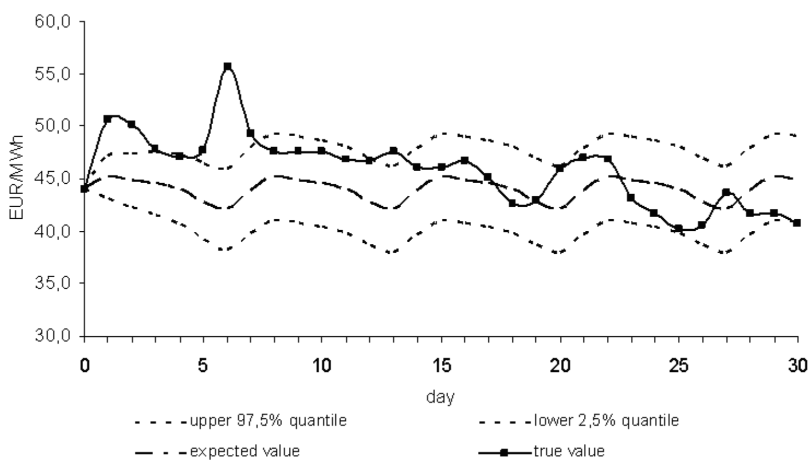


Figure 8: True price, forecast and confidence interval

Model	Return	Forecast evaluation model		
		MSE	**RDSM**	**GFESM**
GBM	**discrete**	14,89	3,86	446,8
	continuous	14,92	3,86	447,53
M-R	**discrete**	14,81	3,85	444,25
	continuous	49,2	7,01	1475,96
Econometric		16,62	4,08	498,59

Table 5: Forecasting evaluation

criterions, measures described in Chapter 3 were employed, the results are summarized in the following Table 4.

It is obvious from the results that the best models for electricity forecast is mean reversion model with discrete returns regardless what forecast evaluation model is applied. Even if in the short-time period the forecasted variable can behave randomly, in the long-time period have the electricity prices tendency to revert to the equilibrium level. It is apparent here, where the forecast errors of GBM and mean-reversion models is close.

In the end, proposed diffusion models were used for simulation of the price distribution on the last day (January 30th) the horizon analyzed. Results of the price density function including confidence interval are graphically presented in Table 3 and Figure 11. It is apparent from the results that in the case of mean-reversion model, the true value lies outside of the confidence interval. It confirms the fact stated above, that applicability and reliability of models can differ in short and long time period.

Acknowledgement

This paper is solved within the project GACR 402/07/P121. The support is appreciated.

Reference

[1] Bunn, D.W. *Modelling prices in competitive electricity markets.* Wiley: England; 2004. ISBN 0-470-84860-X.

[2] Clements, M. P., Hendry, D. F. *On the limitations of comparing mean square forecast errors.* Journal of Forecasting; 12; pp. 617-676.

[3] Clewlow, L., Strickland, CH. *Energy derivatives: pricing and risk* management. Lacima Publications, 2000. ISBN 0-953-8896-0-2.

[4] Čulík, M., Valecký, J. *Application of diffusion and econometrics models for daily electricity price modelling at the European electricity market.* IASK Global management, Portugal; 2007; p. 388. ISBN: 978-972-99397-4-7.

[5] Dixit, A. K., Pindick, R. S. *Investment under Uncertainty.* University Press; 1994. ISBN 0-691-03410-9.

[6] Dowd, K.: *Measuring Market Risk*. 2nd edition. Wiley: England; 2006. ISBN

[7] Eydeland, A., Wolyniec, K. *Energy and Power Risk Management. New development in Modeling, Pricing and Hedging.* Wiley: New Jersey; 2003. ISBN 0-471-10400-0.

[8] Greene, W. H. *Econometric Analysis*. 6th edition, Prentice Hall: New York; 2008. ISBN 978-0-13-513245-6.

[9] Kaminsky, V. *Managing energy price risk: the new challenges and solutions.* 3rd Edition, Risk Books: London; 2004. ISBN 1-904-339-19-0.

[10] Koop, G. *Introduction to Econometrics*. Wiley: New York; 2007. ISBN 978-0-470-03270-1.

[11] Mazer, A. *Electric power planning for regulated and deregulated markets*. Wiley: New York; 2007. ISBN 978-0-470-11882-5.

[12] Ramanathan, R. *Introductory econometric with applications.* 3rd Edition, Dryden Press: U.S.A.; 1995. ISBN 0-03-015228-3.

[13] Weron, R. *Modelling and Forecasting Electricity Loads and Prices: A Statistical Approach (The Wiley Finance Series).* Wiley: New Jersey; 2006. ISBN 0-470-05753-X.

[14] Zmeškal, Z. *et al*. *Financial models*. EkF VŠB-TU Ostrava; 2004. ISBN 80-248-0754-8.

Jan Fábry

Department of Econometrics
University of Economics Prague
W. Churchill sq. 4, 130 67 Prague 3

e-mail: fabry@vse.cz

Capacited Messenger Problem

Abstract

In a messenger problem the origins of the packages and their destinations are specified for a set of customers. Some services assume a size of each package is inconsiderable in reference to the capacity of the vehicle; however in most of real messenger problems the vehicle's capacity has to be respected. In the paper, one vehicle with limited capacity is considered for service of advanced customers. The optimization model is formulated for small-sized capacited messenger problems. If time windows are given for all pick-ups and deliveries, more vehicles have to be involved to assure all requirements of customers. In real problems the earliest possible time for pick-up and the latest acceptable time for delivery of each package are obviously announced. XPRESS-MP was used for computational experiments.

Keywords: *Capacited messenger problem, dial-a-ride problem, pick-up and delivery problem, traveling salesman problem, XPRESS-MP*

Introduction

In a messenger problem each customer specifies an origin where the driver has to pick up the package and a destination for package's delivery. Distances between all points in the distribution network are given. In the paper, one vehicle with unlimited capacity is considered for applications without time windows. If the size of each package is relevant, the capacity of the vehicle has to be involved in the model. When time windows have to be respected, more vehicles have to be used for service. A messenger problem is analogous to dial-a-ride problem or pickup-and-delivery problem described in literature (Cordeau, 2006), (Lu and Dessouky, 2004). The optimization models for small-sized messenger problems will be formulated in the paper.

1 Uncapacited messenger problem

First, one uncapacited vehicle and no time windows are considered in the problem. The objective is to minimize total length of the route for pick-up and delivery of all packages.

Example 1. Figure 1 shows an illustrative example of a messenger problem with 4 customers. Each arc in the network corresponds to the requirement of a customer for delivery. Customers are purposely situated in even nodes, while destinations in odd nodes. If i is the number of the customer (and the origin of its package), $i + 1$ is a number of the package destination. A depot of the vehicle is located in the node 1.

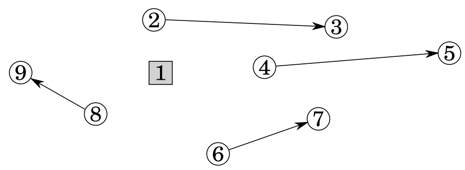


Figure 1: Example of messenger problem

Let us have n customers; each customer requires a delivery of one package from its office to a specific destination. Considering one depot in the problem, there are $(2n + 1)$ locations in the distribution network. Let c_{ij} denote the shortest distance between locations i and j. The mathematical model of a messenger problem is defined as follows (Fábry, 2007):

$$\text{Minimize } z = \sum_{i=1}^{2n+1} \sum_{j=1}^{2n+1} c_{ij} x_{ij}, \tag{1}$$

subject to

$$\sum_{j=1}^{2n+1} x_{ij} = 1, \quad i = 1, 2, ..., 2n + 1, \tag{2}$$

$$\sum_{i=1}^{2n+1} x_{ij} = 1, \quad j = 1, 2, ..., 2n + 1, \tag{3}$$

$$u_i - u_j + (2n + 1)x_{ij} \leq 2n, \quad i = 1, 2, ..., 2n + 1, \quad j = 2, 3, ..., 2n + 1, \quad i \neq j, \tag{4}$$

$$u_{2i} \leq u_{2i+1}, \quad i = 1, 2, ..., n, \tag{5}$$

$$u_1 = 0, \tag{6}$$

$$x_{ij} \in \{0, 1\}, \quad i, j = 1, 2, ..., 2n + 1. \tag{7}$$

Binary variable x_{ij} equals 1, if the vehicle goes from location i to location j, 0 otherwise. The objective function (1) corresponds to the total length of the vehicle's route. Sets of equations (2) and (3) assure that each location is visited exactly once. Constraints (4) including variables u_i are Miller-Tucker-Zemlin's inequalities that avoid partial cycles in the solution. Because each package has to be picked up before its delivery inequalities (5) must be respected.

Figure 2 illustrates the feasible route in the example given above. The package is delivered to location 9 immediately after its pick-up in location 8, while the package being picked up in location 6 is delivered to location 7 after several visits to other locations.

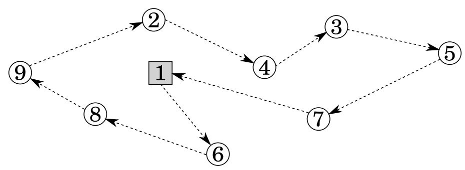


Figure 2: Feasible route

2 Capacited messenger problem

In real applications, the size of each package is considerable in reference to the capacity of the vehicle. Let us denote q_i the requirement of customer i and V the vehicle's capacity. Optimization model is defined as follows:

$$\text{Minimize } z = \sum_{i=1}^{2n+1} \sum_{j=1}^{2n+1} c_{ij} x_{ij}, \tag{8}$$

subject to

$$\sum_{j=1}^{2n+1} x_{ij} = 1, \quad i = 1, 2, ..., 2n+1, \tag{9}$$

$$\sum_{i=1}^{2n+1} x_{ij} = 1, \quad j = 1, 2, ..., 2n+1, \tag{10}$$

$$v_i + q_{2j} - V(1 - x_{i,2j}) \leq v_{2j}, \quad i = 1, 2, ..., 2n+1, \quad j = 1, 2, ..., n, \quad i \neq j, \tag{11}$$

$$v_i - q_{2j+1} - V(1 - x_{i,2j+1}) \leq v_{2j+1}, \quad i = 1, 2, ..., 2n+1, \quad j = 1, 2, ..., n, \quad i \neq j, \tag{12}$$

$$0 \leq v_i \leq V, \quad i = 2, 3, ..., 2n+1, \tag{13}$$

$$u_i - u_j + (2n+1) x_{ij} \leq 2n, \quad i = 1, 2, ..., 2n+1, \quad j = 2, 3, ..., 2n+1, \quad i \neq j, \tag{14}$$

$$u_{2i} \leq u_{2i+1}, \quad i = 1, 2, ..., n, \tag{15}$$

$$u_1 = 0, v_1 = 0, \tag{16}$$

$$x_{ii} = 0, \quad i = 1, 2, ..., 2n+1, \tag{17}$$

$$x_{ij} \in \{0,1\}, \quad i,j = 1,2,...,2n+1. \tag{18}$$

The objective function (8) and constraints (9), (10), (14) and (15) have the same meaning as statements (1) – (5) in uncapacited problem. Variable v_i is introduced for balancing load on the vehicle. Constraints (11) balance loading packages, while (12) unloading of them. Inequalities (13) avoid exceeding the vehicle's capacity.

Example 2. There are 5 customers with requirements 4, 6, 7, 3 and 3 units. In the following matrix distances between all locations are given. Indices of locations are organized in the way described in the Example 1. The vehicle's capacity is 12 units.

	1	2	3	4	5	6	7	8	9	10	11
1	0	25	24	59	45	33	37	55	51	16	55
2	25	0	11	71	70	46	59	61	55	40	80
3	24	11	0	78	67	38	53	69	63	40	75
4	59	71	78	0	65	89	81	15	21	49	82
5	45	70	67	65	0	46	27	71	71	32	16
6	33	46	38	89	46	0	21	86	83	40	46
7	37	59	53	81	27	21	0	83	81	35	25
8	55	61	69	15	71	86	83	0	7	48	87
9	51	55	63	21	71	83	81	7	0	46	87
10	16	40	40	49	32	40	35	48	46	0	46
11	55	80	75	82	16	46	25	87	87	46	0

Table 1: Matrix of minimal distances

XPRESS-MP was used for the optimization. The optimal route is 1 – 2 – 3 – 4 – 8 – 9 – 10 – 5 – 11 – 6 – 7 – 1 with the length of 334.

3 Capacited messenger problem with time windows

If time windows are given for all requirements, the optimization model has to be modified. Let us suppose the earliest possible time for pick-up and the latest acceptable time for delivery are specified for each package. These values are denoted a_i (i = 2,3,. . . ,2n+1). Traveling times d_{ij} between all pairs of locations i and j have to be known. In this type of problem, more vehicles must be mostly used to satisfy all requirements. Hence, binary variable x_{ij}^k is introduced: it equals 1, if the vehicle on route k goes directly from location i to location j, 0 otherwise. Value V_k is a capacity of the vehicle on route k. We suppose K available vehicles (possible routes). Then, mathematical model is

$$\text{Minimize } z = \sum_{k=1}^{K} \sum_{i=1}^{2n+1} \sum_{i=1}^{2n+1} c_{ij} x_{ij}^k, \tag{19}$$

subject to

$$\sum_{k=1}^{K}\sum_{j=1}^{2n+1} x_{ij}^k = 1, \quad i = 2,3,...,2n+1, \tag{20}$$

$$\sum_{i=2}^{2n+1} x_{1i}^k \le 1, \quad k = 1,2,...,K, \tag{21}$$

$$\sum_{i=1}^{2n+1} x_{ij}^k = \sum_{l=1}^{2n+1} x_{jl}^k,\ j = 1,2,...,2n+1,\ k = 1,2,...,K, \tag{22}$$

$$v_i^k + q_{2j} - V_k(1 - x_{i,2j}^k) \le v_{2j}^k,\ i = 1,2,...,2n+1,\ j = 2,3,...,n,\ i \ne j,\ k = 1,2,...,K, \tag{23}$$

$$v_i^k - q_{2j+1} - V_k(1 - x_{i,2j+1}^k) \le v_{2j+1}^k,\ i = 1,2,...,2n+1,\ j = 2,3,...,n,\ i \ne j,\ k = 1,2,...,K, \tag{24}$$

$$0 \le v_i^k \le V_k, \quad i = 2,3,...,2n+1, \quad k = 1,2,...,K, \tag{25}$$

$$t_i^k + d_{ij} - M(1 - x_{ij}^k) \le t_j^k,\ i = 1,2,...,2n+1,\ j = 2,3,...,2n+1,\ i \ne j,\ k = 1,2,...,K, \tag{26}$$

$$t_{2i}^k \le t_{2i+1}^k, \quad i = 1,2,...,n, \quad k = 1,2,...,K, \tag{27}$$

$$t_{2i}^k \ge a_{2i}\sum_{j=1}^{2n+1} x_{2i,j}^k, \quad i = 1,2,...,n, \quad k = 1,2,...,K, \tag{28}$$

$$t_{2i+1}^k \le a_{2i+1}\sum_{j=1}^{2n+1} x_{2i+1,j}^k, \quad i = 1,2,...,n, \quad k = 1,2,...,K, \tag{29}$$

$$\sum_{j=1}^{2n+1} x_{2i,j}^k = \sum_{j=1}^{2n+1} x_{2i+1,j}^k, \quad i = 1,2,...,n,\ \ k = 1,2,...,K, \tag{30}$$

$$v_1^k = 0, t_1^k = 0, \quad k = 1,2,...,K, \tag{31}$$

$$x_{ii}^k = 0, \quad i = 1,2,...,2n+1, \quad k = 1,2,...,K, \tag{32}$$

$$x_{ij}^k \in \{0,1\}, \quad i,j = 1,2,...,2n+1, \quad k = 1,2,...,K. \tag{33}$$

The objective (19) corresponds to the total length of all vehicles' routes. Sets of equations (20) assure that each location (excluding depot) is visited exactly once. Inequalities (21) enable at most one vehicle on each route. Respecting constraints (22), the vehicle entering certain location has to leave it. Inequalities (23) and (24) set the balance of loaded and unloaded packages on each route. Inequalities (25) respect vehicle's capacity on each route. The value of variable t_i^k gives the time of visit to location i by the vehicle on route k. Inequalities (27) have to be respected because the time of pick-up of each package has to precede to the time of its delivery. Constraints (28) and (29) define time windows which have to be satisfied by the visiting vehicle.

Example 3. Let us suppose the following time windows extending the problem in Example 2: the earliest possible times for pick-up of 5 packages are 4, 3, 6, 3 and 4, the latest acceptable times for their delivery are 12, 13, 10, 15 and 14. In Table 2 traveling times matrix is given.

	1	2	3	4	5	6	7	8	9	10	11
1	0	2	2	5	4	3	3	5	4	1	5
2	2	0	1	6	6	4	5	5	5	3	7
3	2	1	0	7	6	3	4	6	5	3	6
4	5	6	7	0	5	7	7	1	2	4	7
5	4	6	6	5	0	4	2	6	6	3	1
6	3	4	3	7	4	0	2	7	7	3	4
7	3	5	4	7	2	2	0	7	7	3	2
8	5	5	6	1	6	7	7	0	1	4	7
9	4	5	5	2	6	7	7	1	0	4	7
10	1	3	3	4	3	3	3	4	4	0	4
11	5	7	6	7	1	4	2	7	7	4	0

Table 2: Matrix of traveling times

As the result, 3 vehicles have to be used for pick-up and delivery of all packages:

1. 1 – 10 – 6 – 7 – 11 – 1,
2. 1 – 2 – 3 – 1,
3. 1 – 8 – 9 – 4 – 5 – 1.

Total distance is 410.

4 Conclusions and future work

In this paper, the optimization model for a capacited messenger problem is presented. Introducing time windows brings the model closer to real problems. Because of the NP-hardness of this type of distribution problems, the proposed model can be used for small-sized problems only. In case of huge problems, heuristic algorithms have to be investigated.

If a messenger company enables its customers to call for service through day, the problem must be studied as the dynamic one. In Fabry (2007), models for uncapacited dynamic messenger problem can be found. In additions, multiple depots are often considered instead of one.

This publication was supported by the project 402/06/0123 founded by the Grant Agency of the Czech Republic.

References

[1] CORDEAU, J.-F. 2006. A Branch-and-Cut Algorithm for the Dial-a-Ride Problem. *Operations Research* 54, pp. 573-586.

[2] FÁBRY, J. 2007. Dynamic Messenger Problem. *Communications*. Vol. 9, No. 4, pp. 66-69.

[3] LU, Q., M. DESSOUKY. 2004. An Exact Algorithm for the Multiple Vehicle Pickup and Delivery Problem. *Transportation Science*. Vol. 38, No. 4, pp. 503-514.

Eleonora Fendeková, Michal Fendek

Faculty of Economic Informatics
University of Economics
Dolnozemská cesta 1/b, 852 35 Bratislava

e-mail: nfendek@dec.euba.sk, fendek@dec.euba.sk

Comparison of the Effectiveness of Models for Network Industries Regulation

Abstract

For the regulation of subjects of the chosen industries, which are in light of their market share monopolies or natural monopolies, the state is creating institution, so-called regulator, which task, under the state authority, is to create such legislative environment and regulatory mechanisms, which will ensure market equilibrium for observed commodities while granting fair profit for regulated subjects. Among these industries particularly belong network industries, which subjects often share characteristics of monopolies and state has therefore a desire to regulate them.

In the paper we will analyze effectiveness comparison of the two frequently used schemes for monopoly price regulation—rate of return model and return over costs model. Return over costs is a scheme of natural monopoly regulation, which is in principle different form the model of regulation on the basis of rate of return. It derives the barrier for not exceeding the reasonable profit only from the part of the regulated entity input activities, namely from the volume of investment. This undesirably motivates monopoly to disproportionate increase the capital investments, which is, of course, contra productive.

Return over costs regulation model sets the maximum profit margin for regulated firm on the basis of its total costs. There is a certain analogy between this form of regulation and regulation on the basis of the rate of return. However the difference is that return over costs regulation does not prefer particular cost group, but uses the total costs.

Keywords: *Network industries, regulated prices, reasonable profit in regulated industries*

1 Regulation on the basis of the rate of return

Traditional methodological tool for price regulation applied by price regulators to set maximum price of network industries products is the regulation on the basis of the rate of return—Rate of Return Regulation, by which are the prices of electricity, gas and other companies regulated in the most of the developed countries.

The aim is to ensure that the regulated entity will set the price of commodity or service for its consumers in a way the revenues will cover all its *reasonable and provident costs incurred* as well as regulated return on its *provident*[1] investment.

Let us now analytically derive the allowable rate of cost return for investment, co called *RoR* parameter of the regulated entity. Let us suppose that the firm is producing a homogenous product in production volume q, which it realizes on a relevant market for the price p.

Let us further suppose that the firm uses two production factors, namely labor force with consumption level L by labor price w and the capital with consumption level K by the capital price r.

The profit of the firm is generally defined as the difference between the yields and costs

$$\pi(q) = t\left(q\right) - n\left(q\right),$$

where

$t(q) = p \times q$ – function of revenues of the firm, $t: \ R \to R$,

$n(q) = nv(q) + n_f$ – function of the total costs of the firm, $n : R \to R$,

$nv(q)$ – function of the variable costs of the firm, nv: R→ R,

n_f – fixed costs of the firm, $n_f \in R$.

If we substitute general cost function on the basis of consumption of production factors, we get a profit function in a following form

$$\pi(q) = p \times q - w \times L - r \times K.$$

If we further express the production volume q on the basis of the production function in the form

$$q = f(K, L)$$

and the production price p on the basis of the price-demand function in the form

$$p = p(q),$$

then we can express the profit function in the form

$$\pi(q) = p(q) \times q - w \times L - r \times K$$

and after further modification in the form

[1] It should be noted, that under term provident we understand situations, when certain decision about investment or expenditures has been made in conditions of verified and relevant information available in the time of decision making.

$$\pi(q) = p(f(K,L)) \times f(K,L) - w \times L - r \times K.$$

A non-regulated firm can set its endogenous decision parameters in any way. So it chooses an optimum output volume $q*$, an acceptable optimum price $p*$ and corresponding consumption levels of the production factors labor L and capital K in a way to reach maximum profit. Optimum output and optimum price will be calculated by solving the following mathematical programming task

$$\pi(q) = p(f(K,L)) \times f(K,L) - w \times L - r \times K \to \max$$

$$K, L \ \in \ R_{\geq 0}.$$

In this case the non-regulated firm has no formal boundaries for setting the parameters guaranteeing its maximum profit. On the other hand, the regulated firm must respect boundaries given by the regulator.

Price regulation regime on the basis of the rate of return lies in a fact, that through the use of exogenously defined control variable RoR the allowable level of quotient of the revenues $p \times q$ reduced by its non-capital expenditures $L \times w$ and the volume of consumed capital K is regulated.

In other words, a firm can optimize or freely determine the consumption levels of labor L, capital K by the market prices of production factors w, r and on the other side the level of its production q but also the production price p. However firm has to respect the rate of return defined by the regulator i.e. the validity of the relation

$$RoR \geq \frac{p \times q - w \times L}{K}.$$

Let us further in detail explore the relation between the rate of return of the capital expenditures and the profit of the regulated entity. Profit can be analytically expressed as the difference between the proceeds and the costs of the firm in the form

$$\pi(q) = p \times q - w \times L - r \times K.$$

After another modification we get

$$(RoR - r) \times K \geq \pi(q).$$

However this form of price regulation is hiding one serious risk. It often motivates firm to use higher volume of variable input capital than non-regulated firm. Firm could produce with price p_R^* regulated production volume q_R^* also by other combination of variable inputs labor and capital than the optimum regulated combination labor L_{Reg}^* and capital K_{Reg}^* is.

2 Regulation on the basis of the rate of costs

Return over costs regulation sets the maximum profit margin for regulated firm on the basis of its overall costs. We can see that there is certain analogy between this form of regulation and regulation on the basis of the rate of return. However the difference is that return over costs regulation does not prefer particular cost group, but uses the overall costs.

In short, the keystone of return over costs regulation is that regulator as the base for regulated entity's reasonable profit definition sets its overall costs and defines reasonable profit as a certain allowed percentage RoC of its costs. Analytically we can express this condition as

$$RoC \times n\left(q\right) \geq \pi\left(q\right)$$

or

$$RoC \times \left(w \times L + r \times K\right) \geq \pi\left(q\right).$$

Regulated output and regulated price in the return over costs environment are calculated by solving the following mathematical programming task

$$\pi(q) = \pi\left(f\left(K, L\right)\right) = p\left(f\left(K, L\right)\right) \times f\left(K, L\right) - w \times L - r \times K \rightarrow \max \tag{1}$$

subject to

$$p\left(f\left(K, L\right)\right) \times f\left(K, L\right) - w \times L - r \times K - RoC \times \left(w \times L + r \times K\right) \leq 0 \tag{2}$$

$$K, L \ \in \ R_{\geq 0}. \tag{3}$$

Solution of the optimization task (1) ... (3) is optimal consumption volume of production factors labor L^* and capital K^*, on the basis of which, with the help of the production function, the regulated optimal volume of output q^*_{RoC} is quantified

$$q^*_{RoC} = f(K^*, L^*)$$

and regulated optimal price p^*_{RoC} with the help of the price-demand and production function on the basis of the relation

$$p^*_{RoC} = p(q^*_{RoC}) = p(f(K^*, L^*).$$

While also in this regulation approach the rate of return on revenues defined by the parameter RoC is respected, i.e. exogenous control parameter set by the regulator.

Let us now transform the optimization task (1) ... (3) with two variables L, K into a task with one variable q in a following way

$$\pi(q) = p(q) \times q - n(q) \to \max \tag{4}$$

subject to

$$p(q) \times q - n(q) - RoC \times n(q) \leq 0 \tag{5}$$

$$q \in R_{\geq 0}, \tag{6}$$

where

$t(q) = p \times q$ – function of revenues of the firm, $t: R \to R$

$n(q) = nv(q) + n_F$ – function of the total costs of the firm, $n: R \to R$

$nv(q)$ – function of the variable costs of the firm, $nv: R \to R$

n_F – fixed costs of the firm,$n_F \in R$

RoC – reasonable profit margin set by the regulator corresponding to the unit costs.

On the basis of the substitution $t(q) = p(q) \times q$ we can reformulate the optimization task (4) ... (6) to

$$\pi(q) = t(q) - n(q) \to \max \tag{7}$$

subject to

$$t(q) - n(q) - RoC \times n(q) \leq 0 \tag{8}$$

$$q \in R_{\geq 0}. \tag{9}$$

Regulated firm has the tendency to set its decision making parameters in a way the limit set by the regulator would allow it to reach maximum profit.

The comparison of the products market price and production costs of the firm also leads to interesting conclusion while using this type of the regulation. Let us explore the reasonable profit margin in return over costs regulation from this aspect again.

$$RoC \times TC(Q) \geq \pi(Q). \tag{10}$$

In this condition we express the total cost function of the firm analytically. We get reformulated condition expression (10):

$$p(q) \leq (1 + RoC) \times np(q), \tag{11}$$

where

$np(q) = \frac{(nv(q)+n_F)}{q}$, $\quad q > 0 \quad$ are total average costs of the firm.

Relation (11) represents substantial feature of the return over costs regulation, which also explains already mentioned mystification about direct relation of the regulated reasonable profit margin and total costs of the natural monopoly.

3 Conclusions

On the basis of the model analysis several important conclusions about the rate of return regulated firms' behavior are elaborated:

1. Rate of return on capital regulated firm in effort to raise its allowed reasonable profit is motivated to inappropriate and wasteful increase of capital investments.

2. When the rate of return on capital investments is decreased, assuming $RoR > r$ is still valid, firm in order to maintain profit volume rises capital expenditures.

   On the basis of the relation (11) we can formulate the following important conclusions about firm behavior in the condition of the return over costs regulation:

3. In general, return over costs regulation constructs reasonable profit margin for the regulated entity on the basis of the proportional portion of its total expended costs. This proportional portion is defined by the RoC parameter. So primary it encourages the producer to produce greater volume of supply by the lower price, which is increasing social welfare.

4. The particular optimal position of the regulated firm is determined by the characteristics of the cost function, which is directly related to the character of the profit function on one side, because

$$\pi(q) = t(q) - n(q)$$

   and on the other side by the characteristics of the price-demand function $p(q)$, which specifies elastic and inelastic demand zones.

5. From the relation (11) we can see that regulated firm can set its production parameters, production prices and consumption of production factors only in a manner for its production market price to be up to RoC percent greater than the average unit costs of production.

We can see that ineffective cost increase of the firm, in accordance with regulatory relation of this method, would be albeit creating room for reasonable profit increases however the validity of the relation (11) needs to be ensured and such combination of supply production price found, that would ensure its consumption.

References

[1] BIEN, F.: *Systemwechsel im Europischen Kartelrecht.* In: Der Betrieb N' 46/2000. Dsseldorf: Verlagsgruppe Handelsblatt, 2000.

[2] FENDEK, M, FENDEKOVÁ, E: *Mikroekonomická analýza.* Bratislava: IURA Edition, 2008.

[3] FENDEKOVÁ, E.: *Podmienky rovnováhy firmy v rôznych typoch trhových štruktúr. In: Transformácia ekonomiky a rozvoj podnikania v SR.* Bratislava: FPM EU, 1995

[4] FENDEKOVÁ, E: *Oligopoly a regulované monopoly.* Bratislava: IURA Edition, 2006.

[5] O'SULLIVAN, A., SHEFFRIN, S., PEREZ, P.: *Microeconomics: Principles, Applications, and Tools.* New York: Prentice Hall, 2006.

[6] PEPALL, L., RICHARDS, D. J.D. J., NORMAN, D.: *Industrial Organization: Contemporary Theory and Practice (with Economic Applications).* New York: South-Western College Publishing, 2004

[7] SHY, OZ.*: The Economics of Network Industries.* Cambridge: Cambridge University Press, 2001

[8] TRAIN, K., E.: Optimal Regulation. London: The MIT Press, 1995

Petr Fiala, Václav Kořenář

Department of Econometrics

University of Economics Prague

W. Churchill sq. 4, 130 67 Prague 3

e-mail: pfiala@vse.cz, korenar@vse.cz

Supply Chain Formation by Combinatorial Auctions

Abstract

Supply chain management has generated a substantial amount of interest both by managers and by researchers. Supply chain management is now seen as a governing element in strategy and as an effective way of creating value for customers. Complex business negotiations often involve interrelated exchange relationships among multiple levels of production. The paper describes an approach for modeling and solving the supply chain formation problem. Supply chain formation is the problem of determining the production and exchange relationships across a supply chain. To respond to rapidly changing market conditions, companies must be able to dynamically form and dissolve business interactions, requiring automated support for supply chain formation. The problem can be modeled as the task dependency network. A task dependency network is a directed acyclic graph, representing dependencies among agents and goods. Using of combinatorial auctions is promising for solving the supply chain formation problem. Combinatorial auctions are those auctions in which bidders can place bids on combinations of items. Agents submit bids reporting costs and values, and then the auction computes an allocation that maximizes the reported value and informs the agents of results. There are some ways to extend the bidding policies to accommodate more general production capabilities and consumer preferences.

Keywords: *supply chain management, supply chain formation, task dependency network, combinatorial auction*

Introduction

Supply chain management has generated a substantial amount of interest both by managers and by researchers. Supply chain management is now seen as a governing element in strategy and as an effective way of creating value for customers. There are many concepts and strategies applied in designing and managing supply chains (see Simchi-Levi, Kaminsky, Simchi-Levi, 1999). The expanding importance of supply chain integration presents a challenge to research and focus more attention on supply chain modeling (see Tayur, Ganeshan & Magazine, 1999). Supply chain management is more and more affected by network and dynamic business environment. The overall business environment is becoming increasingly dynamic. Demand and supply for custom products can be very dynamic. Supply chains operate in network environment as supply networks. Dynamic information and decision-making models are called to accommodate this new changes and uncertainties. Complex business negotiations often involve interrelated exchange relationships among multiple levels of production. The paper describes an approach for modeling and solving the supply chain formation problem. The problem is formulated in Section 2. The model as the task dependency network is introduced in Section 3. In Section 4, using of combinatorial auctions for solving the supply chain formation problem is presented. Results and possible extensions are discussed in Section 5.

1 Supply chain formation problem

Supply chain formation is the problem of determining the production and exchange relationships across a supply chain (Walsh, Wellman, Ygge, 2000, Walsh, Wellman, 2003). To respond to rapidly changing market conditions, companies must be able to dynamically form and dissolve business interactions, requiring automated support for supply chain formation. Agents in the supply chain are characterized in terms of their capabilities to perform tasks, and their interests in having tasks accomplished. A central feature of the model of the problem is hierarchical task decomposition. In order to perform a particular task, an agent may need to achieve some subtasks, which may be delegated to other agents. These may in turn have subtasks that may be delegated, forming a supply chain through a decomposition of task achievement. Constraints on the task assignment arise from resource contention, where agents require a common resource to accomplish their tasks. Tasks are performed on behalf of particular agents. If two agents need a task then it would have to be performed twice to satisfy them both. In this way, tasks are the same as any other discrete, rival resource. Hence, there is no distinction in the model, and use the term “good” to refer to any task or resource provided or needed by agents. The assumption that goods cannot be shared or reused is necessary for analysis.

2 Task Dependency Network

The problem can be formulated as so called a task dependency graph (Walsh, Wellman, Ygge, 2000). A task dependency network is a directed acyclic graph $G = (V, E)$, representing dependencies among agents and goods. $V = G \cup A$, where G is the set of goods and $A = C \cup \Pi$ is the set of agents, comprised of consumers C, and producers Π. Edges, E, connect agents with goods they can use or provide. There exists an edge $< g, a >$ from $g \in G$ to $a \in A$ when agent a can make use of one unit of g, and an edge $< a, g >$ when a can provide one unit of g. When an agent can use or provide multiple units of a good, separately indexed edges represent each unit. The goods can be traded only in discrete quantities.

Figure 1 shows an example task dependency network for a supply chain problem. Here the goods are indicated by circles, and agents by boxes and triangles. Suppliers and consumers are indicated by boxes and producers by triangles. The numbers under agent boxes represent production costs and consumption values. An arrow from an agent to a good indicates that the agent can provide that good, and an arrow from a good to an agent indicates that the agent can make use of the good.

An allocation is a subgraph $(V', E') \subseteq (V, E)$. For $g \in G$, an edge $< a, g > \in E'$ means that agent a provides g, and $< g, a > \in E'$ means a acquires g. A producer is active iff it provides its output. A producer is feasible iff it is inactive or acquires all its inputs. Consumers are always feasible. An allocation is feasible iff all agents are feasible and all goods are in material balance, that is the number of edges into a good equals the number of edges out.

The value of allocation (V',E') is:

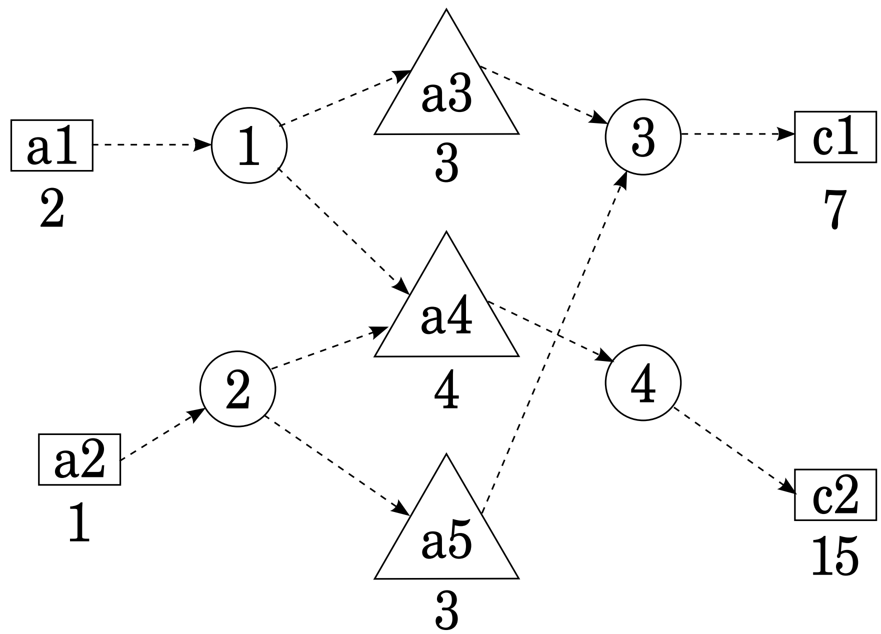


Figure 1: Task dependency graph

$$value((V', E')) \equiv \sum_{c \in C} v_c(E') - \sum_{\pi \in \Pi} k_\pi(E'),$$

where

v_c is consumption value for the consumer c, k_π is production cost for the producer π.

The set of efficient allocations contains all feasible allocations (V*, E*) such that

$$\text{value}((V*, E*)) = max_{V', E'})\{\text{value}(V', E'))|(V', E')\text{is feasible}\}.$$

A solution is a feasible allocation such that one or more consumers acquire a desired good. If $c \in C \cap V'$ for solution (V', E'), then (V', E') is a solution for consumer.

Figure 2 presents a suboptimal solution and Figure 3 presents an efficient solution of the supply chain formation problem.

Value of allocation is: 7 – 2 – 1 – 3 – 3 = – 2.

Value of allocation is: 15 – 2 – 1 – 4 = 8.

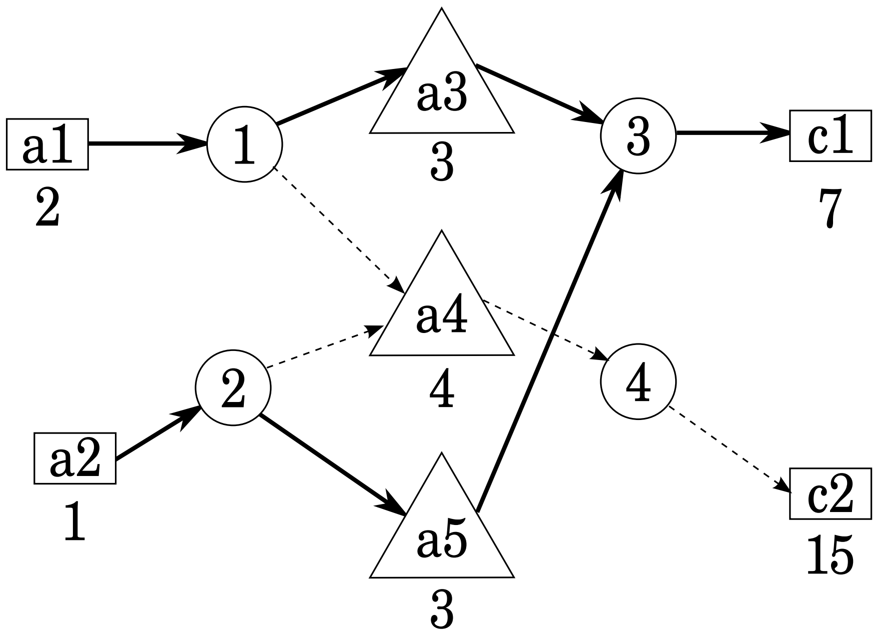


Figure 2: Sub-optimal solution

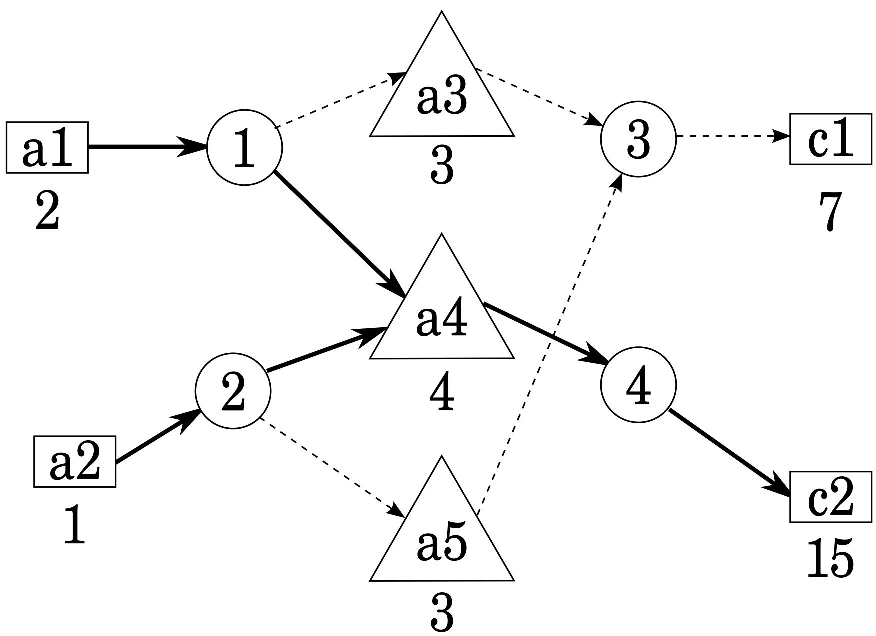


Figure 3: Efficient solution

3 Combinatorial Auction Mechanism

Auctions are important market mechanisms for the allocation of goods and services. Design of auctions is a multidisciplinary effort made of contributions from economics, operations research, informatics, and other disciplines. Combinatorial auctions are those auctions in which bidders can place bids on combinations of items, so called bundles. The advantage of combinatorial auctions is that the bidder can more fully express his preferences. This is particularly important when items are complements. The auction designer also derives value from combinatorial auctions. Allowing bidders more fully to express preferences often leads to improved economic efficiency and greater auction revenues. However, alongside their advantages, combinatorial auctions raise a host of questions and challenges (Cramton et al., 2006).

Many types of combinatorial auctions can be formulated as bivalent mathematical programming problems. Complexity is a fundamental question in combinatorial auction design. There are some types of complexity:

- computational complexity,
- valuation complexity,
- strategic complexity,
- communication complexity.

Computational Complexity covers such questions: How much computation is expected of the mechanism to compute an outcome given the bid information of the bidders. This is an extremely important question because winner determination problem is an NP-complete optimization problem. The winner determination problem turns out to be an instance of a weighted set packing problem. The weighted set packing problem is a problem of finding a disjoint collection of weighted subsets of a larger set with maximal total weight. Weighted set packing is a classical NP-complete problem.

Valuation complexity deals with such questions: How much computation is required to provide preference information within a mechanism? Estimating every possible bundle of items requires exponential space and hence exponential time. Bidders need to determine valuations for $2^m - 1$ possible bundle.

Strategic complexity concerns such questions: Which of the $2^m - 1$ bundles to bid on? What is the best strategy for bidding? Must bidders model behavior of other bidders and solve problems to compute an optimal strategy? For instance, in a sealed bid combinatorial procurement scenario, sellers will need to not only take their valuation of the bundles into consideration but also the bidding behavior of their competitors. This requires sophisticated bidding logic.

Communication complexity concerns such questions: How much communication is required to exchange between bidders and auctioneer until an equilibrium price is reached the mechanism to compute an outcome. The amount of communication between the bidders and the auctioneer can become quite high. For instance, in an iterative combinatorial auction, where individual valuations are revealed progressively in an iterative manner, the communication costs could be high if the auction were conducted in a distributed manner over space and/or time. The problem of communication complexity can be addressed through the design of careful bidding languages that provide expressive but concise bids.

Using of combinatorial auctions is promising for solving the supply chain formation problem (Walsh, Wellman, Ygge, 2000, Walsh, Wellman, 2003). Combinatorial auction mechanism is one-shot mechanism. Agents submit bids reporting costs and values, and then the auction computes an allocation that maximizes the reported value and informs the agents of results. An agent pays the price it bids for the allocation it receives. If the auction receives more money than it pays out, the proceeds are distributed evenly among all consumers.

3.1 Bidding Language

Agent a places a bid b_a of the form

$$< r_a, < g_1, q_a^1 >, \ldots, < g_n, q_a^n >>$$

where

q_a^i is the integer quantity that agent a demands (positive for input demands and negative for output demands) for good g_i, and

r_a is its reported willingness to pay (or be paid, in the case of negative numbers) for the demanded bundle of goods.

Given a set of bids B, the auction computes the winning allocation from:

$$f(B) = \max_x \sum_{b_a \in B} r_a x_a$$

subject to

$$\sum_{b_a \in B} q_a^i x_a = 0, \; i = 1 \ldots n,$$

where $x_a = 1$ if agent a wins the bid, and $x_a = 0$ otherwise.

3.2 Combinatorial Bidding Policies

If agents behave non-strategically (i.e., bid their true valuations) in the combinatorial auction mechanism, then the result will always be an efficient allocation.

Analysis for the case when agents behave strategically

- Assumption that it is common knowledge that consumers bid their true values.
- Producers play Bayes-Nash equilibrium (BNE) strategies. Computing a BNE is generally very difficult. In order to find a solution the setting is often constrained by making simplifying assumptions.
- N buyers one seller case with producer's cost k_π (to provide its output) and consumer's values drawn from uniform probability distribution [0,1], and buyer i has value v_i for the good, then a Bayes-Nash equilibrium bidding policy is

$$r_\pi = -k_\pi - \frac{1}{N}(1 - k_\pi).$$

Plausible bidding policy for complicated networks:

- A producer bids to obtain a fraction of the expected available surplus scaled by the expected proportion of its contribution to the global value.
- Let $\Pi^* \subseteq A$ be the producers participating in the efficient allocation. The contribution of these producers to the value of the allocation is Δ^* , where

$$\Delta^* = f^*(A) - f^*(A - \Pi^*).$$

  The function $f(.)$ is defined by the above formulation of the winning allocation problem.
- The contribution Δ_π of a producer to the value of an allocation is the difference between the efficient global value, and the global value with π excluded from the allocation,

$$\Delta_\pi = f^*(A) - f^*(A - \{\pi\}).$$

- A producer's relative contribution can then be defined in terms of its expected proportional contribution, conditional on its being part of the efficient allocation

$$r_\pi = -k_\pi - E\left[\frac{\Delta_\pi \Delta^*}{\sum_{\pi \in \Pi} \Delta_\pi} | \pi \in \Pi^*\right].$$

4 Conclusions

Supply chain formation is a very important problem of determining the production and exchange relationships across a supply chain. Whereas typical research in supply chain management focuses on optimizing production and delivery in a fixed supply chain structure, the approach is concerned with ad hoc establishment of supply chain relationships in response to varying needs, costs, and resource availability. The task dependency network model provides a basis for understanding the automation of supply chain formation. Using of combinatorial auctions is promising for solving the supply chain formation problem. There are some ways to extend the bidding policies to accommodate more general production capabilities and consumer preferences. With these extensions it can be modeled capabilities and preferences on multi-attribute goods (e.g., goods with multiple features such as quality and delivery time, in addition to price and quantity) by simply representing each configuration as a distinct good in the network.

Acknowledgements

The research project was supported by Grant No. 402/07/0166 from the Grant Agency of Czech Republic "Combinatorial Auctions—Modeling and Analysis".

References

[1] Babaioff, M., Walsh, W. E. (2003): Incentive-Compatible, Budget-Balanced, yet Highly Efficient Auctions for Supply Chain Formation. In *Fourth ACM Conference on Electronic Commerce*, 64–75.

[2] P. Cramton, Y. Shoham and R. Steinberg (eds.) (2006): *Combinatorial Auctions.* MIT Press, Cambridge.

[3] Karimi, R., Lucas, C., Moshiri, B. (2007): New Multi Attributes Procurement Auction for Agent-Based Supply Chain Formation. *International Journal of Computer Science and Network Security* 7, 255-261.

[4] Simchi-Levi, D., Kaminsky, P., Simchi-Levi, E. (1999): *Designing and Managing the Supply Chain: Concepts, Strategies and Case studies.* Boston: Irwin/ Mc Graw-Hill.

[5] Tayur, S., Ganeshan, R., & Magazine, M. (1999): *Quantitative models for supply chain management.* Boston: Kluwer.

[6] Walsh, W. E., Wellman, M. P. (2003): Decentralized Supply Chain Formation: A Market Protocol and Competitive Equilibrium Analysis. *Journal of Artificial Intelligence Research* 19, 513-567.

[7] Walsh, W. E., Wellman, M. P., Ygge, F. (2000): Combinatorial Auctions for Supply Chain Formation. In *Second ACM Conference on Electronic Commerce*, 260–269

Vladimir Fishman, Yuri Galperin

Decision Analytics Division

Experian Inc.

10 Talcott Notch Road Suite, 105 Farmington CT, U. S. A.

e-mail: `Vladimir.Fishman@experian.com, Yuri.Galperin@experian.com`

Scalable Customer Level Optimization in Marketing and Risk Management

Abstract

The record level constrained scalable optimization solution has been developed to address the problem of determining the best decision strategy for marketing campaigns. The goal of optimization is to determine which decision options should be selected for each prospective customer, so that the resulting decisions across the millions of prospects under consideration maximizes a business objective—such as profitability—while satisfying business and operational constraints. Constraints can be subdivided into two separate classes: 1) campaign level (global, resource) constraints like marketing budget or other resource limitations; and 2) customer level (record, household) restrictions (contact rules) like eligibility conditions or limitations on the total number of assignments within a "peer group" of similar decision options.

The problem under consideration is an integer linear optimization problem. In typical business cases the size of the problem is huge: number of customers $N = O(10^8)$, number of decision options $M = O(10^2)$, number of resource constraints $K = O(10^2)$. Obviously, a problem of such size cannot be solved using any standard approach. To address this problem, we developed a new iterative scalable algorithm based on the Branch-and-Bound technique and the Lagrange relaxation method.

We will also discuss the algorithmic approach and practical usage of a more general formulation of the customer level problem: offer parameter optimization and sliding time windows; and general contact rules: arbitrary nested Boolean expressions over linear conditions or conditional assignments. In addition, we will consider non-linear assignment interactions in the form of cannibalization and saturation effects. And finally we will discuss the multi-objective nature of the problem.

Keywords: *large scale optimization, Lagrange relaxation, marketing, risk management*

1 Introduction

The optimization approach described in this paper was originally developed to address the problem of determining the best decision strategy for marketing campaigns. Our method provides a precise and scalable solution for the problem at the customer level. In the context of cross-sell marketing campaigns, the goal is to find which of the dozens of potential offers should be extended to which of the thousands or even millions of prospective customers, so that the resulting decisions maximize a business objective (such as profitability or sales penetration) while satisfying a range of complex business and operational constraints:

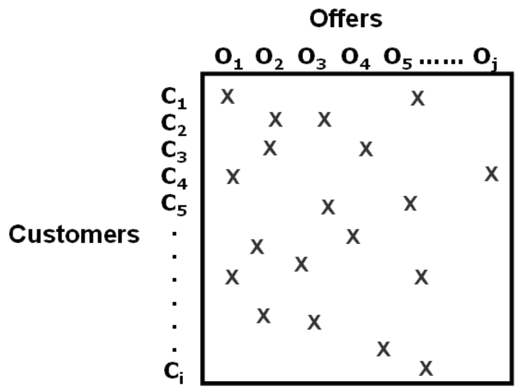


Figure 1: Schematic structure of the optimization problem

Constraints can be sub-divided into two separate classes: 1) campaign level (global, resource) constraints like marketing budget or other resource limitations; and 2) customer (record, household) level restrictions like eligibility conditions and limitations on the number of offers (total, or within a "peer group" of similar promotions).

The problem under consideration in its simplest form can be expressed as follows:

The objective is to maximize utility function:

$$\begin{aligned} &U(X) = \textstyle\sum_{i=1}^{N} \sum_{j=1}^{M} u_{ij} a_{ij} \\ &a \in [0,1] \end{aligned} \tag{1}$$

where:

u_{ij} – is expected value of ij customer/offer contribution to the utility function

Subject to constraints:

Global (resource) constraints:

$$B_k(X) = \sum_{i=1}^{N} \sum_{j \subset I_k} v_{ij}^{k} a_{ij} < C_k k = 1, \ldots, K \tag{2}$$

The "ratio type" global constraints:

$$\sum_{i=1}^{N} \sum_{j \subset I_k} v_{ij}^{k} a_{ij} / \sum_{i=1}^{N} \sum_{j \subset J_k^1} w_{ij}^{k} a_{ij} < C_k^1 \tag{3}$$

fall into the same category.

Where:

v_{ij}^k and w_{ij}^k – are expected value of ij customer/offer contribution to k metric value in numerator and denominator expressions respectively

Customer level constraints:

$$\sum_{j=1}^{M} a_{ij} \leq h_i i = 1, \ldots, N \quad \text{(maximum offers per record)} \tag{4}$$

$$\sum_{j \in J_G} a_{ij} \leq 1 \tag{5}$$

where:

h_i – is the maximum number of offers that can be solicited to a customer

J_G – is a set of promotions which belong to a group of similar promotions

Most practitioners realize that in today's highly competitive environment the optimization of marketing strategies is a necessity rather than a luxury. Because of dimensionality, other software products typically use a simplified approach based on segmentation ([1], [2]). The idea behind the segmentation approach is to aggregate similar customer accounts into clusters and then treat all accounts in a cluster as identical. The optimization problem thus becomes low dimensional and can be solved using a standard technique. If data is highly segmented and there are no record level conditions, this approach can be successful.

However, if models predicting customer behavior are more sophisticated, predictions will be different for different customers; as a result, data is no longer segmented. In such cases the segmentation solution becomes significantly sub-optimal. An additional drawback is that the level of sub-optimality cannot be reliably evaluated. Our comparison of segment-based and "true" optimization shows that for real business cases the loss due to sub-optimality is about 20%. Besides, since segment-based optimization treats all customers in a segment as identical, it cannot properly handle even simple record level contact rules and conditions.

2 Scalable record level optimization: basic approach overview

The problem under consideration is an optimal assignment problem. It has a block-diagonal structure; each of N blocks corresponds to customer level constraints, and additional K rows correspond to resource constraints.

Our approach is a modification of the generalized Lagrange multiplier technique. It can be formulated as follows:

$$Min_{\lambda} Max_{a_{ij}} (\sum_{i=1}^{N} \sum_{j=1}^{J} a_{ij} u_{ij} + \sum_{k} \lambda_k (\sum_{i=1}^{N} \sum_{j=J_k}^{J} a_{ij} v_{ij}^k - C_k)) \tag{6}$$

$$
\begin{aligned}
&A_m = \sum_{i \in I_m}^{N_m} \alpha_i a_i - G_m > 0 \\
&i = 1...N \qquad\qquad\qquad\qquad \text{Row level conditions} \\
&m = 1...M
\end{aligned}
\tag{7}
$$

For each set of λ the problem is separable and can be solved in a straightforward way by solving each of the record level sub-problems. Essentially, the problem is transformed into finding the minimum of the convex K-dimensional continuous piecewise differentable function. It is done by iterations using the sub-gradient to compute the next point in λ space. Each point computation involves solving N sub-problems, and each is an integer optimization problem of a relatively small size and can be solved using the Branch-and-Bound or similar technique [4]. On each iteration λ values are incremented in a sub-gradient direction. The increments are proportional to gradient values and are adjusted depending on the stage of conversion. The iteration process continues until a solution is feasible and the Karush-Kuhn-Tucker conditions:

$$
\lambda_k(\sum_{i=1}^{N}\sum_{j=J_k}^{J} a_{ij} v_{ij}^k - C_k) = 0 \tag{8}
$$

are satisfied within a predefined tolerance. The number of required iterations depends on the tolerance level and the complexity of the project. For a 1% tolerance it normally takes less than 100 iterations to converge. Initial values of λ are important for a faster convergence. To find them we use a random subset of the original dataset and solve the original problem for this subset as a continuous optimization problem with properly scaled conditions. We then use dual variables which correspond to resource constraints as initial values for λ.

3 More advanced formulation of the record level problem

The optimization approach outlined above has proved to be successful for a variety of business cases. The software based on it is an effective optimization tool widely used in the marketing and risk management analytical community. However, as practitioners become more and more sophisticated they use more advanced approaches to predict customer behavior and apply complex business rules, and the formulation described above becomes overly simplistic. To address this challenge we utilized the structural properties of the task. Within the framework of a generalized Lagrange multiplier technique, on each step of the iteration search the problem is separable, and each sub-problem can be solved independently. Therefore, the record level (relatively low-dimensional) sub-problem can be generalized and made more complex without modification of the overall framework.

Below are outlined some areas where we made the sub-problem level formulation more general and added enhancements to the optimization software.

3.1 Household functionality

If two or more customers are member of the same household, promotion selection for them is no longer independent as there may be restrictions that link them together. From a structural point of view it means that sub-problems can be of different sizes with overlapping linear restrictions.

3.2 Offer parameter optimization

The complexity of the record level sub-problem obviously depends on the number of possible decision options. It is common that metrics associated with decision options are multidimensional functions of parameters. As a typical example, credit card promotion property metrics are characterized by credit line and APR. One can certainly consider each combination of parameters as an individual decision option; but in this case the sub-problem becomes very large, not easily traceable, and computationally time consuming.

Alternatively, we may formulate the problem as selecting an assignment for one dimension (credit line, for example) along with the best value associated with this assignment parameter P_{ij}. We assume that $u_{ij}(P_{ij})$ and $v_{ij}^k(P_{ij})$ are known functions and can be discrete or continuous.

$$\begin{aligned}&Min_\lambda Max_{a_{ij},P_{ij}}(\textstyle\sum_{i=1}^N \sum_{j=1}^J a_{ij}u_{ij}(P_{ij}) + \sum_k \lambda_k(\sum_{i=1}^N \sum_{j=J_k}^J a_{ij}v_{ij}^k(P_{ij}) - C_k)) = \\ &Min_\lambda Max_{a_{ij}}(\textstyle\sum_{i=1}^N(\sum_{j=1}^J a_{ij}(u_{ij}(P_{ij}^*) + \lambda_k v_{ij}^k(P_{ij}^*) - C_k)))\end{aligned} \tag{9}$$

For the current iteration value λ and each record-promotion combination, the optimal parameter values $P_{ij}^*(\lambda)$ is the solution the of one-dimensional maximization problem.

For the above example, the optimization procedure will identify for each customer the optimal credit line allocation and the associated APR value that should be extended.

3.3 Sliding time window

This functionality is designed to support multi-stage campaign optimization. It allows the user to select a group of promotions and specify maximum and minimum conditions over a number of consecutive time intervals, regardless of their beginnings and ends.

3.4 Contact rules

There are many business cases in which, as a part of a record or household level sub-problem, one must satisfy some "contact rules" which usually have the form of *if-then-else*, or *either-or* conditions as in following example:

The Boolean condition can be quite complex and nested.

Formulation:

Let us define $\sum_{i\in I}^N \alpha_i a_i > M$ elementary linear condition for a group of promotions, where α_i are arbitrary coefficients (not necessarily integer), and $a_i \in [0,1]$ are assignments.

Let P_1, P_2, P_3 be arbitrary nested combination of OR- and AND- groups of elementary conditions.

The contact rule is defined as follows:

$$\begin{aligned}&If(P_1)\\&then(P_2)\\&else(P_3)\end{aligned}$$

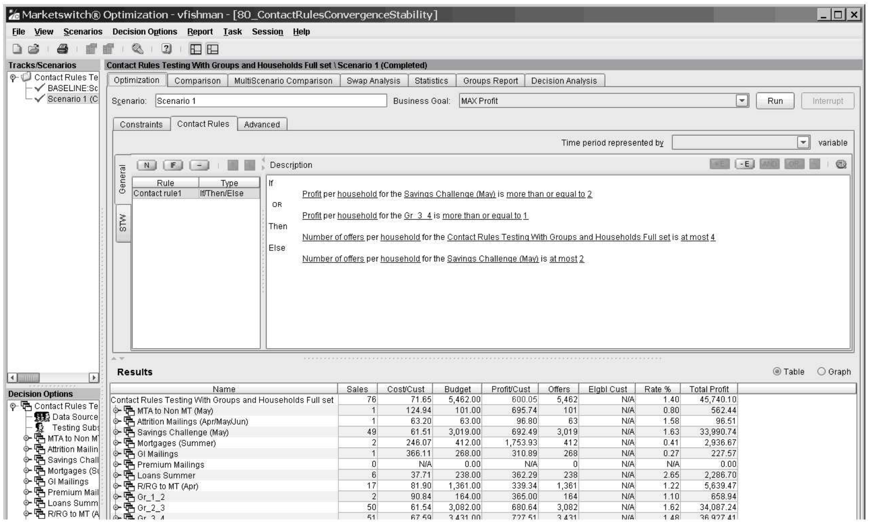


Figure 2: If-Then-Else Contact Rules

An arbitrary number of contact rules may be specified on a record or household level. Promotion groups for different contact rules may intersect. This feature can be combined with sliding time window functionality.

The goal is to find the optimal solution while satisfying contact rules as well as other record and household level conditions and resource constraints. This issue can be addressed within the generalized Lagrange multiplier framework. For solving the corresponding sub-problem we developed a factorization procedure to convert Boolean expression into linear constraints over extended set of binary variables.

3.5 Promotion interaction: saturation and cannibalization effects

In the current approach we assume that all decision options make an independent contribution to the utility function and metrics. Under this assumption, for example, for any two decision options A and B the expected profitability (or response rate) of A does not depend on whether or not B is selected. As many practitioners agree, in many actual business cases this assumption is over-simplistic, and it is essential to take into account so called cannibalization and saturation effects. Solving this problem, however, involves difficult issues from the standpoints of both optimization and modeling. On the record level we now have a non-linear integer optimization problem. Besides, in many business cases interaction is order dependent and time dependent. So the optimal solution for each customer must be a set of decision options in the proper order or within specified time intervals.

Normally, the number of potential decision options for each customer can be in the hundreds, and solving the problem without any assumptions is unrealistic. To develop an efficient branch-and-bound type algorithm, we assumed that decision options can be sub-divided into interacting groups and the only interaction across these groups is a limitation on the total number of decision options for a customer or household. Interactions within a group can be of different types depending on the business problem and modeling approach used to take the interaction into account. We considered three cases: order independent cannibalization, expressed by a cannibalization matrix; sequencing; and order optimization. In the latter two cases the saturation effect was expressed as degradation coefficients depending on the order or time interval. In our search algorithm we also assumed a "convexity" property, which essentially means that interaction always results in cannibalization rather than synergy.

Our first experiments proved that this approach could be applied to real business cases and produced significant increase in profit.

With the generalization described above the optimization problem (no longer linear) can be formulated as follows:

$$\max_{a_i, P_{ij}} \left(\sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{k=1}^{K} u_{ij}^{k}(a_i, P_{ij}) \right) \tag{10}$$

s. t.

Resource constraints:

$$\begin{array}{l} \sum_{i=1}^{N} \sum_{j \subset I_k} v_{ij}^{k}(a_i, P_{ij}) < C_k \\ \text{or} \\ \sum_{i=1}^{N} \sum_{j \subset I_k} v_{ij}^{k}(a_i, P_{ij}) / \sum_{i=1}^{N} \sum_{j \subset I_k^1} w_{ij}^{k}(a_i, P_{ij}) < D_k \end{array} \tag{11}$$

Sub-problem level constraints:

$$B_m\left(\sum_{i \in I_k}^{N_K} \alpha_{ijk} a_{ijk} < (>)(=) M_{iK}\right) \tag{12}$$

Both utility u_{ij}^{k} and metric v_{ij}^{k} (and w_{ij}^{k}) individual contributions are now functions of parameter values P_{ij} and can be influenced by other assignments through cannibalization effects. Sub-problem level constraints (12) can be arbitrary nested Boolean expressions B_m over linear conditions.

4 Multi-objective optimization

Most practical optimization problems in marketing risk management are multi-objective in nature. Normally, profit is not the only goal; risk considerations, market penetration, and cost are also important. These objectives are often conflicting and the ability to analyze trade-offs is very important. One way of handling the problem is by expressing it through a single combined objective function with weight coefficients for each term according to its relative importance. Obviously, this is not very convenient since the weight coefficients are not known in advance.

We therefore provide the decision maker with the multi-objective optimization tool, which essentially generates a Pareto set or efficient frontier [5]. This multi-objective curve contains a number of points, representing different compromises between two objectives. This makes it easy to analyze trade-offs and select the preferred solution, as in the following example:

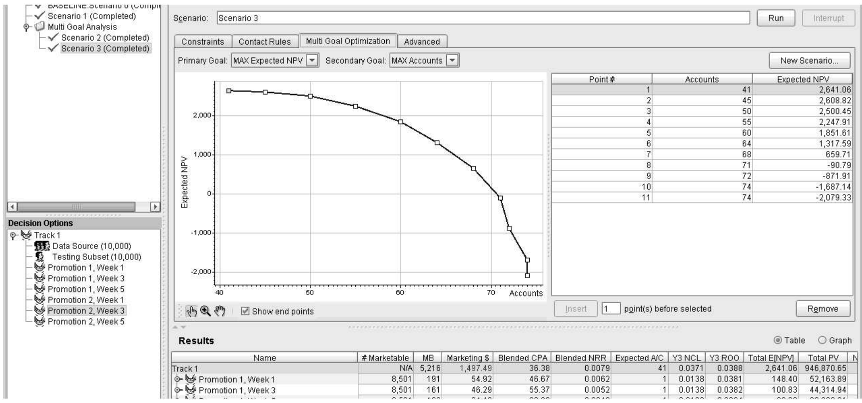


Figure 3: Multi-objective curve representing different compromises between two objectives

5 Areas of future development

As can be seen, the record level optimization approach has significant potential for improving strategies in marketing and risk management. There are, however, issues that deserve intensive future research.

The current economic situation makes abundantly clear that to develop a successful strategy, we must assess its sensitivity to a number of key characteristics which are known only with some level of uncertainty, and which may change in the future as the economic and demographic environment evolves. Such key characteristics can be interest or inflation rates, average income, etc. In multistage campaigns, where decisions are made on a stage-by-stage basis, at the end of each stage some of these uncertainties are resolved, which means the problem we have to deal with is a high dimensional multiple recourse optimization problem.

In making marketing decisions, practitioners normally deal with individual characteristics known only with uncertainty. To predict uncertain parameters like expected profit, delinquency, response, etc., one needs to build predictive models. Thus another important issue related to uncertainty is assessing to what extend better prediction can improve the optimal results of marketing strategy.

Yet another important area of possible improvement is making decision strategy more transparent and intuitive. It is important for a practitioner not simply to arrive at an optimal solution, but also to understand it, to be able to express it in a transparent way, to see if it is in sync with his/her practical experience, and to be able to develop new intuitions. To address this challenge, we recently developed a hybrid "Tree Optimization" approach and made it the core of new software product.

References

[1] *Practical Challenges of Portfolio Optimization*: Fair Isaac White Paper (2004) `http://www.fairisaac.com/NR/rdonlyres/`

/5AF87014-557C-4391-BA1B-34C5D0F827FC/0/
/PracticalPortfolioOptimizationWP_oct04.pdf

[2] Storey, N., and Cohen, M. *Offer Optimization, Optimizing Cross-Sell and Up-Sell Opportunities in Banking.* SUGY27 Proceedings (2002)
http://www2.sas.com/proceedings/sugi27/p112-27.pdf

[3] Everett, H. *Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources.* Operations Res. No. 11 339 – 417 (1963)

[4] Nemhauser, G., and Wolsey L. *Integer and Combinatorial Programming.* New York: John Wiley & Sons, Inc. (1999) ISBN 047182819X

[5] Steuer, R.E. *Multiple Criteria Optimization: Theory, Computations, and Application.* New York: John Wiley & Sons, Inc. (1986) ISBN 047188846X

Lenka Flusserová, Petr Fiala

Department of Econometrics

University of Economics Prague

W. Churchill sq. 4, 130 67 Prague 3

e-mail: `lflusser@vse.cz, pfiala@vse.cz`

Iterative Combinatorial Auctions

Abstract

Combinatorial auctions are those auctions in which bidders can place bids on combinations of items. One way of reducing some of the computational burden in solving combinatorial auctions is to set up a fictitious market that will determine an allocation and prices in a decentralized way. The paper is devoted to analyzing an iterative approach to solving combinatorial auctions. In the iterative approach, there are multiple rounds of bidding and allocation and the problem is solved in an iterative and incremental way. Iterative combinatorial auctions are attractive to bidders because they learn about their rivals' valuations through the bidding process, which could help them to adjust their own bids. There is a connection between efficient auctions for many items, and duality theory. The primal-dual algorithm can be taken as a decentralized and dynamic method of determine the pricing equilibrium. A primal-dual algorithm usually maintains a feasible dual solution and tries to compute a primal solution that is both feasible and satisfies the complementary slackness conditions. If such a solution is found, the algorithm terminates. Otherwise the dual solution is updated towards optimality and the algorithm continues with the next iteration. Several auction formats based on the primal-dual approach have been proposed.

Keywords: *combinatorial auction, iterative approach, dual theory, primal-dual algorithm*

1 Introduction

Auctions are important market mechanisms for the allocation of goods and services. Design of auctions is a multidisciplinary effort made of contributions from economics, operations research, informatics, and other disciplines. Combinatorial auctions are those auctions in which bidders can place bids on combinations of items, so called bundles. The advantage of combinatorial auctions is that the bidder can more fully express his preferences. This is particular important when items are complements. The auction designer also derives value from combinatorial auctions. Allowing bidders more fully to express preferences often leads to improved economic efficiency and greater auction revenues. However, alongside their advantages, combinatorial auctions raise a lot of questions and challenges (Cramton et al., 2006).

The problem, called the winner determination problem, has received considerable attention in the literature. Iterative auctions are considered as an alternative for solving the winner determination problem. Iterative formats such as the English auction are very popular in electronic commerce applications. They allow bidders to learn about their competitors' bids, which is an important aspect if bidders' valuations are affiliated (Milgrom, 1987). In iterative auctions, bidders do not have to submit bids on all possible bundles at once, but can bid only on a small number of bundles in each round.

Unfortunately, designing of iterative combinatorial auctions leads to a number of difficulties:

1. threshold problem,
2. exposure problem,
3. ties,
4. communicative complexity,
5. determining feedback prices.

The well-known threshold problem refers to the difficulty that multiple bidders desiring small bundles that constitute a large bundle may have in outbidding a single bid for the large bundle. The exposure problem considers the risk of winning items at prices above the valuations, which usually happens if a bidder with a super-additive valuation of bundle wins only a part of this bundle. Though the exposure problem is usually typical only for pseudo-combinatorial auctions, it can also become relevant for combinatorial auctions in case of OR-bidding. Avoiding or resolving ties can become a problem, because allocations can be composed of multiple winners. Communicative complexity refers to problem that the amount of communication between the bidders and the auctioneer can become quite high. However, the most fundamental problem in the iterative combinatorial auctions design is determining feedback prices in each iteration. The primal-dual algorithm can be taken as a decentralized and dynamic method of determine the pricing equilibrium. Several auction formats based on the primal-dual approach have been proposed.

2. Winner determination problem

The problem is formulated as: Given a set of bids in a combinatorial auction, find an allocation of items to bidders that maximizes the seller's revenue. Let us suppose that one seller offers a set M of m items, $j = 1, 2, \ldots, m$, to n potential buyers. Items are available in single units. A bid made by buyer i, $i = 1, 2, \ldots,$ n, is defined as

$$B_i = \{S, v_i(S)\},$$

$S \subseteq M$, is a combination of items,

$v_i(S)$, is the valuation or offered price by buyer i for the combination of items S.

The objective is to maximize the revenue of the seller given the bids made by buyers. Constraints establish that no single item is allocated to more than one buyer and that no buyer obtains more than one combination. Bivalent variables are introduced for model formulation:

$x_i(S)$ is a bivalent variable specifying if the combination S is assigned to buyer $i(x_i(S) = 1)$.

The winner determination problem can be formulated as follows

$$\sum_{i=1}^{n} \sum_{S \subseteq M} v_i(S) x_i(S) \quad \rightarrow \quad \max$$

subject to

$$\sum_{S \subseteq M} x_i(S) \leq 1, \ \forall i, i = 1, 2, \ldots, n,$$

$$\sum_{i=1}^{n} \sum_{S \subseteq M} x_i(S) \leq 1, \ \forall j \in M, \tag{1}$$

$$x_i(S) \in \{0, 1\}, \ \forall S \subseteq M, \ \forall i, i = 1, 2, \ldots, n.$$

The objective function expresses the revenue. The first constraint ensures that no bidder receives more than one combination of items. The second constraint ensures that overlapping sets of items are never assigned.

Complexity is a fundamental question in combinatorial auction design. The algorithms proposed for solving the winner determination problem are exact algorithms and approximate ones. Many researchers consider iterative auctions as an alternative.

2 Pricing Schemes

The key challenge in the iterative combinatorial auctions design is to provide information feedback to the bidders after each iteration (Pikovsky, Bichler, 2005). Pricing was adopted as the most intuitive mechanism of providing feedback. In contrast to the single-item single-unit auctions, pricing is not trivial for iterative combinatorial auctions. The main difference is the lack of the natural single-item prices. With bundle bids setting independent prices for individual items is not obvious and often even impossible. Different pricing schemes are introduced and discussed their impact on the auction outcome.

A set of prices $p_i(S), \ i = 1, 2, \ldots, n, \ S \subseteq M$ is called:

- **linear**, if $\forall i, S: \ p_i(S) = \sum_{j \in S} p_i(j)$,
- **anonymous**, if $\forall k, l, S: \ p_k(S) = p_l(S)$.

Prices are linear if the price of a bundle is equal to the sum of the prices of its items, and anonymous if the prices of the same bundle are equal for every bidder. The non-anonymous prices are also called discriminatory prices. The following pricing schemes can be derived using the above definitions:

1. linear anonymous prices,
2. non-linear anonymous prices,
3. non-linear discriminatory prices.

The first pricing scheme is obviously the simplest one. Linear anonymous prices are easily understandable and usually considered fair by the bidders. The communication costs are also minimized, because the amount of information to be transferred is linear in the number of items. The second pricing scheme introduces the non-linearity property, which is often necessary to express strong super- or sub-additivity in the bidder valuations. Unfortunately, non-linear prices are often considered too complex and the communication costs also increase. If even non-linear anonymous prices are not sufficient to lead the auction to competitive equilibrium, the third pricing scheme can be used. However, discriminatory pricing introduces additional complexity and is often considered unfair by the bidders.

A set of prices $p_i(S)$ is called **compatible** with the allocation $x_i(S)$ and valuations $v_i(S)$, if

$$\forall i, S: \; x_i(S) = 0 \;\Leftrightarrow\; p_i(S) > v_i(S) \text{ and } x_i(S) = 1 \;\Leftrightarrow\; p_i(S) \leq v_i(S)$$

The set of prices is compatible with the given allocation at the given valuations if and only if all winning bids are higher than or equal to the prices and all loosing bids are lower than the prices (assuming the bidders bid at their valuations).

Compatible prices explain the winners why they won and the losers, why they lost. In fact, informing the bidders about the allocation $x_i(S)$ is superfluous, if compatible prices are communicated. However, not every set of compatible prices provides the bidder with meaningful information for improving bids in the next auction iteration. Another important observation is the fact that linear compatible prices are harder and often even impossible to construct, when the bidder valuations are super- or sub-additive.

A set of prices $p_i(S)$ is in **competitive equilibrium** with the allocation $x_i(S)$ and valuations $v_i(S)$), if

1. The prices $p_i(S)$ are compatible with the allocation $x_i(S)$ and valuations $v_i(S)$).
2. Given the prices $p_i(S)$, there exists no allocation with larger total revenue than the revenue of the allocation $x_i(S)$.

The idea behind this concept is to define prices characterizing the optimal allocation. The prices may not be too low to violate the compatibility condition 1, but they may not be too high to violate the condition 2. In general, one can show that the existence of competitive equilibrium prices implies optimality of the allocation and that the opposite is also true in case of non-linear discriminatory prices:

1. If an allocation $x_i(S)$ and prices $p_i(S)$ are in competitive equilibrium for the given valuations $v_i(S)$), this allocation is the optimal allocation.
2. For the optimal allocation $x_i(S)$ there always exist discriminatory non-linear competitive equilibrium prices $p_i(S)$. This is not always true for linear and anonymous non-linear prices.

3 Primal-dual algorithms

One way of reducing some of the computational burden in solving the winner determination problem is to set up a fictitious market that will determine an allocation and prices in a decentralized way. In the iterative approach, there are multiple rounds of bidding and allocation and the problem is solved in an iterative and incremental way. Iterative combinatorial auctions are attractive to bidders because they learn about their rivals' valuations through the bidding process, which could help them to adjust their own bids.

There is a connection between efficient auctions for many items, and duality theory. The Vickrey auction can be taken as an efficient pricing equilibrium, which corresponds to the optimal solution of a particular linear programming problem and its dual. The simplex algorithm can be taken as static approach to determining the Vickrey outcome. Alternatively, the primal-dual algorithm can be taken as a decentralized and dynamic method to determine the pricing equilibrium. A primal-dual algorithm usually maintains a feasible dual solution and tries to compute a primal solution that is both feasible and satisfies the complementary slackness conditions. If such a solution is found, the algorithm terminates. Otherwise the dual solution is updated towards optimality and the algorithm continues with the next iteration. The fundamental work (Bikhchandani, Ostroy, 2002) demonstrates a strong interrelationship between the iterative auctions and the primal-dual linear programming algorithms. A primal-dual linear programming algorithm can be interpreted as an auction where the dual variables represent item prices. The algorithm maintains a feasible allocation and a price set, and it terminates as the efficient allocation and competitive equilibrium prices are found.

For the winner determination problem we will formulate the LP relaxation and its dual. Consider the LP relaxation of the winner determination problem (1):

$$\sum_{i=1}^{n} \sum_{S \subseteq M} v_i(S) x_i(S) \rightarrow \quad \max$$

subject to

$$\begin{aligned} \sum_{S \subseteq M} x_i(S) \leq 1,\ \forall i,\ i = 1, 2, \ldots, n, \\ \sum_{i=1}^{n} \sum_{S \subseteq M} x_i(S) \leq 1,\ \forall j \in M, \\ x_i(S) \geq 0, \forall S \subseteq M,\ \forall i,\ i = 1, 2, \ldots, n. \end{aligned} \tag{2}$$

The corresponding dual to problem (2)

$$\sum_{i=1}^{n} p(i) + \sum_{j \in S} p(j) \rightarrow \quad \min$$

subject to

$$p(i) + \sum_{j \in S} p(j) \quad \geq \quad v_i(S) \ \forall i, S, \tag{3}$$

$$p(i), p(j) \quad \geq \quad 0, \ \forall i, j, \tag{4}$$

The dual variables $p(j)$ can be interpreted as anonymous linear prices of items, the term $\sum_{j \in S} p(j)$ is then the price of the bundle S and $p(i) = \max_S [v_i(S) - \sum_{j \in S} p(j)]$ is the maximal utility for the bidder i at the prices $p(j)$.

Following two important properties can be proved for the problems (2) and (3):

1. The complementary-slackness conditions are satisfied if and only if the current allocation (primal solution) and the prices (dual solution) are in competitive equilibrium.
2. The formulation (2)–(3) is weak. For the optimal allocation there no always exist anonymous linear competitive equilibrium prices.

The formulation (2)–(3) can be strengthened. Additional variables $y(K)$ are introduced for each personalized item set partition k. In the integer version only one of the variables $y(K)$ can be set to 1, which means that the allocation xi(S) is compatible with the personalized partition k. The problem relaxation and its dual are then formulated as follows

$$\sum_{i=1}^{n} \sum_{S \subseteq M} v_i(S) \, x_i(S) \ \rightarrow \quad \max$$

subject to

$$\sum_{S \subseteq M} x_i(S) \leq 1, \ \forall i,$$

$$x_i(S) \leq \sum_{k:[i,S] \in k} y(k), \ \forall, S, \tag{5}$$

$$\sum_{k} y(K) \leq 1,$$

$$x_i(S) \geq 0, y(K) \geq 0, \forall i, S, k.$$

$$\sum_{i=1}^{n} p(i) + \pi \ \rightarrow \quad \min$$

subject to

$$
\begin{aligned}
p(i) + p_i(S) &\geq v_i(S), \ \forall i, S, && (6)\\
\pi - \sum_{[i,S]\in k} p_i(S) &\geq 0, \ \forall k, && (7)\\
p(i), \ p_i(S), \ \pi &\geq 0, \ \forall i, S. && (8)
\end{aligned}
$$

The dual variables $p_i(S)$ for each bundle S and each bidder i can be interpreted as discriminatory non-linear prices, $p(i) = \max\limits_S [v_i(S) - p_i(S)]$ is the maximal utility for the bidder i at the prices $p_i(S)$ and $\pi = \max\limits_k \sum_{[i,S]\in k} p_i(S)$ is the maximal utility of the auctioneer at the prices $p_i(S)$.

Following two important properties can be proved for the problems (5) and (6):

1. The complementary-slackness conditions are satisfied if and only if the current allocation (primal solution) and the prices (dual solution) are in competitive equilibrium.
2. The formulation (5)–(6) is strong. It proves the existence of discriminatory non-linear competitive equilibrium prices for the optimal allocation.

4 Auction formats

Several auction formats based on the primal-dual approach have been proposed in the literature. Though these auctions differ in several aspects, the general scheme can be outlined as follows:

1. Choose minimal initial prices.
2. Announce current prices and collect bids. Bids have to be higher or equal than the prices.
3. Compute the current dual solution by interpreting the prices as dual variables. Try to find a feasible allocation, an integer primal solution that satisfies the stopping rule. If such solution is found, stop and use it as the final allocation. Otherwise update prices and go back to 2.

Concrete auction formats based on this scheme can be implemented in different ways. The most important design choices are the following:

- bid structure,
- pricing scheme,
- price update rule,

- bid validity,
- feedback,
- way of computing a feasible primal solution in each iteration,
- stopping rule.

Two concrete iterative combinatorial auction formats (Combinatorial Clock Auction and iBundle) are compared.

4.1 Combinatorial Clock Auction (CC)

The Combinatorial Clock auction proposed in (Porter et al., 2003) can be seen as some kind of a primal-dual auction algorithm. It utilizes anonymous linear prices which are called item clock prices. In each round bidders submit which packages they would purchase at the current prices. If overdemand holds for at least one item the price clock "ticks" for all overdemanded items (the item prices are increased by a fixed price increment), and the auction goes to the next iteration. If there is no excess demand and no excess supply, the items are allocated corresponding to the last iteration bids and the auction terminates. If there is no excess demand but there is excess supply (all active bidders on some item did not resubmit their bids in the last iteration), the auction solves the winner determination problem considering all bids submitted during the whole auction. If the computed allocation does not displace any active last iteration bids the auction terminates with this allocation, otherwise the prices of the respective items are increased and the auction continues. The key design features of the CC auction are summarized in the Table 1.

The authors do not provide any theoretical analysis of the auction efficiency, they only claim that the auction is "simply a greedy algorithm to discover pseudo-dual upper-bound prices: The lowest prices at which everyone who submitted a bid is definitively declared a winner". The authors also report very good experimental efficiency results. The advantages of the CC auction are its cognitive, computational and communicative simplicity. However this can result in efficiency losses. One kind of inefficiency can be due to the exposure problem, since OR-bidding is used.

4.2 iBundle

The iBundle (Parkes, 2001) is an ascending-price combinatorial auction, which follows quite directly from the primal-dual algorithm for the combinatorial allocation problem. The best-response information provided by agents has a natural interpretation as a utility-maximizing bidding strategy for a myopic agent, i.e. an agent that takes the current prices as fixed and does not look beyond the current round. The iBundle is the iterative auction to provably terminate with an efficient allocation for a reasonable agent bidding strategy, without any restrictions on agents' valuation functions. The main design decisions in iBundle are:

- Exclusive-or bids over bundles of items.

	CC	**iBundle**
Bid structure	bundles, OR	bundles, XOR
Pricing scheme	anonymous linear	discriminatory non-linear
Price used as	bid price	minimal bid price or bid price
Price updates	increase on all overdemanded items	increase on all overdemanded items
Bid validity	whole auction	current iteration previously winning
Feedback	prices	prices own winning bids
Stopping rule	no overdemand no last iteration bid is displaced	no overdemand or no new bids

Table 1: Comparisons of iterative combinatorial auctions (CC and iBundle)

- A simple price-update rule with minimal consistency requirements on prices across different bundles.
- A dynamic method to determine when non-anonymous prices are required to price the efficient allocation in competitive equilibrium.

Myopic best-response need not be an agent's optimal sequential strategy in iBundle, and the basic auction design is not strategy-proof like the Generalized Vickrey Auction (GVA). The auctioneer in iBundle solves one winner determination problem in each round, compared to one for each agent in the final allocation in the GVA. Although the problem of computing a provisional allocation in each round remains NP-hard the problem instances in iBundle are much smaller than in the GVA because the agents only bid for a small subset of bundles in each round. The key design features of the iBundle auction are summarized in the Table 1.

Experimental results confirm the efficiency of iBundle across a set of combinatorial allocation problems from the literature. The auction computes efficient solutions, even with quite large bid increments. Results also demonstrate that non-anonymous prices are only important above 99% allocative efficiency.

5 Conclusions

The paper is devoted to analyzing an iterative approach to solving combinatorial auctions. Iterative combinatorial auctions are a promising subject for research and for practical exploitations. The key challenge in the iterative combinatorial auctions design is to provide information feedback to the bidders. There is a connection between efficient auctions for many items, and duality theory. The primal-dual algorithm can be taken as a decentralized and dynamic method to determine the pricing equilibrium. Several auction formats based on the primal-dual approach have been proposed in the literature. Comparisons of iterative combinatorial auctions give us an opportunity to design modifications and new versions of these auctions.

Acknowledgements

The research project was supported by Grant No. 402/07/0166 from the Grant Agency of Czech Republic "Combinatorial Auctions - Modeling and Analysis".

References

[1] Bikhchandani, S., Ostroy, J. M. (2002): The package assignment model. *Journal of Economic Theory*, 107(2), 377–406.

[2] Cramton, P., Shoham, Y., Steinberg, R. (eds.) (2006): *Combinatorial Auctions*. MIT Press, Cambridge.

[3] Milgrom, P. R. (1987): Auction theory. In Bewley, T.(ed.) *Advances in Economic Theory: Fifth World Congress*, Cambridge University Press, Cambridge.

[4] Parkes, D. C. (2001): *Iterative Combinatorial Auctions: Achieving Economic and Computational Efficiency*. PhD thesis, University of Pennsylvania.

[5] Pikovsky, A, Bichler, M. (2005): Information Feedback in Iterative Combinatorial Auctions. In *Conference proceedings Wirtschaftsinformatik*, Bamberg, Springer.

[6] Porter D., Rassenti, S., Roopnarine, A., Smith,V. (2003). Combinatorial auction design. Proceedings of the National Academy of Sciences of the United States of America (PNAS).

Wojciech Gamrot

Department of Statistics
University of Economics in Katowice
ul. 1-Maja 50, 40-287 Katowice

e-mail: gamrot@ae.katowice.pl

Representative Sample Selection via Random Search with Application to Surveying Communication Lines

Abstract

The Municipal Transport Union of the Upper-Silesian Industrial District (KZK-GOP) organizes the public transport system in 23 municipalities of the Silesian Region, covering the area of 1400 square kilometers with over two million inhabitants. Its main sources of financing include the fare income as well as subsidies from member municipalities. Since the year 2008 the amount of subsidy from each member depends on the profitability of communication lines that cross its territory. To calculate the profitability passenger loads on these lines must be assessed by means of a sample survey. Most lines in the system cross the area of several municipalities. Moreover, specific requirements concerning the composition of the line sample must be met to ensure the representativeness of estimates and preserve the credibility of the survey. This makes the use of simple random sampling designs problematic. In this paper a specialized random search algorithm for selecting representative samples is presented. It incorporates random sampling as well as purposive selection of sample units in a way that makes the evaluation of sampling errors possible. The procedure attempts to minimize them.

Keywords: *Sampling, variance, optimization, random search*

1 Introduction

The Municipal Transport Union of the Upper-Silesian Industrial District (KZK-GOP) was founded in the year 1992 by 11 municipalities of the Katowice agglomeration. Since then several new municipalities have joined the Union including neighbouring large cities of Gliwice and Bytom to bring the total number of members to 23. The mission of the Union is the effective rendering of transport services for inhabitants of the agglomeration, while applying market mechanisms and in accordance with transport policy executed by local governments. Accordingly, its statutory task is to satisfy the needs of member municipalities within the range of local mass transport. To offer affordable public transportation services within the area of member municipalities as well as connections to locations outside the Union area the Union signs contracts with private, municipal and state-owned transport companies that operate over 290 bus lines and over 30 tram lines. Transport companies are fully paid for conducted services. The payments are based on negotiated rates for elementary vehicle kilometre. The difference between the income (from fares and other sources) and total costs is covered by subsidies from member municipalities. The Union statute regulates the division of subsidy payments between member municipalities. Since the year 2008 the amount of subsidy from each municipality depends on the profitability of communication lines within its territory. Hence, the subsidy also indirectly depends on annual passenger loads on these lines and the need appears to assess them. Due to high cost and low reliability of automatic monitoring solutions and the infeasibility of a complete count these quantities must be estimated on the basis of a sample survey. The survey has multiple purposes and the sample has to be selected in a way that guarantees a sufficient representation of each municipality. Most lines in the system cross the area of several municipalities which complicates the task of controlling sample composition in a way that makes the computation of estimator variances possible. The aim of this paper is to illustrate the prospects of applying random search methods to select representative samples, while retaining the possibility of evaluating precision of estimates. In the following paragraphs a general estimation procedure based on well-known stratified sampling design is presented. Then one possible interpretation of sample representativeness is considered and the general procedure is applied to satisfy this requirement by using a fixed, exhaustively sampled stratum. Finally the random search algorithm dedicated to optimally construct this fixed strata is proposed.

2 Sampling and estimation

Let U denote the population of N communication lines to be sampled. Each line crosses the area of at least one of G municipalities (currently $G = 23$ for KZK-GOP). Denote the annual number of passengers carried by the i-th communication line (its passenger load) by X_i. It is a random variable with unknown expectation:

$$E(X_i) = \mu_i. \tag{1}$$

Define the coverage ratio of the g-th municipality by the i-th line as:

$$r_{ig} = \frac{w_{ig}}{w_i} \tag{2}$$

where wig represents known annual transportation work (measured in vehicle-kilometers traveled) of the i-th line within the territory of g-th municipality, while wi is the total annual transportation work of this line irrespective of municipalities' borders. Hence we have $r_i g = 0$ when i-th line does not cross the area of the g-th municipality, $r_i g = 1$ when the whole line is contained this area and $0 < r_i g < 1$ when only some part of the line crosses the area. Define the annual passenger load of the i-th line within the g-th municipality as:

$$X_{ig} = r_{ig} \cdot X_i \tag{3}$$

Obviously in reality this does not have to be met exactly, but we may expect that the deviations from true passenger loads will cancel themselves for large number of lines and that $X_i g$ is a good approximation of the true passenger load. The expectation of $X_i g$ is:

$$\mu_{ig} = r_i g \cdot \mu_i \tag{4}$$

Hence the total annual number of passengers transported (by all lines) in the g-th municipality is:

$$X_g = \sum_{i=1}^{N} X_{ig} \tag{5}$$

The objective of the survey is to estimate for any g-th municipality the expectation of X_g:

$$\mu_g = \sum_{i=1}^{N} \mu_{ig} \tag{6}$$

For convenience we shall divide the problem of surveying communication lines into two stages: the selection of a line sample and the assessment of passenger loads on individual lines. A thorough discussion of the second stage exceeds the scope of this paper. Hence, we treat estimation results obtained for individual lines as given and concentrate on the first stage: the line sample selection. Assume that an unbiased estimate $\hat{\mu}_i$ of μ_i is obtained when i-th line is examined. Moreover, assume that $\hat{\mu}_i$ and $\hat{\mu}_j$ are independent for $i \neq j \in U$ and that standard deviations of these estimates are at least roughly proportional to μ_i and hence $D(\hat{\mu}_i) = \alpha\mu_i$ for $i \in U$. The constant α may depend on the sampling design and sample size used for selection of days of the year, courses of the line and the stations or vehicles to be surveyed when investigating individual lines. It may also depend on the scope of various non-sampling errors (especially measurement error) as well as the choice of estimators and modeling assumptions. It is assumed to be known from the sources external to the survey, e.g. assessed in similar surveys or in a controlled experiment. Consequently we may compute unbiased estimates of $\hat{\mu}_{ig}$'s in the form:

$$\hat{\mu}_{ig} = r_{ig} \cdot \hat{\mu}_i \tag{7}$$

For line selection let us consider the well-known stratified sampling design [3]. The population is divided into H disjoint subsets (strata) $U_1, \ldots, U_H$ of sizes $N_1, \ldots, N_H$. Then H independent samples $s_1, \ldots, s_H$ of sizes $n_1, \ldots, n_H$ are respectively drawn from strata using simple random sampling without replacement. These constitute the full sample $s = s_1 \cup \ldots \cup s_H$. Estimates $\hat{\mu}_i$ of expected passenger loads are computed for each $i \in s$. Then estimates $\hat{\mu}_{ig}$ are computed for each $i \in s$ and $g = 1 \ldots G$. Consequently, the statistic:

$$\hat{\mu}_g = \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{i=1}^{n_h} \hat{\mu}_{ig} \tag{8}$$

is an unbiased estimator of μ_g and its variance is equal to:

$$D^2(\hat{\mu}_g) = \alpha^2 \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{i=1}^{N_h} r_{ig}^2 \mu_i^2 + \sum_{h=1}^{H} N_h^2 \frac{N_h - n_h}{N_h n_h} S_h^2(\mu_{ig}) \tag{9}$$

where

$$S_h^2(\mu_{ig}) = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(\mu_{ig} - \frac{1}{N_h} \sum_{i=1}^{N_h} \mu_{ig} \right)^2 \tag{10}$$

The first term in (9) corresponds to errors associated with estimating passenger loads on individual lines and the second term reflects the error due to sampling the population of lines.

3 Sample representativeness and optimization.

The results of the survey serve as a justification for payments. Hence it is important to make them as indisputable as possible and special care should be taken to guarantee at least minimum representation of each municipality in the sample so that direct estimates may be obtained via the statistic (8). It appears reasonable to associate with each g-th municipality some minimum number mg of lines to be included in the sample. The procedure should guarantee that a number of sampled units covering at least partially the g-th municipality is not lower than mg. However, the use of strata corresponding to specific municipalities is problematic as most lines cross more than one municipality. This contradicts the independence of sampling and greatly complicates the calculation of line inclusion probabilities as well as estimator properties. We consider another method of stratification. Let the population be divided into two strata U_1 and U_2 such that U_1 contains the lines that provide the minimum coverage of all municipalities and U_2 contains remaining lines. The sample of a pre-determined size n is drawn in the following way: all units from U_1 are included in the sample (hence $s_1 = U_1$ and $n_1 = N_1$) and remaining $n - n_1$ units are drawn from U_2 using simple random sampling without replacement [6]. Such a procedure is a special case of stratified sampling discussed above. It retains the unbiasedness of estimates, while guaranteeing the minimum coverage of each municipality. However, usually the set U_1 may be chosen in many ways and this choice determines the variances for all municipalities via the formula (9). Hence one may search for the set U_1 minimizing these variances. One potential difficulty with this approach comes from the fact that variances $V(\hat{\mu}_1) \dots V(\hat{\mu}_G)$ depend on unknown expectations $\mu_1 \dots \mu_H$. However, in the case of the KZK system their reasonable approximations $\mu_1^*, \dots, \mu_N^*$ are available from previous surveys (partially imputed) and may be used instead.

Let $M = [m_1, \dots, m_G]$ denote the vector of minimum sample line numbers corresponding to all municipalities with $m_g > 0$ for $g = 1 \dots G$. Let $B = [b_1, \dots, b_N]$ be a vector of binary indicators such that $b_i = 1$ when $i \in U_1$ and $b_i = 0$ otherwise for $i \in U$. Let us also define matrix:

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1G} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NG} \end{bmatrix} \tag{11}$$

where $a_i g = 1$ when i-th line crosses the area of the g-th municipality and $a_i g = 0$ otherwise, for $i = 1 \ldots N,\ g = 1 \ldots G$. Denote the i-th column of A by A_g. The sufficient representation of all municipalities is guaranteed when $B \cdot A_g \geq b_g$ for $g = 1 \ldots G$ or simply when

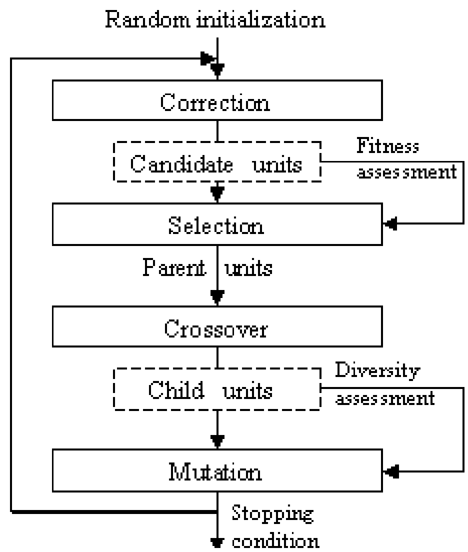
$$B \cdot A \geq M \tag{12}$$

We now formulate the following optimization problem:

$$\begin{cases} B \cdot A \geq M \\ Q(B) = \max\limits_{g=1\ldots G} V(\hat{\mu}_g) \rightarrow \min \end{cases} \tag{13}$$

To find the solution a genetic algorithm mimicking the natural evolution process [4],[2] is adopted. It processes vectors of binary numbers (genes) that naturally represent the possible (candidate) solutions $B_1, \ldots, B_K$ to the problem (13) called units (or chromosomes). The algorithm starts with assigning random initial values to the population of units. Then it sequentially executes three steps: selection, crossover and mutation [1]. In the selection phase the fitness $Q(B)$ of each unit is evaluated. Then a new population of units is created by drawing units from the original population in such a way that units with relatively high fitness have high probability of selection while units with low $Q(B)$ have lesser chances of survival. The new population replaces the old one. In the crossover phase units are randomly assigned into pairs and some of their genes are randomly interchanged. In the mutation phase some genes in the population are replaced with randomly generated values, to introduce new information into the population. Mutation happens with some pre-determined, low probability. These phases are repeated sequentially until some termination condition is satisfied. The unit with highest fitness represents the solution to the optimization problem.

We modify the original algorithm to facilitate the handling of the limit. In the original algorithm most units obtained by random crossing do not satisfy it. To compensate for this unwelcome effect we introduce another phase of the process—a “Correction Phase” which is always executed before the fitness assessment. During this phase the procedure tries to alter all units that do not satisfy the limit by randomly setting the elements of the candidate vector to ones (adding lines to U_1) until the limit is satisfied. Vectors that satisfy the limit but contain more lines than is needed are also altered in this phase by setting randomly chosen bits to zero (removing lines from U_1) until the set of lines cannot be further reduced. This is done to avoid the domination of candidates containing unreasonably large number of lines. At selection phase the Bernoulli sampling scheme [5] is used with the probability of selection proportional to $Q(B)$. Crossover phase is carried out by randomly arranging all units in pairs and then randomly deciding for each pair of corresponding genes (bits) if they should be interchanged or not (with the swapping probability equal to 0.5). In the mutation phase the value of some genes is randomly negated. To avoid stalling in some local minimum, the probability of mutation reflects the uniformity of candidate solutions. It is proportional to the ratio: k/K where k is the greatest number of duplicate chromosomes (units) in population and K is its size. The best solution in each iteration is recorded. It is stored separately and does not influence the evolution process. In subsequent iterations the corresponding chromosome may even disappear from the population. The procedure stops when some pre-determined number of iterations is completed, and the best one of stored solutions is reported as the final outcome. A block diagram of the modified algorithm is shown on the graph 1.



Graph 1: Block diagram of the algorithm

4 Conclusions

Thanks to proposed modifications the algorithm produces solutions that are representative in the sense discussed above. However, it does not guarantee the global optimality of obtained solutions with respect to criterion function. This property was sacrificed in order to improve flexibility and applicability to more complicated problems—an advantage especially valuable in the case of further modifications of estimation procedure such as introduction of additional strata, the use of unequal probability sampling designs or regression estimators. Anyway, the genetic algorithm seems to be well suited for finding in a reasonable time parameter vectors providing attractive values of a criterion function. The proposed approach may be used with various methods of assessing expected passenger loads on individual lines. It is also possible to use varying methodologies for individual lines—the extension of proposed procedure for such a situation is straightforward.

References

[1] DAVIS, L. *Handbook of Genetic Algorithms.* Van Nostrand, New York, 1991.

[2] GOLDBERG, D.E. *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison-Wesley, 1989.

[3] HEDAYAT, A.S., SINHA, B.K., *Design and Inference in Finite Population Sampling,* Wiley, New York 1991.

[4] HOLLAND, J.H. *Adaptation in natural and artificial systems,* Ann Arbor: Michigan University Press, 1975.

[5] SÄRNDAL, C.E., SWENSSON, B., WRETMAN, J. *Model Assisted Survey Sampling,* Springer, New York , 1992.

[6] THOMPSON, M.E., *Theory of Sample Surveys*, Chapman & Hall 1992.

Alicja Ganczarek

Department of Statistics
University of Economics in Katowice
ul. 1-Maja 50, 40-287 Katowice

e-mail: alicjaganczarek@wp.pl

Choosing Between the Skewed Distribution and the Skewed Model in General Autoregressive Conditional Heteroscedasticity[1]

Abstract

The Polish Energy Exchange was established in July 2000. The Day Ahead Market (DAM) was the first market to be established on the Polish Energy Exchange. This whole-day market consists of the twenty-four separate, independent markets where participants may freely buy and sell electricity. The advantage of the Exchange is that all participants of the market can buy and sell electric energy, irrespective of whether they are producers or receivers. This is a market of heightened volatility in level of prices and demand, so producers and receivers are forced to protect themselves against losses. Financial decisions always include a number of risks including the changes of prices. In order to estimate the future risk we must measure it. Good choice of volatility process is very important for risk measurement. The aim of this paper is to compare results of estimating GARCH models with skewed Student's t-distribution to results of estimating skewed GARCH models with symmetric Student's t-distribution. GARCH models include stationary models such as: GARCH, , EGARCH, APARCH and nonstationary models such as: IGARCH, FIGARCH, FIEGARCH, FIAPARCH, HYGARCH. Their parameters are estimated on the basis of the hourly logarithmic rates of return on the Day Ahead Market (DAM) of the Polish Power Exchange. The results of estimation are compared by information criteria such as the Akaike, Schwarz and Hannan-Quinn. Practical implications of these results in forecasting on DAM are discussed.

Keywords: *symmetric and skewed Student's t-distribution, stationary and nonstationary GARCH models, information criteria*

1 Introduction

A lot of empirical results on electric energy markets and on financial markets indicate, that empirical distributions have a non-normal distribution. They have fat tails, like a Student's t- distribution. Also very often empirical distributions are skewed and leptokurtic. Moreover first and second moments of these time series depend on earlier observations and rates of return are negatively correlated with prices due to leverage effects. Hence, in this article the problem

[1] Research supported by grant N111 003 32/0262

of choosing the proper model describing the distribution of rates of return is considered. Two alternative approaches involving the use of skewed distribution as well as the skewed model in general autoregressive conditional heteroscedasticity were discussed based on empirical time series quoted on the Day Ahead Market (DAM) from July to December 2007.

1.1 The Day Ahead Market

The Day Ahead Market (DAM) was the first market which was established on the Polish Energy Exchange. This market is a whole-day market, which consists of the twenty-four separate, independent markets. The contracts for energy delivery are traded in all days of the week. The electric energy cannot be stored. It is delivered once there is demand for it. So this is a market of highly volatile prices dependent on supply and demand. This greatly increases the market risk forcing producers and receivers to seek protection against losses. To properly assess the most likely direction of price changes, an accurate model is needed.

2 Time series on DAM and their distributions

There are a lot of different measures of risk. We can divide them into three groups: measures of volatility, measures of sensitivity and measures of downside risk. Most of them require the knowledge of the distribution of rates of return. We will now consider the approaches to estimate it.

The distribution of time series of hourly logarithmic rates of return of electric energy prices noted on DAM from July to December 2007 is presented on Figure 1. It is positively skewed (skewed = 0.6107, Table 2.) and leptokurtic (kurtosis = 2.5078,.Table 2).

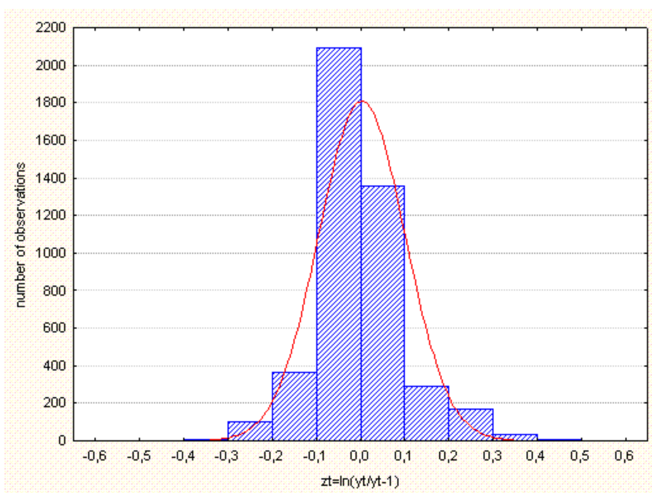


Figure 1: Distributions of time series of hourly logarithmic rates of return of electric energy prices noted on DAM from July to December 2007

The electric energy volumes and prices are characterized by daily, weekly and yearly seasonal peaks and lows. Autocorrelation Coefficient Function (ACF) and Partial Autocorrelation

Coefficient Function (PACF) with lag length equal to 48 are presented on Graph 2. The values of ACF and PACF clearly show daily seasonal autocorrelation. So in the next step of time series analysis on DAM Seasonal Autoregressive Integrated Moving Average (SARIMA) model [5] was used with daily seasonal lag 24 and another one with both daily and weekly seasonal lag of 168 hours (daily seasonal lag of 24 hours and weekly seasonal lag of 7 days). Yearly seasonality was ignored because the data set is too short.

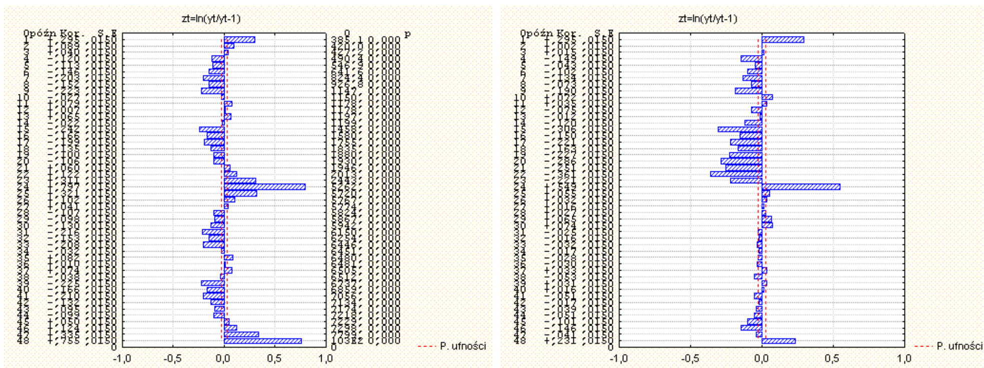


Figure 2: The ACF and the PACF of time series of hourly logarithmic rates of return of electric energy prices noted on DAM from July to December 2007

The parameters of these two models are presented in the Table 1. Maximum likelihood methods are used to estimate this parameter. According to t-Student's statistics all parameters are significant. The distributions' parameters of SARIMA models residuals are presented in Table 2. The distributions are negatively skewed and leptokurtic. Residuals are less volatile than original data but they are more leptokurtic. Besides, there are negatively skewed. And the extent of skewed is greater for the first model. The plots of residuals of this two models are presented on Graph 3. We can observe a variance clustering effect (Graph 3) and autocorrelation of square of residuals (Graph 4), which proves the heteroscedastity of variance. So in risk measurement we can not forget about autocorrelation of fist and second moments. In the next section the models with autoregressive variance are discussed.

3 General autoregressive conditional heteroscedasticity

GARCH models include stationary models such as: GARCH [8, 9], EGARCH [13] APARCH [7] and nonstationary models such as: IGARCH [9], FIGARCH [2], FIEGARCH [4], FIAPARCH [2], HYGARCH [6]. Their parameters are estimated by maximum likelihood method on the basis of time series of SARIMA's residuals $\varepsilon_t = \xi_t \sigma_t$, where ξ_t is a white noise with standardized symmetric or skewed Student's t-distribution. We will now compare results of estimating GARCH models with skewed Student's t-distribution to results of estimating skewed GARCH models with symmetric Student's t-distribution.

Model	Parameters		t	p
SARIMA $(1,0,1)(1,1,1)_{24}$	p(1)	0.7480	34.0309	0.0000
	q(1)	0.9492	73.6522	0.0000
	Ps(1)	0.1041	5.7528	0.0000
	Qs(1)	0.8351	96.4843	0.0000
SARIMA $(1,0,1)(1,1,1)_{168}$	p(1)	0.6905	36.3948	0.0000
	q(1)	0.9305	88.2569	0.0000
	Ps(1)	0.6871	25.4714	0.0000
	Qs(1)	0.4312	13.3018	0.0000

Table 1: Parameters of SARIMA

Parameters	$zt = \ln(yt/yt-1)$	SARIMA $(1,0,1)(1,1,1)_{24}$	SARIMA $(1,0,1)(1,1,1)_{168}$
T	4415	4391	4247
Mean	0.0000	–0.0001	0.0000
Min	–0.4308	–0.4127	–0.4148
Max	0.4463	0.3366	0.3388
standard deviation	0.0974	0.0475	0.0469
skeweness	**0.6107**	**–0.6091**	**–0.2320**
kurtosis	2.5078	10.5100	10.3758

Table 2: Parameters of rates of return and residuals of SARIMA

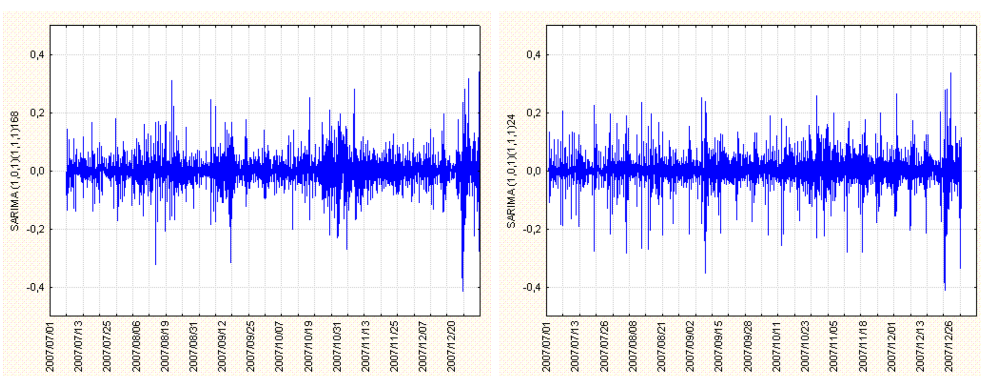


Figure 3: Plots of residuals of SARIMA $(1,0,1)(1,1,1)_{24}$ *and SARIMA* $(1,0,1)(1,1,1)_{168}$

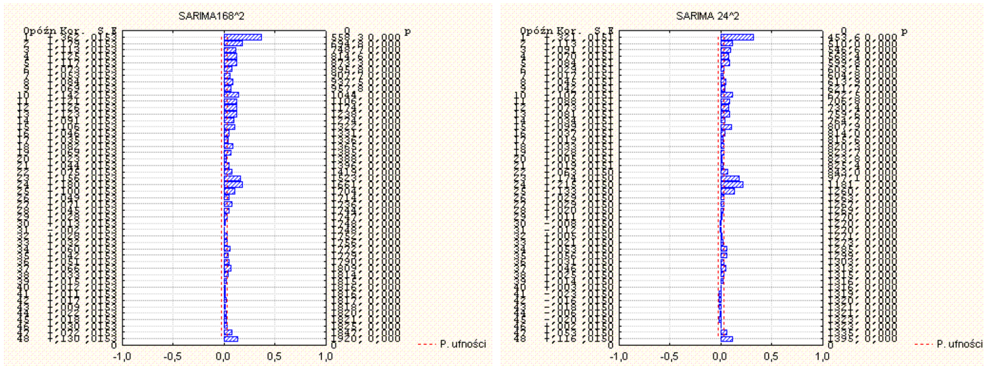


Figure 4: ACF of square of residuals of SARIMA $(1,0,1)(1,1,1)_{24}$ *and SARIMA* $(1,0,1)(1,1,1)_{168}$

A lot of different model selection criteria are proposed for selecting an optimal model. The most of them involve minimizing some loss function. One of the most popular models is Akaike's [1] information criterion, which takes the form:

$$AIC = -2\frac{\ln L}{T} + 2\frac{k}{T}. \tag{1}$$

Schwarz [14] developed a consistent criterion based on Bayesian arguments:

$$SC = -2\frac{\ln L}{T} + 2\frac{\ln(k)}{T}. \tag{2}$$

Hannan and Quinn [11] proposed the consistent criterion for autoregressive models based on the law of iterated logarithm:

$$HQ = -2\frac{\ln L}{T} + 2\frac{k\ln(\ln(T))}{T}. \tag{3}$$

In formulas (1-3) T is the sample size, k is number of the coefficients, L – maximum likelihood. The Akaike's (1) and Schwarz's (2) criterions are most popular and very often used. The comparison of various alternative criteria is given by Mitchell and McKenzie [12].

The values of the information criteria are presented in Table 3 and 4. For every model the smallest values of information criteria are underlined. For SARIMA $(1,0,1)(1,1,1)_{24}$ – GARCH (table 3) and stationary basic ARCH model symmetric Student's t-distribution is a little better then skewed one. But asymmetric APARCH model with symmetric distribution provides worse performance than EGARCH model with skewed distribution. In the case of non-stationary models all models with skewed distribution are better than symmetric ones with the notable exception of IGARCH. And the best of all models for this time series is SARIMA $(1,0,1)(1,1,1)_{24}$ – HYGARCH(1,1) with skewed Student's t-distribution (Table 3 – dark cell).

Models		Symmetric Student's t-distribution			Skewed Student's t -distribution		
		AIC	SC	HQ	AIC	S.C.	HQ
stationary	ARCH	**<u>4.2073</u>**	**<u>4.2102</u>**	**<u>4.2084</u>**	4.2074	4.2118	4.2090
	GARCH	2.2564	**<u>2.2607</u>**	2.2579	**<u>2.2563</u>**	2.2622	**<u>2.2584</u>**
	EGARCH	2.2590	2.2663	2.2616	**<u>2.1612</u>**	**<u>2.1699</u>**	**<u>2.2584</u>**
	APARCH	**<u>2.2324</u>**	**<u>2.2397</u>**	**<u>2.2349</u>**	2.2324	2.2411	2.2355
non-stationary	IGARCH	**<u>2.2750</u>**	**<u>2.2779</u>**	**<u>2.2761</u>**	2.2752	2.2796	2.2767
	FIGARCH	2.1715	2.1736	2.1703	**<u>2.1702</u>**	**<u>2.1723</u>**	**<u>2.1691</u>**
	FIEGARCH	2.2595	2.2626	2.2576	**<u>2.1412</u>**	**<u>2.1448</u>**	**<u>2.1389</u>**
	FIAPARCH	2.1540	2.1570	2.1521	**<u>2.1532</u>**	**<u>2.1568</u>**	**<u>2.1509</u>**
	HYGARCH	2.1353	2.1379	2.1338	**<u>2.1344</u>**	**<u>2.1375</u>**	**<u>2.1325</u>**

Table 3: Information criteria of SARIMA $(1,0,1)(1,1,1)_{24}$ *– GARCH*

Concerning SARIMA $(1,0,1)(1,1,1)_{168}$ – GARCH (table 4) we conclude that for stationary basic GARCH model, symmetric Student's t-distribution is better than the skewed one. The EGARCH model with skewed distribution should be preferred to asymmetric APARCH model with symmetric distribution.

Analysis of results obtained for non-stationary SARIMA $(1,0,1)(1,1,1)_{168}$ – GARCH leads to different results than the previous example. The symmetric distribution seems to be better than skewed one in terms of AIC and SC. But based on Hannan and Quinn's criterion for this time series SARIMA $(1,0,1)(1,1,1)_{168}$ –FIEGARCH (1,1) with skewed Student's t-distribution is the best of all models (Table 4 – dark cell).

4 Analysis of residuals

In this section we provided additional analysis of residuals for models considered above. The Kolmogorow satistic was used to verify if the residuals distribution for particular models differ significantly. In table 5 the values of test statistics are presented. The statistics value which are higher than critical value on significance level 0,05 are marked (on significance level 0,05 the critical value of Kołmogorow's statistics equal 1,358).

The results for time lag 24 show that only the SARIMA24-APARCH model with symmetric Student's t-distribution differs significantly from others with respect to distribution of residuals. Differences among other models are not significant.

The results for time lag 168 show no significant differences between any models, probably as a result of rather weak asymmetry

4.1 Conclusion

Based on two time series of SARIMA's residuals, we can observe, that for highly skewed time series GARCH models with skewed distribution are much better than asymmetric GARCH

Models		Symmetric Student's t-distribution			Skewed Student's t-distribution		
		AIC	SC	HQ	AIC	S.C.	HQ
stationary	ARCH	**4.2498**	**4.2528**	**4.2508**	4.2503	4.2547	4.2518
	GARCH	**2.2936**	**2.2981**	**2.2952**	2.2940	2.3000	2.2962
	EGARCH	2.3420	2.3495	2.3447	**2.2353**	**2.2443**	**2.2385**
	APARCH	**2.2867**	**2.2942**	**2.2894**	2.2872	2.2961	2.2903
non-stationary	IGARCH	**2.3051**	**2.3081**	**2.3062**	2.3056	2.3101	2.3072
	FIGARCH	2.2362	**2.2383**	2.2350	**2.2361**	**2.2383**	**2.2349**
	FIEGARCH	2.2419	2.2451	2.2400	**2.2193**	**2.2230**	**2.2170**
	FIAPARCH	**2.2282**	**2.2314**	**2.2262**	2.2286	2.2323	2.2262
	HYGARCH	**2.2189**	**2.2216**	**2.2173**	2.2194	2.2225	2.2174

Table 4: Information criteria of SARIMA $(1,0,1)(1,1,1)_{168}$ *– GARCH*

Models	SARIMA APARCH symmetric St	SARIMA EGARCH skewed St	SARIMA HYGARCH symmetric St	SARIMA HYGARCH skewed St
SARIMA24- APARCH symmetric St	-	**1,8888**	**1,3872**	**1,4299**
SARIMA24-EGARCH skewed St	**1,8888**	-	0,8430	0,8003
SARIMA24-HYGARCH symmetric St	**1,3872**	0,8430	-	0,0427
SARIMA24-HYGARCH skewed St	**1,4299**	0,8003	0,0427	-
SARIMA168- APARCH symmetric St	-	0,6402	0,4232	0,4232
SARIMA168-EGARCH skewed St	0,6402	-	0,2279	0,2170
SARIMA168-HYGARCH symmetric St	0,4232	0,2279	-	0,0109
SARIMA168-HYGARCH skewed St	0,4232	0,2170	0,0109	-

Table 5: Kolmogorow's statistics

models.. When the skewness is not significant, GARCH models incorporating skewness in the model and not in the distribution seem to be more appropriate. The similar result was obtained for five other time series from Polish electric energy market [10] which are not discussed here.

In conclusion, the possible asymmetry of distribution of rates of return greatly influences the risk management, and justifies the incorporation of asymmetry in models. One could ask how strong it should be in order to take special care of it. This question will be addressed in the next paper.

References

[1] AKAIKE H. *Information theory and an extension of the maximum likelihood principle* 2nd. Int. Symp. On Information Theorized B N Petrov and F Csaki (Budapest: Akademiai Kiado), 1973, 267-281.

[2] BAILLIE R., BOLLERSEV T., MIKKELSEN H. O. *Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity*, Journal of Econometrics, 1996, 74, 3-30.

[3] BOLLERSLEV T. *A Conditionally Heteroscedastic Time Series Model of Security Prices and Rates of Return,* Review Economics Statistics, 1987, 59, 542 – 547.

[4] BOLLERSLEV T., MIKKLELSEN H. O. *Modeling and Pricing Long-Memory in Stock Market Volatility*, Journal of Econometrics, 1996, 73, 151-184.

[5] BROCKWELL P. J., CityDAVIS R. A. *Introduction to Time Series and Forecasting*, Springer – Verlag, New York, 1996.

[6] DAVIDSON J. *Moment and Memory Properties of Linear Conditional Heteroscedastity Models*, Manuscript, Cardiff PlaceTypeUniversity, 2001.

[7] DING Z., GRANGER C.W.J., ENGLE R. F. *A Long Memory Property of Stock Market Returns and a New Model*, Journal of Empirical Finance, 1993, 1, 83-106.

[8] ENGLE R. F. *Autoregressive Conditional Heteroscedasticy with Estimates of the Variance of placecountry-regionUnited Kingdom Inflation*, Econometrica, 1982, 50, 987-1007.

[9] ENGLE R. F., BOLLERSLEV T. *Modeling the Persistence of Conditional Variance,* Econometric Review, 1986, 5, 1-50.

[10] GANCZAREK A. *Weryfikacja modeli z grupy GARCH na dobowo-godzinnych rynkach energii elektrycznej w Polsce, Rynek Kapitałowy. Skuteczne inwestowanie, Zeszyty Naukowe Uniwersytetu Szczecińskiego, Szczecin 2008, in press.*

[11] HANNAN, E. J., AND B. G. QUINN. *The Determination of the Order of an Autoregression*, Journal of the Royal Statistical Society, B, 41, 1979, 190-195.

[12] MITCHELL H., MCKENZIE M. *GARCH model selection criteria*, Quantitative Finance, 3, 2003, 262-284.

[13] NELSON D. *Conditional Heteroskedasticity in Basset Returns: a New Approach,* Econometrica, 1991, 59, 347-370.

[14] SCHWARZ G. *Estimating the Dimension of a Model,* The Annals of Statistics, 7978, 6, 461-464

Gianluigi Giorgioni, Ken Holden

Faculty of Business and Law
John Moores University, Liverpool
98 Mount Pleasant, L3 5UZ Liverpool

e-mail: G.Giorgioni@ljmu.ac.uk, K.Holden@ljmu.ac.uk

Jana Hančlová

Finance department
Faculty of Economics
VŠB – Technical University Ostrava
Sokolská 33, 701 21 Ostrava

e-mail: jana.hanclova@vsb.cz

A Review of the Current Methods of Econometric Modelling

Abstract

This paper reviews the different methods used in econometric modelling. The traditional approach, as developed by the Cowles Commission in the USA, is illustrated by considering the Klein III model. This model of the USA economy, developed in 1950, had 21 observations, simple dynamics, no optimising behaviour and was based on the Keynesian approach. Since then, thanks to increased computer power and the availability of longer and more accurate macroeconomic time series, it has been possible to develop larger models that allow greater disaggregation of the economy and detailed simulations. Moreover, the poor forecasts of, in particular, inflation and exchange rates in the 1970s led to the modification of the Keynesian framework, the development of models inspired by monetarist theories and the introduction of "model consistent" expectations. We then consider in more detail two international models - the NiGEM model of 35 countries and the OECD Global Model - and two modern single-country models - for the Bank of England and the Czech National Bank. For the future, any modelling has to deal with the increasing importance of the European Union, and, from the point of view of the Czech Republic, with the possible adoption of the Euro as the currency.

Keywords: *econometric modelling, macroeconomic models, Cowles Commission, Klein III model, NiGEM model, OECD Global Model.*

1 Introduction

The purpose of this paper is to review some recent econometric models and consider the current methods of modelling. Econometric models have two main purposes: forecasting the future state of the economy and predicting the effects of alternative policies. While spreadsheets can be useful for extrapolative forecasting they cannot explain the economy.

The real world is complicated and, to be understood, it must be simplified. According to the Bank of England ([1] ch.1 page 1) "all models are imperfect, precisely because they are simplifications of reality". However, they are useful because they formalise economic assumptions and linkages in a systematic way. The paper first considers the Klein III Model, constructed in 1950, and then the developments of traditional structural modelling since then. Two modern international models and two single country models are discussed. Finally, some comparisons are made and the implications for modelling the Czech Republic are considered.

2 Klein III Model

To provide an historical perspective on how macro-modelling has evolved over the last half a century, we outline the main features of the Klein III model (see [4]) published in the Cowles Commission. This is a prototype for later models. It was considered at the time to be a "large structural" model, although it had only 12 stochastic equations and four non-stochastic relationships (identities and definitions). Because it focuses on the United States during the inter-war period, the foreign sector and exchange rate determination are ignored, and interest rates have a minimal effect. The equations are summarised in the table.

The equations are mainly linear in the levels of variables and estimation is by ordinary least squares and limited information maximum likelihood. The main features are the establishment of a circular flow to determine expenditures and income, and the inclusion of both real variables (exogenously deflated) and nominal ones. Aggregate demand is split into five components: consumption, investment, inventories, owner-occupied housing and rented housing. There are further equations for the proportion of non-farm housing units which are occupied, and rental payments. There are three adjustment equations: rents adjust with the level of disposable income and the level of occupation of housing; interest rates adjust with the tax rate and a time trend; and output adjusts to price changes and unexpected changes in inventories. The wage bill is determined by output. Aggregate supply and demand are assumed to balance, except for a random disturbance. Since annual data are used, and there are only 21 observations, there are simple dynamics, with the maximum lag being generally one year. Lagged values of the endogenous dependent variables are included in the inventories equation and the demand for idle cash balances equation. A time-trend is included in six equations.

Stochastic Equations	Main Explanatory Variables
Demand for consumption	Income, time trend
Demand for investment	sales income, capital stock
Demand for inventories	price level, net private output
Demand for owner-occupied houses	rent, income
Demand for rented housing	rent, bond yield
Occupied housing units	income, rent, time trend
Rent adjustment	income, occupied housing units, rent
Wage bill	output, time trend
Demand for active balances	income, tax revenue, time trend
Demand for idle cash balances	bond yield, time trend
Interest rate adjustment	interest rate, tax rate, time trend
Output adjustment	price change, inventories shock
Identities	
Net national product	income, taxes, demand variables
Private output excluding housing	output, rental income, farm incomes
Capital stock	investment
Rental payments	rent, occupied housing units

There are a number of weaknesses. Each item of domestic demand is deflated with a different (exogenous) price deflator, and only a broad price level for output is endogenous. The two money demand variables, for active balances and for idle cash balances, are not integrated with the rest of the model. Likewise, some of the other endogenous variables only occur in one equation and the national income identity. There is no production function and no focus on optimising behaviour. Following the publication of the Klein III model there have been many developments in econometric modelling which we now consider.

3 Developments of Traditional Structural Models

The developments in econometric modelling we consider are, inevitably, inter-related. The most important ones are:

a. More Equations

It was assumed forecasts would be improved if the models had more equations. Following the Klein III model several larger models were developed for the USA, in the 1960s, including the 203 equation Brookings model (see [3]). This was constructed by a team of 19 economists and included high levels of disaggregation into industrial sectors where output, prices, wages etc are explained, and linked to categories of expenditure. Similar models were developed for other countries. These models had a number of common features: they were intended to test economic theory, produce forecasts and have desirable simulation properties. They required data on a large number of variables and their upkeep was expensive. While some companies developed models, in the main it was academic and government institutions, supported by research grants and sales of their forecasts, which came to dominate the field.

b. Simulation Methods Developed

With the massive increase in computer power the prospect of extensive simulation analysis of models has been realised. Estimated econometric models are dynamic difference equations and the short-term and long-term properties of models can be examined. These include the effects of shocks to individual equations and the model as a whole, as well as stochastic simulation to account for the uncertainty in the parameter estimates.

c. Analysis of Forecast Errors

During the 1970s, following the large oil price rises of 1972 and 1975, the forecast record for inflation and growth in Western economies, was poor. The basic problem was that models estimated using data from the low inflation world of the 1950s and 1960s did not apply to the high inflation world of the 1970s. This resulted in attempts to improve the models. One approach was to add extra equations to include influences which had been neglected—for example, the original, 1968 Keynesian version of the London Business School model, with 24 equations, became, following the floating of the pound in 1972, a 350 equation "international monetarist model", which included traded/non-traded goods sectors and monetary equilibrium.

d. Monetarist Ideas Accepted

The role of money in macroeconomics became important in the 1970s and Western governments controlled inflation through the money supply. This provided an incentive for models to include monetarist concepts linking money supply, interest rates and inflation. A related development was the revision of models following the widespread introduction of floating exchange rates in the 1970s.

e. Model Consistent Expectations

Another factor that became important in economic modelling was the role of expectations. Survey data established some links between expected inflation and actual inflation. Models were developed which included equations to predict prices. Initially, the forecasts of inflation were adaptive (or backward-looking) but "model consistent" expectations developed. These required the expectations of future inflation to be the same as the model's forecasts of inflation, with random errors. While the assumption of strict "rational expectations" (i.e. that the forecasts of inflation would be perfect) was not always accepted because of learning and delays in information flows, model consistent expectations became a common property of models.

f. A theoretical Models Developed

From the early days of modelling the accuracy of forecasts has been judged relative to benchmark forecasts. For stationary variables, a naïve "no change" forecast was used, while for trending variables the naïve model was a "same change" forecast. An econometric model was expected to be more accurate than these naïve models. The development of Box-Jenkins univariate and multivariate models resulted in the atheoretic approach of Sims [7]. This rejects the usefulness of information arising from economic theory and starts with a list of economic variables, suitably differenced to be stationary. Then unrestricted vector autoregressive (VAR) models are fitted to the data and the system is used for forecasting. A variation, which avoids the over-parameterisation common in unrestricted VARs, is to assume each series can be represented by a simple random walk, and then use Bayesian estimation methods to determine whether it is useful to add extra variables. These small atheoretic models produce relatively accurate short-term forecasts but, being reduced forms, are not useful for policy simulation. It is difficult to relate their equations to the underlying economic behaviour of agents and markets.

g. Sample sizes and Estimation Methods Improved

The early models had very limited amounts of data. As mentioned previously, the Klein III model had only 21 annual observations. Now, for many countries, consistent annual data goes back to the 1950s, giving over 55 observations, and for some countries, quarterly data from the 1960s gives over 180 observations. This provides a wider range of economic experience and allows sophisticated model selection methods to be used in selecting equations. The early models were estimated by ordinary least squares (OLS) or limited information maximum likelihood, rather than by the theoretically preferred systems methods. However, the major problem of systems methods of estimation is that the assumption of correct specification of the whole model is likely to be invalid. Single equation methods, such as OLS and instrumental variables, avoid this problem.

4 Some Recent Structural Models

a. NIESR NiGEM Model [5]

The NiGEM model is an inter-related set of 35 country models and 13 regional blocks which are used for forecasting and policy analysis. It is the core of the European Commission financed European Forecasting Network in EUROFRAME (see www.euroframe.org). The model is based on economic theory and emphasises international linkages. As well as the obvious links through trade (affected by demand and relative competitiveness) there are further effects from exchange rates and the patterns of domestic and foreign asset holding. The model contains the determinants of domestic demand, exports and imports, GDP and prices, current accounts and net assets. Consumption depends on income and total wealth in the long run. Countries are linked in their financial markets via the structure and composition of wealth, emphasising the role and origin of foreign assets and liabilities. There are complete demand and supply sides, and there is an extensive monetary and financial sector.

Labour markets have rational expectations. Forward-looking variables are included in the determination of the exchange rate, short-term interest rates, consumption, equity prices, prices, wages, long rates and the weights in long real interest rates. Budget deficits are controlled by tax changes. It is assumed that the authorities use short-term interest rates to stabilise the price level (or the inflation rate) in the long run. The model has 3677 equations, of which 448 are estimated. The trade equations and the labour market equations are estimated in error-correction (or equilibrium-correction) form which determines their long-run properties.

b. OECD New Global Model [6]

The OECD New Global Model is smaller and more regionally aggregated than the OECD Interlink model and it uses quarterly data. It has been developed to analyse the effects of shocks on different countries. There are three country models - for the USA, Japan and China - and 6 economic blocks models - for the Euro Area, Other OECD Europe (which includes the Czech Republic), Other OECD, Other non-OECD Asia, Non-OECD Europe, and Africa, the Middle East and Latin America. Each country or bloc model combines short-term Keynesian-type dynamics with a consistent neo-classical supply-side in the long run. A constant returns to scale Cobb-Douglas production function determines output. The demands for labour (hours worked) and capital are derived from the first-order profit maximising conditions under imperfect competition. Output is largely demand-driven in the short term, but supply-driven in the longer term, with the long-run capital-labour ratio being equal to the ratio of real producer wages to the real cost of capital. Potential output depends on trend employment, trend productivity, trend hours worked and the capital stock.

Wage inflation is affected by price inflation, trend productivity growth and the gap between the unemployment rate and the NAIRU (which is exogenous). The price equations are homogenous with respect to domestic costs and foreign prices, and have error-correction terms so that they determine long-run price levels. Domestic prices and inflation are determined by domestic costs, import prices and the output gap. Export prices reflect domestic and competitors output prices. Trade volumes depend on income and relative prices. Household consumption depends on real disposable income and real net wealth (i.e. net housing and financial assets). The long-run savings ratio is a function of the net wealth to income ratio. The different models are linked by international trade, international capital markets and returns on international assets. There is stock-flow consistency both within and across models.

c. The Bank of England Quarterly Model [1]

The Bank of England uses monetary policy to achieve an inflation target (currently 2% inflation in the UK consumer prices index) through changes in the short-run nominal interest rate. The model produces forecasts of output growth and inflation up to three years ahead. While these projections are important, the MPC is keen to understand the economics of the risks and uncertainties around the central projection. It finds the use of the model, based on economic theory, helpful in combining the different assumptions and judgements being made.

The model consists of a theoretical core to which data-driven dynamics are added. The theoretical core is a general equilibrium model in which endogenous variables are determined by the equilibration of demand and supply in markets for goods, labour and assets. Households maximise utility and firms maximise profits, while the government reacts to ensure debt sustainability and the monetary authority controls the short-term interest rate. The "rest of the world" is assumed to be large, relative to the UK. Goods markets are monopolistically competitive. Labour market equilibrium is not perfectly competitive. Asset market prices reflect standard no-arbitrage conditions. The model is explicit about the motivation of agents, the constraints they face, and the conditions in the markets they interact in.

The core model is intentionally theoretical and simple, so that it provides a plausible framework for analysing economic issues, and has a well-defined steady state long-run sustainable equilibrium.

To model the properties of the data, a layer of ad hoc non-core dynamics is added to the core model, to give the full forecasting model. While this has the potential effect of violating the assumptions behind the core model, it allows adjustments to be made which result in a good fit to the data. As well as lagged variables, proxies for missing effects are included, as are error-correction terms and modifications from the direct application of judgement. The basic requirement is that, after these modifications, in the long run the values of the variables in the full forecasting model converge to the values of the long-run path of the variables in the core model. However, in the short term, the imposition of judgement means that behaviour is potentially quite flexible. Projections from the non-core equations are not allowed to feed back into the core model, to avoid violating the assumptions of the core model.

Agents' expectations are model consistent, so they agree with the core model solutions. The parameter values of the steady state of the model are determined by judgement and the past history of the relationships. Then the parameters of the core model are determined in the same way. Finally the non-core equations are estimated using OLS.

The core model has 130 endogenous variables, 35 exogenous and 30 working variables. There are 160 equations. The non-core model has 168 endogenous variables, 49 exogenous and 176 equations. The data start in 1978 quarter 1, giving 104 observations to 2003 quarter 4, which are used for the OLS estimation of the non-core model.

d. Czech National Bank Quarterly Model [2]

The Czech Republic was founded in 1993 and so there is limited historical data. Also, the structure of the economy has been changing continuously, so a large model estimated using historical data would not be useful.

The Czech National Bank Quarterly Projection Model is a small theory-based model, which explains output, employment, interest rates, exchange rates and inflation, with a link to monetary policy. The short-term forecast is almost entirely judgemental, while in the longer term economic relationships are important. The model offers consistency and the power to address risks and uncertainties. It provides an economic framework for the medium-term baseline scenario, around which the sensitivity of the assumptions can be investigated. While the model has little to say about the general equilibrium, it attempts to explain the dynamics of disequilibrium by focussing on the "gaps" between the actual and trend values of output, unemployment, and interest rates. However, there are no stock/flow links to avoid inconsistencies.

There are 14 behavioural equations and identities bringing the total number of equations up to 85. The behavioural equations include three identities which define the output gap, the unemployment gap and the link between the Czech-German exchange rate with Czech and German prices. The other equations are reduced forms (or quasi-reduced forms) based on economic theory. There is no supply side.

The fundamental role of monetary policy is to anchor expectations to the target rate of inflation. The short-term interest rate is the policy instrument. This is determined by the difference between the forecast (expected) rate of CPI inflation and the target rate, four quarters ahead. If the expected inflation rate exceeds the inflation target, the short-term interest rate increases. In the short term there is some inertia but, in equilibrium, the nominal rate equals the real interest rate plus the expected rate of inflation.

Because of the short data series and changing structure (and the developing monetary policy) the model is not estimated. Instead, calibration is preferred, using judgement and the simulation properties of the model. Potential output and NAIRU are determined by using filters (Hansen-Hodrick and Kalman) on historical data.

The model is used, with judgement, to determine the "baseline" scenario which is consistent with the assumptions and the reaction function. The model is almost linear and so shocks have a symmetric effect. Stochastic simulation is used to determine confidence intervals for projections.

e. Conclusions on the Models

It is useful to try to identify similarities and contrasts between the four quarterly models just reviewed. There is a common reliance on economic theory, for optimising behaviour determining factor demands and for checking the steady state properties of the models. International links between markets are important in all the models. Only the OECD model does not include model-consistent expectations, preferring adaptive ones in financial markets. All the models except the CNB impose stock/flow consistency and include a production function. There are obvious similarities between the two one-country models (BE and CNB), which are calibrated, and the two global models (NI and OECD), which are estimated from the data. The models are used for policy analysis.

5 Implications for modelling the Czech Republic

There are three problems in economic modelling:

- To include economic theory that is internally consistent and easily understood by the user.
- To relate the theoretical model to the data so that it is useful.
- To make the model reliable under different forecasting assumptions, and amenable to judgemental adjustments.

Decisions are needed on the purpose of the model, the theory for a changing economy (e.g. the full impact of membership of the EU, the move towards the euro, the changing world economy), which variables are to be modelled, whether annual or quarterly data are to be used for estimation or calibration. Finally, are there any lessons from the performance of the Czech National Bank model?

Acknowledgement

The research work on this paper was supported by the Czech Science Foundation, grant number 402/08/1015 - Macroeconomic Models of the Czech Economy and Economies of the other EU Countries.

References

[1] Bank of England Quarterly Model (2005) See
http://www.bankofengland.co.uk/publications/other/beqm/index.htm

[2] The Czech National Bank's Forecasting and Policy Analysis System (March 2003) Edited by Warren Coats, Douglas Laxton and David Rose. `http://www.cnb.cz/en/public/publications`

[3] Duesenberry, J. S. et al. (1965*) The Brookings Quarterly Econometric Model of the United States Economy* (Chicago: Rand McNally)

[4] Klein, L. R. (1950). *Economic Fluctuations in the US 1921-1941* (New York: John Wiley)

[5] NIESR NiGEM Model `http://www.niesr.ac.uk/research/research.php`

[6] OECD New Global Model (2007) Annex A of `http://www.olis.oecd.org/olis/2007doc.nsf/linkto/eco-wkp(2007)12`

[7] Sims, C.A. (1980). Macroeconomics and Reality. *Econometrica*, vol. 48, 1-48.

Agata Gluzicka, Donata Kopańska-Bródka

Department of Operations Research
University of Economics in Katowice
ul. 1-Maja 50, 40-287 Katowice

e-mail: gluzicka@ae.katowice.pl, broda@ae.katowice.pl

Some Measures of the Investment Risk and the Coherency

Abstract

Risk is one of fundamental terms in the theory of uncertain investment decisions. The risk can be defined in many ways. Some authors define risk as the changes in values between two moments in time. We can also say that risk is a random profit or loss of the future worth.

Many different measures can be used to analysis of the risk. Investors are seeking still this best and universal measure of the risk. In 1999, Artzner, Delbaen, Eber and Heath proposed the axioms of the coherent measure. They say that the good measure of the risk should fulfill these axioms. The Worst Conditional Expectation, Conditional Value at Risk or Expected Shortfall are examples of coherent measures of the risk.

However these axioms of coherency haven't fulfilled by traditional measures of the risk. In literature, we can find some theorems which let convert some measures of dispersion to the coherent form. These rules were proposed e.g. for the standard semideviation, the mean semideviation and for the Gini's Mean Difference.

The theory of coherent measures and theorems about the relation of measure of dispersion with the coherency will be presented in the article. The converted form of some risk measures will be applied to the construction of the optimal portfolios.

Keywords: *measure of the investment risk, coherency, optimal portfolio*

1 Introduction

In portfolio analysis we can use many different measures. We can also select optimal portfolio according to different optimization models. Many researchers still seeking the best measure or the best methods of the optimal portfolio selection.

For modeling the choice among uncertain prospects we can use two methods: stochastic dominance relation and the mean-risk approaches. Mean- risk approaches are not capable to use all information about the risk-averse preferences. Moreover, for the typical risk measures, the mean-risk approach may lead to inferior conclusions. Many authors have pointed out that the mean-variance model is, in general, not consistent with stochastic dominance.

The relation of stochastic dominance is one of the fundamental concepts of the decision theory. Taking into account the axiomatic models of preferences for choice under risk, we can introduce a partial order in the space of real random variables. For some risk measures used in mean-risk models we can prove the stochastic dominance consistency.

On the other side we have the class of coherent risk measures. This class is defined by few axioms, which every good measure of risk should fulfill. Coherent measures are usually applied to the insurance risk.

In this paper we will present the axioms of theory of coherency and the relationship between typical, traditional measures of risk and the coherency. In empirical examples we will analyze optimization models with these measures of risk which fulfill the axioms of coherency.

2 Coherent measure of risk

Artzner [1] have introduced a class of risk measures by means of several axioms. The axioms present the most important issues in the risk comparison for economic decisions. Therefore, they are being regarded as the standard requirements for risk measures. Denote Ω – the set of states of nature, and assumed it is finite; $\mathcal{G}$ - the set of all real-valued functions (set of all risks) on Ω. Every mapping ρ from $\mathcal{G}$ into R is called a measure of risk. For any measure of risk ρ we can consider several possible properties:

Axiom of translation invariant: $\rho(X + \alpha) = \rho(X) - \alpha$ for all $X \in \mathcal{G}$ and real number α

Axiom of subadditivity: $\rho(X + Y) \leq \rho(X) + \rho(Y)$ for all $X, Y \in \mathcal{G}$

Axiom of positive homogeneity: $\rho(\lambda X) = \lambda\rho(X)$ for all $\lambda \geq 0$ and all $X \in \mathcal{G}$

Axiom of monotonicity: for $X \leq Y$ we have $\rho(Y) \leq \rho(X)$, for all $X, Y \in \mathcal{G}$

Axiom of risk relevant: for $X \leq 0$ and $X \neq 0$ we have $\rho(X) > 0$.

Note that if the performance function is positively homogeneous, then axiom of subadditivity is equivalent to the standard convexity requirement:

Axiom of convexity: $\rho(\alpha X + (1 - \alpha)Y) \leq \alpha\rho(X) + (1 - \alpha)\rho(Y)$ for any $0 \leq \alpha \leq 1$

A risk measure satisfying the axioms of translation invariance, subadditivity, positive homogeneity and monotonicity is called coherent measure.

Examples of the coherent measures are: The Worst Conditional Expectation, Conditional Value at Risk or Expected Shortfall. However these axioms of coherency haven't fulfilled by traditional measures of the risk (like variance, standard deviation, semideviation etc.).

3 Relation between stochastic dominance and the coherency

Stochastic dominance is one of the criterions which could be applied to analyze the investment risk. However in the case when we have rates of return with normal distribution, mean-variance model is consistent with the second stochastic dominance rules. In general, this consistent isn't valid and application this model can lead to select dominated portfolio in the SSD and FSD sense. The Markowitz model is frequently criticized as not consistent with axiomatic models of preferences for choice under risk.

This problem was solved by introducing a new form of risk measure. Baumol [2] proposed considering a measure $\mu(X) - \lambda\sigma(X)$ to be maximized instead of the minimization of standard deviation $\sigma(X)$ (X – random variable, $\mu(X)$ – the mean rate of return of portfolio, λ is some constant). He called this measure the expected gain-confidence limit criterion. Similarly, Yitzhaki [10, 11] considered maximization of the criterion $\mu(X) - \Gamma(X)$, where Γ means the Gini's Mean Difference.

For any dispersion type risk measures $\rho(X)$, the function $S(X) = \mu(X) - \rho(X)$ can be defined as the corresponding safety measures. We say that the safety measure $\mu(X) - \rho(X)$ is SSD consistent or that risk measure $\rho(X)$ is SSD safety consistent if [5]

$$X \succeq_{SSD} Y \Rightarrow \mu(X) - \rho(X) \geq \mu(Y) - \rho(Y)$$

for two random variable X, Y. The relation of SSD (safety) consistency is called strong if, in addition to above condition, the following holds

$$X \succ_{SSD} Y \Rightarrow \mu(X) - \rho(X) > \mu(Y) - \rho(Y).$$

The following risk measures are SSD safety consistent: the standard semideviation [7]; the mean semideviation [2, 3]; the conditional β-semideviation [8]; the maximum semideviation [8]; the Gini's mean difference [10, 11]. The Gini's mean difference is one of these measures satisfying the strong SSD consistency.

Mentioned SSD safety consistent measures are convex, positive homogeneity and fulfill the translation invariant condition. For these measures of risk we can find the relationship with the coherency. This relation presents the following theorem:

Theorem 1 [5]. Let $\rho(X) \geq 0$ be a convex, positively homogeneous and translation invariant (dispersion type) risk measure. If the measure satisfies additionally the SSD safety consistency for random variables X and Y, then the corresponding performance function $C(X) = -S(X) = \rho(X) - \mu(X)$ fulfills the coherency axioms.

Theorem 2 [5]. Let $\rho(X) \geq 0$ be a convex, positive homogeneity and translation invariant (dispersion type) risk measures. If the measure is additionally bounded by mean rate of return:

$$X \geq 0 \Rightarrow \rho(X) \leq \mu(X)$$

then the corresponding performance function $C(X) = \rho(X) - \mu(X)$ fulfills the coherency axioms.

As a measure of risk we can consider also the combination of some risk measures. For the safety consistent measures we have valid the following condition:

Theorem 3 [4]. If all risk measures $\rho_k (k = 1, 2, \ldots, m)$ are SSD safety consistent, then every combined risk measures

$$\rho_c(X) = \sum_{k=1}^{m} w_k \rho_k(X) \quad \sum_{k=1}^{m} w_k(X) \leq 1 \quad w_k > 0,\ k = 1, 2, \ldots, m$$

is also SSD safety consistent in the sense that

$$X \succeq_{SSD} Y \Rightarrow \mu(X) - \rho_c(X) \geq \mu(Y) - \rho_c(Y)$$

Moreover, if at least one measures ρ_{k0} is strongly SSD safety consistent, then every combined risk measure ρ_c is strongly SSD safety consistent.

The weighted combination allows us to regularize any SSD (safety) consistent risk measures to achieve the strong consistency. Recall that the Gini's mean difference satisfies the strong SSD (safety) consistency. According to theorem mentioned above, combination of any risk measures with the Gini's mean difference is an example of measure satisfying the strong SSD consistency.

Theorem 4 [4]. For any SSD safety consistent risk measures $\rho(X)$ and any $\varepsilon \in (0;1)$, the combined risk measures

$$\rho_c(X) = (1-\varepsilon)\rho(X) + \varepsilon\Gamma(X)$$

is strongly SSD safety consistent.

It is obvious that the combination of convex measures of risk is convex. If measures of the risk are positive homogeneity the combination such measures is also positive homogeneity. Moreover, if measures fulfill the translation invariant axioms, combination of these measures fulfill this axiom too. So, the combination of risk measures realizes the assumptions of the theorem 1. Therefore this combination is an example of measure which fulfills the axioms of coherency.

4 Empirical example

In empirical example we will construct and analyze optimal portfolio by using the following measures:

- The Gini's mean difference: $\Gamma(X) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|\, p_i p_j$
- The mean absolute semideviation: $\overline{\delta} = \sum_{i=1}^{n}\overline{\delta_i}$ where $\overline{\delta_i} = \begin{cases} |x_i - \mu| & if\ \ x_i - \mu < 0 \\ 0 & if\ \ x_i - \mu \geq 0 \end{cases}$
- The standard semideviation: $\overline{\sigma} = \sum_{i=1}^{n}\overline{\sigma_i}$ where $\overline{\sigma_i} = \begin{cases} (x_i - \mu)^2 & if\ \ x_i - \mu < 0 \\ 0 & if\ \ x_i - \mu \geq 0 \end{cases}$

where X is a random variable such as $Pr\{X = x_i\} = p_i$ for $i = 1, 2, \ldots, n$ and $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$ is a mean of X. For these measures we use optimization models in the following forms:

$$2\sum_{i=1}^{n} x_i \ \mathrm{cov}[R_i; F_i(R_i)] - \sum_{i=1}^{n} \mu_i x_i \to \min,\ \sum_{i=1}^{n} \mu_i(x) x_i \geq \mu_0,\ \sum_{i=1}^{n} x_i = 1,\ x_i \geq 0$$

Model 1 (with the Gini's mean difference [10, 11])

$$\frac{1}{T}\sum_{t=1}^{T} y_t - \sum_{i=1}^{n} \mu_i x_i \to \min,\ y_t + \sum_{i=1}^{n}(r_{it} - \mu_i) x_i \geq 0,\ y_t - \sum_{i=1}^{n}(r_{it} - \mu_i) x_i \geq 0,\ \sum_{i=1}^{n} \mu_i x_i \geq \mu_0$$

$$\sum_{i=1}^{n} x_i = 0,\ x \geq 0,\ y_t \geq 0 \text{ for } i = 1, 2, \ldots n \text{ and } t = 1, 2, \ldots T$$

Model 2 (with the mean absolute semidevition [3])

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \overline{\sigma_{ij}} x_i x_j - \sum_{i=1}^{n} \mu_i x_i \to \min,\ \sum_{i=1}^{n} \mu_i x_i \geq \mu_0,\ \sum_{i=1}^{n} x_i = 1,\ x_i \geq 0$$

Model 3 (with the standard semideviation [7])

where

x_i – the share of the i-th stocks in the portfolio,

μ_i – expected rate of return for i-th stocks,

μ_0 – parameter representing the minimal rate of return required by an investor,

r_{it} – realization of random variable r_i during period t (the historical data),

$R_i = [r_{i1}, r_{i2}, ..., r_{iT}]$ – vector of realization of random variable r_i,

F_i – cumulative distribution function of the R_i,

y_t – auxiliary variables,

$\overline{\sigma_{ij}} = \begin{cases} (r_{it} - \mu_i)(r_{jt} - \mu_j) & if \quad (r_{it} - \mu_i)(r_{jt} - \mu_j) < 0 \\ 0 & if \quad (r_{it} - \mu_i)(r_{jt} - \mu_j) > 0 \end{cases}$ - matrix of semivariances - semi-covariances.

In the next part of examples, as a criterion of optimization we consider the combinaton of risk measures. In model 4 as a measure of risk we use $(1 - \varepsilon)\overline{\delta}(x) + \varepsilon\Gamma(x)$and in the model 5 we minimize $(1 - \varepsilon)\overline{\sigma}(x) + \varepsilon\Gamma(x)$ (where $\overline{\sigma}$ - standard semideviation, Γ - Gini's Mean Difference, $\overline{\delta}$ - mean-absolute semideviation and constant $\varepsilon \in (0; 1)$). These models have been analyzed for different value of ε. Note that every measures in these models are SSD consistent measure and fulfill the axioms of coherency.

The data from Warsaw Stock Exchange will be used in this example. We analyse a few diffrent groups of stocks: group of stocks from WIG20 index (group of 20 stocks), group of stocks from TECHWIG index (group of 30 stocks), group of stocks noted continously in 2007 (group of 100 stocks), group of stocks noted continously during January 2006–December 2007 (group of

70 stocks). Daily date (i.e. daly rates of return) from selected stocks were used to generated set of portfolios based on the models presented above. For every considered group of stocks we received the similar results, so we present conclusions for two first group.

We constructed the next portfolios for different required level of rate of return of portfolio. We received different optimal portfolios by using different models. We compared optimal portfolios, calculated according to different measure and for the same assumed level of return. These portfolios have different structure and the level of diversification. But using each of examinated models, we can select only the efficient portfolio. The pictures belows are presenting the efficient frontier for portfolios with the stocks from TECHWIG index (figure 1, 2, 3) and for portfolios with the stocks from WIG20 index (figure 4, 5, 6).

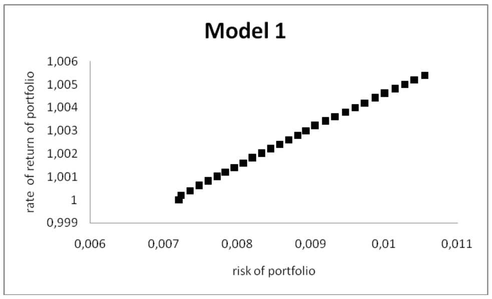


Figure 1. Efficient frontier for model 1 (TECHWIG index)

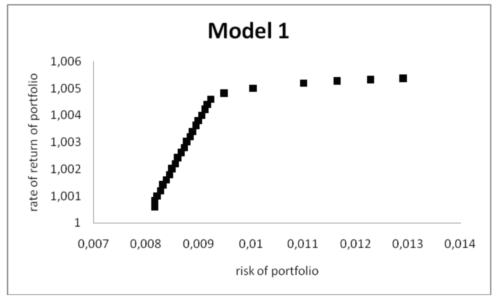


Figure 2. Efficient frontier for model 1 (WIG20 index)

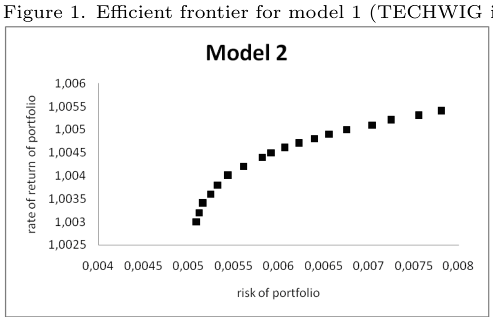


Figure 3. Efficient frontier for model 2 (TECHWIG index)

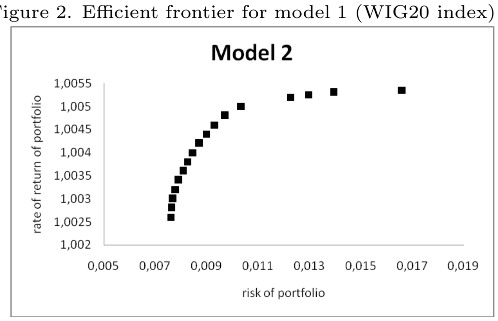


Figure 4. Efficient frontier for model 2 (WIG20 index)

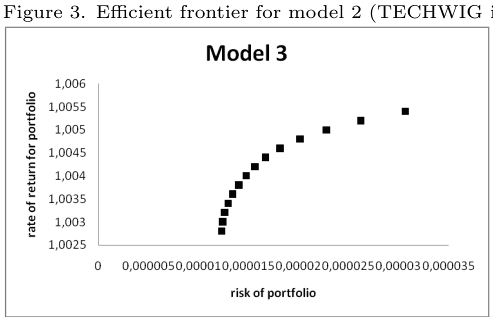


Figure 5. Efficient frontier for model 3 (TECHWIG index)

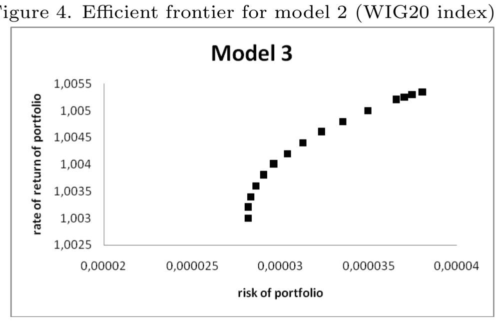


Figure 6. Efficient frontier for model 3 (WIG20 index)

Model 4 was used for different value of coefficient ε. Based on the results we suggest that: if the mean-absolute semideviation is applayed as the measure of risk, the portfolios with the highest risk are received for the higher value of the coefficient ε. In the case when the

<table>
<tr><td rowspan="2">data</td><td colspan="4">Portfolios for the lowest required level of return</td><td colspan="4">Portfolios for the highest required level of return</td></tr>
<tr><td>Model 1</td><td>Model 2</td><td>Model 3</td><td>Model 4</td><td>Model 1</td><td>Model 2</td><td>Model 3</td><td>Model 4</td></tr>
<tr><td>02/01/08</td><td>1,0239</td><td>0,9623</td><td>0,9208</td><td>0,9629</td><td>0,9753</td><td>0,8568</td><td>0,7945</td><td>0,8501</td></tr>
<tr><td>09/01/08</td><td>1,0040</td><td>0,9309</td><td>0,9200</td><td>0,9545</td><td>0,9995</td><td>0,8855</td><td>0,8205</td><td>0,8787</td></tr>
<tr><td>16/01/08</td><td>1,0191</td><td>0,9313</td><td>0,9085</td><td>0,9581</td><td>1,0222</td><td>0,8980</td><td>0,8298</td><td>0,8917</td></tr>
<tr><td>23/01/08</td><td>1,0274</td><td>1,0880</td><td>1,1018</td><td>1,0647</td><td>0,9966</td><td>1,1161</td><td>1,0133</td><td>1,1218</td></tr>
<tr><td>30/01/08</td><td>1,0112</td><td>1,0120</td><td>1,0133</td><td>1,0126</td><td>1,0240</td><td>1,0195</td><td>1,0184</td><td>1,0212</td></tr>
<tr><td>06/02/08</td><td>0,9889</td><td>0,9780</td><td>0,9744</td><td>0,9782</td><td>0,9930</td><td>0,9779</td><td>0,9756</td><td>0,9779</td></tr>
</table>

Table 1: Real rates of return for indicated optimal portfolios from model 1, 2, 3, 4 (for ε=0,2) (TECHWIG index).

standard semideviation is a risk measure, the differences between risk of optimal portfolio are very small (the value of the standard semideviation is much smaller then value of the Gini's mean difference). The results for model 4 are presented on the pictures below (figure 7, 8).

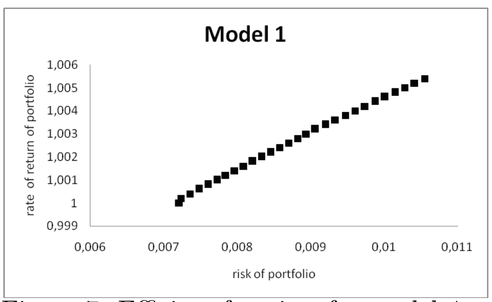


Figure 7. Efficient frontiers for model 4 (for different value of ε) – group of 30 stocks

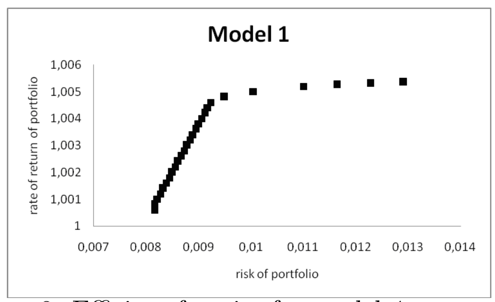


Figure 8. Efficient frontier for model 4 (for different value of ε) – group of 20 stocks

For different efficient portfolios, with the same required rate of return, the real rates of return is compared. These real of return were calculated on the base of the data from the next few weeks after the last observation (starting from 2nd January 2008). Results are presented in the below table 1.

We received the best – the highest real rates of return for portfolios calculated on the base of model 1. The lowest real rates of return we have for portfolios from model 3.

5 Conclusions

In this article we analyzed the measures of risk investment which are SSD safety consistent. Moreover these all measures satisfy the axioms of coherency. So they could be examples of good risk measures.

For presented models the efficient portfolio was selected. When we using different models we received different optimal portfolios according to composition and the level of diversification.

We can obtain the highest real rates of return for the optimal portfolios selected according to model with the Gini's mean difference as a measure of risk,.

Results of our research are inducing to further research concerning of the measure of the risk of investment. The further research will concentrate on the combination of different measures of the risk and they application in optimization models.

References

[1] ARTZNAER P., DELBAEN F., EBER J., HEATH D., *Coherent Measures of Risk*, Mathematical Finance, vol.9, July 1999, str.203-228,

[2] BAUMOL W.J., *An expected gain-confidence limit criterion for portfolio selection*, Management Science, vol. 10, 1964,174-182,

[3] KONNO H., YAMAZAKI H., M*ean-absolute deviation portfolio optimization model and its applications to Tokyo Stock Market*, Management Science, vol. 37, No.5, May 1991, str. 519-531

[4] KRZEMIENIOWSKI A., OGRYCZAK W., *On Extending the LP Computable Risk Measures to Account Downside Risk*, Computational optimization and Applications, 32, 2005, str.133-160,

[5] OGRYCZAK W., OPOLSKA-RUTKOWSKA M., *On Mean-Risk Model Consistent with Stochastic Dominance*, Report of the Institute of Control and Computation Engineering Warsaw University of Technology, May 2004

[6] OGRYCZAK W., OPOLSKA-RUTKOWSKA *SSD consistent criteria and coherent risk measures*, 2004

[7] OGRYCZAK W., RUSZCZYŃSKI A., *From stochastic dominance to mean-risk models: Semideviations as risk measures*, European Journal Optimization Research, 116, 1999, str. 33-50,

[8] OGRYCZAK W., RUSZCZYŃSKI A., *Dual stochastic dominance and related mean-risk models*, SIAM Journal Optimization, 13, 2002, str. 60-78

[9] OKUNEV J., *The Generation of Mean Gini Efficient Sets,* Journal of Business Finance and Accounting, January 1991, str.209-218,

[10] SHALIT HAIM, YITZHAKI SHLOMO, *The Mean - Gini Efficient portfolio Frontier*, The Journal of Financial Research, Vol. XXVII, 2005, str.59-75,

[11] YITZHAKI S., *Stochastic Dominance, Mean Variance and Gini's Mean Difference*, The American Economic Review, 1982, str.178-185,

Kateřina Gurinová, Vladimíra Valentová

Department of Department of Economic Statistics
Faculty of Economics
Technical University of Liberec
Studentská 2, 461 17 Liberec

e-mail: katerina.gurinova@tul.cz, vladimira.valentova@tul.cz

Using of Harmonic Analysis at Research of Economic Time Series

Abstract

The article deals with possibilities of using of harmonic analysis for research of time series of economic indicators. The harmonic analysis is a method which considers a frequency factor to be the most important for the analysis. The time variable is not so important for that just like in other methods. This approach is based on mathematical methods like Fourier analysis. The basic element is a trigonometric polynom. This conception allows to describe periodic properties of a time series explicitly and makes possible to identify the components of periodicity which participate in subject properties of analysing economic process significantly. A suitable instrument for the detection such periodic components is the periodogram. It allows to get an idea about the occurence intensity of particular frequences in the researched time series. The periods, which significance is proved by any test of periodicity, can be designated as a holder of periodicity in behaviour of the economic indicator. In this paper the methods of harmonic analysis were aplied on the time series of monthly data of inflation rate in 2003 – 2007 in the Czech Republic. The statistical software STATGRAPHICS Centurion XV was used for the periodogram construction. This programme provides both tabular and graphic outputs. The Fishers test was used for testing of particular periodic components significance. The calculations showed that one of the periods detected in the time series of inflation rate is statistically significant. This period can be considered to be a holder of periodicity.

Keywords: *harmonic analysis, Fourier analysis, frequence, analysis of periodogram, Fishers test*

1 Introduction to Harmonic Analysis

In reference to an extreme importance of researching of economic events and processes dynamics the time series analysis is number one in economics. Economic time series, e.g. macroeconomic indicators time series such as gross domestic product, inflation rate, unemployment rate etc., have got unique characteristics. It is necessary to respect these characteristics while working with them. Today a variety of methods and procedures exists which make description and extrapolation of time series behaviour possible. Except classic methods other tools for researching of time series develop in process of time. These methods are numerically-intensive and almost depend on using of a computer technology. Various approaches to working with time series exist. First, classic model describes particular forms of a movement with emphasis on a systematic component with ignoring of knowledge of real reasons. Another possibility is the Box-Jenkins methodology which insists on an irregular component. The harmonic analysis is a different approach. The time variable is not the fundamental factor like at previous two approaches. It is a frequency factor here.

2 Essence of Harmonic Analysis

The researched time series $y_t, t = 1, 2, \ldots, n$, is considered as a linear combination of sine and cosine curves with various amplitudes and frequencies in the harmonic analysis. It is possible to get an idea about intensity of particular frequencies occurrence in researched time series by special statistical tools, such as the periodogram. This conception enables a description of a periodic behaviour of a time series explicitly and especially identification of the periodicity components that participate in real properties of analysed economic process significantly. Harmonic analysis especially uses mathematical methods from the Fourier analysis. A time series can be described by a goniometric way with a help of the Fourier series that serves to a formulation of functions development through goniometric functions sine and cosine. Thus, the Fourier series enables a notation of any periodical behaviour through sine and cosine curves. It is possible to analyze also very complicated functions, which is difficult to portray, by this method. The mathematical basis of this conception is a trigonometric polynon commonly written as:

$$y_t = \mu_0 + \sum_{j=1}^{H} \alpha_j \sin \omega_j t + \sum_{j=1}^{H} \beta_j \cos \omega_j t + \varepsilon_t, \quad t = 1,\ 2,\ \ldots,\ n, \tag{1}$$

where $H = \frac{n}{2}$ (n even) and $H = \frac{n-1}{2}$ (n odd),

$\omega_j = \frac{2\pi j}{n}$ for $j = 1,\ 2,\ \ldots,\ H$ is the j-th frequency (angular frequency),

$\mu_0,\ \alpha_j,\ \beta_j$ are parameters which values must be estimated.

The y_t process need not be always expressed just in a time domain, i.e. a process changing its values in depending on time, but it is possible to express it also in a frequency (spectral) domain. We can consider both the approaches to be equivalent but the spectral domain is not so usual approach. The y_t process, expressed by formula 1.1, is a result of the sum of H harmonic terms. Actually, it is a mixture of waves with ω_j different frequencies and different amplitudes, i.e. a process formed by a lot of goniometric curves blending together. Some of them are complementary to other ones, others in contrary eliminate which results in the y_t process.

The ω_j frequency is given in radians per a time unit. The time unit is represented by the time interval between consecutive time series observations. Thus, the 2π radians per a time unit frequency means that just one cycle becomes in the time unit. If the frequency is π radians per a time unit, the whole cycle becomes in 2 time units. Thus, the frequency is higher, the cycles circulate more often in a behaviour of the time series. The period length, i. e. time during which one cycle becomes, has got a time dimension t and is called Fourier period $\tau_j = \frac{2\pi}{\omega_j} = \frac{n}{j}$. Too high frequencies can blend with other ones and this is the reason why they can be unidentifiable. This event is named aliasing. The shortest identifiable period is $\tau = 2$. Thus, the shortest time which one cycle becomes in is 2 time units. It corresponds to Nyquist frequency $\omega = \pi$ that is the highest frequency which enables researching of periodic time series behaviour.

The formula (1) can also be written as:

$$y_t = \mu_0 + \sum_{j=1}^{H} F_j(t) + \varepsilon_t,$$

where $F_j(t) = A_j \sin\{\omega_j t + \varphi\}$, $t = 1,\ 2, \ldots, n$, is the j-th harmonic term, A_j is an oscillation amplitude for the j-th period, φ_j is a shift phase. Because of relation $\alpha_j^2 + \beta_j^2 = A_j^2\left(\sin^2\varphi_j + \cos^2\varphi_j\right) = A_j^2$ it is possible to express the oscillation amplitude A_j by harmonic coefficients α_j a β_j. Particular harmonic terms in (1) have got greater profit for a time series description, their coefficients α_j a β_j are more different from 0. Then, the oscillation amplitude for the j-th period is higher.

2.1 Model with Hidden Periodicities

It is necessary to realize just some of Fourier periods are significant while others not. If supposing the first m $m \leq H$ Fourier periods τ_j are significant and the trend component is constant, i. e. $T_t = \mu_0$, it is possible to write the relation (1) as:

$$Y_t = T_t + P_t = \mu_0 + \sum_{j=1}^{m} \alpha_j \sin\omega_j t + \sum_{j=1}^{m} \beta_j \cos\omega_j t \tag{2}$$

where T_t is the trend component

P_t is the periodic component (It consists of the seasonal and cyclical component.)

The model (2) is named as the model with hidden periodicities with the constant trend. In reference to it is a regression model linear in parameters, it is possible to get estimates of the parameters by the least squares method. Solving the normal equations we obtain following formulas:

$$m_0 = \frac{1}{n}\sum_{t=1}^{n} y_t = \bar{y},$$

$$a_j = \frac{2}{n}\sum_{t=1}^{n} y_t \sin\omega_j t, j = 1,\ 2, \ldots, m,$$

$$b_j = \frac{2}{n}\sum_{t=1}^{n} y_t \cos\omega_j t, j = 1,\ 2, \ldots, m. \tag{3}$$

If the model with hidden periodicity is not constant, i. e. it changes in time, it can be written as:

$$Y_t = T_t + P_t = T_t + \sum_{j=1}^{m} \alpha_j^* \sin \omega_j t + \sum_{j=1}^{m} \beta_j^* \cos \omega_j t, t = 1,\, 2,\, \ldots,\, n,$$

where parameters α_j^* and β_j^* are related to another trend situation then parameters α_j and β_j in (2). Trend T_t is understood to be a dominant component and it is necessary to do its description, e. g. analytical, mechanical or another smoothing. A formation of model with hidden periodicity with changing trend is based on an idea of easy transformation of such a time series to a time series with constant trend. Next steps are same like at model (2).

Describing periodical behaviour of a time series by the model with hidden periodicity it is necessary to determine the model variance which enables to review a quality of the model. Furthermore, it is important for formation and analysis of the periodogram. This variance can be defined as:

$$var\ \hat{Y}_t = \frac{1}{2} \sum_{j=1}^{m} \left(a_j^2 + b_j^2 \right) = \frac{1}{2} \sum_{j=1}^{m} A_j^2. \tag{4}$$

Regression and correlation analysis are bases for a determination of coefficient of determination value. It informs how many percent from the whole variability is possible to explain by the used model with hidden periodicity. We can define it as:

$$I^2 = \frac{var\ \hat{Y}_t}{var\ y_t},$$

whereas var y_t is the variance of empirical values y_t.

The conception of modelling of periodical variations in time series, mentioned above, supposes the respective period (or periods) is known and significant. However, this hypothesis is usually not accepted in practice because of complicated economical reality and the result cannot be gotten intuitively or by using previous experiences often. Then, the most important question is to find out if the time series ever contains a significant periodical component. This is a very serious problem because of making very important decisions based on a correct identification of periodicity of the researched process.

2.2 Periodogram Analysis

The periodogram analysis occupies an important position among objective methods of a research of periodicity. It serves to a detection of significant periodical components in time series. The periodogram is usually expressed by a chart or table. The method is very difficult, but it is possible to get results easy by using of any statistical software. $I\left(\omega_j\right)$ periodogram of the time series $y_1,\, y_2,\, \ldots,\, y_n$ is a view of model variances values expressed by the oscillation A_j amplitude. And it is defined as a function of the angular frequency:

$$I\left(\omega_j\right) = \frac{1}{2} \left(a_j^2 + b_j^2 \right) = \frac{1}{2} A_j^2, \tag{5}$$

where $a_j = \frac{2}{n} \sum_{t=1}^{n} y_t \sin \omega_j t$, $j = 1,\, 2,\, \ldots,\, H$,

$$b_j = \frac{2}{n}\sum_{t=1}^{n} y_t \cos \omega_j t, j = 1,\, 2,\, \ldots,\, H.$$

In specialized publications it is possible to find other forms of an explicit formulation of the periodogram than (5). E. g. statistical programme Statgraphics used in next chapters of this paper multiplies furthermore particular j-th terms of the periodogram by number of observations n. But such modifications have an insignificant influence on forming of the periodogram and interpretation of results.

In reference to the fact we do not distinguish significant and insignificant frequencies by construction of the periodogram we have to determine which periodogram components we can consider to be significant for an explanation of total magnitude of variance of the y_t variable. Significant periods are signed to be a holder of periodicity in behaviour of the researched economic indicator. We have a lot of tools for searching of local extremes of the periodogram but the most popular is Fishers test or Siegels test. Exactly, tests of periodicity are objective methods how to determine on significance of particular periodical components.

2.3 Fishers Test

Fishers test is based on testing the null hypothesis $H_0 : y_t = \varepsilon_t$. The alternative hypothesis is defined as $H_1 : non\ H_0$. The null hypothesis specifies that $y_1,\, y_2,\, \ldots,\, y_n$ is a sequence of independent random variables with normal distribution with mean equal to 0. It means the time series does not contain a periodical component. The alternative hypothesis indicates the time series contains a significant periodicity.

The test uses periodogram values $y_1,\, y_2,\, \ldots,\, y_n$ calculating for each ω_j frequency, $j = 1,\, 2,\, \ldots,\, H$. If the null hypothesis is true, no periodogram value should be significantly greater than others. First, periodogram values are standardized to the form:

$$\eta_j = \frac{I\left(\omega_j\right)}{\sum_{j=1}^{H} I\left(\omega_j\right)}, j = 1,\, 2,\, \ldots,\, H \tag{6}$$

The test statistics is defined as:

$$W = \max_j \eta_j, j = 1,\, 2,\, \ldots,\, H. \tag{7}$$

The null hypothesis is rejected at significant level α if $W > g_F\left(\alpha\right)$, where $g_F\left(\alpha\right)$ represents a Fishers test critical value at α. Critical values $g_F\left(\alpha\right)$ are tabulated for various H, e. g. in [6]. If the greatest periodogram value is significant, we can continue with testing of significance of the second greatest periodogram value etc. The principle of testing is same but we leave out the greatest periodogram value and H is about one unit smaller. It was found out that the power function of Fishers test becomes smaller in the case of composed periodicity. It is an event where a time series contains several significant periods. E. g. Siegels test, a modification of Fishers test, is considered to be more suitable. The test statistics is different (in detail e. g. in [7]) and the critical values for Siegels test are tabulated, e. g. in [7]. In reference to the null hypotheses formulations for the both tests the empirical data should not contain a trend because its presence depresses a sensitivity of these tests. Therefore the elimination of a trend is recommended. Statistical software usually does it as well.

2.4 Periodogram Analysis of Inflation Rate in the Czech Republic in 2003 – 2007

For an application of the methods noticed previously there were used monthly data about inflation rate in % in 2003 – 2007 in the Czech Republic. Inflation rate is one of the most important indicators of each country. This indicator is recently highly followed in the Czech Republic because of its considerable increase during the last few months. The inflation rate in January 2007 was 1,3% and assumed the highest value in January and February 2008, namely 7,5%. Now its value is decreasing since March 2008. The periodogram analysis can provide an idea about a regularity of development of this indicator and then it is possible to make an idea about the development of the indicator in future.

The periodogram analysis was done which provides both results in table form (see Table 1) and graphic outputs (see Figure 1).

Frequency	Ordinate	Frequency	Ordinate	Frequency	Ordinate
0,0166667	10,7537	0,183333	1,51176	0,35	0,435943
0,0333333	12,777	0,2	2,24582	0,366667	0,331248
0,05	34,193	0,216667	0,646831	0,383333	0,338161
0,0666667	4,00002	0,233333	0,294907	0,4	0,408515
0,0833333	3,64421	0,25	0,563333	0,416667	0,130456
0,1	0,625176	0,266667	1,40096	0,433333	0,0804686
0,116667	2,85175	0,283333	0,880338	0,45	0,477748
0,13333	6,08561	0,3	0,347158	0,466667	1,08708
0,15	3,12663	0,316667	0,490659	0,483333	0,225445
0,166667	1,993	0,333333	0,343	0,5	0,216

Table 1: Periodogram of inflation rate (in %) in the Czech Republic in 2003 – 2007

Periodogram values are calculated for all the frequencies ω_j, $j = 1, 2, \ldots, H$. It is necessary to determine which of them are significant. Just the periods corresponding to such frequencies are periodicity holders in the time series. We use Fishers test. The greatest periodogram value is 34,193 (for j = 2). The period 20 months corresponds to this value. According to (6) $\eta_3 = \frac{34{,}193}{92{,}506} = 0,3696301$. We can find $g_F\,(0,01) = 0,214669$ for $H = 30$ in tables of F distribution critical values. Because of $W = 0,3696301 > 0,214669$ the null hypothesis is rejected. The presence of the 20-month periodicity was proved at the 0,01 level of significance. We continue with testing of the second greatest periodogram value 12,777 (for $j = 2$) which corresponds to the 30-month period. $\eta_2 = \frac{12{,}777}{58{,}318} = 0,2191109$, the critical value $g_F\,(0,01)$ for $H = 29$ is 0,248852. This critical value was determined by a linear interpolation. Because of $W = 0,2191109 < 0,248852$ the null hypothesis is accepted. Thus the 30-month period is not statistically significant. Therefore Fishers test will be finished because the time series does not contain a composed periodicity. To summarize the analysis we can say that there is the only one significant period in the time series of inflation rate in 2003 – 2007. The length of the periodicity is 20 months.

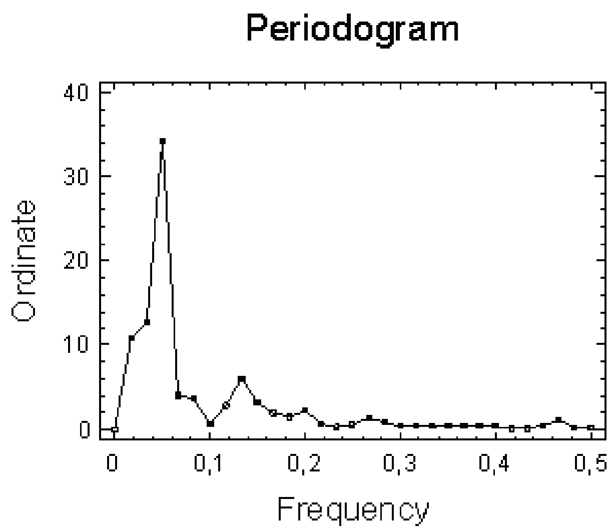


Figure 1: Periodogram of inflation rate (in %) in 2003 – 2007

In the end we can say that an application of selected harmonic analysis methods on inflation rate time series in the Czech Republic in 2003 – 2007 approved that this approach gives very important information about time series periodicity. Obtained results are easily interpretable and represent a significant benefit to a whole view on a character and a development of researched time series.

References

[1] ANDĚL, J.: *Statistická analýza časových řad.* SNTL, Praha, 1976. L11-B3-IV-41f/11740

[2] CIPRA, T.: *Analýza časových řad s aplikacemi v ekonomii.* SNTL/Alfa, Praha, 1986. ISBN 99-00-00157-X

[3] HINDLS, R.; KAŇOKOVÁ, J.; NOVÁK, I.: *Metody statistické analýzy pro ekonomy.* Management Press, Praha, 1997. ISBN 80-85943-44-1

[4] KOZÁK, J.; HINDLS, R.; ARLT, J.: *Úvod do analýzy ekonomických časových řad.* VŠE, Praha, 1994. ISBN 80-7079-760-6

[5] PRIESTLY, M. B.: *Spectral Analysis and Time Series.* Academic Press, London, 1983. ISBN 978-0-12-564922-3

[6] SEGER, J.; HINDLS, R.: *Statistické metody v tržním hospodářství.* Victoria Publishing, Praha, 1995. ISBN 80-7187-058-7

[7] SIEGEL, A. F.: *Testing for periodicity in a time series.* J. Amer. Statist. Association, Vol. 75, pp. 345-348, 1980.

[8] `http://www.czso.cz`

Jan Hanousek, Evžen Kočenda

CERGE-EI, Charles University

Politických vězňů 7, POB 882, 111 21 Praha

e-mail: `jan.hanousek@cerge-ei.cz, evzen.kocenda@cerge-ei.cz`

Intraday Effect of Macroeconomic Announcements on CEE Financial Markets

Abstract

We estimate the impact of macroeconomic news on composite stock returns in three emerging European Union (EU) financial markets (Budapest, Prague, and Warsaw), using intraday data and macroeconomic announcements. Our contribution is twofold. We employ a larger set of macroeconomic data releases than used in previous studies; due to the fact that most of the local news were not announced during active trading, we concentrate on the effect of foreign macroeconomic news (US and EU) on local stock returns. Composite stock returns are computed based on five-minute intervals (ticks) in Budapest (BUX), Prague (PX-50), and Warsaw (WIG-20. Macroeconomic news are measured based on the deviations of the actual announcement values from their expectations. We find that all the three emerging EU stock markets are subject to significant spillovers either directly via composite index returns or indirectly through the transmission of macroeconomic news. EU-wide announcements have the biggest impact on the Budapest market, followed by the Prague market, while they have no impact on the Warsaw market. Overall, our evidence indicates that the impact of macroeconomic news on stock returns may be sensitive to institutional structures of the markets studied as well the presence of foreign investors. We also discuss the implications of the findings for financial stability in the three emerging European markets.

Keywords*stock markets, intraday data, macroeconomic announcements, European Union, volatility, excess impact of news*

1 Introduction, Motivation and Related Literature

Modern stock market research draws attention to the use of intraday data that are able to reveal the effect of macroeconomic announcements on stock market movements. In this paper we analyze the impact of macroeconomic announcements on stock market returns using the intra-day data from the most liquid, three emerging European Union (EU) markets: the Czech Republic, Hungary, and Poland. Our paper contributes to the related literature on several ways. To our knowledge, there are no studies investigating the direct impact of macroeconomics news on emerging EU stock markets using intraday data. Previous studies focus on advanced stock markets, especially on German, U.S., and U.K. stock markets. The only study that is closely related to ours is a recent work by [9] who analyze the behavior of volatilities around ten important scheduled U.S. macroeconomic announcements on stock markets in several world regions, including the Czech Republic, Poland, Hungary, Slovakia, and placeRussia.

Using cross-sectional monthly data, they find that these transition markets as a group are not affected by placeU.S. announcements. Monthly data they use may not capture the true spillovers effects of the foreign news announcements on local markets. We instead provide time series evidence using intraday data.

Second, the majority of the studies focus only on a few macroeconomic announcements. In particular, most of them analyze only one event, namely the impact of monetary policy news on stock returns.[1] However, if there are other major announcements in the same time frame, then focusing only on monetary policy or only a few announcements may bias the estimated coefficients and hence may explain the poor performance of macroeconomic announcements in explaining asset returns. This paper uses the most comprehensive domestic macroeconomic news. To our knowledge, [4] have so far employed the most comprehensive set of local announcements, which include 17 placeU.S. macroeconomic data releases. We use a larger set of five different types of news releases than used in previous studies.

Third, previous studies mainly focus on local macroeconomic announcements. Investors may also react to foreign macroeconomic announcements if there is significant trade and financial linkages or institutional arrangements between countries.[2] This limited evidence indicates that foreign news arrival may have important implications for stock valuation. Therefore, besides local news, we use regional and foreign macro news. In particular, we use news from other emerging EU markets, Germany, and the placeU.S. These sets of news is interesting because Germany, apart from being important trading partner, is part of the European monetary union, so news from Germany may affect the emerging EU markets substantially as the latter strive to enter the union as soon as possible

Fifth, previous studies tend to investigate the impact of macro news only on conditional returns, assuming that stock returns do not exhibit time-varying volatility. In this study, we model both conditional returns and the conditional variance of returns simultaneously in a time-varying (SUR-GARCH) framework to better capture the impact of macroeconomic announcements of stock returns and stock market volatility in the three emerging EU markets.

Finally, our paper is related to the literature on cross-market correlations between stock markets (e. g., [1]) Our findings can shed light on the question of whether the source of any observed correlations among stock markets is due to economic fundamentals contained in public announcements, rather than "contagion". For example, if U. S. macroeconomic announcements have an explanatory power in the variations of stock prices in the emerging EU markets, this would suggest that economic fundamentals are responsible for the co-movements between the placeU. S. and the emerging EU markets. Limited studies summarized above on Central and European stock markets do not provide comprehensive evidence regarding cross market linkages between these markets and the rest of the world.

2 Data and Methodology

We analyze impact of news on stock returns[3] in the new EU stock markets where we concentrate on stock exchanges in Budapest, Prague, and Warsaw in particular. These markets are

[1] These studies include among others [6], [7], [12] and [5]

[2] Few studies that investigate impact of both local and foreign announcements on stock market returns are [9] and [1].

[3] Returns are stationary series. Results of stationarity tests not reported.

the largest European emerging markets in terms of market capitalization as well as the most liquid ones.

3 Data Set: Stocks and News

We constructed our dataset from intraday data recorded by Bloomberg for the stock markets of three emerging EU markets. Stock exchange index quotes are available in five-minute intervals (ticks) for stock indices at the stock markets in Budapest (BUX), Prague (PX-50), and Warsaw (WIG-20). The time period of our data starts on the 2nd of June 2003, at 9:00 and ends on the 31st of December 2006 at 16:00 Central European Daylight Time (CEDT). Descriptive statistics of the stock indices are presented in Table 1.

Further, we compiled an extensive data set on the macroeconomic announcements (news) in the above markets as well as announcements in the EU and US. The announcements are defined in the following way. There is news i in the form of various macroeconomic releases or announcements that are known ahead to materialize on specific dates t.[4] The extent of such news is not known but expectations on the market from its forecast whose values are factored in. Thus, the impact of such news appears through their difference from market expectations than from the extent of the news. Therefore, we define the news in our data set from the *excess impact* perspective. Formally this type of news is labeled as xn_{it} and defined as $xn_{it} = (sn_{it}) - E_{t-1}[sn_{it}])$, where sn_it stands for value or extent of the scheduled news or announcement. For the purpose of our analysis we collected macroeconomic news for which there exists a Bloomberg or Reuters survey including clearly defined calendar of releases with defined timing of news as well as their expectations. Among these we distinguish five categories of announcements according to the following characteristics: nominal aggregates, real economy, monetary policy, fiscal policy, and economic confidence measures.

From the practical perspective, we consider the new announcement within 5 minutes after its release and account for its effect for another 10 minutes. Following the description above, we differentiate positive (negative) impact if the announcement is above (below) market expectations and zero impact if it is in line with the market. The time difference between the markets is accounted for by setting the CEDT time for all news releases, which eliminates the time difference between the US and continental Europe.

3.1 Estimation Methodology

As behavior on these markets has been documented to follow periods of lower and higher activity during a trading day, our analysis is performed by using periods of lower volatility during standard trading period and separating periods of high volatility during the first 15 minutes after the opening and last 5 minutes before closing of the trading session.[5] This approach avoids mixing periods of varying volatility during the trading day and reflects the U-shape pattern documented for volatility in various markets including the three emerging markets under research (see [3]). Further, we allow for various types of news to impact the value of the stock indices, as well as for spillover effect from remaining new EU markets under research.

[4]There is also news i in the form of unexpected announcement that can be understood as truly exogenous shock or surprise appearing at time t. The number of such news that are recorded is negligible and we do not consider them in the present study.

[5] These times were chosen based on sensitivity analysis. Further adjustments did not change results.

To test empirically for stock market volatility, we employ an augmented generalized autoregressive conditional heteroskedasticity (GARCH) model due to [2]. Further, we augment the mean specification by the parameters to account for the effect of market spillovers and macroeconomic news in the form of the deviations of scheduled releases from market expectations. Purposely we use a GARCH-in-mean (GARCH-M) specification which includes a conditional variance in the mean equation so that we can analyze the process with the path-dependent rather than the zero-conditional mean.

Thus, to reflect preceding description the mean and volatility specification of the GARCH-M model is specified in the following form:

$$\Delta SI_{i,t}^{E} = \sum_{i=1}^{2}\sum_{j=1}^{p} \pi_i \Delta SI_{i,t-j}^{E} + \sum_{i=1}^{2}\sum_{j=1}^{q} \gamma_i xn_{i,t-j} + \theta \ln h_{i,t}^{E} + \varepsilon_t \tag{1}$$

$$h_{i,t}^{E} = \omega + \sum_{m=1}^{k} \alpha_m \varepsilon_{t-m}^2 + \sum_{m=1}^{l} \beta_m h_{i,t-m}^{E}. \tag{2}$$

The variables used are coded as follows. The parameter ΔSI_t^E is return on a specific emerging market stock index (Budapest, Prague, Warsaw), and *xnit* is vector of news as defined above. Coefficients π capture the effects of market spillovers and coefficients γ capture the contemporaneous effects of the news on the stock indices. The numbers of lags p, q, k, and l are chosen by the lag selection information criteria. The log of the conditional variance in the mean equation, i.e. $\ln h_t^E$, allows for an exponential rather than quadratic effect of observed volatility.

In the above specification, the ARCH term, i.e., $\alpha\varepsilon_{t-1}^2$, primarily reflects the impact of news or surprises from previous periods that affect stock price volatility. A significant and positive value for α that is less than one characterizes the extent to which shocks do not destabilize volatility. When α is greater than one, shocks from the past are destabilizing. The GARCH term, i.e.,βh_{t-1}^E, measures the impact of the forecast variance from previous periods on the current conditional variance, or volatility. Hence, a significant value for β that is close to one indicates a high degree of persistence in stock price volatility. The sum of both coefficients, i.e., α plus β, indicates the speed of convergence of the forecast of the conditional volatility to a steady state. The closer its value is to one, the slower is the convergence.

To avoid the risk of overestimating volatility, we do not impose the normality condition on distribution of errors. Rather, we allow for the generalized error distribution (GED) following [8]. The volatility of the stock prices is likely to follow a leptokurtic data distribution that is reflected by an actual GED parameter considerably lower than 2, which is the value in the case of normal distribution. Leptokurtosis implies that daily stock price volatility tends to concentrate around the mean during tranquil market periods but that shocks to volatility are large during turbulent times.

The above specification accounts for the effect of various types of news on the firms' market value, hence the value of the market index. The emerging European stock markets are documented to be influenced by EU news but also by U.S. macroeconomic announcements at 14:30 CEDT and by the opening of the U.S. stock markets at 15:30 CEDT. The news from these two regions is hypothesized to exhibit most direct influence on the new EU stock markets. The specification also accounts for the spillover effects through lagged indices of neighboring stock markets. Since trading hours in different markets span over different time periods we treat this difference by estimating the set of mean equations for each of the three emerging markets as seemingly unrelated regressions (SUR). This way we also allow for common features in the CEE markets to affect our estimates. The coefficients of the volatility equation are estimated via maximum likelihood estimation.

4 Empirical Findings

Our empirical findings are presented in Tables 2-4 for each of the three countries separately. They show considerable spillover effects among the new EU markets and the news impacts the index returns in general. However, the impact differs with respect to the origin of the news as well as impacted stock market. Majority of the local macroeconomic news of essential importance (e.g. GDP, inflation, etc.) in the Czech Republic, Hungary and Poland is announced before the trading session or at its very beginning. This institutional arrangement means that at the opening the markets already reflect the announcements from a large extent. For this reason we report results of specifications that include only foreign news as local news turned out to be statistically insignificant.[6] In case of the US news we do not include the news on monetary and fiscal policy measure as these are not released during the time when stock markets in the continental Europe are in trading session. In any event, foreign news and announcement seem to have stronger impact than the local ones. We first review the results for each market separately.

4.1 Czech Republic

The returns on the Czech stock index PX-50 are affected by the developments of both Hungarian and Polish indices but the effect is small (Table 1). Announcements originating in the EU affect the returns only at the segment of news related to the consumer, business, investment and economic confidence. The U.S. announcements exhibit a similar impact as only economic confidence related releases are statistically significant. The impact of the EU confidence news is negative, while that of the US confidence news is positive. This is somewhat puzzling since given a positive sign, interpretation is simple: upward movement of the return follows good news and vice versa. Overall, the economic confidence related news represents a strong signal on the Czech stock market and the effect of the foreign news has to be paired with the heavy presence of the foreign investors who tend to put more weight on the foreign news.[7] Path dependent behavior of the returns is confirmed by the significant coefficient of the conditional variance in the mean equation. The impact of the lagged volatility should not be overestimated as its magnitude is smaller than that of the combined effect of the news.

In our specification, the ARCH term, i.e., $\alpha\varepsilon_{t-1}^2$, reflects the impact of news or surprises from previous periods that affect stock price volatility. We find the relatively small coefficient α, meaning that news are well accounted for in the mean equation, past news do not affect volatility in a great extent, and they do not destabilize volatility. On other hand, the β coefficient in the GARCH term, i.e. βh_{t-1}^E, is quite large and indicates that no matter how

[6] Results are available upon request.

[7] Local Czech announcements during the day usually do not reach the magnitude of those made public before the trading starts and this is evidenced by the insignificant coefficients (not reported).

Parameter(coeff.)	Estimate	Std. error	t-statistic	p-value
Mean equation				
D_BUX(-1) (π_1)	0,002***	0,000	8,465	0,000
D_BUX(-2) (π_2)	0,001***	0,000	4,316	0,000
D_BUX(-3) (π_3)	0,001***	0,000	3,757	0,000
D_WIG20(-1) (π_7)	0,007***	0,001	6,679	0,000
D_WIG20(-2) (π_8)	0,006***	0,001	5,565	0,000
D_WIG20(-3) (π_9)	0,003***	0,001	2,885	0,004
US_nominal (γ_1)	-0,002	0,016	-0,146	0,884
US_real economy (γ_2)	0,005	0,014	0,358	0,721
US_econ.confidence (γ_3)	0,167**	0,071	2,362	0,018
EU_nominal (γ_4)	0,018	0,030	0,601	0,548
EU_real economy (γ_5)	-0,027	0,022	-1,215	0,224
EU_monetary policy (γ_6)	-0,234	0,170	-1,379	0,168
EU_fiscal policy (γ_7)	-0,030	0,459	-0,065	0,948
EU_econ.confidence (γ_8)	-0,059*	0,037	-1,604	0,100
ln ht (θ)	0,022***	0,006	3,697	0,000
Volatility equation				
Intercept (ω)	0,001***	0,000	1,761	0,078
ARCH term (α)	0,026***	0,008	3,294	0,001
GARCH term (β)	0,973***	0,008	115,162	0,000

*Table 1: Czech Republic: Impact of News on Stock Market (Note: ***, **, and * denote significance at 1, 5, and 10%.)*

large or small the effect of the news on volatility is, the impact of the forecast variance from previous periods on the current conditional variance, or volatility, is considerably persistent. Finally, sum of both coefficients close to one, i.e., α plus β, indicates that convergence of the conditional volatility to a steady state is very slow.

4.2 Hungary

The Hungarian stock index exhibits considerable spillovers from both neighboring markets, with effect of the Prague market being stronger than that of Warsaw. The latter effect is more stable in time, however (Table 2). Hungarian index is strongly impacted by the EU fiscal policy announcements. This is an interesting finding since Hungary is known to be having problems with fiscal discipline for considerable time and its future entry to the Eurozone is far from settled due to lack of fiscal convergence to the EU standards. The Budapest market also reacts very strongly to the US news related to real economy. Presence of the foreign investors on Hungarian market is somewhat smaller than in the Czech Republic, but still significant enough for the substantial foreign news effect to materialize as the above results suggest.

Path dependence of the returns with respect to volatility is not confirmed as the volatility coefficient in the mean equation is insignificant. Further, the past news affects volatility in a moderate extent and they are not destabilizing. The volatility of the Hungarian stock index exhibits lower persistence than the Czech one but higher than Polish index. Finally, the speed of convergence of the forecast of the conditional variance to a steady state is very slow.

4.3 Poland

Polish stock index is moderately affected by spillovers from Prague and only a little from Budapest. On other hand, the effect of Budapest market is more stable in time. From the perspective of the news transmission, the Polish stock market represents the least affected in the region under research (Table 3). There is no effect of the US news. The EU announcements related to nominal macroeconomic aggregates exhibit moderate negative effect. Other announcements are not found to leave a trace on the developments of the returns as none of the associated coefficients was significant. This finding can be explained by the fact the Polish stock market has the smallest participation of the foreign investors, only about one third, among its regional counterparts. The rest of the market is captured by the local traders and pension funds. We conjecture that for this reason the foreign announcements are not processed with the highest impact as the foreign investors are a minority.

Path dependent behavior of the returns with respect to the own volatility is not present as the coefficient of the conditional variance in the mean equation is insignificant. In terms of volatility alone, we find that past news does affect volatility to a largest extent among all three markets, but they are not destabilizing. On other hand, the volatility of the Polish stock index exhibits the lowest persistence with respect to its regional counterparts. Similar to the Prague and Warsaw indices, the speed of convergence of the forecast of the conditional variance to a steady state is very slow.

5 Concluding remarks

We estimate the impact of macroeconomic announcements on stock market returns and volatility in the most liquid three emerging European Union markets: the Czech Republic, Hungary,

Parameter(coeff.)	Estimate	Std. error	t-statistic	p-value
Mean equation				
D_PX50(-1) (π_4)	0,509***	0,120	4,239	0,000
D_PX50(-2) (π_5)	0,137	0,119	1,145	0,252
D_PX50(-3) (π_6)	0,299***	0,135	2,208	0,027
D_WIG20(-1) (π_7)	0,146***	0,024	6,008	0,000
D_WIG20(-2) (π_8)	0,100***	0,024	4,201	0,000
D_WIG20(-3) (π_9)	0,079***	0,024	3,358	0,001
US_nominal (γ_1)	-0,045	0,426	-0,105	0,917
US_real economy (γ_2)	0,461*	0,279	1,654	0,098
US_econ.confidence (γ_3)	-0,107	1,278	-0,084	0,933
EU_nominal (γ_4)	0,378	0,379	0,997	0,319
EU_real economy (γ_5)	-0,185	0,412	-0,450	0,652
EU_monetary policy (γ_6)	-2,733	4,169	-0,656	0,512
EU_fiscal policy (γ_7)	8,143**	4,154	1,960	0,050
EU_econ.confidence (γ_8)	-0,484	0,576	-0,841	0,400
ln ht (θ)	0,003	0,004	0,618	0,536
Volatility equation				
Intercept (ω)	1,091***	0,190	5,749	0,000
ARCH term (α)	0,088***	0,008	11,058	0,000
GARCH term (β)	0,911***	0,008	113,956	0,000

*Table 2: Hungary: Impact of News on Stock Market (Note: ***, **, and * denote significance at 1, 5, and 10%.)*

Parameter(coeff.)	Estimate	Std. error	t-statistic	p-value
Mean equation				
D_BUX(-1) (π_1)	0,006***	0,001	7,315	0,000
D_BUX (-2) (π_2)	0,003***	0,001	4,201	0,000
D_BUX (-3) (π_3)	0,002***	0,001	3,092	0,002
D_PX50(-1) (π_4)	0,069***	0,016	4,427	0,000
D_PX50(-2) (π_5)	0,019	0,015	1,246	0,213
D_PX50(-3) (π_6)	0,028*	0,015	1,919	0,055
US_nominal (γ_1)	-0,070	0,098	-0,712	0,477
US_real economy (γ_2)	0,117	0,144	0,809	0,418
US_econ.confidence (γ_3)	-0,005	0,626	-0,008	0,994
EU_nominal (γ_4)	-0,179*	0,108	-1,658	0,097
EU_real economy (γ_5)	0,050	0,062	0,805	0,421
EU_monetary policy (γ_6)	-0,024	0,470	-0,051	0,959
EU_fiscal policy (γ_7)	-0,232	0,647	-0,358	0,720
EU_econ.confidence (γ_8)	-0,034	0,109	-0,316	0,752
ln ht (θ)	0,005	0,004	1,134	0,257
Volatility equation				
Intercept (ω)	0,155***	0,017	9,184	0,000
ARCH term (α)	0,124***	0,007	16,926	0,000
GARCH term (β)	0,860***	0,008	101,944	0,000

*Table 3: Poland: Impact of News on Stock Market (Note: ***, **, and * denote significance at 1, 5, and 10%.)*

and Poland. We use intra-day data and the *excess impact* approach, as well local and foreign news to provide more comprehensive evidence and reliable inferences on the impact of news in emerging markets. Our evidence indicates that the importance of macroeconomic news on stock returns may be sensitive to the institutional structures of the markets studied as well the presence of foreign investors. We find the impact of foreign news stronger in countries with a larger proportion of foreign investors.

We find that the deviations from expectations play an important role in the EU emerging stock markets. The finding validates the use of our *excess impact* approach, yielding important insights to arguments of [11] who claim that the "detachment" of monetary policy expectations and asset prices from the incoming economic news is partly related to the difficulties associated with measuring the surprise component of that news. Using *excess impact approach,* we attempt to reduce the difficulties of measuring "news" correctly.

Regarding the impact of foreign and local macroeconomic announcements on stock returns, news originating in the EU affects the returns in Prague, Budapest and Warsaw markets, while U.S. announcements have an impact on the Prague and Budapest markets. The results for the Czech stock market do reflect the composition of investors which consists of about 85% foreign investors. These market players put more weight on the foreign announcements which shows in the direct impact of U.S. news releases on the Prague stock index. This finding is consistent with previous studies that find a reduced role for macroeconomic news when composite indices are included in regressions. Local macroeconomic news does not have immediate (intra-day) effect on returns in all markets as the majority of scheduled announcements are released before the trading on the market begins. This however does not rule out that domestic news affect the markets between subsequent trading sessions.

In terms of the volatility effects, stock returns show strong path-dependency on its own conditional variance (volatility) which is highly persistent in the Prague market but insignificant in the Budapest and Warsaw markets. This may reflect the higher degree of financial spillovers effects to the Prague market originating from both EU and U.S. markets directly through the composite index movements and indirectly via macroeconomic announcements. News affects volatility as shown by the statistically significant ARCH term in the conditional variance equation. News has a larger effect on volatility in the Budapest and Warsaw markets than the Prague market, indicating that the latter market is calmer than the former markets. Perhaps, this is due to more informed trading in the Czech market due to dominant foreign investor presence.

Overall, we find that all the three new EU stock markets are subject to significant spillovers directly via the composite index returns from neighboring markets, or indirectly through the transmission of macroeconomic news. Impact of the EU-wide announcements depends on the news segment and varies across the three markets. Effect of the U.S. macroeconomic news is similar.

The results have therefore important implications for diversification and risk management strategies in these European emerging markets. As these countries prepare to enter the euro zone in the future, we expect that their financial markets will be more sensitive to macroeconomic shocks, especially those originating from the European Union. This means that investors should price this expected higher future volatility (risk) in investing these markets now. From a broader perspective, we find that the emerging markets in our sample countries seem to react similarly to macroeconomic news compared to those in advanced industrial markets. This finding suggests that the emerging European countries have made significant progress in terms of financial market development by successfully integrating their financial markets into the world economy and in a relatively short period. Hence, the results have implications for the other emerging economies in the region to speed up economic reforms. This is encouraging news for lagging economies in the region, especially for the candidate EU countries.

References

[1] ALBUQUERQUE, R. AND VEGA, C. (2006) *Asymmetric Information in the Stock Market: Economic News and Co-movement, CEPR Discussion Paper* no. 5598, C.E.P.R. Discussion Papers.

[2] BOLLERSLEV, T. (1986) *Generalized autoregressive conditional heteroscedasticity,* Journal of Econometrics 31: 307–327.

[3] ÉGERT, B. AND KOČENDA, E. (2007) *Interdependence between Eastern and Western European Stock Markets: Evidence from Intraday Data,* Forthcoming in Economic Systems.

[4] FLANNERY, M. J. AND PROTOPAPADAKIS A. A. (2002) *Macroeconomic Factors Do Influence Aggregate Stock Returns*, Review of Financial Studies, 15: 751–782.

[5] GURKAYNAK, R., SACK, B., AND SWANSON, E. (2004). *Do Actions Speak Louder than Words?: The Response of Asset Prices to Monetary Policy Actions and Statements*, Finance and Economics Discussion Series 2004-66, Federal Reserve Board, Washington, D.C.

[6] JENSEN, G. R., AND JOHNSON, R. R. (1995) *Discount Rate Changes and Security Returns in the US, 1962-1991*, Journal of Banking and Finance, 19: 79–95.

[7] JENSEN, G. R., MERCER, M. J., AND JOHNSON, R. R. (1996). *Business Conditions, Monetary Policy, and Expected Security Returns*, Journal of Financial Economics, 40: 213–237.

[8] NELSON, D. B. (1991) *Conditional heteroskedasticity in asset returns: A new approach. Macroeconomic News and Stock Valuation in Europe*, Journal of Multinational Financial Management, 14: 201-215.

[9] NIKKINEN, J. AND SAHLSTRÖM, P. (2004) *Scheduled Domestic and US Main Short Sterling Futures*, Journal of Risk 5(1): 59–74.

[10] NIKKINEN, J., OMRAN, M., SAHLSTRÖM, P. AND ÄIJÖ, J. (2006) *Global stock market reactions to scheduled U.S. macroeconomic news announcements*, Global Finance Journal, 17(1): 92–104.

[11] RIGOBON, R., AND SACK, B. (2006) *Noisy Macroeconomic Announcements, Monetary Policy, and Asset Prices,* NBER Working Paper No. W12420.

[12] SIKLOS, P. L. AND ANUSIEWICZ J. (1998) *The Effect of Canadian and U.S. M1 Announcements on Canadian Financial Markets: The Crow Years*, Journal of Economics and Business, 50: 49–65.

Jan Hanousek, Evžen Kočenda

CERGE-EI, Charles University

Politických vězňů 7, POB 882, 111 21 Praha

e-mail: jan.hanousek@cerge-ei.cz, evzen.kocenda@cerge-ei.cz

Potential of the State to Control Privatized Firms

Abstract

The privatization strategy in many transition economies involved the creation of a special government agency that administered state property during privatization programs as well as after the privatization was declared complete. The National Property Fund (NPF) was the agency in the Czech Republic. In many firms the state kept property long after the privatization was completed. We analyze the control potential of the state exercised through the NPF via the control rights associated with capital stakes in firms along with special voting rights provided by law. With respect to the capital stakes and extent of implied control we distinguish majority ownership, blocking minority ownership and legal minority ownership. However, we also analyze the extent state control through special mechanisms such as golden shares. Based on a complete data set on assets as well as the means of control in privatized firms our analysis shows that a substantial part of the private sector composed of medium- and large-sized companies privatized in the large-scale privatization were not truly in the private economy over the period 1994–2005. These firms became truly private only after the sale of the remaining shares possessed by the state, the liquidation of golden shares and the consequent decline of the state control potential. We conclude that for most of the 1994–2005 period, the state control potential was extensive and certainly larger than has been found by earlier research.

Keywords: *privatization, state control potential, golden share, strategic firms.*

1 Introduction and Motivation

At the beginning of the transformation process in Central Europe, privatization was largely considered the foundation of the entire transition process. The transfer of ownership rights was seen by most academics and policy makers as crucial for the efficient allocation of resources and economic growth [3]. Truly, transition economies carried out a major ownership transformation that made the share of the private sector in stocktickerGDP increase from extremely low levels to between 60% and 90% [2].

Privatization schemes in many European emerging economies involved the creation of a special government agency that assumed the role of the administrator of state property. The state often remained the ultimate owner of numerous firms long after privatization was concluded, a situation that has been documented for the Czech Republic [8]. In the Czech Republic, the National Property Fund (NPF) acted as the governmental administrator of the property that remained in the possession of the state. As the NPF was dissolved at the end of 2005, it is relevant to study the extent of state control potential in privatized firms from a long-term perspective. In this paper, we examine the dynamics of changes in the structure of ownership, the extent of state ownership and the potential for state control in privatized Czech firms over the period 1994–2005, i.e. from the end of privatization to the dissolution of the NPF.

The state control in privatized firms may have potentially negative effects due to several reasons. In [10], the authors argue that inferior performance (in terms of value maximization) of state firms can be expected because state officials are prone to impose on the management of state companies a variety of goals that are not necessary targeted to the economic efficiency. In [1] the authors argue that the state would choose to retain some ownership in order to raise its bargaining power for the later sale of the residual state property and to increase its price. This way a higher sale price was at the expense of corporate efficiency due to delayed restructuring. Further, direct or indirect state control can provide grounds for retaining or creating a monopoly. This poses a dilemma for the state as owner because a firm that is privatized with monopoly power can be sold for a higher price than if it is divided into units whereas on the other hand monopoly consitutes serious burden to efficiency in the market economy. Finally, the hardening of budget constraints (i.e. curtailing firms' access to formal or informal state subsidies) is widely accepted to have a positive effect on corporate restructuring. However, soft budget constraints can be practiced between state-controlled credit providers and firms while being hidden by less-than-evident links among firms and financial institutions under state control. In [4] the authors show that shortly after privatization, firms owned by the state in Central Europe are often unable to perform on par with private firms and are riskier to their lenders as they are less able to repay their debts.

In this paper, we employ a transparent method that enables us to quantify the extent of potential state control over the assets of privatized firms.. Based on the data analysis, we suggest taking into account the potential of the state to control the assets of privatized firms, since firms in the economy under such potential control cannot be considered truly private.

The rest of the paper is organized as follows. In section 2 we review the privatization process in the placeCzech Republic and in section 3 we introduce our data. Section 4 provides analysis of the direct, indirect, and combined control of the state over the firms in the NPF portfolio. In section 5 we provide an assessment of the extent of the state control in the economy. Conclusions are presented in section 6.

2 Privatization and State Ownership

A massive privatization program was administered in the placecountry-regionCzech Republic in the first half of the 1990s under three different schemes: restitution, small-scale privatization, and large-scale privatization. We only reiterate the main aspects of the Czech privatization that are relevant for this study since the process has been extensively described in the literature, e.g. [6] among others. The first two schemes began in 1990 and were most important during the early years of the transition. Large-scale privatization, by far the most important scheme, began in 1991, was completed in early 1995, and allowed for various privatization techniques. Small firms were usually auctioned or sold in tenders. Many medium-sized businesses were sold in tenders or to predetermined buyers in direct sales. Most large and many medium-sized firms were transformed into joint stock companies and their shares were distributed through voucher privatization (almost one-half of the total number of all the shares of all joint stock companies were privatized in the voucher scheme), sold in public auctions or to strategic partners, or transferred to municipalities.

The voucher scheme was part of the large-scale privatization process. Two waves of voucher privatization took place, in 1992–93 and 1993–94. Both waves were administered in the same manner and there were no differences in their set-up. During the scheme, a total of 1664 firms were privatized: 988 in the first wave and 676 firms in the second wave; from this number 185 firms were privatized in both waves in various proportions of their assets. All Czech citizens over the age of 18 who resided in the placeCzech Republic could participate in the voucher process. For each wave every eligible citizen was authorized to buy a voucher book that contained 1000 investment "points" for 1000 Czech crowns (CZK), about a week's wage. Before the privatization started, individuals had the option of assigning none, some, or all of their points to Privatization Investment Funds (PIFs): newly established financial firms vaguely similar in their scope of activities to closed-end mutual funds.

At the beginning it was the Ministry of Privatization that executed the privatization process. To administer the property that remained in the state's possession after the privatization, the NPF was established as a state institution that was entitled with legal power to exercise property rights over companies that were fully or partially owned by the state. By the end of the scheme in 1994 the NPF held on average about a 25% stake in privatized firms, but the extent varied greatly. The NPF was dissolved at the end of 2005 and the remaining agenda was transferred to the Ministry of Finance.

3 Data

We have assembled a large data set on the extent of ownership in a large sample of Czech firms over the period 1994–2005. The beginning of our data sample coincides with the end of the privatization schemes, e.g. the starting point from which such data is meaningful and available. The end of our sample coincides with the end of the NPF as an institution. The data come from the archives of the NPF, the former Ministry of Privatization, the Prague Stock Exchange, the Center for Securities in Prague, the commercial database Aspekt, and the Commercial Register of the placeCzech Republic. Details on the number of firms, the value of their assets, and the extent of the ownership stakes of the state are given below in Tables 1 and 2.

From our data we are able to isolate the specific extent of ownership held by the state as well as to distinguish various means of its direct and indirect control along with the amount of assets under control. In this respect we are able to trace the development of state control in a number of firms and also control over their assets over time.

4 Persistency of State Ownership and Extent of Control

4.1 Direct control

In the divisions of stakes that allow for effective control we follow [5]. According to their taxonomy, depending on the extent of the stake, different blockholders have under Czech law different opportunities to influence corporate governance. In particular, the law provides important rights of ownership and control for owners with majority ownership (more than 50% of shares), blocking minority ownership (more than 33% but not more than 50% of shares) and legal minority ownership (at least 10% but not more than 33% of shares).

Majority ownership grants the owner the right to staff management and supervisory boards, to alter and transfer firm assets and to make crucial strategic decisions at general shareholder meetings. Blocking minority ownership gives the right to block a number of decisions, such as those related to increasing or reducing assets and implementing major changes in business activities that the majority shareholder may strive to implement at the general shareholder meeting. Finally, legal minority ownership is potentially important because the law entitles the holder of this stake to call a general shareholder meeting and obstruct its decisions by delaying their implementation through lengthy court proceedings.

In Table 1 we present the value of total assets of firms in the NPF portfolio categorized according to divisions of stakes as outlined above. The total assets of the firm can be considered as a proxy for the size of each firm in the NPF portfolio. Hence, it allows inferring the extent of control over the large and important firms in the economy. This enables deriving a perception of the economic power of the companies and consequently the extent of wealth that is controlled by the state through direct ownership channels, e.g. via the amount of shares.

For each ownership category in Table 1 the value in the left column indicates the sum of the absolute value of the total assets of firms from a given category in a given year. The number of ownership categories (five) does not change over time but the number of firms in the portfolio differs from year to year. The overall value of a portfolio in a given year is the sum of the absolute values of the total assets of firms across the categories.

The evidence in Table 1 shows that the state kept tight control over the largest firms. We consider the state to fully control a firm if the share owned by the state is bigger than 50%. This way we prevent speculation about the proportion of state ownership sufficient to effectively control a firm. The absolute value of the total assets of the firms where the state kept more than a 50% but less than a 100% stake is larger than the absolute value of the total assets of firms in any other single category for any given year with the exception of 1995 and 1996. Even more important is the fact that the value of the assets in firms with more than a 50% stake held by the state (including 100%) was during 2002–2005 larger than the combined value of the assets of the firms in the remaining categories.

Year	Total	Less than 10%		[10% to 33%)		[33% to 50%)		[50% to 100%)		100%	
	absolute	absolute	relative	absolute	relative	absolute	relative	absolute	relative	absolute	relative
1994	353	104	29%	101	29%	33	9%	115	32%	1	0%
1995	579	224	39%	72	12%	71	12%	206	36%	7	1%
1996	490	181	37%	71	15%	92	19%	141	29%	6	1%
1997	408	121	30%	38	9%	105	26%	138	34%	7	2%
1998	350	98	28%	12	4%	101	29%	133	38%	6	2%
1999	330	94	29%	10	3%	79	24%	141	43%	5	2%
2000	315	88	28%	8	3%	75	24%	138	44%	5	2%
2001	289	69	24%	8	3%	72	25%	131	45%	9	3%
2002	270	66	24%	5	2%	52	19%	141	52%	5	2%
2003	229	56	24%	23	10%	35	15%	111	48%	5	2%
2004	185	52	28%	15	8%	0	0%	110	59%	8	4%
2005	136	49	36%	14	10%	0	0%	68	50%	5	3%

Note: Absolute values are expressed in billions of CZK

Table 1: Absolute Value of Total Assets of Firms in the NPF Portfolio

The reason behind this development is that the early post-privatization ownership structure emerged when shares from the second wave were distributed in early 1995 and rapid reallocation of shares across new owners took place in 1995–96 during the so-called "third wave" of privatization as new owners, including privatization funds and the state, reshaped their initial post-privatization portfolios of acquired companies.

Depending on the investor, the swapping of shares in 1995–96 was aimed at (a) portfolio diversification, (b) obtaining concentrated ownership in specific firms and industries or (c) achieving conformity with legal requirements to prevent excessive stakes being held by privatization funds. Investors, especially PIFs, engaged in direct swaps of large blocks of shares, and off-market share trading was common. The 1995–96 ownership changes were massive, unregulated and frequently unobservable to outsiders. The absolute value of the total assets of firms with a clear state majority decreased only very slowly, from 141 in 1996 to 110 in 2004. Also, the state kept a relatively stable portfolio of firms with 100% control and even strengthened its positions. In the categories of legal and blocking minorities the state has been decreasing its stakes in general, which is in line with the decreasing amount of total assets in these categories.

The analysis would be incomplete without inspecting the developments of the control over the firms' assets in proportions over time. In Table 1, the right column presents the relative value of the total assets of firms in each category of direct control as a proportion of the total for a given year. Similar to previous findings, extensive direct control by the state is confirmed. Further, relative control increases: as the portfolio reduces in absolute terms, state control over the assets in firms increases. The reason is that state kept the largest and most important firms under its direct control for the entire period of economic transformation. One cannot avoid concluding that, despite the voucher privatization, the state maintained its influence over a significant part of the Czech economy.

4.2 Indirect control

State control over a firm may be ensured by various means. The simplest way of control is described in the preceding section: control through the number of shares held by the state that represent the associated voting rights. Another way is embodied in the "golden" share. The golden share was introduced by Act No. 210/1993, modifying Act No. 92/1991. The act set the conditions for property transfer from the state to others with the aim of protecting the special interests of the state in firms privatized in large-scale privatization. When the state sells its golden share, it gives up its rights in the company and the golden share ceases to exist.

The golden share, as an instrument in the form of a single share with a special status, allows the state to prevent any major changes in a company where the state holds such a share. Utility companies are a typical example of state control through the golden share. Further, a number of companies were declared strategic firms and enjoyed a special status that was embedded in related legal provisions. Similarly to the golden share instrument, in legally declared strategic firms the state was able to exercise greater control than would correspond to its ownership rights. In some companies the two instruments of indirect control were combined.

Table 2 shows the potential of indirect control measured by the size of the assets of firms under a particular form of indirect control. The amount of assets controlled via indirect instruments is substantially higher than the amount of assets of the firms without such control. In 2004, the state was able to control the assets of 124 billion CZK out of a total 185 billion CZK through indirect instruments. As the golden share or strategic classification was associated with companies vital to the economy, their independence from the state did not fully reflect reality. Despite being privatized, the firms and their assets were part of the public economy.

Year	Total	Gold		Strategic		Gold&Strategic		Standard	
		absolute	relative	absolute	relative	absolute	relative	absolute	relative
1994	353	Data not available							
1995	579								
1996	490								
1997	408	37	9%	177	43%	57	14%	138	34%
1998	350	36	10%	175	50%	56	16%	83	24%
1999	330	34	10%	165	50%	56	17%	75	23%
2000	315	22	7%	158	50%	56	18%	78	25%
2001	289	18	6%	151	52%	56	19%	64	22%
2002	270	19	7%	145	54%	43	16%	62	23%
2003	229	19	8%	118	51%	24	10%	69	30%
2004	185	17	9%	107	58%	0	0%	61	33%
2005	136	17	13%	65	48%	0	0%	53	39%

Note: Absolute values are expressed in billions of CZK.

Table 2: Absolute Value of Total Assets of Firms in the NPF Portfolio

4.3 Combined control

In order to evaluate the effective control power of the state over companies we combined all the feasible means of control together. Figure 1 illustrates the developments of the control potential through direct control, i.e. an ownership stake greater than 50% and the associated voting rights, together with indirect control via the golden share, strategic firm classification or both.

The combined data are in line with the previous findings. Figure 1 shows that the control potential in terms of the assets of the firms in the NPF portfolio is highest in 1997, when 68% of the assets of all firms in the NPF portfolio were controlled by the state. Assets worth 277 billion CZK out of a total of 408 billion CZK were in this group. The extent decreases slightly over time, however, a significant decline is visible only after 2002. In 2005, i.e. at the end of the relevant period, about 90 billion CZK out of a total 136 billion CZK is under combined control. Based on the shares in the portfolio, the relative ratio of assets under the control of the state was stable from 1998.

The size of the assets under combined control decreases over time. However, its magnitude document that medium and large companies privatized in the early 1990s remained under the control potential of the state dramatically longer than the state suggested. This is an outcome not specific only to the Czech Republic, though. In [7], the authors showed that, despite reforms, state intervention in firms' decisions continued well after privatizations were completed in many transition economies.

Another way the state can further increase its control in privatized firms is through pyramids or cross-holdings (see [9] for an overview). In the Czech setting this could happen through privatization investment funds (PIFs) or banks. Due to this, the numbers presented in the preceding sections may actually underestimate the state control in the economy. However, at the moment, lack of data prevents us from analyzing this interesting issue further.

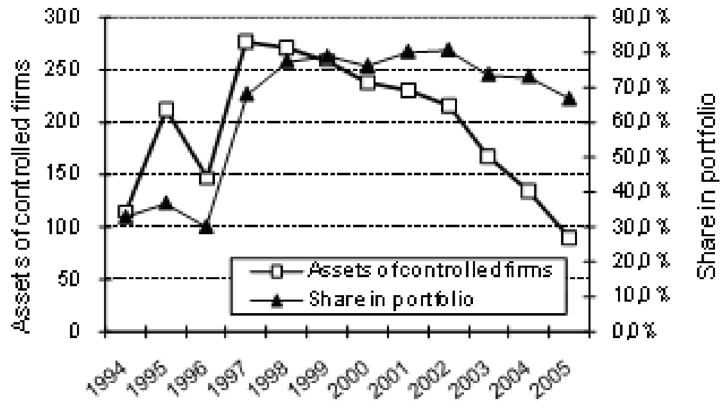


Figure 1: Combined control potential: Absolute and relative values of total assets of firms in the NPF portfolio (in billions of CZK)

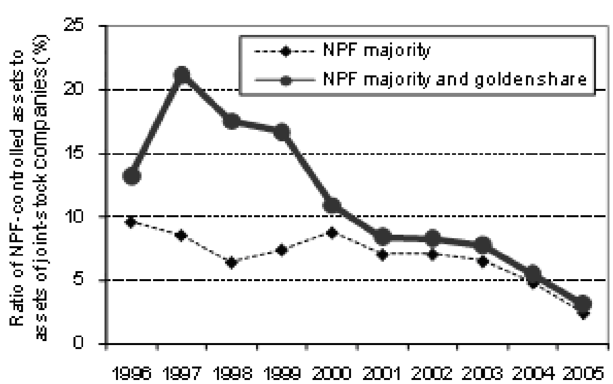


Figure 2: Ratio of assets of firms controlled by NPF (majority and golden share) to assets of joint-stock companies in the placeCzech Republic

5 The Extent of the Control in the Economy

So far, we have focused on the extent of the control potential within firms in the portfolio of the NPF. In Figure 2, we present the proportion of the control potential relative to other firms not under the influence of the state. Year by year, we present the development of two ratios. The first ratio measures the ratio of the value of the assets of the firms in the NPF portfolio controlled by a share majority over the value of all joint stock companies in the Czech Republic. The second ratio measures the ratio of the value of the assets of the firms in the NPF portfolio controlled by a share majority or indirectly via a golden share over the value of all joint stock companies in the Czech Republic. Czech joint stock companies and the value of their assets thus serve as a proxy for the size of the economy. Since the data on the contributions of individual firms to stocktickerGDP are not available, we use the value of the assets of Czech companies as a supplementary measure of the size of the economy.

Figure 2 suggests that the peak of the control potential was reached in 1997, when the ratio of controlled and non-controlled assets jumped over 20%. The ratio declines over time; however, it still exceeds 10% in 2000. In the same year, the two ratios converge, which supports the tendency of the state to control firms primarily by the standard instrument of control, via shares.

Despite the substantial extent of control potential, it should not be understood as a gigantic network of day-to-day management. Rather, it is a real control potential in cases of important decisions about the economic development of firms, their assets or cash flows between firms and the state. In the same way, the extent of control should not be overemphasized from the viewpoint of firm performance. In [5], the authors document that over the period 1996–1999, which coincides with the period of strongest state control potential, the performance of firms privatized in the large-scale privatization was very similar to that of firms with state ownership.

6 Conclusion

We have analyzed the extent of the control potential of the state in Czech firms privatized in the mass privatization scheme. The property remained for an extended period in the hands of the state even after the privatization scheme was completed, administered by the National Property Fund (NPF), which was set up as a special agency to protect the economic interests of the state. More than a decade of data allows us to explore the behavior of the state in its position as an important owner and co-owner.

The state control in a firm may be exercised by various means. The simplest method of control is through the number of shares held by the state that represent the associated voting rights. Another way is embodied in the "golden" share. Such an instrument, in the form of a single share with a special status, allows the state to prevent any major changes in a company where the state holds such a share. Further, a number of companies were declared strategic firms and enjoyed a special status that was embedded in related legal provisions. Similarly to the golden share instrument, in legally declared strategic firms the state was able to exercise greater control than would correspond to its ownership rights derived from the extent of its share holdings. Also, in some companies the two instruments of indirect control were combined. All of the feasible means of control combined together allow an evaluation of the effective control power of the state over the companies. At its peak, nearly 60% of firms and close to 90% of the assets of the firms in the NPF portfolio were under the effective control potential of the state. The extent decreases slowly and only after 2002 the amount of assets of firms within the state control potential decreases faster.

To conclude, one can say that the extent of state control potential should not be overestimated. On the other hand, our analysis shows that a substantial part of the private sector composed of medium- and large-sized companies privatized in the large-scale privatization were not truly in the private economy over the period 1994–2005. These firms became truly private only after the sale of the remaining shares possessed by the state, the liquidation of golden shares and the consequent decline of the state control potential.

References

[1] BENNETT J., ESTRIN S., MAW J. (2005) *Why did transition economies choose mass privatization?* Journal of the European Economic Association, 3(2-3): 567–75

[2] EBRD (2006) Transition report 2006 - Finance in transiton. European Bank for Reconstruction and Developement, London

[3] ESTRIN S., HANOUSEK J., KOČENDA E., SVEJNAR J. (2007) *Effects of privatization and ownership in transition economies,* IPC Working Paper 30, International Policy Center, University of Michigan

[4] FRYDMAN R., HESSEL M., RAPACZYNSKI A. (2006) *Why ownership matters? Entrepreneurship and the restructuring of enterprises in central Europe.* In: Fox M., Heller M. (ed): Corporate governance lessons from transition economies, Princeton University Press

[5] HANOUSEK J., KOČENDA E., SVEJNAR J. (2007) *Origin and concentration: Corporate ownership, control and performance in firms after privatization.* Economics of Transition, 15(1): 1–31

[6] HANOUSEK J., KROCH E. (1998) *The two waves of voucher privatization in the Czech Republic: A model of learning in sequential bidding*, Applied Economics, 30(1): 133–143

[7] HELLMAN J., SCHANKERMAN M. (2000) *Intervention, corruption and capture: The nexus between enterprises and the state,* Economics of Transition, 8(3): 545–576

[8] KOČENDA E. (1999) *Residual state property in the Czech Republic,* Eastern European Economics, 37(5): 6–35

[9] MORCK R., WOLFENZON D., YEUNG B. (2005) *Corporate governance, economic entrenchment and growth.* Journal of Economic Literature, 43(3): 657–722

[10] SHLEIFER A., VISHNY R.W. (1994) *Politicians and firms.* Quarterly Journal of Economics 109(4): 995–102

Milan Hladík

Faculty of Mathematics and Physics
Charles University Prague
Malostranske namesti 25, 118 00 Praha 1

e-mail: milan.hladik@matfyz.cz

Tolerances in Portfolio Selection via Interval Linear Programming

Abstract

We consider a linear programming problem and develop an effective method for computing tolerances for input data. The tolerances are determined such that the input quantities can simultaneously and independently vary within these tolerances while the optimal value does not exceed given lower and upper bounds. In our approach we are able to take into account all the input quantities or some selected ones. The procedure runs in polynomial time. Although the tolerances are not the best possible (due to dependencies between quantities) in general, the results are satisfactory. We illustrate the procedure on a simple portfolio selection problem modelled as a linear program.

Keywords : *Portfolio selection, linear programming, generalized fractional programming, tolerance analysis, interval analysis.*

1 Introduction

The aim of this paper is to develop a tool for computing tolerances of possibly all input quantities of a linear programming problem. Our approach is based on interval analysis [1], so we first introduce some basic notations and theorems.

An interval matrix is defined as

$$\boldsymbol{A} := [\underline{A}, \overline{A}] = \{A \in \mathbb{R}^{m\times n};\ \underline{A} \le A \le \overline{A}\},$$

where $\underline{A}$, $\overline{A} \in \mathbb{R}^{m\times n}$, $\underline{A} \le \overline{A}$, are given matrices. By

$$A_c := \frac{1}{2}(\underline{A} + \overline{A}), \quad A_\Delta := \frac{1}{2}(\overline{A} - \underline{A})$$

we denote the midpoint and the radius of $\boldsymbol{A}$, respectively. Interval vectors are defined analogously.

Let an interval matrix $\boldsymbol{A} \subset \mathbb{R}^{m\times n}$ and interval vectors $\boldsymbol{b} \subset \mathbb{R}^m$, $\boldsymbol{c} \subset \mathbb{R}^n$ be given. By an interval linear programming problem we understand a family of problems

$$\min c^T x \ \text{ subject to } \ Ax \ge b, x \ge 0, \tag{1}$$

where $A \in \boldsymbol{A}$, $b \in \boldsymbol{b}$ and $c \in \boldsymbol{c}$. Interval linear programs of this or similar type were studied e.g. in [2], [3], [6], [8], [11]. Denote by

$$f(A,b,c) := \inf c^T x \ \text{ subject to } \ Ax \geq b, x \geq 0$$

the optimal value of the linear program, also infinite values are possible. The lower and upper bound of the optimal value is denoted, respectively, by

$$\text{“}f'' := \inf f(A,b,c) \ \text{ subject to } \ A \in \boldsymbol{A}, b \in \boldsymbol{b}, c \in \boldsymbol{c},$$
$$\overline{f} := \sup f(A,b,c) \ \text{ subject to } \ A \in \boldsymbol{A}, b \in \boldsymbol{b}, c \in \boldsymbol{c}.$$

These bounds are easily computable by two linear programs, as stated it the following theorem. For the proof see e.g. [6].

Theorem 1. *We have*

$$\text{“}f'' = \inf \underline{c}^T x \ \ \textit{subject to} \ \ \overline{A}x \geq \underline{b}, \ x \geq 0. \tag{2}$$

Let “f'' be finite or let the right-hand side of the equation (3) *be positively infinite. Then*

$$\overline{f} = \sup \overline{b}^T y \ \ \textit{subject to} \ \ \underline{A}^T y \leq \overline{c}, \ y \geq 0. \tag{3}$$

The formula (2) holds without any assumption. To use the formula for the upper bound we must be sure that there is no duality gap between primal and dual linear programs (i.e., at least one of them to be feasible). Therefore we have the assumption in Theorem 1. If this assumption is not satisfied, we can process according to [6]: If “$f'' = \infty$ then clearly also $\overline{f} = \infty$. If “$f'' = -\infty$ then check feasibility of the linear system of inequalities

$$\underline{A}x \geq \overline{b}, \ x \geq 0.$$

If it is feasible then (3) is true, otherwise $\overline{f} = \infty$.

2 Computing the tolerances

We use the previous results for solving the inverse problem: We are given a linear program and the bounds for the optimal value. The aim is to compute intervals for all the interested quantities such that all of them can simultaneously and independently vary inside their intervals while the corresponding optimal value does not exceed the prescribed bounds.

Formally, consider the linear program

$$\min c_c^T x \ \text{ subject to } \ A_c x \geq b_c, x \geq 0,$$

where A_c, b_c and c_c are known and fixed. By f^* denote its optimal value. Let A_Δ, be a nonnegative matrix representing scale of demanded perturbations. Its entries $(a_\Delta)_{ij}$ are usually set to ones (for absolute tolerances) or to $|(a_c)_{ij}|$ (for relative tolerances). If the quantity $(a_c)_{ij}$ is out of focus then we put $(a_\Delta)_{ij} = 0$. In the same manner introduce nonnegative vectors b_Δ and c_Δ.

Let “f” and $\overline{f}$ be lower and upper bound of the optimal value, respectively, and suppose that “f” $< f^* < \overline{f}$. We are seeking for a maximal $\delta > 0$ such that optimal value to (1) lies within the interval [“f”$, \overline{f}$] for all $A \in \boldsymbol{A}_\delta$, $b \in \boldsymbol{b}_\delta$ and $c \in \boldsymbol{c}_\delta$, where

$$\begin{aligned} \boldsymbol{A}_\delta &:= [A_c - \delta \cdot A_\Delta, A_c + \delta \cdot A_\Delta], \\ \boldsymbol{b}_\delta &:= [b_c - \delta \cdot b_\Delta, b_c + \delta \cdot b_\Delta], \\ \boldsymbol{c}_\delta &:= [c_c - \delta \cdot c_\Delta, c_c + \delta \cdot c_\Delta]. \end{aligned}$$

Theorem 2. *Let*

$$\delta^1 := \inf \delta \ \textit{subject to} \ -c_c^T x + \delta \cdot c_\Delta^T x + \text{“}f\text{''} \geq 0, \qquad (4)$$
$$A_c x - b_c + \delta \cdot \big(A_\Delta x + b_\Delta\big) \geq 0, \ x \geq 0,$$
$$\delta^2 := \inf \delta \ \textit{subject to} \ b_c^T y + \delta \cdot b_\Delta^T y - \overline{f} \geq 0, \qquad (5)$$
$$-A_c^T y + c_c + \delta \cdot \big(A_\Delta^T y + c_\Delta\big) \geq 0, \ y \geq 0,$$

Denote $\delta^* := \min(\delta^1, \delta^2) - \varepsilon$ with an arbitrarily small $\varepsilon > 0$. If the linear system

$$\big(A_c - \delta^* \cdot A_\Delta\big) x \geq b_c + \delta^* \cdot b_\Delta \qquad (6)$$

is feasible then $f(A, b, c) \in$ [“f”$, \overline{f}$] for all $A \in \boldsymbol{A}_{\delta^*}$, $b \in \boldsymbol{b}_{\delta^*}$ and $c \in \boldsymbol{c}_{\delta^*}$.

Proof. The value of δ^1 is defined such that the optimal value function $f(A, b, c)$ does not get over the lower bound “f” while perturbing the input data within $[0, \delta^1)$ tolerances. Let $\delta > 0$. According to Theorem 1, the minimal optimal value of the linear programs (1) over $A \in \boldsymbol{A}_\delta$, $b \in \boldsymbol{b}_\delta$ and $c \in \boldsymbol{c}_\delta$ is achieved for

$$\inf\, (c_c - \delta \cdot c_\Delta)^T x \ \text{ subject to } \ (A_c + \delta \cdot A_\Delta) x \geq b_c - \delta \cdot b_\Delta, \ x \geq 0. \qquad (7)$$

We are seeking for a maximal $\delta > 0$ such that the linear program (7) does not exceed the lower bound “f”. This is not an easy problem in general, so we transform this problem to the another yielding the same or smaller value: We compute a minimal $\delta > 0$ such that the linear program (7) has a feasible point with the objective value being at most “f” (since with the increase of δ the feasible set is expanding and the optimal value function is nonincreasing). Thus δ^1 can be obtained by solving the optimization problem

$$\delta^1 = \inf \delta \ \text{ subject to } \ c_c^T x - \delta \cdot c_\Delta^T x \leq \text{“}f\text{''}, \ \big(A_c + \delta \cdot A_\Delta\big) x \geq b_c - \delta \cdot b_\Delta, \ x \geq 0.$$

Nevertheless, the optimal value of (7) with $\delta = \delta^1$ can exceed the lower bound “f”. Therefore, we must subtract from δ^1 an arbitrarily small $\varepsilon > 0$.

The second part of the proof concerning δ^2 and the upper bound $\overline{f}$ is analogous and the condition (6) ensures that the formula (3) is applicable. Let $\delta > 0$. By (3), the maximal optimal value of the linear programs (1) over $A \in \boldsymbol{A}_\delta$, $b \in \boldsymbol{b}_\delta$ and $c \in \boldsymbol{c}_\delta$ is achieved for

$$\sup\, (b_c + \delta \cdot b_\Delta)^T y \ \text{ subject to } \ (A_c - \delta \cdot A_\Delta)^T y \leq c_c + \delta \cdot c_\Delta, \ y \geq 0, \qquad (8)$$

and we want to maximize δ such that this linear program does not exceed the upper bound $\overline{f}$. Again, we transform a problem and compute a minimal $\delta > 0$ such that the linear program (8) has a feasible point with the objective value being at least $\overline{f}$. Thus δ^2 will be achieved by solving the optimization problem

$$\delta^2 = \inf \delta \ \text{ subject to } \ b_c^T y + \delta \cdot b_\Delta^T y \geq \overline{f}, \ \left(A_c - \delta \cdot A_\Delta\right)^T y \leq c_c + \delta \cdot c_\Delta, \ y \geq 0.$$

□ □

Theorem 2 gives the formulae for computing the demanded tolerances. The value of resulting δ^* is not the best possible in general, but the cases when δ^* is not optimal are very rare.

Note that in the case when (6) is not satisfied, we can easily decrease δ^* such that the inequality system would be satisfied. The procedure is quite similar to the one proposed in the previous proof.

Note also that the optimization problems (4)–(5) belong to the class of so called generalized linear fractional programs. These programs have the form of

$$\sup \left(\inf_i \frac{P_{i\bullet} x}{Q_{i\bullet} x} \right) \ \text{ subject to } \ Qx > 0, \ Rx \geq r,$$

(where $P_{i\bullet}$ and $Q_{i\bullet}$ denotes i-th row of P and Q, respectively), or,

$$\sup \alpha \ \text{ subject to } \ Px - \alpha \cdot Qx \geq 0, \ Qx \geq 0, \ Rx \geq r,$$

and they are solvable in polynomial time using an interior point method [4], [12]. In the cases (4) and (5), the nonnegativity condition $Qx \geq 0$ is always satisfied, and therefore we omit it.

3 Application in portfolio selection problem

Interval analysis in portfolio selection problem was used e.g. in [7], [9], [10], [13]. Our approach is different; we exhibit results from the previous section to compute tolerances for the problem quantities.

Consider the following portfolio problem. We have J possible investments for T time periods and r_{jt}, $j = 1, \dots, J$, $t = 1, \dots, T$, stands for return on investment j in time period t. Estimated reward on investment j using historical means is defined as

$$R_j := \frac{1}{T} \sum_{t=1}^{T} r_{jt}.$$

In order to get a linear programming problem we measure risk of investment j by sum of absolute values instead of the historical variances:

$$\frac{1}{T} \sum_{t=1}^{T} |r_{jt} - R_j|.$$

Let μ be a risk aversion parameter (upper bound for risk) given by a user, and the variable x_j, $j = 1, \dots, J$, denotes a fraction of portfolio to invest in j. Then the maximal allowed risk is expressed by the constraint

$$\frac{1}{T}\sum_{t=1}^{T}\Big|\sum_{j=1}^{J}(r_{jt}-R_j)x_j\Big| \leq \mu,$$

or, by converting to linear inequality system

$$-y_j \leq \sum_{j=1}^{J}(r_{jt}-R_j)x_j \leq y_t,\ \forall t=1,\ldots,T,\quad \frac{1}{T}\sum_{t=1}^{T}y_t \leq \mu.$$

The portfolio selection problem takes the form

$$\begin{aligned}
&\max\ \sum_{j=1}^{J}R_jx_j\\
&\text{subject to } -y_j \leq \sum_{j=1}^{J}(r_{jt}-R_j)x_j \leq y_t,\quad \forall t=1,\ldots,T,\\
&\qquad\qquad \sum_{j=1}^{J}x_j=1,\ \frac{1}{T}\sum_{t=1}^{T}y_t \leq \mu,\\
&\qquad\qquad\qquad\qquad x_j \geq 0,\quad \forall j=1,\ldots,J.
\end{aligned}$$

Computing tolerances is demonstrated by the following example.

Consider a portfolio selection problem with $J=4$ investments and $T=5$ time periods. The risk aversion parameter is set as $\mu := 10$. The returns are displayed below:

time period t	reward on investment			
	1	2	3	4
1	11	20	9	10
2	13	25	11	13
3	10	17	12	11
4	12	21	11	13
5	12	19	13	14

The optimal solution is

$$\begin{aligned}
x^* &= (0,\ 0.1413,\ 0.8587,\ 0)^T,\\
y^* &= (10.3891,\ 9.9043,\ 9.9587,\ 10.0174,\ 9.7304)^T,
\end{aligned}$$

and the corresponding optimal return is 12.5.

Now, we compute tolerances for the given rewards r_{jt} such that the optimal return does not exceed the upper bound $\overline{f} := 20$ and the lower bound "f'' := 6. Notice that the portfolio selection problem has an opposite direction of optimization in comparison with (1) (maximization instead of minimization).

First, we compute tolerances only for one quantity, say r_{21}. This quantity appears in the objective function and also several times in the constraints, and the traditional sensitivity analysis techniques are not applicable. Our approach works, though the results are not optimal just because of the multiple appearance of r_{21} causing dependences between the coefficients. We set the radii of intervals as $(r_\Delta)_{21} = 1$ and $(r_\Delta)_{jt} = 0$ otherwise. By solving the optimization problems (4) and (5) we get the results

$$\delta^1 = \infty, \quad \delta^2 = 14.8069,$$

and the condition (6) is fulfilled. Hence, the upper bound $\overline{f}$ can be possibly exceeded by perturbing r_{21} at least by $\delta^2 = 14.8069$, while the lower bound will never be achieved or exceeded.

Now, we take into account the quantities r_{2t}, $t = 1, \ldots, T$. Define the radii $(r_\Delta)_{2t} = 1$, $t = 1, \ldots, T$, and $(r_\Delta)_{jt} = 0$, $j \neq 2$, $t = 1, \ldots, T$. By solving (4) and (5) we obtain

$$\delta^1 = \infty, \quad \delta^2 = 2.9614,$$

and the condition (6) is again fulfilled.

In the last case we take into account all the quantities r_{jt}, $j = 1, \ldots, J$, $t = 1, \ldots, T$. The radii are set accordingly as $(r_\Delta)_{jt} = 1$, $j = 1, \ldots, J$, $t = 1, \ldots, T$. Naturally, computations yield much smaller values

$$\delta^1 = 0.04545, \quad \delta^2 = 0.1575.$$

The resulting tolerance is $\delta^* = 0.04545 - \varepsilon$ with an arbitrarily small $\varepsilon > 0$, and the condition (6) is fulfilled. Therefore, as long as all the rewards r_{jt} vary within intervals $[r_{jt} - \delta^*, r_{jt} + \delta^*]$, the optimal return will never get outside the interval $[6, 20]$.

References

[1] Alefeld, G. and Herzberger, J.: *Introduction to interval computations*, Academic Press, London, 1983.

[2] Chinneck, J. W.; Ramadan, K.: Linear programming with interval coefficients, *J. Oper. Res. Soc.* **51**, No. 2 (2000), pp. 209–220.

[3] Fiedler, M., et al., *Linear optimization problems with inexact data*, Springer, New York (2006).

[4] Freund, R.W. and Jarre, F.: An interior-point method for multifractional programs with convex constraints, *J. Optim. Theory Appl.* 85, No. 1 (1995), pp. 125–161.

[5] Giovea, S. , Funaria, S. and Nardelli, C.: An interval portfolio selection problem based on regret function, *Eur. J. Oper. Res.* 170, No. 1 (2006), pp. 253–264.

[6] Hladík, M.: Optimal value range in interval linear programming, submitted to Fuzzy Optim. Decis. Mak., available as a preprint KAM-DIMATIA Series 2007-824, Department of Applied Mathematics, Prague, 2007.

[7] Ida, M. Portfolio selection problem with interval coefficients, *Appl. Math. Lett.* 16, No. 5 (2003), pp. 709–713.

[8] Inuiguchi, M.; Ramik, J.; Tanino, T. and Vlach, M.: Satisficing solutions and duality in interval and fuzzy linear programming, *Fuzzy Sets Syst.* **135**, No. 1 (2003), pp. 151–177.

[9] Inuiguchi, M. and Sakawa, M.: Minimax regret solution to linear programming problems with an interval objective function, *Eur. J. Oper. Res.* 86, No. 3 (1995), pp. 526–536.

[10] Lai, K.K., Wang, S.Y., Xu, J.P., Zhu, S.S. and Fang, Y.: A class of linear interval programming problems and its application to portfolio selection, *IEEE Trans. Fuzzy Syst.* 10 (2002), pp. 698–704.

[11] Mráz, F.: Calculating the exact bounds of optimal values in LP with interval coefficients, *Ann. Oper. Res.* **81** (1998), pp. 51–62.

[12] Nesterov, Yu.E. and Nemirovskij, A.S.: An interior-point method for generalized linear-fractional programming, *Math. Program.* 69, No. 1(B) (1995), pp. 177-204.

[13] Wang, S. and Zhu, S.: On fuzzy portfolio selection problem, *Fuzzy Optim. Decis. Mak.* 1 (2002), pp. 361–377.

Miroslav Hloušek, Osvald Vašíček[1]

Department of Applied Mathematics and Computer Science
Faculty of Economics and Administration of Masaryk University[2]
Lipová 41a, 602 00 Brno

e-mail: `osvald@econ.muni.cz, hlousek@econ.muni.cz`

Sensitivity Analysis of DSGE Model of Open Economy with Nominal Rigidities

Abstract

This paper deals with sensitivity analysis of DSGE model estimated on data of Czech economy using Bayesian techniques. The main goal is comparison of different competing models in terms how they fit the data. The benchmark model includes four types of nominal rigidities which are based on Calvo price setting mechanism. The competing models assume flexibility in some of these prices or combinations of them. The measure of fit depends on marginal likelihood obtained from Bayesian estimation. The emphasis is also put on the sensitivity of estimated parameters for different specifications of the model. The results of estimation show that frictions in import prices and wages play the most significant role in the model of Czech economy. Sticky wages are far more important than sticky prices. The most sensitive parameters are Frisch elasticity of labor supply and openness of the economy. On the other hand, the measurement errors are quite stable across the models and have negligible effect on the results.

1 Introduction

This paper presents results of sensitivity analysis made on DSGE model of open economy with nominal rigidities. The model includes four types of nominal rigidities which are based on Calvo price setting mechanism. The question of interest is comparison of different competing models in terms how they fit the data. These competing models assume flexibility in some of these prices or combinations of them. The measure of fit is marginal likelihood based on Bayesian estimation. The emphasis is also put on the sensitivity of estimated parameters for different models – how much the parameters change to compensate the lack of rigidity in particular sector of the modeled economy. Special attention is devoted to relative importance of price and wage stickiness in the Czech economy. The rest of the paper is structured as follows. Section 2 briefly describes main features of the model. The results of estimation and sensitivity analysis are presented and discussed in section 3. Section 4 summarizes main findings and concludes with prospects for further research.

[1]We thank to Dan Němec for useful comments and suggestions and to Honza Čapek for technical help.

[2]This work is supported by MŠMT project Research centers 1M0524 and funding of specific research at ESF MU. The access to the METACentrum supercomputing facilities provided under the research intent MSM6383917201 is highly appreciated.

2 Model structure

This section provides brief overview of the model structure. Detailed description of the model can be found in Hloušek (2008). It is adjusted and slightly extended model developed by Maih (2005). The model economy consists of the following agents: households, firms and aggregators, government, central bank and foreign economy that is modeled exogenously. There is monopolistic competition on the market of various goods and on the labor market. This market structure implies that goods and services are imperfect substitutes and the firms (households) have certain market power and can set price of their goods (wages). This allows existence of frictions which are modeled in the form of Calvo (1983) contracts with indexation.

Specifically, there are rigidities in the following sectors: labor market, market of domestic intermediate goods and imported and exported intermediate goods. The optimization problem of agents results in Phillips curve, which has following form (for the prices of imported goods; similarly for other prices/wages)

$$\hat{\pi}_{ft} = \frac{\beta}{1+\beta} E_t \hat{\pi}_{ft+1} + \frac{1}{1+\beta} \hat{\pi}_{ft-1} + \frac{(1-\delta_f)(1-\beta\delta_f)}{\delta_f(1+\beta\delta_f)} \; \widehat{mc}_t^f$$

The hybrid Phillips curve shows that contemporary inflation (of imported intermediate goods) depends on both lagged and expected future inflation and on the gap of the real marginal cost. Parameter β is subjective discount factor and δ_f is Calvo parameter that expresses degree of nominal rigidity, $\delta_f \in [0, 1)$. This parameter is quite crucial in our analysis; it says how often the contracts are reset.[3]

Similar Phillips curves characterize behaviour of inflation in all sectors with monopolistically competitive markets. The production structure of the model is illustrated in Figure 1. Nominal rigidities in price setting are depicted by dashed line. The model also includes one type of real rigidity that is expressed by habit in consumption. Thus the benchmark model contains four types of nominal rigidities and one real rigidity.

3 Sensitivity analysis

The competing models are derived from the benchmark model – they allow flexible prices in particular sector. The term flexibility (or no habit persistence) means that corresponding parameter is set to 0.2.[4]

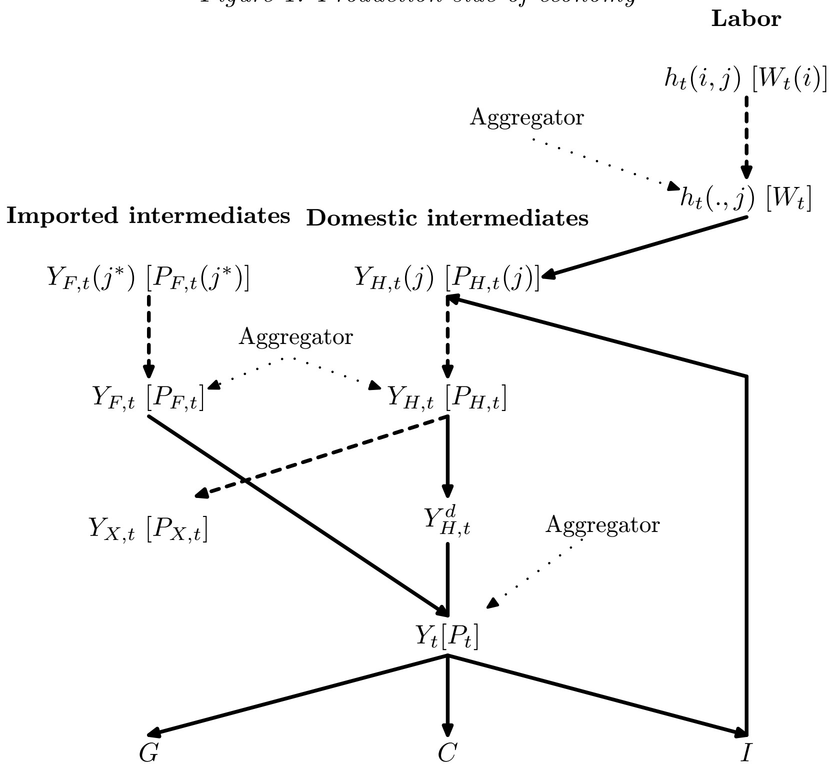
The competing models are evaluated in terms how they fit the data. Specifically, marginal likelihood calculated from Bayesian estimation is used. The Bayesian estimation overcomes Maximum Likelihood estimation because it takes into account the uncertainty that comes from the models or the estimates of shocks and parameters. Because the priors play key role in Bayesian estimation and can influence results in important way, it is necessary to set the priors of the estimated parameters for all models to the same value. It is exceptionally important for the sensitivity analysis.

The analysis considers following specifications: the benchmark model is denoted BM, the model with flexible import prices FIP, the model with flexible domestic prices FDP, the model with flexible wages FW and the model with flexible export prices FEP. Then some combinations

[3] Average length of contract is calculated using formula: $\frac{1}{1-\delta_f}$.

[4] For behaviour of prices it means that the contract is changed once in 1.25 quarters.

Figure 1: Production side of economy



with rigidities in two or more sectors are: the model with flexible domestic and export prices FDEP, the model with flexible domestic, import prices and no habit formation FDIPH and model without nominal rigidities FWDIEP. [5]

Table 1 presents estimated parameters of competing models together with Laplace approximation of the log data density.[6] The models are ordered according to the value of log data density which measures fit of the data.

The results of the analysis are quite surprising. The model with flexible domestic prices fits the data better than the richer benchmark model. Also model with assumption of flexible domestic and export prices overtook the benchmark model. With focus on relative importance between price and wage stickiness it is obvious that these rigidities are not interchangeable. Specifically, wage rigidity is more important than price rigidity. It is confirmed both by triumph of FDP

[5]Some other specifications as the models with flexible domestic and import prices FDIP plus flexible export prices FDIEP were also analyzed. Their performance to fit the data is not significantly different from FDIPH model, so they are not presented here.

[6]Posterior inference was carried out by Random Walk Chain Metropolis-Hastings algorithm incorporated in Dynare toolbox. The algorithm generated 1,000,000 draws from the posterior distribution. They were computed in two chains with 500,000 replications each, 50 % of replications were discarded so as to avoid influence of initial conditions. MCMC diagnostics were used for verification of the algorithm.

Table 1: Comparison of the models

Param.	*Prior*	*FDP*	*FDEP*	*BM*	*FEP*	*FDIPH*	*FIP*	*FW*	*FWDIEP*
Nominal rigidities									
δ_f	0.75	0.7596	0.7657	0.7530	0.7564	[0.2]	[0.2]	0.7571	[0.2]
δ_h	0.75	[0.2]	[0.2]	0.4181	0.4242	[0.2]	0.4902	0.3988	[0.2]
δ_w	0.75	0.8351	0.8579	0.8569	0.8299	0.8498	0.8382	[0.2]	[0.2]
δ_x	0.75	0.6323	[0.2]	0.6329	[0.2]	0.7007	0.7016	0.6745	[0.2]
Miscellaneous (CPI index, risk premium)									
γ	0.27	0.1374	0.1351	0.1629	0.1508	0.5133	0.5128	0.0940	0.2172
φ	0.05	0.0203	0.0197	0.0192	0.0190	0.0238	0.0203	0.0230	0.0243
Preferences									
hab	0.75	0.6892	0.7010	0.6937	0.6884	[0.2]	0.6537	0.5461	0.4972
υ	3.00	2.7051	2.7843	2.7062	2.8382	2.5112	2.5180	3.9320	3.7854
Monetary policy reaction function									
α	0.80	0.8991	0.8969	0.8972	0.8957	0.8281	0.8503	0.8964	0.8408
ω_π	1.70	1.7014	1.7002	1.7026	1.7036	1.7109	1.7174	1.6877	1.7384
ω_{yh}	0.10	0.1025	0.1024	0.1070	0.1028	0.1105	0.1103	0.0988	0.0984
Production and export function									
ψ	0.60	0.6522	0.6406	0.6498	0.6370	0.6961	0.6724	0.6019	0.6094
ϱ_{xp}	0.10	0.2521	0.1175	0.2500	0.1178	0.2283	0.2400	0.1453	0.0702
Shock persistence									
ρ_{π^T}	0.40	0.4167	0.4092	0.4184	0.4138	0.4446	0.4637	0.4159	0.4449
ρ_{θ_h}	0.40	0.4448	0.4523	0.3151	0.3151	0.4903	0.2876	0.3182	0.4574
ρ_{θ_w}	0.40	0.3824	0.398	0.3918	0.3817	0.3918	0.3751	0.6070	0.5972
ρ_{θ_f}	0.40	0.1711	0.1722	0.1705	0.1707	0.5234	0.5223	0.1714	0.5464
ρ_{zc}	0.40	0.4491	0.455	0.4440	0.4487	0.5134	0.3803	0.5025	0.5160
ρ_{zy}	0.40	0.4207	0.4049	0.4193	0.4063	0.3524	0.3680	0.4198	0.3929
ρ_{zrp}	0.40	0.5785	0.5815	0.5719	0.5705	0.5919	0.5767	0.5807	0.6136
Standard deviation of shocks									
σ_{Z_c}	0.01	0.0275	0.0297	0.0281	0.0285	0.0096	0.0236	0.0197	0.0167
$\sigma_{Z_{ls}}$	0.01	0.0401	0.2022	0.1415	0.0228	0.1405	0.0146	0.0064	0.0066
$\sigma_{Z_{mp}}$	0.01	0.0036	0.0037	0.0036	0.0037	0.0032	0.0030	0.0036	0.0033
σ_{π^T}	0.01	0.0065	0.0064	0.0064	0.0064	0.0058	0.0065	0.0064	0.0073
σ_{Z_y}	0.01	0.0031	0.0032	0.0031	0.0032	0.0029	0.0031	0.0032	0.0030
$\sigma_{Z_{rp}}$	0.01	0.0090	0.0088	0.0091	0.0090	0.0077	0.0078	0.0087	0.0069
σ_{θ_h}	0.01	0.0078	0.0079	0.0163	0.0170	0.0082	0.0227	0.0156	0.0082
σ_{θ_w}	0.01	0.1331	0.0168	0.0878	0.1417	0.0430	0.1452	0.0307	0.0303
σ_{θ_f}	0.01	0.1358	0.1526	0.1321	0.1352	0.0317	0.0318	0.1279	0.0264
Standard deviation of measurement errors									
π_{yt}	0.01	0.0031	0.0034	0.0031	0.0033	0.0043	0.0041	0.0033	0.0034
π_t	0.01	0.0101	0.0101	0.0102	0.0101	0.0123	0.0123	0.0104	0.0117
log data density		978.57	976.53	972.79	969.39	949.23	947.52	935.20	900.71

model and by poor fit of FW model. It also corresponds to empirical fact that real wages behave countercyclically in the Czech economy.

Next, the results indicate that extension of the model by rigidities in export market (compared to Maih's (2005) original model) has only small effect. Difference between benchmark model and that with flexible export prices is not so significant. The model without nominal rigidities and with only real rigidity (habit in consumption) has the worst fit of all models. It suggests that nominal stickiness is important phenomenon. However, it is not universal for all prices. The most important nominal frictions in the Czech economy are rigidities in wages and import prices.

Interesting fact is how the model parameters compensate the lack of rigidity in particular sector (compared to the benchmark model). The most striking difference is in estimates of Frisch elasticity of labor supply ($\frac{1}{\upsilon}$) which decreases from $\frac{1}{2.7062}$ for the benchmark model to $\frac{1}{3.9320}$ for the flexible wage model. It implies that labor supply is more unresponsive to changes in the real wage. In other words, the nominal rigidity in wages is transferred to rigidity in labor supply.

Another structural parameter that is highly volatile for different specifications is γ. It expresses openness of the economy; more precisely $(1-\gamma)$ should capture share of imports to output. However, γ also regards to the price index, it expresses weight of domestic prices in the CPI. Because the nominal time series (inflation of domestic prices, import prices and CPI) are used for estimation this influence is probably more important for the results. Value of γ fluctuates from 0.1629 for the benchmark model (very open economy) up to 0.5128 for flexible import prices model. The domestic prices thus increased their ability in explaining (rigid) behaviour of CPI index. Low value of 0.094 for flexible wage model shows extreme openness to trade. This result is quite puzzle and can stem from mutual influence of more parameters for this specification.

Autoregressive parameter of shock to import prices is also highly volatile. It increases (from 0.17) to value around 0.52 for all models with flexible import prices. The model tries to explain higher persistence in import prices by shock behaviour.

This approach also enables to assess importance of measurement errors introduced in the estimation. Measurement errors remain quite stable for all specifications of the models; the only exception is measurement error in output growth which slightly increased for models which include flexible import prices. However, in general the results indicate that measurement errors have negligible effect for the fit of the models.

Similar analysis could be applied to shocks and their volatility. The most important shocks (with largest standard deviation) are labor supply shock and markup shocks to import prices and to wages. These shocks also vary most of all which indicates quite high sensitivity to the type of rigidity. It is quite hard to find some stable pattern across the models. However, it seems that higher flexibility of particular prices help to explain behaviour of relevant variables by the model and reduces the standard deviation of shocks.

4 Conclusion

This paper analyzed DSGE model with nominal rigidities estimated on Czech data using Bayesian techniques. The unrestricted model included four types of nominal rigidities and one real rigidity. Several model specifications with flexibility in particular sector were estimated and compared. The marginal likelihood provided quantitative measure of data fit.

The result shows that only some nominal rigidities are important characteristics of the model for Czech economy. The benchmark model (with all set of rigidities) is not the winner. Models with flexible domestic prices and flexible domestic and export prices fit the data better. The issue whether price or wage stickiness is more important has been resolved in favour of wage rigidity. Model with flexible wages had poor outcome. Thus sticky import prices and wages are the core of the model.

The most sensitive parameters are Frisch elasticity of labor supply and openness of the economy. On the other side, the measurement errors are quite stable across the models and do not influence obtained result in an important way.

Further research will be focused on comparison of DSGE model with VAR. Behaviour of selected variables for several restricted models will be examined and compared by vector autocorrelations and impulse response functions.

References

[2005] Adolfson, M., Laseen, S., Linde, J. and Villani, M.: Bayesian estimation of an open economy DSGE model with incomplete pass-through, *Working Paper 179*, Sveriges Riksbank, (2005).

[1983] Calvo, G.: Staggered prices in a utility-maximizing framework, *Journal of Monetary Economics 12*, 383398, (1983).

[2008] Hloušek, M.: Nominal rigidities in DSGE model with imperfect exchange rate pass-through, *Working Paper*, Brno: CVKSČE MU, forthcoming (2008).

[2003] Koop, G.: *Bayesian Econometrics* Chichester: Wiley, (2003).

[1] Maih, J.: *Asymmetric Trade and Nominal Rigidities in a DSGE Perspective: Implications for the Transmission of Shocks and Real Exchange Rate Dynamics*, PhD thesis, The University of Oslo, (2005).

Michal Houda

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4, 182 08 Praha 8

e-mail: houda@karlin.mff.cuni.cz

Weak Structural Dependence in Chance-Constrained Programming

Abstract

In chance-constrained optimization problems, a solution is assumed to be feasible only with certain, sufficiently high probability. For computational and theoretical purposes, the convexity property of the resulting constraint set is treated. It is known, for example, that a suitable combination of a concavity property of the probability distribution and concavity of constraint mappings are sufficient conditions to the convexity of the resulting constraint set. Recently, new concavity condition of the probability distribution - r-decreasing density - has been developed. Henrion and Strugarek (2006) show, under the assumption of independence of constraint rows, that this condition on marginal densities allows us, on the other side, weaken the concavity of constraint mappings. In this contribution we present a relaxation of the independence assumption in favour of a specific weak-dependence condition. If the independence assumption is not fulfiled, the resulting constraint set is not due to be convex. However, under a weak-dependence assumption, the non-convex problem can be approximated by a convex one. Applying stability results on optimal values and optimal solutions, we show that optimal values and optimal solutions remain stable under assumptions common in stochastic programming. This implies desirable consequences, because convex problems are easiest to compute and also many theoretical results are based on convexity assumptions. We accompany the shown results by simple example to illustrate the concept of the presented approximation.

1 Introduction

Theory of optimization is very important area to model real-life economic and engineer problems. These can be modelled as deterministic, with ad-hoc selected or estimated parameters but there are many known cases leading to false results due to presence of uncertainty that cannot be neglected. In such cases various approaches dealing with uncertainty are applied, e. g., traditional sensitivity analysis, parametric programming, robust programming techniques. In this paper we discuss stochastic programming approach, notament its important part: chance-constrained programming.

Consider an optimization problem of the form

$$\text{minimize } c(x) \text{ subject to } h(x;\xi) \geq 0 \tag{1}$$

where $x \in \mathbb{R}^m$ is a decision vector, $\xi \in \Xi \subset \mathbb{R}^s$ is a parameter of the problem, $c\colon \mathbb{R}^m \to \mathbb{R}$ is real and $h\colon \mathbb{R}^m \times \mathbb{R}^s \to \mathbb{R}^d$ vector-valued mappings. The parameter ξ is not known precisely – suppose that it is a random vector on some underlying probability space $(\Omega, \mathcal{A}, \Pr)$ with known probability distribution μ defined on the support Ξ. We can now require the constraints of the problem to be satisfied only with some prescribed but sufficiently high probability $p \in [0;1]$. We obtain so-called chance-constrained optimization problem:

$$\text{minimize } c(x) \text{ subject to } \Pr\{h(x;\xi) \geq 0\} \geq p \tag{2}$$

There are many known results from both the theoretical and application area of chance-constrained programming, see e. g. [4]. One of two main known issues discussed by the stochastic programming community is the question of convexity of the feasible set in (2). Denote

$$M(p) \coloneqq \Pr\{h(x;\xi) \geq 0\} \geq p \tag{3}$$

Convexity of the set $M(p)$ is of high importance from the theoretical as from the computational point of view. There are known results starting with the trivial one: if $h(\cdot, \xi)$ is convex for all ξ then the sets $M(0)$, $M(1)$ are convex. Unfortunately, these sets are not of our usual interest (they represent all possible, and all almost sure realizations respectively).

An answer to the question posed above, nowadays considered as classical one, is first developed in [3]. The author uses a concept of parameterization of the concavity of a function.

Definition 1. *A function $f\colon \mathbb{R}^d \to (0;+\infty)$ is called r-concave for some $r \in [-\infty;+\infty]$ if*

$$f(\lambda x + (1-\lambda)y) \geq [\lambda f^r(x) + (1-\lambda) f^r(y)]^{1/r} \tag{4}$$

is valid for each $x, y \in \mathbb{R}^d$ and each $\lambda \in [0;1]$. The cases $r = -\infty, 0, +\infty$ are treated by continuity.

Our interest focuses on values of $r \leq 1$. For $r = 1$, f is concave in its classical sense. 0-concave function is also called log-concave (as $log f$ is concave in such case), $-\infty$-concave function is known as quasi-concave function. Further, in [3] you can find the definition of r-concave probability measure which is used as assumption for the following result:

Proposition 1 ([5], Theorems 2.5 and 2.11)**.** *If μ is absolutely continuous (with respect to Lebesgue measure), log-concave (or r-concave for $r \geq -1/s$) measure, and the one-dimensional components of h are quasi-concave functions of (x, ξ) then $M(p)$ is convex set.*

For our purposes it is sufficient to say that a log-concave (r-concave) measure is implied by a log-concave ($\frac{r}{1-rs}$-concave) density. Many multivariate distributions (normal, beta, Wishart, etc.) share this property hence many chance-constrained problems can be solved by means of convex optimization.

In this paper we focus on the chance-constrained problem with random right-hand side only:

$$\min c(x) \text{ subject to} \Pr\{g(x) \geq \xi\} \geq p. \tag{5}$$

Here, we have set $h(x;\xi) = g(x) - \xi$ with g having appropriate dimensions, and according to Proposition 1 we require h to be quasi-concave. Unfortunately, quasi-concavity is not preserved under addition and we have to require $g(x)$ to be concave (quasi-concave is not sufficient) to satisfy this requirement (see [4] again).

Recently, Henrion and Strugarek [1] proposed an idea to relax concavity condition of g and make more stringent concavity condition on the probability distribution μ. They defined the notion of so-called r-decreasing density as follows:

Definition 2 ([1], Definition 2.2)**.** *A function $f : \mathbb{R} \to \mathbb{R}$ is called r-decreasing for some $r \in \mathbb{R}$ if*

1. *it is continuous on $(0; +\infty)$, and*
2. *there exists a threshold $t^* > 0$ such that $t^r f(t)$ is strictly decreasing for all $t > t^*$.*

In [1], the notion of r-decreasng densities is used to replace the original assumption on r-concavity of the distribution of μ as follows:

Theorem 3 ([1], Theorem 3.1)**.** *If*

1. *there exist $r_i > 0$ such that the components g_i of g are $(-r_i)$-concave,*
2. *the components ξ_i of ξ have $r_i + 1$-decreasing densities, and*
3. *the components ξ_i of ξ are independently distributed,*

then $M(p)$ is convex for all $p > p^* := \max_i F_i(t_i^*)$ where F_i denotes the distribution function of ξ_i (one-dimensional marginals of ξ) and t_i^* refer to the definition of r_i+1-decreasing probability density.

2 Weak dependence of the rows

An important assumption – the row independence (assumption 3 in the previous theorem) – could be relaxed in the following way. We define an α coefficient of dependence as

$$\alpha := \sup_z \left| F(z) - \prod_i F_i(z_i) \right| \tag{6}$$

where F is the distribution function of the vector ξ, F_i are the corresponding one-dimensional marginal distribution functions and $z = (z_1, \ldots, z_s) \in \mathbb{R}^s$.

It is a modified version of the strong-mixing coefficient, well known from the theory of random processes. To simplify the notation we use the notion of α-dependence and α coefficient in the sense of the above definition and not in the classical sense which we will not need anymore in the remaining part of the paper.

In Problem (5) we allow for a small structural dependence in the following way. Recall that the set $M(p)$ of feasible solution can be written as

$$M(p) := \left\{ x \in X | F(g(x)) \geq p \right\} \tag{7}$$

where F is the distribution function of the random right-hand side ξ. For our purposes we replace $M(p)$ in (5) by another set defined as

$$M'(p) = \{x \in X | \prod_{i=1}^{s} F_i(g_i(x)) \geq p\} \tag{8}$$

If the components ξ_i of ξ are independently distributed then the two sets $M(p)$ and $M'(p)$ are equal. This is not true for the dependent case but the following proposition is valid

Proposition 2. *If the components ξ_i of ξ in (5) are α-dependent (in the sense of (6) then*

$$M'(p+\alpha) \subset M(p) \subset M'(p-\alpha) \subset M(p-2\alpha). \tag{9}$$

The proof of this proposition is given in [2]. For sufficiently high p the (possibly) non-convex set is bounded from both sides by convex sets by the following theorem:

Theorem 4. *If*

1. *there exist $r_i > 0$ such that the components g_i of g are $(-r_i)$-concave,*
2. *the components ξ_i of ξ have $r_i + 1$-decreasing densities,*
3. *the components ξ_i of ξ are α-dependently distributed, and*
4. $p > \max_i F_i(t_i^*) + \alpha$

*then $M(p)$ is bounded (from both sides) by **convex** sets $M'(p+\alpha)$ and $M'(p-\alpha)$.*

For small values of α, classical stability results of stochastic programming apply. For example, if $c(x)$ is Lipschitz continuous, the constraints are metrically regular and α sufficiently small, then the optimal values and optimal solutions remains (locally) stable. For details see [2] again.

3 Comparison of the dependent and independent case

Consider the following optimization problem

$$\text{minimize } x + y \text{ subject to} \tag{10}$$

$$\begin{aligned} g_1(x,y) &= \frac{1}{x^2+y^2+0.1} &\geq \xi_1 \\ g_2(x,y) &= \frac{1}{(x+y)^2+0.1} &\geq \xi_2 \end{aligned}$$

and assume that the random vector ξ is normally distributed with zero mean and the variance matrix Σ. We consider two cases:
(a) independent case with

$$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

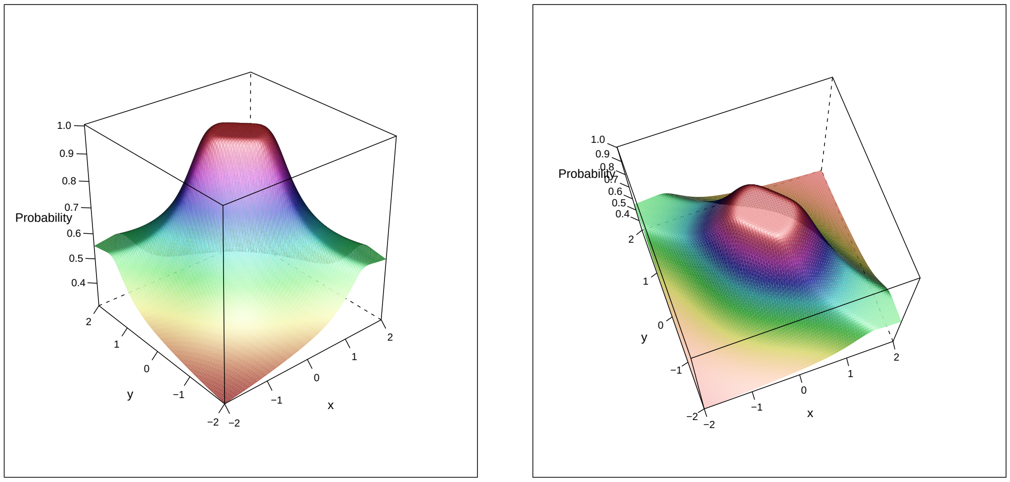


Figure 1: Collection of sets $M'(p)$

and (b) weak dependent case with

$$\Sigma = \begin{pmatrix} 1 & 0.1 \\ 0.1 & 1 \end{pmatrix}$$

In this example the g_i's are (-1)-concave, ξ_i's have 2-decreasing densities with the treshold $t^* = \sqrt{2}$, the critical probability level is $p^* = \Psi(\sqrt{2}) = 0.921$ and the weak-dependence coefficient for the dependent case is $\alpha = 0.017$.

The overall shape of the collection of sets $M'(p)$ is given on Figure 3. Each individual set $M'(p)$ is given as horizontal cut on the specified level p (z-axis). The contour lines of these sets are depicted on the following figure; the symbol φ denotes the corresponding optimal values.

For the chosen normal distribution, convexity of the feasible set is assured theoretically at the probability level of 0.921 in the independent case. As Figure 3 shows, the actual probability level in the example is much more smaller, around the value of 0.7. In the weak dependent case, these tresholds (theoretical and actual) are shifted towards the center of feasibility sets (center of image), and the optimal values and optimal solutions (depicted as points on the last figure) remain stable as the value of α-coefficient is small. Interesting behaviour around the actual treshold value is still open question from the theoretical point of view.

References

[1] Henrion, R., Strugarek, C.: Convexity of chance constraints with independent random variables. Stochastic Programming E-Print Series (SPEPS) **9** (2006) To appear in Computational Optimization and Applications.

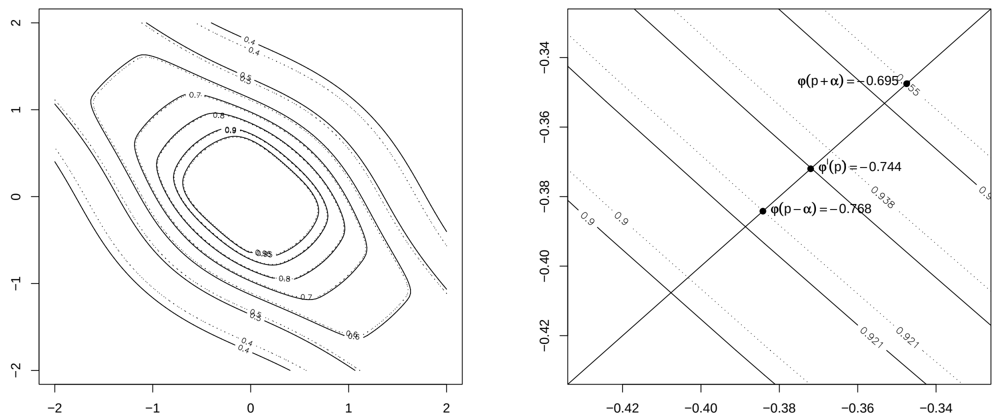


Figure 2: Contour lines for $M'(p)$ (solid) and $M(p)$ (dotted) sets

[2] Houda, M.: Convexity and dependence in chance-constrained programming. Research Report 2190, Institute of Information Theory, Academy of Science of the Czech Republic (August 2007)

[3] Prékopa, A.: A class of stochastic programming decision problems. Mathematische Operations forschung und Statistik (1972) 349–354

[4] Prékopa, A.: Stochastic Programming. Akadémiai Kiadó, Budapest (1995)

[5] Prékopa, A.: Probabilistic programming. In Ruszczynski, A., Shapiro, A., eds.: Stochastic Programming. Handbooks in Operations Research and Management Science. Volume 10. Elsevier, Amsterdam (2003) 267–352

Roman Hušek[1], Václava Pánková[2]

Department of Econometrics
Faculty of Informatics and Statistics
University of Economic
W. Churchill Sq. 4, 130 67 Praha 3

e-mail: husek@vse.cz, pankova@vse.cz

Analysis of Financial Structures In the Czech and Slovak Republics

Abstract

The implications of the financial structure of the private sector for the monetary transmission mechanism are analyzed. Co-movements of selected macroeconomic variables conditional on domestic monetary shocks are studied by the help of VAR system estimates conducted separately on Czech Republic and Slovakia. A comparison of both economies is given. It is possible to argue that in the both countries the effects of monetary policy on business cycle fluctuations does not differ significantly from that observed in current members of the Eurozone. As a benchmark for assessing the convergence of these two countries financial structures we choose the Eurozone economies.

Keywords: *financial structure, transmission of monetary shocks, VAR models, impulse response functions*

1 Introduction

The structure of financial assets and liabilities in the corporate and household sectors has important impact to the transmission mechanism of monetary policy. The financial pattern of EU average shows (see Anzuini, Levy, 2004) that

- total financial assets represent 8 times GDP
- the ratio of financial assets held by banks to financial assets held by all other sectors is 44%
- the size of financial liabilities of firms is 80% of GDP
- the household debt is 60% of GDP

[1]Financial support of GAČR 402/06/0190 is gratefully acknowledged
[2]Financial support of GAČR 402/06/0049 is gratefully acknowledged

There is an evidence of a less maturity of financial markets of new EU members, including Czech Republic and Slovakia which are analyzed here (Ihnat, Prochazka, 2002 and Janosik, Malina, 2002). A question arises if a lower level of financial development of this economies does or does not influence a reaction of their main macroeconomic variables to a monetary shock. An answer will be given by the help of VAR models constructed for both countries. The same methodology was used for a study of monetary transmission mechanisms (MTM) in the Eurozone (Peersman, Smets, 2001). With the help of VAR models we analyze shocks of a monetary policy which are given as a unit increase of the short – term interest rate and its influence on all other variables in the model.

2 Theoretical Model

The five variable VAR(1) model for each country is formulated as

$$\begin{bmatrix} CPI_t \\ IP_t \\ M_t \\ R_t \\ CP_t \end{bmatrix} = \begin{bmatrix} c_{CPI} \\ c_{IP} \\ c_M \\ c_R \\ c_{CP} \end{bmatrix} + C(L) \begin{bmatrix} CPI_{t-1} \\ IP_{t-1} \\ M_{t-1} \\ R_{t-1} \\ CP_{t-1} \end{bmatrix} + \begin{bmatrix} u_{CPI,t} \\ u_{IP,t} \\ u_{M,t} \\ u_{R,t} \\ u_{CP,t} \end{bmatrix} \tag{1}$$

with consumer price index (CPI), industrial production (IP), money supply M2 (M), interest rate (R), commodity prices (CP), disturbances (u) and lagged variables coefficients matrix $C(L)$.

The estimated models are just identified systems. All variables are expressed in the first differences of log levels (except of interest rates). Their first differences are covariance stationary or near stationary (according to ADF and PP tests). Only one lag of each variable is comprised despite to the results of AIC or SBC. All time-series (source: database "IFS" of IMF) are seasonally adjusted, the analyzed period is from 1996(Q1) to 2007(Q1). The results of the estimates and relevant impulse response functions are presented in the Appendix part. Single equations were estimated by the help of OLS in the software product EViews.

For identifying monetary policy shocks, the standard Choleski decomposition was used. The recursive structure of identification scheme is designed to recover only the effects of a monetary policy shock and assumes that the Central Bank (CB) chooses the interest rate by looking at the current level of prices and output and that output and price level do not change in the impact period but react only with one period delay.

The commodity prices are the world prices in domestic currency, i.e. the product of the nominal exchange rate and the world export prices in US dollars. The inclusion of this variables is motivated to control for an endogenous increase in the interest rate. We consider all impulse response functions (IRF) with a 20 quarter horizon after the shock.

3 The comparison of monetary shocks in CR and SR

On the base of a MTM analysis of Central and East European Countries (CEEC), e.g. Hericourt, Matei (2004) or Anzuini, Levy (2004) have found, that effectiveness of monetary policy is closely related to the transmission of a monetary impulse from the CB key interest rates to money market and then other fixed income interest rates. The degree of development of the financial markets is therefore crucial for the propagation of monetary impulses. The effects of the interest rates on the exchange rate is obviously an another important channel.

Since we expect the credit channel plays a central role in a firm's response to changes in monetary policy, the higher the share of bank loans to total liabilities, the more firms should be affected by monetary impulses. In the papers of Ihnat, Prochazka (2002) and Janosik, Malina (2002) we can see that this share in CR and SR is high in comparison to Eurozone.

As an indicator of the expected reaction of households we use the share of mortgage payments to total payments. The higher this share the greater the impact of monetary police. In the both countries households debt is very low compared with the Eurozone average.

We can say that the CR and SR have low levels of financial depth and financial intermediation. Moreover an important role play the non – residents. However, despite the lack of financial development, the effects of a contractionary monetary policy shocks in the CR and SR are qualitatively similar to that found in the Euroarea economies and conditional on a monetary shock, macroeconomic variables display standard behavior.

4 Empirical results

From IRF and their confidence bands at 95% significance level (Figures 1, 2) we can see, that despite of the lack of financial development, for both countries, conditional on a monetary shock, macroeconomic variables display standard behavior. In both economies impulse responses recovered, following an increase in interest rates, industrial production declines and the exchange rate tends to appreciate. Almost all variables move in the expected direction and all of them at a 95% level of significance. The price level follows closely the movement of the monetary aggregate and both variables decrease significantly and reach a negative peak with only few quarters delay with respect to industrial production.

Since we are dealing with small open economies we assume that the monetary shock has no impact on world export quantities. Therefore two channels are operating at the same time. The interest rate channel and the exchange rate one and increase in the interest rate should depress an economic activity through tighter credit conditions, while the increased interest rate appreciates the currency. The nominal appreciation of the currency may depress the economy even more if it affects the relative prices of domestic and foreign goods. The shapes of the dynamic IRF for the CR and SR are very similar to what we can observe in advanced economies, but estimates for SR are surrounded by higher uncertainty which enlarges confidence bands.

5 Conclusions

Despite the lack of financial development, the results of the econometric analysis suggest that the co-movement of macroeconomic variables in both countries, conditional on a domestic monetary shocks, is not different from the standard behavior of Euroarea economies. Also the impact of monetary policy on the exchange rate is the expected one. Despite of the relative small number of observations in time-series we can confirm that in CR and SR the contribution of monetary policy to business cycle fluctuations does not differ significantly from developed countries. This signals the absence of a strong asymmetry in the effects of monetary policy. This would suggest that the cost of the adopting of Euro currency may be in both countries relatively low. A similar conclusion is to be read in Hušek, Formánek (2006). We can conclude that prior accession to the Eurozone some changes should be made: First of all, monetary police should be more oriented to price level stability by using the interest rate transmission channel, which is closely tied to further development of financial markets.

References

[1] Anzuini, A., Levy, A.: "Financial structure and the transmission of monetary shocks: preliminary evidence for the Czech Republic, Hungary and Poland," Temi di discussione (Economic working paper 514), Banca d' Italia, Economic Research Department. 2004.

[2] ECB: Financial sectors in EU accession countries. C. Thimann, ed., Frankfurt, 2002

[3] Hericourt, J., Matei, I.: A VAR description of the effects of monetary policy in CEECs' team. 5th Doctoral meeting in international Trade and Finance. Working paper, 2004

[4] Hušek, R., Formánek, T.: Konvergence české ekonomiky k eurozóně a odhad nákladů přijet eura v ČR. Statistika 1, 2006

[5] Ihnat, P., Prochazka, P.: The financial sector in the Czech Republic: an assessment of its current state of development and functioning. ECB, 2002

[6] Janosik, J., Malina, L.: Financial sector situation and development in the Slovak Republic. ECB, 2002

[7] Peersman, G., Smets, F.: The monetary transmission mechanism in Euro Area: more evidence from VAR analysis. ECB, Working paper No. 91, 2001

Internet: `http://ifs.apdi.net`, `http://www.ecb.int`

Appendix

Continues on next page – editor's remark.

Vector Autoregression Estimates					
Sample: 1996:1 2007:1					
Included observations: 45					
Standard errors in ( ) & t-statistics in [ ]					
	CZ_CPI	CZ_IP	CZ_M	CZ_R	CZ_CP
CZ_CPI(-1)	0.639968	-0.263414	-0.421591	37.37454	-2.099111
	(0.10221)	(0.30014)	(0.47305)	(13.9114)	(0.48213)
	[6.26143]	[-0.87764]	[-0.89122]	[2.68661]	[-4.35380]
CZ_IP(-1)	0.149774	0.697572	0.360889	16.07087	0.118062
	(0.04352)	(0.12779)	(0.20142)	(5.92325)	(0.20528)
	[3.44163]	[5.45857]	[1.79176]	[2.71318]	[0.57511]
CZ_M(-1)	-0.001570	-0.107607	0.699908	-5.373127	0.028288
	(0.01826)	(0.05363)	(0.08453)	(2.48576)	(0.08615)
	[-0.08595]	[-2.00646]	[8.28031]	[-2.16156]	[0.32836]
CZ_R(-1)	0.000830	-0.000357	0.001521	0.945978	0.005161
	(0.00031)	(0.00092)	(0.00145)	(0.04278)	(0.00148)
	[2.64219]	[-0.38694]	[1.04551]	[22.1136]	[3.48128]
CZ_CP(-1)	0.075142	0.196850	0.174963	-1.452287	0.892816
	(0.01613)	(0.04738)	(0.07468)	(2.19610)	(0.07611)
	[4.65710]	[4.15464]	[2.34294]	[-0.66130]	[11.7305]
C	-0.005213	0.011160	-0.010201	-0.216368	-0.024255
	(0.00230)	(0.00675)	(0.01064)	(0.31283)	(0.01084)
	[-2.26829]	[1.65351]	[-0.95892]	[-0.69165]	[-2.23719]
R-squared	0.966594	0.871340	0.834424	0.995520	0.878188
Adj. R-squared	0.962311	0.854845	0.813196	0.994946	0.862572
Sum sq. resids	8.99E-05	0.000775	0.001926	1.665635	0.002001
S.E. equation	0.001518	0.004459	0.007027	0.206661	0.007162
F-statistic	225.6929	52.82477	39.30822	1733.465	56.23332
Log likelihood	231.4236	182.9480	162.4750	10.31802	161.6191
Akaike AIC	-10.01883	-7.864356	-6.954446	-0.191912	-6.916404
Schwarz SC	-9.777938	-7.623468	-6.713557	0.048976	-6.675516
Mean dependent	0.009875	0.012022	0.009688	8.278308	-0.002660
S.D. dependent	0.007821	0.011703	0.016259	2.907025	0.019320
Determinant Residual Covariance		9.41E-22			
Log Likelihood (d.f. adjusted)		770.0790			
Akaike Information Criteria		-32.89240			
Schwarz Criteria		-31.68796			

Table 1: Estimation of VAR(1) model for CR

Vector Autoregression Estimates					
Sample: 1996:1 2007:1					
Included observations: 45					
Standard errors in ( ) & t-statistics in [ ]					
	SK_CPI	SK_IP	SK_M	SK_R	SK_CP
SK_CPI(-1)	0.807353	0.096605	-0.755015	-56.41275	-0.382478
	(0.08683)	(0.14768)	(0.28604)	(18.4623)	(0.16877)
	[9.29785]	[0.65417]	[-2.63955]	[-3.05556]	[-2.26632]
SK_IP(-1)	-0.048760	0.832862	0.435499	-21.45741	-0.260961
	(0.06039)	(0.10270)	(0.19893)	(12.8396)	(0.11737)
	[-0.80746]	[8.10955]	[2.18926]	[-1.67119]	[-2.22345]
SK_M(-1)	-0.002922	0.037586	0.775606	2.523240	-0.163203
	(0.02687)	(0.04571)	(0.08853)	(5.71409)	(0.05223)
	[-0.10874]	[0.82234]	[8.76103]	[0.44158]	[-3.12452]
SK_R(-1)	0.000133	-0.000644	0.000728	1.033493	0.001200
	(0.00019)	(0.00033)	(0.00064)	(0.04110)	(0.00038)
	[0.69007]	[-1.96045]	[1.14271]	[25.1460]	[3.19341]
SK_CP(-1)	0.046372	0.133362	0.324299	-9.007573	0.826080
	(0.03328)	(0.05661)	(0.10964)	(7.07672)	(0.06469)
	[1.39323]	[2.35599]	[2.95784]	[-1.27285]	[12.7700]
C	0.001784	0.008610	0.001550	0.526759	-0.003943
	(0.00225)	(0.00382)	(0.00740)	(0.47761)	(0.00437)
	[0.79417]	[2.25364]	[0.20943]	[1.10290]	[-0.90310]
R-squared	0.838736	0.760063	0.870073	0.984988	0.903445
Adj. R-squared	0.818061	0.729302	0.853416	0.983063	0.891066
Sum sq. resids	0.000383	0.001107	0.004152	17.29862	0.001445
S.E. equation	0.003132	0.005327	0.010318	0.665999	0.006088
F-statistic	40.56792	24.70857	52.23370	511.7783	72.98278
Log likelihood	198.8364	174.9393	145.1897	-42.34143	168.9324
Akaike AIC	-8.570509	-7.508415	-6.186208	2.148508	-7.241439
Schwarz SC	-8.329620	-7.267526	-5.945320	2.389397	-7.000550
Mean dependent	0.015936	0.012468	0.018806	12.90747	-0.000251
S.D. dependent	0.007344	0.010239	0.026951	5.117498	0.018446
Determinant Residual Covariance		3.19E-19			
Log Likelihood (d.f. adjusted)		638.9597			
Akaike Information Criteria		-27.06488			
Schwarz Criteria		-25.86043			

Table 2: Estimation of VAR(1) model for SR

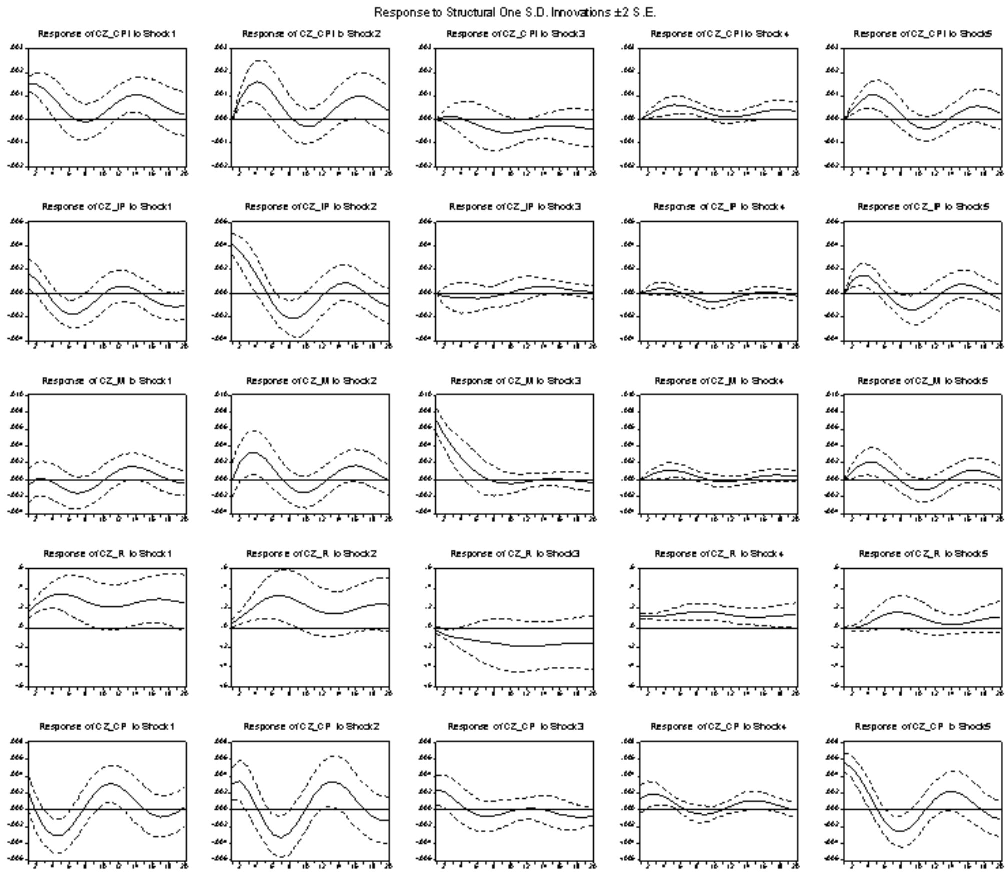


Figure 1: Impulse response functions – recursive identification scheme – CR

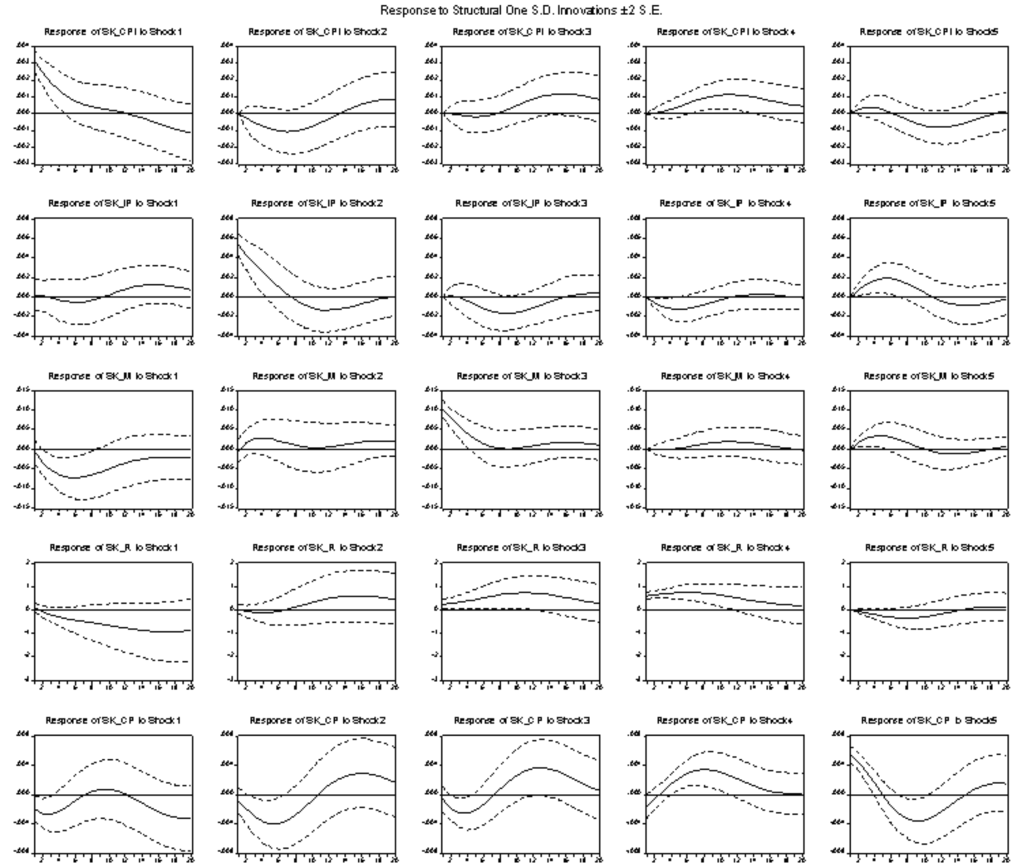


Figure 2: Impulse response functions – recursive identification scheme – SR

In both figures, unit shocks of interest rate (R) are plotted in the fourth column.

Josef Chudoba

Faculty of Mechatronics and Interdisciplinary Engineering Studies
Technical University of Liberec
Studentská 2, 461 17 Liberec

e-mail: josef.chudoba@tul.cz

Reliability Analysis of a Natural Gas Compression Station, Prediction of a Profit and Loss Arise from Estimation of Non-delivery of Agreed Amount of the Gas

Abstract

Reliability estimation of a natural gas compression station is one of the important input parameters for the prediction of a profit/loss during the gas transportation. The total profit is given by the volume of transported gas and reliability of the system. The Czech Republic represents one of the European transit countries for distribution of natural gas. Moreover, the transit of medium is governed by similar regulations either we take into account gas or electricity. We speak about distant transitions in which the Czech Republic is "only" the transit country.

The aim of my contribution is to provide the mathematical model (Markov's process) of actual natural gas compression station which is situated in the Czech Republic. Reliability of the system is dependent on the volume of transported gas. The volume of gas is function of a number of connected lines and turbocompressors. The compression station and input/output lines are described by dynamical model. This model contents a time degradation of components, preventive and periodical maintenance.

It is possible to determine estimation of profit/loss during the gas transportation by the knowledge of unitary profit/loss and probability of every state in time and number of transits between the states in transient diagram.

Keywords: *Reliability analysis, compression station, Markov chains, maintenance, profit and loss*

1 Introduction

A reliability estimation of a natural gas compression station is one of the important input parameters for the availability of the system $A(t)$ (it indicates probability, that system is in function state) and subsequently for estimation of the profit and loss from transport of gas. The instantaneous availability is related on volume of the transported gas and the dependability particularly items. The availability of whole components is related on parameters of reliability, maintainability and maintenance support.

The Markov model [1] is considered for reliability modelling of the compression station. The main advantage of Markov models is the fact that we can obtain information about system availability as a function of time: $A(t)$. In order to increase the reliability of the system, we can also assume maintenance operations in our model. It is worth noting that conventional reliability methods, such us Fault Tree Analysis (FTA) or Reliability Block Diagrams (RBDs) estimate only the asymptotic system availability.

The Markov's models are able to describe the system by using more than two states. It is possible to define every state if it is in function state, degradation state or failure state. It is append the information about the degradation to every state. It is possible to differentiate states of the system by the volume of transported gas. The system is described by 440 states (It is described by measure of the degradation and volume of transported gas) and it is possible to estimate profit or loss for every states during the time, if the system is in specific state.

In our contribution, we extend Markov models for modeling of periodical maintenance of the system by assuming Weibulls models [3, 4]. We will also assume several various power configurations of the system by multi-state Markov model. Our model is dynamic in sense of time and system power:

1) We can estimate the system availability as a function of time.

2) We can model the several scenarios of performance changes. These assumed performance changes are based on consumer wishes which are changing during time and are provided and supervised by a human dispatcher.

This model will be expanded of the modelling profit or loss of the system. The principle of the method is the estimation of profit for every state of system and the estimation of profit from transient states. The estimation of profit for whole states describes unit's profit from the transported gas. The estimation of profit from transient between states represents failure or restoration of the components.

The Czech Republic represents one of the European transit countries for distribution of natural gas. Possible failures in transition networks can influence not only the locality of the failure, but also surrounding regions even all country.

2 System analysis

2.1 System description

We assume a compressor station and also its surrounding gas pipeline network in our mathematical model. All safety-valves, valve piece, pipeline,... will be called by an expression "lines" and similarly a compressor, turbine, energy management, cooling management, oil management will be called under one title "turbocompressor". Lines are divided into input and output. We assume a system with 7 identical turbocompressors and with 3 input lines and with 3 output lines in our model. A state of the assumed system is called safety state, if all turbocompressors and also all lines work well.

A state of the assumed system is called degradated/failure, if there is a failure in a turbocompressor or a failure in a line and this failure has/has not influence to a consumer. The consumers get/refused the agreed amount of the gas in time and in agreed quality.

3 Power configurations of the model

The transit gas pipeline network (and also the other transit networks) does not work in a constant level of the power. The pipe lines and compressors are turned on/off in order to satisfy consumer wishes which are changing during time.

If consumer wishes are low, the assumed scenario should be different than in a maximum power scenario. The presented model assumes 10 performance configurations of the compressor station of a transit gas pipeline. The configurations are identified by numbers 1, 2, ..., 10, see Table 1. We will demonstrate the behaviour of the Configuration No. 1 and the behaviour of the Configuration No. 7 in our contribution.

Scenario indication	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Numer of used lines	1	1	2	2	2	3	3	3	3	3
Numer of used turbocompressors	1	2	2	3	4	3	4	5	6	7

Table 1: List of solved schemes

3.1 Block Diagrams of non-failure power configurations

The reliability block diagram (RBD) of the compressor station for configuration No.1 means that only one input line, one output line and one turbocompressor is enough for a degradated state. Other lines or compressors are in standby redundancy (expressed by arrows, see Figure 1). In case of a failure the device with the standby redundancy sets off and replaces the device with the failure. Indication "1 out of "3" means that for satisfying a system function at least one working device is needed. The model assumes that the component starts operating immediately and every time.

3.2 Markov model of power configurations

In this section, we describe RBD diagrams and Markov models of the compressor station for configurations No.1 and No.7, see Figure 3 and Figure 4.

Failure states are marked in red (grey). The rest of states (i.e. safety or degradated states) are marked in white. For example a state 132 represents a failure of the line on the input, 3 turbocompressors being in a failure state and two broken-down lines on the output.

3.3 Transition intensities of power configurations

Markov models assume standby redundancy **m** out of **n**, see Figure 3 and Figure 4. In case of a failure the device with the standby redundancy sets off and replaces the device with the failure.

In individual scenarios there are different multiplying coefficients at transition intensities in a transition matrix. The value of a multiplying coefficient can be found in Figure 5. The names of states also describe the number of failure components. The transition intensity from state 0 to state 1 is $m\lambda$, because of there is exactly m non-failure components.

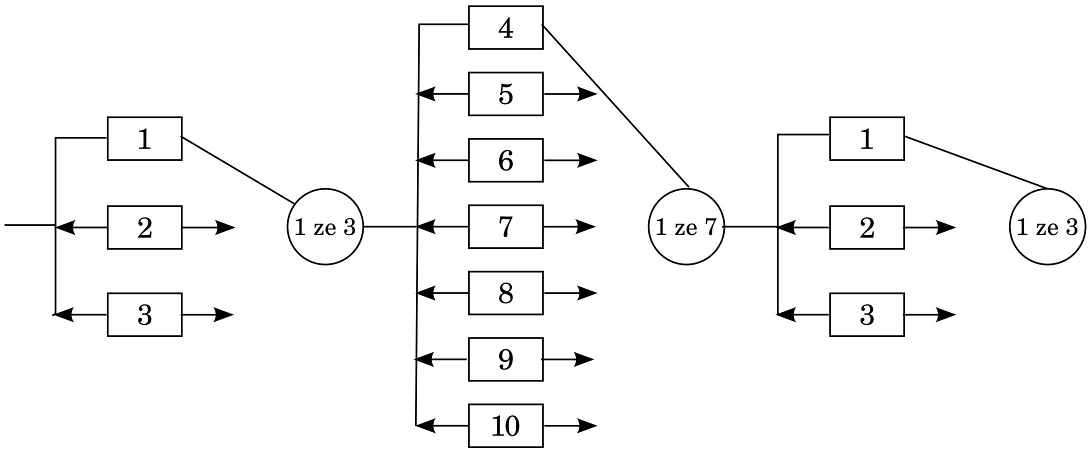


Figure 1: The reliability block diagram (RBD) of the compressor station for Configuration No.1

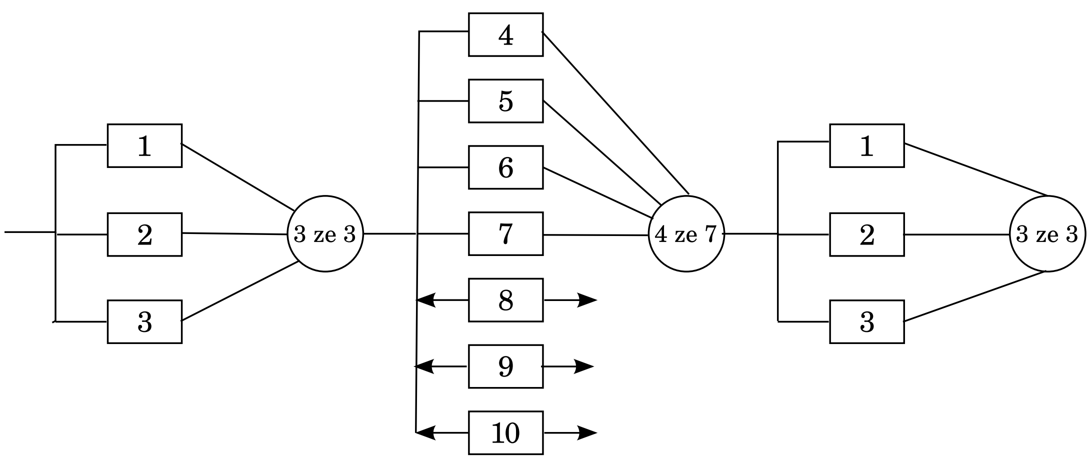


Figure 2: The reliability block diagram (RBD) of the compressor station for Configuration No.7

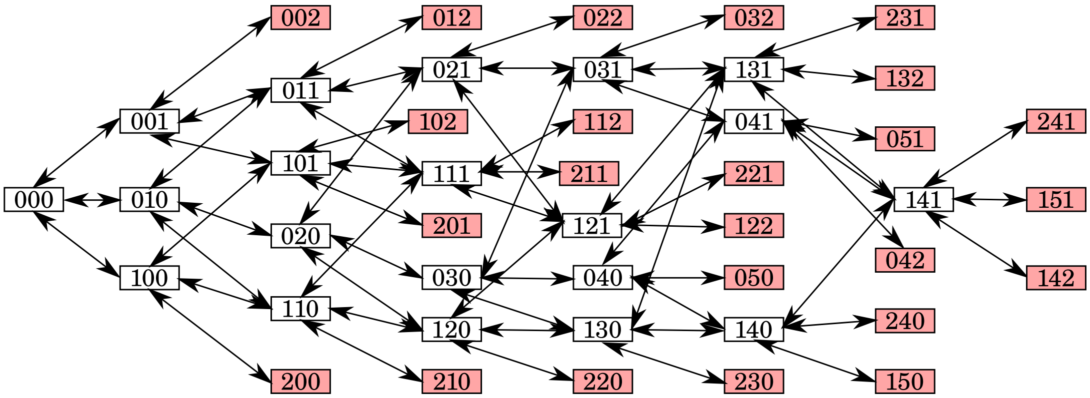


Figure 3: Markov model of the compressor station for Configuration No.1.

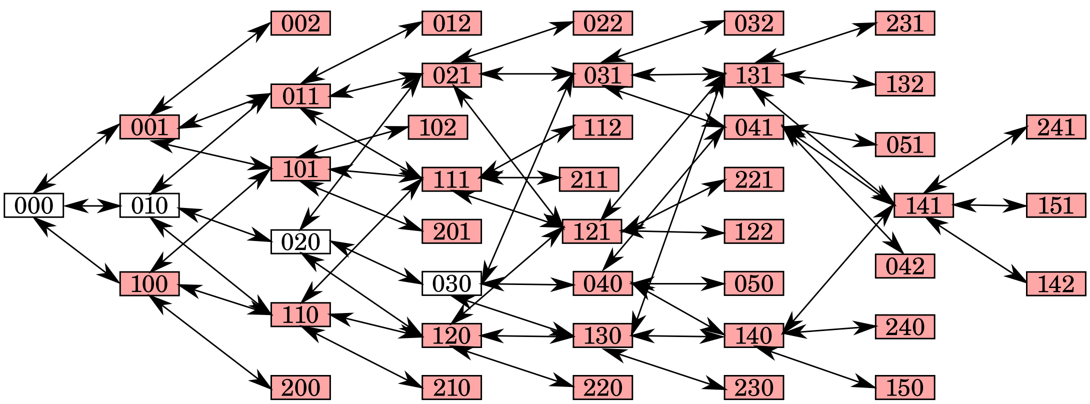


Figure 4: Markov model of the compressor station for Configuration No.7.

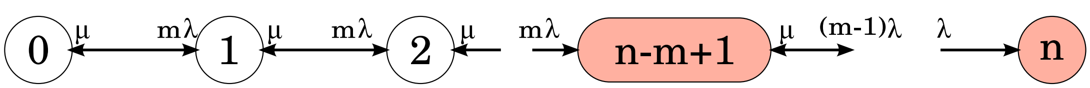


*Figure 5: Transition diagram for components with standby redundancy in the **m** out of **n** system*

3.4 Sestavení soustavy diferenciálních rovnic

Let us assume a transition diagram with n states. We can construct the $n \times n$ transition matrix $\mathbf{h}$ as follows: The entry h_{ij} represents the transition intensity between states i and j. Next, i-th component of a vector $\boldsymbol{p}(t)$ represents a probability that physical system will be in the time t in the state indicated by a symbol i. We also define the initial conditions by a vector $\boldsymbol{p}(t)$ for $t = 0$. Than we solve numerically the system

$$\frac{d\boldsymbol{p}(t)}{dt} = \boldsymbol{p}(t)h.$$

The result of modeling is an estimation of the availability of the station $A(t)$ in the time t. The availability in the time t is the sum of probabilities that the system is in the time t in a non-failure state.

4 Results

The input failure intensities and repair intensities presented in our contribution are not related to RWE Transgas. In order to minimize space of our contribution, we present results related to the minimal configuration (denoted by Configuration No.1) and the strongly loaded Configuration No.7. The Configuration No.7 is used only rarely.

4.1 System with constant failure rate and without periodic maintenance

The assumed failure intensity of lines and turbocompressor is $\lambda = \frac{1}{87600}h^{-1}$. The assumed renewal intensity of lines and turbocompressor is $\mu = \frac{1}{720}h^{-1}$.

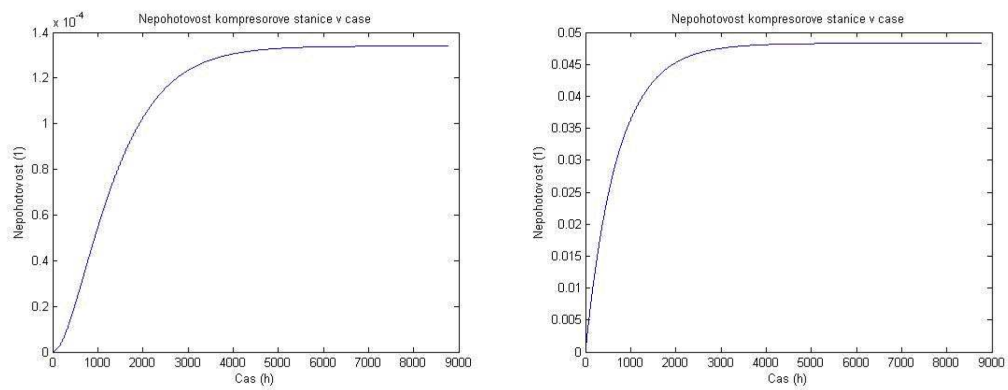


Figure 6: A system unavailability for Configuration No. 1 (left) and for Configuration No. 7 (right) with assuming the exponential distribution of time to failure.

Results are presented in Figure 6. We can observe that the history and also the asymptotic value of unavailability depend significantly on the assumed configurations. Of course, the system unavailability for Configuration No. 1 is lower than the system unavailability for Configuration No. 7 (because of standby redundancy of lines and turbocompressors in Configuration No. 1).

4.2 System with non-constant failure rate and with periodic maintenance

Non-constant failure rates (caused by degradation of components) can be described by Weibull models. Maximum likelihood estimation (MLE) can be used to calculate the parameters of the Weibull model from failure data [3, 4, 6]. In order to obtain a better approximation of the bathtub shaped failure rate model, we assume the following failure rate in our model: $h(t) = \lambda + \frac{\beta \cdot t^{\beta-1}}{\alpha^{\beta}}$. We used the following parameter values: $\alpha = 800000$, $\beta = 1.2$ and also parameters from Section 3.1.

We also assume the periodic system maintenance in our model. The periodic system maintenance of lines is half of year (4380 h). The periodic maintenance of turbocompressors is 3 months (2190 h). The unavailability of a system with non-constant failure rate and with periodic maintenance is presented in Figure 7. We can observe that the asymptotic unavailability does not exist in this case. The periodic maintenance is indicated by a step change of the system unavailability in the maintenance times. System unavailability depends significantly on the assumed configurations.

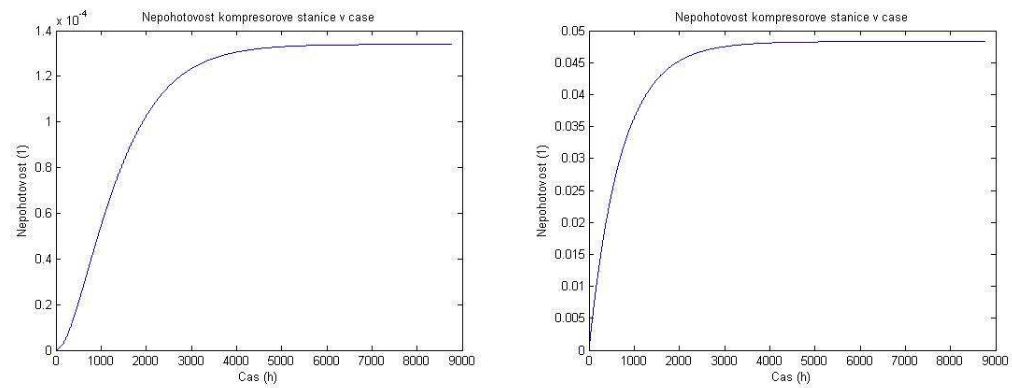


Figure 7: A system unavailability for Configuration No. 1 (left) and for Configuration No. 7 (right) with assuming the Weibull model of ageing with periodic system maintenance.

5 Modelling of the profit and loss

An every states of the transition diagram, which is mentioned at pic. 3 and 4 for some performance configurations, are described by number of failure lines and failure turbocompressors and volume of transported gas. The every states are evaluated by profit or loss during transportation of gas in unit's time. If the system is in function state, usually produces a profit. From the result of the differential equation mentioned in the chapter 2.6 is known probability in which is the system in time t in state k. It is possible to estimate profit of whole system by the sum of all states. See the first term in equation (1).

Transitions between the states are described by failure or restoration of components in the system. The failure or restoration of components becomes in one moment. The evaluation of the profit or loss is dependent on probability, that the system is in state i and transitions rate from state i to state j. See the second term in equation (1).The profit in time t is described by equation:

$$Cost(t) = \sum_{k \in \Omega} P_k(t) \cdot Cost_k + \sum_{i,j \in \Omega} P_i(t) \cdot h_{ij}(t) \cdot Cost_{ij}(1), \tag{1}$$

where$Cost(t)$whole unit profit in time t

$Cost_k$	unit profit of state k
$Cost_{ij}$	unit profit of transition between states i and j
$P_k(t)$	probability of the state k in time t
$h_{ij}(t)$	matrix of the transition rate between states

From the equation (1) we obtain the profit or loss in time t. For calculation of whole profit and loss to time t is necessary to integrate equation (1) by time. We obtain equation (2).

$$Cost_M(t) = \frac{1}{t} \int_0^t Cost(\tau) d\tau \tag{2}$$

By results of the profit to time t is possible to determine the best scenario for the transported gas in time. The resulted profit from the transportation to time t is function only of reliability analysis of system.

6 Conclusion

The aim of our contribution was to show approaches for modeling of system unavailability by Markov models. If only asymptotic values of the system unavailability are needed, then Markov models are not necessary. On the other hand, dynamic reliability models should be assumed in cases, where highly reliable systems are required (chemical industry, nuclear devices, railways, ...).

In conclusion it should be stressed the importance of reliability modelling of transit networks and thus to describe problems as well as to increase the availability of the whole system. It was shown on the example that by not using one line from threes the availability of the whole system increases greatly. It is worth noting that 25th of July 2006 a state of electric emergency lasted for all nine hours. It was announced due to an extremely big electricity failure in the Czech Republic, the biggest in the last 30 years. The transit of medium is governed by similar regulations either we take into account gas or electricity. We speak about distant transitions in which the Czech Republic is “only” the transit country. A few device failures can cause a failure of the whole system.

Acknowledgement

This research was supported by the Ministry of Education, Youth and Sports of the Czech Republic, project No. 1M0554 - Advanced Remedial Technologies and Processes.

References

[1] Praks P., Chudoba J., Bris R., Koucky M.: *Reliability analysis of a natural gas compression station and surrounding gas pipeline network with assuming of performance changes by a dispatcher*, In: Proceedings of the European Safety and Reliability Conference 2007 (ESREL 2007). Ed. Terje Aven&Jan Erik Vanen, London: Tailor&Francis Group, 2007, ISBN 978-0-415-44786-7

[2] ČSN IEC 60605-4 (01 0644), *Zkoušení bezporuchovosti zařízení – Část 4: Statistické postupy pro exponenciální rozdělení – Bodové odhady, konfidenČní intervaly, předpovědní intervaly a toleranČní intervaly.* ČNI Praha, 2002

[3] ČSN IEC 60605-6 (01 0644), *Zkoušení bezporuchovosti zařízení – Část 6: Testy platnosti předpokladu konstantní intenzity poruch nebo konstantního parametru proudu poruch.* ČNI Praha, 1998

[4] ČSN EN 61164 (01 0647), *Růst bezporuchovosti – Metody statistických testů a odhadů*, ČNI Praha 2005

[5] ČSN EN 61165 (01 0691), *Použití Markovových metod.* ČNI Praha, 2007

[6] `http://www.weibull.com`

Josef Jablonský

Department of Econometrics
University of Economics Prague
W. Churchill sq. 4, 130 67 Prague 3

e-mail: jablon@vse.cz

Data Envelopment Analysis Models for Resource Allocation

Abstract

The main aim of the paper is to verify alternative approaches for allocation of resources among departments within an organization. The paper introduces the approach based on standard DEA models. The main input is the amount of resources allocated to the departments by the current system. The outputs are the criteria influencing the performance of the department. Our approach is based on solving standard DEA model with constant or variable returns to scale. By the model the departments are split into two groups—efficient and inefficient. Our approach offers how to change inputs of efficient (increase of inputs, i.e. resources) and inefficient (decrease of inputs) units under the assumption that the outputs remain unchanged in order to move all the originally inefficient departments to the efficient frontier. The DEA results are compared to the current practice and to the results of several other multicriteria decision making methods. Numerical experiments with DEA models are realized on our own MS Excel based application that allows solving problems up to 200 decision making units and 20 inputs and 20 outputs. The paper contains a brief information about our DEA solver which covers basic DEA models including super-efficiency models, models with uncontrollable inputs and outputs and models with undesirable inputs and outputs. The system can be simply downloaded from my personal web pages and used without any restrictions.

Keywords: *data envelopment analysis, efficiency, allocation of resources, multiple criteria decision making*

1 Introduction

The problem of allocation of resources (financial funds or other kinds of resources) among given number of units (firms, regions, departments, etc.) for a next planning period can be recognized in many real situations. This allocation can be demanded on the departmental, regional, governmental or business levels. It is often based on evaluation and mutual comparison of performance and efficiency of the set of units taking into account in one or several past planning periods. Performance and efficiency are characteristics that cannot be defined easily and they always depend on several partial characteristics that can be more or less easily measured by decision makers. If the decision maker considers just a pair of characteristics—one of the output and one of the input natures—the efficiency can be defined by means of a simple ratio output/input. The situation that corresponds better to reality is that the decision maker considers more input and more output characteristics simultaneously. In this case more sophisticated tools have to be used. Multiple criteria decision making (MCDM) methods or data envelopment analysis (DEA) models which are specially developed for analysis of efficiency and performance of the set of decision making units can be used in this context. In most cases MCDM methods assign to evaluated units (alternatives) their utility which is defined specifically by each of the methods. This utility can be used as a base for allocation of resources among evaluated units. The paper aims at alternative approaches for allocation of resources based on DEA models and discussion of their wider applications. The proposed approach is demonstrated on a simple numerical example of a real nature. The paper is completed by information about an original software tool for solving DEA models within MS Excel environment developed at the University of Economics Prague.

2 Data Envelopment Analysis

DEA is a non-parametric method for measuring the relative efficiency and comparison of decision making units (DMU). Let us consider the set E of n decision making units $E = \{U_1, U_2, \ldots, U_n\}$. Each of the units produces r outputs and for their production spends m inputs. Let us denote $\mathbf{X}_j = \{x_{ij}, i = 1, 2, \ldots, m\}$ the vector of inputs and $\mathbf{Y}_j = \{y_{ij}, i = 1, 2, \ldots, r\}$ the vector of outputs for the DMU$_j$. Then $\mathbf{X} = (x_{ij}, i = 1, 2, \ldots, m, j = 1, 2, \ldots, n)$ is the matrix of inputs and $\mathbf{Y} = (y_{ij}, i = 1, 2, \ldots, r, j = 1, 2, \ldots, n)$ is the matrix of outputs. The measure of efficiency of the unit U_q can be expressed as follows:

$$\frac{\sum_j u_j y_{qj}}{\sum_p v_p x_{qp}}, \tag{1}$$

where $v_p, p = 1, 2, ..., m$ is the weight assigned to the the p-th input and $u_j, j = 1, 2, ..., r$ is the weight assigned to the i-th output. The level of efficiency defined by (1) can be estimated by means of standard DEA models.

DEA models evaluate the relative efficiency of the decision making units, i.e. they evaluate the efficiency of the given units comparing them to the other units of the set. The aim of the model is to estimate an efficient frontier which consists of so called efficient units. The standard DEA model with the assumption of constant returns to scale and with input orientation can be formulated as follows:

minimise

$$z(U_q) = \theta_q - \varepsilon(\mathbf{e}^T\mathbf{s}^+ + \mathbf{e}^T\mathbf{s}^-),$$

subject to

$$\mathbf{Y}^T\boldsymbol{\lambda} - \mathbf{s}^+ = \mathbf{Y}_q, \tag{2}$$

$$\mathbf{X}^T\boldsymbol{\lambda} + \mathbf{s}^- = \theta_q\mathbf{X}_q,$$

$$\boldsymbol{\lambda}, \mathbf{s}^+, \mathbf{s}^- \geq \mathbf{0},$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n)$, $\boldsymbol{\lambda} \geq 0$ is the vector of weights assigned to decision making units, $\mathbf{s}^+$ a $\mathbf{s}^-$ are vectors of slack variables in constrains for outputs and inputs respectively, $\mathbf{X}_q$ a $\mathbf{Y}_q$ are vectors of inputs and outputs of the evaluated unit U_q, θ_q is the radial variable for the unit U_q, $\mathbf{e}^T = (1, 1, \ldots, 1)$ and ε is the infinitesimal constant. The model can be explained as follows. When the unit U_q is evaluated the model looks for a virtual unit described by inputs $\mathbf{X}^T\boldsymbol{\lambda}$ and outputs $\mathbf{Y}^T\boldsymbol{\lambda}$ that are created by linear combination of inputs and outputs of the remaining units of the set and have better (or not worse) inputs and outputs than the evaluated unit U_q. For the inputs and outputs of the virtual unit the following relations must obviously hold: $\mathbf{X}^T\boldsymbol{\lambda} \leq \mathbf{X}_q$ and $\mathbf{Y}^T\boldsymbol{\lambda} \geq \mathbf{Y}_q$. The unit U_q is recognized as efficient (lying on the efficient frontier) iff there does not exists any virtual units with the above mentioned properties or in better words the virtual unit is equal to the evaluated unit, i.e. the relations $\mathbf{X}^T\boldsymbol{\lambda} = \mathbf{X}_q$ and $\mathbf{Y}^T\boldsymbol{\lambda} = \mathbf{Y}_q$ hold. This situation is recognized when the model (2) is solved as follows:

1. The optimal value of the radial variable θ_q is equal to 1,
2. The optimal values of all slacks equal to zero, i.e. $\mathbf{s}^+ = \mathbf{0}$ and $\mathbf{s}^- = \mathbf{0}$.

The unit U_q is efficient if the optimal value of the objective function of the model (2) $z(U_q)* = 1$. Otherwise the unit is not efficient. The optimal value $z(U_q)*$ is denoted as the efficiency score of the unit U_q. Lower value of the efficiency score means lower efficiency within the set of the evaluated units (the unit is farther from the efficient frontier). The value of the radial variable θ_q lower than 1 can be explained as the need of the relative reduction (i.e. improving) of inputs in order the unit U_q reaches the efficient frontier. This property of DEA models can be conveniently used in solving problems of allocation of resources. The units those are not efficient, i.e. they have reached with their given resources comparing to other units unsatisfactory outputs in the past planning period, have to reduce their inputs in order to be efficient in the future period, i.e. the allocated inputs for them have to be reduced for the next period.

The evaluation of efficiency of all DMUs of the decision set needs solving of n linear programming optimisation problems (2) or its modifications. Each of the LP programs (2) contains $(n + m + r + 1)$ variables and $(m + r)$ constraints. The LP programs are relatively small but their repeated solution for all the units can be time consuming especially for a higher number of units. Sometimes it can be more useful to formulate one bigger LP program that can find out the efficiency of all units by one optimisation run. This problem has very big number of variables and constraints—$n(n + m + r + 1)$ variables and $n(m + r)$ constraints—but its solving can be quite efficient by using a high quality professional optimisation solvers. This aggregated model can be formulated as follows:

minimize

$$\sum_{q=1}^{n} \left(\theta_q - \varepsilon \left(\sum_{p=1}^{r} s_{qp}^{+} + \sum_{j=1}^{m} s_{qj}^{-} \right) \right)$$

subject to

$$\sum_{i=1}^{n} y_{ij}\lambda_{iq} - s_{qp}^{+} = y_{qp}, \qquad p = 1, 2, ..., r, \;\; q = 1, 2, ..., n, \tag{3}$$

$$\sum_{i=1}^{n} x_{ij}\lambda_{iq} + s_{qj}^{-} = \theta_q x_{qj}, \qquad j = 1,2,...,m, \;\; q = 1,2,...,n,$$

$$\lambda_{qj} \geq 0, \;\; s_{qj}^{+} \geq 0, s_{qj}^{-} \geq 0, \theta_q \geq 0.$$

One of the important results of DEA models is information about level of efficiency of the evaluated units (efficiency score). Very important information for our purposes is to know how to improve inputs and/or outputs in order to reach the efficient frontier. This improved inputs and outputs for the unit U_q can be easily computed as the weighted sum of other inputs and:

$$\mathbf{X'}_q = \mathbf{X}^T \boldsymbol{\lambda}_q^*, \tag{4}$$

$$\mathbf{Y'}_q = \mathbf{Y}^T \boldsymbol{\lambda}_q^*,$$

where $\boldsymbol{\lambda}_q^*$ is the vector of optimal weights $\boldsymbol{\lambda}$ for pro the unit U_q. given by the model (2).

3 DEA Model for Resource Allocation

We suppose that each unit is characterised by the outputs of the last planning period and we look for the values of inputs (resources) that ensure the maximum efficiency of all units. In the first step we can use the inputs of the same values as in the last period. The basic idea for application of DEA models for allocation of resources consists in modification of resources (inputs) in this way that all the units will lay on the efficient frontier after this modification. We will formulate a DEA model that ensures the efficiency for all the units with the minimal changes in inputs (reduction) for inefficient units and outputs (expansion) for efficient ones. In order to ensure the efficiency of all the units their efficiency scores have to be equal to 1, i.e. $\theta_q = 1, \; q = 1,2,\ldots,n$ and all the slacks $\mathbf{s}^+$ a $\mathbf{s}^-$ equal to zero. Let us denote $alpha_{ij}, \; i = 1,2,\ldots, \; n, j = 1,2,\ldots,m$ the reduction of inputs for the units recognised as inefficient by the model (2) and $\beta_{ij}, \; i = 1,2,\ldots,n, \; j = 1,2,\ldots,m$ the expansion of resources for units that are identified as efficient by the model (2). The mathematical model of the optimization problem mentioned above can be formulated as follows:

minimize

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_{ij} + \beta_{ij})$$

subject to

$$\sum_{i=1}^{n} y_{ij}\lambda_{iq} = y_{qp}, \qquad p = 1,2,...,r, \;\; q = 1,2,...,n, \tag{5}$$

$$\sum_{i=1}^{n}(x_{ij} - \alpha_{ij})\lambda_{iq} = x_{qj} - \alpha_{qj}, \qquad j = 1,2,...,m, \;\; q \in \mathrm{U_e},$$

$$\sum_{i=1}^{n}(x_{ij}+\beta_{ij})\lambda_{iq}=x_{qj}+\beta_{qj}, \qquad j=1,2,...,m, \;\; q\in \mathrm{U}_n,$$

$$\lambda_{qj}\geq 0, \; \alpha_{ij}\geq 0, \beta_{ij}\geq 0,$$

where U_e is the set of indices of efficient units and Un is the set of indices of inefficient units. The model (5) is quite simple but unfortunately it is not linear model (there is a product of variables in constraints). Another problem is that the model (5) has not always a feasible solution. Due to the computational problems with solving of the model (5) we suggest for allocation of resources a simple iteration procedure. It can be described in the following steps:

1. The efficiency score $z^*(U_q)$ for all the units of the set $U_q, \; q=1,2,\ldots,n$ is derived by solving the problem (2). The efficiency score is either lower than 1 for inefficient units or equal to one for efficient ones. If all the units of the set are efficient the iteration process finishes.
2. For the units identified as efficient in the first step their supper-efficiency scores are computed—any of the DEA super-efficiency models can be applied (see [5]). It is a characteristics with value greater than 1 that allows discriminate among the efficient units because all they have their efficiency score the same (1). The greater value of the super/efficiency indicates higher efficiency level.
3. The reduction of all inputs (resources) for all inefficient units relatively by using of their efficiency scores. By this reduction some of the currently allocated resources among inefficient units are freed.
4. The free resources from the previous step will be assigned to the efficient units relatively according to their super-efficiency scores.
5. Go to step 1.

4 A Numerical Example

We will illustrate the proposed approach on a simple example with a real background. It is the problem of allocation of financial funds among departments of a faculty. Let us consider that the funds are allocated according to the outputs of the past teaching periods and we will take into account the most important three output characteristics—the number of hours of direct and indirect teaching and the quality of research measured by the publication activity of departments measured by publication points. Table 1 contains hypothetical data for 9 departments. In the first column is an initial allocation of funds given by the faculty custom practise. The next columns show the mentioned characteristics describing the direct and indirect teaching and the quality of research. The pre-last column presents the efficiency score given by the model (2) for inefficient units (lower than 1) and super-efficiency score for efficient units computed by Andersen and Petersen model—see [5]. The last column shows how to improve (reduce) the single input (funds) in order the unit reaches the efficient frontier—the values are computed by (4). These values give the needed reduction of funds for inefficient departments in order to reach the full efficiency. The data in Table 1 show that this reduction frees the total amount of 1030 thousands CZK. This amount can be split among the originally efficient units. This secondary splitting can be done according to the super-efficiency scores presented in Table 1. The results of a new allocation with efficiency scores computed by model (2) with new data set, i.e. with the new allocation of funds, are presented in Table 2.

Dept.	**Funds [thous. CZK]**	**Direct [hours.]**	**Indirect [hours]**	**Research [points]**	$\mathbf{z * (U_q)}$	**Reduction of funds**
K1	1500	1598	571	4.25	0.7885	1183
K2	2000	2006	552	9.78	0.8427	1685
K3	4000	4540	2422	15.75	0.9654	3862
K4	7000	7627	3794	32.75	0.9628	6740
K5	4000	5387	1587	22.47	1.1854	4000
K6	1000	474	93	6.65	1.1838	1000
K7	7000	8601	4194	26.00	1.0015	7000
K8	3000	4302	1212	11.33	1.0648	3000
K9	5000	5980	3136	22.83	1.0801	5000
Sum	**34500**					**33470**

Table 1: Input data and DEA results

The final allocation of the given amount of funds computed by the presented approach after three-step process is presented in the last two columns of Table 2.

The results given by our procedure can be compared to other approaches. They are compared to the results of the DEA model with uncontrollable outputs—DEA_NO. It is the model (2) with additional assumption that all the positive slacks $\mathbf{s}^+ = 0$. Except the DEA_NO results we will present in Table 3 the results of splitting of funds computed by custom practice. It is based on MCDM methodology with weights for research 30 % and for teaching (direct plus indirect) 70 %. In case of DEA model with uncontrollable outputs the above described splitting procedure is used. Table 3 does not describe how the results were derived but compares them each other only.

The comparison of results shows that the allocation is very close in the first and the third case except the department K6 that takes much more funds in the first case comparing to the MCDM methodology. Big differences in allocation of funds comparing to the remaining two approaches are identified in the second approach (DEA_NO). A detailed analysis of reasons of presented differences in allocation will be a subject for further research.

The presented approach needs solving of standard DEA models. In our experiments we have worked with our original DEA application that was created as an add-in application for MS Excel. This application can be downloaded form my personal web page `http://nb.vse.cz/~jablon`. It covers the most often used DEA models as presented in the following list:

Dept.	**Funds step 2**	$\mathbf{z*(U_q)}$	**Reduction step 3**	**Funds step 3**	$\mathbf{z*(U_q)}$
K1	1183	1.0264	1183	1190	1.0000
K2	1685	1.1247	1685	1736	1.0000
K3	3862	1.0496	3862	3908	1.0000
K4	6740	1.0461	6740	6814	1.0000
K5	4499	0.9667	4350	4350	1.0000
K6	1124	1.0198	1124	1129	1.0000
K7	7007	1.0088	7007	7022	1.0000
K8	3131	1.0170	3131	3144	1.0000
K9	5270	0.9883	5208	5208	1.0000
Sum	**34500**		**34288**	**34500**	

Table 2: Allocation of resources—steps 2 and 3.

Dept.	**DEA Model** (1)	**DEA_NO model**	**MCDM WSA**
K1	1190	1523	1193
K2	1736	1834	1730
K3	3908	4061	3969
K4	6814	6774	6982
K5	4350	4061	4432
K6	1129	1015	689
K7	7022	7108	7093
K8	3144	3046	3065
K9	5208	5077	5347
Sum	**34500**	**34500**	**34500**

Table 3: Allocation of resources—comparison of results.

- envelopment models with constant, variable, non-decreasing and non-increasing returns to scale including super-efficiency option,
- additive models (slack based models),
- models with uncontrollable inputs and outputs, and
- models with undesirable inputs and outputs.

The application is user-friendly and offers solving DEA models up to 200 decision making units and 20 inputs and 20 outputs. The application needs not any other special requirements except standard internal MS Excel optimisation solver that have to be activated before one uses the application. The application was described in more detail in [7].

5 Conclusions

Allocation of resources is an interesting class of problems. Solving of such problems consists in comparison of past outputs of the set of evaluated units and based on this comparison the new resources are allocated to the units for the next period. In the paper we have formulated models for allocation of resources based on DEA models. The proposed approaches were tested on a simple numerical example and compared to the results given by MCDM methodology (a simple additive utility function). The results given by both methodical approaches show their relative closeness. The paper can be considered as a contribution to discussion about alternative approaches for solving this class of problems.

Acknowledgements

The research is supported by the Grant Agency of the Czech Republic—grant no. 402/06/0150.

References

[1]] CHARNES, A., COOPER, W.W., LEWIN, A., SEIFORD, L.: *Data Envelopment Analysis: Theory, Methodology and Applications.* Boston, Kluwer Publ. 1994.

[2]] COOPER, W.W., SEIFORD, L.M, TONE, K.: *Data Envelopment Analysis. Boston.* Kluwer Publ. 2000.

[3]] FANG, L., ZHANG, C.Q.: *Resource Allocation Based on the DEA Model.* Journal of the Opera-tional Research Society, to appear.

[4]] JABLONSKÝ, J.: *Modely hodnocení efektivnosti produkčních jednotek.* Politická ekonomie, č. 2, 2004, s. 206-220.

[5]] JABLONSKÝ, J., DLOUHÝ, M.: *Modely hodnocení efektivnosti produkčních jednotek* (*Models for evaluation of efficiency of production units*). Professional Publishing, Praha 2004.

[6]] JABLONSKÝ, J.: *Measuring the Efficiency of Production Units by AHP Models.* Mathematical and Computer Modelling, 46(2007), pp. 1091-1098.

[7]] JABLONSKÝ, J.: A *MS Excel Based Support System for Data Envelopment Analysis Models*. In: SKALSKÁ, Hana (ed.). *Mathematical Methods in Economics 2005.* Hradec Králové : Gaudeamus, 2005, s. 175–181.

[8]] ZHU, J.: Quantitative Models for Performance Evaluation and Benchmarking. Boston, Kluwer Publ. 2003.

Tomáš Jurík

Faculty of Mathematics, Physics and Informatics
Comenius University in Bratislava
Mlynska dolina, 842 48 Bratislava, Slovakia

e-mail: tomas.jurik@fmph.uniba.sk

Mixed Integer Linear Programming in Credit Scoring System

Abstract

Each prospective client is obliged to fulfill a questionnaire based on which a credit score is calculated. The score reflects a quality of the client: the higher score, the better client. One of the management goals is to divide all clients into a few groups according to their scores. These groups should be kind of intrinsically homogenous (the clients within one group have approximately coincident probability of default) and mutually heterogeneous (the clients from different groups have significantly different probability of default). Historical records provide an estimate of the probability of default for every single score. The aim is to calculate the critical score values that split the clients into the homogeneous groups.

This model leads to a min-max linear programming problem that can be transformed into a mixed integer linear programming (MILP) problem. In general, solving MILP problems even with hundreds of variables is hard, computationally exhaustive. Although a corresponding exact model has an excessive number of integer variables, some further derivations can significantly reduce the number of variables and tighten the set of feasible solutions. These derivations can be performed due to the special structure of the MILP problem. With these modifications we achieved a considerable reduction in the computation time.

Keywords: *optimization, mixed integer linear programming, credit score, risk management, probability of default*

1 Introduction

Since the early 1950s, many researchers have studied many mixed integer linear programming (MILP) problems. Those problems arose from different applications, mainly in transportation and manufacturing (see, e.g. [1], [7]) but also in digital electronics [6], biocomputing [3] and economics [2]. It is known that a general MILP problem is NP-complete but despite of this fact the easiest solution to some large scale optimization problems is provided by the MILP algorithms.

In this paper we present one optimization problem from a risk management that leads to a MILP problem. An institution assings to each client a score from the set $\{s_j, 1 \leq j \leq N\}$. The management is faced to the problem to divide all clients into a few groups G_k, $(1 \leq k \leq K)$ according to their scores. The ratios of the number of defaulted clients to the number of all clients within a group G_k have to be close to the given values p_k $(1 \le k \le K)$. The aim is to calculate the critical score values a_k $(0 \le k \le K)$ which split the clients into these groups. Although a corresponding exact model has an excessive number of integer variables, some further derivations can significantly reduce the number of variables and tighten the set of feasible solutions.

In the rest of this section we establish some notation. Section 2 describes a corresponding mathematical model and clarifies our modifications. Section 3 presents the computational results and the last section is devoted to the remarks and suggestions.

For each index $j = 1, \ldots, N$ we denote the number of clients with a score s_j by n_j and the number of clients with a score s_j that have defaulted by d_j $(0 \le d_j \le n_j)$. For a simplicity in the following, we set the scores to $s_j = j - 1/2$ for $j = 1, \ldots, N$.

The problem is to find the critical integer score values $0 = a_0 \le a_1 \le \cdots \le a_K = N$ which define the groups G_k $(1 \le k \le K)$ of clients with a score s_j such that $a_{k-1} < s_j < a_k$. The number of defaulted clients has to be close to the expected number of defaulted clients within a group G_k provided by the probability of default equal to p_k, i.e.

$$\sum_{j: a_{k-1} < s_j < a_k} d_j \quad \sim \quad p_k \sum_{j: a_{k-1} < s_j < a_k} n_j$$

for each group G_k $(1 \le k \le K)$.

2 Mathematical Model

For a fixed integer k we define the auxiliary variables $y_{j,k}$ $(1 \le j \le N,\ 1 \le k \le K)$ which decide whether a score s_j fulfills $a_{k-1} < s_j < a_k$, i.e. whether the clients with a score s_j belong to the group G_k. We define the values of the auxiliary variables $y_{j,k}$ by the formula

$$y_{j,k} = \begin{cases} 1, & \text{if } a_{k-1} < s_j < a_k \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

then a mathematical model for the aforementioned problem is to minimize the sum

$$\sum_{k=1}^{K} \left| y_{1,k} \left(d_1 - p_k n_1 \right) + y_{2,k} \left(d_2 - p_k n_2 \right) + \cdots + y_{N,k} \left(d_N - p_k n_N \right) \right| \tag{2}$$

for the given number of clients n_j $(1 \le j \le N)$, the number of defaulted clients d_j $(1 \le j \le N)$ and the apriori probability of default p_k $(1 \le k \le K)$ within each group G_k. The described mathematical model is not a MILP problem. We show in the following how to rewrite it into the desired form. In section 3 we solve it with a standard routine for MILP problems.

2.1 Mixed Integer Linear Programming Model

In the rest of the paper we consider $j = 1, \ldots, N$ and $k = 1, \ldots, K$ unless stated otherwise, where N is a number of different scores and K is the desired number of groups. The objective function (2) is linear in its variables $y_{j,k}$, we have to handle the definition (1): we have $y_{j,k} = 0$

if and only if (iff) both numbers $a_k - s_j$, $s_j - a_{k-1}$ are positive (due to definition of s_j and since a_k are integers). Let us denote

$$t_{j,k} = 2\min\{a_k - s_j, s_j - a_{k-1}\}. \tag{3}$$

We remark that $t_{j,k}$ is for all indices j and k an nonzero integer such that if $t_{j,k} \leq -1$ then $y_{j,k} = 0$ and if $t_{j,k} \geq 1$ then $y_{j,k} = 1$. Here, if we set

$$u_{j,k} = \max\{0, t_{j,k}\} \tag{4}$$

then $u_{j,k} = 0$ iff $t_{j,k} \leq -1$. Finally, we can define $y_{j,k}$ equivalently to the definition (1) by the formula

$$y_{j,k} = \min\{1, u_{j,k}\}. \tag{5}$$

We have added the auxiliary variables $t_{j,k}$, $u_{j,k}$ to the model and the last step in order to reach a MILP model is to write the max (min) constraints as some inequalities. It is known that a constraint $x = \max\{y, z\}$ for the variables x, y, z is equivalent to the four inequalities (cf. [6])

$$\begin{array}{rclcrcl} x & \geq & y & \qquad & x - y & \leq & M \cdot B \\ x & \geq & z & \qquad & x - z & \leq & M \cdot (1 - B), \end{array} \tag{6}$$

where M is a sufficiently large constant and B is a new binary integer variable ($B \in \{0, 1\}$). By the sufficiently large value of M we mean the largest possible value of $\max\{x - y, x - z\}$ among all feasible values of x, y, z. We note that if $y \neq z$ then $B = 0$ iff $x = y$. To avoid ambiguity, we set $B = 0$ in the case $y = z$. Using of aforementioned equalities we rewrite our problem into a MILP problem

$$\begin{array}{rcl} w_1 + w_2 + & \cdots & + w_K \to \min \\ -w_k \leq y_{1,k}\,(d_1 - p_k n_1) + y_{2,k}\,(d_2 - p_k n_2) + & \cdots & + y_{N,k}\,(d_N - p_k n_N) \leq w_k \end{array} \tag{7}$$

$$\begin{array}{rclcrcl} t_{j,k} & \leq & 2(a_k - s_j) & \qquad & u_{j,k} & \geq & t_{j,k} \\ t_{j,k} & \leq & 2(s_j - a_{k-1}) & & u_{j,k} & \geq & 0 \\ 2(a_k - s_j) - t_{j,k} & \leq & BT_{j,k} M_T & & u_{j,k} - t_{j,k} & \leq & BU_{j,k} M_U \\ 2(s_j - a_{k-1}) - t_{j,k} & \leq & (1 - BT_{j,k}) M_T & & u_{j,k} & \leq & (1 - BU_{j,k}) M_U \\ & & & & & & \\ y_{j,k} & \leq & 1 & & & & \\ y_{j,k} & \leq & u_{j,k} & & s_j & = & j - 1/2 \\ 1 - y_{j,k} & \leq & BY_{j,k} M_Y & & 0 = a_0 \leq a_1 \leq & \cdots & \leq a_K = N \\ u_{j,k} - y_{j,k} & \leq & (1 - BY_{j,k}) M_Y & & & & \end{array}$$

for the binary variables $BT_{j,k}$, $BU_{j,k}$, $BY_{j,k}$, integer variables $y_{j,k}$, $t_{j,k}$, $u_{j,k}$, $a_0, a_1, \ldots, a_K$, and real value variables $w_1, w_2, \ldots, w_K$. All constraints are considered for $k = 1, \ldots, K$, $j = 1, \ldots, N$.

Any solution $a_0, a_1, \ldots, a_K$ of this problem represents the optimal critical scores for the groups of clients.

j	1	2	$\cdots$	$a_{k-1}-1$	a_{k-1}	$a_{k-1}+1$	$\cdots$	a_k	a_k+1	a_k+2	$\cdots$	N
$y_{j,k}$	0	0	$\cdots$	0	1	1	$\cdots$	1	0	0	$\cdots$	0
$u_{j,k}$	0	0	$\cdots$	0	1	3	$\cdots$	1	0	0	$\cdots$	0
$t_{j,k}$	1–2a_{k-1}	3–2a_{k-1}	$\cdots$	-1	1	3	$\cdots$	1	-1	-3	$\cdots$	
$BY_{j,k}$	1	1	$\cdots$	1	0	0	$\cdots$	0	1	1	$\cdots$	1
$BU_{j,k}$	1	1	$\cdots$	1	0	0	$\cdots$	0	1	1	$\cdots$	1
$BT_{j,k}$	1	1	$\cdots$	1	1	1	$\cdots$	0	0	0	$\cdots$	0

Table 1: Structure of the problem variables

Estimations of "Sufficiently Large Constants". In order to make the feasible region of the MILP problem (7) smaller we set constants M_T, M_U, M_Y as small as possible. As a consequence of the inequalities $0 = a_0 \le a_1 \le \cdots \le a_K = N$, $0 < s_j = j - 1/2 < N$ and the equalities (3), (4), (5) we have

$$-2(N-1/2) \le t_{j,k} \le 2(N/2-1/2), \quad 0 \le u_{j,k} \le N-1, \quad 0 \le y_{j,k} \le 1.$$

Therefore we set the values $M_T = 4N-2$, $M_U = 3N-2$ and $M_Y = N-1$.

2.2 Improved Model

The main disadvantage of the problem (7) is that it has $6KN + K + 1$ integer variables. A deeper insight into the problem reveals that these variables keep a special structure which is defined only by the values a_k $(0 \le k \le K)$ as shown in Table 1 and that some of the variables are reundant. As a result we conjecture some simple observations which are summarized in Theorem 5.

Theorem 5. *Let $t_{j,k}$, $u_{j,k}$, $y_{j,k}$, $BT_{j,k}$, $BU_{j,k}$, $BY_{j,k}$, a_k, w_k for $j = 1,\ldots,N$, $k = 1,\ldots,K$ and $a_0 = 0$ be any feasible solution of the problem (7). Then*

$$1 - y_{j,k} = BY_{j,k} = BU_{j,k} \tag{8}$$

$$\sum_{j=1}^{N} y_{j,k} = a_k - a_{k-1} \text{ for } 1 \le k \le K, \quad \sum_{k=1}^{K} y_{j,k} = 1 \text{ for } 1 \le j \le N \tag{9}$$

$$\sum_{j=1}^{N} BT_{j,k} = \left\lfloor \frac{a_{k-1}+a_k}{2} \right\rfloor \text{ for } 1 \le k \le K. \tag{10}$$

where $\lfloor x \rfloor$ denotes the integer part of x. Moreover, if $y_{j,k}, a_k \in \mathbb{Z}$ and $s_j = j - 1/2$, then $t_{j,k}, u_{j,k} \in \mathbb{Z}$.

Proof. The basic idea in a proof is to reuse the definitions and the inequalities in the model (7) properly. From the definition (1) we have that $y_{j,k}$ can gain only two values, the same as binary variables $BY_{j,k}$ and $BU_{j,k}$ do.

First we prove that $y_{j,k} = 0$ iff $BY_{j,k} = 1$. If $y_{j,k} = 0$ then from the inequality $1 - y_{j,k} \leq BY_{j,k}M_Y$ we have $BY_{j,k} = 1$. Reversely, if $BY_{j,k} = 1$ then from the definition of $BY_{j,k}$ due to (6) we have $y_{j,k} = u_{j,k}$ and then from (5) we have $1 \neq u_{j,k} = y_{j,k}$ therefore $y_{j,k} = 0$. We just proved that $1 - y_{j,k} = BY_{j,k}$. Similarly can be proved the equality $1 - y_{j,k} = BU_{j,k}$.

The value of $y_{j,k}$ for a fixed index k equals 1 iff a value of s_j fulfills $a_{k-1} < s_j < a_k$, which is the same case as $s_j \in A_k = \{a_{k-1}+1/2, a_{k-1}+3/2, \ldots, a_k-1/2\}$. Obviously $|A_k| = a_k - a_{k-1}$, which proves the first equality in (9). Consider the case $y_{j,k} = 1$ for a fixed index j. This happens only if $s_j \in [a_{k-1}, a_k)$. Since the intervals $[a_{k-1}, a_k)$ for $k = 1, \ldots, N$ are disjoint and its union is $[0, N)$ then $s_j \in [a_{k-1}, a_k)$ is fulfilled for exactly one index k. We proved the second equality in (9).

By using (3) and inequalities for the variable $BT_{j,k}$ in the model (7) we have that $BT_{j,k} = 1$ iff for an index j we have $a_k - s_j > s_j - a_{k-1}$. This is true for indices $j = 1, \ldots, \lfloor a_{k-1} - a_k \rfloor /2$ and the equality (10) is proved.

If $y_{j,k}$ and a_k are integer variables for all possible indices j and k then obviously $t_{j,k} \in \mathbb{Z}$ since $BT_{j,k}$ is a binary variable provided that $s_j = j - 1/2$ for $j = 1, \ldots, N$.

Since $y_{j,k}$ is a binary variable then from the inequalities (11) we have that $u_{j,k} \in \mathbb{Z}$.

$\square$

As a consequence of Theorem 5 the MILP problem (7) can be reduced to the improved form

$$
\begin{aligned}
w_1 + w_2 + \cdots + w_K &\to \min \\
-w_k \leq y_{1,k}\,(d_1 - p_k n_1) + y_{2,k}\,(d_2 - p_k n_2) + \cdots + y_{N,k}\,(d_N - p_k n_N) &\leq w_k
\end{aligned}
\tag{11}
$$

$$
\begin{array}{rcl}
t_{j,k} & \leq & 2(a_k - s_j) \\
t_{j,k} & \leq & 2(s_j - a_{k-1}) \\
2(a_k - s_j) - t_{j,k} & \leq & BT_{j,k}M_T \\
2(s_j - a_{k-1}) - t_{j,k} & \leq & (1 - BT_{j,k})M_T
\end{array}
\qquad
\begin{array}{rcl}
u_{j,k} & \geq & t_{j,k} \\
u_{j,k} & \geq & 0 \\
u_{j,k} - t_{j,k} & \leq & (1 - y_{j,k})M_U
\end{array}
$$

$$
\begin{array}{rcl}
y_{j,k} & \leq & 1 \\
y_{j,k} & \leq & u_{j,k} \\
u_{j,k} - y_{j,k} & \leq & y_{j,k}M_Y
\end{array}
\qquad
\begin{array}{rcl}
s_j & = & j - 1/2 \\
0 = a_0 \leq a_1 \leq & \cdots & \leq a_K = N
\end{array}
$$

also with the inequalities (9) and (10) in the binary variables $BT_{j,k}$, integer variables a_0, a_k, $y_{j,k}$ and real value variables $t_{j,k}$, $u_{j,k}$ and w_k ($j = 1, \ldots, N$, $k = 1, \ldots, K$). The improved model contains only $2KN + K + 1$ integer variables and has "tighter" feasible set.

3 Numerical Results

In this section we describe the numerical tests that we provided with help of the *lp_solve* routine - a free MILP solver based on the revised simplex method and the Branch-and-bound method (e.g. [4]).

In our test cases we set $K = 4$ and solved the problem to find the critical score values a_k for a different values of N. The desired ratios in the groups were $p_1 = 3\%$, $p_2 = 5\%$, $p_3 = 10\%$,

Table 2: Computational effort for the improved MILP model

N	*problem size*	*time(s)*	*iter.*	*opt. value*
10	474×173	*0.097*	*209*	*1.53*
15	699×253	*4.777*	*7902*	*3.27*
20	924×333	*17.429*	*22248*	*4.60*
25	1149×413	*64.544*	*67878*	*5.10*
30	1374×493	*167.370*	*148514*	*2.39*
35	1599×573	*329.678*	*266271*	*4.40*
40	1824×653	*529.026*	*354382*	*0.87*
45	2049×733	*922.863*	*568454*	*1.60*
50	2274×813	*2065.350*	*1199625*	*3.93*

$p_4 = 25\%$. We assume that the number of clients with a same score is normally distributed, therefore we set

$$n_j = \left\lfloor 1000 \cdot \varphi\left(\frac{4\ |(N+1)/2 - j|}{N}\right)\right\rfloor, \quad d_j = \left\lfloor 100 \cdot \varphi\left(\frac{4\ |2/3(N+1) - j|}{N}\right)\right\rfloor$$

where $\varphi(x)$ denotes the density function of the standard normal distribution. The results are summarized in Table 2: the first two columns describe the size of the problem, the next two columns denote the CPU time and the number of iterations. For a comparison to the original model (7) - the problem for $N = 10$ uses 678 seconds of CPU time (and 5 525 710 iterations) and the case $N = 15$ has not been solved within one hour. Notice that the number of integer variables for the largest test case is 405. The MILP problem in such dimensions is considered as medium size problem and solving it can take a few hours[1].

4 Remarks

The described model calculates the values of critical scores for a single scorecard. This simple problem can be solved more efficiently using dynamic programming approach (see [5]). But this is not a case when some additional criteria are demanded, for example: the number of clients within the smallest group could not be smaller than 10% of the number of clients within the largest group. On the other side, these criteria can be easily included in our model.

The model could be also easily widen to a case that each client has more scores due to the different scorecards. Then an additional objective is to set the critical scores such that all scorecards reflect a similar distribution of clients.

The constraint matrix of the described MILP problem has a special sparse block structure, therefore we conjecture that the developing of the special branch-and-bound algorithm suited for this problem could shorten the calculation time a little more.

[1] http://miplib.zib.de/miplib2003.php

Acknowledgement

The author is grateful to Juraj Lörinc and Róbert Zvonár for their helpful support during the evolution of the proposed model and also for their comments and remarks on an earlier draft of the paper.

References

[1] Heilporn, G., Giovanni, L.D., Labbé, M.: Optimization models for the single delay management problem in public transportation, European Journal of Operational Research 189 (3), 762–774 (2008)

[2] Benatia, S., Rizzi, R.: A mixed integer linear programming formulation of the optimal mean/Value-at-Risk portfolio problem, European Journal of Operational Research 176 (1), 423–434 (2007)

[3] Dasika, M.S., Gupta, A., Maranas, C.D.: A mixed integer linear programming (MILP) framework for inferring time delay in gene regulatory networks, The Pacific Symposium on Biocomputing, 474–485 (2007)

[4] Linderoth, J.T., Ralphs, T.K.: Noncommercial Software for Mixed-Integer Linear Programming, Integer Programming: Theory and Practice, John Karlof (ed.), CRC Press Operations Research Series, 253–303 (2005)

[5] Hubert, L.J., Arabie, P., Meulman, J.: Combinatorial data analysis: Optimization by dynamic programming, SIAM, Philadelphia (2001)

[6] Burks, T.M., Sakallah, K.A.: Min-max linear programming and the timing analysis of digital circuits, ICCAD '93: Proceedings of the 1993 IEEE/ACM international conference on Computer-aided design, 152–155 (1993)

[7] Nemhauser, G. L., Wolsey, L. A.: Integer and Combinatorial Optimization, John Wiley & Sons, New York, (1988)

Revaz Kakubava, Nino Svanidze

Department of Mathematics
Rustaveli State University
Ninoshvili 35, Batumi, 6010, Georgia

e-mail: `r.kakubava@gmail.com, n.svanize@gmail.com`

Profit and Probabilistic Analysis of Warm Standby System

Abstract

We consider the profit and probabilistic analysis of standby system, in which m-number of main elements operating simultaneously in parallel, and n- number of warm-standby elements. All of them are undergoing to failures. The failure times are assumed exponentially distributed. There is a single server for renewal in the system. Renewal time is generally distributed.

The analytical model of system is constructed in the form of closed queuing system with two types of request sources. The model is studied in steady-state condition. We have obtained the explicit expressions for transient probability and expected profit function per unit time for the investigated system. The independent variables of this function are initial characteristics of the system and output probabilistic characteristics.

Economical analysis of the system is formulated as follow: some initial characteristics of the considered system are to be selected so that the profit function would accept the extreme value in the permissible range on change for these initial characteristics. The model enables to choose optimal number of standby element as well as the parameters of renewal time distribution for standby system on economical criterion.

We complete our analysis by numerical examples for different distribution functions of renewal times.

Keywords: *probabilistic characteristics, warm standby, queieng system, distribution, renewal.*

1 Description of the system

We consider a standby technical system with m identical elements operating simultaneously in parallel and n warm-standbys. All of them are undergoing to failures. Each of elements fails independently of state of others and has an exponential time-to-failure distribution for main elements with parameters α, and with - β,$(0 < \beta < \alpha)$ for the spares elements. For normal functioning of the system all main elements should be maintained in operating condition. Although the system continues to function when main elements number is reduced, the effectiveness of system functioning drops down. As a result it becomes necessary to replace the failed main element with a standby one.

The failed main element is immediately replaced by a standby element if one is available. It is assumed that when a spare moves into an operating state, its failure characteristics will be that of an operating element.

Both main and standby failed elements are transmitted for renewal. After service they are treated as a spare. There is one server performing renewal.

The renewal times are generally distributed. Distribution functions and probability density function of renewal times are $G(t)$ and $g(t)$, $(t \geq 0)$.

2 Mathematical model of the system

The closed queuing system with two sources of requests represents a mathematical model of the considered standby system.

The state of the system at time t is given by:

$i(t)$ is the number of failed element at moment t;

$\xi(t)$ is time passed after the beginning of renewal at moment t;

Let's denote intensities of replacement by $\rho(u)$. Then $\rho(u) = \frac{g(u)}{1-G(u)}$.

And define probability characteristics:

$$P(0,t) = P\{i(t) = 0\};$$

$$p(i,t,u) = \lim_{h\to 0}\left(\frac{1}{h}\mathrm{P}\{i(t) = i,\ u \leq \xi(t) < u+h\}\right), i = \overline{1,\ m+n}$$

Suppose, function $P(0,t)$ a continuous derivative when $t > 0$and functions $p(i,t,u)$ have continuous partial derivatives when $t > 0$, $u \geq 0$.

Relating to the states of the system at time t and $t+dt$, we obtain

Theorem 1. In the mentioned functions $P(0,t)$ and $p(i,t,u)$ satisfy the following system of integral-differential and partial differential equations:

$$\frac{dP(0,t)}{dt} = -[m\alpha + n\beta]P(0,t) + \int_0^t p(1,t,u)\rho(u)\,du$$

$$\frac{\partial p(1,t,u)}{\partial t} + \frac{\partial p(1,t,u)}{\partial u} = -[m\alpha + (n-1)\beta + \rho(u)]p(1,t,u)$$

$$\frac{\partial p(i,t,u)}{\partial t} + \frac{\partial p(i,t,u)}{\partial u} = -[m\alpha + (n-i)\beta + \rho(u)]p(i,t,u) + [m\alpha + (n-(i-1))\beta]p(i-1,t,u) \tag{1}$$

$$2 \leq i \leq n$$

$$\frac{\partial p(i,t,u)}{\partial t} + \frac{\partial p(i,t,u)}{\partial u} = -[(m+n-i)\alpha + \rho(u)]p(i,t,u) + (m+n-i+1)\alpha p(i-1,t,u)$$

$$n < i < m+n,$$

$$\frac{\partial p\left(m+n,t,u\right)}{\partial t}+\frac{\partial p\left(m+n,t,u\right)}{\partial u}=-\rho\left(u\right)p\left(m+n,t,u\right)+\alpha p\left(m+n-1,t,u\right).$$

Theorem 2. Functions $p\left(i,t,u\right)$ satisfy the following equalities:

$$p\left(1,t+h,0\right)h=\left[m\alpha+n\beta\right]hP\left(0,t\right)+h\int_0^t p\left(2,t,u\right)\rho\left(u\right)du+o\left(h\right)\hat{}$$

$$p\left(i,t+h,0\right)h=h\int_0^t p\left(i+1,t,u\right)\rho\left(u\right)du+o\left(h\right)\qquad 2\le i\le m+n-1, \tag{2}$$

$$p\left(m+n,t+h,0\right)h=o\left(h\right).$$

Consider the system in steady-state conditions and introduce the following notations:

$$\lim_{t\to\infty}P\left(0,t\right)=P\left(0\right),\lim_{t\to\infty}p\left(i,t,u\right)=p\left(i,u\right),i=\overline{1,m+n}.$$

The latter values are expressed from system (1)-(2) as follows:

$$P\left(0\right)=\frac{C\left(1\right)}{m\alpha+n\beta}\overline{g}\left(m\alpha+\left(n-1\right)\beta\right);$$

$$p\left(1,u\right)=C\left(1\right)\left[1-G\left(u\right)\right]e^{-\left[m\alpha+\left[n-1\right]\beta\right]u} \tag{3}$$

$$p\left(i,u\right)=\left[1-G\left(u\right)\right]\sum_{j=0}^{i-1}\left(-1\right)^j C\left(i-j\right)\frac{\prod_{k=n-(i-1)}^{n-(i-j)}\left(m\alpha+k\beta\right)}{j!\beta^j}e^{-\left[m\alpha+\left(n-\left(i-j\right)\right)\beta\right]u},2\le i\le n,$$

$$p\left(i,u\right)=\left[1-G\left(u\right)\right]\left[\sum_{j=0}^{i-n}\left(-1\right)^j C\left(i-j\right)\frac{\left[m+n-\left(i-j\right)\right]!}{j!\quad\left[m+n-i\right]!}e^{-\left[m+n-\left(i-j\right)\right]\alpha u}+\right.$$

$$\left.+\sum_{j=i-n+1}^{i-1}\left(-1\right)^j C\left(i-j\right)\frac{m!\quad\alpha^{i-n}\quad\prod_{k=1}^{n-(i-j)}\left[m\alpha+k\beta\right]}{\left(m+n-i\right)!\ \left[n-\left(i-j\right)\right]!\quad\beta^{\left[n-\left(i-j\right)\right]}\quad\prod_{k=1}^{i-n}\left[k\alpha+\left(n-\left(i-j\right)\right)\beta\right]}e^{-\left[m\alpha+\left(n-\left(i-j\right)\right)\beta\right]u}\right],$$

$$n < i < m+n;$$

$$p\left(m+n,u\right)=\left[1-G\left(u\right)\right]\left[\sum_{j=0}^{m}\left(-1\right)^j C\left(m+n-j\right)e^{-j\alpha u}+\right.$$

$$+\sum_{j=m+1}^{m+n-1}(-1)^j C(m+n-j)\frac{m!\quad \alpha^m \quad \prod_{k=1}^{j-m}[m\alpha+k\beta]}{[j-m]!\quad \beta^{j-m}\quad \prod_{k=1}^{m}[k\alpha+(j-m)\beta]}e^{-[m\alpha+(j-m)\beta]u}\Bigg],$$

where $C(i)$ coefficients are unknown values, and $\overline{g}(s)$- SnLaplace transform of $g(t)$.

To find $P(t)$ and $C(i)$ values is possible to equalize expressions from system (3) to $p(i,t,u)$ from (2) when $u=0$. We obtain the following system of algebraic equations:

a) if $n>1$,

$$C(2)\,\overline{g}(m\alpha+(n-2)\beta)-C(1)\left[1+\frac{m\alpha+(n-2)\beta}{\beta}\overline{g}(m\alpha+(n-1)\beta)\right]=0$$

$$C(i+1)\,\overline{g}(m\alpha+(n-(i+1))\beta)+\sum_{j=0}^{i-1}\Bigg[\quad(-1)^{j+1}C(i-j)\frac{\prod_{k=n-(i-1)}^{n-(i-j)}(m\alpha+k\beta)}{j!\quad \beta^j}\times$$

$$\times\left(1+\frac{m\alpha+(n-i)\beta}{(j+1)\beta}\overline{g}(m\alpha+(n-(i-j))\beta)\right)\Bigg]=0\qquad 2\le i<n,$$

$$C(i+1)\,\overline{g}((m+n-(i+1))\alpha)+\sum_{j=0}^{i-n}(-1)^{j+1}C(i-j)\frac{[m+n-(i-j)]!}{j!\quad [m+n-i]!}\cdot\left(1+\frac{m+n-i}{j+1}\overline{g}((m+n-(i-j))\alpha)\right)+$$

$$+\sum_{j=i-n+1}^{i-1}\Bigg[(-1)^{j+1}C(i-j)\frac{m!\quad \alpha^{i-n}\quad \prod_{k=1}^{n-(i-j)}(m\alpha+k\beta)}{(m+n-i)!\ [n-(i-j)]!\quad \beta^{[n-(i-j)]}\quad \prod_{k=1}^{i-n}[k\alpha+(n-(i-j))\beta]}\times$$

$$\times\left(1+\frac{(m+n-i)\alpha}{(i-n+1)\alpha+(n-(i-j))\beta}\overline{g}(m\alpha+(n-(i-j))\beta)\right)\Bigg]=0\qquad n\le i<m+n-1,$$

$$\sum_{j=0}^{m}(-1)^j C(m+n-j)+\sum_{j=m+1}^{m+n-1}(-1)^j C(m+n-j)\frac{m!\quad \alpha^m\quad \prod_{k=1}^{j-m}[m\alpha+k\beta]}{[j-m]!\quad \beta^{j-m}\quad \prod_{k=1}^{m}[k\alpha+(j-m)\beta]}=0$$

b) if $n=1$;

$$C(2)\,\overline{g}((m-1)\alpha)-C(1)[1+(m-1)\,\overline{g}(m\alpha)]=0$$

$$C\left(i+1\right)\overline{g}\left(\left(m+n-\left(i+1\right)\right)\alpha\right)+\sum_{j=0}^{i-n}\left(-1\right)^{j+1}C\left(i-j\right)\frac{\left[m+n-\left(i-j\right)\right]!}{j!\quad\left[m+n-i\right]!}\cdot\left(1+\frac{m+n-i}{j+1}\overline{g}\left(\left(m+n-\left(i-j\right)\right)\alpha\right)\right)=0$$

$$2\leq i<m+n-1,$$

$$\sum_{j=0}^{m}\left(-1\right)^{j}C\left(m+n-j\right)=0$$

To the end of the system is added the normalizing equation:

$$P\left(0\right)+\sum_{i=1}^{m+n}\int_{0}^{\infty}p\left(i,u\right)du=1$$

.

From where we have:

$$C\left(m+n\right)=\frac{1-\frac{C(1)}{m\alpha+n\beta}\overline{g}\left(m\alpha+\left(n-1\right)\beta\right)}{E\xi}$$

.

Here $E\xi=\int_{0}^{\infty}\left[1-G\left(u\right)\right]du$ is the mean time of element renewal.

3 Economical Analysis of the System

In this section we first determine some system's some probabilistic characteristics.

$E_1(n)$ - the expected number of main elements;

$E_2(n)$- the expected number of spare elements;

$E_3(n)$ - the expected number of renewal facilities at steady-state condition.

It is easy to set the following equations

$$E_1\left(n\right)=m-\sum_{i=n+1}^{m+n}\int_{0}^{\infty}\left(i-n\right)p\left(i,u\right)du,$$

$$E_2\left(n\right)=nP\left(0\right)+\sum_{i=1}^{n}\int_{0}^{\infty}\left(n-i\right)p\left(i,u\right)du,$$

$$E_3\left(n\right)=1-P\left(0\right).$$

From expressions (3) we have obtained:

$$E_1(n) = mP(0) + C(m+n-1)\frac{1-\overline{g}(\alpha)}{\alpha} + \sum_{j=1}^{n-1}\left((-1)^j C(n-j)\frac{\prod_{k=0}^{j-1}(m\alpha+k\beta)}{\beta^{j-1}(j-1)!\,((m-1)\alpha+j\beta)} \times \right.$$

$$\left. \times \frac{1-\overline{g}(m\alpha+j\beta)}{m\alpha+j\beta}\right)$$

$$E_2(n) = nP(0) + \sum_{i=1}^{n}\left[(n-i)\sum_{j=0}^{i-1}(-1)^j C(i-j)\frac{\prod_{k=n-(i-1)}^{n-(i-j)}(m\alpha+k\beta)}{j!\beta^j}\cdot\frac{1-\overline{g}(m\alpha+(n-(i-j))\beta)}{m\alpha+(n-(i-j))\beta}\right]$$

$$E_3(n) = 1 - P(0).$$

Before we introduce expected profit functions per unit time for the system, we introduce the following symbols:

p =income per unit time from one serviceable main element;

c_1 = cost per unit time for one serviceable main element

c_2 =cost per unit time for one serviceable standby element;

c_3 = cost per unit time for operating renewal facility;

c_4 = cost per unit time for non-operating renewal facility;

On the basis of the introduced symbols and calculations, present the function describing the effectiveness of the system

$$F(n) = (p-c_1)E_1(n) - c_2E_2(n) - c_3E_3(n) - c_4(1-E_3(n)) \tag{4}$$

When m is fixed function F depends on argument n. Therefore, the problem is stated: for the given m it is necessary to find n^*, that gives the function F the maximum value.

4 Numerical Illustration

As a numerical example we consider some distribution functions of renewal times with the same mean time

I. $G_\mu(t) = \begin{cases} 0, & t \le 0, \\ 1-e^{-\mu t}, & t > 0. \end{cases};$

II. $G_\mu(t) = \begin{cases} 0, & t \le 0, \\ 1-e^{-2\mu t}(1+2\mu t), & t > 0. \end{cases}$

III. $G_\mu(t) = \begin{cases} 0, & t \le \frac{1}{\mu}, \\ 1, & t > \frac{1}{\mu}. \end{cases};$

(5)

IV. $G_\mu(t) = \begin{cases} 0, & t \leq \tau, \\ 1 - e^{-\frac{\mu}{1-\mu\tau}(t-\tau)}, & t > \tau. \end{cases}, \tau = \frac{\mu}{2}.$

We set the fixed cost $p = 450$, $c_1 = 90$, $c_2 = 70$, $c_1 = c_2 = 100$.

Case 1. We fix $m = 10, \quad 1 \leq n \leq 10, \quad \mu = 1, \quad \alpha = 0,02, \quad \beta = 0,01.$

In Table 1 is shown how profit function is dependent on the number of standby elements. We can easily see that $n* = 1$ is the optimal value for all distribution functions:

n	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
I	3426,9	3374,5	3309,2	3241,1	3172,5	3103,7	3035	2966,2	2887,5	2828,8
II	3431,2	3375,6	3308,8	3240,1	3171,3	3102,3	3033,4	2964,6	2895,7	2826,9
III	3435,5	3376,2	3308	3239	3170	3100,9	3031,9	2962,9	2893,9	2824,9
IV	3433,3	3375,9	3308,4	3239,6	3170,6	3101,6	3032,7	2963,7	2894,8	2825,9

Table 1:

Case 2. Table 2 show the effect of initial parameters on the profit function and optimal n^* value, when $\alpha = \overline{0,01;\ 0,1}$ and $\beta = \frac{\alpha}{2}$:

$\boldsymbol{\alpha}$		**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**	**0,1**
I	n*	1	1	1	1	2	2	3	4	5	7
	F(n)*	34433,2	3426,9	3408,9	3377,2	3344,2	3301,9	3252,1	3186,2	3100,7	2990,8
II	n*	1	1	1	1	2	2	3	4	5	6
	F(n)*	3434,2	34431,2	3418,9	3395,3	3366	3337,3	3294,7	3238,5	3163,2	3057,5
III	n*	1	1	1	1	2	2	2	3	4	6
	F(n)*	3435,3	3435,5	3429,4	3414,8	3389,5	3373,4	3342,8	3302,2	3238,1	3137,8
IV	n*	1	1	1	1	2	2	3	3	4	6
	F(n)*	3434,7	344433,3	3424,1	3405	3376	3355	3315	3268,7	3198,3	3095,8

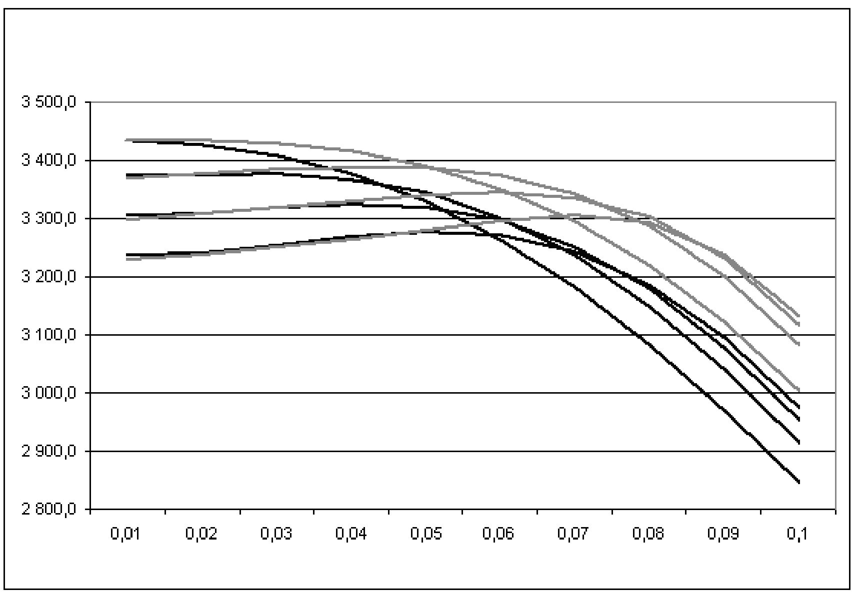
Table 2:

Case 3. We fix $n* = 1$, and select $m = \overline{1, 10}$. In Table 3 is shown, that when number of main elements is small, value of profit functions for different distribution functions are not too must different. When number of main elements increase there are differences between the profit functions.

Case 4. We select $n = 1, 2, 3, 4$. We can express on the graph below how the profit function changed when $\alpha = \overline{0,01;\ 0,1}$, and assumed $\beta = \frac{\alpha}{2}$. The graph is obtain from distribution functions I and III.

m	1	2	3	4	5	6	7	8	9	10
I	191,8	552,7	913,3	1273,6	1633,6	1999,9	2352,1	2710,8	3069,1	3426,9
II	191,9	552,9	913,7	1274,2	1634,5	1994,5	2354,1	2713,5	3072,5	3431,2
III	192,0	553,1	914,1	1274,9	1635,5	1995,9	2356,2	2716,2	3076,0	3435,5
IV	191,9	553	913,9	1274,5	1365	1995,2	2355,2	2714,8	3074,2	3433,3

Table 3:



5 Conclusion

In this paper, we constructed closed queuing system with warm standby elements and renewal service to study the profit analysis of the investigated system above. We performed an economical analysis for the optimal value of $n$with various parameters values and cost elements. The results of this paper make it possible to analyze many queueing models, both simple and complicated, where sequential repair is required.

References

[1] Sivarzlan B.D.,Wang K.H. *Economic analysis of the M/M/R machine repair problem.* Microelectron. Reliab.1989, vol. 29, 1, pp.25-35

[2] Kuo-Hsiung Wang, Yu-Ju Lai, Jyh-Bin Ke. *Reliability and Sensitivity Analysis of a System with Warm Standbys and a Repairable Service Station.* Inter. Journal of Oper. Res. 2004. Vol. 1, No. 1, 61-70

[3 Khurodze R., Kakubava I., Svanidze N., Kubetsia I. *Closed Queuing System with two Sources of Requests.* Bull. Georg. Acad. Sci., 2004, 170, 1.

Jana Kalčevová

Department of Econometrics
University of Economics Prague
W. Churchill sq. 4, 130 67 Prague 3

e-mail: kalcevov@vse.cz

Combinatorial Auctions Test Suite

Abstract

The classical auction problem is to allocate goods when the amount that bidders are willing to pay is unknown or changeable over time with respect to maximal revenue of seller. The combinatorial auction is the problem in which bidders place bids for bundles of goods. This optimization problem is not easy to solve and so researchers are focused on algorithms for finding an optimal winning bid or at least approximately optimal one. For quality testing of combinatorial auctions algorithms and comparison of them the set of appropriate data is necessary. Therefore there was developed CATS (Combinatorial Auctions Test Suite) for generating of an appropriate data set for combinatorial auction problem, created by Kevin Leyton-Brown, Mark Pearson and Yoav Shoham at the Stanford University, Department of Computer Science. This paper analyses CATS, a suite for generating realistic bids. CATS need only three type of information: the type of distribution, the number of bids and the number of goods. The problem of bids generating is equivalent to graph generating, where nodes represent goods, edges represent relationships between goods and subset of nodes represents the bundle of goods. CATS then generates an appropriate graph and after it a set of asked number of realistic bids.

Keywords: *combinatorial auctions, bids, graph, test suite*

1 CATS

The combinatorial auction is the multi-object auction in which bids name bundles of goods. The researchers are focused on algorithms for finding an optimal winning bid or at least approximately optimal one. In this type of auction a seller has a set of price offers for various bundles of goods and his goal is to allocate the goods in the way that maximizes his revenue [8].

The Combinatorial Auction Test Suite (CATS) was created for generating of an appropriate data set for combinatorial auction problem. The data set is given by the number of bids, number of goods and prices of bids. The price is defined as the amount a bidder is made to pay for a bundle.

The CATS is also a set of distributions that attempt to model realistic bidding behavior. All distributions are parameterized by number of goods and number of bids. The CATS has several positive properties. It assumes [7]:

- certain goods are more likely to appear together than others,
- the number of goods appearing in the bundle is often related to which goods appear in the bundle,
- valuations are related to which goods appear in the bundle; where appropriate, valuations can be configured to be subadditive, additive or superadditive in the number of goods requested,
- it is build for both Windows and Linux operation system.

2 Assumptions

The CATS is built under several assumptions. The first one says that all bundles of the same size are equally likely to be requested. It is clear that in practice this assumption is often violated because bundles with complementarities are more likely than bundles without them (e.g. a bundle with lemons and oranges is more likely than a bundle with lemons and computers).

The second assumption relates to a number of goods. The number of goods in a bundle is independent on the fact which goods appear in the bundle. In practice also this assumption can be violated (e.g. a bundle with computers can contain also monitors, keyboards, printers, scanners etc. whereas a bundle with woods contains only short bid).

The third assumption is related to price. There are two ways how to generate prices in [3]. The first one chooses the prices randomly from [0,1]. This is unreasonable because the price is unrelated to the size of bundle. The problem consists in a fact that a bid for a large bundle and the bid for a small subset of the same bundle have the same expected price. Therefore the second way was suggested the prices are drown randomly from $[0,g]$ where g is a number of goods in the bundle. In this case both mean and range are parameterized by the same variable (number of goods). The third way for prices generation is presented in [4]. Here prices are drown from $[0.5g,\ 1.5g]$ that avoids the problems mentioned above. Also in this case prices are unrelated to which goods are in a bundle and it is unrealistic. Much more realistic approach to prices generation is described in [5]. The best way of solving this problem is to make bundle prices superadditive in a number of goods. In [6] is introduced a quadratic function of the prices of bids for subset.

3 Combinatorial auction graph

The problem of bids generating is equivalent to graph generating, where nodes represent goods, edges represent relationships between goods and subset of nodes represents the bundle of goods. CATS then generates an appropriate graph and after it a set of asked number of realistic bids.

The main idea is to create graph with a number of nodes grater than number of goods. Each node then represents one good. The second step is to generate some edges those represent relationships between goods. Two connected nodes represent two related goods. It is more probable that related goods will be in the same bundle than unrelated goods (those are not connected by edge). In the third step we choose subgraph that represents bundle of goods.

So, in the first step we generate number of nodes given by number of goods and place them on a plane (it means to generate 2 values – for 2 axes vertical and horizontal). Then edges to this graph are included – starting by connecting each node to a fixed number of its nearest neighbours. After that the shortest path between two random nodes is searched. Then the new edge with penalization (length times penalty) between given nodes is added in the graph and the new shortest path is calculated. If the new edge lies on the shortest path it is added in the graph (without penalty) and replaces the existing edges in graph.

This graph is used to generate bids. The target is generated bids over a desired number of goods. Note that number of nodes is grater than number of goods that we want to generate. So small disconnected components of the graph are removed in order to the total number of goods equal to required one. Now the edges those are used only in singleton bids or isolated groups of bids are deleted.

The detailed description of this process is presented in [7].

4 Parameters of CATS

The user need not know how the CATS works but he/she needs to use it. The CATS is available at `http://robotics.stanford.edu/CATS/` together with information about how to use CATS and up-to-date releases. The user can download ZIP file with *CATSguide.pdf* [9] file and *cats.exe* file and extract it. There is no need of installation, user only runs *cats.exe* file.

4.1 Required parameters of CATS

CATS is called by command line. This command line must contain three type of information: the type of distribution, the number of bids and the number of goods. The distributions are defined in [1]. For each run the number of bids and the number of goods must be specified (*–bids* and *–goods*). These numbers can be also choose from uniform distribution over a specified range (*–random_bids* and *–random_goods*). In this case it is necessary to set minimum and maximum values of the range.

The second possibility is to use *–default_hard* for using of default values of all parameters. For user-specified hybrid distributions (*–dist_dist*) it is not possible to use *–default_hard* or the hybrid models must be restricted to 256 goods and 1000 bids as in [2].

Other optional parameters are presented at `http://robotics.stanford.edu/CATS` and in [9].

4.2 Parameters of CATS distribution

Every CATS distribution has from one to seven parameters. The user can set their values individually or can choose the value from their ranges (function *–random_parameters*) with given distribution function (uniform distribution and polynomially defined joint distribution). Also it is possible to use default values for parameters.

4.3 Optional flags

The behaviour of CATS is possible to affect by optional flags. The CATS data file can be written also in CPLEX *.lp file (*–cplex*) and the user can set filename prefix (*–filename*) for all generated data. For data generation the user can use default values (*–default_hard*) or can choose parameters uniformly from their ranges (*–random_parameters*). If the user wants integer prices in bids instead of real values he/she can use *–int_prices*. The dominated bids can be removed by *–no_dom_check*. The flag *–no_output* does not write unnecessary output to the screen, *–help* prints a usage screen and alike *–model_file_help* prints help screen for model file features.

Other information about CATS option can be found in [9].

5 Conclusion

The CATS is a test suite for combinatorial auction optimization algorithms. It is based on the fact that the problem of bids generating is equivalent to graph generating. It is freeware and it can be run in Microsoft Windows, UNIX and Linux. The user must know only three information: the type of distribution, the number of bids and the number of goods (it is possible to run CATS from the command line by typing *cats –default_hard*). It makes CATS user-friendly. For the more interested users hybrid models and large user option are implemented. For users is available current version 2.0 at the websites.

Acknowledgement

This work was supported by the Grant Agency of the Czech Republic – GA CR 402/07/0166.

Bibliography

[1] LEYTON-BROWN, K., PEARSON, M., SHOHAM, Y. *Towards a universal test suite for combinatorial auction algorithms.* ACM EC, 2000.
`http://robotics.stanford.edu/~kevin1b/CATS.pdf`

[2] LEYTON-BROWN, K., NUDELMAN, E., ANDREW, G., MCFADDEN, J., SHOHAM, Y. *A portfolio approach to algorithm selection.* IJCAI, 2003.
`http://robotics.stanford.edu/~kevin1b/portfolio-IJCAI.pdf`

[3] SANDHOLM, T. *An algorithm for optima Winter determination in combinatorial auctions.* IJCAI-99, 1999.

[4] FUJISHIMA, Y., LEYTON-BROWN, K., SHOHAM, Y. *Taming the computational complexity of combinatorial auctions: Optimal and approximate approaches.* IJCAI, 1999.

[5] ANDERSON, A., TENHUNEN, M., YGGE, F. *Integer programming for combinatorial auction winner determination.* ICMAS, 2000, pp. 39-46.

[6] DE VRIES, S., VOHRA, R. *Combinatorial auctions: A survey.* INFORMS Journal on Computing, 15(3). 2003.

[7] LEYTON-BROWN, K., SHOHAM, Y., STEINBERG, R. Combinatorial Auctions. Chapter 18, MIT Press, 2006.

[8] DE VRIES, S., VOHRA, R. *Combinatorial auctions: A survey.* 2000

[9] ANDREW, G. *CATS 2.0 User Guide.* `http://robotics.stanford.edu/CATS/`

Vlasta Kaňková

Department of Econometrics
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4, 182 08 Praha

e-mail: kankova@utia.cas.cz

Multiobjective Stochastic Programming via Multistage Problems

Abstract

Economic activities developing over time are very often influenced simultaneously by a random factor (modelled mostly by a stochastic process) and a "decision" parameter that has to be chosen according to economic possibilities. Moreover, it is necessary often to evaluate the economic activities simultaneously by a few "utility" functions. Evidently, the mentioned economic situations can lead to mathematical models corresponding to multistage multiobjective stochastic programming problems. Usually, the multiobjective (one–stage) problems and multistage (one–objective) problems have been investigated separately. The aim of this contribution will be to try to analyze a relationship between these two approaches.

Keywords: *Multistage stochastic programming problems, multiobjective stochastic programming, efficient points, weight approach*

1 Introduction

Multistage stochastic programming problems belong to optimization problems depending on a probability measure. Usually, the operator of mathematical expectation appears in the objective function and constraints set can depend on the probability measure also. The multistage stochastic programming problems correspond to applications that can be considered with respect to some finite "discrete" (say $(0, M)$; $M \geq 1$) time interval and simultaneously there exists a possibility to decompose them with respect to the individual time points. A decision, at every individual time point say k, can depend only on the random elements realizations and the decisions to the time point $k-1$ (we say that it must be nonanticipative). To define the multistage stochastic programming problems we employ an approach in which the multistage stochastic programming problem is introduced as a finite system of parametric (one–stage) optimization problems with an inner type of dependence (for more details see e.g. [1]). The multistage stochastic programming problem can be then introduced in the following form.

Find
$$\varphi_{\mathcal{F}}(M) = \inf \{\mathsf{E}_{F^{\xi^0}} g^0_{\mathcal{F}}(x^0, \xi^0) |\, x^0 \in \mathcal{K}^0\}, \tag{1}$$

where the function $g^0_{\mathcal{F}}(x^0,\, z^0)$ is defined recursively

$$\begin{array}{lcl} g^k_{\mathcal{F}}(\bar{x}^k,\, \bar{z}^k) & = & \inf\{\mathsf{E}_{F^{\xi^{k+1}|\bar{\xi}^k=\bar{z}^k}}\, g^{k+1}_{\mathcal{F}}(\bar{x}^{k+1},\, \bar{\xi}^{k+1})\,|x^{k+1} \in \mathcal{K}^{k+1}_{\mathcal{F}}(\bar{x}^k,\, \bar{z}^k)\}, \\ k = 0,\, 1,\, \ldots,\, M-1, & & \\ g^M_{\mathcal{F}}(\bar{x}^M,\, \bar{z}^M) & := & g^M_0(\bar{x}^M,\, \bar{z}^M), \quad \mathcal{K}_0 := X^0. \end{array} \tag{2}$$

$\xi^j := \xi^j(\omega),\ j = 0,\, 1,\, \ldots,\, M$ denotes an s–dimensional random vector defined on a probability space $(\Omega,\, \mathcal{S},\, P)$; $F^{\xi^j}(z^j),\ z^j \in R^s,\ j = 0,\, 1\, \ldots,\, M$ the distribution function of the ξ^j and $F^{\xi^k|\bar{\xi}^{k-1}}(z^k|\bar{z}^{k-1}),\ z^k \in R^s,\ \bar{z}^{k-1} \in R^{(k-1)s},\ k = 1,\, \ldots,\ M$ the conditional distribution function (ξ^k conditioned by $\bar{\xi}^{k-1}$); $P_{F^{\xi^j}},\, P_{F^{\xi^{k+1}|\bar{\xi}^k}},\ j = 0,\, 1,\, \ldots,\, M,\, k = 0,\, 1,\, \ldots,\, M-1$ the corresponding probability measures; $Z^j := Z_{F^{\xi^j}} \subset R^s,\ j = 0,\, 1,\, \ldots,\, M$ the support of the probability measure $P_{F^{\xi^j}}$. Furthermore, $g^M_0(\bar{x}^M,\, \bar{z}^M)$ denotes a continuous function defined on $R^{n(M+1)} \times R^{s(M+1)}$; $X^k \subset R^n,\ k = 0,\, 1,\, \ldots,\, M$ is a nonempty compact set; the symbol $\mathcal{K}^{k+1}_{\mathcal{F}}(\bar{x}^k,\, \bar{z}^k) := \mathcal{K}^{k+1}_{F^{\xi^{k+1}|\bar{\xi}^k}}(\bar{x}^k,\, \bar{z}^k)\, (\mathcal{K}^{k+1}_{F^{\xi^{k+1}|\bar{\xi}^k}}(\bar{x}^k,\, \subset X^k),\ \ k = 0,\, 1,\, \ldots,\, M-1$ denotes a multifunction mapping $R^{n(k+1)} \times R^{s(k+1)}$ into the space of subsets of R^n. $\bar{\xi}^k(:= \bar{\xi}^k(\omega)) = [\xi^0,\, \ldots,\, \xi^k];\ \bar{z}^k = [z^0,\, \ldots,\, z^k],\ z^j \in R^s;\ \bar{x}^k = [x^0,\, \ldots,\, x^k],\ x^j \in R^n;\ \ \bar{X}^k = X^0 \times X^1 \ldots \times X^k;\ \bar{Z}^k := \bar{Z}^k_{\mathcal{F}} = Z_{F^{\xi^0}} \times Z_{F^{\xi^1}} \ldots \times Z_{F^{\xi^k}},\ j = 0,\, 1,\, \ldots, k,\ k = 0,\, 1,\, \ldots,\, M.$ Symbols $\mathsf{E}_{F^{\xi^0}},\, \mathsf{E}_{F^{\xi^{k+1}|\bar{\xi}^k=\bar{z}^k}},\ k = 0,\, 1,\, \ldots,\, M-1$ denote the operators of mathematical expectation corresponding to $F^{\xi^0},\ F^{\xi^{k+1}|\bar{\xi}^k=\bar{z}^k},\ k = 0,\, \ldots,\, M-1$.

The problem (1) is a "classical" one–stage, one–objective stochastic problem, the problems (2) are (generally) parametric one–stage, one–objective stochastic optimization problems. Let us assume a special case when the function $g^M_0(\bar{x}^M,\, \bar{z}^M)$ fulfils the following assumption.

i.1 there exist continuous functions $\bar{g}^j(x^j,\, z^j),\ j = 0,\, 1,\, \ldots,\, M$ defined on $R^n \times R^s$ such that

$$g^M_0(\bar{x}^M,\, \bar{z}^M) = \sum_{j=0}^{M} \bar{g}^j(x^j,\, z^j). \tag{3}$$

Evidently, under i.1, the function $\bar{g}^j(x^j,\, z^j)$ corresponds to an evaluation of the economic activity at the time point $j \in \{0,\, 1,\, \ldots,\, M\}$. However, it happens rather often that it is reasonable to evaluate this economic activity simultaneously by several "utility" functions, say $\bar{g}^j_i(x,\ ,z),\ i = 1,\, \ldots,\, l$. Including this reality we can see that the "underlying" common objective function $g^M_0(\bar{x}^M,\, \bar{z}^M)$ (in (1) and (2)) has to be replaced by the following multiobjective criterion function.

$$g^M_{0,\,i}(\bar{x}^M,\, \bar{z}^M) = \sum_{j=0}^{M} \bar{g}^j_i(x^j,\, z^j), \quad i = 1,\, \ldots,\, l. \tag{4}$$

Consequently, assuming the same inner time dependence as it was assumed in the problem (1) and (2), we obtain formally the following multistage, multiobjective problem.

Find

$$\varphi_{\mathcal{F}}(M,\, i) = \inf\, \mathsf{E}_{F^{\xi^0}} g^{0,\, i}_{\mathcal{F}}(x^0,\, \xi^0),\ \ i = 1,\, \ldots,\, l \quad \text{subject to} \quad x^0 \in \mathcal{K}^0, \tag{5}$$

where the function $g^{0,\, i}_{\mathcal{F}}(x^0,\, z^0)$, $i = 1,\, \ldots,\, l$ are defined recursively

$$\begin{aligned} g^{k,\, i}_{\mathcal{F}}(\bar{x}^k,\, \bar{z}^k) \quad &= \quad \inf\, \mathsf{E}_{F^{\xi^{k+1}|\bar{\xi}^k = \bar{z}^k}}\ g^{k+1,\, i}_{\mathcal{F}}(\bar{x}^{k+1},\, \bar{\xi}^{k+1}),\, i = 1,\, \ldots,\, l \\ &\qquad \text{subject to} \quad x^{k+1} \in \mathcal{K}^{k+1}_{\mathcal{F}}(\bar{x}^k,\, \bar{z}^k), \quad k = 0,\, 1,\, \ldots,\, M-1, \\ g^{M,\, i}_{\mathcal{F}}(\bar{x}^M,\, \bar{z}^M) \quad &:= \quad g^M_{0,\, i}(\bar{x}^M,\, \bar{z}^M), \quad i = 1,\, \ldots,\, l, \quad \mathcal{K}_0 := X^0. \end{aligned} \tag{6}$$

The problem (5) is formally a problem of one–stage multiobjective optimization theory. The problems (6) are one–stage multiobjective parametric optimization problems. It is known that there doesn't exists (mostly) an optimal solution simultaneously with respect to all criteria. Consequently, the optimal solution has to be mostly replaced by a set of efficient points. Consequently, $g^{k,\, i}_{\mathcal{F}}$, $i = 1,\, \ldots,\, l$, $k = 0,\, \ldots,\, M-1$ have to be replaced by multifunctions $\mathcal{G}^{k,\, i}_{\mathcal{F}}$, $i = 1,\, \ldots,\, l$, $k = 0,\, \ldots,\, M-1$ corresponding to the function values in efficient points (for definition of the efficient points see [2] or the following section).

2 Some Definitions and Auxiliary Assertion

2.1 One–Stage Deterministic Multiobjective Problems

A multiobjective deterministic optimization problem can be introduced as the problem.

Find

$$\min f_i(x),\ \ i = 1,\, \ldots,\, l \quad \text{subject to } x \in \mathcal{K}. \tag{7}$$

f_i, $i = 1,\, \ldots,\, l$ are functions defined on R^n, $\mathcal{K} \subset R^n$ is a nonempty set.

Definition 1. [4] The vector x^* is an efficient solution of the problem (7) if and only if $x^* \in \mathcal{K}$ and if there exists no $x \in \mathcal{K}$ such that $f_i(x) \le f_i(x^*)$ for $i = 1,\, \ldots,\, l$ and such that for at least one i_0 one has $f_{i_0}(x) < f_{i_0}(x^*)$. We denote the set of efficient points of the problem (7) by $\mathcal{K}_E$.

Definition 2. [4] The vector x^* is a properly efficient solution of the multiobjective optimization problem (7) 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

$$\frac{f_i(x^*) - f_i(x)}{f_j(x) - f_j(x^*)} \le M. \tag{8}$$

We denote the set of properly efficient points of problem (7) by $\mathcal{K}_{PE}$.

Definition 3. [2] The vector x^* is called weakly efficient solution of the problem (7) if and only if $x^* \in \mathcal{K}$ and if there exists no $x \in \mathcal{K}$ such that $f_i(x) < f_i(x^*)$ for every $i = 1,\, \ldots,\, l$. We denote the set of weakly efficient points of the problem (7) by $\mathcal{K}_{wE}$.

Furthermore, let us define the following parametric optimization problem.

Find
$$\begin{aligned}&\min_{x\in\mathcal{K}} f^{\lambda}(x), \quad \lambda\in\Lambda, \quad \text{where} \quad f^{\lambda}(x)=\sum_{i=1}^{l}\lambda_i f_i(x) \quad \text{and}\\ &\Lambda=\{\lambda\in R^l:\ \lambda=(\lambda_1,\,\dots,\,\lambda_l),\ \lambda_i\in\langle 0,\,1\rangle,\ i=1,\,\dots,\,l,\ \sum_{i=1}^{l}\lambda_i=1\}.\end{aligned} \tag{9}$$

Proposition 1. [2] If

1. $\hat{x}\in\mathcal{K}$ is a solution of (9) for $\lambda\in\Lambda$, then $\hat{x}\in\mathcal{K}_{wE}$,
2. $\mathcal{K}$ is a convex set, f_k, $k=1,\,\dots,\,l$ are convex functions, then

$$\hat{x}\in\mathcal{K}_{wE} \quad\Longleftrightarrow\quad \text{there exists} \quad \lambda\in\Lambda \quad \text{such that} \quad \hat{x} \quad \text{is optimal in} \quad (9).$$

Proposition 2. [4] If

1. $\hat{x}$ is optimal in the problem (9) for some fixed $\lambda=(\lambda_1,\,\dots,\,\lambda_l)\in\Lambda$ with $\lambda_i>0$, $i=1,\,\dots,\,l$, then $\hat{x}\in\mathcal{K}_{PE}$,
2. $\mathcal{K}$ is a convex set and f_i, $i=1,\,\dots,\,l$ are convex functions on $\mathcal{K}$, then

$$\begin{aligned}\hat{x}\in\mathcal{K}_{PE} \quad\Longleftrightarrow\quad &\text{there exists} \quad \lambda\in\Lambda \quad \text{with} \quad \lambda_i>0,\, i=1,\,\dots,\,l \quad \text{such that}\\ &\hat{x} \quad \text{is optimal in (9).}\end{aligned}$$

If we denote by the symbols $f(\mathcal{K}_E)$, $f(\mathcal{K}_{PE})\subset R^l$ the image of $\mathcal{K}_E$, $\mathcal{K}_{PE}\subset R^n$ obtained by the vector function $f=(f_1,\,\dots,\,f_l)$, then the following implication has been recalled in [4].

$$\begin{aligned}&\mathcal{K} \quad \text{closed and convex,} \quad f_i,\, i=1,\,\dots,\,l \quad \text{continuous and convex on} \quad \mathcal{K}\\ &\Longrightarrow \quad f(\mathcal{K}_{PE})\subset f(\mathcal{K}_E)\subset \bar{f}(\mathcal{K}_{PE}), \quad \text{where} \quad \bar{f}(\mathcal{K}_{PE}) \quad \text{denotes a closure of} \quad f(\mathcal{K}_{PE}).\end{aligned} \tag{10}$$

Lemma 1. [10] Let $\mathcal{K}\subset R^n$ be a nonempty set, f_i, $i=1,\dots,l$ be functions defined on R^n. Let, moreover, the function f^{λ} be defined by the relation (9). If $f_i, i=1,\dots,l$ are bounded functions on $\mathcal{K}$, ($|f_i(x)|\le\bar{M}$, $x\in\mathcal{K}$, $i=1,\,\dots,\,l$, $\bar{M}>0$), then for every $x\in\mathcal{K}$, f^{λ} is a Lipschitz function on Λ with a Lipschitz constant not greater then $l\bar{M}$.

2.2 Deterministic Parametric Optimization

Definition 4. [5] Let $h(x)$ be a real–valued function defined on a convex set $\mathcal{K}\subset R^n$. $h(x)$ is a strongly convex function with a parameter $\rho>0$ if

$$h(\lambda x^1+(1-\lambda)x^2)\le\lambda h(x^1)+(1-\lambda)h(x^2)-\lambda(1-\lambda)\rho\|x^1-x^2\|^2 \quad \text{for every} \quad x^1,\,x^2\in\mathcal{K},\,\lambda\in\langle 0,\,1\rangle.$$

Lemma 2. [6] Let $\mathcal{K} \subset R^n$ be a nonempty, compact, convex set. Let, moreover, $h(x)$ 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 $x^0 = \arg\min_{x \in \mathcal{K}} h(x)$, then

$$\|x - x^0\|^2 \leq \frac{2}{\rho}|h(x) - h(x^0)| \quad \text{for every } x \in \mathcal{K}.$$

Lemma 3. [10] Let $\mathcal{K} \subset R^n$ be a nonempty convex set, $\varepsilon \in (0, 1)$. Let, moreover, f_i, $i = 1, \ldots, l$ be convex functions on $\mathcal{K}$. If

1. f_1 is a strongly convex (with a parameter $\rho > 0$) function on $\mathcal{K}$, then f^λ defined by (9) is for $\lambda = (\lambda_1, \ldots, \lambda_l) \in \Lambda$ with $\lambda_1 \in (\varepsilon, 1)$ a strongly convex function on $\mathcal{K}$,
2. f_i, $i = 1, \ldots, l$ are strongly convex function on $\mathcal{K}$ with a parameter ρ, then f^λ defined by (9) is a strongly convex function on $\mathcal{K}$.

Lemma 4. [7] Let $\mathcal{K} \in R^n$, $Y \in R^m$, $n, m \geq 1$ be nonempty convex sets. Let, furthermore, $\bar{\mathcal{K}}(y)$ be a multifunction mapping Y into the space of nonempty closed subsets of $\mathcal{K}$, $h(x, y)$ function defined on $X \times Y$ such that

$$\bar{\varphi}(y) \; = \; \inf\{h(x, y) | x \in \bar{\mathcal{K}}(y)\} > -\infty \quad \text{for every} \quad y \in Y.$$

If

1. $h(x, y)$ is a convex function on $X \times Y$ and simultaneously
$$\lambda(\bar{\mathcal{K}}(y(1)) + (1 - \lambda)(\bar{\mathcal{K}}(y(2)) \; \subset \; \bar{\mathcal{K}}(\lambda y(1) + (1 - \lambda)y(2)) \quad \text{for every} \quad y(1), \, y(2) \in Y,$$
then $\bar{\varphi}(y)$ is a convex function on Y.
2. $h(x, y)$ is a Lipschitz function on $X \times Y$ with the Lipschitz constant L and simultaneously
$$\Delta[\bar{\mathcal{K}}(y(1)), \, \bar{\mathcal{K}}(y(2))] \leq \bar{C}\|y(1) - y(2)\| \quad \text{for every} \quad y(1), \, y(2) \in Y \text{and a} \, \bar{C} \geq 0$$
then $\bar{\varphi}(y)$ is a Lipschitz function on Y with the Lipschitz constant not greater then $L(\bar{C} + 1)$.
(the symbol $\Delta[\cdot, \cdot]$ denotes the Hausdorff distance, for the definition see e.g. [12].)

3 Problem Analysis

In this section, we return to the problem introduced by (6) and (7). In particular, the aim of this section will be to try to characterize points obtained by "weight" approach and consequently to approximate efficient points sets corresponding to the individual problems (6), (7). To this end we assume.

A.1 $\bar{g}_i^j(x^j, z^j)$, $i = 1, \ldots, l, \quad j = 0, \ldots, M$ are for every $\bar{z}^M \in \bar{Z}^M$ convex functions of $\bar{x}^M \in \bar{X}^M$, A.2 $\Delta[\mathcal{K}_{\mathcal{F}}^{k+1}(\bar{x}^k(1),\, \bar{z}^k),\, \bar{x}^k(2),\, \bar{z}^k)] \le C\|\bar{x}^k(1) - \bar{x}^k(2)\|$ for every $\bar{x}^k(1),\, \bar{x}^k(2) \in \bar{X}^k,\, \bar{z}^k \in \bar{Z}^k$ and some $C > 0$,

A.3 $\lambda\mathcal{K}_{\mathcal{F}}^{k+1}(\bar{x}^k(1),\, \bar{z}^k) + (1-\lambda)\mathcal{K}_{\mathcal{F}}^{k+1}(\bar{x}^k(2),\, \bar{z}^k) \subset \mathcal{K}_{\mathcal{F}}^{k+1}(\lambda\bar{x}^k(1) + (1-\lambda)\bar{x}^k(2),\, \bar{z}^k)$ for every $\bar{x}^k(1),\, \bar{x}^k(2) \in \bar{X}^k,\, \bar{z}^k \in \bar{Z}^k$,

A.4 $\mathcal{G}_{\mathcal{F}}^{k,\, i}(\bar{x}^k,\, \bar{z}^k) \in (-\infty,\, +\infty)$ for every $\bar{x}^k \in \bar{X}^k,\, \bar{z}^k \in \bar{Z}^k$.

The cases under which the assumption A.2, A.3 and A.4 are fulfilled can be found e.g. in [8] or [11].

Evidently, according to the assertions of Proposition 1, Proposition 2 and the relation (10) it is reasonable to set a weight approach to the multiobjective function (5). We obtain by this one–objective multistage parametric problem of the type (1), (2) in which

$$g_0^M(\bar{x}^M,\, \bar{z}^M) \quad \text{is replaced by} \quad g_{\mathcal{F}}^{M,\, \lambda}(\bar{x}^M,\, \bar{z}^M) = \sum_{j=0}^{M}\sum_{i=1}^{l} \lambda_i \bar{g}_i^j(x^j,\, z^j), \quad \lambda = (\lambda_1,\, \ldots,\, \lambda_l) \in \Lambda.$$

Obviously, for given $\lambda \in \Lambda$ the values of the component λ_i for every $i \in \{1,\, \ldots,\, l\}$ corresponds to the relevance of the component $\sum\limits_{j=1}^{M} \bar{g}_i^j(x^j,\, z^j)$ in the multiobjective criterion $\sum\limits_{j=1}^{M} \bar{g}_1^j(x^j,\, z^j),\, \ldots,\, \sum\limits_{j=1}^{M} \bar{g}_i^j(x^j,\, z^j)$ (for more details see e.g. [3] or [4]). Consequently, we obtain (by this approach) the following multistage stochastic parametric programming problem.

Find
$$\varphi_{\mathcal{F}}^{\lambda}(M) = \inf\{\mathsf{E}_{F^{\xi^0}} g_{\mathcal{F}}^{0,\, \lambda}(x^0,\, \xi^0)|\; x^0 \in \mathcal{K}^0\}, \tag{11}$$
where the function $g_{\mathcal{F}}^{0,\, \lambda}(x^0,\, z^0)$ is defined recursively for $k = 0,\, 1,\, \ldots,\, M-1$ by

$$g_{\mathcal{F}}^{k,\, \lambda}(\bar{x}^k,\, \bar{z}^k) \quad = \quad \inf\{\mathsf{E}_{F^{\xi^{k+1}|\bar{\xi}^k = \bar{z}^k}}\, [\sum_{i=1}^{l} \lambda_i \bar{g}_i^k(x^k,\, z^k) + g_{\mathcal{F}}^{k+1,\, \lambda}(\bar{x}^{k+1},\, \bar{\xi}^{k+1})]\, |x^{k+1} \in \mathcal{K}_{\mathcal{F}}^{k+1}(\bar{x}^k,\, \bar{z}^k)\}. \tag{12}$$

According to Lemma 3, if for every $k \in \{0,\, \ldots,\, M\}$ at least one $\bar{g}_i^k(x^k,\, z^k)$, $i \in \{1,\, \ldots,\, l\}$ is a strongly convex function of $x^k \in X^k$, then for $\lambda \in (0,\, \varepsilon)$ (ε arbitrary small) there exists only one solution of every individual problem (12). Employing, the assertions of Lemma 1 and Lemma 4 we can see that the optimal function of $\varphi_{\mathcal{F}}^{\lambda}(M)$ is (under general assumptions) a Lipschitz function of $\lambda \in \langle 0,\, 1\rangle$ and x. Completed this consideration by scenario approach based on some stability results (for details see e.g. [11]), according to Lemma 2 and Lemma 4 we can obtain a relatively "good" approximation of criteria value functions and approximation of efficient points sets also.

Remark. Evidently, the introduced approach doesn't introduce completely efficient points in the multiobjective, multistage problems. However, we obtain a "reasonable" solution corresponding to the evaluation of individual criteria corresponding to the relevance of "underlying" economic problem. Maybe that some others results can be obtain employing the definition of multistage problems as a problem in some abstract mathematical space. However this consideration is over the possibilities of this contribution.

Acknowledgement

This work was supported by the Grant Agency of the Czech Republic under grants 402/07/1113, 402/08/0107 and 402/06/0990.

References

[1] J. Dupačová: Multistage Stochastic Programs: the State–of–the–Art and Selected Bliography. Kybernetika *31* (1995),2, 151–174.

[2] M. Ehrgott: Multicriteria Optimization. Springer, Berlin 2005.

[3] A. M. Geoffrion: Solving Biccriterion Mathematical Programs. Operations Research *15* (1967), 1, 39–54.

[4] A. M. Geoffrion: Proper Efficiency and the Theory of Vector Maximization. J. Math. Anal. Appl. *22*, (1968), 3, 618–630.

[5] J.-B. Hiriart–Urruty, C. Lamaréchal: Fundamentals of Convex Analysis. Springer, Berlin 2000.

[6] V. Kaňková: Stability in Stochastic Programming–the Case of Unknown Location Parameter. Kybernetika *29* (1993), 1, 97–112.

[7] V. Kaňková: Convexity, Lipschitz Property and Differentiability in Two–Stage Stochastic Nonlinear Programming Problems. Proceedings of the 3rd International Conference on Approximation and Optimization in the Carribean, Puebla, Mexico 1995. Aportaciones Matematicas, Serie Comunicaciones *24* (1999), 135–149.

[8] V. Kaňková: A Note on Multistage Stochastic Programming. In: Proc. 11th Joint Czech–Gerany–Slovak Conference: Mathematical methods in Economy and Industry. University of Technology , Liberec 1998, 45–52.

[9] V. Kaňková: Stochastic Programming Approach to Multiobjective Optimization Problems with Random Elements I. Research Report UTIA 2000, No. 1990.

[10] V. Kaňková: A Note on the Relationship between Strongly Convex Functions and Multiobjective Stochastic Programming Problems. In: Operations Research Proceedings 2004 (Heiin Fleuren, Dick den Hertog and peter Kort, eds.), Springer, Berlin 2005, 305–312.

[11] V. Kaňková: Multistage Stochastic Programs vis Autoregressive Sequences and Individual Probability Constraints. Kybernetika *44* (2008), 2, 151–170 (to appear).

[12] G. Salinetti and R. J.-B. Wets: On the Convergence of Sequences of Convex Sets in Finite Dimension. SIAM Review *21* (1979), 18–33.

Jan Kodera, Tran Van Quang

Centre for Dynamic Economics and Econometrics
University of Economics Prague
W. Churchill sq. 4, 130 67 Prague 3

e-mail: kodera@vse.cz, tran@vse.cz

Model Narmax and Its Application to Exchange Rate Dynamics

Abstract

Box and Jenkins model ARIMA (p,d,q) has contributed a great deal to explanation of behaviour of economical and financial time series. Nevertheless, in many cases, dynamics of an economic or financial variable is affected not only by its endogenous relationships, but there also are various exogenous variables which massively influence its path. Furthermore, its dynamic is not always linear which is not in accordance with non-linear nature of the relationships among economic and financial factors.

NARMAX (Nonlinear Auto-Regressive Moving Average polynomial model with eXogenous variables) is an extension of the ARIMA model. On the one hand, the inclusion of exogenous variable and nonlinear relationships into the model may substantially improve its performance and, on the other hand, an algorithm is used to select the most suitable candidates from a set of explanatory variables that should be included in the model. Using this algorithm, the model structure is determined not arbitrarily, but by an objective criterion.

The applicability of the model NARMAX will be demonstrated on time series of indices of Prague Stock Exchange PX and NYSE S&P500, series of exchange rate of CZKUSD and CZKEUR and series of interest rate differentials of CZK PRIBOR and USD LiUSD or EURIBOR respectively. Some interesting results were achieved.

Keywords: *exchange rate, non-linear methods, NARMAX model, interest rate differential, interest rate parity theory, Householder orthogonal transformation*

1 Introduction

Exchange rate is an important economic variable. On the one hand, it is one of four tools of economic policy. Besides it, they are taxes, government expenditures, interest rate through which an economy could be managed even though nowadays in most economies exchange rate is left freely floated or at least partially floated. On the other hand, money are transferred into various exchanges by means of exchange and they are restlessly moving around the globe to seek for higher returns at money markets. Therefore, study of exchange rate behaviour has become a central interest of both researchers and practitioners for long time.

Many of them believe that exchange rate behaviour follows a random walk, others try to find factors which influence its behaviour. As results of their research effort, there are several theories explaining exchange rate movements. One of them is the interest rate parity theory. This theory postulates that expected returns on interest bearing asset across countries must be equal if exchange rate movements are taken into account. If it is not the case, then market forces are set into motion and thus restore the equality. This theory is simple and elegant, nevertheless, it often does not hold in practice. Furthermore, this theory and others also omit the intrinsic dynamics of exchange rate and its reactions to the changes in economic fundamentals.

In this work, we try to explain exchange rate behaviour by connecting both inside and outside factors, namely interest rate differentials, causing its movements. To reach this goal, model NARMAX is used. In this model, not only linear relationship of endogenous and exogenous variables, but their nonlinear relationships are also examined. Our work is structured as following: in section 2, a brief summary of interest rate parity theory is presented. In section 3, the essence of model NARMAX is explained. In section 4, exchange rate series CZKEUR and CZKUSD and series of interest rate differentials between Czech crown and Euro and US dollar respectively are tested on model NARMAX. In the last section some conclusions are given.

2 Uncovered interest rate parity theory

Uncovered interest rate parity theory postulates that the rate of return from bearing interest assets dominated in various currency must be equal:

$$(1 + IR_{D,t}^{t+n}) = (1 + IR_{F,t}^{t+n})\frac{E(ER_{t+n})}{ER_t}$$

where $IR_{D,t}^{t+n}$,$IR_{F,t}^{t+n}$ are the domestic and foreign interest rate for the time period t to $t+n$, $E(ER_{t+n})$ is the expected value of spot exchange rate at time t for time $t+n$ and ER_t is the spot exchange rate at time t. If for any reason this equation does not hold, then market forces immediately set into motion and the equilibrium will be restored. For instance, if domestic interest rate is higher than interest rate abroad, that would lead to an increased inflow of foreign capital to domestic economy and force the domestic currency to appreciate and the equilibrium will be rebalanced. The only barrier of the equilibrium restoration is the impossibility of free capital movement.

In reality, the restriction in capital movement is not the only reason that prevents the equilibrium from its restoration. In developed economies, there are no restrictions in capital mobility but uncovered interest rate parity theory does not hold as well. Besides transaction costs, opponents of the uncovered interest rate parity theory often point out that not all interest bearing assets are homogeneous. They are rather heterogeneous with respect to risk. And the more risky the assets are the higher returns they should offer. The other reason why the reality differs from the uncovered interest rate parity theory is that investors might need to diversify their portfolios and do not want to invest to only one currency.

Other economists come up with different arguments for explanation why the reality deviates from the uncovered interest rate parity theory. They are uncertainty, non-ergodicity[1] and endogenous money. The first two arguments assume that the investors' level of confidence in their

[1]A system is called ergodic if it tends in probability to a limiting form that is independent on the initial conditions.

own future forecast drawn from historical information is very important. As not all market expectations are equal, and when all other reasons for deviation from the uncovered interest rate parity theory are excluded, then if the investors' confidence in their own expectations is high, interest rate parity theory tends to hold, otherwise, it tends to fail.

As far as the argument of endogenous money is concerned, its reasoning is following. When money is exogenous, only central monetary authority can affect the money supply. One of its possible tools is sterilization of capital inflow into the economy. The sterilization dampens capital flows created by interest rate differentials and might act as an indirect impediment to capital mobility and prevent the exchange rate adjustment. In the case of endogenous money, when money supply is driven by demand and banks loan to the credit worthy clients first and find reserves later, the private bank sector adjusts money supply by purchasing securities from central monetary authority with its unwanted funds or overdrafts its funds from central bank by reducing its liability to the central monetary authority. Acting that way, the private banks sterilize the capital inflow from abroad.

3 NARMAX model

NARMAX (Nonlinear AutoRegressive Moving Average with eXogenous variables) model is an extension of the traditional linear Box–Jenkins ARIMA model. In this extension, not only intrinsic dynamics of dependent variable, but also its relationships with other exogenous variables are examined. Furthermore, the model can capture the influences of both linear and nonlinear terms of independent variables which is impossible to be caught in linear models. But the model as the whole is linear in the parameters.

Let's consider a dependent variable

$y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$, where N is the number of observations,

and a matrix of exogenous and independent variables

$X = \begin{bmatrix} x_{11} \cdots x_{1M} \\ \vdots \ddots \vdots \\ x_{N1} \cdots x_{NM} \end{bmatrix}$, where M is the number of exogeneous and independent variables,

and a vector of noise which is a zero mean independent sequence

$$e = \begin{bmatrix} e_1 \\ \vdots \\ e_N \end{bmatrix}.$$

The dynamics of the dependent variable is a function of its own lagged value, of the exogenous and independent variables and their lagged value, of the noise and its lagged values and of the mutual combinations of all these three parts.

$$y_t = f(y_{t-k}, X_{t-l}, e_{t-m}, C(y_{t-k}, X_{t-l}, e_{t-m})),$$

where *k, l, m* are the numbers of lags which generally are not equal.

As model is linear in its parameters, the least squares technique can be used to estimate a vector of parameters $\hat{\theta}$:

$y = \psi\hat{\theta} + \xi$, where ψ is a matrix of regressors of dimension (N x P) (P is the number of regressors) and ξ is the vector of residuals. It is necessary to mention that the matrix $\psi^{-1}\psi$ is of a relatively huge dimension and when inverting it one might encounter such difficulties as its determinant can be close to zero and rounding-off might cause certain inaccuracies later.

To avoid these problems, orthogonal least squares technique is used. An important feature of this orthogonal least squares technique by Householder transformation is that it enables simultaneously to select the most important regression variables and estimate the values of the corresponding parameters. The essence of this algorithm is following.

When carrying out the regression $y = \psi\hat{\theta} + \xi$, we try to find a vector of parameters $\hat{\theta}$ such that the sum of squares of residuals ξ is minimized. It is well known that the matrix of regressors can be factorized into two matrices: $\psi = QR$, where Q is a unitary matrix, that is $Q^T Q = I$, and R is an upper triangular matrix. As Q is a unitary matrix, we get:

$y^T Q^T Q y = y^T I y = y^T y$ and $\xi^T Q^T Q \xi = \xi^T I \xi = \xi^T \xi$. This means that no matter whether the regression is carried out on the original matrix of regressors $y = \psi\hat{\theta} + \xi$or on its orthogonalized form $Q^T y = Q^T \psi\hat{\theta} + Q^T \xi$, the sum of squares of their corresponding residuals are the same. For $\psi = QR$, substituting it into the orthogonalized equation $Q^T y = Q^T \psi\hat{\theta} + Q^T \xi$, we get: $Q^T y = Q^T QR\hat{\theta} + Q^T \xi$ or $Q^T y = R\hat{\theta} + Q^T \xi$ using the property that Q is a unitary matrix. It can be rewritten as:

$$\begin{bmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_N \end{bmatrix} = \begin{bmatrix} r_{11} r_{12} \cdots r_{1P} \\ 0 r_{22} \cdots r_{2P} \\ \cdots\cdots \\ 00 \cdots r_{PP} \\ 000 \\ \cdots\cdots \\ 00 \cdots 0 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_P \end{bmatrix} + \begin{bmatrix} \tilde{\xi}_1 \\ - \\ \tilde{\xi}_2 \end{bmatrix}$$

It is absolutely clear from the above equation that because of the zeros under the main diagonal of matrix R, the sum of squares of residuals $\tilde{\xi}_2$ by no means can be explained by any regressor in the matrix of regressors ψ. On the other hand, adding a new regressor to the model can reduce the sum of squares of residuals in residuals $\tilde{\xi}_1$.

Let us consider instead of original NARMAX model we carry out the regression with regressors w_i which are orthogonal over the data set as following:

$$y = Wg + \Xi,$$

where W is a matrix of orthogonalized regressors, g is a vector of parameters and Ξ is a vector of residual.

Then the error reduction ratio thank to i-th regressor w_i is defined as the proportion of the dependent variable explained by w_i:

$$ERR_i = \frac{g_i^2 \langle w_i, w_i \rangle}{\langle y_i, y_i \rangle},$$

where parameters g_i is defined as:

$$g_i = \frac{\langle w_i, y \rangle}{\langle w_i, w_i \rangle}.$$

The error reduction ratio serves as a criterion for model selection. The algorithm for model selection procedure is following:

At the beginning of the p-th stage of the selection procedure, the regressor matrix ψ has been transformed into the shape: $\psi^{(j-1)} = [\psi_1 \cdots \psi_j^{p-1} \cdots \psi_P^{p-1}]$ and these steps are taken:

1. for $p \leq j \leq P$ calculate $g_p^{(j)}$ and $ERR_p^{(j)}$
2. find maximal value $ERR_p^{(j)}$
3. interchange column ψ_j^{p-1} and column with max $ERR_p^{(j)}$
4. carry out Householder transformation.

This procedure is terminated when the required number of parameters is reached or required sum of $ERR_p^{(j)}$ is reached.

4 Verification of the interest rate parity theory

Using model NARMAX explained above, we examine the dynamics of exchange rate of Czech crown (CZK) to US dollar (USD) and common currency euro (EUR). As exogenous input, we choose interest rate differential between interest rate PRIBOR 3M and LIUSD 3M and EURIBOR 3M respectively. Time series of exchange rate CZKUSD and CZKEUR and interest rate differentials are of period from March 2006 to March 2008. Figures $1-4$ show the evolution of these variables in the period. We can observe that in this period of time exchange rate CZKUSD exhibits a constant decreasing trend whereas with the exchange rate CZKEUR the same trend is seen at the last part o the period. On the other hand, interest rate differentials between Czech crown and US dollar has an increasing trend and the differentials between Czech crown and common currency Euro does not show any clear trend at all. In most cases interest rate on Czech crown market is lower than on US dollar or Euro markets.

Because exchange rate time series are not stationary, they are transformed into series of relative changes and the relationship between them and interest rate differentials among others is examined. A matrix of regressors is created. It includes lagged series of exchange rate and interest rate differentials up to order of 2 and their mutual combinations up to order of 3 (the

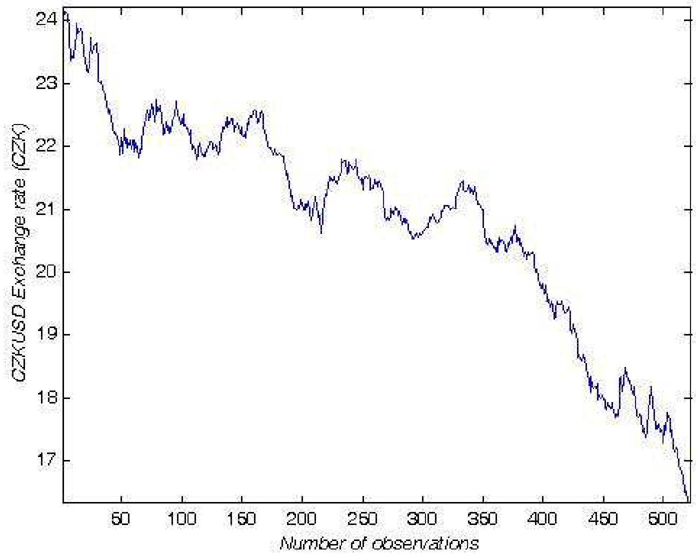


Figure 1: The market trend of CZKUSD exchange rate

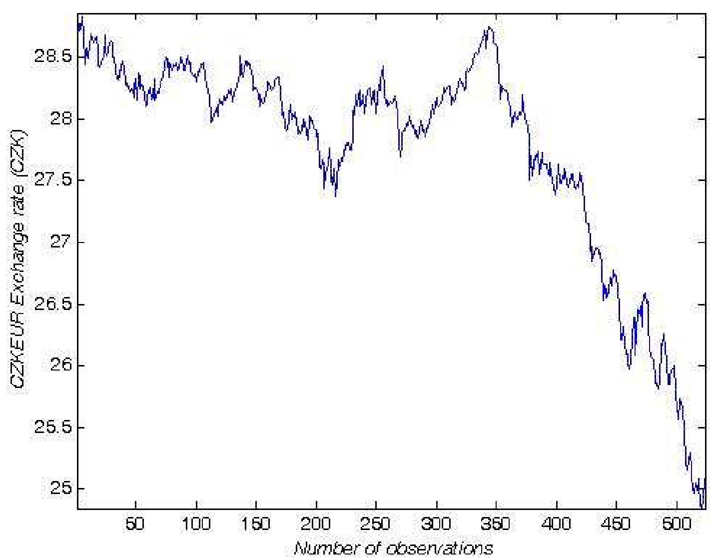


Figure 2: The market trend of CZKEUR exchange rate

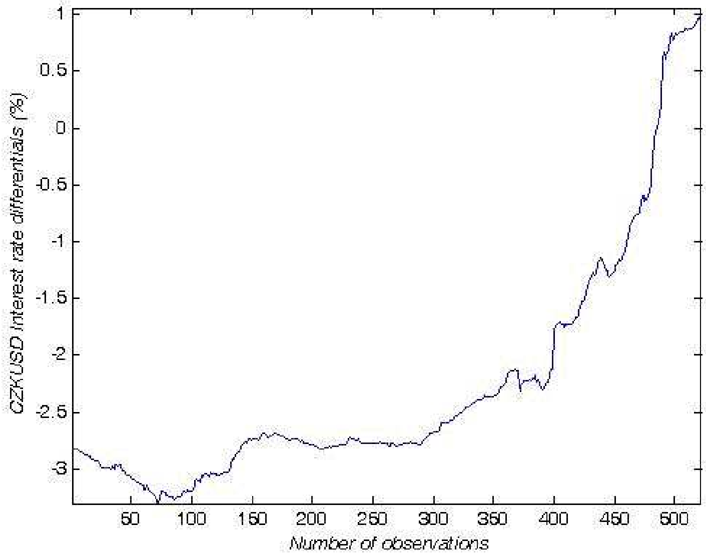


Figure 3: The market trend of CZKUSD interest rate

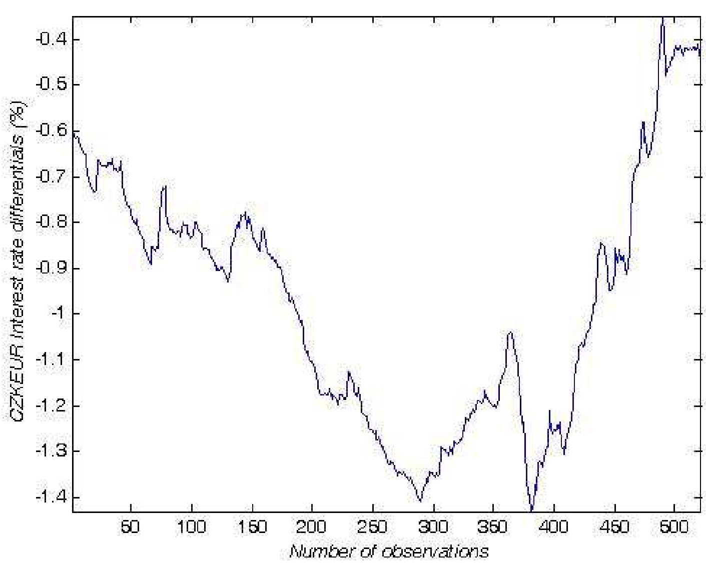


Figure 4: The market trend of CZKEUR interest rate

Regressor	Estimated parameter	ERR
ER(-1)	0.85742917201924	0.88230823646010
Constant	-1.66428961909276	0.00106765677093
IRD(-1)	-0.00461471137514	0.00088410736772
ER(-2)	0.06722671469531	0.00052226691696

Table 1: The results of model selection and parameter estimation for series CZKUSD

Regressor	Estimated parameter	ERR
ER(-1)ˆ3	-0.00002099573330	0.01133956815829
Constant	-4.77167106111829	0.00940581021151
ER(-2)ˆ2*IRD(-2)	-0.00035278635031	0.00820618478917
ER(-2)ˆ2* IRD(-1)	0.00033484636549	0.00737408905279

Table 2: The results of model selection and parameter estimation for series CZKEUR (ER = exchange rate; IRD = interest rate differentials)

repetitive combinations of them are also included). A script in MATLAB is written to realize the algorithm for model selection and its parameter estimation. In tables 1 and 2 the results for four most relevant regressors are shown.

From the table 1 we can see that a vast majority of variability in the changes of exchange rate CZKUSD in the period of our interest is explained by its lagged values of order 1 (88.2 %). The contribution of other variables taken into account is very small. On the other hand, no variable of our interest explains the dynamics of the change of exchange rate CZKEUR in the same period. This result is in accordance with the movement of exchange rate CZKUSD. On the contrary, similar movement is not observed at CZKEUR. The validity of the uncovered interest rate parity theory is not proven.

5 Conclusions

Because of the relevant of the exchange rate in economies and capital markets, we model the behaviour of series of exchange rate CZKUSD and CZKEUR in the last 2 years from March 2006 to March 2008 by using nonlinear autoregressive moving average with exogenous inputs (NARMAX) model. This model can capture not only intrinsic dynamics of exchange rate, but also the influence of exogenous economic variables. Another advantage of NARMAX model is that it can capture the influence of mutual combination of endogenous and exogenous variables. And we use an algorithm which enables to choose the most relevant regressors and their parameter estimation simultaneously.

We choose interest rate differentials between Czech crown and US dollar or common currency Euro as an exogenous input in accordance with uncovered interest rate parity theory. No matter how elegant the uncovered interest rate parity theory is, we find that the rate differentials between Czech crown and US dollar or common currency have no influence on the behaviour of exchange rate CZKUSD and CZKEUR. This conclusion is in line with the results of a variety of works dealing with this topic which assume that uncovered interest rate parity theory is useless. On the other hand, in this period, we find very strong endogenous influence on dynamics of the exchange rate CZKUSD which accounts for 88% of its variability. Nevertheless, this influence is not found in the dynamics of exchange rate CZKEUR.

References

[1] Chen S., Bilings, S.A., Luo W.(1989): *Orthogonal least squares methods and their application to nonlinear system identification*, International Journal of Control, vol. 50, No 5, pages 1873 – 1896.

[2] Aguirre, L.A., Aguirre A. (2002): *Analysis of economic time series using NARMAX polynomial models*, in Soofi A.S., Cao L. (2002): *Modelling and Forecasting Financial Data, Technique of Nonlinear Dynamics*, Kluwer Academic Publisher, StateMassachusetts, USA.

[3] Durčáková J., Mandel M., Tomšík V. (2005): *Puzzle in the theory of uncovered interest rate parity – empirical verification for transitive countries,* Finance India, Vol. 19, No 2, pp 449 – 464.

[4] Harvey J. T. (2004): *Deviation from uncovered interest rate parity: A post-Keynesian explanation*, Journal of Post Keynesian Economics, Vol.27, No. 1, pp. 19 – 35.

Miloš Kopa

Department of Probability and Mathematical Statistics
Faculty of Mathematics and Physics
Charles University Prague

e-mail: kopa@karlin.mff.cuni.cz

Irena Šindelářová, Vladislav Chýna

Department of Econometrics
University of Economics Prague
W. Churchill sq. 4, 130 67 Prague 3

e-mail: isin@vse.cz, chyna@vse.cz

Financial Data Prediction via Wavelet Transform—Exchange Rate Study

Abstract

The paper deals with modelling of financial time series using various econometric models. We analyse and forecast weekly logarithmic returns of EUR/CZK exchange rate. Firstly, we transform the given data using the discrete Daubechies wavelet decomposition [1]. It divides the original data into two segments (low frequencies and high frequencies). We discuss the choice of threshold level in each of these two segments separately. Low frequencies are not changed. For high frequencies several threshold levels are used. Then the time series is de-noised by means of the proposed thresholding. Secondly, we estimate ARMA and GARCH model for time prediction of future values of each reconstructed segment for each choice of threshold level. Then we merge two predicted series into final data forecast and we find the optimal threshold level, that is level which minimize total prediction error. Finally, we compare these results with approach used in [2].

Keywords: Time series forecasting, ARMA, GARCH models, discrete wavelet transform, signal thresholding

1 Introduction

Wavelet transform is a well know approach to de-noising of signals and time series. It was used in many applications, see e.g. [1], [7]. It is alternative concept to classical ARIMA models with possible GARCH extensions for modelling the varying conditional heteroskedasticity introduced in [3], [6] and [5]. In this paper we focus on financial time series application of wavelet techniques. One of the possible ways of employing wavelet transform and signal thresholding in financial data prediction was suggested in [2]. The aim of [2] was to compare the time series forecast via ARMA – GARCH models for origin data with that for de-noised data. De-noising was done for various threshold levels using both soft and hard thresholding. The best performance of the s-period-ahead forecast was reached for 100% level of thresholding. In this case soft thresholding is equivalent to hard thresholding because all high frequencies are removed. The prediction was done after the de-noising procedure was finished.

In this paper we extend methods published in [2]. The main difference is in the way of prediction constructions. While in [6], the prediction was done after the de-noising procedure was finished, in this paper, we model and forecast low frequencies and high frequencies separately. Firstly, we decompose the time series in four segments (low frequencies and three groups of high frequencies). The highest frequencies are considered as white noise. Therefore we use 100 % threshold level for this segment. For the other two groups, we apply several choices of threshold level. As it is usual, low frequencies are not changed. After thresholding procedure, we reconstruct each segment separately, that is, we obtain two time series (a reconstruction of low and high frequencies separately). Than we model and forecast each time series via ARMA – GARCH models and we merge it in final prediction.

The rest of the paper is structured as follows. Section 2 deals with wavelet transform and signal thresholding. It introduces the basic terms as Daubechies wavelets, scaling function, wavelet coefficients, etc. Section 3 recalls ARMA and GARCH models with focus on prediction constructions. The last section illustrates these techniques on CZK/EUR exchange rate series. We compare the predictions for several threshold levels. When 100% threshold level is used our approach is equivalent to that in [6]. We will show that 50% level gives better performance for data predictions.

2 Discrete Wavelet Transform and Signal Denoising

Wavelet functions are pulses created an orthogonal system used as the base for integrated functions. Wavelets are functions $w_{j,k}(t)$ with parameters of time and scale. The orthonormal set of functions is derived from the so called mother wavelet defined by relation (see [3] for more details):

$$w_{j,k}(t) = \frac{1}{\sqrt{j}} w\left(\frac{t-k}{j}\right)$$

Varying of the wavelet frequency information (scale) is referred to the dilation of the wavelet. Low frequencies in the wavelet representation are related to a scaling function. Scaling function $\phi(x)$ and wavelet function $w(x)$ can be defined by the dilation equation in the following form:

$$\phi(x) = c_0\phi(2x) + c_1\phi(2x-1) + c_2\phi(2x-2) + c_3\phi(2x-3)$$

$$w(x) = -c_3\phi(2x) + c_2\phi(2x-1) - c_1\phi(2x-2) + c_0\phi(2x-3)$$

where $c_0 = (1+\sqrt{3})/4, c_1 = (3+\sqrt{3})/4, c_2 = (3-\sqrt{3})/4, c_3 = (1-\sqrt{3})/4$. The mother wavelet defined by this described manner is Daubechies wavelet function for four coefficients.

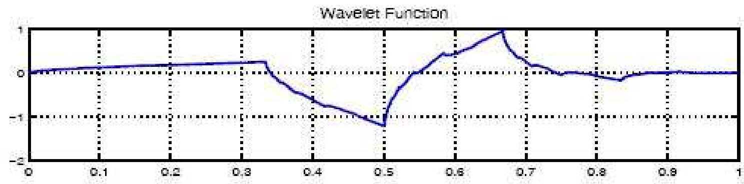


Figure 1: Daubechies wavelet function

The discrete wavelet transform of a signal $f(t)$ is the time and scale signal representation expressed by the wavelet coefficients. These coefficients $a_{j,k}$ can be obtained as the product of the signal and orthogonal matrices V and W, respectively.

$$V = \begin{bmatrix} c_3 & c_2 & c_1 & c_0 & 0 & 0 & ... \\ 0 & 0 & c_3 & c_2 & c_1 & c_0 & ... \\ 0 & 0 & 0 & 0 & c_3 & c_2 & ... \\ ... & ... & ... & ... & ... & ... & ... \end{bmatrix} W = \begin{bmatrix} c_0 & -c_1 & c_2 & -c_3 & 0 & 0 & ... \\ 0 & 0 & c_0 & -c_1 & c_2 & -c_3 & ... \\ 0 & 0 & 0 & 0 & c_0 & -c_1 & ... \\ ... & ... & ... & ... & ... & ... & ... \end{bmatrix}$$

Resulting coefficients $a_{j,k}$ are the representation of the given signal in the system of the wavelets. In this paper we use this multiresolution representation organised signal frequency information into two levels (various signal scales).

DWT is applied in signal de-noising using thresholding of wavelet coefficients (Fig 2.). The main idea is in the fact that signal frequency components are different from the noise corresponding to the highest frequencies of the observed time series. The algorithm follows these steps.

- Decomposition is DWT of the given signal providing wavelet coefficients.
- Thresholding is the modifying of the obtained coefficients. In the paper we use the soft thresholding with the selected threshold level.
- Reconstruction is the inverse discrete wavelet transform of the modified series.

3 Time Series Modelling and Forecasting

One of the most popular approach to time series modelling and forecasting uses ARMA(p,q) models in combinations with GARCH(m,n) models. A centred mixed autoregressive moving average process, denoted ARMA(p,q), satisfies:

$$y_t = \sum_{k=1}^{p} \varphi_k y_{t-k} + \varepsilon_t + \sum_{l=1}^{q} \theta_l \varepsilon_{t-l}$$

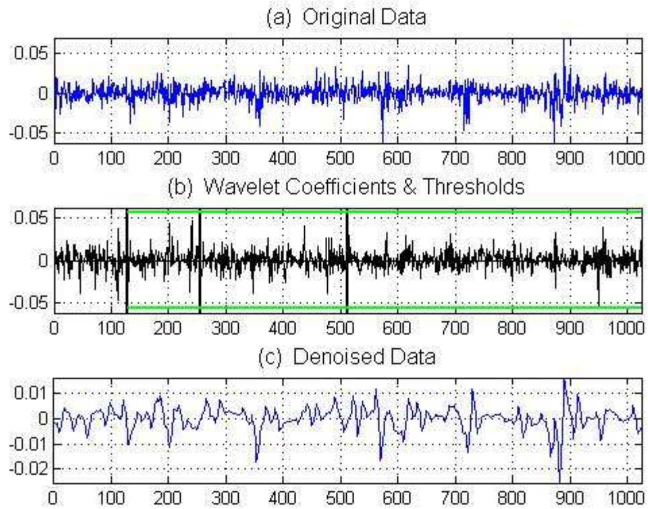


Figure 2: The principle of signal thresholding

where $\{\varepsilon_t\}$ is a white noise process, that is $E\varepsilon_t = 0$, $var(\varepsilon_t) = \sigma^2$ and $E\varepsilon_t\varepsilon_s = 0$ for$t \neq s$. To guarantee the stationarity of ARMA processes $\{y_t\}$, we need to assume that the roots of characteristic polynom:

$$1 - \varphi_1 z - \varphi_2 z^2 - ... - \varphi_p z^p = 0$$

lie outside the unit circle. Moreover, for the purpose of forecasting we will assume that $\{y_t\}$ is an invertible process, that is the roots of polynom:

$$1 - \theta_1 z - \theta_2 z^2 - ... - \theta_q z^q = 0$$

lie outside the unit circle too.

In ARMA processes, the unconditional variance of ε_t is constant and the conditional variance of ε_t is also constant in time. However, the financial time series are usually much more volatile at some times than at others. Therefore we might be interested in modelling not only the level of the series $\{y_t\}$ but also the conditional variance of residuals. One approach is to describe ε_t^2as itself following an AR(m) process:

$$\varepsilon_t^2 = \zeta + \sum_{k=1}^{m} \alpha_k \varepsilon_{t-k}^2 + \eta_t$$

where $\{\eta_t\}$ is a new white noise process. A process$\{\varepsilon_t\}$ satisfying this formula is described as an auto-regressive conditional heteroskedastic process of order m, denoted $\varepsilon_t \sim$ ARCH(m). This class of processes was introduced in [5]. It is convenient to use an alternative representation.

Let

$$\varepsilon_t = \sqrt{h_t}\nu_t$$

where $\{\nu_t\}$ is an i.i.d. sequence with zero mean and unit variance. If $h_t$evolves according to

$$h_t = \zeta + \sum_{k=1}^{m} \alpha_k \varepsilon_{t-k}^2$$

then $\varepsilon_t \sim$ ARCH(m). More generally we can consider a process for which ε_t^2 follows ARMA(m,n) process and equivalently:

$$h_t = \zeta + \sum_{k=1}^{m} \alpha_k \varepsilon_{t-k}^2 + \sum_{l=1}^{n} \beta_l h_{t-l}.$$

This is the generalized autoregressive conditional heteroskedasticity model (GARCH(m,n)) proposed in [2]. Both ARCH and GARCH models will be used in order to improve the quality of ARMA models.

For a centred, stationary and invertible ARMA process, an optimal s-period-ahead linear forecast $\widehat{y}_{t+s|t}$ based on $\{y_t, y_{t-1}, \ldots\}$, that is on the data till t-th period, can be obtained from the formula:

$$\hat{y}_{t+s|t} = \sum_{k=1}^{p} \varphi_k \hat{y}_{t+s-k|t} + \sum_{l=1}^{q} \theta_l \hat{\varepsilon}_{t+s-l}$$

with $\hat{y}_{\tau|t} = y_\tau$ for $\tau \leq t$ and generated recursively from

$$\begin{aligned} \hat{\varepsilon}_k = & \quad y_k - \hat{y}_{k|k-1} && \text{for } k \leq t \\ \hat{\varepsilon}_k = & \quad 0 && \text{for } k > t \end{aligned}$$

Since we are interesting only in point forecast of $\{y_t\}$ we can use this prediction also in the case of changing conditional heteroskedasticity. However, the presence of conditional heteroskedasticity can significantly change the estimations of ARMA coefficients. More details about prediction in ARMA – GARCH models can be found in [1].

4 Numerical application

In this numerical study, we illustrate the advantages of our approach on CZK/EUR exchange rate time series. First we transform the weekly closed data (1.1.1999 – 4.7.2007) into logarithmic returns

$$r_t = \log \frac{P_t}{P_{t-1}}.$$

It is well known, that the mathematics models give better results after this simply transformation. Since we want to use ARMA-GARCH models to describe the behaviour of the data, first we use two tests [4] to avoid the linear trend and periodicity. The null hypothesis of randomness was not rejected by two tests of randomness (test based on signs of differences, test of changing points). The results (p-value) are described in the following table:

Table In the first step we decompose the time series into four segments (low frequency and

	test based on signs of differences (p-value)	test of changing points (p-value)
EUR/CZK	0.298	0.798

Table 1: The p-values of randomness tests

three groups of high frequencies). The highest frequencies are considered as white noise. Therefore we use 100 % threshold level for this segment. For the other two groups, we apply three choices of threshold level: 100 %, 50 % and 20 %. When 100% threshold level is used our approach is equivalent to that in [6] because all high frequencies are removed and only low frequencies are reconstructed. To save information hidden in two groups of high frequencies we also apply 50 % and 20 % threshold level. After reconstruction we obtain two time series: one for low frequency, second for high frequencies with 50 % threshold level, 20 % threshold level, respectively.

The following table shows the best models for low and high frequencies when 50 % and 20 % threshold level is used.

	Low frequencies	High frequencies
x = *100%*	Model AR(2) - GARCH(1,1) : $y_t = 1.6y_{t-1} - 0.64y_{t-2} + \varepsilon_t$ $\varepsilon_t = \sqrt{h_t}\nu_t$ $h_t = 0.74\varepsilon_{t-1}^2 + 0.36h_{t-1}$	0
x = *50%*	Model ARMA(2,1) : $y_t = 0.4y_{t-1} - 0.4y_{t-2} + \varepsilon_t - 0.98\varepsilon_{t-1}$	Model ARMA(1,1) : $y_t = 0.08y_{t-1} + \varepsilon_t - 0.8\varepsilon_{t-1}$
x = *20%*	Model ARMA(1,1) : $y_t = 0.29y_{t-1} + \varepsilon_t - 0.97\varepsilon_{t-1}$	Model MA(2): $y_t = \varepsilon_t + 0.43\varepsilon_{t-1} + 0.44\varepsilon_{t-2}$

Table 2: The best ARMA – GARCH models for low and high frequencies using various threshold levels

For each model, we construct one to four-steps-ahead forecast. Then we merge low and high frequencies prediction into final forecast.

As can be easily seen in Figure 3 the best prediction performance is reached for 50 % threshold.

	100% threshold level	50% threshold level	20% threshold level	Original data
1 step ahead	0,001024	-0,0068	-0,0032	-0,013510
2 steps ahead	0,001260	-0,0068	-0,0032	-0,002120
3 steps ahead	0,001372	-0,0035	0,0011	-0,008170
4 steps ahead	0,001400	-0,0035	0,0011	0,000535

Table 3: Comparison of original data and one to four-steps-ahead forecasts for 100 %, 50 % and 20 % threshold level

Alternatively we can compare the accuracy of these point predictions using total prediction error defined as:

$$\mathrm{TPE}(y, t, s) = \sum_{k=t+1}^{t+s} (y_k - \hat{y}_{k|t})^2$$

The following table shows that using 50 % threshold level instead of 100 % threshold level suggested in [6] leads to the 66 % reduction of TPE.

100% threshold level	50% threshold level	20% threshold level
0,000314	0,000105	0,000194

Table 4: Comparison of total prediction errors when 100 %, 50 % and 20 % threshold level is used

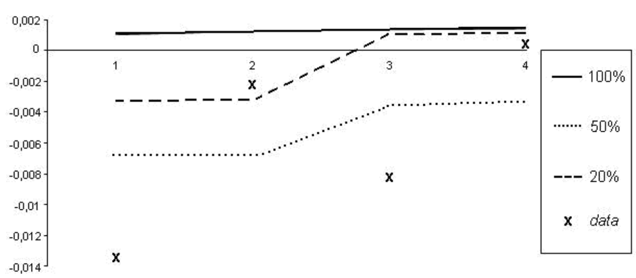


Figure 3: The graphical comparison of original data and one to four-steps-ahead forecasts when 100 %, 50 % and 20 % threshold level is used

5 Conclusion

Numerical results of real data modelling show that the given time series can be successfully predicted after its thresholding. In the paper, we decomposed the CZK/EUR exchange rate time series into four segments applying DWT. As usual, low frequencies were not changed while for the highest frequencies 100 % threshold level was used. For the other two high frequency segments, several thershold levels were proposed, in order to find the best one. The accuracy of one to four-steps-ahead prediction using ARMA – GARCH methodology was considered as the optimality criterion. The optimal thershold level was found as 50 % of maximal possible value where maximum threshold level removes all the highest signal frequencies. The accuracy of the forecasts can be seen in Figure 3, Table 3 and Table 4.

Acknowledgement

This work was partially supported by grant MSMT 0021620839.

References

[1] C. S. Burrus, R. A. Gopinath, and H. Guo. *Introduction to Wavelets and Wavelet Transform.* Prentice Hall, New Jersey, 1998.

[2] M. Kopa, I. Šindelářová, V. Chýna: Thresholding in Financial Data Prediction, Proceedings of Academic Internations Conference Increasing Competitiveness of Regional, National and International Markets Development – New Challenges, 1-8, 2007.

[3] R. T. Baillie and T. Bollerslev. *Prediction in dynamic models.* Journal of Econometrics 52 (1992) 91-113.

[4] T. Cipra: *Analýza časových řad s aplikacemi v ekonomii*, SNTL, 1986, Praha.

[5] R. F. Engle. *Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation.* Econometrica 50 (1982) 987-1007.

[6] T. Bollerslev. *Generalized autoregressive conditional heteroskedasticity.* Journal of Econometrics 31 (1986) 307-327.

[7] G. Strang a T. Nguyen. *Wavelets and Filter Banks.* Wellesley-Cambridge Press, Wellesley, USA, 1996.

Donata Kopańska-Bródka, Renata Dudzińska-Baryła

Department of Operations Research
University of Economics in Katowice
ul. 1-Maja 50, 40-287 Katowice

e-mail: broda@ae.katowice.pl, dudzinska@ae.katowice.pl

The Influence of the Reference Point's Changes on the Prospect Value

Abstract

The idea of the value function in the Prospect Theory differs significantly from the utility function considered in the Expected Utility approach. Based on the empirical evidence that humans are often more sensitive to how their situation differs from some reference level the value function is constructed in particular manner. Hence, instead of the utility function of the absolute wealth the value function of the changes in relation to the reference point is considered as the adequate model of the decision maker's behaviour. The reference level can be determined in different way, for example by value of past gain or temporary wealth or expectation of future wealth. On the other hand, the reference point can also change along the utility function. Having in mind that the shape of the utility function in the Expected Utility Model varies according to the decision maker's attitude towards risk, the value function also ought to depend on this attitude. The objective of our presentation is to show, how the shape of value function depends on the changes of the reference point and the measure of risk. We assumed, that attitude towards risk is implied by the formal properties of the utility function and the properties of the absolute risk aversion function are known.

Keywords: *prospect theory, utility function, changes of the reference point*

1 Introduction

The Prospect Theory [6] becomes more and more popular theory describing the decision makers' choices in the situation of risk. Results of many experiments and tests, which were inconsistent with classical rationality expressed in the Expected Utility Theory, were justified by the principles of the Prospect Theory or the Cumulative Prospect Theory [14]. The behaviour of the decision makers described by the Allais paradox or Ellsberg paradox, as well as the coexistence of gambling and insurance was explained by the Prospect Theory. There are also new competing theories, e.g. the Rank Dependent Utility Theory by Quiggin, which are trying to characterize the process of decision making at risk.

In the Prospect Theory the value of the lottery depends on the value function $v(x)$ and the probability weighting function $g(p)$. The scale g assigns to each probability p a decision weight $g(p)$. This decision weight reflects the impact of p on the over-all value of the prospect. The scale v assigns to each outcome x a number $v(x)$ which reflects the subjective value of that outcome. It is assumed that $v(0) = 0$, $g(0) = 0$ and $g(1) = 1$.

The difference between the Prospect Theory and the Cumulative Prospect Theory is in the formulation of the weights. In the Cumulative Prospect Theory the value of the prospect is multiplied by the difference between transformed adequately cumulated probabilities instead of the simple transformed probabilities.

In Prospect Theory outcomes are understood as gains and losses, rather than the final states. Gains and losses are the changes in the wealth from a reference point. This point can be interpreted as a current wealth of a decision maker. The value function in the Prospect Theory is:

- defined for gains and losses in relative to some reference point;
- concave for gains ($v''(x) < 0$ for $x > 0$) and convex for losses($v''(x) > 0$ for $x < 0$). It is consistent with the diminishing marginal value of gains and losses. The concavity of the value function for gains means that decision makers are risk avers and they choose certain decision and do not take risky action connected with even higher gain. Whereas considering losses decision makers are risk-prone (convexity of the value function) and they choose to take risky loss than certain loss;
- steeper for losses and than for gains, because the decision makers feel the loss more than the gain of the same absolute value.

The decision weight measures the impact of event on the desirability of prospect. The weighting function g is an increasing function of p and $g(0) = 0$ and $g(1) = 1$. Furthermore, for small probabilities $g(p) > p$, it means that the decision makers overweight low probabilities. And the next property of the probability weighting function is that for all $p \in (0, 1)$, $g(p)+g(1-p) < 1$. As a consequence of this property we have that $g(p) < p$ for high probabilities.

In the field of the Prospect Theory researchers focus on the following group of issues:

- the analytic form of the value function and the probability weighting function, e.g. [8; 3; 13];
- the choice of the reference point, e.g. [9];
- the loss aversion, e.g. [12; 10; 1];
- the risk aversion, e.g. [11; 4; 2].

Based on the Prospect Theory the decision maker individually values changes of wealth in relation to some reference point which is understood as a status quo. Then the level of wealth which states such a reference point is of great importance in decision modelling and its value directly affects the valuation of the risky decision. In the evaluation of the reference point, the von Neumann and Morgenstern utility is applied. Hence it is told that reference point travels along the utility curve of the decision maker, moreover the subject's utility curve includes the information about strength of his aversion toward risk. In the following part we show, how the change of the reference point influences the value function as well as the valuation of the risky alternative.

2 The value function in Prospect Theory

Kahneman and Tversky have proposed the model of the decision maker's behaviour in the face of risky choice. In this model the loss aversion is included by assuming specific form of the value function. The value function above the reference point (concerning the section of gains) is concave and it reflects the subject's aversion toward risk. Whereas below the reference point the value function is convex, which determines the of risk-prone attitude when faced of losses. So the decision maker's attitude toward risk is changing near the reference point.

The purpose of further research is the investigation of the relationship between the value function and the reference point and the strength of the risk aversion. We further consider the increasing value function measuring the value of the changes in wealth relative to some reference point w_i:

$$v_{w_i}(x) = \begin{cases} v_i(x) & for\, x \geq 0 \\ -\lambda v_i(-x) & for\, x < 0 \end{cases} \tag{1}$$

where λ is the degree of loss aversion and $v_i'(x) > 0$and $v_i''(x) < 0$ for $x > 0$.

Values of $v_i(x)$ for x > 0 determine the valuation of the relative gains (in the section of gains) while values of $-\lambda v_i(x)$ for $x < 0$ determine the valuation of the relative losses. The value function in the work of Kahneman and Tversky [14] was defined by means of the power function:

$$v(x) = \begin{cases} x^\alpha & for\, x \geq 0 \\ -\lambda(-x)^\alpha & for\, x < 0 \end{cases} \tag{2}$$

where $\lambda > 1$ and $0 < \alpha < 1$.

In the section of gains the value function reflects the aversion toward risk while in the section of losses the inclination toward risk.

It is beyond all doubt that the valuation of the relative gains depends on the reference point. In the last years many researchers have investigated the changes of the loss aversion in dependence on the reference point. D.R. Just and S. Wu [5] have showed that the reward contracts are optimal when reservation opportunities are independent of the reference point. In the work of G.B. Davies and S.E. Satchell [2] the sufficient conditions to ensure risk aversion in the CPT have been proved.

The change of the reference point can lead to such a change of valuation of the prospect that the preference relationship of two prospects changes. The example mentioned below illustrates such situation. Let's assume that the subject evaluates the risky choices by means of the PT values, where the value function $v(x)$ and the probability weighting function $g(p)$ have the form proposed by Kahneman and Tversky in their work [14]. Further let's consider two lotteries:

$$\begin{aligned} L1 &= \{(\mathbf{2}, 0.1); (\mathbf{3}, 0.2); (\mathbf{4}, 0.3); (\mathbf{5}, 0.2); (\mathbf{8}, 0.1); (\mathbf{10}, 0.1)\} \\ L2 &= \{(\mathbf{1}, 0.5); (\mathbf{4}, 0.1); (\mathbf{6}, 0.1); (\mathbf{8}, 0.1); (\mathbf{9}, 0.1); (\mathbf{10}, 0.1)\} \end{aligned}$$

In lottery $L1$, for example, the subject can receive **2** with probability 0.1, **3** with probability 0.2 and so on. When the status quo of the subject equals 5 then according to the Prospect Theory he gets the following prospects:

$$\begin{aligned} L1_{w=5} &= \{(-\mathbf{3}, 0.1); (-\mathbf{2}, 0.2); (-\mathbf{1}, 0.3); (\mathbf{0}, 0.2); (\mathbf{3}, 0.1); (\mathbf{5}, 0.1)\} \\ L2_{w=5} &= \{(-\mathbf{4}, 0.5); (-\mathbf{1}, 0.1); (\mathbf{1}, 0.1); (\mathbf{3}, 0.1); (\mathbf{4}, 0.1); (\mathbf{5}, 0.1)\} \end{aligned}$$

The values of these prospects are $PT(L1_{w=5}) = -1,55$ and $PT(L2_{w=5}) = -1,76$, so the subject prefers the lottery $L1$ than $L2$.

When the status quo of the subject equals 6 then he gets two prospects:

$$\begin{aligned} L1_{w=6} &= \{(-\mathbf{4}, 0.1); (-\mathbf{3}, 0.2); (-\mathbf{2}, 0.3); (-\mathbf{1}, 0.2); (\mathbf{2}, 0.1); (\mathbf{4}, 0.1)\} \\ L2_{w=6} &= \{(-\mathbf{5}, 0.5); (-\mathbf{2}, 0.1); (\mathbf{0}, 0.1); (\mathbf{2}, 0.1); (\mathbf{3}, 0.1); (\mathbf{4}, 0.1)\} \end{aligned}$$

The values of these prospects are $PT(L1_{w=6}) = -3,78$ and $PT(L2_{w=6}) = -3,45$, so now the subject prefers the lottery $L2$ than $L1$.

The conclusion is that the change of the reference point influences the subject's preferences.

3 The relationship between the value function and the reference point

Pratt [7] has shown that the appropriate measure of local risk aversion at a particular wealth position can be defined as

$$ARA\,(w) = \frac{-u''\,(w)}{u'\,(w)} \tag{3}$$

where $u(w)$ is a twice differentiable function in the sense of the von Neumann-Morgenstern utility function.

Pratt makes the normative observation that many decision makers are willing to pay less for insurance against a given risk with the increase of their assets. The behaviour of such decision maker is described by means of utility function for which $ARA(w)$ is decreasing. The utility functions with such property are often called the DARA function. If the decision makers' payments for insurance are independent of the wealth, the utility function is the CARA type. It means that $ARA(w) = \text{const}$. Whereas in the situation when with the increase of the wealth also the aversion toward risk increases, i.e. subject is willing to pay more for the insurance, then his $ARA(w)$ function is increasing (denoted by IARA).

Our analysis apply to the mentioned above classical measures of risk aversion. Under expected utility theory (EU) risk aversion is entirely interpreted by concavity of the utility function. The value function $v(x)$ for gains is the concave function thus the $ARA(x)$ measure can be interpreted in the same way.

Now the risk measure will concern the changes of the gains (in relation to the reference point) instead of the changes of the wealth. We will distinguish between $ARA(w)$ defined in the wealth domain (based on the utility function $u(w)$) and $ARA(x)$ defined in the gains relative to some level of wealth w, that is

$$ARA\left(x\right)=\frac{-v''\left(x\right)}{v'\left(x\right)}\qquad for\, x>0 \tag{4}$$

The risk seeking which is observed in the section of losses of the value function can be measured in the similar way. Then the appropriate transformation of the value function is analyzed, i.e. $-\lambda v_i(-x)$.

Below we assume that the reference point is valuated by means of the utility function which characterizes the subject's preferences and that with the change of the reference point, the valuation of relative gains and the strength of the risk aversion also change.

3.1 The case of constant absolute risk aversion

Let's assume that for the reference point w_i the value function $v_i(x)$ has the form of equation (1) and $u(w)$ is the subject's utility function of the CARA type, in other words

$$ARA\left(w\right)=\frac{-u''\left(w\right)}{u'\left(w\right)}=\text{const} \tag{5}$$

It means that with the increase of wealth decision maker doesn't change the amount of the risky securities in this portfolio. Therefore the strength of the risk aversion concerning the relative gain x also remains unchanged, what makes that if $w_i < w_j$ and for the reference point w_j the value function has the form of:

$$v_{w_j}(x)=\begin{cases} v_j(x) & for\, x\geq 0\\ -\lambda v_j(-x) & for\, x<0\end{cases} \tag{6}$$

Then

$$ARA_i(x)=ARA_j(x) \tag{7}$$

So we are searching for $v_j(x)$ which satisfies the above condition for $i\neq j$.

PROPOSITION 1.

For the twice differentiable function $v_i(x)$ for $x\neq 0$ and $\alpha>0$ the function $v_j(x)=\alpha\cdot v_i(x)$ satisfies the following condition:

$$\frac{-v_j''(x)}{v_j'(x)}=\frac{-v_i''(x)}{v_i'(x)} \tag{8}$$

The proof of the above equality is obvious, since $v_j'(x)=\alpha\cdot v_i'(x)$ and $v_j''(x)=\alpha\cdot v_i''(x)$, now then $ARA_i(x)=ARA_j(x)$.

In the situation when the measure of risk aversion in the domain of the reference points is constant ($ARA(w) = \text{const}$) then the value functions $v_i(x)$ makes up the family of function of the form:

$$v_{w_i}(x) = \begin{cases} \alpha \cdot v_i(x) & for\, x \geq 0 \\ -\alpha \cdot \lambda v_i(-x) & for\, x < 0 \end{cases} \tag{9}$$

where $\lambda > 1$ and $\alpha > 0$.

3.2 The case of decreasing absolute risk aversion

Now let's assume that the strength of the risk aversion decreases with the increase of the wealth possessed. This is expressed by the decreasing function of $ARA(w)$. In other words, the investment portfolio of such decision maker with DARA utility function together with the increase of wealth contains more risky securities. It means that the valuation of relative gains changes in the specific way. If only $w_j > w_i$ then for any gain x we have

$$ARA_j(x) < ARA_i(x) \tag{10}$$

PROPOSITION 2.

Given the concave, increasing and twice differentiable function $v_i(x)$ for $x \neq 0$, if the function g is twice differentiable and increasing convex function, and $v_j(x)$ is the composite concave function of g and $v_i(x)$ having the following form:

$$v_j(x) = g[v_i(x)] \tag{11}$$

then

$$ARA_j(x) < ARA_i(x) \tag{12}$$

Proof of this inequality is based on the derivative of the composite function.

$$ARA_j(x) = \frac{-v_j''(x)}{v_j'(x)} = \frac{-\left[g''\left[v_i\left(x\right)\right] \cdot \left(v_i'\left(x\right)\right)^2 + g'\left[v_i\left(x\right)\right] \cdot v_i''\left(x\right)\right]}{g'\left[v_i\left(x\right)\right] \cdot v_i'(x)} \tag{13}$$

Hence

$$ARA_j(x) = -\frac{g''\left[v_i\left(x\right)\right]}{g'\left[v_i\left(x\right)\right]} \cdot v_i'(x) + \frac{-v_i''(x)}{v_i'(x)} \tag{14}$$

what results in

$$ARA_j(x) = -\frac{g''\left[v_i\left(x\right)\right]}{g'\left[v_i\left(x\right)\right]} \cdot v_i'(x) + ARA_i(x) \tag{15}$$

From the assumption about the function g we have that $g' > 0$ (increasing), $g'' > 0$ (convex) and $v_i' > 0$. Therefore the first term of the sum is negative, thus $ARA_j(x) < ARA_i(x)$.

To sum up, if only $w_i < w_j$ then each composite concave function defined by $v_j(x) = g[v_i(x)]$ is the value function satisfying the inequality $ARA_j(x) < ARA_i(x)$.

3.3 The case of increasing absolute risk aversion

The last situation considered in the context of $ARA(w)$ properties concerns the increasing risk aversion. Along with the increase of the level of possession the decision maker is willing to pay more for the insurance against a risk and his utility function $u(w)$ is the IARA type.

So if $w_i < w_j$ then for any $x > 0$ the following inequality is fulfilled:

$$ARA_i(x) < ARA_j(x) \tag{16}$$

PROPOSITION 3.

The increasing function $v_i(x)$ is concave and twice differentiable for $x \neq 0$ and the function g is increasing, convex and twice differentiable function and $g'(x) \neq 0$. The composite function

$$v_j(x) = g^{-1}[v_i(x)] \tag{17}$$

satisfies the condition (16) and g^{-1} is the inverse function of g.

The proof is similar to the proof of the proposition 2. Noting that if (17) then $v_i(x) = g[v_j(x)]$ the inequality (16) can be proofed in the same manner .

Now the analyzed decision makers' profile toward risk (characterized by the utility function) we will illustrate for the value function proposed by Kahneman and Tversky in the form of

$$v(x) = \begin{cases} x^{0.88} & for\, x \geq 0 \\ -\lambda(-x)^{0.88} & for\, x < 0 \end{cases} \tag{18}$$

For the profile of CARA type, if $w_j > w_i$ and $a > 0$ the value function for the reference point w_j has the following form

$$v_j(x) = \begin{cases} ax^{0.88} & for\, x \geq 0 \\ -\lambda a(-x)^{0.88} & for\, x < 0 \end{cases} \tag{19}$$

For the profile of DARA type the value function for the reference point w_j can be as follows

$$v_j(x) = \begin{cases} x^{0.96} & for\, x \geq 0 \\ -\lambda(-x)^{0.96} & for\, x < 0 \end{cases} \tag{20}$$

where $g\left[v\left(x\right)\right] = \left[v\left(x\right)\right]^{1.1} = x^{0.96}$ and the function g is increasing and convex and the composition $g \circ v$ is the concave function.

For the profile of IARA type the value function for the reference point w_j can be

$$v_j(x) = \begin{cases} x^{0.8} & for\, x \geq 0 \\ -\lambda(-x)^{0.8} & for\, x < 0 \end{cases} \tag{21}$$

where $g^{-1}\left[v\left(x\right)\right] = \left[v\left(x\right)\right]^{\frac{1}{1.1}} = x^{0.8}$ and the function g^{-1} is increasing and concave.

4 Conclusion

Our goal in this paper has been to show what the effects of the assumption of the decision maker's attitude toward risk aversion are for the curvature of the graph of the value function. It is important to fully understand the Prospect Theory approach in the decision making process. We showed that the value function depends on the reference point and the properties of the risk aversion's measure.

The families of the value function we have showed do not exhaust the set of function satisfying the properties characterized in the propositions. In our examples we have showed only the functions of the same class i.e. the power functions. Further we are going to investigate the interesting dependence between the change of the reference point and the value function.

References

[1] Brooks P., Zank H. *Loss Averse Behavior.* Journal of Risk and Uncertainty, 31:3; p.301-325, 2005

[2] Davies G.B., Satchell S.E. *The behavioral components of risk aversion.* Journal of Mathematical Psychology, 51, p.1-13, 2007

[3] Gonzales R., Wu G. *On the Shape of the Probability Weighting Function.* Cognitive Psychology, 38, p.129-166, 1999

[4] Hilton R. *Risk Attitude under two alternative theories of choice under risk.* Journal of Economic Behavior and Organization, 9, p.119-136, 1988

[5] Just D.R., Wu S. *Loss Aversion and Reference Points in Contracts.* Paper presented at the Conference "Economics and Management of Risk in Agriculture and Natural Resources", 2005. http://purl.umn.edu/28727

[6] Kahneman D., Tversky A. *Prospect Theory: An Analysis of Decision Under Risk.* Econometrica, 47:2, p.263-291, 1979

[7] Pratt J.W. *Risk Aversion in the Small and in the Large.* Econometrica, 32, p.122-136, 1964

[8] Prelec D. *The Probability Weighting Function.* Econometrica, 66:3, p.497-527, 1998

[9] Schmidt U. *Reference Dependence in Cumulative Prospect Theory.* Journal of Mathematical Psychology, 47, p.122-131, 2003

[10] Schmidt U., Traub S. *An Experimental Test of Loss Aversion.* Journal of Risk and Uncertainty, 25:3; p.233-249, 2002

[11] Schmidt U., Zank H. *Risk Aversion in Cumulative Prospect Theory.* Discussion Paper 0207, School of Economic Studies, University of Manchester, 2002

[12] Schmidt U., Zank H. *What is Loss Aversion?* Journal of Risk and Uncertainty, 30:2; p.157-167, 2005

[13] Stott H.P., *Cumulative Prospect Theory's Functional Menagerie*. Journal of Risk and Uncertainty, 32, p.101-130, 2006

[14] Tversky A., Kahneman D. *Advances in Prospect Theory: Cumulative Representation of Uncertainty*. Journal of Risk and Uncertainty, 5; p.297-323, 1992

Andrej Kopčaj

SILMA, a. s.

Úvozní 5, 735 32 Rychvald

e-mail: silma@silma.cz

The Potential of Success—the More Objective Metrics of Managers' Performance

Abstract

The article deals with a critical judgement of managers performance assesment practice.The proposed Potential of success (PS), expresses a dynamic stability or the probability of positive fluctuation, both the economic results, *economic success* (ES) and the long – term performance or *the time success* (TS).

Its value can be defined by the relation : $PS = 1 - e_c$, where e_c = the total entropy. The solution for the Potential of success (PS) definition is the physical fundament of labour or $PS = A = E \cdot \eta$, where A represents the labour, E – social energy, or social capital SC and η is a coefficient of transformation of energy into labour or *knowledge capital* KC.The derivation of the total entropy comes out from the relation of probability product, $SC = 1 - e_{int}$ and $KC = 1 - e_{ext}$, where e_{int} – is internal social entropy, e_{ext} is the external knowledge entropy so the outcome is $e_c = e_{int} + e_{ext} - e_{int} \cdot e_{ext}$.

The benefit of this new managers performance methodology for diagnostic purposes lies in the link between the potential of success defined like this and other regularities of success, e.g. *Pareto law about wealth division, synergetic definition of success, streamline of changes efficiency, the level of cooperation complexity and the spiral growth of wealth* as well. The desired level of PS defines the improvement of spiral growth on the other hand the non – desired PS leads to a decadent dynamics.

The practical use is proved by measuring the real data what enables comparing the levels of management in various fields and thus take into account the impact of recession and boom, for instance.

Key words: *potential of success, managers performance, dynamic stability*

1 The dynamic stability

Increasing dynamic stability is the target orientation of all living systems anywhere in the universe. The specification of the space of dynamic stability is shown in Fig. 1 in the direction from the decreasing parameter S/S_max, where S is the real value of entropy and S_{max} is the highest value of entropy. If$S/S_{max} = 1$, such a state is defined as an equilibrium state with zero energetic gradients, i.e. death. If $S/S_{max} = 0$, this state would mean absolute stability or an unachievable state representing the victory of life over death.

There exists a certain regularity of behaviour in life according to which, with increasing distance from an equilibrium state (death), the probability of positive fluctuation (i.e. life subgoals) increases.The increasing distance from the equilibrium state = the growth of dynamic stability can be expressed as an increase in success potential $SP = 1 - x$. The lower the value of entropy x, the higher the quality of life, and there exists a higher probability that a better existence will be guaranteed and will last longer.

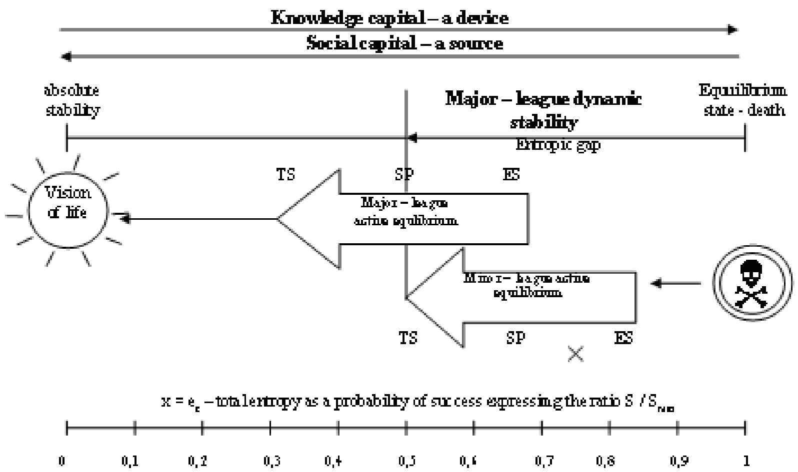


Fig1

2 Three goals

In a managerial approach, these three sub-goals of life—i.e. quality, guarantee and length of existence—represent a competitive position of SP = (quality), economic prosperity ES (guarantee), and temporal success TS (length).This is in accordance with Ilya Prigogin's three principles of existence and life development – the effort to decrease entropy (SP), to increase energy (ES) and to develop complexity (TS). The interaction of the achieved levels of SP, TS and ES creates the culture of active equilibrium.

These three life goals, demonstrated in the triangle in Fig. 2 as its vertices, are complementarily interconnected with the sides: contrast, unity and indefiniteness.

Between an increase in energy (ES) and a decrease in entropy (SP) is a contrast creating competitive turbulence. Between a decrease in entropy (SP) and a decrease in danger at time of life (TS) is a harmony creating the complexity of cooperation. Between increasing energy of economics (ES) and length of existence (TS) is a principle of certainty seeking an optimal solution between non-lasting and lasting success, which results in accelerating periodicity.

The dynamics of the rising vertices of the triangle and the rotating forces of the complementarily interconnected sides produce the resulting effect of a spiral of wealth increase. Managers focusing on the increase in dynamic stability = success potential $SP = 1-x$ are thus rewarded with an increase in both ES and TS including growing turbulence, complexity as well as a rate

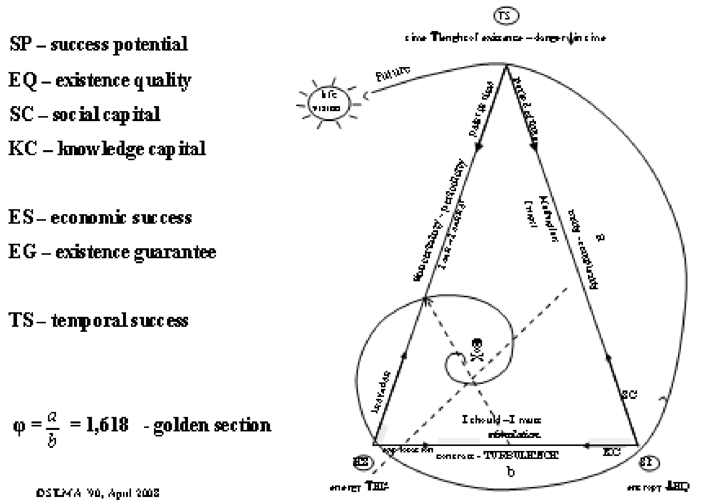


Fig.3

of cycles searching for the optimum position of the uncertainty principle. The optimum position is given by the complexity and turbulence ratio, i.e. the quotient of sides b/a, which finds the harmony of the dynamics of a spiral increase in wealth at the value of the golden section of $\varphi = 1.618$.

Fig. 3 shows the increase in ES in dependence on the increase in SP. The increase in economic success ES can be expressed as the growth of the probability of the specific positive fluctuation of the success potential increase, or: $ES = \frac{SP}{x}$. The inverted value $\frac{x}{SP}$ by contrast, expresses the growth of the probability of specific negative fluctuation, i.e. an increase in risk.

3 Dynamic stability zones

The intersection at point $x = 0.5$ divides the field of dynamic stability into the domain $x < 0.5$ where the possibility of positive fluctuation is higher than its risk, and $x > 0.5$ where the risk is greater than the effect of the positive fluctuation. Thus, the field of dynamic stability is divided into the "major league" zone which creates originals with the effects of Pareto's wealth distribution, where 20 % of the participants receive 80 % of the wealth, and the "minor league" zone, copying the these originals, however with a 80/20 wealth distribution, which means that 80 % of the participants receive only 20% of the wealth.

Economic success ES is the positive fluctuation of the success potential SP, and thus trying to create an economic growth strategy without success potential growth is both nave and entirely confusing. It would be the same as striving for higher incomes without any effort to improve one's position in the ATP ranking in tennis. The improved ATP ranking (the increase in SP) logically results in the probable increase of financial incomes (ES) in accordance with the

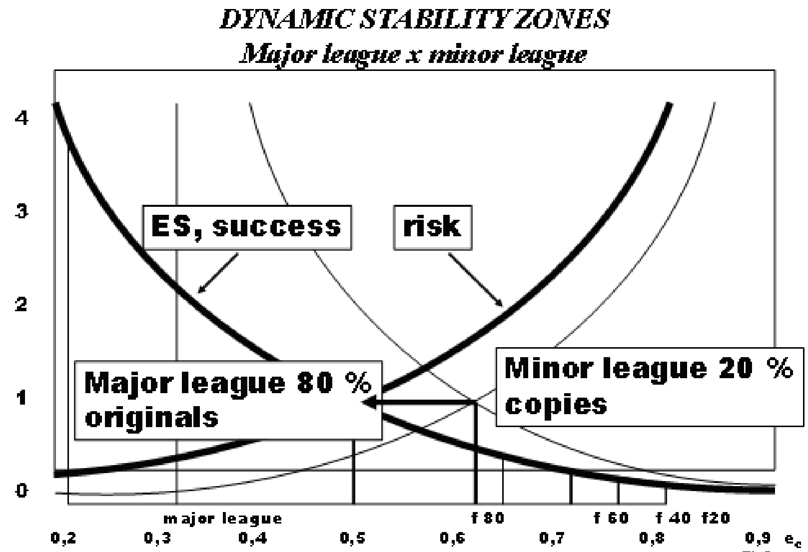


Fig.3

level of the player's improved ranking. The craving to be the best tennis player in the world will bring a better financial profit than the effort to create a maximum financial profit when playing tennis.

Similarly, the aim of conducting business must be defined as an effort to be the best among the competition in one's own business, rather than as a simplifying or vulgarizing aim of maximizing earnings.Achieving a success potential increase SP means respecting the heart of the matter, i.e. the law of the impossibility of building a perpetuum mobile, in which it holds that $A = E \cdot \eta$. Energy E stands for the level of social capital development (SC) as a source of success. The coefficient of the conversion of energy to work η is knowledge capital (KC), as a means of success.

The principle of success potential was excellently formulated by Tomas Bata, who stated: "Everything is possible (ES + TS) if you *want it* (SC) and you *know how* (KC)". The positive fluctuation of durable success results from success potential development consisting of social and knowledge potential development. The social capital source (to want) decreases the total entropy $e_c = x$, while the entropy increases for the knowledge capital in time: and that is why it has to be cyclically decreased by the source.

If we express both social capital and knowledge capital by means of the success probability $SC = 1-e_{int}$ and $KC = 1-e_{ext}$, where e_{int} stands for internal – social entropy and e_{ext} stands for external – process entropy, then $SP = (1 - e_c) = (1 - e_{int})(1 - e_{ext})$. From this relation the principal enemy of success may be expressed as follows: $x = e_c = e_{int} + e_{ext} - e_{int}.e_{ext}$.

4 The practical example

An example of the values developed in time and measured at the joint stock company Bonatrans a.s. Bohumín is given in Tab. 1. [1]

YEAR		**2001**	**2003**	**2005**	**2008**
KNOWLEDGE CAPITAL	External entropy (e_{ext})	0,38	0,40	0,38	0,32
SOCIAL CAPITAL	Internal entropy (e_{int})	0,49	0,56	0,50	0,47
SUCCESS POTENTIAL	$SP = (1 - e_C)$	0,32	0,25	0,32	0,36
Total entropy	e_C	0,68	0,75	0,68	0,64

Table 1: Bonatrans a.s. Bohumín

Measuring success potential offers an excellent opportunity to compare managerial productivity in various types of organizations. The standard of managing profit-making organizations in various sectors is thus comparable with the standard of managing hospitals, insurers, tax offices or municipal authorities while also filtering out the influence of boom or bust.

5 Time

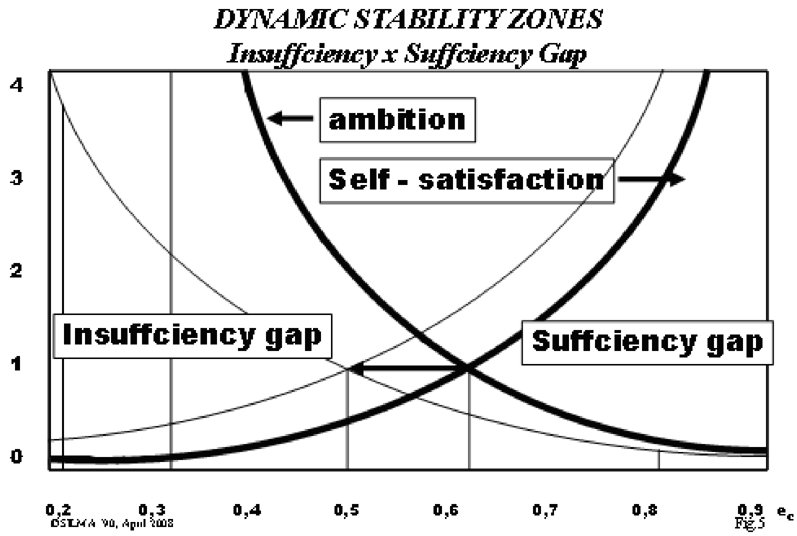
An interesting derivative (fluctuation) of success potential is temporal success (TS). Looking beyond the "bend" in the road of time enables us to move into the future more successfully—and this is the basis and the goal of all strategies. The logic for deriving TS may be deduced from Socrates' statement "the more I know, the more new and hitherto unknown facts I discover", i.e. the greater ES as a consequence of increasing SP, the broader are the horizons for TS. The first derivative of $ES = \frac{1-x}{x}$ is $-\frac{1}{x^2}$, which is the direction of the tangent to the ES. How far it is possible to see in time is thus given by the relation $TS = \frac{SP}{x^2}$, (Fig.5).

With a growing quantity of knowledge gained—especially that knowledge which was hitherto not even suspected but is newly discovered – humility also grows, generated by the insufficiency gap between the ambition to learn and the self-satisfaction resulting from the learning.

The reciprocal value $\frac{1}{TS} = \frac{x^2}{SP}$, is the level of self-satisfaction, which decreases with increasing success potential and vice versa—it increases with the decrease of potential. Only a fool thinks that he knows everything and experiences the arrogance of the sufficiency gap, compared with a giant who feels humility and a growing ambition to fight against the increasing insufficiency gap. The intersection of ambition and self-satisfaction is at the value of $x = \frac{1}{\varphi} = 0.618$

It is the value ec at which management is able to see the niche of the major league beyond the bend of time and reveals the ambition to reach the major league—which is a necessary precondition for launching a spiral increase in wealth.

The major league sees beyond the bend of time into the so-called niche of time in which the major league original is created, bringing 80 per cent of ES for 20 per cent of the participants of the major league. Thus the manager's success resides in his/her abilities to draw ideas from the niche of time of the major league (KC) and to win all people (SC) over to these ideas. The motivation energy of all people creates the inevitable condition for long-lasting wealth.



In this way, switching the evaluation of managers' performance from maximization of profit to maximization of success potential results both in increasing one's own profit and securing its better orientation to a long-lasting and ethically more acceptable profit.

6 The comparison of the past and the modern approach

MAXIMIZATION OF PROFIT X MAXIMIZATION DYNAMIC STABILITY		
INDICATOR	**TRADITIONAL ECONOMY**	**SPIRAL MANAGEMENT**
	MAXIMIZATION OF PROFIT	MAXIMIZATION OF DYNAMIC STABILITY
	MONEY	SUCCESS POTENTIAL
1. The basis of wealth expressed as		
- Economic success	increases	maximizes
- Temporal success	endangers	maximizes
- Success potential	ignores	maximizes
2. Factors of success		
- Objective *	misrepresents and decreases	is the goal of all living systems
- Source *	money more often than people	predominantly people
- Device *	employs people as a device	Profit is a device – an expense of future time periods
- Surrounding	non – ambiguity of the aritmetics of money	multiple meanings of opportunities and risk
	strongly misrepresents the influence of the boom / bust cycle	probability of success without the influence of boom / bust
- Future time (Strategy)	regularities of the time period	pegularities of the time period + randomness of the point in time
3. Universality	only profit – making organizations	all organizations
4. Application in the management process	for remuneration	for diagnostics
5. Ethics in entrepreneurship	often abuses	unses
6. Energization regime	stimulation	motivation

Notes(*):

1. Money cannot be a source, it is a device and if it is a goal, its quality is loewered.
2. People can be both a source and a device and if they are a device, they achieve better money creation..

References

[1] KOPČAJ, A., *Spirálový management*, Alfa Publishing, 2007, ISBN 978-80-86851-71-6

Norbert Kriedel

Hamburg Institute of International Economics
Heimhuder Straße 71, 20148 Hamburg

e-mail: kriedel@hwwi.org

Modelling Private Consumption in Germany— —a State Space Approach

Abstract

In Germany, the cyclical upswing that followed the New Economy recession at the beginning of the decade was mainly driven by foreign trade and to a lesser extent by private investment. Private Consumption generally did not contribute to this upswing. The most prominent explanation for this fact stresses the likewise faint development of private income and monetary wealth. Although these two factors can explain a good portion of the sluggishness, there is some rest that has to be explained by other factors. The aim of this paper is to measure this 'extra loss' in private consumption which cannot be attributed to the development of real private income and real monetary wealth since the year 2003. It can be identified by estimating a time variant absolute term within a state space model in which the consumption function acts as the measurement equation. The estimated absolute and relative amount of this 'extra loss' is very small, though significant. It has arisen mainly in the period between 2003 and 2007. There are several possible factors that might drive the course of this unobservable component. Among them, a persistence or shock effect, originating in the burst of the New Economy bubble and the following decline in the stock market value has to be seen as one important component.

Keywords: *Private Consumption, State Space Model, Stock Market Effect*

1 Introduction

From a long run perspective, private consumption in Germany has been extraordinarily weak since the year 2002. This is shown by the fact that during this period, the average growth rate of private consumption was lower than in every time span of equal length since the 1950s, with exception to the period at the beginning of the 1980s.[1] As can be seen from figure 1, the deflated consumption expenditures of private households have been more or less constant since 2002.[2] The relevant series fluctuates around a value slightly below 300 billion EUR. Looking at figure 2, which shows the year on year growth rate of the quarterly series for private consumption, three different phases can be distinguished.

Phase 1, reaching from 1992 until the second quarter of the year 1998, is characterised by a falling trend in the growth rate of private consumption. The average year on year rate in that

[1] ARGE (2006), p. 45f.

[2] All the series in this analysis were seasonally adjusted using the TRAMO SEATS programme for seasonal adjustment, deflated by the deflator for private consumption and transformed in logarithms. An exception is the unemployment rate which was only seasonally adjusted as it does not contain nominal monetary values.

time was 1,33 %. In phase 2, ranging from the third quarter of 1998 until the end of the year 2001, the falling trend of phase 1 was stopped, and the growth rates of private consumption were on average rather high, amounting to 1,97 %. At the beginning of the year 2002, a level shift has taken place, so that a third phase 3 started, in which the growth rates of private consumption were on average much lower than in phase 2. For this phase 3, the average relative change of private consumption was in fact negative (- 0,15 %).

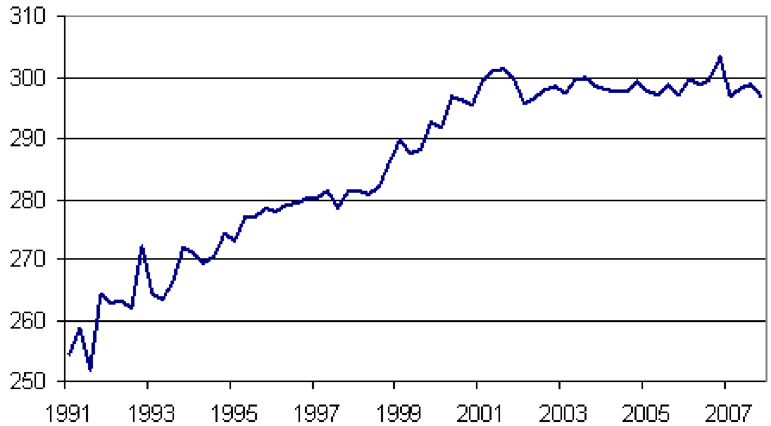


Figure 1: Real Private Consumption in Germany (Q1/1991 – Q4/2007)
Source: German Statistical Office (2008), own calculations; unit: billion EUR

According to keynesian consumption theory, private consumption is mainly dependant on real private income. The latter has also been more or less stagnant within the last years, as is shown in figure 3. Hence, from a superficial point of view, it would be tempting to explain the weakness of private consumption by the sluggishness of disposable income. However, figure 4, which presents private consumption as a percentage of disposable income, shows that the evolution of private income cannot explain fully the weakness in private consumption. This is due to the fact that the part of income that is used for consumption has fallen sharply around the year 2001 and has only recovered in recent time.

Apart from income, private consumption is also dependant upon private wealth. Wealth can be accumulated in three different ways: First, and maybe most important, by saving. Secondly, by the rise in the value of wealth. Thirdly, by transfers and donations. The importance of these three different sources for wealth can change over time.[3] The following figure 4, which

[3] Schulmeister (2004) found evidence, that from 1960 until 1985, the increase in wealth of private households in the USA was almost completely caused by saving. In contrast, the rise in the value of private wealth between 1985 and 1999 acted as the main driving force for the increase in private wealth in that time. Also the following downturn was mainly brought about by the fall in the valuation of wealth.

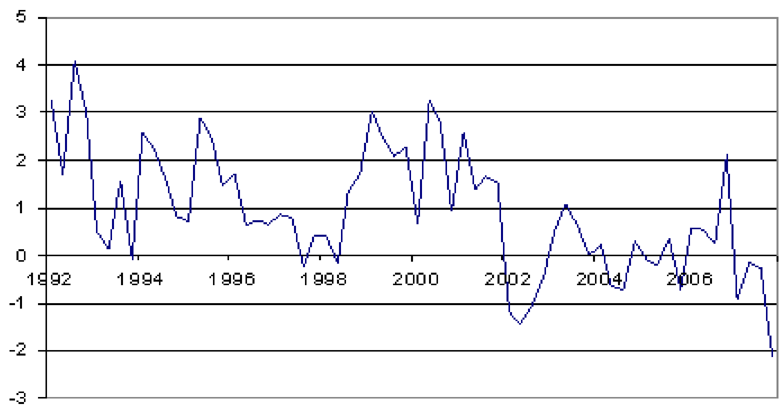


Figure 2: Growth rate of real private consumption in Germany (year on year, quarterly)
Source: German Statistical Office (2008), own calculations; unit: %

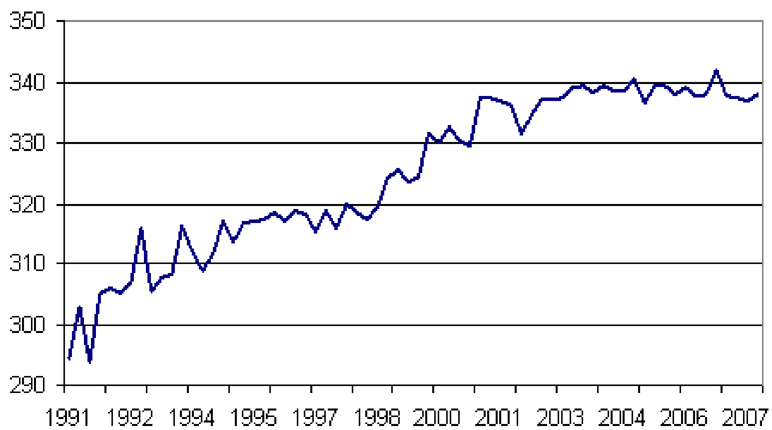


Figure 3: Real disposable income in Germany (Q1/1991 – Q4/2007)
German Statistical Office (2008), own calculations; unit: billion EUR

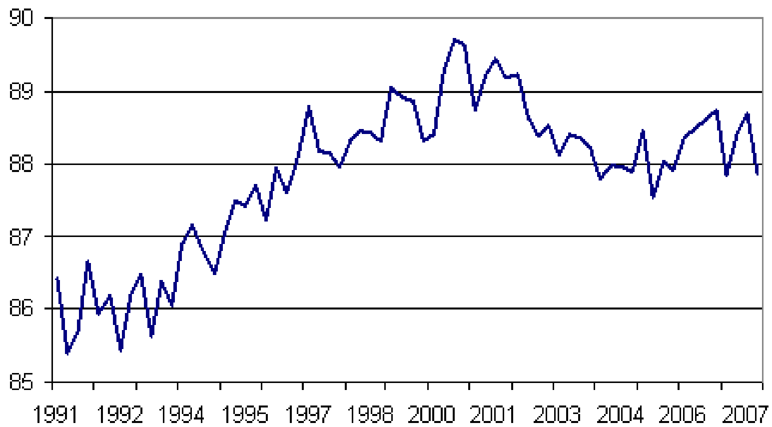


Figure 4: Private consumption as percentage of private income in Germany (Q1/1991 – Q4/2007)
Source: German Statistical Office (2008), own calculations; unit: %, deflated data

shows the series of private wealth from 1991 to 2007, indicates the occurrence of a break in the evolution of private wealth around the year 2000.

There is no doubt that this break is connected with the declining dynamics in the stock market, namely with the burglary in the course of the burst of the "New Economy" bubble at the beginning of the year 2000. Stocks are one important component of private wealth. Though the decline in the value of stocks – and therefore also in private wealth – was overcome at the beginning of the year 2003, private consumption remained sluggish afterwards. There are different possible explanations for this phenomenon. On the one hand, this could point to a kind of persistence or shock effect, which was brought about by the burst of the bubble in the year 2000, and lead to a decline of the consumption rate and an increase in the saving rate. As can be seen from figure 10, the saving rate has been rising constantly from the end of the year 2000.

One the other hand, economic literature has stressed the concept of permanent income as one possible determinant for private consumption. Following Andro and Modigliani (1964), current income is substituted by normal income, which also contains expectations concerning the future income. These expectations can be dependant upon a variety of economic, political and also psychological factors. Therefore, it might be possible to identify further economic and political trends that started around the year 2001 or intensified in that time, and lead to a declining consumption rate.

One aim of the present paper is to measure the proportion of private consumption which cannot be explained by variations in current income and current wealth. After the estimation of this proportion, different possible determinants for it will be discussed. For this purpose,

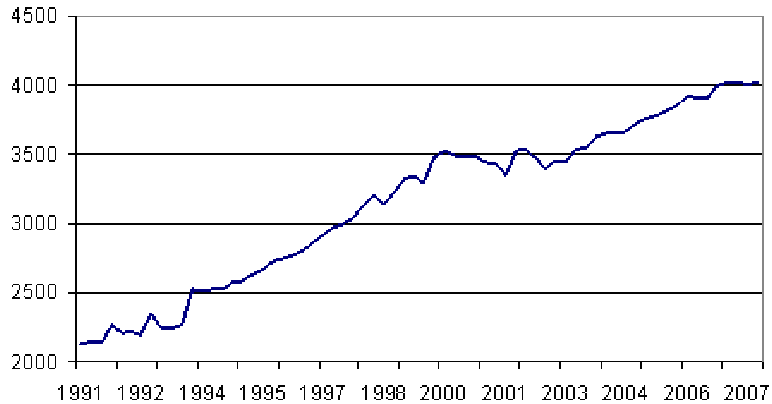


Figure 5: Real private wealth in Germany (Q1/1991 – Q4/2007)
Source: German Statistical Office (2008), unit: billion EUR, deflated data

a consumption function is estimated, with real private consumption as dependant variable and real private income and real private wealth as independent variables. In contrast to the usual specification of a consumption function, the constant or absolute term of the function will be regarded as time variant. It will be modelled as a Markov process of first order. Hence, the whole problem can be formulated as a system of two equations, representing a state space model that is containing one measurement or observation equation and one state equation. The model is estimated with the Kalman filter. The resulting estimation of the time variant absolute term allows us to draw conclusions concerning the quantitative effect that the shock or persistence effect originating in the bubble burst of the year 2000 might have had. Alternatively, also further economic and political determinants of private consumption could have played a role for the weak development of consumption. In chapter 4, the empirical evidence which can be found in favour for the shock or persistence effect as well as for other possible determinants for private consumption are discussed.

2 The state space model

If one is computing the OLS estimator for a classical linear regression model recursively, this can be regarded as an application of the Kalman filter. A regression equation can then be seen as the measurement or observation equation of a state space model. [4] The state space representation of a regression function for private consumption allows us to treat the absolute term of the consumption function as a variable. In a regression function with the two regressors private income and private wealth, a variable absolute term comprises all the influence of other determinants for private consumption that are not part of these two regressors. The general

[4] Harvey (1993), p. 98f.

form of the state space representation of a regression function is given by: [5]

$$y_t = X_t \cdot a_t + G_t \cdot \varepsilon_t \tag{1}$$

$$a_{t+1} = a_t \tag{2}$$
$$t = 1, 2, \dots, T$$

where equation (1) is the measurement equation and equation (2) the state equation. In the case of a multivariate state space model, y_t, a_t and ε_t are vectors, while X_t and G_t are matrices. The time index of G_t can be dropped when the system matrix is constant over time. In the following analysis, a univariate state space model is specified, so that y_t, at, ε_t as well as X_t and G_t are scalars.

The following linear state space model consists of two equations. First, of a regression function (3) that is identical with the measurement or observation equation of a state space model. In equation (3), private consumption C_t acts as the dependant variable of the regression function as well as the observable variable of the state space model. The observable variable C_t is related to the unobservable or state variable via the regression function. The unobservable variable is represented by the time variant absolute term a_t of the consumption function. In addition, there are two exogenous regressors (private income YV_t and private wealth W_t). Both of them enter the observation equation with a lag of one quarter. For the error term u_t the usual assumption holds, that $\{u_t\}$ is a series of independently distributed error terms with mean zero and variance σ^2.

The state equation (4), which describes the law of motion of the unobserved variable, shows the usual first-order Markov process. This is an effective way to describe the autoregressive structure of the state variable.[6] As linear state space models are most appropriate for cases where the distribution of data can be approximated by a normal distribution, it is further assumed that u_t is a normally distributed variable.

$$C_t = a_t + b_1 \cdot YV_t(-1) + b_2 \cdot W_t(-1) + u_t \tag{3}$$

$$a_t = a_{t-1} \tag{4}$$
$$u_t \sim NID(0; \sigma^2)$$

Furthermore, the specification of a state space model requires determining the mean and the variance of the initial state a_0. This is done by setting $E(a_0) = \alpha_0$ and $Var(a_0) = Q_0$.

[5] Koopman, Shephard, Doornik (1999), p. 123.
[6] Harvey (1993), p. 83; Koopman, Shephard, Doornik (1999), p. 115; Fahrmeir (1991), p. 55.

3 Results

The procedure to obtain estimations for the unobserved term a_t is as follows: First, a normal regression is run for equation (3). The estimation yields the following results (t-values are in brackets in the line under equation 5).[7]

$$\begin{array}{l} C_t{=}1,43\ {+}0,55\cdot YV_t(-1){+}0,13\cdot Wt(-1) \\ \quad (2,82)\ (4,63) \qquad\qquad (5,21) \end{array} \tag{5}$$

$$R^2 = 0,94 \quad R^2_{adj} = 0,94 \quad \text{Durbin-Watson} = 1,97$$

Breusch-Godfrey LM Test (p-values): $LM(-1) = 0,71; LM(-2) = 0,10$

The usual tests for stability (CUSUM test, One-step-forecast-test) indicate that there is no structural break present in the parameters of the regression function.[8] The Durbin-Watson-statistic and the p-values of the Breusch-Godfrey LM Test show that there is no autocorrelation of first or second order in the residuals. The estimated marginal propensity to consume amounts to 0,55. The marginal effect of an increase of private wealth is 0,13.[9] The remaining changes of private consumption which are not explained by these two regressors are caught by the absolute term at.

Looking at figure 11 in the appendix, it gets clear that estimating a time variant absolute term can be of great interest. As can be seen from figure 11, there is a negative trend in the residuals starting approximately in the year 2002 and reaching to the end of the year 2007. This means that a constant absolute term cannot capture the development of private consumption over the whole period of time under study. Therefore, a time variant absolute term has to be estimated.

In a further step, equation (3) together with equation (4) is estimated by applying the Kalman filter. The calculated values for the marginal propensity to consume and the marginal wealth effect are the same as in the regression analysis (5). The following equation gives the results of the Kalman filter estimation.[10] The z-Statistics are in parentheses in the next line under equation (6). The final value of the state variable is given below, together with some details of the state space estimation.

$$\begin{array}{l} C_t = a_t{+}0,54\cdot YV_t(-1){+}0,13\cdot W_t(-1) \\ \qquad\qquad (5,06) \qquad\qquad (5,61) \end{array} \tag{6}$$

[7] In the appendix, figure 11 shows the empirical series of private consumption together with the fitted values and the residuals.

[8] Figures showing the results of CUSUM and One-step-forecast-tests are given in the appendix.

[9] Similar values were obtained in recent estimations of the private consumption function for Germany. Ripp and Schulze (2004) estimate a consumption function for the period 1991 – 2002. Their estimation for the marginal propensity to consume is 0,68, and for the marginal effect of wealth 0,185. An analysis of the long run consumption function for the period 1952 - 2005 [ARGE (2006), p. 49] results in a value of 0,50 for the short run marginal propensity to consume, and a value of 0,03 for the marginal effect of an increase of private wealth.

[10] For the estimation, the Maximum-Likelihood method (Marquardt) that is implemented in EViews has been applied.

Final State Value: 1,48 (998,68)   Root Mean Square Error = 0,001

Log Likelihood: 188,02   Convergence achieved after 8 iterations

The calculated series for the state variable is depicted in the following figure 5. The time variant term a_t contains all other determinants for private consumption besides income and wealth. Figure 6 shows that a_t has increased from 1992 until 1997 and has remained stable in the year 1998. Afterwards it has risen again until the third quarter of the year 2003. Since that time there has been a continuous decline. In order to explain the development of the term at, all possible factors (besides current income and current wealth) which have an influence on private consumption have to be considered. This will be done in the next chapter. Before, the quantitative effect of the time variant term a_t shall be analyzed. The aim is to calculate the loss of private consumption brought about by the decrease of the term at. The first quarter showing a decrease is the third quarter of the year 2003.

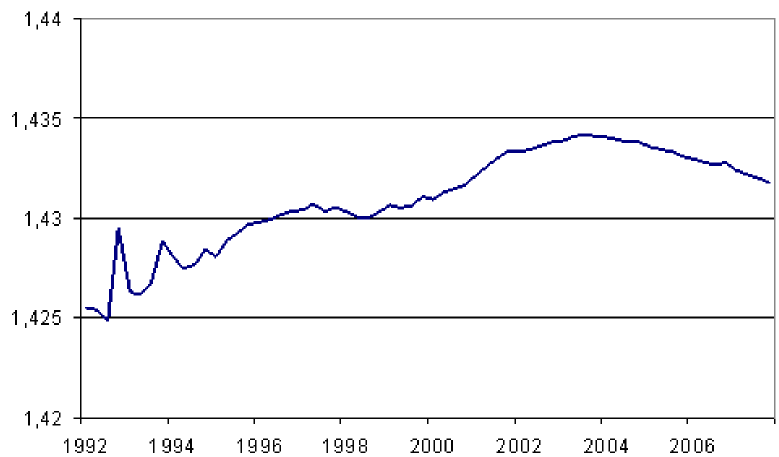


Figure 6: Kalman filter estimation for the time variant absolute term of the consumption function (billion EUR)
Source: own calculations

Figure 7 shows the extra loss of private consumption for each quarter since the fourth quarter 2003.[11] This loss is due to the decrease of the absolute term at. Thus it can be attributed to all other influencing factors for private consumption besides current income and current wealth.

The total loss of consumption (sum of losses per quarter) that is attributed to the reduction of the absolute term a_t (fourth quarter 2003 – fourth quarter 2007) amounts to 10,0 million

[11] In order to calculate this loss per quarter, the estimated values for a_t which are shown in figure 5, had to be delogarithmised. This had to be done as all variables entered the regression function and the state space model in logarithmic terms.

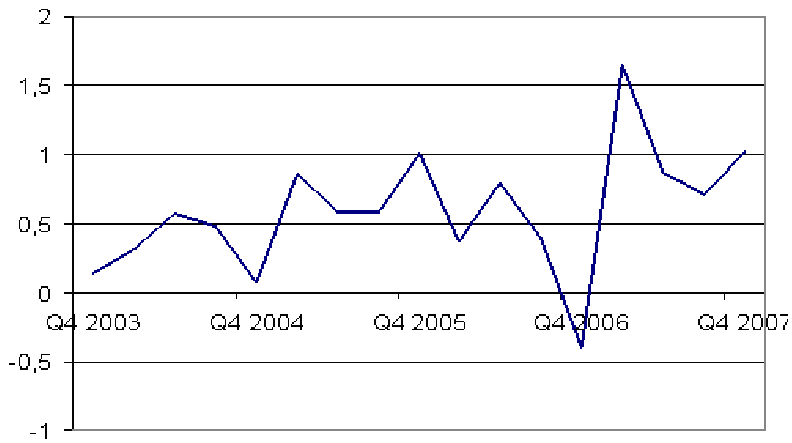


Figure 7: Loss of private consumption per quarter attributed to the absolute term (million EUR)
Source: own calculations

EUR. The total net reduction in real private consumption in that time amounts to 2,97 billion EUR.[12] This would imply that only 0,33 % of the net reduction of private consumption which took place since the fourth quarter 2003 can be attributed to other factors than private income and private wealth. In the following chapter some of those further possible determinants shall be discussed.

4 Possible determinants for the time variant absolute term

Apart from current income and wealth, economic literature has stressed the concept of permanent income as one possible determinant for private consumption. Following Andro and Modigliani (1964), current income is substituted by normal income, which also contains expectations concerning the future income.[13] It can be supposed that the expected future income of a person is influenced by macroeconomic indicators like the unemployment or a country's growth rate of aggregate production. In the following, a variety of possible factors that could be seen as determinants of permanent income are discussed and viewed from an empirical point of view for Germany.

4.1 Persistent unemployment

A high rate of unemployment can increase the (perceived) risk of losing the job among employed persons. This could in turn lower their consumption expenditures, as their expected future income decreases. A similar effect could result from a persistent weak economic situation.

[12] This number was calculated by adding the reductions in private consumption per quarter and subtracting the increases per quarter whenever those occurred.

[13] Andro, A.; Modigiliani, F. (1963)

The perception of low growth rates of aggregate production could be regarded as the result of low productivity growth. Therefore, consumers could expect a lower increase in wages for the future, and in turn lower consumption possibilities.

The rate of unemployment in Germany has undergone rather intense changes, as can be seen from figure 8. Parts of these fluctuations can be attributed to institutional changes concerning the definition of unemployed persons, etc. However, the sharp increase in unemployment that took place from 2001 until 2005 must also be seen in the light of weak economic conditions which were present after the financial crisis of the year 2000. As the decline of the absolute term a_t started in the fourth quarter of the year 2003, it would be tempting to attribute this decline in consumption to the pessimistic sentiments of consumer due to high unemployment and low growth.

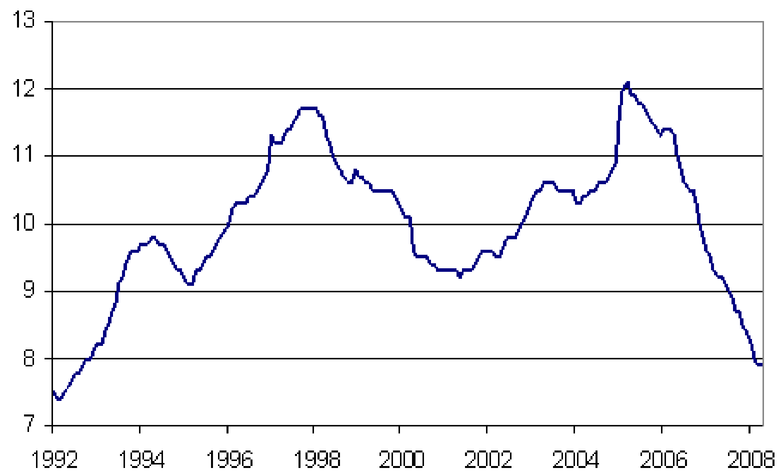


Figure 8: Rate of unemployment in Germany (%)Source: German Central Bank (2008)

4.2 Shock effect of declining dynamics in the stock market

The burst of the bubble in the New Economy was followed by a dramatic fall in the stock market's index DAX, which lasted until the end of the year 2002. This fall in the DAX index was accompanied by huge losses of wealth on behalf of the consumers. In addition, it can be supposed that the dramatic loss of wealth within a very short period of time has represented a shock that decreased the willingness to consume persistently. An argument that supports this view is the fact that although private wealth regained dynamics in 2003, private consumption still remained sluggish. This can be seen from figure 1 which shows that absolute private consumption more or less stagnated from the year 2002 onwards. An argument against this thesis concerns the property structure. The importance of shareholdings within finance property is substantially lower in Germany than for example in the USA where in 2002 15,2%

of the whole finance property of private households was held in the form of stocks, compared to only 5,2 % in Germany.[14]

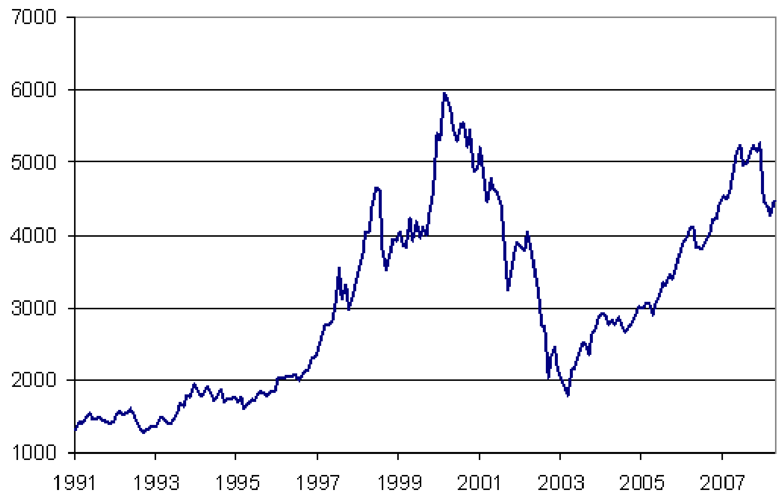


Figure 9: German stock market index DAX (1/1991 – 4/2008)
Source: German Bundesbank (2008), monthly data; 1987 = 1000

4.3 Rising life expectancy and uncertain pension systems

The rise of the statistical life expectancy requires that an individual has to build up more reserves for a longer life time. This is apparently becoming more important regarding the background of increasing uncertainties within the legal pension system. Both factors would explain a rise of the savings rate and therefore also a sinking proportion of income used for private consumption. As can be seen from figure 10, the saving rate has increased constantly since the end of the year 2000.

However, the history of reforms in the german pension system is characterised by a more or less continuous discussion rather than a unique event that might have changed the attitudes of consumers at once. Therefore, the particular rise in the saving rate (and the corresponding fall in the consumption rate) might not directly be related to the problem of an increased uncertainty in pension systems.

4.4 Increasing income disparity and heterogeneous consumption rates

Every five years, the German Statistical Office is conducting a survey about income and expenditures of private households. Within this survey, the volume and structure of income and consumption expenditures is collected in detail for a random sample of consumers. According

[14] Schulmeister, S. (2004), p. 30 and p. 87.

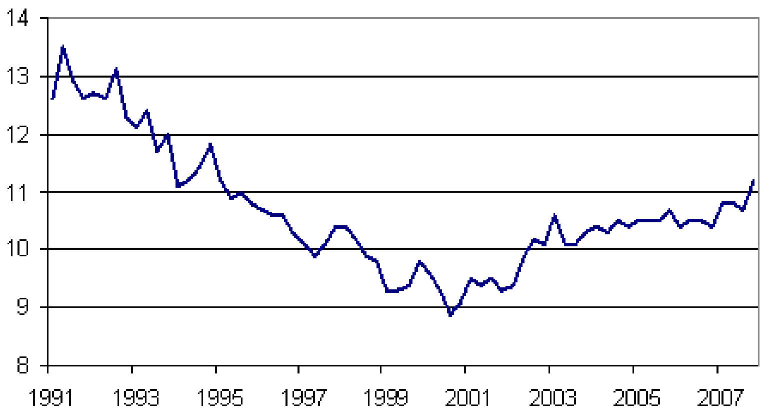


Figure 10: Saving rate in Germany in percent of disposable income
Source: German Statistical Office (2008)

to the last survey which was conducted in the year 2003, the proportion of income used for consumption is negatively dependant upon the level of the net income of a person.[15] This means that the higher the net income of a person, the lower is the proportion of income that is spent on consumption. Against the background of an increase in the disparity of incomes, this consumption pattern can contribute to lower consumption expenditures in total. This is because a rising disparity in personal incomes increases the proportion of people with very high incomes, low consumption and high saving rates. Evidence for a rising disparity in personal incomes in Germany can be found from Goebel and Krause (2007). According to their analysis based on the Socio-Economic Panel, they find evidence that income disparity has increased since the year 2000, while it had been more or less stable in the 1990s.

4.5 Increase in other issues not relevant for consumption

According to the survey about income and expenditures of private households of the German Statistical Office, the proportion of expenditures for purposes that are not relevant for consumption has increased between 1998 und 2003 from 10,8 % to 13,4 %. These purposes include:[16]

- Contributions for private insurances (health, pension, car, furniture, liabilitiy, accident insurances, etc.)
- Interest rates for loans
- Car taxes, capital transfer taxes etc.

[15] Kott, K. (2005), p. 1313.
[16] Kott, K. (2005), p. 1313.

The most important part of these expenditures are contributions for private insurances (35%), followed by interest rate payments for construction loans (24%). On average, these expenditures not relevant for consumption have increased from 288 EUR in the year 1998 to 386 EUR in the year 2003. This amounts to a relative increase of 34 %. With respect to different cohorts, empirical evidence shows that people aged between 35 and 55 spend the highest proportion of their income for non consumption relevant purposes, namely 14,5 %. At the same time, this age group has the highest saving rate and the lowest consumption rate within all age groups. Furthermore, the people aged from 35 to 55 years have the highest net incomes per household and month in comparison to the remaining age cohorts.[17] In other words: Those age cohorts, that possess the highest net incomes exhibit the lowest consumption rates, the highest saving rates and expend the highest proportion of their income for non consumption relevant purposes. This pattern may be partly explained by the need to create financial reserves for the old age, which can be of high relevance for the age group 35-55.

5 Summary and conclusions

A good deal of discussion has arisen because of the sluggishness of private consumption in Germany during the last years. The present analysis wants to shed some light on possible explanations for this stagnation. Statistical evidence shows that by far the largest part of the phenomenon can be explained by the weak development of real disposable income as well as real private monetary wealth. Especially private income, which contributes to a large part to private consumption, has been more or less stagnant since the beginning of the year 2003. However, there are two features that cannot be explained solely by the course of disposable income and monetary wealth. One feature concerns the rebound of monetary wealth since the year 2003, which was not accompanied by an adequate increase in private consumption. Secondly, the proportion of consumption to private income has fallen quite sharply since the beginning of the year 2001. Therefore, one aim of this paper was to estimate how much other factors besides income and wealth contributed to the development of private consumption.

There seems to be evidence that only a small part of the weakness in consumption can be explained by those additional factors. Their contribution amounts to less than one percent of the total net reduction of private consumption in the time from the fourth quarter of 2003 till the fourth quarter 2007.

One prominent explanation for these additional factors is based on the idea of a shock effect that originated in the burst of the stock market bubble of the year 2000. Another possible explanation stresses the possible fall in permanent income that was brought about by weak economic conditions, namely the high unemployment rate in the years 2002 until 2005. This might have dampened income expectations and depressed the willingness to consume. Furthermore, the need to save a higher proportion of income because of the rising life expectancy and uncertain pension systems (precautionary saving) might have also played a certain role. Finally the increase in income disparity and rising expenditures for purposes not relevant for consumption can explain a decline in consumption rates. It is difficult to disentangle the importance of all these effects. However, the dramatic decline of the stock market index during

[17] Households, where the head of household was aged between 35 and 55 years in 2003, earned the highest net incomes and had the lowest consumption rates. See: Kott, K. (2005), p. 1314f.

the years 2000 until 2002 certainly must have provoked quite strong shock effects on behalf of the consumers relying on monetary wealth. Because all these factors have reduced the expected permanent income quite clearly, it seems rather astonishing that the effect they had on consumption was rather small.

References

[1] Andro, Albert; Modigliani, Franco (1964), The 'Life Cycle' Hypothesis of Saving: Aggregate Implications and Tests, in: The American Economic Review, Vol. 53, No.1, Part 1, pp. 55-84

[2] ARGE (2006), Die Lage der Weltwirtschaft und der deutschen Wirtschaft im Frühjahr 2006, HWWA Report 261, Hamburg Institute of International Economics 2006

[3] Fahrmeir, Ludwig (1991), Zustandsraummodelle: Filtern, Glätten und Prognose dynamischer Systeme, in: Allgemeines Statistisches Archiv / Journal of the German Statistical Association, Vol. 75 (1991),1, p. 53-74

[4] Goebel, Jan; Krause, P. (2007), Gestiegene Einkommensungleichheit in Deutschland, in: Wirtschaftsdienst Vol. 87/12, p. 824 – 832

[5] Harvey, Andrew C. (1993), Time Series Models; Harvester Wheatsheaf, New York - London – Toronto et. al., 2nd ed.

[6] Koopman, Siem Jan; Shephard, Neil G.; Doornik, Jurgen A. (1999), Statistical Algorithms for Models in State Space Using SsfPack 2.2, in: The Econometrics Journal Vol. 2/1999, 1, p. 107-160

[7] Kott, Kristina (2005): Einnahmen und Ausgaben privater Haushalte – Jahresergebnisse der Einkommens- und Verbrauchsstichprobe 2003, in: Wirtschaft und Statistik Vol. 12/2005, p. 1309-1323

[8] Ripp, Kristina; Schulze, Peter M. (2004), Konsum und Vermögen: eine quantitative Analyse für Deutschland; Working Paper Nr. 29 of the Institute of Statistics and Econometrics of the Johannes-Gutenberg-University Mainz

[9] Schulmeister, Stephan (2004): Aktienkursdynamik und privater Konsum in den USA und in Deutschland; Österreichisches Institut für Wirtschaftsforschung (WIFO), Wien 2004

[10] Velde, François R. (2006), An Alternative Measure of Inflation, in: Economic Perspectives, Research Dept. of the Federal Reserve Bank of Chicago, Vol. 30/1 (2006), p. 55-65

Appendix

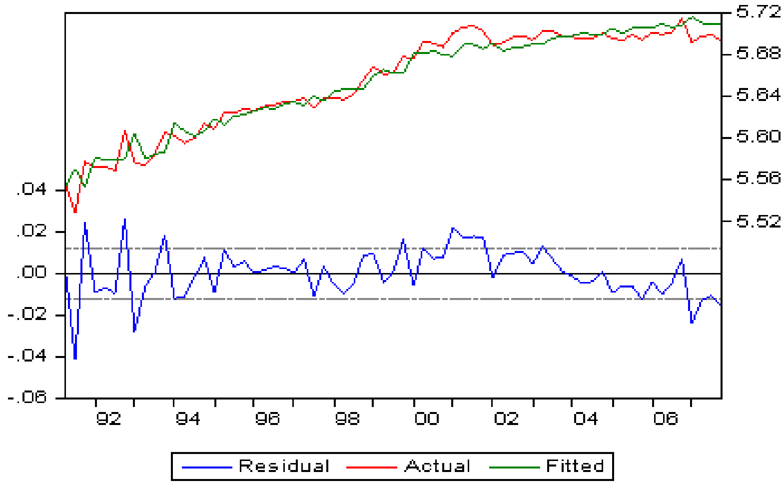


Figure 11: Private consumption, fitted series and residuals for regression function(5)
Source: own calculations

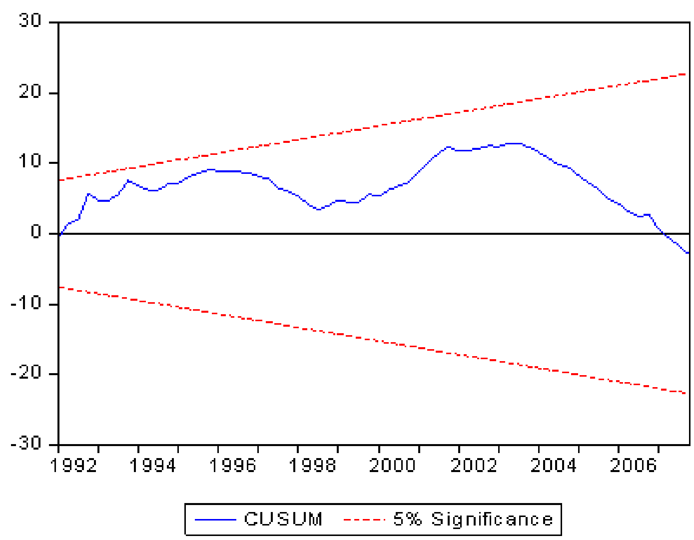


Figure 12: CUSUM-Test for regression function(5)
Source: own calculations

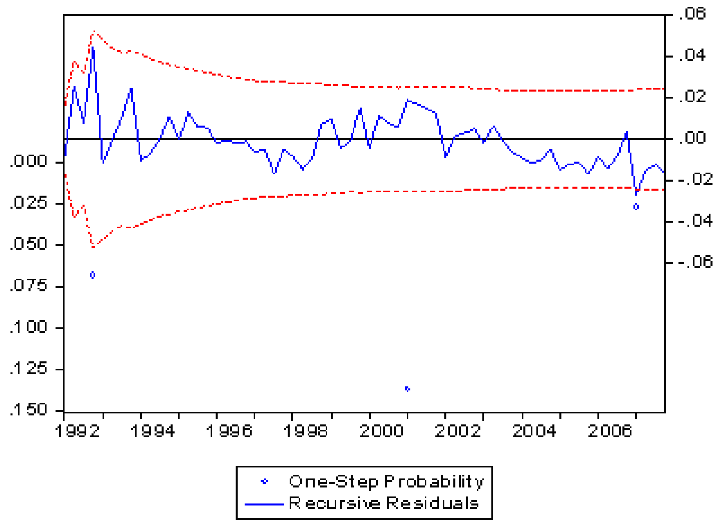


Figure 13: One-step-forecast-Test for regression function (5)
Source: own calculations

Petr Kučera, Milan Houška,
Martina Houšková Beránková

Department of Operational and Systems Analysis
Faculty of Economics and Management
Czech University of Agriculture in Prague
Kamýcká 129, 165 21 Praha 6 - Suchdol

e-mail: kucera@pef.czu.cz, houska@pef.czu.cz, berankova@pef.czu.cz

Selected Methods for the Time Limited Vehicle Routing Problem

Abstract

The time limited vehicle routing problem (TLVRP) is a special kind of the vehicle routing problem (VRP) where the routes are paths (instead of cycles as usual), i.e. vehicles do not return to the central city. Costs are given for straight routes between each pair of cities and represent time necessary for going through. Each path must not exceed a given time limit. The sum of time for all routes is to be minimized.

In this contribution two following methods are tested on several instances of various size:

The former one is the tree approach. It is based on combining the Mayer method for the VRP and the Christofides method for the traveling salesman problem (TSP). The Mayer method, choosing cities for single routes, properly constructs a spanning tree for all cities on each route and these trees can be utilized in the first step of the Christofides method. In addition, the Christofides method achieves the best approximation ratio among all methods for the TSP.

The latter method is based on approach using the Habr frequencies, which are values assigned to single edges (straight routes between couples of the cities), specifying a given edge in comparison with all edges non-incident with this edge.

Keywords : *Time limited vehicle routing problem, vehicle routing problem, traveling salesman problem, Mayer method, Habr frequencies.*

1 Introduction

The time limited vehicle routing problem (TLVRP) is defined as follows: One central city and other n (ordinary) cities are given and for each pair of cities a cost is given, representing time necessary for going through the straight route between them. The cost matrix is supposed to be symmetric. The goal is to find a set of paths so that each of them has one of its endpoints in the central city, its length does not exceed a given time limit and each city except for the central one lies on exactly one of the paths.

This problem has many practical instances, e.g. transportation newspapers from the printers' to shops, grocery products (dumplings etc.) from the manufactory to restaurants, daily reports from affiliated branches to the headquarters etc. Each vehicle is required to visit all the cities on its route until a given time, but we do not mind how it gets from the end back to the start of its route to realize it next time.

Nevertheless, TLVRP has not been sufficiently studied. It belongs to the NP-hard problems, for which there is no efficient algorithm finding their theoretical optimum. It is related to the classical vehicle routing problem (VRP), where routes are cycles instead of paths, and so to the traveling salesman problem (TSP), i.e. the task to construct one cyclic route containing all the cities, too. Thus heuristics (approximation methods) for the TLVRP can be derived from the methods used for solving the VRP and the TSP.

Let us introduce some notation. The central city will be indexed by 0 and the other cities by numbers from 1 to n. The cost matrix will be denoted by $\mathbf{C}$ (and so single costs $c_{ij},\ i,j = 0,\dots,n$).

2 Tree Approach

For the VRP, more preciously for its special case where capacities (demands) of cities and vehicles are specified (so called multipletours traveling salesman problem, cf. e.g. [5]), the Mayer method is often used. It only separates cities into groups so that each group contains cities on a route for one vehicle and then some of the methods for the TSP must be used to determine the order of the cities on the single cycles.

In fact, this method provides more outputs than only groups of the cities for single routes. It works in the same way as the Prim (Jarník) algorithm starting in the currently remotest city from the central one and finishing a group when capacities do not enable to add another city to the route. Thus it finds for each group a minimum cost spanning tree.

A well-known Christofides method for the TSP [1] goes out just from the minimum spanning tree. It adds to this tree a minimum cost matching of its odd degree vertices and in this way an Eulerian graph is obtained. Then the Eulerian walk is found in it and it is transformed into a cyclic route by preserving the first occurrence of each city and deleting all other occurrences.

These two methods fit together to design a method for the TLVRP. Its detailed description is here:

1. Choose from the cities, which have not been put onto any route, such a city i so that c_{0j} is maximum possible (i.e. the remotest city from the central one) as the first city of the currently created route. Set T to the graph consisting of the only vertex i.

2. Let P denote the set of non-central cities in the currently created route and Q the set of non-central cities, which have not been put onto any route

   If $Q = \emptyset$

   <u>then</u> stop

   <u>otherwise</u> find the minimum c_{ij} so that $i \in P$ and $j \in Q$.

3. Add the vertex j and the edge $\{i, j\}$ to T. Derive T_0 from T by adding the vertex 0 and the edge $\{0, k\}$ such that $k \in P \bigcup \{j\}$ and c_{0k} is minimum possible. (Note that both T and T_0 are trees).

4. Find the minimum cost "nearly perfect" matching M of all such cities of T that have odd degree in T_0. Obviously, there is odd number of such cities, so exactly one of them, say l, is to be isolated in this matching.

5. Now there are exactly two vertices of odd degree in $T \bigcup M$: l and 0. Find a walk from l to 0 containing all the edges of $T \bigcup M$.

6. Create a path by preserving the first occurrence of each city in the walk and deleting all other occurrences.

7. If the total cost (the time for going through) of this path does not exceed the time limit given in the input of this task

   then declare the path from the last executing of the step 6 to be the currently created route and go to 2 to search for other cities for this route

   otherwise declare the path obtained *before* the last executing of the steps 2 to 6 to be a finished route and go to 1 to start creating a new path.

Let us remark that all paths are supposed to contain at least two non-central cities in this formulation of the algorithm.

3 Habr Frequencies Approach

Habr frequency for the edge (a straight route between two cities) is the value

$$F_{ij} = \sum_{k=1}^{n} \sum_{l=1}^{n} (c_{ij} + c_{kl} - c_{il} - c_{kj}).$$

This form obviously shows its sense of comparing this edge with the others. Habr (author of e.g. [2], however, in Czech only) applied these frequencies in approximation methods for different transportation problems.

There exists another form, called modified frequency, more suitable for computations: $F'_{ij} = c_{ij} - r_i - s_j$, where r_i and s_j are the arithmetic means of the costs of i-th row and j-th column of $\mathbf{C}$, respectively. F'_{ij} can be derived from F_{ij} by linear transformation.

Habr frequencies consider all edges with the same importance. But in the case of the TLVRP the edges incident to the central city are more important (more frequently and often used) than the others. Now we show how big this difference is: Let us suppose that the solution will consist of p paths (p vehicles will be used). Let us consider a randomly chosen (with a uniform probability distribution, without respect to the costs) solution with p cycles. The probability of the choice of an edge non-incident to the central city is $\frac{2(n-p)}{n(n-1)}$ while for the edges incident to the central city this probability is equal to p/n. So the edges incident to the central city are $\frac{p(n-1)}{2(n-p)}$-times more important than the others (they occur in the solution with $\frac{p(n-1)}{2(n-p)}$-times bigger probability). Thus the frequencies for the TLVRP will be computed by the formula

$$F_{ij} = \sum_{k=1}^{n} \sum_{l=1}^{n} (c_{ij} + c_{kl} - c_{il} - c_{kj}) + \frac{p(n-1)}{2(n-p)} \sum_{m=1}^{n} (2c_{ij} + c_{m0} - c_{i0} - c_{mj} + c_{0m} - c_{im} - c_{0i}) \quad (1)$$

or modified frequencies F'_{ij} can be computed by a formula derived from (1) by an analogous linear transformation as in the general case above, which we do not mention here.

The algorithm of the Habr frequencies approach for the TLVRP looks as follows:

1. For all pairs of the non-central cities (i, j) compute the frequencies according to (1).
2. Process edges according to the ascending order of the frequencies. If after adding this edge all the edges so far added form the set of vertex disjoint paths and for each path the sum of the costs (times) of its edges does not exceed the time limit given in the input of the task then add this edge to the solution (else reject it definitely). Repeat this until each city lies on some of the paths and adding the central city 0 to the arbitrary path the time limit is exceeded.
3. In the end join the central city 0 to closer ends of all the paths.

4 Test Computations and Their Results

For testing two types of randomly generated cases were taken. In both the types all the cities were located in a circle with 100 time unit diameter (the time necessary for going through a given route is supposed to be directly proportional to the distance) and the central city was in the middle of this circle. The time limit for routes was set to 250.

In the former type, first, the area between a circle given above and another circle with the diameter of 20 time units with the same center is considered. In this area 20 cities are originally randomly generated with the uniform distribution. Then the closest pairs of cities are joined into "regions" so that at most four of them may form one "region" and the final number of non-central cities ("regions") is 12. (The same type of test cases was used in [3] and [4].)

The latter type of test cases contains 24 (non-central) cities randomly generated with the uniform distribution with no additional conditions or modifications.

As far as the first type of test cases is concerned, ten of them were computed using both the tree approach and the Habr frequencies approach. The results are summarized in the table metricconverterProductID1 in1 in the percentage form, where 100 p.c. is the result of the Habr frequencies approach, which always gave a better result than the tree approach. However, it is interesting to notice the difference between the results by these methods at single cases. In the case 4 the tree approach gave relatively good result comparable with the Habr frequencies approach while in the cases 5, 7 and 9 it completely failed. Thus some properties, which would have influence on the quality of the tree approach solution, were searched for. It was found that there is an important dependence of the results on the ratio between the farthest and the nearest city from the central one among cities lying on the convex hull of the set of all cities. Big values of this ratio properly indicate that the central city does not lie near the middle of the actually attended region. This property will be called eccentricity and it is added to the table 1.

Only three cases of the latter type were generated and tested to show that the tree approach is not suitable for so large cases in comparison to the Habr frequencies approach. In all the cases it gave at least by 10 p.c. worse result and in one of them even by more than 40 p.c. (cf. table 2).

Case	Tree approach	Habr	excentr.
1	109,9%	100,0%	1,684518
2	108,5%	100,0%	1,714486
3	105,6%	100,0%	1,763561
4	104,4%	100,0%	2,10142
5	127,7%	100,0%	1,264336
6	112,4%	100,0%	1,304033
7	117,6%	100,0%	1,324511
8	104,5%	100,0%	1,521423
9	130,3%	100,0%	1,241651
10	107,1%	100,0%	1,657835

Table 1: Test case results – the former type

Case	Tree approach	Habr
1	142,2%	100,0%
2	112,8%	100,0%
3	114,6%	100,0%

Table 2: Test case results – the latter type

5 Conclusions

In all the test cases the Habr frequencies approach gave a better solution than the tree approach. But for small tasks with high eccentricity the tree approach is also useful. It provides solutions practically of the same quality, only by several p.c. worse. In addition, in all the cases these two methods have found different solutions and so the user can choose between them according to eventual additional criteria. Unfortunately, for larger tasks the tree approach has shown to be by far not as suitable as the Habr frequencies approach.

The computation of the Habr frequencies is relatively complicated and so it was possible to expect that their utilization would bring very good results. Possibly, in comparison with some of other methods the tree approach would manage better than it was shown here.

References

[1] Christofides, N. *Worst-case Analysis of a New Heuristic for the Traveling Salesman Problem.* Technical Report CS-93-13, Carnegie Mellon University, 1976

[2] HABR, J. *Jednoduché optimalizační metody pro ekonomickou praxi.* Prague, SNTL, 1964

[3] KUČERA, P. *Tree Approach to the Time Bounded Transportation Problem.* Mathematical Methods in Economics, Hradec Králové, UHK, 2005, pp. 217 – 220

[4] KUČERA, P., DÖMEOVÁ, L. *Comparing Various Heuristics for Special Types of the Vehicle Routing Problem.* Increasing Competitiveness or Regional, National and International Markets, Ostrava, VŠB-TU, 2007, 8p.

[5] KUČERA, P., DÖMEOVÁ, L. *Solution of the Multipletours Traveling Salesman Problem with Vehicle Capacity Preference.* Mathematical Methods in Economics, Prague, CULS, 2003, pp. 171 – 176

Alexandr Kuchynka

Department of Statistics and Operations Research
Faculty of Economic
University of West Bohemia in Pilsen
Husova 11, 306 14 Plze

e-mail: alexk@kso.zcu.cz

Applications of Stochastic Volatility Models

Abstract

This contribution focuses on the modelling of volatility of returns in Czech and country-regionplaceUS stock markets using a two-factor stochastic volatility model, i.e. the volatility process is modeled as a superposition of two autoregressive processes. As the volatility is not observable, the logarithm of the daily range is employed as the proxy. The log range exhibits several desirable properties: it is more efficient due to lower variance of the measurement errors relative to log absolute returns, and moreover, its distribution is known to be nearly Gaussian. The estimation of parameters and volatility extraction are performed using the Kalman filter. We have obtained a meaningful decomposition of the volatility process into one highly persistent factor and another quickly mean-reverting factor. Moreover, we have shown that although the overall *level* of the volatility of returns is roughly the same in both markets, the US market exhibits substantially lower *volatility* of the volatility process.

Keywords: *volatility, stochastic volatility models, Kalman filter*

Acknowledgement

This research has been supported by Czech Science Foundation GACR (grant nr. 402/07/0465) and Ministry of Education, Youth and Sports of the Czech Republic (Center for Basic Research in Dynamic Economics and Econometrics, LC06075).

1 Introduction

The fact that volatility is not constant in most financial time series such as stock prices and exchange rates is widely recognised. From a practical point of view, estimating volatility is important in several fields within finance such as option pricing, portfolio optimization or risk management. There exist two prominent approaches to deal with time-dependent variances: GARCH and stochastic volatility (SV) approaches. The GARCH model ([6], [4]) focuses on capturing the clustering of volatility in returns when the conditional variance at time t is modelled as a deterministic function of lagged values of conditional variances and squared returns. On the other hand, the stochastic volatility models understand the time-varying variance as a stochastic process which can be a continuous-time diffusion [11] or a more general Lvy process [3]. For econometric purposes, it is convenient to work with some discretized version of the model (pioneered by the seminal paper [16]).

This paper follows the approach proposed in [1] and employs the decomposition of the volatility process into a sum of two independent components. We investigate the usefulness of this model in the sample of three Czech and three placecountry-regionUS stocks.

2 Two-factor model of stochastic volatility

Stochastic volatility models can be conveniently written in the state space form. Following [1] we will focus on two-factor model for the logarithm of the daily volatility h_t:

$$h_t = \bar{h} + h_{1t} + h_{2t} \quad \text{(1a)}$$
$$h_{1t} = \rho_1 h_{1t-1} + \eta_{1t} \quad \text{(1b)}$$
$$h_{2t} = \rho_2 h_{2t-1} + \eta_{2t} \quad \text{(1c)}$$
$$R_t = b + h_t + \varepsilon_t \quad \text{(1d)}$$

Equation (1a) describes the log volatility as a sum of two component processes h_{1t}, h_{2t} and a common mean $\bar{h}$. Equations (1b) and (1c) are transition equations which specify the dynamics of the latent variables whereas the observation equation (1d) relates the log volatility and its (observable) proxy R_t. Transition errrors η_{1t}, η_{2t} and measurement errors ε_t are supposed to be mutually uncorrelated i.i.d. disturbances.

It is important to note that the log-volatility is a latent variable and therefore is not directly observable. Instead, we are able to observe some of its proxy which is contaminated by some amount of measurement error. Typical examples of proxies for h_t include log range

$$\log\left(\sup_{s\in[0,1]} p(s) - \inf_{s\in[0,1]} p(s)\right) \quad (1)$$

and log absolute return

$$\log\left|p(1) - p(0)\right|. \quad (2)$$

where p(s) denotes the log price and supremum and infimum are taken over the daily interval which is normalized to unity purely for ease of notation.

Log absolute returns (or squared returns) have historically constituted a first-choice proxy (see for instance [10] or [15]), however, we will follow [1] and make use of the log range instead. The asymptotic distribution of the log range has been studied in [1]. Based on the result of Feller [7], they find out that log range can be well aproximated by the normal distribution with mean 0.43 + h_t and variance 0.08. On the other hand, the distributional properties of the log absolute return are quite different. Assuming that daily returns are given by

$$r_t = u_t \exp(h_t) \quad (3)$$

where $u_t \sim nid(0,1)$, then

$$\log(|r_r|) = -0.64 + h_t + \xi_t \quad (4)$$

where $\xi_t = \log(|u_t|) - E\log(|u_t|)$ has zero mean and variance $\pi^2/8 \doteq 1.23$ and is highly skewed. Therefore, comparing to the log absolute returns, the superiority of the log range stems from its lower variance and the fact that its distribution is nearly Gaussian which turns out to be extremely useful if the Kalman filter algorithm is employed.

However, a word of caution is needed here. In finite samples, the distribution of range estimators depends also on the number of observations per unit of time (day in this case). Therefore, we investigated the impact of discretization on mean and variance of the log range by a Monte Carlo simulation (results are reproduced in Table 1). The pattern is clear: reducing the number of observations during a trading day results in lower mean and higher variance. Nevertheless, the variance of the proxy seems to be quite close to the asymptotic one even for 50 trades per day.

N	*5*	*10*	*50*	*100*	*200*	*500*	*1000*
mean	*–0.115*	*0.097*	*0.300*	*0.340*	*0.366*	*0.401*	*0.415*
variance	*0.233*	*0.152*	*0.104*	*0.097*	*0.092*	*0.086*	*0.084*

Table 1: Mean and variance of log range for a Wiener process with zero drift and unit variance observed N times during a unit period. The Monte Carlo simulation was performed with 1 million replications.

There exists a closely related estimator proposed by Parkinson [13] and several modifications thereof: Garman and Klass [8] included the open and close prices in addition to the high and close prices, and CityplaceRogers and Satchell [14] suggested an estimator which allowed for a nonzero drift and also investigated the discretization bias (see also [5]).

If higher efficiency is needed, intraday data should be used and appropriated volatility estimator constructed (realized volatility, see [2] or realized range, see [12]).

3 Empirical application

EZ (CZ0005112300)	*Prague Stock Exchange (SPAD)*
Telefnica 02 C.R. (CZ0009093209)	*Prague Stock Exchange (SPAD)*
Erste Bank (AT0000652011)	*Prague Stock Exchange (SPAD)*
General Electric Co.	*New York Stock Exchange*
Microsoft Corp.	*NASDAQ*
Intel Corp.	*NASDAQ*

Table 2: Stocks included in the dataset

We use daily high and low prices of six stocks (see Table 2) for the period from September 16, 2005 until November 13, 2007 (543 and 544 observations for Czech and US markets, respectively). The average number of transactions per day for CEZ, Telefnica O2 and Erste

Bank was 231, 102 and 86, respectively. In the case of Erste Bank there were only three transactions during the trading day in April 14, 2006 giving rise to the observed range very close to zero and a corresponding large negative outlier in the log range. Therefore, this observation was excluded from our analysis.

	mean	*std deviation*	*skewness*	*kurtosis*
EZ	–3.7541	0.5796	0.1082	3.1717
Telefnica 02	–4.2106	0.6215	0.2519	3.3516
Erste Bank	–4.2254	0.5840	–0.0179	2.9515
General Electric	–4.4138	0.3962	0.4316	3.1181
Microsoft	–4.2230	0.4194	0.3390	3.0718
Intel	–3.9355	0.3821	0.1501	2.7630

Table 3: Unconditional moments of the observed log range

The data are depicted in Figure 1 together with their sample autocorrelation functions and QQ plots. The autocorrelation functions clearly show a certain degree of persistency in the volatility proxy. Empirical moments of the log range are reported in Table 3. The empirical coefficients of skewness and kurtosis roughly correspond to their theoretical values which are 0.17 and 2.80, respectively.

Now we proceed to estimate the stochastic volatility model given by equations (1a) through (1d) using the Kalman filter (see [9]). All the calculations were carried out using MATLAB 7.1. Estimation results are shown in Tables 4: there exists a strong evidence that the volatility process can be meaningfully decomposed into one highly persistent factor and another quickly mean-reverting factor. In order to obtain a better interpretation of our results, we computed estimated variances for both volatility factors and the total variance of the log-volatility process as their sum (due to zero cross-correlation) (see Table 5). The variance of both factors is roughly the same with notable exceptions of Erste Bank and Intel where the second (less persistent component) seems to be much more volatile. Comparing Czech and US markets, the most striking feature is substantially lower *volatility* of the log-volatility process for US stocks, even if the overall *level* of the volatility of returns (estimated by $\bar{h}$) is roughly the same.

4 Conclusion

We have applied a two-factor model of stochastic volatility along the lines in [1]. The analysis in this paper can be extended to a multivariate setting which could be highly relevant for risk management purposes. However, this issue is left for further research.

	ρ_1	ρ_2	$\bar{h}$	$var(\eta_1)$	$var(\eta_2)$
EZ	*0.9427*	*0.3992*	*-4.1688*	*0.0156*	*0.0941*
Telefnica O2	*0.9403*	*0.4938*	*-4.6317*	*0.013*	*0.1427*
Erste Bank	*0.9395*	*0.4235*	*-4.6493*	*0.0081*	*0.1542*
General Electric	*0.9551*	*0.1602*	*-4.8295*	*0.0039*	*0.0299*
Microsoft	*0.946*	*0.1086*	*-4.6366*	*0.0056*	*0.0404*
Intel	*0.9799*	*0.3444*	*-4.3664*	*0.0004*	*0.0457*

Table 4: Quasi-maximum likelihood estimates of the two-factor model

	first factor	*second factor*	*total variance*
EZ	*0.1401*	*0.1119*	*0.2520*
Telefnica O2	*0.1122*	*0.1887*	*0.3009*
Erste Bank	*0.0690*	*0.1879*	*0.2569*
General Electric	*0.0444*	*0.0307*	*0.0751*
Microsoft	*0.0533*	*0.0409*	*0.0942*
Intel	*0.0101*	*0.0518*	*0.0619*

Table 5: Variances of individual factors and the total variance of the estimated log-volatility

References

[1] ALIZADEH, S., BRANDT, M., DIEBOLD, F. (2002): *Range-based estimation of stochastic volatility models.* Journal of Finance 57, pp. 1047–1091.

[2] ANDERSEN, T.G., BOLLERSLEV, T., DIEBOLD, F.X., LABYS, P. (2001): *The distribution of realized exchange rate volatility.* Journal of the American Statistical Association 96, pp. 42–55.

[3] BARNDORFF-NIELSEN, O. E. , SHEPARD, N. (2001): *Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economic..* Journal of the Royal Statistical Society, Series B 63, pp. 167-241.

[4] BOLLERSLEV, T. (1986): *Generalized autoregressive conditional heteroskedasticity.* Journal of Econometrics 31, pp. 307–327.

[5] CHRISTENSEN, K., PODOLSKIJ, M. (2006): *Realized range-based estimation of integrated variance,* Journal of Econometrics (forthcoming).

[6] ENGLE, R.F. (1982): *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation.* Econometrica 50, pp. 987–1007.

[7] FELLER, W. (1951): *The asymptotic distribution of the range of sums of independent random variables.* Annals of Mathematical Statistics, 22, pp. 427-432.

[8] GARMAN, M. , KLASS, M.J. (1980): *On the estimation of security price volatilities from historical data.* Journal of Business, 53, pp. 67-78.

[9] HAMILTON, J. (1994): *Time Series Analysis.* Princeton, Princeton University Press.

[10] HARVEY, A.C. , RUIZ, E. , SHEPARD, N. (1994): *Multivariate stochastic variance models.* Rev. Economic Studies, 61, pp. 247-264.

[11] HULL, J. , WHITE, A. (1987): *The pricing of options on assets with stochastic volatilities.* Journal of Finance 42, pp. 281–300.

[12] MARTENS, M. , VAN DIJK, D. (2007): *Measuring volatility with the realized range.* Journal of Econometrics, 138, pp. 181-207.

[13] PARKINSON, M. (1980): *The extreme value method for estimating the variance of the rate of return.* Journal of Business, 53(1), pp. 61-65.

[14] ROGERS, L. C. G., SATCHELL, S. E. (1991): *Estimating variances from high, low, and closing prices*, Annals of Applied Probability 1(4), pp. 504–512.

[15] RUIZ, E. (1994): *Quasi-maximum likelihood estimation of stochastic volatility models.* Journal of Econometrics, 63, pp. 289–306.

[16] TAYLOR, S. J. (1982): *Financial returns modelled by the product of two stochastic processes—a study of daily sugar prices 1961-79.* In O. D. Anderson (ed.), Time Series Analysis: Theory and Practice, 1, pp. 203-226. Amsterdam: North-Holland.

Figures

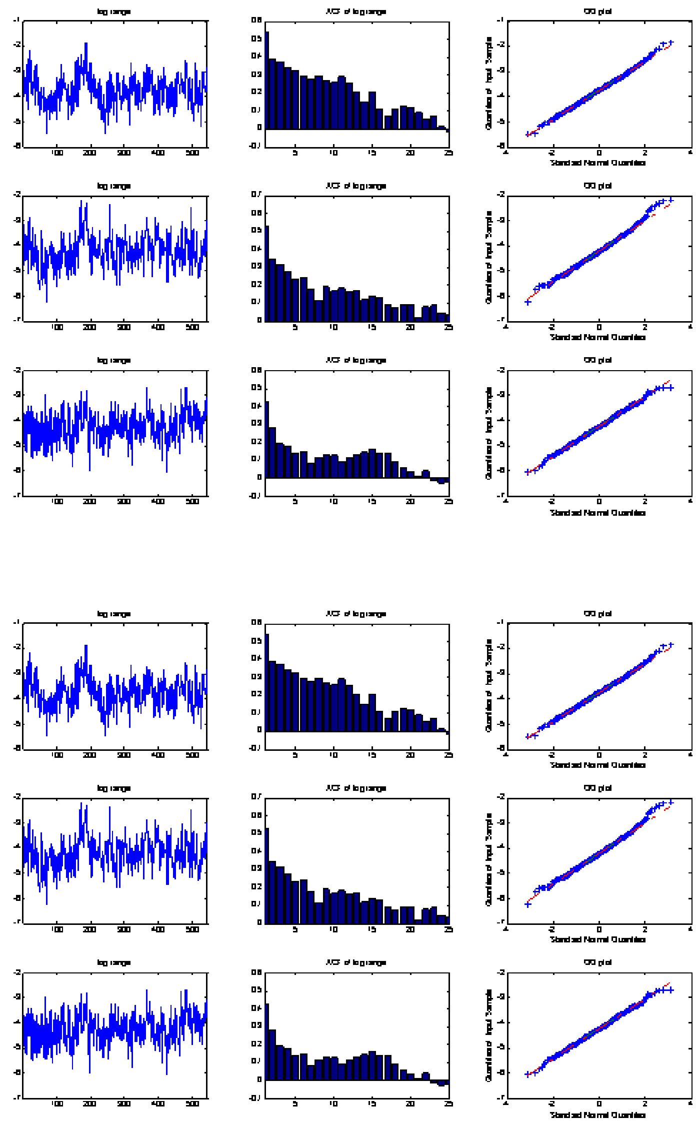


Figure 1: Log range, its sample autocorrelation function and QQ plots for CEZ, Telefonica O2, Erste Bank, General Electric, Microsoft and Intel (from top to bottom)

Milada Lagová, Jana Kalčevová

Department of Econometrics
University of Economics Prague
W. Churchill sq. 4, 130 67 Prague 3

e-mail: lagova@vse.cz, kalcevov@vse.cz

Automatic Testing Framework for Linear Optimization Problems

Abstract

Student examination is both time-consuming and boring, error-prone task for teachers. Nowadays a modern trend is to leave as much work as possible on computers. This paper proposes LPTest, a tool for automatic testing of linear programming problems solving. LPTest has been developed as plugin for LP System that is currently used at University of Economics in Prague, Department of Econometrics in Linear models, Operation research and other optimization courses (this LP System was introduced last year on Mathematical methods in Economics Conference). LP System consists of six modules (simplex method with dual simplex method and manually controlled computation, transportation problem, postoptimal changes with parametric programming, lower and upper bounded variables, integer programming and multicriteria programming), but in detail, LPTest is limited only to simplex method and dual simplex method algorithm and checks for proper content of simplex tables. LPTest is not intended only for teachers, students use both LPTest and LP System also for their exercises and homeworks, and not only for knowledge testing, because of its numerical and also algorithmic checks. LPTest as well as LP System has been created in Microsoft Office Access and it reads help from Microsoft Office Word document and exports simplex tables, results and also LPTest protocol into Microsoft Office Excel sheets.

Keywords: *Linear programming, testing, simplex method*

1 Introduction of LPTest

Student examination is both time-consuming and boring, error-prone task for teachers. Nowadays a modern trend is to leave as much work as possible on computers. The University of Economics in Prague, Department of Econometrics has developed software called LPTest for automatic testing and educational purposes based on simplex and dual simplex method. It is build as a program appendix to "LP System" (it was introduced last year [1], [2]) that is used in Linear models, Operation research and other optimization courses.

Program LPTest is as well as LP System implemented in Microsoft Access environment because its appropriate format of input and output screens and user's interactive contact with LPro system play the role. Operation system Windows XP, Microsoft Office Access 2003 and monitor with diagonal at least metricconverterProductID15 inches15 inches are necessary for the run of LPTest. For help is necessary Microsoft Office Word 2003 and saving of protocol needs Microsoft Office Excel 2003.

The main function of LPTest.mdb is an automatic check of student's errors—both numerical and methodical. The calculation process is divided into logical blocks called Akce (Action). Action is for example “insert slack or surplus variables” in the initial solution, “selection of pivot row and pivot column” in iteration, “table transformation” etc.

2 Check of LPTest

There are implemented two ways of calculation check in the program: check after each action (kontrola 1—check 1) and check in the end of calculation (kontrola 2—check 2). The first way informs and records errors immediately after each action and it is no possible to continue without correction of errors. The program gives notice of error and waits for correction. This way is suitable for testing of knowledge and for practice. The second way checks only the initial solution during calculation process (the wrong initial solution means no sense of calculation). The correctness of calculation and numerical values are checked after the end of calculation process. Student so works individually as well as in the manually calculation process.

For each student the LPTest creates result report (protocol) in which all errors sorted by type (below) and iterations are presented. The protocol is available on the screen and also in the unique Excel file. The teacher can so documented protocols of each student during the whole semester.

Program input data are created and kept by program LinPro.mdb or LPPro.mdb. It is not possible to insert or edit input data by using a screen in LPTest. The user's file with standard name LPdata.mdb must be placed in the directory with LPTest.mdb in the case of the first software run. The data file is individual for each user and it contains input data for solved problems. After the first software run the user can loaded new data files from arbitrary files, places and dictionaries (only *.mdb extension is necessary).

The data unit of LPTest is “Úloha” (Problem) as well as in the case of LP System. Since this application is focused only on testing the problem dimension is restricted on maximally three constraints and four variables. The example is presented in the Figure 1.

Vstupní údaje úlohy 3

	Interpretace	x1	x2	x3	Rel.	b
1	Omezení 1	1	2	0	>=	90
2	Omezení 2	2	-2	2	>=	340
41	Účelová funkce	120	35	100	**min**	

Figure 1: Input data

LPTest is able to indicate eight types of errors—two are numerical and the others are related with algorithm application. As we said before errors are displayed either during calculation process (check 1) or after calculation process (check 2)—see Figure 2. Information about the type of each error and number of errors is saved in protocol.

Zjištěné chyby

Poř.	Iter.	Kontrola	Chyba	Druh
1	1	Přidání proměnných	Nesprávně přidaná proměnná	2

Figure 2: Errors window

3 The work with LPTest

The problem selected by student from data file is displayed in the form in Figure 3.

								Metoda:	SM
		Ceny p			0	0	0	Extrém:	min
			Ceny z		120	35	100	Iterace:	1
				Prom.	x1	x2	x3	b	t
	1				1	2	0	90	
	2				2	-2	2	340	
	3			z(1)					
	4			p(1)					

Figure 3: Initial solution of user's problem

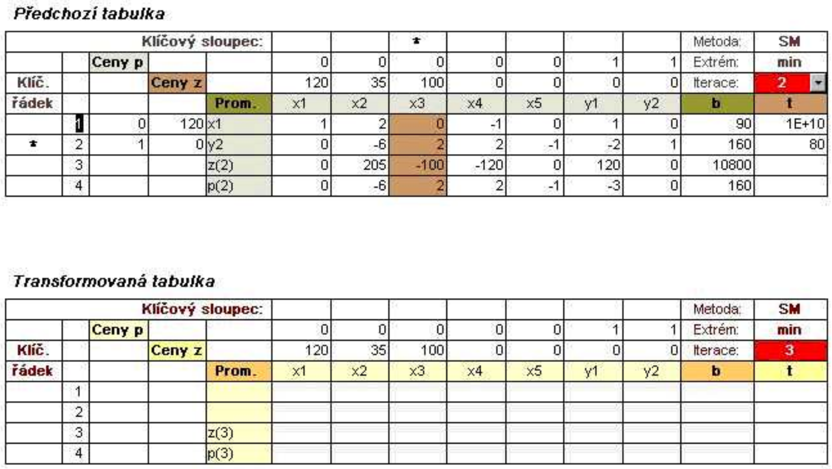
The table in Figure 3 is in this time closed (gray color). Student then selects from the menu slack/surplus or artificial variables those have been inserted in the table. If the selection is wrong user has to select again (and this error is registered).

If all slacks, surpluses and artificial variables are included the table is opened for edit. Student then inserts numerical data and so creates initial solution. After correct initial table the iteration process starts.

Each iteration screen consists of two tables (see Figure 4). The first one is the last table of previous iteration with numerical values (Předchozí tabulka—Previous table). This table is closed for numerical changes and it is given for selection of pivot column and pivot row (marked by star). In the case of check 1, if the user presses "Konec akce" (End of Action) the check is done, errors are displayed and selection is repeated. In the case of check 2 errors are not displayed. Then is presented the second table (see Figure 4)—Transformovaná tabulka (Transformed table) in which user fills a new solution (numerical values and other data). After press of "Konec akce" (End of Action) and check—in the case of check 1—the next iteration starts.

The end of solution and the type of the end (optimal solution, unfeasible solution, unbounded solution and alternative solution) selects student also. After it program creates error protocol (see Figure 5) and saves it into Microsoft Office Excel sheet. The numbers of errors are sorted by type marked from E01 to E08 in protocol.

The list of error types is possible to see in menu "Druhy chyb" (Errors types), for example:



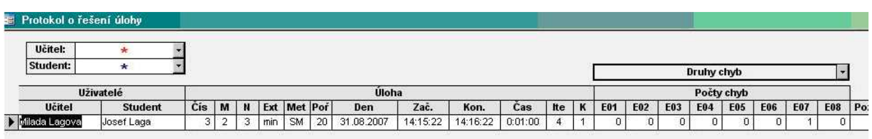
Předchozí tabulka

			Klíčový sloupec:				*					Metoda:	SM
		Ceny p			0	0	0	0	0	1	1	Extrém:	min
Klíč.			Ceny z		120	35	100	0	0	0	0	Iterace:	2
řádek				Prom.	x1	x2	x3	x4	x5	y1	y2	b	t
	1	0	120	x1	1	2	0	-1	0	1	0	90	1E+10
*	2	1	0	y2	0	-6	2	2	-1	-2	1	160	80
	3			z(2)	0	205	-100	-120	0	120	0	10800	
	4			p(2)	0	-6	2	2	-1	-3	0	160	

Transformovaná tabulka

			Klíčový sloupec:									Metoda:	SM
		Ceny p			0	0	0	0	0	1	1	Extrém:	min
Klíč.			Ceny z		120	35	100	0	0	0	0	Iterace:	3
řádek				Prom.	x1	x2	x3	x4	x5	y1	y2	b	t
	1												
	2												
	3			z(3)									
	4			p(3)									

Figure 4: Iteration process



Protokol o řešení úlohy

Učitel: *

Student: *

Druhy chyb

Uživatelé		Úloha												Počty chyb								
Učitel	Student	Čís	M	N	Ext	Met	Poř	Den	Zač.	Kon.	Čas	Ite	K	E01	E02	E03	E04	E05	E06	E07	E08	Po:
Milada Lagova	Josef Laga	3	2	3	min	SM	20	31.08.2007	14:15:22	14:16:22	0:01:00	4	1	0	0	0	0	0	0	1	0	

Figure 5: Error protocol

E01: the selected solution method is wrong

E05: there is a numerical error in the table (in the variables or right-hand-sides columns)

E08: the calculation was ended in the wrong time (too early, too late) or the type of the end is wrong

4 Conclusion

Program LPTest was developed at the University of Economics in Prague, Department of Economics as an extension of LP System—program that is used in optimization courses. All terms used in LPTest are identical to terms used in LP System and english version of these terms was published in [1]. Unfortunately, only czech version of this program is available because this software is determined for students of optimization courses taught in Czech language. The primary goal of this research was to create software for Czech students and it is determined for testing of mathematical knowledge and not language knowledge. LPTest is not used in lessons in this time because it is now is tested phase. The authors hope it will be as useful as last year presented LinPro and LPPro—the parts of LP System. In the case of financial support the authors will consider to crate the English translation of the whole LP System with LPTest.

References

[1] KALČEVOVÁ, Jana, LAGOVÁ, Milada. Computer Support of Courses of Linear Optimization Models. Pilsen 13.09.2006 – 15.09.2006. In: *Mathematical Methods in Economics 2006 [CD-ROM].* Pilsen : University of West Bohemia in Pilsen, 2006, s. 333–338. ISBN 80-7043-480-5.

[2] LAGOVÁ, Milada, KALČEVOVÁ, Jana. System LPPro for Computer support of courses of linear optimization. Praha 08.07.2007 – 11.07.2007. In: *EURO XXII Prague.* United Kingdom : OptiRisk Systems, 2007, s. 170.

[3] LAGOVÁ, Milada, JABLONSKÝ, Josef. *Lineární modely.* 1. vyd. Praha : Oeconomica, 2004. 287 s. ISBN 80-245-0816-8.

Martin Lukáčik, Karol Szomolányi, Adriana Lukáčiková

Department of Operation Research and Econometric
Faculty of Business Informatics
University of Economics
Dolnozemská cesta 1/b, 852 35 Bratislava

e-mail: lukacik@euba.sk

Credibility of Slovak National Bank—Measuring of Inflation Bias in Slovak Economy[1]

Abstract

By entrance to European Monetary Union, Slovak economy will give up its own monetary policy. How credible this policy is? Is it necessary to remove it by common European monetary policy?

Kydland and Prescott (1977) [7] introduced a time-consistency problem of the economic policy. The time consistency problem is also called as credibility problem. A monetary authority can by time-inconsistent policy through the private sector expectations produce an inflation bias. By measuring of the inflation bias we can evaluate the monetary policy. The most of papers (Barro and Gordon [2]) dealing with time-consistency problem of the economic policy assume quadratic preferences of agents. In flavor of quadratic preferences, Ireland [5] found a way how to verify time-inconsistency problem of the monetary policy. Using this approach in our previous papers (Lukáčik, Szomolányi [8]) we found that there is no time-consistency problem of the Slovak monetary policy for short run. Surico [12] offered a way how to measure an average inflation bias under the asymmetric preferences assumption. Using this approach, in the paper we will measure time-inconsistency problem of the Slovak monetary policy. In comparison with our previous papers we will find that asymmetric preferences better correspond to Slovak economy than quadratic preferences. Slovak National Bank lacks commitment and produces inflation bias.

Keywords: *time consistency of the monetary policy, asymmetric preferences, inflation bias*

1 Introduction

Nobel Prize laureates in 2005, Kydland and Prescott [7], offered an explanation of inflation behavior by introducing a model of the time-inconsistency of the monetary policy problem. Society's preferences to low inflation and output higher than potential product, expressed by a quadratic loss function, makes an incentive of monetary authority, who cannot commit its policy, to deviate from original plans. By assumption that the money growth has effect on

[1] This paper is supported by the Grant Agency of Slovak Republic – VEGA, grant no. 1/4652/07 "The Analysis of Actual Problems of Slovak Economy Progress before the Entrance to European Monetary Union – Econometrical Approach".

inflation (see Romer [11] e.g.) and under rational expectations, time inconsistent monetary policy produces an inflation bias.

Economic environments and monetary policies differ by timing of the decision making by agents. We refer their as "with commitment" and "without commitment". In environment without commitment agents behave sequentially rational (Chari, Kehoe and Prescott [4]). Central bank announces target inflation, before it chooses the monetary policy, but, as it cannot commit its policy, it can only choose the actual inflation level after agents choose their expectations. The marginal cost of slightly higher inflation rate is zero and marginal cost of slightly higher output level is positive. Since there is no uncertainty and agents understand the time-inconsistency problem of monetary authority, the expected level of inflation is higher and monetary authority has to deviate from the announced policy. This is time-inconsistency problem of the monetary policy and it implies high inflation even though output does not rise.

Even if, potentially, there is a recourse of the time-inconsistency problem of the monetary policy in every time and in all economies, there have been economies, both in the past and in the presence, with low inflation (USA in 50's of the last century, Germany in the most of the post-war period, significant inflation problem neither have been at Slovakia since its origin). These phenomena economists have explained by reputation (see Barro [1]), delegation (see Rogoff [10]), punishment equilibrium (see Barro and Gordon [3]) and incentive contracts (see Persson and Tabellini [9]).

There is no place in the paper to present more detailed overview of the time-inconsistency problem of the monetary policy. To learn more about the problem we refer reader to study papers mentioned above. More comfortable Slovak or Czech reader can study the problem from papers of authors: Szomolányi ([13], [14]) and Szomolányi, Lukáčiková and Lukáčik ([15], [16]).

There are econometric techniques helping us to verify and measure time-inconsistency problem of the monetary policy in the economy. country-regionIreland [5] answered the question whether the time-consistency problem explained the behavior of inflation in the United States economy in 1960-1997. He used Barro and Gordon's [2] theory, where preferences of agents are expressed by quadratic square function.

The Ireland's approach using Slovak data was realized in our latest paper (Lukáčik, Szomolányi [8]). Following Ireland, the presence of cointegration relationship between unemployment and inflation rates implied long-run time-inconsistency problem of the Slovak monetary policy. On the other hand a special VARMA (1,2) structure generated by Barro and Gordon theory failed in explanation of the whole dynamics of Slovak inflation and unemployment. This suggests that there is no short-run time-inconsistency problem of the Slovak monetary policy.

Surico [12] removed in the Barro and Gordon's [2] theory quadratic preferences by asymmetric and offered a way how to explain inflation in economy with no commitment and how to measure an average inflation bias. The aim of the paper is to measure an average inflation bias of the Slovak monetary policy using Surico's approach and Slovak data. According to the approach there is no incentive of the society to produce more than potential product, but there is incentive to depreciate the positive deviation of the current production from the potential product in comparison with the same negative deviation. Positive output gap means lower loss of the society's welfare than the same negative output gap—society's preferences are asymmetric. We will show that asymmetric preferences of Slovak monetary authorities are valid and we will find a significant average inflation bias of the Slovak monetary policy.

The paper can be divided into sections: the short introduction and review of papers dealing with the time-inconsistency problem of the monetary policy has been in introduction; in the second section is Surico's modification of Barro and Gordon's theory; in the third section are empirical results using Slovak data and the paper is closed by conclusion.

2 Time-Inconsistency Problem of the Monetary Policy under Asymmetric Preferences

Let us assume that the relation between inflation and unemployment is given by expectation augmented Phillips Curve:

$$y_t = \theta(\pi_t - \pi_t^e) + u_t,\ \theta > 0 \tag{1}$$

The left side of the term (1) is output gap expressed by deviation of the current output and potential product, inflation rate is denoted as π_t, symbol $\theta > 0$ denotes Phillips Curve parameter, the random error u_t express a supply shock that is autoregressive according to $u_t = \rho u_{t-1} + \varepsilon_t$, where ρ is positive and less than one,ε_t is shock with zero mean and standard deviation σ_ε. Output gap is random, since it depends on u_t, and its standard deviation σ_y is increasing function of ρ and σ_y. The private sector has rational expectations:

$$\pi_t^e = E_{t-1}\pi_t \tag{2}$$

where E_{t-1} are expectations in time t upon an information in time $t-1$.

Surico's model differs from Barro and Gordon's one by the form of the objective function. The desired output does not exceed potential product, but the same differences of the current production from potential product with different signs generate different loss function values. The time-inconsistency problem of the monetary policy is not characteristic by an effort of the central bank to over valuate potential product, but by different evaluating of the same negative and positive output gap. Positive output gap is for monetary policymaker a smaller loss than the same negative one. The loss function will be therefore expressed in lin-exp form:

$$L_t = \frac{1}{2}(\pi_t - \pi^*)^2 + \mu\frac{e^{\nu y_t} - \nu y_t - 1}{\nu^2} \tag{3}$$

where p^* is desired inflation, parameters $\mu > 0$ and ν express relative wages, or asymmetric preferences respectively, on output stabilization. Consider negative parameter $\nu < 0$. It is clear that whenever output gap is negative $y_t < 0$, the exponent component of the loss function dominates the linear component, while the opposite is true for the positive output gap, $y_t > 0$. Under negative parameter $\nu < 0$ the positive output gap generates lower loss than the same negative output gap. This approach differs from Barro and Gordon's theory, where preferences are quadratic and desired production is higher than potential product.

Using L'Hôpital rule it can be showed that quadratic form is a special case of lin-exp form of loss function. If ν tends to zero the function (3) can be written in the form: $L_t = 1/2(\pi_\tau - \pi^*)^2 + \mu y_t^2$. The test of the statistical significance of the parameter ν therefore will show, whether the central bank's preferences are asymmetric. As we will see in section 3 National Bank of Slovakia has asymmetric preferences.

Monetary authority set the policy that minimizes all expected losses over time according to:

$$\min E_{t-1} \sum_{\tau=0}^{\infty} \delta^{\tau} Lt + \tau \tag{4}$$

where δ is the discount factor. Constraints of the problem differ according to different timing of making decisions—according to different economical environment.

By the commitment assumption we assume that central bank can commit its policy—it set inflation before agents make expectations. With commitment central bank minimizes the (4) subject to the (3), the Phillips Curve (1) and the rational expectations condition (2). The solution of the problem can be written as:[2]

$$E(\pi_t) = \pi^* \tag{5}$$

Inflation target on average equals the socially desirable inflation rate and therefore is independent of the output gap.

Monetary authority without commitment chooses target inflation π_t after setting of inflation expectations π_t^e by private sector, but before the realization of the real shock u_t. The central bank minimizes the (4) subject to the (3) and subject to the Phillips Curve (1). The first order conditions are:

$$(\pi_t - \pi^*) + E_{t-1}\left[\frac{\mu\theta}{\nu}(e^{\nu y_t} - 1)\right] = 0 \tag{6}$$

Let us compare solutions (6) under quadratic and under asymmetric preferences. Using the L'Hôpital rule, it can be showed that if parameter ν tends to zero the (8) tends to:

$$(\pi_t - pi^*) = -\mu\theta E_{t-1}(y_t) \tag{7}$$

This implies that under quadratic preferences there is linear relation between inflation and output gap. Moreover under white noise assumption of the supply shock, e.g. $\rho = 0$, inflation bias vanish. Since society's desired output equals to potential product and since the symmetric preferences to the output gap, the system does not generate systematic inflation bias.

After some modification[3] of the (6) we can get monetary policy without commitment rule:

$$E(\pi_t);\ \pi^* - \frac{\mu\theta\nu}{2}\sigma_y^2 \tag{8}$$

The difference between expected inflation rate with commitment and expected inflation rate without commitment is average inflation bias $-\mu\theta\nu\sigma_y^2/2$. The source of inflation bias is not a

[2] See Surico [12] for more details.
[3] See Surico [12] for more details.

Barro and Gordon's incentive of the monetary authority to select slightly higher inflation in flavor of higher output, but it is an incentive of the monetary authority to react more aggressive if current output is below the potential. Since agents know the structure of the economy and they understand time-inconsistency problem of monetary authority, their expectations are higher by the inflation bias. Under perfect expectations central bank must deviate from inflation target π^*.

As Surico writes: "Possible improvements to the discretionary solution include the appointment of a more conservative central banker, i.e. one endowed with a lower relative weight μ in the spirit of Rogoff [10] or with lower inflation target than society. Alternatively, the appointment of a more symmetric policy-maker, endowed with a smaller absolute value of ν, can also enhance welfare."

If monetary policy is time-inconsistent characterized by asymmetric preferences of the society, inflation bias is given by the (6). The term can be written by reduced econometric form[4]:

$$\pi_t = \pi^* + \alpha y_t + \beta y_t^2 + \nu_t \tag{9}$$

where parameters π^* is an average inflation target of the monetary policy, $\alpha = -\mu\theta < 0$ and $\beta = -\mu\theta\nu/2 > 0$ and error term is orthogonal to any variable in the information set available at time $t-1$. These orthogonal properties determine the GMM as appropriate estimator of the (9).

From parameters of the (9) we can derive asymmetric preferences parameter as $\nu = 2\beta/\alpha$. Testing of statistical significance of ν will answer the question, whether preferences are asymmetric. Average inflation bias we can estimate as $\beta\sigma_y^2$.

Surico realised estimation of (9) using two different samples of country-regionplaceU.S. data: sample corresponding to pre-Volcker regime (1960Q1 - 1979Q3) and sample corresponding to post-Volcker regime (1982Q4 - 2005Q2). He found out statistically significant asymmetric parameter ν with average inflation bias of about one percent in pre-Volcker regime, and statistically insignificant asymmetric parameter ν in post-Volcker regime.

3 Slovak Monetary Policy

We used seasonally adjusted quarterly time series of Slovak economy before the revision realised in the end of 2007; our data are from 1995Q1 till 2007Q2. The inflation is measured by three methods; the first way is equivalent to Surico's one—it means the annualized change in the log of the GDP deflator; the second way is annualized change in the log of the CPI and the third one is annualized change in the log of the HCPI. Output gap is difference of log real and potential product. Data are from the SLOVSTAT database published by The Statistical Office of the PlaceNameSlovak PlaceTypeRepublic, HCPI is from the EUROSTAT database and potential product is from The Financial Policy Institute of Ministry of Finance of the placePlaceNameSlovak PlaceTypeRepublic.

Different estimates of parameters of the (9) are in Table 1. We estimated the equation for all three measurements of the inflation rate. Following Surico, we checked of results to changes in the instrument set. For estimation of inflation rate measured as log changes of GDP deflator

[4] See Surico [12] for more details.

we used two instrument sets: exogenous variables, five lags of inflation, output gap and squared output gap (instrument set I) and exogenous variables, six lags of inflation and output gap and three lags of output gap (instrument set II). For estimation of inflation rate measured of changes of both CPI and HCPI we used same two instrument sets: exogenous variables (including constant), three lags of inflation, output gap and squared output gap (instrument set I) and four lags of output gap and square output gap and one lag of inflation (instrument set II). Together we have six estimations of parameters of the (9). In all cases we realised Hansen's test for over-identifying restrictions. All instrument sets are statistically significant. Estimated values of parameters are in Table 1 with corresponding standard error in round brackets. All estimates are statistically significant at 5% level.

From Table 1 it follows that average inflation target of national Slovak bank was about four and half percent measured by annualized log change of GDP deflator and about 6% measured by annualized

Measurement of the Inflation	**Instrument Set**	π^*	α	β	ν	***p values***
GDP deflator	I	4.72	-0.7	0.12	-0.35	J(21): 0.21
		(0.25)	(0.08)	(0.03)	(0.08)	
	II	4.45	-0.67	0.13	-0.39	
		(0.17)	(0.08)	(0.03)	(0.08)	
CPI	I	6.06	-0.27	0.11	-0.79	
		(0.2)	(0.11)	(0.03)	(0.29)	
	II	5.98	-0.26	0.13	-1.03	
		(0.22)	(0.1)	(0.04)	(0.3)	
HCPI	I	6	-0.26	0.12	-0.93	
		(0.23)	(0.11)	(0.03)	(0.29)	
	II	6.08	-0.28	0.12	-0.87	
		(0.23)	(0.11)	(0.04)	(0.28)	

Table 1: Estimates of parameters of the equation 10

log change of consumer indexes. Statistically significance of the asymmetric parameter means that National Bank of country-regionplaceSlovakia has incentive for different evaluation of the same negative and positive output gap. This incentive leads to systematic inflation bias due to the (8).

Evaluations of the average inflation bias for all six equations are in the Table 2. Average inflation bias is computed as $\beta sig\tilde{m}a_y^2$. We can see that all six equations generate from 0.19% to 0.25% of average inflation bias. Average inflation targets are absolute constants π^*. In the last but one column is sum of average inflation bias and average inflation target. Values in brackets are corresponding standard errors. In the last column are values from descriptive statistics of

Measurement of the Inflation	Instrument Set	Average Inflation Bias	Average Inflation Target	Sum	Inflation Mean
GDP deflator	I	0.23	4.72	4.94	4.95
		(0.05)	(0.25)	(0.22)	(2.25)
	II	0.24	4.45	4.69	4.95
		(0.06)	(0.17)	(0.14)	(2.27)
CPI	I	0.19	6.06	6.25	6.41
		(0.06)	(0.2)	(0.18)	(2.97)
	II	0.25	5.98	6.23	6.38
		(0.08)	(0.22)	(0.17)	(3)
HCPI	I	0.21	6	6.2	6.46
		(0.05)	(0.23)	(0.21)	(3.29)
	II	0.23	6.08	6.3	6.42
		(0.08)	(0.23)	(0.18)	(3.21)

Table 2: Average inflation bias, average inflation target, inflation mean

all three dependent variables. Differences of descriptive statistics in rows corresponding to the same dependent variable are because different instrument sets generated different samples.

4 Conclusion

Our results correspond to results of our latest paper (Lukáčik, Szomolányi [8]). Using Ireland's approach we found co-integration relationship between inflation and unemployment to conclude that there is long-run time consistency problem in the Slovak monetary policy. Using Surico's approach of asymmetric preferences and using GMM we found relation between inflation, output gap and squared output gap. Asymmetric preferences of National Bank of country-regionplaceSlovakia lead to systematic inflation bias.

Is entrance of the Slovak economy to European Monetary Union appropriate? If European Central Bank disposes commitment, that Slovak National Bank lacks and if Euro area is optimal currency area, the answer is yes. For this time, however, we do not have enough information about Euro area and European monetary placeUnion to answer questions above. Furthermore, interesting results could generate a robust model with asymmetric preferences for both output gap and inflation difference. Answering these questions will be subject of future papers of authors.

References

[1] Barro, R. J. *Reputation in a Model of Monetary Policy with Incomplete Information.* Journal of Monetary Economics 17, pp. 1-20, 1986

[2] Barro, R. J., Gordon, D. B. *A Positive Theory of Monetary Policy in a Natural Rate Model.* Journal of Political Economy 91, pp. 589-610, 1983

[3] Barro, R. J., Gordon, D. B. Rules, *Discretion, and Reputation in a Model of Monetary Policy.* Journal of Monetary Economics 12, pp. 101-121, 1983

[4] Chari, V. V., Kehoe, P. J., CityplacePrescott, E. C. *Time Consistency and Policy.* In Barro, R. J.: Modern Business Cycle Theory. PlaceNameHarvard PlaceTypeUniversity Press, Boston, 1989

[5] Ireland, P. N. *Does the Time-Consistency Problem Explain the Behavior of country-regionplaceUS Inflation?* Journal of Monetary Economics 38, pp. 215-220, 1999

[6] Ištvániková, A. *Optimalizačnı prístup k formovaniu hospodárskej politiky na základe ekonometrického modelu.* 7. medzinárodná vedecká konferencia Kvantitatívne metódy v ekonómii a podnikaní, Bratislava 2001

[7] Kydland, F., Prescott, E. *Rules Rather than Discretion: The Inconsistency of Optimal Plans.* Journal of Political Economy, vol. 85, 1977

[8] LUKÁČIK, M., SZOMOLÁNYI, K. *Is Slovak Monetary Policy Time Consistent?* Quantitative Methods in Economics, Multiple Criteria Decision Making XIV, 2008

[9] Persson, T., Tabellini, G. *Designing Institutions for Monetary Stability.* Carnegie-Rochester Conference Series on Public Policy 39, 1993

[10] Rogoff, K. *The Optimal Degree of Commitment to an Intermediate Monetary Target.* Quarterly Journal of Economics 100, pp. 1169-1190, 1985

[11] Romer, D. *Advanced Macroeconomics.* Third edition McGraw-Hill, Berkeley, 2006

[12] Surico, P. *Measuring the Time Inconsistency of US Monetary Policy.* Economica, Vol. 75, Issue 297, pp. 22-38, February 2008

[13] Szomolányi, K. *Časová konzistencia politiky v monetárnej ekonomike.* Dizertačná práca, 2006

[14] Szomolányi, K. *Podstata a povaha problému časovej konzistencie ekonomickej politiky.* Ekonomika a informatika III. ročník, 2/2005

[15] Szomolányi, K., Lukáčiková, A., Lukáčik, M. *Problém časovej konzistencie ekonomickej politiky, zdanenie kapitálu.* Ekonomika a informatika V. ročník, 1/2007

[16] Szomolányi, K., Lukáčiková, A., Lukáčik, M. *Problém časovej konzistencie ekonomickej politiky, model dlhu.* Ekonomika a informatika VI. ročník, 1/2008

Ladislav Lukáš, Alena Motyčková

Department of Statistics and Operations Research
Faculty of Economics
University of West Bohemia in Pilsen
Husova 11, 30614 Plzeň

e-mail: lukasl@kso.zcu.cz, alena.motyckova@centrum.cz

Numerical Simulation of Interest Rate Using Vasicek and CIR Models

Abstract

Paper presents two frequently used stochastic models for numerical simulation of interest rate. In particular, there are Vasicek's model and Cox-Ingersoll-Ross's model. Both models are formulated by stochastic differential equations. The discretization of discussed SDE-s is made by Euler method. The discretized equations are implemented by m-files written in Matlab (MathWorks, Inc.). The numerical experiments obtained are discussed in detail, in order both to compare these often used stochastic models, and to demonstrate their specific features and properties, as well. The paper is focused on numerical aspects of the simulation algorithms, too.

Keywords: *Interest rate, CIR model, Vasicek model, stochastic models, numerical simulation*

1 Introduction

During the last several years an increasing interest was focused upon numerical simulation of stochastic dynamic models in finance and economics. Apart of a large number of models with various topics which are at disposal in literature, there is also a role of increasing power of computers which contributes a lot to the development in this field because providing facilities to accomplish large and massive numerical experiments and simulations. A very typical example of that development trend is quite recognisable in mathematical finance, in the inetrest rate models, in particular. The paper is composed as follows. After a brief introduction, the section two brings a short overview of some basic discrete time approximations of stochastic differential equations (SDE). The section three presents some results of numerical solutions of two well-known models for interest rates based on SDE, the Vasicek's model and Cox-Ingersoll-Ross's model, and gives some comparisons of both. The last section concludes the results and gives notes of ongoing research.

At present, there is a large number of books and papers which are devoted to theoretical foundations, constructions and evaluations of stochastic dynamic models in finance. For the reference, we list just a particular selection – [1],[3],[4],[5] and [7]. The fundamental reference book regarding numerical solution of SDE is [2], and finally, all results of numerical simulations presented in this paper which were produced by Matlab (MathWorks, Inc.) may be found in [6].

2 Descrete time approximation of SDE

We shall consider the simplest time discretization $\{\tau\}_N$ with $0 =< \tau_0 < \tau_1 < \ldots < \tau_N = T$, of a time interval $[0, T]$, where $\delta = T/N$ is an equidistant step size, i.e. $\Delta_n = \tau_{n+1} - \tau_n = \delta, forn = 0, \ldots, N-1$. However, before constructing any approximation scheme, there is reasonable to mention basic types of convergence criteria which are used to measure an accuracy of numerical approximations in some sense. In general, such measures should be concerned with distances between trajectories, i.e. sample paths generated by numerical procedures, and theoretical trajectories of SDE being approximated. It is well-known, that theoretically and practically as well, there is mostly interesting to consider the absolute error of approximation at the final time instant T.

Let us denote $X = \{X_t, t \geq 0\}$ an Ito process

$$X_t = X_0 + \int_0^t a(X_s)ds + \int_0^t b(X_s)dW_s \tag{1}$$

where X_0 is an initial value random variable, the second term represents the drift component and the third one, which is an Ito stochastic integral with respect to the Wiener process $W = \{W_t, t \geq 0\}$, is called the diffusion component. As usual, the process (1) is also called a diffusion process, and it may be written in differential form called an Ito stochastic differential equation (SDE)

$$dX_t = a(X_t)dt + b(X_t)dW_t \tag{2}$$

A discrete time approximation of the X with respect to $\{\tau\}_N$ will be denoted $Y = \{Y_n, t \in \{\tau\}_N\}$. Hence, the most intresting approximation quantity to consider will be $E(|X_T - Y_N|)$, which is the expectation of absolute error at the final time T.

Further, two convergence criteria can be defined

1. *strong convergence*: an approximation process Y converges in the strong sense with order $\gamma \in]0, +\infty]$ to X if exists $\delta_0 > 0,\ K < +\infty$, such that $E(|X_T - Y_N|) \leq K\delta^\gamma$ holds for any time discretization with maximum step size $\delta \in]0, \delta_0[$.
2. *weak convergence*: an approximation process Y converges in the weak sense with order $\beta \in]0, +\infty]$ to X if exists $\delta_0 > 0, K < +\infty$, such that $\left|E(g(X_T)) - E(g(Y_N))\right| \leq K\delta^\beta$ holds for any time discretization with maximum step size $\delta \in]0, \delta_0[$, for the given function $g(x)$.

Usually, the weak convergence concerns with the first two moments $E(X_T)$ and $E((X_T)^2)$ only, which means that Y is to provide a good approximation of corresponding characteristics of probability distribution of X_T rather than a closeness of sample paths to theoretic ones.

Numerical schemes used for approximating solutions of SDE are constructed either by heuristics linked with numerical schemes used for solving deterministic ODE or by truncations of stochastic Taylor formulae, which is more proper approach. The most interesting instrument is so called Wagner-Platten expansion formula, which allows a function $f(X_t)$ of an Ito process to be expanded around random variable $f(X_{t_0})$ into series containing multiple stochastic integrals with constant integrands and the remainder term, for any $t \in [t_0, T]$. For example, the 2-nd order formula takes the following form

$$f(X_t) = f(X_{t_0}) + c_1(X_{t_0}) \int_{t_0}^{t} ds + c_2(X_{t_0}) \int_{t_0}^{t} dW_s + c_3(X_{t_0}) \int_{t_0}^{t} \int_{t_0}^{s_2} dW_{s_1} dW_{s_2} + R, \quad (3)$$

where the coefficients $c_i(x), i = 1, 2, 3$ are evaluated at X_{t_0} and expressed by given functions $a(x), b(x), f(x)$ and their derivatives

$$c_1(x) = a(x)f^(1)(x){+}1/2(b(x))2f^{(2)}(x),\ c_2(x) = b(x)f^{(1)}(x),\ c_3(x) = b(x)(b(x)f^{(2)}(x){+}b^{(1)}(x)f^{(1)}(x)),$$

with $b^{(1)}(x) = db/dx,\ f^{(1)}(x) = df/dx$ and $f^{(2)}(x) = d^2f/dx^2$.

The simplest and the most popular strong Taylor approximation of difussion SDE (2) is the Euler approximation

$$Y_{n+1} = Y_n + a(Y_n)\Delta_n + b(Y_n)\Delta W_n, \quad (4)$$

for $n = 0, 1, \ldots, N - 1$, and given initial value $Y_0 = x_0$,

where ΔW_n are Wiener increments, which are independent $N(0, \Delta_n)$ random variables, i.e. having means and variances $E(\Delta W_n) = 0$ and $D(\Delta W_n) = E((\Delta W_n^2) - (E(\Delta W_n))^2) = E((\Delta W_n^2)) = \Delta_n$, respectively. The approximation Y generated by scheme (4) has strong convergence rate $\gamma = 0.5$if the functions $a(x)$ and $b(x)$ obey usual Lipschitz conditions.

Another well-known approximation derived from (3) with higher convergence rate $\gamma = 1.0$, in particular, is the Milstein scheme

$$Y_{n+1} = Y_n + a(Y_n)\Delta_n + b(Y_n)\Delta W_n + 1/2b(Y_n)b^{(1)}(Y_n)((\Delta W_n)^2 - \Delta_n). \quad (5)$$

Note, that the double Wiener integral has appeared in (3), which yields in this case $\int_{\tau_n}^{\tau_{n+1}} \int_{\tau_n}^{s_2} dW_{s_1} dW_{s_2} = 1/2((\Delta W_n)^2 - \Delta_n)$. Since the scheme (5) contains the first derivative of difussion coefficient b(x) which has to be known, it increases its regularity conditions, and hence may cause some troubles in applications.

In order to avoid the use of derivatives of drift and difussion coefficients in general, there are developed so called Runge-Kutta schemes. The Runge-Kutta scheme which is tightly linked to the Milstein scheme (5) and has $\gamma = 1.0$, too, is following

$$Y_{n+1} = Y_n{+}a(Y_n)\Delta_n{+}b(Y_n)\Delta W_n{+}1/2\{b(Y_n{+}b(Y_n)(\Delta_n)^{1/2}){-}{-}b(Y_n)\}\{(\Delta W_n)^2{-}{-}\Delta_n\}(\Delta_n)^{-1/2}. \quad (6)$$

Another strong approximations, e.g. multi-step schemes or implicit methods, are known, too – for more details see [2].

Weak approximations of SDE (2) can be constructed by truncating W-P formula, too. However, they provide approximations of some functionals of the Ito process, only. Hence, it is sufficient to use an initial value Y_0 just approximating X_0 in some sense, and the Wiener increments ΔW_n replace by simpler random incrementsΔZ_n having similar moment characteristics.

The simplest useful weak approximation having weak convergence rate $\beta = 1.0$ is the weak Euler scheme

$$Y_{n+1} = Y_n + a(Y_n)\Delta_n + b(Y_n)\Delta Z_n, \tag{7}$$

for $n = 0, 1, \ldots, N-1$, and given initial value $Y_0 = X_0$,

where ΔZ_n are independent random increments generated by two-point random variable taking values $\pm(\Delta_n)^{1/2}$ with equal probabilities, i.e. $P\{\Delta Z_n = \pm(\Delta_n)^{1/2}\} = 0.5$.

When we are interested in approximation of $E(g(X_T))$ only, we can use an interesting extrapolation method with $\beta = 2.0$, which combines two estimations obtained for original step size δ and doubled one 2δ, and takes the form

$$E(g(X_T)) \approx 2E(g(Y^{\delta}(T))) - E(g(Y^{2\delta}(T))). \tag{8}$$

Lot of schemes of higher strong or weak order, Runge-Kutta, implicit ones, are already known, but the search for numerical effective and more accurate approximations is still open.

3 Vasicek's model and Cox-Ingersoll-Ross's model

The theoretical formulations of these models are well-known and given in [1],[3],[4],[5] and [7]. The work [6] contains both an overview of interest rate modeling methods and the own numerical experiments and results, and provides a background for this section.

3.1 Introduction to interest rate modeling

The process of short term interest rate is specified by SDE

$$dr_t = \alpha\,(\mu - r_t)\,dt + \sigma r_t^{\gamma} dz_t, \tag{9}$$

where $\alpha\,(\mu - r_t)\,dt$ represents a drift (deterministic element), which ensures mean reversion. The volatility of the process (stochastic element) is $\sigma r_t^{\gamma} dz_t$. The factors μ, α and σ are positive constants, and notations are following

1. μ is a mean reversion level,
2. α is mean reversion rate,
3. σ is the spot interest rate volatility,
4. r_t is the spot interest rate,
5. z_t is Wiener process – note: we introduce this notation instead of W_t used in the previous section, since z_t is almost traditional one in financial applications.

Notice that the exponent γ being applied on current interest rate r_t may cause various influences on the stochastic element of the process. Hence, by changing the value of γ we can get special versions of original SDE (9).

3.2 Vasicek's model

This is one the first known and the most widely-used interest rate model, published by Oldrich Vasicek in 1977 in Journal of Financial Economics. It is based on Ornstein-Uhlenbeck process for the spot interest rate r_t

$$dr_t = \alpha \left(\mu - r_t\right) dt + \sigma dz_t. \tag{10}$$

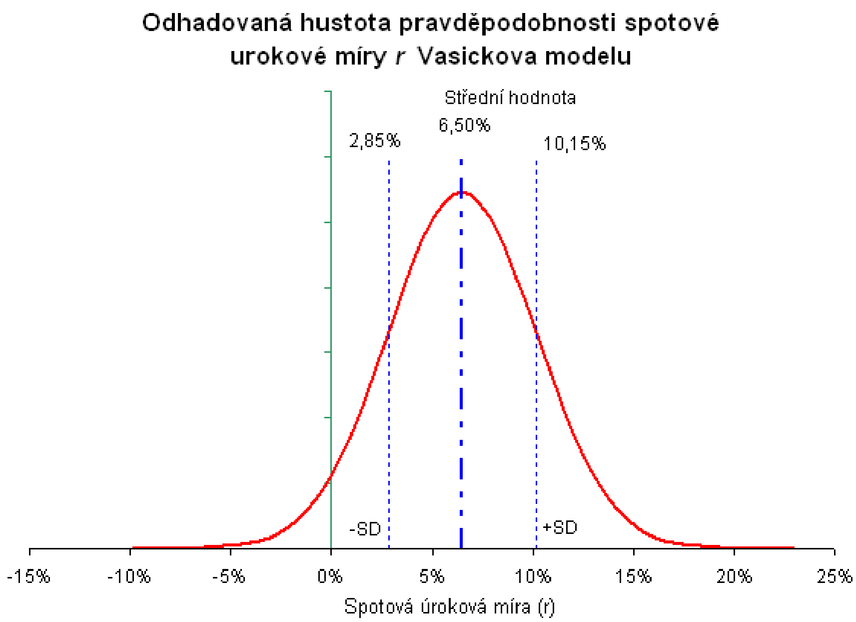


Figure 1: Estimated density function – Vasicek model, $\alpha = 0.15$, $\mu = 6\%$, $\sigma = 2\%$

The parameter γ becomes zero in the Vasicek's model, i.e. $\gamma = 0$. Hence, the equation (9) corresponding with (10) takes the form

$$dr_t = \alpha(\mu - r_t)dt + \sigma r_t^0 dz_t$$

where multiplicative neutrality of the factor r_t^0 in the volatility term allows the spot interest rate to become negative. This main disadvantage of Vasicek's model is shown in the Graph 1, which presents the estimated probability density function. Probability distribution of long term interest rate simulated by Vasicek's model is normal distribution with the density

$$f_r = \sqrt{\frac{\alpha}{\pi}\frac{1}{\sigma^2}} e^{\frac{\alpha(r-\mu)^2}{\sigma^2}}$$

3.3 Cox-Ingersoll-Ross model

The second most common interest rate model is Cox-Ingersoll-Ross (CIR) model. It is a kind of model similar to Vasicek's model, but with $r_t^{1/2}$, i.e.$\gamma = 1/2$, in the volatility of the process

$$dr_t = \alpha(\mu - r_t)dt + \sigma r_t^{1/2} dz_t$$

This factor ensures that rates do not go negative not even for a large volatility of spot interest rates σ. Probability distribution of long term interest rate simulated with CIR model is gamma distribution with the density function

$$f(r) = \frac{\left(\frac{2\alpha}{\sigma^2}\right)}{\Gamma(k)} r^{k-1} e^{-\frac{2\alpha r}{\sigma^2}} = \left(\frac{2\alpha}{\sigma^2}\right)^k r^{k-1} e^{-\frac{2\alpha r}{\sigma^2} - \ln(\Gamma(k))}.$$

Figure 2: Estimated density function – CIR model, $\alpha = 0.15$, $\mu = 6\%$, $\sigma = 5\%$

3.4 Vasicek and CIR model simulations.

When simulating continuous time processes of the spot interest rates by solving SDE (9) we use strong Euler scheme (4), which takes the particular form

$$r_{t+\Delta t} = r_t + \alpha(\mu - r_t)\Delta t + \sigma r_t^{\gamma} \sqrt{\Delta t} \varepsilon_{t+\Delta t}, \tag{11}$$

where the Wiener increments $\Delta W_n = \Delta z_t = (\Delta t)^{1/2}\varepsilon_{t+\Delta t}, \varepsilon_{t+\Delta t} \in N(0, 1)$, and time steps $\Delta_n = \Delta t = \delta$.

Sample paths of the spot interest rates presented on Graphs 3 and 4 were simulated by the scheme (11). The simulations were run with values:$\alpha = 0.03$,$\mu = 6\%$,$\sigma = 8\%$ for Vasicek's model, $\sigma = 12\%$for CIR model, and with initial value $r_0 = 7\%$ for both. Since we assumed the simulation period 1 year, taking 250 business days, we got the time step $\Delta t = 1/250 = 0.004$. The Graphs 3 and 4 besides illustrating possibility the simulated r_t to reach unrealistic negative values by Vasicek's model, they show also its another typical property that the simulated interest rates reach more extreme values than those simulated by CIR model.

This property is illustrated on the Graphs 5 and 6 presenting frequency histograms of both models and documenting either uni- or multi-modality of sample distributions, respectively. However, the negative values of the simulated r_t by Vasicek's model need not appear in any case, which is well documented on the Graphs 7 and 8. In order to show better this weak point of Vasicek's model in comparison with the CIR one we run two numerical experiments with different simulation horizons (1 year and 2 years) with 20 sample paths in each. The results are depicted on the Graphs 9 and 10.

Let's complete our simulations with Graphs 11 – 14. First, the Graphs 11 and 12 show long term simulations, in particular for 10 years. Again, we observe already mentioned and known major disadvantages of Vasicek's model, that simulated values of spot interest rate r_t take much more extreme values and may also reach negative ones, when being compared with corresponding results yielded by CIR model. One of the most important feature of the CIR model is that r_t never decrease bellow zero. Further, we may also observe less fluctuations of simulated r_t by this model in average, but not the extreme values.

The last two graphs, i.e. the Graph 13 and 14, show large number of simulations we run with our Matlab m-files. Both graphs illustrate simulations for one year period with 1000 sample paths each, and allow us to confirm the main differences between Vasicek and CIR model regarding appearance of negative values and extremities of interest rates simulated.

4 Conclusions

Numerical simulations of interest rates using SDE models were presented. The short overview of numerical schemes for approximate solving SDE was focused on the most popular methods both of strong and weak type of convergence.

The gained numerical results have demonstrated the applicability and flexibility of the developed Matlab m-files to generate large number of sample paths of two most popular interest rate stochastic models – the Vasicek model and the CIR model, respectively.

However, still a lot of theoretical work and numerical experiments are to be done both on the field of building stochastic models in mathematical finance and development of effective numerical schemes for solution stochastic differential equations.

Acknowledgement: The first author acknowledges the partial support of this research work both by the grant No. 402/07/1228 of the Grant Agency of Czech Republic and by funds of the project LC06075 of the Ministry of Education of Czech Republic.

References

[1] JAMES, J., WEBER, N. *Interest rate modelling.* Wiley & Sons, 2000. ISBN 0-471-97523-0

[2] KLOEDEN, P.E., PLATEN, E. *Numerical Solution of Stochastic Differential Equations.* Springer, 1999, 3-e. ISBN 3-540-54062-8

[3] KOROLJUK, V.S., PORTENKO, N.I., SKOROCHOD A.V., TURBIN, A.F. *Spravočnik po teorii věrojatnostej i matematičeskoj statistice.* (*Handbook of Probability Theory and Mathematical Statistics*). Nauka, 1985, Moscow

[4] LONDON, J. *Modeling Derivatives in C++* . Hoboken, 2005. ISBN 0-471-65464-7

[5] LYUU, YUH-DAUH *Financial Engineering and Computation: Principles, Mathematics, Algorithms.* e-book, Cambridge University Press, 2001

[6] MOTYČKOVÁ, A. *Stochastické modely úrokových měr.* (*Stochastic models of interest rates*). Master Thesis, Univ. of West Bohemia in Pilsen, Faculty of Economics, Pilsen, 2008

[7] SVOBODA, S. *Interest Rate Modelling.* Palgrave Macmillan. 2004. ISBN: 1-4039-3470-3

Appendix – results of numerical simulations

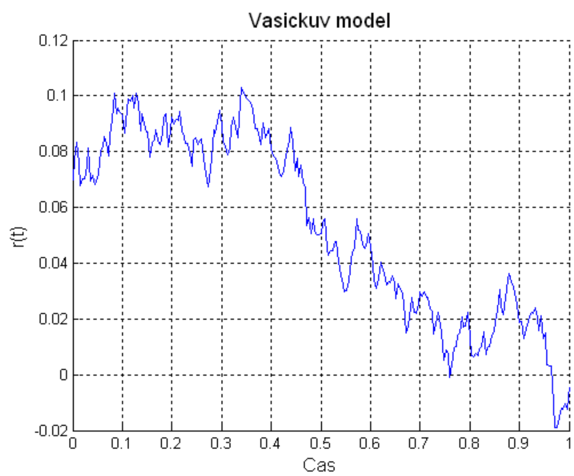


Figure 3: Sample path for Vasicek model – blue color

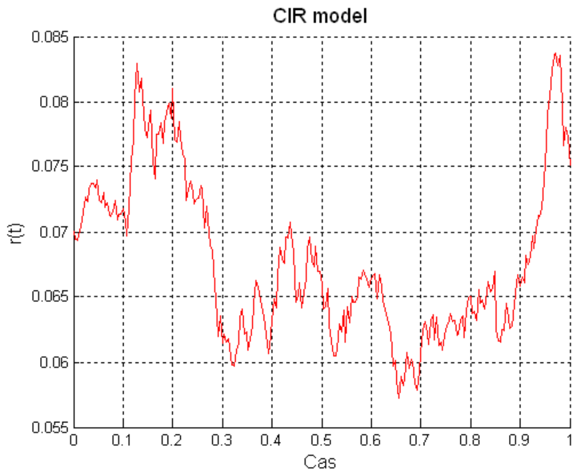


Figure 4: Sample path for CIR model – red color

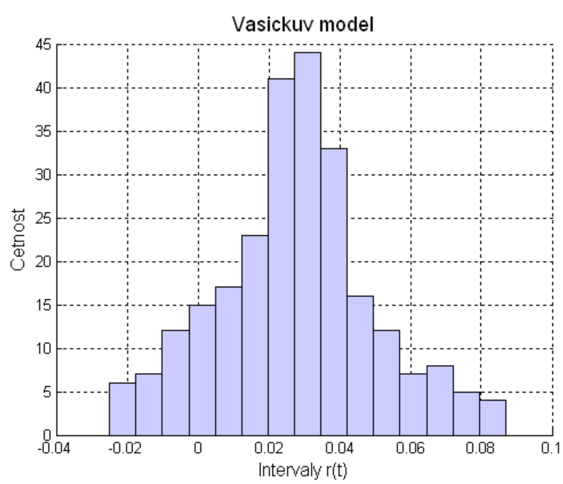


Figure 5: Vasicek model – Frequency histogram

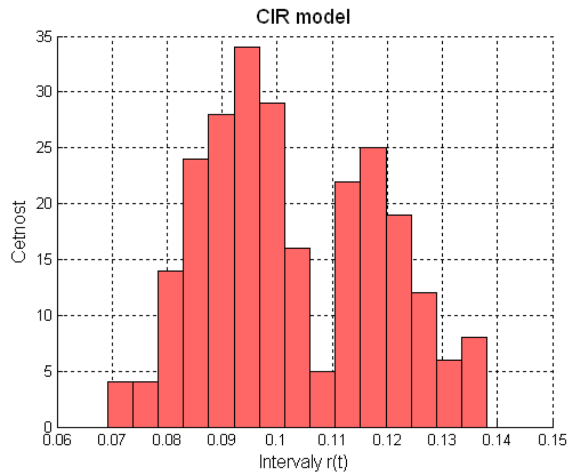


Figure 6: CIR model – Frequency histogram

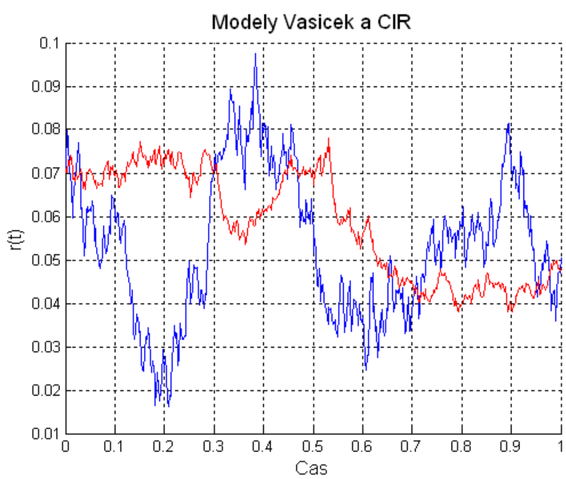


Figure 7: Sample paths for Vasicek and CIR model

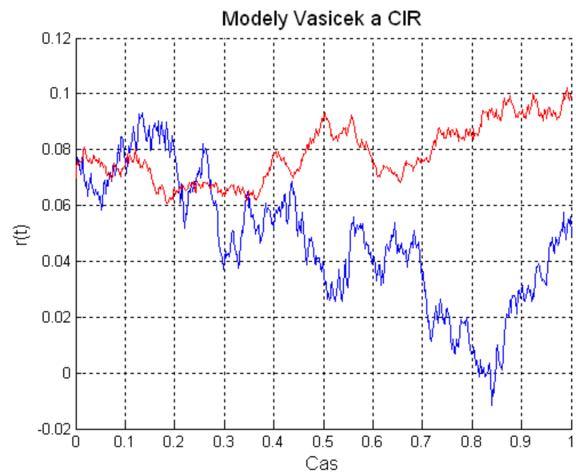


Figure 8: Sample paths for Vasicek and CIR model

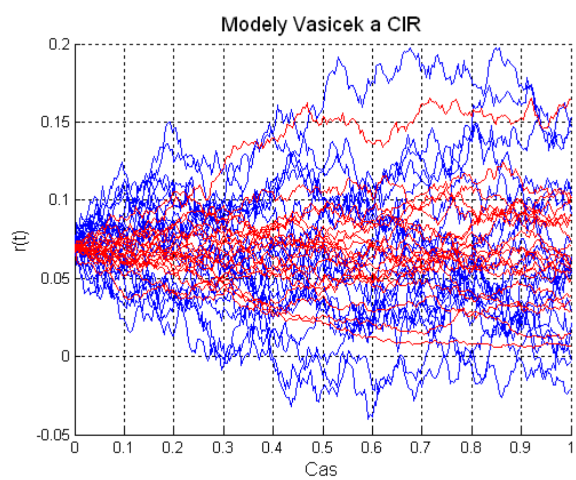


Figure 9: 20 sample paths for both Vasicek and CIR model, 1 year (250 time steps)

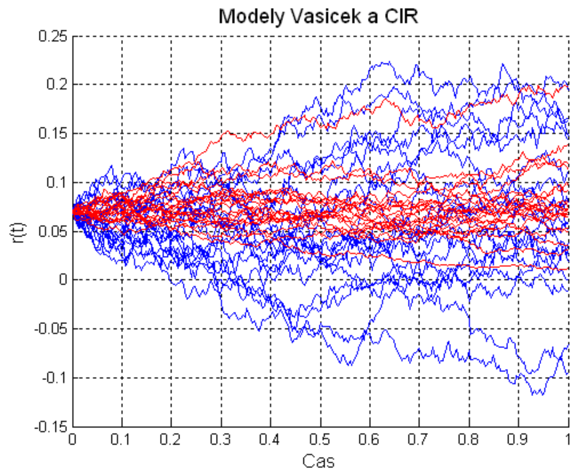


Figure 10: 20 sample paths for both Vasicek and CIR model, 2 years (500 time steps)

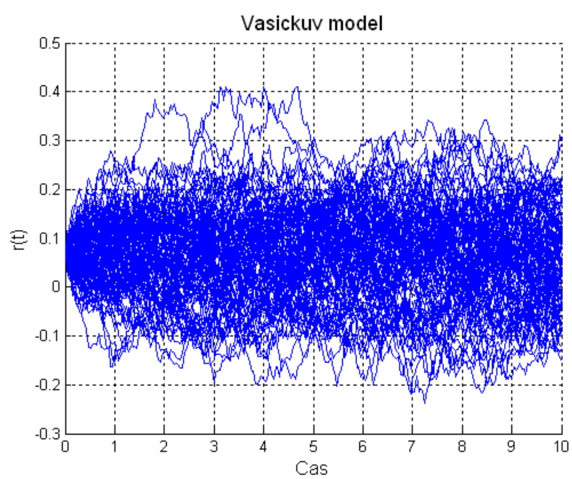


Figure 11: 100 sample paths for Vasicek model, 10 years (2500 time steps)

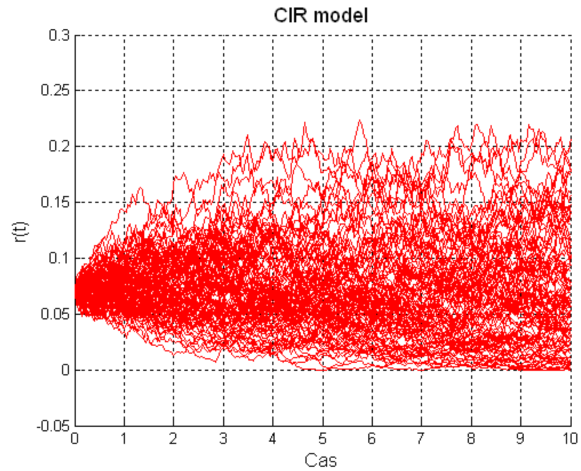


Figure 12: 100 sample paths for CIR model, 10 years (2500 time steps

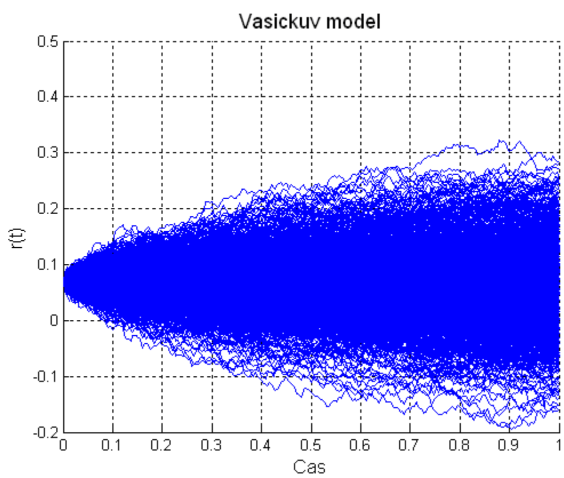


Figure 13: 1000 sample paths for Vasicek model, 1 year (250 time steps)

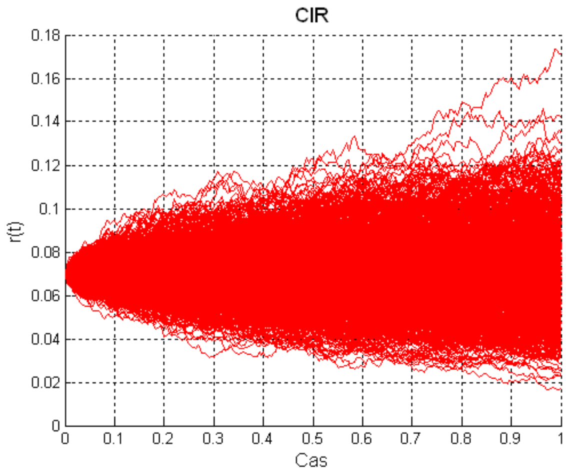


Figure 14: 1000 sample paths for CIR model, 1 year (250 time steps)

Veronika Mit'ková

Institute of Economics
Faculty of Social and Economic Sciences
Comenius University in Bratislava
Odbojárov 10/a, 82005 Bratislava

e-mail: mitkova@fses.uniba.sk

Modeling the Foreign Trade in the Applied CGE Model of Slovakia with Emphasis to the Automotive Sector.[1]

Abstract

The goal of this paper is to analyze the impact of value changes of trade function parameters in applied computable general equilibrium model of Slovakia with emphasis to the automotive sector. This sector is selected as the analyzed one since the Slovak economy is strongly influenced by dynamics in this sector in the presence. The model construction allows analyzing the impact of changes in a selected sector of economy aspect to the aggregated sector of other sectors. It is supposed an imperfect substitutability between imports and domestic output sold domestically and it is captured by a CES aggregation function in which the composite commodity that is supplied domestically is "produced" by domestic and imported commodities. A composite supply of import and domestic use of domestic output is modeled by Armington function. Domestic output as a function of export and domestic use of domestic output is modeled by an output transformation (CET) function. It is supposed an imperfect transformability between domestic output for exports and domestic sales. The experiments with the value changes of elasticity of substitution, elasticity of transformation, shift parameter for composite supply and shift parameter for output transformation were made in order to analyze the sensitivity and stability of the model to the foreign trade function parameters value changes.

Keywords: *foreign trade, applied CGE model, automotive*

1 Introduction and Incentives

The research presented in this paper was inspired by work of Kazybayeva and Tanyeri – Abur [1] about trade policy adjustments in general equilibrium analysis where the parameters of trade functions were changed and the impact to the model results were observed. They conclude that the qualitative results remain the same under a wide range of different CES and CET parameters; thus the interpretations and conclusions apply to the entire spectrum of parameter variation in the model.

[1] This research was supported by Comenius University grant scheme for young scientists No. UK/6/2008.

McDaniel's and Balistreri's [3] research about influence of Armington function parameters to the model results begins with note that a few robust findings emerge from the econometric literature. The first one is that long-run estimates are higher than short-run estimates, the second that more disaggregate analyses find higher elasticities, and the last one that reduced-form time-series analyses find lower elasticities relative to cross-sectional studies. For the experiments they used the steady-state equilibrium as the baseline. The model was greatly simplified by aggregating up to include only three goods and four regions. The standard nested-Armington structure was used in which the lower-tier substitution elasticity (between imported varieties) is twice the substitution elasticity between imported and domestic varieties. This model is similar to the model used in our early work [4] and the goal of this research is to test the model to the sensitivity to the impact of the changes in elasticity of substitution and elasticity of transformation.

2 Modeling the Foreign Trade

A standard computable general equilibrium model (CGE) captures all financial or physical flows in the economy. The database of the model we are working with consists of a social accounting matrix (SAM), which synthesizes these flows on activity, commodity, household, factor, institution, saving and investment, tax and rest of the world accounts. From the mathematical point of view, the model is a set of simultaneous equations describing the agents' behavior; some of them are nonlinear. The first order nonlinear conditions capture the behavior on the production and consumption side; it means that behavior is derived by maximizing profit and utility. The equations include conditions which must be fulfilled for the system, not necessary for individual agents. The equations include factor and commodity markets and such macroeconomic aggregates as savings and investments, government and rest of the world.

We focus on the impact of the value changes in elasticity of substitution and elasticity of transformation to the stability of the applied computable general equilibrium model in this work. The model was developed for analyzing the impact of policy changes to Slovakian automotive sector which was selected for analysis since the Slovak economy is strongly influenced by dynamics in this sector in the past years [4].

For more than twenty years before there was a principal question in the modeling agenda how to deal with export, import, balance of payments and balance of trade in the CGE models. After trying various approaches, a general consensus was reached by adopting the Armington assumption. This assumption regards imperfect substitutability and it is extended to the modeling of exports as well. The use of the Armington function in trade differs from the standard neoclassical trade model in which all goods are tradable and all domestically produced goods are perfectly substitutable with imports. According Zhang [5] the standard treatment is not fully compatible with the empirical results: the domestic relative price of tradables is fully determined by world prices. Neoclassical models result in the full transmission of world price changes and in extreme specialization in production. In the Armington framework, the economy is less responsive to world price changes, thus dampening the move toward specialization [5]. Also Khan [2] quotes that the most common approach now is to specify sectoral constant elasticity of substitution (CES) import demand functions, export transformation functions that assume constant elasticity of transformation (CET) and aggregation functions based on these.

Armington elasticities specify the degrees of substitution in demand between similar products produced in different countries. They are critical parameters which, along with model structure, data and other parameters, determine the results of policy experiments.

3 Modeling the Supply Side

The supply side is modeled as composite supply of import and domestic use of domestic output by Armington function.

$$QQ_c = \alpha q_c \cdot \left[\delta_c^q \cdot QM_c^{-\rho_c^q} + (1-\delta_c^q) \cdot QD_c^{-\rho_c^q}\right]^{\frac{-1}{\rho_c^q}}$$

Equation 1

It is supposed an imperfect substitutability between imports and domestic output sol domestically and it is captured by a CES aggregation function in which the composite commodity that is supplied domestically is "produced" by domestic and imported commodities, and enters the function as inputs. The optimal mix between imports and domestic product is given by Equation 1.

$$\frac{QM_c}{QD_c} = \left(\frac{PD_c}{PM_c} \cdot \frac{\delta_c^q}{1-\delta_c^q}\right)^{\frac{1}{1+\rho_c^q}}$$

Equation 2

The share parameter of composite supply function is derived from Equation 1 and a relationship between elasticity of substitution and an exponent of composite supply $\rho_c^q = \frac{1}{\sigma_c^q - 1}$. Now it is possible to derive the composite supply function form:

$$QQ_c = \alpha q_c \cdot \left\{ \frac{1}{1 + \frac{PD_c}{PM_c} \cdot \left(\frac{QD_c}{QM_c}\right)^{\left(\frac{\sigma_c^q}{\sigma_c^q - 1}\right)}} \cdot QM_c^{-\frac{1}{\sigma_c^q - 1}} + \left[1 - \frac{1}{1 + \frac{PD_c}{PM_c} \cdot \left(\frac{QD_c}{QM_c}\right)^{\left(\frac{\sigma_c^q}{\sigma_c^q - 1}\right)}}\right] \cdot QD_c^{-\frac{1}{\sigma_c^q - 1}} \right\}^{(1+\sigma_c^q)}$$

Equation 3

where:

αq_c shift parameter for composite supply function,

δ_c^q share parameter for composite supply function,

σ_c^q the elasticity of substitution for composite supply function,

ρ_c^q exponent $(-1 < \rho_c^t < \infty)$ for composite supply function,

PD_c domestic price of domestic output,

PM_c import price (in domestic currency),

QD_c quantity of domestic output sold domestically,

QM_c quantity of imports,

QQ_c quantity supplied to domestic commodity demanders (composite supply).

4 Modeling the Domestic Output

Domestic output as a function of export and domestic use of domestic output is modeled by an output transformation (CET) function.

$$QX_C = \alpha t_c \cdot \left[\delta_c^t \cdot QE_c^{\rho_c^t} + \left(1-\delta_c^t\right) \cdot QD_c^{\rho_c^t}\right]^{\frac{1}{\rho_c^t}}$$

Equation 4

It is supposed an imperfect transformability between domestic output for exports and domestic sales as a parallel to imperfect substitutability between imports and domestic output sold domestically used in Armington function. The CET function is identical to CES function except for negative elasticities of substitution. In economic terms, the difference between the Armington and CET functions is that the arguments in the first one are inputs and in the second one are outputs.[2] The optimal mix between exports and domestic sales is given by Equation 5.

$$\frac{QE_c}{QD_c} = \left(\frac{PE_c}{PD_c} \cdot \frac{1-\delta_c^t}{\delta_c^t}\right)^{\frac{1}{\rho_c^t - 1}}$$

Equation 5

The share parameter for output transformation (CET) function is derived from Equation 5 and a relationship between elasticity of transformation and an exponent of output transformation supply$\rho_c^t = \frac{1}{\sigma_c^t + 1}$. Now it is possible to derive the output transformation function form:

$$QX_c = \alpha t_c \cdot \left\{ \frac{1}{1 + \frac{PD_c}{PE_c} \cdot \left(\frac{QE_c}{QD_c}\right)^{\frac{\sigma_c^t + 2}{\sigma_c^t + 1}}} \cdot QE_c^{\frac{1}{\sigma_c^t + 1}} + \left[1 - \frac{1}{1 + \frac{PD_c}{PE_c} \cdot \left(\frac{QE_c}{QD_c}\right)^{\frac{\sigma_c^t + 2}{\sigma_c^t + 1}}}\right] \cdot QD_c^{\frac{1}{\sigma_c^t + 1}} \right\}^{\left(\sigma_c^t + 1\right)}$$

Equation 5

where:

αt_c shift parameter for output transformation function,

δ_c^t share parameter for output transformation function,

σ_c^t elasticity of transformation for output transformation function,

ρ_c^t exponent ($1 < \rho_c^t < \infty$) for output transformation function,

QX_c quantity of domestic output,

QE_c quantity of exports,

[2] There is a clear difference between equations for import demand and export supply: the quantity demanded of the imported commodity is inversely related to the import price and the quantity supplied of the exported commodity is directly related to the export price.

PE_c export price (in domestic currency).

The values of elasticities $\sigma_c^q$and σ_c^t for Armington and constant elasticity transformation functions determine the intensity of the substitution between domestic supply and imports (Armington CES function), and the intensity of the transformation between domestic production and exports (CET function). To evaluate how sensitive the model is to differences in these trade elasticities an experiment with export price increase by 5% was carried out applying series of different elasticity variations as follows in the next chapter.

5 Experiments

Uniform values of elasticities of substitution and transformation were adopted as base for the automotive sector and sector of the aggregated rest of the commodities in this applied CGE model. For value of elasticity of substitution in Armington function was set up 0.7 and for value of elasticity of transformation in CET function was set up 2.0.

In the first run the value of the elasticity of substitution was gradually changed from 0.5 to 2.0 in the sector of automotives while the value in the sector of the aggregated rest commodities remained fixed. The results are shown in the Table 1.

Table 1: Percentage change in selected variables with elasticity of substitution changes

sigmat(auto,rest)=2.0						
sigmaq(rest)=0.7, sigmaq(auto)=	**base**	**0,500**	**0,700**	**0,990**	**1,500**	**2,000**
production of automotives	540 417	1,62%	1,48%	1,34%	1,21%	1,14%
production of the rest of the commodities	2 550 234	-0,23%	-0,23%	-0,24%	-0,24%	-0,24%
intermediate consumption	1 879 977	0,14%	0,10%	0,07%	0,04%	0,02%
GDP	1 210 674	0,56%	0,58%	0,59%	0,61%	0,62%
households consumption of automotives	61 162	-0,19%	-0,43%	-0,65%	-0,87%	-1,00%
households consumption of the rest of the commodities	846 772	0,21%	0,22%	0,23%	0,24%	0,25%
government consumption of automotives	38 069	-0,19%	-0,43%	-0,65%	-0,87%	-1,00%
CPI	1,05	-0,31%	-0,32%	-0,34%	-0,35%	-0,36%
exchange rate	1,00	-1,27%	-1,31%	-1,34%	-1,38%	-1,40%
import of automotives	153 022	-1,27%	-1,31%	-1,34%	-1,38%	-1,40%
export of automotives	248 465	1,62%	1,48%	1,34%	1,21%	1,14%

In the second run the value of the elasticity of transformation was gradually changed from 0.5 to 2.5 in the sector of automotives while the value in the sector of aggregated rest commodities remains fixed again. The results are shown in the Table 2.

Table 2: Percentage change in selected variables with elasticity of transformation changes

sigmaq(auto,rest)=0.7						
sigmat(rest)=2.0, sigmat(auto)=	**base**	**0,500**	**1,000**	**1,500**	**2,000**	**2,500**
production of automotives	540 417	1,85%	1,72%	1,60%	1,48%	1,36%
production of the rest of the commodities	2 550 234	-0,13%	-0,16%	-0,20%	-0,23%	-0,27%
intermediate consumption	1 879 977	0,06%	0,07%	0,09%	0,10%	0,12%
GDP	1 210 674	0,31%	0,40%	0,49%	0,58%	0,66%
households consumption of automotives	61 162	-0,23%	-0,30%	-0,36%	-0,43%	-0,49%
households consumption of the rest of the commodities	846 772	0,12%	0,16%	0,19%	0,22%	0,25%
government consumption of automotives	38 069	-0,23%	-0,30%	-0,36%	-0,43%	-0,49%
CPI	1,05	-0,17%	-0,23%	-0,28%	-0,32%	-0,37%
exchange rate	1,00	-0,70%	-0,91%	-1,11%	-1,31%	-1,49%
import of automotives	153 022	-0,70%	-0,91%	-1,11%	-1,31%	-1,49%
export of automotives	248 465	1,85%	1,72%	1,60%	1,48%	1,36%

The percentage deviations of values of aggregates as domestic consumption on automotives and on aggregated rest of commodities, GDP, government consumption on automotives, domestic and foreign price of automotives, CPI and other aggregates are less than 2% for a wide range of tested values.

6 Conclusion

If we follow the notion that the model is stabile in case the deviations in aggregates are less than 5%, this model is not sensitive on changes in the Armington function elasticity of substitution value and CET function elasticity value according to this experiment. We can conclude that the model is appropriate for foreign trade policy experiments. The changes in both cases were acceptable, in limit up to 2%. Notice that the higher elasticity values the lower percentage changes which can be explained by the higher flexibility of the model at these values.

References

[1] Kazybayeva, S., Tanyeri –Abur, A. (2003) Trade Policy Adjustments in Kazakhstan: A General Equilibrium Analysis. Paper presented at the International *Conference Agricultural policy reform and the WTO: where are we heading?* Capri.

[2] Khan, H.A. (2004) Using Macroeconomic Computable General Equilibrium Models for Assessing Poverty Impact of Structural Adjustment Policies. *ADB Institute Discussion Paper* No.12. Tokyo.

[3] McDaniel, C., Balistreri E. (2003) A review of Armington trade substitution elasticities. *conomie internationale* 94-95, p. 301-313.

[4] Mit'ková, V., Mlynarovič, V. (2007) *Modeling Policy Changes with the CGE Model for Open Economy of Slovakia*. International conference: Improforum. esk Budejovice.

[5] Zhang, X.G. (2006) Armington Elasticities and Terms of Trade Effects in Global CGE Models. *Productivity Commission Staff Working Paper*. Melbourne.

Vladimír Mlynarovič

Faculty of Social and Economic Sciences

Comenius Univerity in Bratislava

Odbojarov 10/a, 82005 Bratislava, Slovak Republic

e-mail: vladimir.mlynarovic@fses.uniba.sk

Some Portfolio Insurance Strategies

Abstract

At first the paper shortly characterizes basic classes of such portfolio insurance strategies as constant proportion portfolio insurance (CPPI) and option based or volatility based portfolio insurance that provide the investor an ability to limit downside risk while allowing some participation in upside markets. The paper presents CPPI as the method for dynamic asset allocation over time where investor chooses a floor equal to the lowest guaranteed value of the portfolio, computes the cushion which is equal to the excess of the portfolio value over the floor and determines amount allocated to the risky asset by multiplying the cushion by the predetermined multiplier. Then some extensions of discrete CPPI methods that introduce risk budget, a stop loss rule, locking of the guaranteed value, the asset management fee and risk measures in the multiplier are presented and illustrated. As the result user procedures in Excel environment that automatize the process of guaranteed strategies construction were developed. Theoretical guaranteed strategies are finally modified to operational strategies on the base of transaction costs implementation. Daily historical data were used to test the strategies and to compare them with the standard CPPI method and some naìve investment strategies.

Keywords: *portfolio insurance, CPPI, multipliers, lock in value*

1 Introduction

Portfolio insurance is designed to give to an investor the possibility to limit downside risk while allowing some participation in upside markets. Such approaches allow the investor to recover a given percentage of the initial investment at maturity, in particular in falling markets. There exists various portfolio insurance models, among them the Constant Proportion portfolio Insurance (CPPI), the Option Based Portfolio Insurance (OBPI) and Volatility Based Portfolio Insurance (VBPI).

The CPPI was introduced by Perold (1986) and Perold and Sharpe (1988) for fixed income instruments and Black and Jones (1987) for equity instruments. The method has been analyzed in Black and Rouhani (1989) and Black and Perold (1992). This method is based on a particular strategy to allocate assets dynamically over time. The investor starts by choosing a floor equal to the lowest acceptable value of the portfolio. Then he computes the cushion which is equal to the excess of the portfolio value over the floor. Finally, the amount invested to the risky asset, exposure, is determined by multiplying the cushion by a predetermined multiplier. The remaining funds are invested in the reserve asset, e.g. a liquid money market instrument. The method is based on the following mathematical background.

The portfolio manager is assumed to invest in two basic assets: a money market tool, denoted by B, and a portfolio of traded assets such as a composite index, denoted by S. The period of time considered is $[0, T]$. The strategy is self-financing. The values of the riskless asset B and risky asset S evolve according to

$$dB_t = B_t r\, dt, \quad dS_t = S_t\, [\mu\, dt + \sigma\, dW_t]$$

where r is the deterministic interest rate, W_t is a standard Brownian motion and μ and σ are positive constants.

The CPPI method consists of managing a dynamic portfolio so that its value is above a floor P_t at any time t. The value of the floor gives the dynamic insured amount. It is assumed to evolve according to

$$dP_t = P_t r\, dt$$

Obviously, the initial floor P_0 is less than the initial portfolio value V_0. The difference $C_0 = V_0 - P_0$ is called the cushion. Its value at any time is given by

$$C_t = V_t - P_t$$

Denote by E_t the exposure, which is the total amount invested in the risky asset. The standard CPPI method consists of letting $E_t = mC_t$, where m is a constant called multiplier. The interesting case is when $m > 1$, that is, when the payoff function is convex.

The OBPI was introduced by Leland and Rubinstein (1976) and analyzed by Martellini, Smisek and Goltz (2005).The method consists basically of purchasing, for simplicity, one share of the asset S and one share of European put option on S with maturity T and exercise price K. Thus, the portfolio value VO is given at the terminal date by

$$V_T^O = S_T + \max\{K - S_T,\ 0\}$$

which is also

$$V_T^O = K + \max\{S_T - K,\ 0\}$$

due to Put/Call parity. This relation shows that the insured amount at maturity is the exercise price K. The value V_t^0 of this portfolio at any time t in the period $[0, T]$ is

$$V_t^O = S_t + P(t, S_t, K) = Ke^{-r(T-t)} + C(t, S_t, K)$$

where $P(t, S_t, K)$ and $C(t, S_t, K)$ are the Black – Scholes values of the European put and call.

The volatility based portfolio insurance (VBPI), introduced in Temporis Presentation (2008), is a systematic investment process, where exposure to equity is adjusted routinely in order to maintain the volatility of the fund close to pre-determined target volatility levels. The guarantee mechanism is based on series of guaranteed bond and at maturity of the bond investor will receive the higher between:

- the final net asset value of the fund,
- the guaranteed net asset value determined on the issued of the bond

It means that each investment is (at least) guaranteed at the maturity. From a view point of a strategic allocation assets of the fund are invested in long – term – assets and/or short term assets (money market tools) according to a specific volatility based risk model. The exposure to the long term assets will be routinely adjusted depending on the prevailing historical volatility of long-term asset and pre-defined target volatility levels which become more conservative as the investment horizon approaches:

- at inception, assets are mainly invested in long term assets to boost returns,
- near to the investment horizon, assets are fully invested in short-term assets.

In a tactical asset allocation, asset mix within long-term assets can be actively managed to increase the exposure to the long-term assets and maximize the expected return for a given target volatility level.

Levels of target volatility on observation date t, $\sigma_{tB}, t = 0, 1, \ldots, T$, are defined at inception, where

$$\sigma_{tB} \to 0,\ as\, t \to T, \quad \sigma_{TB} = 0$$

The fund manager determines the historical volatility σ_{tS} of the long-term asset and the exposure E_t to the risky (long-term) asset on observation date $t,\ t = 0, 1, \ldots, T$, can be then determined, for example, by one of the following approaches:

a) by a simple approach suggested in the Temporis presentation (2008), where

$$E_t = \frac{\sigma_{tB}}{\sigma_{tS}}, \quad E_t \leq h$$

b) by a goal programming approach, where the following problems are solved

$$\min \quad p_t + n_t$$

subject to

$$E_t^2 \sigma_{tS}^2 + (1 - E_t)^2 \sigma_{tM}^2 + 2E_t(1 - E_t)\sigma_{tSM} + n_t - p_t = \sigma_{tB}^2$$

$$p_t n_t = 0, \quad p_t \geq 0, \;\; n_t \geq 0, \quad 0 \leq E_t \leq h$$

where p_t and n_t are deviation variables, σ_{tM} is the historical volatility of the money market tool (riskless asset) and σ_{tSM} is the historical covariance between risky asset and riskless asset on observation date t.

In other words, investment in long-term (risky) asset us routinely adjusted such that the realized volatility of the fund remains close to the target volatility levels and the portion of the assets not invested in long-term assets is allocated to the short-term ("riskless") assets. Among the advantages of such approach there are:

- exposure to long-term assets may temporarily decrease when long-term assets' realized volatility increases (when market conditions are hectic),
- the realized volatility usually converges back to its historical average (well known mean reverting pattern of volatility), which prevents the assets of the fund from being fully and permanently invested in short-term assets.

2 The extensions of the CPPI method

Bertrand and Pringet (2002, 2004) analyze and compare two standard portfolio insurance methods, CPPI a OBPI from a view point such criteria as payoffs at maturity, stochastic or "quantile" dominance of their returns and examine dynamic hedging properties as well. They also compare the performance of two standard methods when the volatility of the stock index is stochastic and provide a quite general formula for the CPPI portfolio value. Boulier and Kanniganti (1995) examine expected performance and risk of various portfolio insurance strategies in a realistic case, where there are constraints on the maximum exposure to the market. The following analysis also concentrates on a realistic, discrete, CPPI method and develops two modification of the basic approach.

2.1. CPPI method with the *stop loss* rule

The following modification of discrete CPPI method takes into account fees for the asset management, which are derived from the starting value of the portfolio at each observation moment, and introduce a lock in of the guaranteed value, floor, when current portfolio value arises, and the stop loss rule for the zero level of the exposure. For the portfolio value at the moment t we have

$$V_t = V_{t-1}(1 + E_{t-1}s_t + (1 - E_t)r_t) + V_0 f \frac{d_t - d_{t-1}}{365}, \quad V_0 = 100$$

where

$$s_t = \frac{S_t}{S_{t-1}} - 1, \quad r_t = \frac{B_t}{B_{t=1}} - 1,$$

and

f –the yearly fee for the asset management, e.g. 2%,
S_0–the official closing price of the risky asset (reference index) on the start date,
S_t –the official closing price of the risky asset (reference index) on observation date t,
B_t–the theoretical price on observation date t of a synthetic zero coupon bond with nominal 1 paid on the terminal date
r_t –the price return of the riskless asset for observation date t,
d_t –the number of calendar day between the start date and observation date t

The actual proportion of the net asset value which is allocated to the risky asset, the exposure, on observation date t equals:

$$E_t = m_t C_t = m_t(V_t - P_t), \quad 0 \leq E_t \leq 1$$

where

$$m_t = \frac{1}{lV_t}$$

while l is the *stop loss*, the percentage loss that the risky asset is allowed to have in 1 day before the exposure in the risky asset drops to 0%, e.g. 15%, and

$$P_t = V_0 B_t \left(k_t + (1+f)^{\frac{d_T = d_t}{365} t} - 1\right)$$

where

$$k_t = \max\left\{k_{t-1}, \frac{V_t}{V_0}\right\}, \quad k_0 = 1$$

is the so called *lock in* value on observation date that locks in the net allocation value, which is guaranteed at maturity. It holds

- if $E_t = 0$ on any observation date, then all subsequent values for exposure will be set to 0%,
- if there is percentage loss higher then the stop loss, the guarantee at maturity would still be

$$k_T = V_0 \max\{1, k_{t-1}\}$$

2.1 CPPI with the risk budget and the multiplier modification

The next modification of the discrete CPPI method, similarly as the previous one, takes into account fees for the asset management, but the value of the fee on each observation date t is derived from the net asset value on observation date $t-1$. It also uses lock in system for the guaranteed floor, but introduce its discrete shifts in both directions (up, down) in dependence from the current net asset value with the restriction that its minimum must not drop under the defined minimal level. The minimal level of the floor is directed through the risk budget or, in other words, through the feasible level of the guaranteed value drop below the starting investment. Observe that an advantage of such kind of the lock in system, which enables limited shifts in the both directions, is that in the case of exposure falls create potential for exploitation of future possible growth on the stock market. The new element of this CPPI method modification consists in introducing the risk in the multiplier construction. The method looks for such multiplier value that exploit the current value of the reduced cushion as much as possible. The reduction size is derived from the risk for period from the observation date to terminal date, or for defined minimal period. As the risk measure the historical CVaR is used.

For the portfolio value on the observation date t we now have

$$V_t = V_{t-1}(1 + E_{t-1}s_t + (1-E_t)r_t + f\frac{d_t - d_{t-1}}{365}), \quad V_0 = 100$$

where

$$s_t = \frac{S_t}{S_{t-1}} - 1, \quad r_t = \frac{B_t}{B_{t=1}} - 1, \quad r_t^{pa} = (1+r_t)^{\frac{365}{d_t - d_{t-1}}}$$

and

B_t – the theoretical price of the riskless asset on observation date t,
r_t^{pa}– annualized riskless asset return on observation date t

The actual proportion of the net asset value which is allocated to the risky asset, the exposure, on observation date t equals:

$$E_t = m_t C_t = m_t(V_t - P_t), \quad d \leq E_t \leq h$$

where

$$P_t = k_t(1 + r_t^{pa})^{-\frac{d_T - d_t}{365}}$$

$$k_t = \begin{cases} \max\{\max\{k_{t-1}, V_0\}(1+q), uV_t\}, & V_t > k_{t-1}(1+q) \\ \max\{k_{t-1}(1-q), V_0(1-b_r), cV_t\}, & V_t < k_{t-1} \\ k_{t-1}, & else \end{cases}$$

while

$$k_0 = V_0(1 - b_r)$$

and

d – the lower bound on the exposure,
h – the upper bound on the exposure,
q – the bound for lock in of the guaranteed value, e.g. 2%,
u – the lock in parameter in “up” direction with possible values 0 or 1
c – the lock in parameter in “down” direction with possible values 0 or 1,
b_r– the risk budget, e.g. 3%.

The multiplier value m_t is computed according to the following iteration procedure:

$$1. m_t^i = 1, \quad i = 1$$

$$2. \omega = \max\{m_t^i(V_t - P_t), 0\} \times CVaR_t\sqrt{\min\{d_T - d_t, o\}}$$

$$3. Q = V_t - \omega$$

$$4. (Q - P_t) \begin{cases} > 0, & m_t^{i+1} = m_t^i + 1, \ i = i + 1, \ \textit{goto step } 2 \\ \leq 0, & m_t = m_t^i, \ \textit{the end} \end{cases}$$

where $CVaR_t$ is the historical conditional VaR on observation date t and o is the minimal length of the period for the risk accounting.

3 Illustrations

In the first illustration we have used the daily data for MSELEMEE commodity index for the period from January 3, 1995 to May 7, 2008 together with daily data for bond a and money market tool to compare Temporis a goal programming VBPI strategies with the standard CPPI strategy with multiplier value equal 2 and a naìve portfolio with 20% exposure. The results are presented in Figure 1, Figure 2 and Figure 3.

In the second illustration we have used the daily data for MSCI World index for period from January 2, 2006 to December 31, 2007 to compare the standard CPPI strategy with multiplier value equals 3 with the CPPI strategy presented above in the part 2.2. The results are presented in Picture 4. In the picture together with the portfolio value for the CPPI strategies there are values of the naive portfolios value, with the exposure 25%, and evolution of guaranteed value and its discounted value as well. In this illustration 3% yearly fee is assumed.

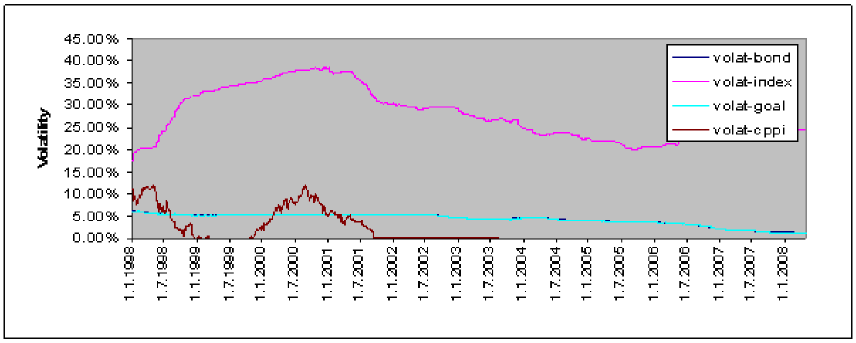


Figure 1: Volatilities of the strategies

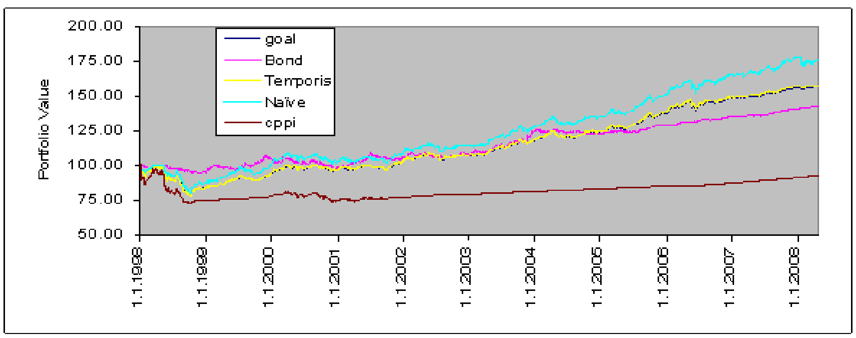


Figure 2: Portfolio values

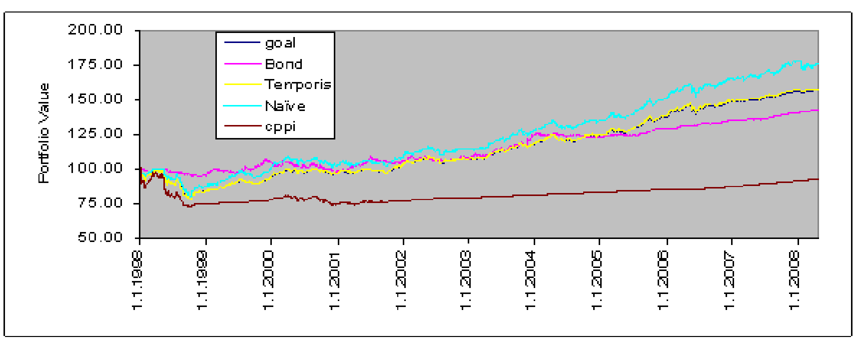


Figure 3: The exposures of the strategies

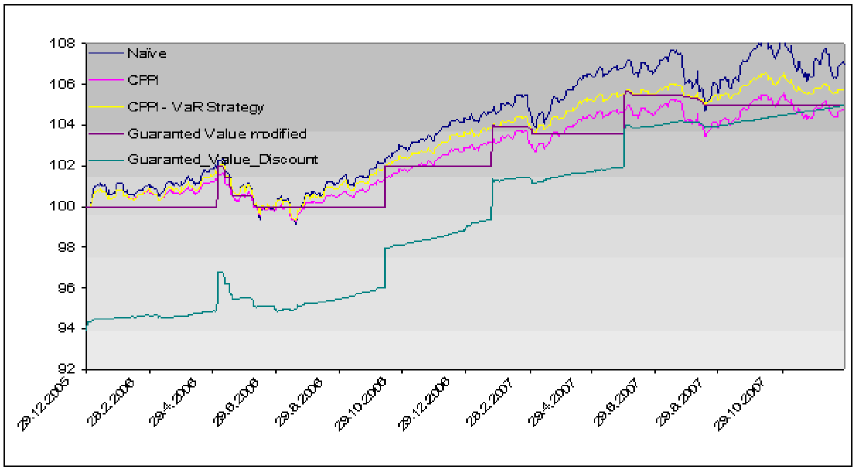


Figure 4: CPPI portfolio values

References

[1] Bertrand, P. J-l. Prigent: Portfolio insurance strategies: OBPI versus CPPI. GREQAM Seminar February 2002, et Universite Montpelier.

[2] Bertrand, P. J-l. Prigent: Portfolio insurance strategies: a comparison of standard methods when the volatility of the stocks is stochastic, GREQAM Seminar November 2004, et Universite Montpelier

[3] Black, F. and R. Jones: Siplifying portfolio insurance. *The Journal of Portfolio Management*, 1987, 48 – 51

[4] Black, F. and R. Rouhani: Constant portfolio insurance and the synthetic put option: a comparison, in Institutional Investor focus on Investment Management, edited by Frank J. Fabozzi, placeCityCambridge, StateMass.: Ballinger, 1989, pp 695 – 708

[5] Black, F. and A. R. Perold: Theory of constant proportion portfolio insurance. *The Journal of Economics, Dynamics and Control*, 16, 403 - 426

[6] Boulier, J-F and A. Kanniganti: Expected performance and risks of various portfolio insurance strategies. Proceedings of the 5th AFIR International Colloquium 1995, 1093 – 1122.

[7] Leland, H.E. and M. Rubinstein: The evolution of portfolio insurance, in: D.L. Luskin, ed.,Portfolio insurance: a guide to dynamic hedging, Wiley 1976

[8] Martellini, L. – K. Simsek – F. Goltz: Structured Forms of Investment Strategies in Institutional Investor's Portfolios. *An Edhec Risk and Asset Management Research Centre Publication. April 2005. EDHEC Bussines School, Lile – Nice.*

[9] Perold, A.: Constant portfolio insurance. Harvard Business School. Unpublished Manuscript, 1986.

[10] Perold, A. and W. Sharpe: Dynamic strategies for asset allocation. Financial Analyst Journal, 1988,January – February, 16 – 27

[11] Global equities and Derivative Solutions. **Temporis** presentation, February 2008, Societe Generale, Corporate and Investment Banking

Jaroslav Mlýnek

Department of Mathematics and Didactics of Mathematics
Faculty of Education
Technical University of Liberec
Studentská 2, 461 17 Liberec

e-mail: jaroslav.mlynek@tul.cz

Business Continuity Planning and Information Evaluation

Abstract

Nowadays, in trading companies, strong emphasis is being placed on ensuring business continuity in the case of an exceptional event. Emergency planning allows to resolve unexpected events promptly and effectively and to minimize incidental losses for a company. This paper is focused on ensuring business continuity in the case of any outage of a company information system or its part and the subsequent unavailability of information. The initial step for the process of ensuring business continuity is the implementation of a business impact analysis. There is presented one of the possible efficient approaches to the evaluation of information based on making a qualified estimation of negative impacts to the company in the case of unavailability of electronic information in this paper. Some areas of negative impact to the company are considered, for example direct financial losses, and the loss of goodwill. A properly selected respondent chooses, on a basis of stated criteria, the importance of an effect for individual areas of negative impact. By means of selected levels and their assigned values, it is possible to assess cumulative and maximum negative impacts. Next, based on the negative impact evaluation, it is possible to move to the processing of corresponding business continuity plans for the case of an accident and subsequent unavailability of electronic information.

Keywords: *Business continuity planning, information unavailability, information unit, business impact analysis, evaluation of negative impacts.*

1 Introduction

Strong emphasis is placed on ensuring business continuity in the case of a special event in trading companies. The summary of reccomendations for the given field is provided in [2], the given issue is also being solved, for example, in [4] and [5]. In the present entry, we shall be focused on the issue of ensuring business continuity in the case of unavailability of electronic information due to the outage of a company information system or its part. For example, the situation can occur in a case of a hardware failure, in presence of an important software error, in occurrence of an operator mistake, and due to a natural disaster or an act of terrorism.

The initial step in processing of business continuity plans is to implement the business impact analysis, i.e. the processing of a thorough analysis regarding financial and non financial negative impacts to the company in dependence on the duration of the outage of individual commercial activities. In doing so, in the paper, we shall come out from the assumption that the business activity outage is caused by the unavailability of necessary electronic information. Based on assessing the negative impacts to the company where information is not available, it is possible to determine the approximate evaluation of given information.

In the present entry, there is presented one of the possible and practically verified way of implementing the evaluation of information from the perspective of its availability. Based of the analysis performed, the company may proceed to the processing and subsequent testing plans for the recovery of a company information system.

2 Evaluation of information from availability stand-point

In this paragraph, we shall state one of the approaches to the information evaluation from the unavailability stand-point, which is used, e.g. in the CRAMM (CCTA Risk Analysis Management Method – see [1]) methodology as well as in the paper of [3].

In practice, an information is usually grouped by us to greater information complexes (e.g. "internet banking", "client information", "staff personal information"), which is often called as the information units. For coming out from appropriate requests to the availability of individual information units in processing business continuity plans related to the company information system, it is necessary to accomplish its evaluation from the availability stand-point.

2.1 Information unit evaluation

The evaluation principle is based on the assessment of qualified assumption for the importance of negative impacts to the company in the case of an information unit unavailability. The qualified assumption shall be accomplished based on the following procedure and collaboration with a properly selected respondent (usually the company's organization unit head, which is the main user or processor of the information unit).

Now, we are considering totally M areas of negative impact to the company in the case of the information unit's unavailability. At the same time, the negative impact areas can be specific for individual companies. For example, the most often areas are: direct financial losses, damages to commercial and economic interests, disturbance of the management and operations of a company, the loss of goodwill, the loss of business opportunity, and the breach of legal and regulatory obligations.

Let us to define the following two-variable function

$$F(m,t) = w(m)\ f(m,t), \tag{1}$$

where m indicates the number of the negative impact area, $m \in N;\ 1 \leq m \leq M$ (i.e. the negative impact areas are numbered with values of 1 to M), w is a weight function, $w\,(m) \geq 0$ for $m = 1,\ \ldots,\ M$; by means of the w weight function, it is possible to correct the importance of individual negative impact areas in the case of necessity. Every negative impact area m contains several levels $t \in Z;\ 0 \leq t \leq T_m,\ T_m \in N$. The evaluation of individual importance levels of the negative impact for the negative impact area m is given by the $f(m,t)$ function.

Number of the negative impact area (m)	Name of the negative impact area	Impact level of the area (t)										
		0	1	2	3	4	5	6	7	8	9	10
1	Loss of goodwill	0	2	3	5	7						
2	Breaking of company management and operations	0	1	3	5	6	7	1				
3	Infingement of legal and regulatory obligations	0	3	4	5	6	7					
4	Infingement of commercial and economic interests	0	1	2	3	4	5	6	7	8	9	10
5	Direct financial loss	0	1	2	3	4	5	6	7	8	10	
6	Unavailability of personal and/or client's information	0	1	2	3	4	5	6				
7	Infingement of law enforcement	0	3	4	7	8						
8	Loss of business opportunity	0	2	3	4	6	7					

Table 1: Example of the evaluation matrix for negative impact levels

In Table 1, there is presented the example of considered negative impact areas, appropriate impact levels, and the evaluation of individual levels of the negative impact for given areas; the number of considered negative impact areas is $M = 8$, and numbers of negative impact levels in individual areas are $T_1 = 4, \quad T_2 = 5, \quad T_3 = 5, \quad T_4 = 10, \quad T_5 = 9, \quad T_6 = 6, \quad T_7 = 4, \quad T_8 = 5$.

From Table 1, we can see that all values of the $f(m,t)$ function meet the $0 \leq f(m,t) \leq 10$ relation. In Table 2, there is presented the example which descripts the levels of impact importance and their evaluation for the "Loss of Goodwill" impact area in the case of $m = 1$.

Based on the respondent's selection of the t negative impact level for individual areas m (using a prepared scenario), we can define the total evaluation of the information unit's unavailability by means of the following relation

$$E_{total} = \sum_{m=1}^{M} w\,(m)\ f(m,\ t) \tag{2}$$

and the maximum evaluation of the information unit's unavailability

$$E_{max} = \max_{1 \leq m \leq M} w\,(m)\ f(m,\ t). \tag{3}$$

Based on the E_{total} and E_{max} calculated values and in the unavailability of the given information unit for established time intervals, it is possible to determine the company requirements

Number of the impact area: **m = 1**. Name of the impact area: **Loss of Goodwill**		
Number of the impact level(t)	Description of the impact level	Evaluation ($f(1\ ,t)$)
1	Adversely affects relations with other parts of the organization.	2
2	Adversely affects relations with other organizations or the public, but with the adverse publicity confined to the immediate geographic vicinity and with no lasting effects.	3
3	Adversely affects relations with other organizations or the public, with the adverse publicity more widespread than just the immediate geographic vicinity.	5
4	Significantly affects relations with other organizations or the public, which results in widespread adverse publicity.	7

Table 2: Example of the impact levels of the "Loss of Goodwill" area

for the information unit's availability in the case of an accident and to come out from these requirements in processing continuity business plans.

3 Practicle example

Let us consider the example of the "internet banking" and "staff personal information" information units. Further, let us suppose that Table 1 with weight function values $w\,(m) = 1$ for $m = 1,\ ...,M$ is used in determination of negative impacts. In table 3, there are shown negative impact levels established by the respondent and calculated values *Etotal* and *Emax* for one-hour and one-week unavailability periods of information units. The evaluation is usually accomplished for different time intervals of information units' unavailability.

Figures 1 and 2 present charts of *Etotal* and *Emax* evaluation values for "internet banking" and "staff personal information" information units in dependence on different time intervals of their unavailability.

In practice, it is suitable to determine two deadlines within which the information unit's availability has to be restored. Nevertheless, the first indicates until which time it is necessary to ensure at least a limited spare availability and the second indicates until which time it is necessary to ensure the information unit's availability in its full extent.

4 Conclusion

Based on the negative impacts' evaluation in the case of the information unit's unavailability and the determination of deadlines within which it is necessary to ensure at least the partial availability and the full availability of the information unit, it is possible to proceed to the

Number of the impact area (m)	Information unit "Internet Banking"				Information unit "Staff Personal Information"			
	Unavailability 1 hour		Unavailability 1 week		Unavailability 1 hour		Unavailability 1 week	
	Level(t)	$f(m,t)$	Level(t)	$f(m,t)$	Level(t)	$f(m,t)$	Level(t)	$f(m,t)$
1	2	3	4	7	1	2	1	2
2	1	1	2	3	1	1	2	3
3	1	3	5	7	0	0	1	3
4	2	2	3	3	0	0	1	1
5	3	3	5	5	0	0	0	0
6	1	1	3	3	0	0	2	2
7	0	0	0	0	0	0	2	4
8	2	3	5	7	0	0	0	0
E_{total} =	16		35		3		15	
E_{max}=	3		7		2		4	

Table 3: Example of determination E_{total} and E_{max} values in case of one-hour and one-week unavailability

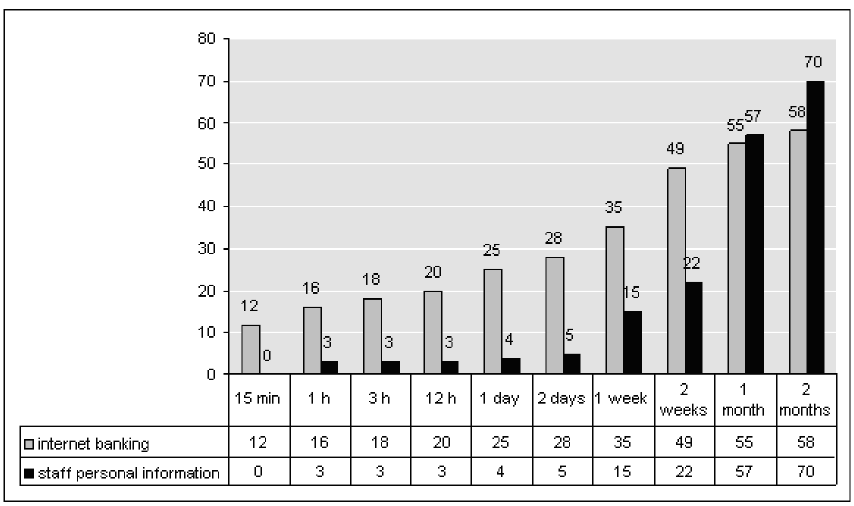


Figure 1: Chart of E_{total} values for different time intervals of information units' unavailability

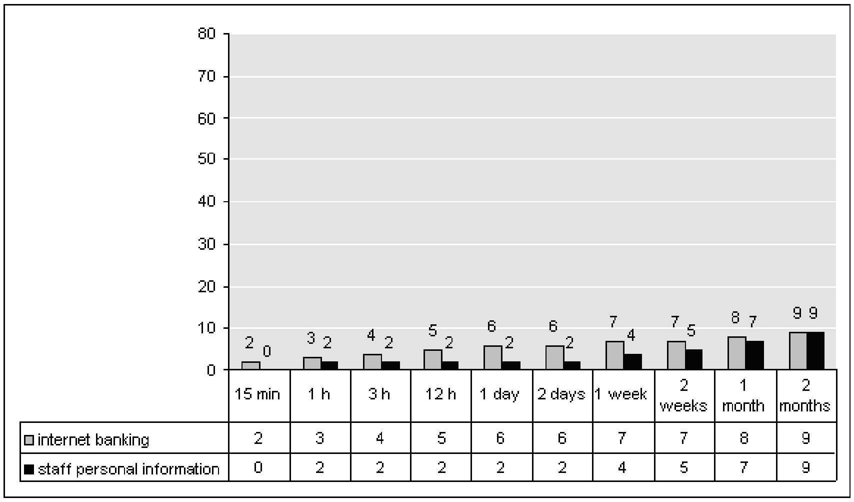


Figure 2: Chart of E_{max} values for different time intervals of information units' unavailability

processing of emergency plans and plans for the recovery of the information system and its parts. The negative impacts' evaluation in the case of the information unit's unavailability allows the company to elaborate emergency plans corresponding with the company needs and ensures only appropriate funds will be expended during recovery plans' processing and testing.

References

[1] *CRAMM User Guide*. `http//www.cramm.com`.

[2] ISO/IEC 2702 *Code of Practice for Information Security Management.*

[3] MLÝNEK, J.: *Securing of Business Information.* Mathematical Methods in Economics 2005, 23rd International Conference, Hradec Králové, pp. 273 – 280. ISBN 80-7041-535-5.

[4] MLÝNEK, J.: *Security of Business Information.* Computer Press, Brno, 2007, 154 pages. ISBN 978-80-251-1511-4.

[5] Smejkal, V., Rais, K.: *Risk Management.* Grada Publishing, 2003. ISBN 80-247-0918-7.

Karel Musil

Faculty of Economics and Administration
Masaryk University
Lipová 41a, 602 00 Brno

Monetary and Statistics Department
Czech National Bank
Na Příkopě 28, 115 03 Praha, Czech Republic[1]

e-mail: karel.musil@cnb.cz

Osvald Vašíček

Faculty of Economics and Administration
Masaryk University
Lipová 41a, 602 00 Brno

e-mail: osvald@econ.muni.cz

A Basic DSGE Model of the Czech Economy[2]

Abstract

The paper introduces a basic New Keynesian Dynamic Stochastic General Equilibrium (NK DSGE) model of the Czech economy. It is a closed economy model with tradable and non-tradable sectors based on microeconomic foundations. The model consists of representative agents. Behavior of a representative household contains rigidities in consumption in a form of a habit formation. Price setting behavior of a representative firm is described by the Calvo style (with a result in a New Keynesian Phillips Curve – NKPC). Central monetary authority implements monetary policy according to a modified Taylor rule (output growth rule model) in the inflation targeting regime. A Bayesian method is used for the estimation of the linearized model equations. A basic analysis of the estimated model is carried out through a description of parameters and impulse response functions analysis. The model seems to be a suitable approximation of the Czech economy behavior and it can be extended to a form of a more complex model structure.

Keywords. *NK DSGE model, output growth rule model, Taylor rule, inflation targeting, Bayesian estimation, rational expectations.*

[1]The opinions expressed are those of the author and do not necessarily reflect the views of the Czech National Bank.

[2]This paper is supported by MŠMT project Research centers 1MO524, funding of specific research at ESF MU, and by CNB.

1 Introduction

To describe the Czech economy in a simple but effective way, it is possible to employee an elementary model framework. On the one hand the framework should detect and introduce basic transmission mechanisms of the economy in a model, on the other hand it should also offer a possibility to extend basic characteristics of the model into a more complex structure.

The basic aim of the paper is to introduce a suitable conceptual model of the Czech economy with a simple structure, which is estimated. The structure and equations describing behavior of agents in the model are required to capture basic characteristics of the Czech economy. Moreover, the structure should offer a scope for a possible extension of the model.

A structure of the employed model proceed from [2][3] and it is extended by [4] and [7]. The model is adjusted and adapted to the conditions of the Czech condition. A closed economy (CE) model is introduced to meet the requirement of a simple structure.

2 Conceptual Model

The employed model is a New Keynesian Stochastic General Equilibrium (NK DSGE) model based on microeconomic foundations. The CE model is composed of domestic representative agents - households, firms, and central monetary authority. There are two production sectors in a form of tradables and non–tradables.[4] The optimizing behavior of agents contains some rigidities.

2.1 Representative Households

A representative household solves the following optimizing problem: it tries to maximize its utility function

$$E_0 \left\{ \sum_{t=0}^{\infty} \beta^t \left[\frac{(CC_t \ A_{W,t})^{1-\tau}}{1-\tau} - N_t \right] \right\} \tag{1}$$

subject to its budget constraint

$$P_t C_t + E_t \left(\frac{D_{t+1}}{R_t} \right) \leq D_t + W_t N_t \tag{2}$$

for $t = 0, 1, 2, \ldots$, where where $0 < \beta < 1$ is a discount factor, $\tau > 0$ is an inverse elasticity of intertemporal substitution. N_t denotes hours of labor, W_t is a nominal wage for labor supply, R_t is a gross nominal interest rate (for payments from a portfolio of assets), P_t is an overall consumption price index (CPI), D_t is a nominal payoff on a portfolio held at t, and the whole term $E_t \left(\frac{D_{t+1}}{R_t} \right)$ expresses a price of a portfolio purchased at time t. The effective consumption (CC_t) depends on a current and last period consumption (C_t and C_{t-1}), on a habit formation parameter $0 < h < 1$, and on parameter γ, which is a steady state growth rate of $A_{W,t}$ (for all t):

$$CC_t = C_t - h\gamma C_{t-1} \tag{3}$$

[3] A simpler structure of the model without tradables and non–tradables is introduced e.g. in [5].

[4] In this sense the model is an extension of [6] in a form of two types of produced and consumed goods.

A long–run technology shock A_t is non–stationary and $z_t = \frac{A_t}{A_{t-1}}$. The evolution of z_t follows a stationary autoregressive process (in a log–linearization form)[5]:

$$\tilde{z}_t = \rho_z \tilde{z}_{t-1} + \epsilon_{z,t}, \tag{4}$$

for all t, with $0 \leq \rho_z < 1$, where $\tilde{z}_t$ is a deviation from its steady state and the serially uncorrelated innovation $\epsilon_{z,t}$ has a standard normal distribution.

The first order conditions (FOCs) derived from the optimizing behavior of the representative household with respect to C_t, N_t and D_{t+1} are:

- for C_t:
$$A_{W,t}\Lambda_t = CC_t^{-\tau} - h\gamma\beta E_t\left(\frac{A_{W,t}}{A_{W,t+1}}CC_{t+1}^{-\tau}\right), \tag{5}$$
- for N_t:
$$\frac{1}{\Lambda_t} = \frac{W_t}{P_t}, \tag{6}$$
- for D_{t+1}:
$$\frac{1}{R_t} = \beta E_t\left(\frac{\Lambda_{t+1}}{\Lambda_t}\frac{P_t}{P_{t+1}}\right). \tag{7}$$

for $t = 0, 1, 2, \ldots$, where Λ_t is a Lagrangian multiplier at time t.

The total consumption C_t is composed of two groups of goods - tradable consumption $C_{T,t}$ and non–tradable consumption $C_{N,t}$ (for all t):

$$C_t \equiv \{(1-\lambda)^{\frac{1}{\nu}} C_{T,t}^{\frac{\nu-1}{\nu}} + \lambda^{\frac{1}{\nu}} C_{N,t}^{\frac{\nu-1}{\nu}}\}^{\frac{\nu}{\nu-1}}, \tag{8}$$

where $0 \leq \lambda \leq 1$ is a share of non–tradable goods in total consumption and $\nu > 0$ is an interteporal elasticity of substitution between tradable and non–tradable consumption.

Then it is possible to derive an overall CPI:

$$P_t \equiv \{(1-\lambda)P_{T,t}^{1-\nu} + \lambda P_{N,t}^{1-\nu}\}^{\frac{1}{1-\nu}}, \tag{9}$$

for $t = 0, 1, 2, \ldots$, where $P_{T,t}$ is a price of tradable and $P_{N,t}$ is a price of non–tradable consumption bundle.

2.2 Representative Producers

Suppose that there is a continuum of monopolistic competitive producers using the following production function:

$$Y_{j,t}(i) = A_t A_{j,t} N_{j,t}(i), \tag{10}$$

[5] Henceforth all variables in small letters with tildas express a deviation of the original variables (in capital letters) from their steady states.

for $t = 0, 1, 2, \ldots$, $j = T, N$ (tradable and non–tradable goods), and for the i–the producer, where $A_{j,t}$ is a stationary sector specific technology shock described by an AR(1) process for $j = T, N$ (in a log–linearized form):

$$\tilde{a}_{j,t} = \rho_j \tilde{a}_{j,t-1} + \epsilon_{j,t}, \tag{11}$$

for all t, where $0 \leq \rho_j < 1$, $\tilde{a}_{j,t}$ is a deviation of $A_{j,t}$ from its steady state and $\epsilon_{j,t}$ is a serially uncorrelated innovation with a standard normal distribution.

A solution of a cost minimizing problem of representative producers in both sectors with respect to their demand constraints follows a development of real marginal costs for $j = T, N$ and for $t = 0, 1, 2, \ldots$ according to:

$$MC_{j,t} = \frac{\Lambda_t^{-1}}{A_{j,t}} \frac{P_t}{P_{j,t}}. \tag{12}$$

The price setting policy of producers is subject to the Calvo style.[6] Under the Calvo pricing only a fraction $(1 - \theta_j)$ of firms sets prices optimally. The rest fraction θ_j of firms in the j–th sector adjusts their prices with respect to the steady state inflation rate π.

The i–th representative firm in the j–th production sector optimizing its price of production at period t maximizes its revenue streams with respect $P_{j,t}(i)$

$$E_0 \left[\sum_{k=0}^{\infty} \theta_j^k Q_{t+k,t} Y_{j,t+k}(i) \left(P_{j,t}(j) - P_{j,t+k} MC_{j,t+k} \right) \right] \tag{13}$$

subject to the demand function in the proper sector (for all t)

$$Y_{j,t}(i) = \left(\frac{P_{j,t}(i)}{P_{j,t}} \right)^{-\omega} C_{j,t}, \tag{14}$$

A result of this kind of behavior implies that prices are set as a (time–varying) mark–up over the marginal costs. The relationship between inflation and marginal costs after aggregation over all individual firms is a New Keynesian Phillips Curve (NKPC) for $j = T, N$ and for all t:

$$\tilde{\pi}_{j,t} = \beta E_t \tilde{\pi}_{j,t+1} + \frac{(1 - \beta\theta_j)(1 - \theta_j)}{\theta_j} \tilde{mc}_{j,t}. \tag{15}$$

2.3 Central Monetary Authority

An implementation of monetary policy is approximated by a modified Taylor rule (in a gap form). The rule describes behavior of a central bank in an inflation targeting regime and can be understood as a growth rule for output. The bank is interested in a development of inflation and output gap. The output gap is used in a form of the adjusted output gap by a true value (not the steady state) growth rate of the output. The rule has the following form:

$$\tilde{r}_t = \rho_r \tilde{r}_{t-1} + (1 - \rho_r) \left[\phi_1 \tilde{\pi}_t + \phi_2 (\Delta \tilde{y}_t + \tilde{z}_t) \right] + \epsilon_{R,t} \tag{16}$$

for all t, where $\tilde{\pi}_t$ is a deviation of inflation from an inflation target and the output gap $(\Delta \tilde{y}_t + \tilde{z}_t)$ is a deviation of adjusted output gap.

[6] For more details see [1].

2.4 Goods Market Clearing Condition

The condition expresses a fact that the total production is divided between domestic tradable and non–tradable consumption (for all t):

$$Y_t = C_{H,t} + C_{N,t}. \tag{17}$$

Consumption in the non–tradables (optimal allocation function for non–tradabel consumption) is a result of the optimizing behavior of a representative household and has the following form:

$$C_{N,t} = \lambda \left(\frac{P_{N,t}}{P_t}\right)^{-\nu} C_t = \lambda \, RP_{N,t}^{-\nu} C_t, \tag{18}$$

for $t = 0, 1, 2, \ldots$, where $RP_{N,t}$ is a relative price of non–tradables.

Similar equation can be derived for the tradables:

$$C_{T,t} = (1-\lambda) \left(\frac{P_{T,t}}{P_t}\right)^{-\nu} C_t = (1-\lambda) \, RP_{T,t}^{-\nu} C_t, \tag{19}$$

for all t, where $RP_{T,t}$ is a relative price of tradables.

A combination of three previous equations together yields the goods market clearing condition:

$$Y_t = \{(1-\lambda) \, RP_{T,t}^{-\nu} + \lambda \, RP_{N,t}^{-\nu}\} \, C_t. \tag{20}$$

3 Model Equations

Log–linearizing and stationarizing (due to a non–stationary variable in long–run technology shock A_t) forms a linearized system. The system is composed of linearized and recalculated equations (3), (4), (5), (7), (1f), (11), (12), (15), (16), and (20). It has the following structure:

$$-\tilde{\lambda}_t = \frac{\tau}{1-h\beta}\tilde{cc}_t - \frac{h\beta}{1-h\beta} E_t(\tau \tilde{cc}_{t+1} + \tilde{z}_{t+1}) \tag{21}$$

$$(1-h)\tilde{cc}_t = \tilde{c}_t - h\tilde{c}_{t-1} + h\tilde{z}_t \tag{22}$$

$$-\tilde{\lambda}_t = -E_t\tilde{\lambda}_{t+1} - (\tilde{r}_t - E_t\tilde{\pi}_{t+1}) + E_t\tilde{z}_{t+1} \tag{23}$$

$$\tilde{\pi}_{j,t} = \beta E_t \tilde{\pi}_{j,t+1} + \kappa_j \tilde{mc}_{j,t} \tag{24}$$

$$\tilde{mc}_{j,t} = -\tilde{\lambda}_t - \tilde{rp}_{j,t} - \tilde{A}_{j,t} \tag{25}$$

$$\tilde{\pi}_t = (1-\lambda)\tilde{\pi}_{T,t} + \lambda \tilde{\pi}_{N,t} \tag{26}$$

$$\tilde{y}_t = -\nu\lambda\{(1-\lambda)\tilde{rp}_{T,t} + \lambda \tilde{rp}_{N,t}\} + \tilde{c}_t \tag{27}$$

$$\tilde{r}_t = \rho_r \tilde{r}_{t-1} + (1-\rho_r)\left[\phi_1 \tilde{\pi}_t + \phi_2(\Delta\tilde{y}_t + \tilde{z}_t)\right] + \epsilon_{R,t} \tag{28}$$

$$\tilde{a}_{j,t} = \rho_j \tilde{a}_{j,t-1} + \epsilon_{j,t} \tag{29}$$

$$\tilde{z}_t = \rho_z \tilde{z}_{t-1} + \epsilon_{z,t} \tag{30}$$

for $t = 0, 1, 2, \ldots$, $j = T, N$, and $\kappa_j = \frac{(1-\beta\theta_j)(1-\theta_j)}{\theta_j}$, where inflation identities are: $\tilde{\pi}_t = \tilde{p}_t - \tilde{p}_{t-1}$, $\tilde{\pi}_{j,t} = \tilde{p}_{j,t} - \tilde{p}_{j,t-1}$ and relative prices hold for $\tilde{rp}_{j,t} = \tilde{p}_{j,t} - \tilde{p}_t$. The parameters of the model for an estimation are θ_T, θ_N, λ, τ, h, ν, ρ_r, ϕ_1, ϕ_2, ρ_z, ρ_T, ρ_N, and steady state value for $\tilde{r}_t$.

4 Solving

The log–linearized model can be rewritten into a linear rational expectations (LRE) model, which can be solved and expressed as a general equilibrium (GE) in a form of the following state space model:

$$s_t = A(\theta)s_{t-1} + B(\theta)\epsilon_t \tag{31}$$

$$y_t = C(\theta) + Ds_t \tag{32}$$

for all t, where s_t is a vector of states, ϵ_t is a vector of innovations of the exogenous processes, $A(\theta)$ and $B(\theta)$ are matrices of models deep parameters, which are collected in a vector θ, representing the dynamic core of the model, y_t expresses a vector of observables and matrices $C(\theta)$ and D.

The Bayesian approach is employed for an estimation of the linearized model. According to the Bayesian formula all inference about the parameter vector is contained in the posterior density. The density is calculated with using a likelihood function and a prior density. To generate draws from the posterior distribution of the model parameters, it is used Markov Chain Monte Carlo (MCMC) method.[7]

5 Estimation Results

A data set for the Czech Republic from I.Q 1999 to IV.Q 2007 is used for the estimation. The quarterly data has the following structure:

- a deviation of the real GDP growth from a development of its balanced growth,
- a deviation of the CPI from a development of its dynamic equilibrium and from I.Q 2006 a deviation from a CPI level corresponding to the inflation target,
- a deviation of the level of tradable prices from a development of its dynamic equilibrium.

With using the data and information included in prior densities, we generate 500 000 draws of Markov Chain. Parameter β was set in the following way: $\beta = \frac{1}{1+rr^{ss}/100}$, where rr^{ss} is a steady state interest rate. Table 1 presents the posterior estimates of parameters and 90 % probability intervals.

The estimation of parameter θ_T is 0.58 or in other words an average duration of price contracts of a representative firm in the tradable sector is more than 0.6 quarter. An average duration of the price contracts in the non–tradable sector is longer – about 0.75 quarter. A share of the non–tradable goods in consumption (λ) corresponds to the prior information (48.87 %). An assumption of a lower possibility of substitution between tradables and non–tradables (parameter ν) is reflected in estimation by 0.63. Behavior of a representative household is characterized by an elasticity of substitution (inverse τ) as 1.82 and by a habit parameter in consumption (h) as 0.85. The estimated value of the discount factor is 0.99.

A monetary policy implementation according to the adjusted Taylor rule is estimated as (for all t):

$$\tilde{R}_t = 0.88\,\tilde{R}_{t-1} + (1 - 0.88)\,[1.29\,\tilde{\pi}_t + 0.45\,(\Delta\tilde{y}_t + \tilde{z}_t)] + \epsilon_{R,t}, \tag{33}$$

[7] The extension of the paper forbids any detailed description of the Bayesian method. For general description of the Bayesian approach see e.g. [3].

Parameter	Prior Mean	Posterior Median	90 % Posterior Interval
θ_T	0.50	0.5805	$\langle 0.4347;\ 0.7211\rangle$
θ_N	0.50	0.3346	$\langle 0.2100;\ 0.4551\rangle$
λ	0.48	0.4887	$\langle 0.4563;\ 0.5204\rangle$
τ	1.00	0.5491	$\langle 0.0890;\ 0.6760\rangle$
h	0.70	0.8497	$\langle 0.7907;\ 0.9566\rangle$
ν	0.60	0.6311	$\langle 0.0001;\ 1.5055\rangle$
ρ_R	0.70	0.8770	$\langle 0.8428;\ 0.9132\rangle$
ϕ_1	1.30	1.2895	$\langle 1.0929;\ 1.4811\rangle$
ϕ_2	0.40	0.4469	$\langle 0.2357;\ 0.6516\rangle$
ρ_z	0.75	0.7556	$\langle 0.6634;\ 0.8663\rangle$
ρ_T	0.60	0.7396	$\langle 0.6324;\ 0.8498\rangle$
ρ_N	0.60	0.6407	$\langle 0.4952;\ 0.7822\rangle$
rr^{ss}	0.25	0.2450	$\langle 0.0868;\ 0.3908\rangle$
σ_{ϵ_T}	1.50	2.3500	$\langle 1.1509;\ 3.5643\rangle$
σ_{ϵ_N}	1.50	1.0791	$\langle 0.6981;\ 1.4224\rangle$
σ_{ϵ_R}	0.30	0.1079	$\langle 0.0830;\ 0.1316\rangle$
σ_{ϵ_z}	0.20	0.1759	$\langle 0.0941;\ 0.2243\rangle$

Table 1: Posterior Estimates of Parameters and Innovations

which indicates a relatively high stress on the interest rate smoothing accompanied by 2.9 times higher importance of the inflation gap compare to the output gap. A standard deviation of the monetary policy shock (σ_{ϵ_R}) is low.

Exogenous processes are characterized by a relatively high persistence in a development of the growth rate of the long–run technology and tradable sector technology shock. A propagation of the non–tradable technology shock is shorter. This is combinated with a high volatility of the tradable technology shock ($\sigma_{\epsilon_T} = 2.4$). On the other hand the standard deviation of the growth rate of the long–run technology shock is relatively small ($\sigma_{\epsilon_z} = 0.2$).

6 Analysis of Behavior

An analysis of behavior is conducted by impulse response functions as a reaction to a unit change of the exogenous variables of the model – an increase in technology in tradable and non–tradable sector, monetary policy shock, and growth rate of the world wide technology shock. All these impulse responses are able to describe economic behavior in an appropriate way. A monetary shock is introduced in Figure 1.

An immediate reaction of the central bank to a unit monetary shock is presented by an increase in nominal interest rate by 0.4 % according to the Taylor rule – see (33). Higher

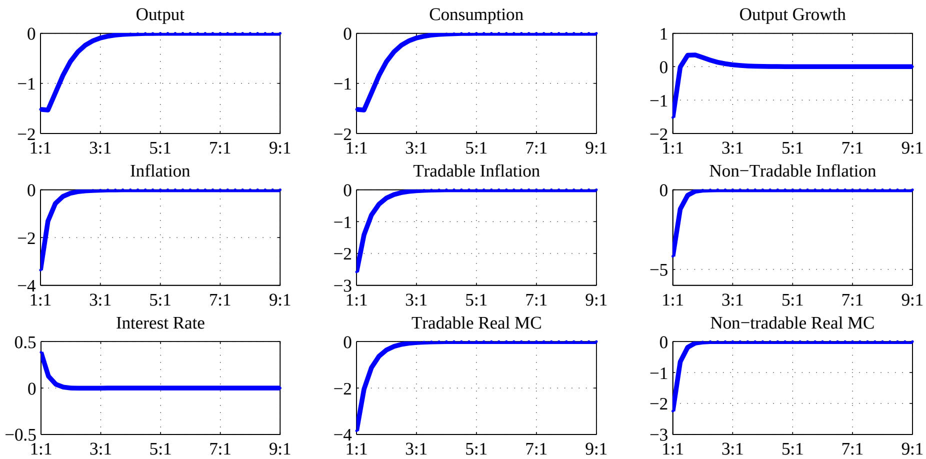


Figure 1: Impulse Response Function from One Unit of Monetary Shock

nominal (and also real) interst rate reduces consumption and output (there is a slowdown in the output growth as well). The output fall reduces overall inflation because of a lower inflation in both production sectors (due to a fall in real marginal costs). The reduction in non–tradable inflation (4.1 %) is higher than in the tradable sector (2.5 %). This corresponds to the estimated values of the Calvo parameter in both sectors.

7 Conclusion

The theoretical background, estimation of the model, and results from the impulse analysis give a suitable approximation of behavior of the Czech economy. Although the model is relatively simple, it is able to describe basic dynamics of the Czech economy. An extended structure of tradable and non–tradable sectors can help to better understood a behavior of the economy. The estimated parameters corresponds to the basic characteristics of the Czech economy. The analysis of behavior aimed at the monetary shock can offer a valuable look at the monetary policy implementation according to the estimated modified Taylor rule (with output growth rule).

The employed model seems to be a suitable starting point for a model with more complex structure.

References

[1] Calvo, G.: Staggered Prices in an Utility Maximizing Framework. *Journal of Monetary Economics* 12 (September), pages 383 – 398, 1983.

[2] Lubik, T. A., Schorfheide, F.: *A Bayesian Look at New Open Economy Macroeconomics*. University of Pennsylvania, F. Schorfheide Web Site, 2005, access from internet <www.econ.upenn.edu/ ˜schorf/papers/nber-final.pdf> (cited to date 12. 5. 2007).

[3] Lubik, T. A., Schorfheide, F.: Testing for Indeterminacy: A Reply to Comments by A. Beyer and R. Farmer. *The American Economic Review*, Vol. 94, No. 1 (Mar.), pages 190 – 217, 2004.

[4] Matheson, T.: *Assessing the Fit of Small Open Economy DSGEs*. Reserve Bank of New Zealand, Discuss Paper 11/2006, 2006, access from internet <www.rbnz.govt.nz/research/discusspapers/dp06 11.pdf> (cited to date 8. 11. 2007).

[5] Monacelli, T.: *Monetary Policy in a Low Pass-through Environment*. European Central Bank, WP 227, 2003, access from internet <www.ecb.int/pub/pdf/scpwps/ecbwp227.pdf> (cited to date 12. 5. 2007).

[6] Musil, K. and Vašíček, O.: A Partial Dynamic Stochastic General Equilibrium Model of the Czech Economy. *Mathematical Methods in Economics 2007*, pages 197 – 204, Ostrava, VŠT – Technical University of Ostrava, ISBN 978-80-248-1457-5, 2006.

[7] Santacreu, A. M.: *Reaction Functions in a Small Open Economy: What Role for Non-traded inflation?*. Reserve Bank of New Zealand, Discuss Paper 04/2005, 2005, access from internet <www.rbnz.govt.nz/research/discusspapers/dp05 04.pdf> (cited to date 31. 11. 2007).

Dana Nejedlová

Department of Informatics
Faculty of Economics
Technical University of Liberec
Studentská 2, 461 17 Liberec

e-mail: Dana.Nejedlova@tul.cz

Classification of Time Series and Visualization of Its Results

Abstract

The article suggests a way how to visually compare a large number of long time series. The output of data processing proposed in this article is a representation that allows humans a comparison of several tens of time series at a glance. Time series represented by graphs where horizontal axis is time can be well compared only if there is some reasonable number of them in a single graph. Representation suggested in this article looks like a table with colored cells. The headings of columns can be names of different kinds of time series. The headings of rows can be time intervals within the whole time series. Cells having the similar colors indicate that the time intervals that they represent have similar values, which means that their Euclidean distance is small.

Data processing is based on the Kohonen self-organizing map (SOM) which is a grid of neurons. In the first phase a set of colors represented as RGB vectors is determined by human. In the second phase this set is spread over the SOM. In the third phase the SOM is reinitialized and used for the classification of time intervals of time series. In the fourth phase the classified time intervals of time series are assigned with colors that were previously assigned to the same neurons that are now assigned to the time intervals of time series. In the fifth phase the time intervals of time series represented by their colors are arranged into table with columns meaning the kinds of time series and rows meaning the time intervals.

The method is demonstrated on data about number of employees in 16 industrial branches from 1995 to 2007 from the Czech Statistical Office.

Keywords: *time series, classification, Kohonen SOM, visualization, business intelligence.*

1 Introduction

We live in the world full of data, which are the output of computers. The search for methods how to obtain knowledge from data is increasingly important. Whilst computers operate in terms of numbers, people think in terms of meanings, colors, sounds, smells, feelings, and so on. It is important to transform the data into these kinds of information, so that people can understand and remember them easily. Of all human senses usually vision is used as a target for data presentation, so the data are being visualized. See [1] for some examples of data visualization.

This paper proposes just another kind of visualization. The results are demonstrated on data from the Czech Statistical Office about the number of employees in 16 industrial branches from 1995 to 2007 taken from [2]. The industrial branches are listed in Table 1.

The visualization was done using a kind of artificial neural network. The artificial neural networks are very simplified models of biological brains. Like neurons in the brain, artificial neurons perform simple signal processing and are exchanging signals with their peers which results in intelligent behavior of the whole system. The most important attribute of both artificial and biological neural networks is the ability to learn, which means that these systems classify the input information into some target response. The neural network that has learned something has the ability of generalization, i.e. correct classification of inputs that it had not seen when it was learning, and the ability of cope with errors both in the input information and in its structure, e.g. missing neurons and synapses for exchanging signals. On the other hand, the neural networks are not the best tools for obtaining the exact values and behaving always the same way when processing the same data. Such needs should be fulfilled by using the traditional computing.

Label	**Branch**
A	Agriculture, hunting, forestry
B	Fishing
C	Mining and quarrying
D	Manufacturing
E	Electricity, gas and water supply
F	Construction
G	Wholesale and retail trade; repair...
H	Hotels and restaurants
I	Transport, storage, communication
J	Financial intermediation
K	Real estate, renting, business activities
L	Public administration and defense; compulsory social security
M	Education
N	Health and social work
O	Other community, social and personal service activities
P	Private households with employed persons

Table 1: Industrial branches used in the statistical data of the Czech Statistical Office and their labels used in this paper

2 Motivation of Research

The research described in this paper focuses on visualization of data in the form of time series. Traditional representation of time series in the form of graph is effective provided the number of time series on the single graph is lower than some limit. See Figure 1 for an example of a graph that has probably too many time series to be well intelligible.

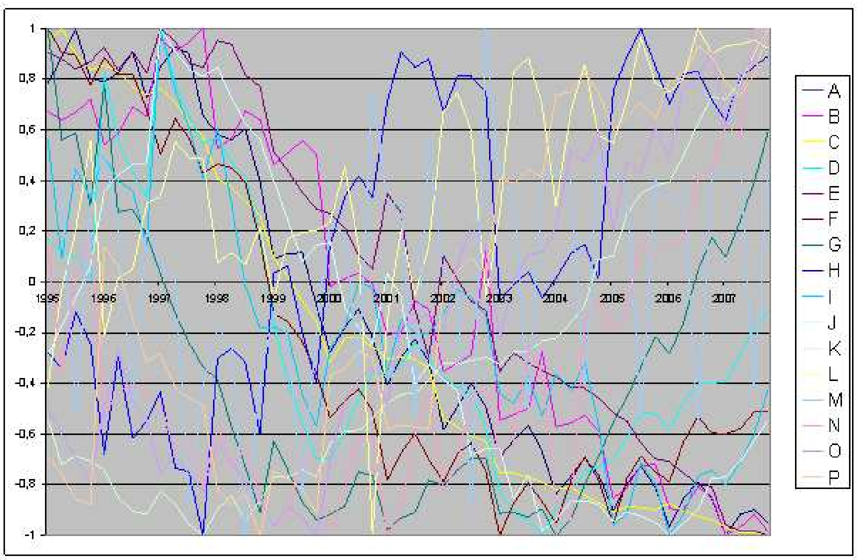


Figure 1: Graph of the normalized number of employees in 16 industrial branches from 1995 to 2007

The method of data visualization proposed in this paper consists in arranging data according to their categories as horizontal dimension and according to time as vertical dimension. Similarities of the parts of time series are expressed using colors. This is probably not very original but the use of neural network for this may be a novelty.

3 Kohonen Self-Organizing Map

Kohonen self-organizing map, also known as SOM, is a kind of artificial neural network originally published in paper [3]. The research presented in this paper is based on it.

SOM is traditionally used for classification or clustering of multidimensional vectors that represent real objects. The typical output of such an analysis is an XY graph. The nearer the objects are in the graph the more similar they are. But the problem with such a kind of graphs is the same as the problem of the graph in Figure 1: too many objects makes the graph less intelligible.

Kohonen self-organizing map is a grid of neurons. When this map has to classify a set of n-dimensional vectors x, each neuron remembers its own n-dimensional vector k which is at the beginning of the processing initialized to small random values. Then each neuron reads the first input vector x to be classified. The squared Euclidean distance (1) is computed for each neuron and the neuron with the smallest Euclidean distance of its vector k from vector x is the winner.

$$d = \sum_{i=1}^{n} (x_i - k_i)^2 \tag{1}$$

The winner makes each of its vector component k_i closer to the responding component x_i of the input vector using the Formula (2). Also neurons which are topologically closer than some radius to the winning neuron in the grid adapt their vectors according to Formula (2).

$$k_i = k_i + a\,(x_i - k_i) \tag{2}$$

The remaining input vectors x are read and processed the same way one after another. The set of input vectors is processed several times until no great changes in winning neurons are observed on the network. Both radius and parameter a in Formula (2) are monotonically decreasing to zero as the number of processing cycles through the set of input vectors rises. Initial radius can be as big as the height or width of the grid and initial parameter a is usually less than 1 and larger than 0.

The result of this process is that winners responding to input vectors form regions of similar vectors k on SOM. When the SOM is a two-dimensional grid, the input vectors can be assigned to the co-ordinates of their winning (i.e. most similar) neurons on SOM and these co-ordinates can be plotted on an XY graph showing the clusters of similar objects represented by input vectors. Neuron's co-ordinates define its location on the grid. Kohonen SOM is shown in Figure 2. Each neuron on SOM reads the input vector, not only the leftmost neurons as it is depicted for simplicity in Figure 2.

To get a better idea how regions of similar vectors on SOM are formed, let us imagine that a certain neuron in SOM is a winner for certain input vector x. Formula (2) makes its weights k_i even closer to x. Then the same formula is applied to the neurons that are close to the winner within some radius. The closeness of neurons can be measured as the minimal number of bars that should be crossed to get from one neuron to the other neuron on the grid. The closeness of neurons should not be confused with the closeness of vectors which is their Euclidean distance. Then let us imagine that the SOM reads some other input vector x and this vector is similar to the previously processed vector. In this case it is probable that the winner for such vector will be close to the previous winner on the grid but it will not necessarily be the same neuron because the neurons' weights k_i were initialized to random values. The winners for the previous and similar current vector can be close on the grid because Formula (2) was applied only to the neurons within some radius from the winner. In this way regions responding to similar vectors are gradually forming on SOM. The radius for applying Formula (2) and the parameter a in (2) are decreasing with the increasing number of readings of the set of input vectors x because in this way the neural net gets stable in terms of its winners for input vectors in the phase when it has already learned existing similarities. Neuroscientists have observed that biological brains have also regions that respond similarly to similar sensory inputs that have the nature of tactile, visual and acoustic information, which has motivated the development of self-organizing maps.

Since the weights k_i of each neuron in SOM are initialized randomly, the resulting mapping of categories to co-ordinates on SOM can be different for different initializations. For example, the grid of neurons on SOM assigned to colors shown in Figure 3 is only one of many possible mappings. And the situation when the colors on SOM had finally formed a continuous ring, as it is shown in Figure 3, was not reached after every initialization of SOM.

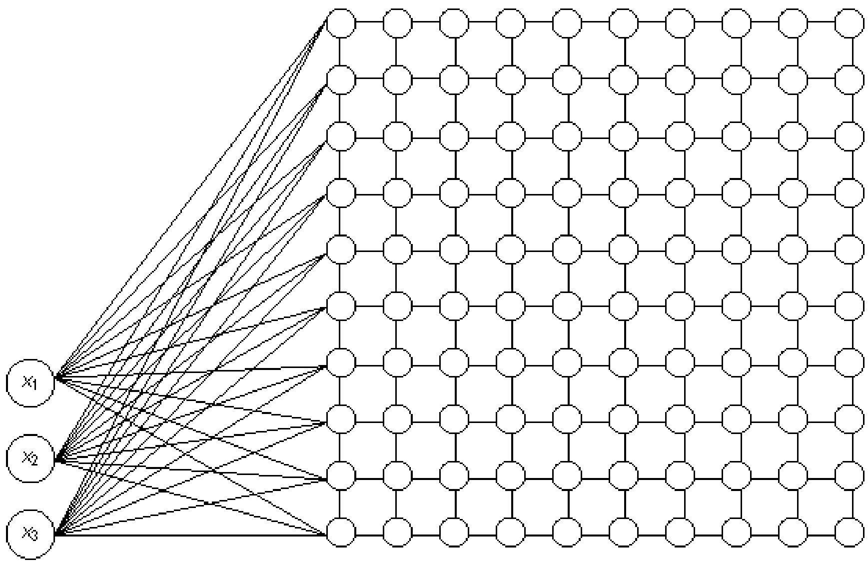


Figure 2: Kohonen self-organizing map for the classification of three-dimensional vectors

4 Data Processing

16 time series from [2] were linearly normalized according to Formula (3) so that their values were ranging from –1 to +1.

$$x_{norm} = \frac{2x - x_{\max} - x_{\min}}{x_{\max} - x_{\min}} \tag{3}$$

The reason for their normalization was a considerable difference in their scales. Branch D has the largest order of millions, branch P has the smallest order of hundreds. After the normalization the time series can be compared in terms of their trends and their graph is shown in Figure 1.

The normalized time series were divided into one-year-long intervals. The resulting vectors had 4 values each for a single quarter of year. The number of vectors is 208 (16 branches times 13 years – 1995 to 2007).

A set of 1530 colors was selected for the representation of nearness of the intervals of time series. The colors were represented by RGB vectors that are standard in computer representation of colors. For example, red color is represented as (255, 0, 0).

The set of colors was placed on the Kohonen self-organizing map in such a way that optimizes the rule that near neurons have similar colors. The placement of colors was computed using the standard algorithm of SOM with the additional rule that no neuron was allowed to be assigned to more than one color. In this way 1530 neurons were assigned to some color and the final placement of colors is shown in Figure 3.

In the next phase the SOM was reinitialized and used for the classification of the 208 vectors representing the 13 parts of 16 time series. The placement of vectors was also computed using the standard algorithm of SOM with the additional rule that only those neurons that were previously assigned to some color were allowed to be assigned to some of the 208 input vectors. A single neuron was allowed to be the winner for more than one input vector. The final placement of the 208 vectors is also shown in Figure 3.

The Kohonen self-organizing map used for this processing had a grid of 85 times 85 neurons.

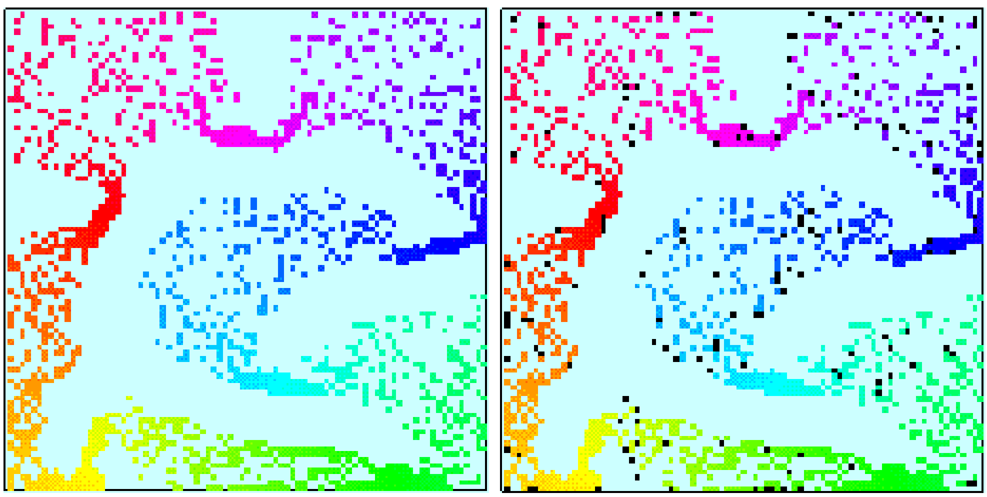


Figure 3: Kohonen SOM with 1530 different colors assigned to some of its neurons (left). The same SOM where some neurons (the black ones) are assigned to 208 parts of time series (right)

We can see in Figure 3 that the input vectors that are placed near each other on the Kohonen self-organizing map have mostly similar colors.

The 208 input vectors are then rearranged into table shown in Figure 4.

Two time series having similar colors in the same row mean that these two time series have similar values in the given year represented by the row. Similar colors of a single time series in several years mean that this time series is on similar levels in these years. In Figure 4 we can easily find pairs of branches that have similar colors in each year. These branches have

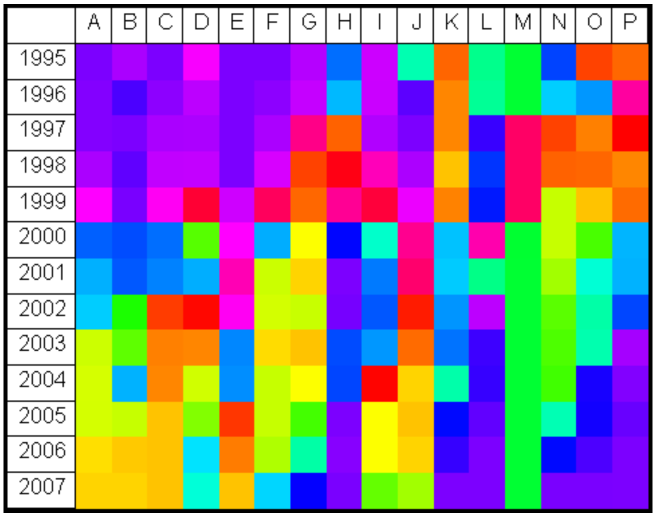


Figure 4: The resulting visualization of the similarities of one-year-long parts of time series of normalized number of employees in 16 industrial branches from 1995 to 2007

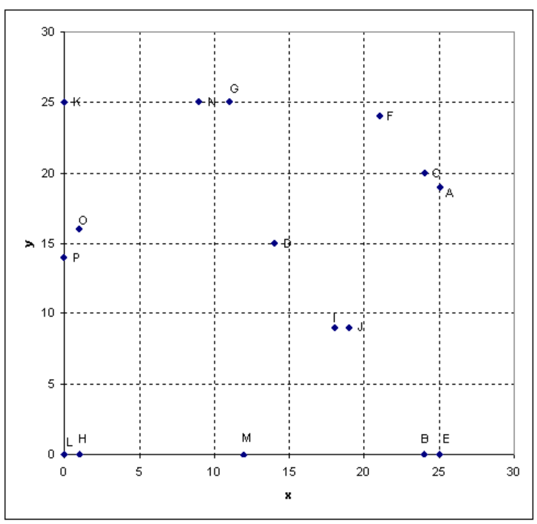


Figure 5: Classification of time series of normalized number of employees in 16 industrial branches from 1995 to 2007 using SOM

similar development. This result can be compared to the result of classification of whole time series by SOM of 26 times 26 neurons shown in Figure 5. In most cases the branches plotted near each other in Figure 5 have similar colors in Figure 4 for each year.

When the time intervals of the analyzed time series are as long as some periodical cycles, then their similarities could be compared correctly. For example, one-year-long cycle can be seen in branch M and as the cycle has the similar development in some years, the color of these years in Figure 4 is similar when the level of the cycles as a whole is similar.

5 Conclusion and Possible Application

A new method of visualization of trends of time series that uses Kohonen self-organizing map was presented. Time series with both similar and opposite trends can be easily seen by comparing their colors in the output of the analysis. The method could be implemented into business intelligence technologies where intelligible presentation of results is very important because it is used for human decision making.

References

[1] `http://www.smashingmagazine.com/2007/08/02/data-visualization-modern-approaches/` accessed on April 4, 2008.

[2] "Tab P1 Zaměstnanci podle odvětví" from `http://www.czso.cz/csu/redakce.nsf/i/hdp_cr` accessed on March 21, 2008.

[3] KOHONEN, T. *Self-Organized Formation of Topologically Correct Feature Maps.* In: Biological Cybernetics, Volume 43, Number 1 January, 1982, Heidelberg, Springer-Verlag, pp. 59 – 69. ISSN 0340-1200 (Print), 1432-0770 (Online).

Daniel Němec

Department of Applied Mathematics and Computer Science
Faculty of Economics and Administrations
Masaryk University
Lipová 41a, 602 00 Brno

e-mail: nemecd@econ.muni.cz

Bayesian Estimation of the Unemployment Gap in the Czech Republic[1]

Abstract

Unemployment gap is important indicator that help monetary authority to pursue good economic policy. This indicators is based on differences between observable unemployment rate and and its equilibrium unobservable counterpart. The equilibrium unemployment rate is usually connected to the non-accelerating inflation rate of unemployment (NAIRU) because the issue of monetary stability is incorporated in this theoretical concept. Alternative estimates of the unemployment gap for a small open economy (Czech Republic) are presented and analyzed in this paper. Estimates are made within the framework of the hysteretic approach and non-hysteretic approach. We use techniques of Bayesian analysis (Gibbs sampler and Metropolis-Hastings algorithm) to identify the models. Unobserved states are estimated using Kalman filter. Non-hysteretic model is solved as Dynamic Stochastic General Equilibrium (DSGE) model with rational expectations using the Dynare toolbox. Model estimates indicate hysteretic patterns of the unemployment and imply that actual lowering unemployment rates may be thus sustainable in the long run.

1 Introduction

Output gap and unemployment gap are important indicators that help monetary authority to pursue good economic policy. Both indicators are based on differences between observable macroeconomic variables (unemployment rate and aggregate product) and their equilibrium unobservable counterparts (equilibrium unemployment rate, potential output). The equilibrium unemployment rate is usually connected to the non-accelerating inflation rate of unemployment (NAIRU) because the issue of monetary stability is incorporated in this theoretical concept. Alternative estimates of the unemployment gap for a small open economy (Czech Republic) are presented and analyzed in this paper. Estimates are made within the framework of the hysteretic and non-hysteretic approach. Hysteresis approach is based on simple Phillips curve with adaptive expectations. Hysteresis hypothesis posits a NAIRU that automatically follows the path of actual unemployment rate. Consequently, law of motion

[1]This work was supported by MŠMT project Research centers 1M0524, and funding of specific research at ESF MU.

of equilibrium unemployment is known. Non-hysteretic approach is implemented into the framework of multi-equation dynamic macroeconomic model which includes basic equations determining economic development. Equilibrium unemployment (NAIRU) follows a random walk. In addition, Phillips curve equation is enriched by expected inflation and the model is solved as a model with rational expectations.

We use techniques of Bayesian analysis (Gibbs sampler and Metropolis-Hastings algorithm) to identify the models. Unobserved states are estimated using Kalman filter. Model is solved as Dynamic Stochastic General Equilibrium (DSGE) model with rational expectations using the Dynare toolbox[2].

2 NAIRU, natural rate hypothesis and hysteresis

Key questions facing economists since World War II have been concerned with the causes of changes in the rates of inflation and unemployment and with the relation between the price level and the level of unemployment and its determinants. These questions are related to the theory of NAIRU, natural rate hypothesis (NRH) and hysteresis hypothesis.

Natural rate of unemployment is an equilibrium level of unemployment determined by structural patterns of labour market (see Friedman [2]). NAIRU is rather an empirical value. The theory of NAIRU is a response to monetarist critique regarding Philips curve and its implication for efficiency of economic (monetary) policy. This concept is based on reinterpretation of Phillips curve in a way that brings closer Keynesian and monetarist views of the unemployment-inflation relation but preserve effects of activist demand policy. NAIRU can be interpreted as a rate below which the actual unemployment causes inflation to accelerate. Mechanism behind this is an extension of traditional Phillips curve mechanism [9].

Theory of the NAIRU implies that low unemployment may cause accelerating inflation regardless of sources of this low unemployment. From Keynesian point of view, theory of the NAIRU successfully reformulates the natural rate hypothesis. The level of NAIRU may be a guide to monetary or fiscal policy: unemployment under NAIRU enables demand incentives of economic policy and actual unemployment above NAIRU requests restrictive policy interventions. NAIRU may be accepted as a practical constraint for economic policy too. From the monetarist perspective, the NAIRU is a synonym for the natural rate. The NAIRU theory is misunderstanding of the ineffectiveness of government demand oriented policy. Espinoza-Vega and Russel [1] discuss these issues in more detail. On other hand, hysteresis hypothesis posits NAIRU that automatically follows the path of the actual unemployment. NAIRU and natural rate of unemployment are the same equilibrium values. Any rate of unemployment may be consistent with the stable inflation. The empirical counterpart of the hysteresis framework is a relation in which inflation depends on the change in unemployment and not on the level of unemployment as in the NRH. Existence of hysteresis in the unemployment has important implications discussed in [8].

3 Hysteretic Phillips Curve and the Gibbs Sampler

The model of the Phillips curve, presented in this section, allows us to prove particular probabilities of the natural rate hypothesis, hysteresis hypothesis and the NAIRU theory. The specificity of hysteretic Phillips curve enables us to estimate the trajectory of the NAIRU (or

[2]Software available on `www.cepremap.cnrs.fr/dynare/`

equilibrium unemployment) based on macroeconomic data of the Czech Republic. Hysteretic Phillips curve offers macroeconomic view on hysteresis in unemployment.

3.1 Theoretical concept

Hysteretic Phillips curve is similar to that one presented in [3]. It is the simplest version of the symmetric natural rate hypothesis relating unemployment and inflation:

$$\pi_t = \alpha\pi_{t-1} + \beta(U_t - U_t^*) \tag{1}$$

Inflation rate (year-on-year) is denoted as π_t, unemployment rate is U_t and U_t^* represents equilibrium rate of unemployment or NAIRU. Parameter α indicates the power of adaptive inflation expectations. Hysteresis hypothesis posits a NAIRU that automatically follows the path of actual unemployment rate. Consequently, law of motion of equilibrium unemployment is known. Hysteresis arises when equilibrium unemployment does not depend only on its microeconomic determinants represented by Z_t but also on the lagged unemployment rate U_{t-1}. These microeconomic determinants correspond to actual structural characteristics of the labour and commodity markets expressed by Friedman. Assuming that hysteresis exists, we obtain

$$U_t^* = \eta U_{t-1} + Z_t \tag{2}$$

Full hysteresis occurs when $\eta = 1$, which means that there is not unique U_t^*. Substituting equation (2) into (1) results in

$$\pi_t = \alpha\pi_{t-1} + \beta(U_t - \eta U_{t-1} - Z_t) \tag{3}$$

By transforming (3), we can see that full hysteresis implies a relation between inflation and the change of unemployment, not the level of unemployment. That is why hysteresis implies that any rate of unemployment may be consistent with steady inflation.

$$\pi_t = \alpha\pi_{t-1} + \beta(1-\eta)U_t + \beta\eta(U_t - U_{t-1}) - \beta Z_t \tag{4}$$

Estimated model is a normal regression model with parameters λ_i and Normally distributed error term ϵ_t with mean 0 and variance σ^2.

$$\pi_t = \lambda_1 + \lambda_2\pi_{t-1} + \lambda_3 U_t + \lambda_4(U_t - U_{t-1}) + \epsilon_t \tag{5}$$

3.2 Estimation techniques - Gibbs sampler

We estimate the parameters of the normal linear regression model with independent normal-gamma prior using Gibbs sampler. The Gibbs sampler is a powerful tool for Bayesian posterior simulation.

3.2.1 The likelihood function

We assume $\epsilon_t \sim N(0, h^{-1})$. Parameter h denotes the error precision which is defined as $\frac{1}{\sigma^2}$. The likelihood function may be written as:

$$p(y|\lambda, h) = \frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}} \left\{ \exp\left[-\frac{h}{2}(y - X\lambda)'(y - X\lambda) \right] \right\}$$

Vector y is the $N \times 1$ vector of dependent variable and X is the $N \times k$ matrix of the k explanatory variables and λ is the $k \times 1$ vector of the parameters.

3.2.2 The prior

We assume $p(\lambda, h) = p(\lambda)p(\beta)$ with $p(\lambda)$ being Normal and $p(h)$ being Gamma:

$$\lambda \sim N(\underline{\lambda}, \underline{V})$$

$$h \sim G(\underline{\nu}, \underline{s}^{-2})$$

The prior density may be enhanced by the information that $\eta \in (0,1)$ (i.e. we define set A where $\eta \in (0,1) \Leftrightarrow \lambda \in A$). We use indicator function $1(\lambda \in A)$, which equals 1 if $\lambda \in A$ and 0 otherwise.

$$p(\lambda) = \frac{1}{(2\pi)^{\frac{k}{2}}} |\underline{V}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\lambda - \underline{\lambda})'\underline{V}^{-1}(\lambda - \underline{\lambda})\right] \cdot 1(\lambda \in A)$$

$$p(h) = c_G^{-1} h^{\frac{\underline{\nu}-2}{2}} \exp\left(-\frac{h\underline{\nu}}{2\underline{s}^{-2}}\right)$$

where c_G is the integrating constant for the Gamma p.d.f. We use following notation: $\underline{\lambda} = E(\lambda|y)$ is the prior mean of λ and the prior mean and degrees of freedom of h are $\underline{s}^{-2}$ and $\underline{\nu}$ respectively. $\underline{V}$ is simply the prior covariance matrix of λ.

3.2.3 The posterior

The posterior is proportional to the prior times the likelihood:

$$p(\lambda, h|y) \propto \left\{\exp\left[-\frac{1}{2}\left\{h(y - X\lambda)'(y - X\lambda) + (\lambda - \underline{\lambda})'\underline{V}^{-1}(\lambda - \underline{\lambda})\right\}\right]\right\} \cdot 1(\lambda \in A)$$

$$\times h^{\frac{N+\underline{\nu}-2}{2}} \exp\left[-\frac{h\underline{\nu}}{2\underline{s}^{-2}}\right]$$

This joint posterior density does not take the form of any known density. We are able to to carry out posterior inference using Gibbs sampler. For further details see [6] and note that our indicator function must be implemented in the appropriate equations. Bayesian approach allows us to derive Bayes factors comparing probabilities of particular model. We estimate model probabilities (Table 2) using Savage-Dickey density ratio (see [6]). This method is a way of writing the Bayes factor for comparing nested models.

4 Dynamic macroeconomic model and the Metropolis-Hastings algorithm

4.1 Theoretical concept

This section presents a multi-equation dynamic (monetary) macroeconomic model of the economy with rational expectations. There are not any explicit assumptions about possible hysteretic patterns of the unemployment in the economy. It is an equilibrium model which connects output dynamic, unemployment dynamic and inflation dynamic. The variables NAIRU and potential output are modelled as unobserved states of the economic system. The model concept is based on approach proposed by Laxton and Scott [7]. It consists of system of indisputable causal equations.

Output gap definition

$$ygap_t = 100 \cdot (gdp_t - \overline{gdp}_t) \tag{6}$$

The definition of output gap ($ygap_t$) corresponds to difference between logs of gross domestic product (gdp_t) and potential output ($\overline{gdp}_t$).

Output gap dynamics

$$ygap_t = \alpha_1 ygap_{t-1} - \alpha_2 rgap_{t-1} - \alpha_3 zgap_{t-1} + \epsilon_t^{ygap} \tag{7}$$

This equation implies relation between actual output gap and its lagged value. Output gap is influenced by lagged real interest rate ($rgap_{t-1}$) and real exchange rate ($zgap_{t-1}$). Shock to output gap is denoted as ϵ_t^{ygap}.

Unemployment gap definition

$$ugap_t = u_t - \overline{u}_t \tag{8}$$

Unemployment gap ($ugap_t$) is defined as deviation of actual unemployment (u_t) from its equilibrium ($\overline{u}_t$) represented by the NAIRU.

Stochastic process of potential output

$$\overline{gdp}_t = \gamma_t + \overline{gdp}_{t-1} + \epsilon_t^{\overline{gdp}} \tag{9}$$

where

$$\gamma_t = \beta\gamma^{ss} + (1-\beta)\gamma_{t-1} + \epsilon_t^{\gamma}$$

Unobserved potential output ($\overline{gdp}_t$)is modelled as random walk. This specification assumes that there is shock ($\epsilon_t^{\overline{gdp}}$), directly affecting the level of potential output each time period. In addition, there can be persistent deviations in the trend growth rate of potential output (γ_t) from a constant steady-state growth rate (γ^{ss}).

Stochastic process of the NAIRU

$$\overline{u}_t = \overline{u}_{t-1} + \epsilon_t^{\overline{u}} \tag{10}$$

The NAIRU is modelled as random walk with the shock $\epsilon_t^{\overline{u}}$.

Unemployment gap dynamics

$$ugap_t = -\phi_1 ygap_t + \phi_2 ugap_{t-1} + \epsilon_t^{ugap} \tag{11}$$

This equation is a dynamic Okun's law which connects output gap and unemployment gap. Possible persistence and exogenous shock are represented by lagged unemployment gap and by ϵ_t^{ugap} respectively.

Inflation equation (Phillips curve)

$$\pi_t = \delta_1\pi_t^m + \delta_2 E_t\pi_{t+4} + (1-\delta_1-\delta_2)\pi_{t-1} - \delta_4 ugap_t - \delta_5 \Delta ugap_t + \epsilon_t^{\pi} \tag{12}$$

The last equation is a standard open-economy inflation equation. In this equation year-on-year inflation depends on past changes in import price inflation (π_t^m), expectations of inflation over the next year $E_t\pi_{t+4}$ as well as on the unemployment gap and its change (pressures from labour market). Thus, the model is backward-looking and forward-looking as well. Possible inflation shocks are added by ϵ_t^{π}.

Table 1: Estimation results - Hysteretic model

Parameter	*Prior mean (std. deviation)*	*Posterior mean (std. deviation)*	*Geweke's CD*	*Bayes factor* ($\lambda_i = 0$)
λ_1	2.0000	2.3339	0.2637	0.0528
	(1.5000)	(0.8058)		
λ_2	0.5000	0.7450	-0.5451	0.0000
	(0.2500)	(0.0747)		
λ_3	0.5000	-0.2443	-0.0514	0.2726
	(1.0000)	(0.0921)		
λ_4	0.5000	-0.7848	0.3689	0.5026
	(1.0000)	(0.4259)		
	α	β	η	Z
Structural parameters	0.7450	-1.0291	0.7627	2.268

4.2 Estimation techniques

Due to unobservable variables in the model, we need to use filtration techniques. Model parameters and the trajectories of unobservable states are estimated using Dynare toolbox for Matlab. Dynare is a powerful tool which use techniques of Bayesian inference, in particular Random Walk Metropolis-Hastings algorithm (see [6] for details). Unobservable states (including shocks) are filtered and smoothed by implemented Kalman filter. The model is taken as dynamic stochastic general equilibrium model (DSGE). This allows us to solve rational expectations occured in the model. Model parameters and the NAIRU have been estimated (filtered and smoothed) separately by extended Kalman filter (as described in [4]). Both NAIRU estimates are presented in the figures below.

5 Estimation results

We use quarterly seasonally adjusted macroeconomic data of the Czech Republic from the first quarter 1996 to the third quarter 2007. Observable variables are year-on-year net inflation, unemployment rate, real output, import price inflation and the gaps of real interest rate and real exchange rate.

Estimation results for the hysteretic model are presented in Table 1. Empirical first and second moments of marginal densities of the parameters are presented (second and third column). Geweke's convergence diagnostic CD (see citeKoop2003) is an indicator of convergence of the Gibbs sampler. It seems that the parameters have converged to the joint posterior density which we are interested in. Bayes factor (post odds ratio) is calculated using Savage-Dickey density ratio (last column of the table). It shows ratio of probabilities of a restricted and unrestricted model. The restricted model is the model with $\lambda_i = 0$. The Table 2 presents probabilities of particular models, i.e. the probabilities that natural rate hypothesis ($\eta = 0$), full hysteresis ($\eta = 1$) or theory of the NAIRU ($\eta \in (0;1)$) can be validated using the data.

Table 2: Posterior models probabilities

$\eta = 0$	$\eta = 1$	$\eta \in (0, 1)$
0.2831	0.1537	0.5632

Table 3: Estimation results - Dynamic model

Parameter	*Prior mean*	*Posterior mean*	*HPDI*		*Prior*	*Prior s.d.*
α_1	0.400	0.6499	0.5277	0.8032	*Beta*	0.1000
α_2	0.200	0.1763	0.0578	0.3084	*Beta*	0.1000
α_3	0.150	0.1507	0.0437	0.2796	*Normal*	0.3000
β	0.800	0.8178	0.6792	0.9526	*Beta*	0.1000
ϕ_1	0.200	0.1615	0.0666	0.2448	*Beta*	0.1000
ϕ_2	0.850	0.9795	0.9613	0.9993	*Beta*	0.1000
δ_1	0.200	0.0847	0.0477	0.1174	*Beta*	0.0500
δ_2	0.300	0.1229	0.0669	0.2246	*Beta*	0.1000
δ_4	0.500	-0.0351	-0.0694	-0.0001	*Normal*	0.2000
δ_5	0.500	0.4051	0.1221	0.7165	*Normal*	0.2000
ϵ^{ygap}	1.000	0.7181	0.3143	1.1675	*Inv. Gamma*	30.0000
$\epsilon^{\overline{gdp}}$	0.500	0.0841	0.0664	0.1015	*Inv. Gamma*	30.0000
ϵ^{γ}	0.500	0.0842	0.0666	0.1027	*Inv. Gamma*	30.0000
ϵ^{NAIRU}	0.500	0.1234	0.0867	0.1553	*Inv. Gamma*	30.0000
ϵ^{ugap}	0.500	0.1213	0.0862	0.1557	*Inv. Gamma*	30.0000
ϵ^{π}	0.500	0.6705	0.4768	0.7955	*Inv. Gamma*	30.0000

We are able to simulate the trajectory of the NAIRU and unemployment gap using the values of estimated structural parameters. The signs of the structural parameters are in accordance with economic theory. Increasing unemployment leads do decreasing inflation rates, high unemployment tends to cut inflation pressures (arising from labour market) as well. The value of parameter η suggests that there is hysteresis in unemployment in the Czech economy. Relative high value of estimated parameter α indicates strong adaptivity in inflation expectations. This may imply high confidence in inflation targeting and in the credibility of the Czech National Bank.

Table 3 present estimation results for model parameters (dynamic model) and standard deviations of the shocks. The signs of the parameters are in accordance with economic theory. The values of estimated parameters ϕ_1 and ϕ_2 suggests that there is high persistence of unemployment gap. The influence of output gap is relatively weak. From this point of view, empirical validity of Okun's law is not confirmed in the Czech Republic.

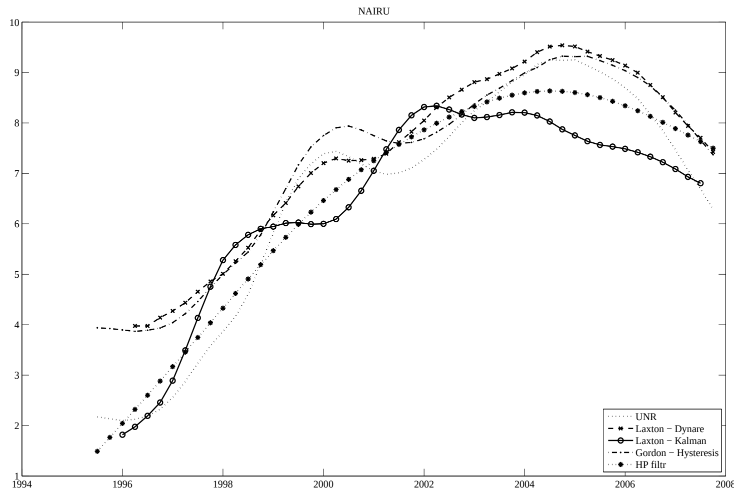


Figure 1: Estimates of the NAIRU

Estimated parameters in the inflation equation indicate similar dependence of the lagged inflation as hysteretic Phillips curve implies. Influence of the imported inflation is very small. That is amazing because openness of the Czech Republic is quite large. Explanation may be seen in the long term appreciation of the Czech currency. Standard deviations of the shocks characterize variability of the related quantities. Output gap and inflation are the most variable of them.

Estimates of the NAIRU made by Hodrick-Prescott (HP) filter (see [5]) are added to the set of our alternative estimates. We can see that estimated paths of NAIRU (unemployment gap) differ in levels (the difference amounts to 2 percentage points). Dynamic of these unobserved states are very similar. All estimates imply similar changes and turning points.

As mentioned in the introduction, the level of the NAIRU determines the equilibrium unemployment. Alternative estimates of the NAIRU are presented in Fig. 1. Equilibrium unemployment has been increasing significantly since 1996. This may be connected with the transformation process and with structural changes in the economy. It must be pointed out that the trajectory of NAIRU in the framework of hysteretic Phillips curve (denoted as Gordon-hysteresis) is very similar to Laxton approach (Laxton-Dynare). It seems that hysteresis hypothesis play essential role in the description of unemployment pattern in the Czech economy. Monetary shock (or restriction) in 1997 influenced low unemployment rates that prevailed before 1996. This shock has amplified structural changes in the economy. Inflation had been decelerating and unemployment rate grown. Since 2000 the inflation has been stabilised at relative low rates but high and persistent unemployment prevailed as well as equilibrium unemployment (in accordance with hysteresis hypothesis).

There is a break in the trend of unemployment in 2005. One factor causing decreasing unemployment may be changes in methods of calculations of unemployment. More plausible explanation is connected with the demand-oriented government policy (including investment

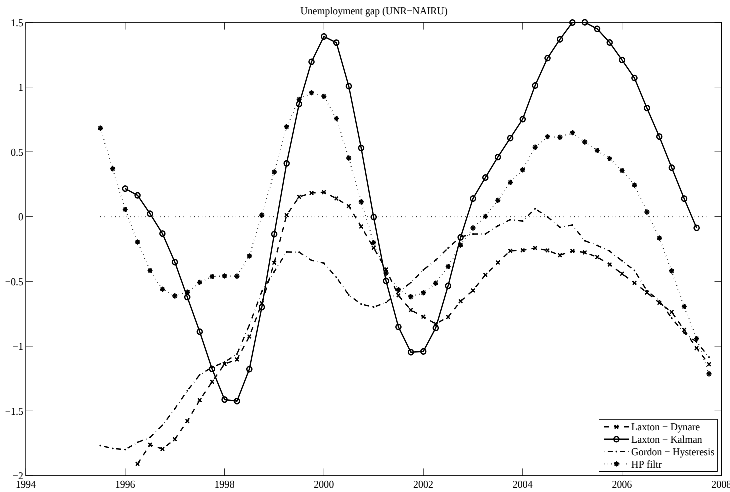


Figure 2: Unemployment Gap

incentives and debt financing of the government spending). Unemployment has decreased and inflation has not accelerated. That is next reasoning in favour of hysteresis effects in the Czech economy.

Unemployment gap (see Fig. 2) in the last two years suggests that actual unemployment is sustainable. Moderate inflation pressures resulting from negative unemployment gap are induced especially by lack of employees. Aged employees retire and the size of labour force is shrinking. This is a natural process not included in the framework of presented models. Higher rate of decrease of the unemployment rate (compared with development of equilibrium unemployment) may be understand as a temporally shock which disappears in the near future.

6 Conclusion

Bayesian estimates of equilibrium unemployment (NAIRU) and unemployment gap imply that the state of the Czech economy in present corresponds to the potential state. Unemployment gap does not mean that the labour market is the source of monetary instability. Actual high growth rates of GDP and lowering unemployment rates may be thus sustainable in the long run.

Moreover, the model estimates indicate hysteretic patterns of the unemployment. Because the hysteresis is identified in the Czech economy, we can accept that unemployment could be reduced (and probably has been reduced in the case of the Czech Republic) by using expansionary policies without negative inflation consequences. The NAIRU is thus compatible with any level of unemployment as the hysteresis hypothesis suggests.

References

[1] Espinoza-Vega, M. A., Russell, S.: History and Theory of the NAIRU: A Critical Review. Economic Review, Vol. 82, No. 2, 4–25 (1997)

[2] Friedman, M.: The Role of Monetary Policy. The American Economic Review, Vol. 58, No. 1, 1–17 (1968)

[3] Gordon, R.J.: Hysteresis in History: Was There Ever a Phillips Curve? The American Economic Review, Vol. 79, No. 2, May, 220–225 (1989)

[4] Hamilton, J.: Time Series Analysis.Princeton University Press (1994)

[5] Hodrick, R., Prescott, E.C.: Post War U.S. Business Cycles: an Empirical Investigation. Journal of Money, Credit and Banking, Vol. 29, 1-16 (1997)

[6] Koop, G.: Bayesian Econometrics. Wiley (2003)

[7] Laxton, D., Scott, A.: On developing a Structured Forecasting and Policy Analysis System Designed to Support Inflation Targeting (IFT), Inflation Targeting Experiences: England, Finland, Poland, Mexico, Brazil, Chile, The Central Bank of Turkey (2000)

[8] Němec, D., Moravanský, D.: Testing of Hysteresis in Unemployment. In Proceedings of the 24th International Conference Mathematical Methods in Economics 2006. Plzeň : University of West Bohemia in Pilsen, 407–414 (2006)

[9] Phillips, A. W.: The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861-1957. Economica, November, 283–297 (1958)

Václava Pánková[1]

Department of Econometrics
Faculty of Informatics and Statistics
University of Economic
W. Churchill Sq. 4, 130 67 Praha 3

e-mail: pankova@vse.cz

FDI Inflows and Exchange Rate

Abstract

Foreign direct investment (FDI) inlow is an important factor of an economic development. A special role of exchange rates in explaining FDI is stressed in this paper. FDI is studied from the point of view of capital and technology acceptors while respecting a basic fact that investor maximizes his profit by allocating FDI according to the exchange rate volatility, i.a. after a sudden large devaluation of the host country currency large FDI inflows follow as future appreciation is expected. Large exchange rate shocks are described by the help of skewness of exchange rate changes. Having host country currency (*hcc*) to USD ratio, positive skewness means that depreciations of *hcc* occur more often. A model explaining FDI by mean, standard deviation and skewness of changes of exchange rate is specified, application to four selected economies is presented using panel data. The results support the hypothesis of a crucial influence of large changes in exchange rate on FDI.

Keywords: *Foreign direct investment, statistical moments of exchange rate changes, panel data*

1 Introduction

Foreign direct investment (FDI) is an important phenomenon of international financial relations. Usually, FDI is explained by the help of labor cost, income difference, purchasing power, export and exchange rate. Studying FDI with regard to one economy only, we can use the exchange rate as one of important explanatory variables. If we intend to handle the exchange rate as a particular quality of each economy which enables to distinguish for an expected profit and hence, affects the decision whether to invest or not in a certain country, the equation should comprise relevant characteristics of all countries to be compared. A special role of exchange rates in explaining FDI applies to the new members of European Union. Those countries, still having their own currencies, converge to introducing euro in a horizon of five years approximately. In principle, a European Monetary Union (EMU) accession is an interesting subject also for UK economists.

Here, FDI is studied from the point of view of capital and technology acceptors while respecting a basic fact that profit is the main investor's interest. Investor can drive for maximizing the

[1]Financial support of GAČR 402/06/0049 is gratefully acknowledged

expected profit by allocating FDI according to the exchange rate volatility, i.a. after a sudden large devaluation of the host country currency large FDI inflows follow as future appreciation is expected. Large exchange rate shocks are described with the help of skewness. Negative skewness means that large differences occur more often. Reasoning of the model explaining FDI by mean, standard deviation and skewness of changes of exchange rate is provided. Application to two new EU members and two ASEAN countries is presented using panel data and seemingly unrelated regression techniques.

Models explaining FDI only upon exchange rate changes clearly do not belong to the main stream of FDI models. A strong point of the model dealing with statistical moments of exchange rate changes only, is a low number of explanatory variables which are based purely on the knowledge of one variable.

Chakrabarti and Scholnick [3] give a reasoning of relations between FDI and exchange rate changes and show that the higher is the current rate (USD / HostCountryCurrency), the lower is (*i*) the expected appreciation of HCC, (*ii*) the optimal scale of project, (*iii*) the expected profit, (*iv*) the FDI inflow.

They argue that FDI adjust to large shocks of exchange rate and not to small shocks. Shocks, in fact, are changes in the exchange rate. A mechanism allowing for distinguishing between small and large shocks is described by Ball and Mankiw [1] who utilize the skewness of changes inprices as a relevant explanatory variable of inflation in the following manner.

(i) A positively skewed distribution results from many small relative price decreases which cannot compensate more large relative price increases and inflation rises (see Figure 1).

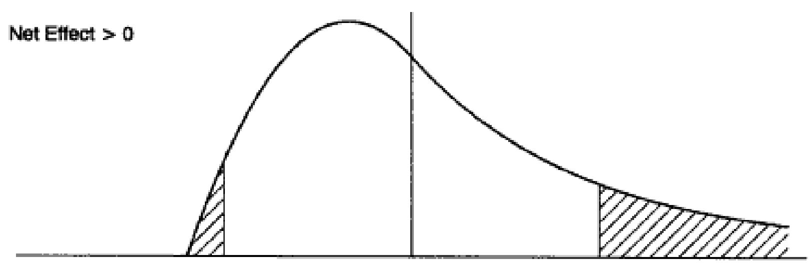


Figure 1

(ii) On the opposite, negative skewness represents an excess of relative price decreases, hence inflation falls (see Figure 2).

In the context of monthly changes of the exchange rate, positive skewness means that depreciations occur more frequently than appreciations.

2 Model and results

An econometric model is proposed in [3] in the form

$$FDI_{it} = \beta_0 + \beta_1 Mean_{it} + \beta_2 StDev_{it} + \beta_3 Skew_{it} + u_{it} \tag{1}$$

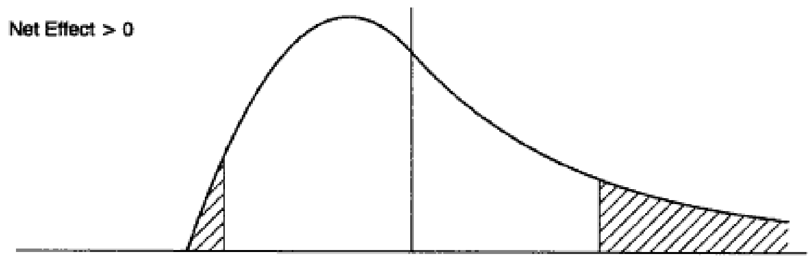


Figure 2

in which $Mean_{it}$, $StDev_{it}$, $Skew_{it}$ respectively are the mean, standard deviation and skewness of differences in monthly exchange rate values of the currency of country i during the year $t-1$. FDI refers to country i and year t. βs are parameters of the model and $u_{it} \approx N(0, \sigma^2)$ represent disturbances.

The theory proposed above should result in a stronger influence of $Skew_{it}$ variable than that of the other exogenous variables.

Data refers to years 1994–2005. Except of FDI inflows to the Czech Republic, the following countries are studied: Hungary as an economy of comparable dimension and similar historical experience; both countries are new members of the EU. Besides, two members of ASEAN: Thailand which is in the move economically and Philippines which provides a different experience as its FDI inflow is rather weak. Another interesting link between the chosen countries is the fact that at the beginning of the analyzed period, Czech Crown, Thai Baht and Philippine Peso were near to parity.

In Table 1, a survey of correlations between exchange rates, respective FDIs, is given

	Philippines		Hungary		Czech Republic	
	ER	FDI	ER	FDI	ER	FDI
Thailand	0.398	0.607	0.407	−0.312	0.741	−0.233
Philippines			0.347	0.041	0.322	−0.075
Hungary					0.670	0.435

Table 1: Correlations between exchange rates, resp. FDI

Using panel data allowing for distinguishing in constants for the countries (or group of countries), the estimates which are summarized in Table 2 were obtained. The results correspond to the findings of Chakrabarti and Scholnick [3] and show no importance of mean and standard deviation of exchange rate devaluations while relevant skewness plays a reasonably significant role. Having Thailand as a reference country, there are shifts in constant computed. The results correspond to the evidence that there is significantly less FDI in the Philippines compared to Thailand. Czech Republic is in a slightly better position and Hungary is indifferent.

Using the same technique to distinguish between EU and ASEAN countries, no difference is found. Hence, we do not confirm a rather often used opinion that a "new member of EU" characteristic could be a special quality in gaining FDI inflows.

	Thailand (ref.category)	Philippines	Hungary	Czech Republic
Constant (st.err.)	**2826**(303)	–**1836**(112)	**66**(798)	**1826** (185)
t-value (t-prob)	9.33 (0.000)	–16.4 (0.000)	0.083(0.93)	9.85(0.000)
Mean (st.err.)	–**20.3**(0.177)			
t-value (t-prob)	–0.177(0.860)			
Stand.dev.(st.err.)	**256**(258)			
t-value (t-prob)	0.992 (0.328)			
Skew (st.err.)	**533**(283)			
t-value (t-prob)	1.88 (0.068)			

Table 2: Results of panel data estimates

Seemingly unrelated regression (SUR), performed under an assumption of common world economic environment, gives a possibility to find individual parameters for each country. In this case we see (Table 3) more detailed and unfortunately also less unambiguous results. Evidently, skewness plays no role in case of Thailand and Czech Republic which are exemplary FDI acceptors.

	Thailand	Philippines	Hungary	Czech Republic
Constant (st.err.)	**242**(122)	**89**(62)	**229**(172)	**104**(252)
t-value (t-prob)	1.98(0.057)	1.42(0.165)	1.33 (0.195)	0.413 (0.682)
Mean (st.err.)	**2307**(445)	**1035**(339)	**156**(124)	–**2346**(1299)
t-value (t-prob)	5.18(0.000)	3.05(0.005)	1.26(0.218)	1.81(0.080)
Stand.dev.(st.err.)	**2148**(232)	**716**(106)	**749**(79)	**5801**(691)
t-value (t-prob)	9.26(0.000)	6.75(0.000)	9.37(0.000)	8.39(0.000)
Skew (st.err.)	–**355**(481)	–**390**(175)	**1618**(468)	**94**(753)
t-value (t-prob)	–0.738(0.466)	–2.22(0.034)	3.45(0.002)	0.125(0.901)

Table 3: Results of SUR estimates

3 Conclusions

An idea of a crucial influence of large changes in exchange rate on an FDI inflow is recapitulated with a target to verify it in case of two middle European and two ASEAN economies. Using relevant data, it is found that in general, the hypothesis is not to be rejected. But, following the countries as single cases with one world economy as a connecting factor, there is an evidence in case of Thailand and Czech Republic, that successful FDI acceptors are probably viewed from more complex perspectives which are hardly to be reduced in a small number of quantified variables. Also a frequent opinion that being a new member of EU represents a special quality in gaining FDI inflows can not be confirmed here.

References

[1] Ball, L.; Mankiw, N.: Relative Price Changes as Aggregate Supply Shocks, *Quarterly Journal of Economics* 110, 1, 1995, pp 161 - 193

[2] Barrell, R.; Gottschalk, StateS.D.; Hall, S.G.: Foreign Direct Investment and Exchange Rate Uncertainty in Imperfectly Competitive Industries, `www.niesr.ac.uk/pubs/dps/dp220.pdf`, 2006

[3] Chakrabarti, R.; Scholnick, B.: Exchange Rate Expectations and FDI Flows, Faculty of Business, University Alberta, `www.prism.gatech.edu/~rc166Ex.FDI.pdf`, 2006

[4] Chowdhury, A.; Mavrotas, G.: FDI & Growth: What Cause What, WIDER Conference. Helsinky, 2003

[5] Hušek, R., Pánková, V.: Exchange Rate Changes Effects on FDI, Prague Economic Papers, 2, 2008

[6] Patrawimolpon Pichit.; Pongsaparn Runchana: Thailand in the New Asian Economy, Bank of Thailand Symposium 2006

[7] Sung, H.; Lapan, H. E.: Strategic Foreign Direct Investment and Exchange-Rate Uncertainty, *International Economic Review*, Vol. 41, No 2, 2000, pp 411 - 423

Jan Pelikán, Jan Fábry, Václav Kořenář

Department of Econometrics
University of Economics Prague
W. Churchill sq. 4, 130 67 Prague 3

e-mail: pelikan@vse.cz, fabry@vse.cz, korenar@vse.cz

Pickup and Delivery Problem

Abstract

Vehicle routing problem and traveling salesman problem are classical problems in operational research, this modification of those problems consists of a transport among nodes of the communication network using cyclical routes of vehicles with given capacity. A transportation demand is given by the place of pickup, the place of delivery and quantity of goods. The goal is to find cyclical routes of minimal length which ensure the transport requirements. In the paper there are two models proposed for the problem, both are demonstrated on an example. The problem is based on a case study from practice.

Keywords: *Pickup and delivery problem, integer programming, heuristic methods*

1 Introduction

Traveling salesman problem and vehicle routing problem and their modifications are frequently solved practical examples. In such problems the goal is to assure pick-up or/and delivery of goods in a distribution network. While in those applications it is necessary to find cyclical routes starting and ending in a depot, in the studied problem it is required to transport goods between nodes of the network. Each requirement is specified by the pickup point, the delivery point and the amount of goods which has to be transported. In [1] and [2] the problem is called pickup and delivery problem.

In the practice time windows are often given for each node in the network and the order of pickup and delivery processes has to be defined, because the goods being loaded as the last one will be unloaded as the first one. The origin and the destination nodes need not to be identical. Vehicles with different capacities can assure the transport of goods. The objective is to minimize the total transportation cost. Column generation method is used for selecting routes in the mathematical model.

In the paper, no time windows and no conditions for order of pickup and delivery are considered. All vehicles have the same capacity. Each route has to be cyclical for all included nodes. Two models are proposed in the paper; the first one is based on the optimal flow theory, the second one uses a model of set covering problem.

Let a distribution network is given by $G = \{V, E\}$, where V is a set of n nodes and E is a set of undirected arcs. Each arc (i, j) is evaluated by minimal distance between nodes i and j. Let us denote q_{kl} the amount of goods that has to be transported from node k to node

l. Vehicles with capacity V are used for pickup and delivery and they can start in any node. All routes have to be cyclical, each vehicle has to come back to the node it starts from. The objective is to minimize the length of all the routes.

Example 1. *Let us consider the distribution network with 4 nodes; the distance matrix C and the transportation requirements matrix Q are given:*

$$C = \begin{bmatrix} 0, 30, 40, 60 \\ 30, 0, 60, 50 \\ 40, 60, 0, 20 \\ 60, 50, 20, 0 \end{bmatrix} \qquad Q = \begin{bmatrix} 0, 0, 7, 5 \\ 0, 0, 5, 7 \\ 5, 6, 0, 0 \\ 8, 0, 8, 0 \end{bmatrix}.$$

Vehicle capacity is $V = 12$.

2 Optimal multi-product flow model

Let us define two additional nodes in the network: the node 0 as the source and the node $n+1$ as the sink. The following variables are defined in the model:

$y_{ij} \geq 0$, integer – number of vehicles going through the arc $(i,\ j)$ in the direction from i to j $(i, j = 0, 1, ..., n+1,\ i \neq j)$,

$x_{ij}^{kl} \geq 0$, amount of goods (part of the total amount q_{kl}) transported from node i to node j $(i, j = 0, 1, ..., n+1,\ i \neq j; k, l = 1, 2, ..., n,\ k \neq l)$.

Mathematical model follows:

$$z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} y_{ij} \rightarrow \ \min \tag{1}$$

$$\sum_{i=1}^{n} y_{ij} = \sum_{i=1}^{n} y_{ji},\ j = 1, 2, ..., n, \tag{2}$$

$$\sum_{i=0}^{n} x_{ij}^{kl} = \sum_{i=1}^{n+1} x_{ji}^{kl},\ j = 1, 2, ..., n; k, l = 1, 2, ...n, k \neq l, \tag{3}$$

$$\sum_{k,l} x_{ij}^{kl} \leq V y_{ij}, i, j = 1, 2, ..., n, i \neq j, \tag{4}$$

$$\begin{array}{l} x_{0,k}^{kl} = q_{kl}\ ,\ k, l = 1, 2, ..., n, k \neq l; \quad x_{0,j}^{kl} = 0,\ k, l = 1, 2, ..., n, k \neq l; j = 1, 2, ..., n+1, j \neq k, \\ x_{l,n+1}^{kl} = q_{kl}\ ,\ k, l = 1, 2, ..., n, k \neq l; \quad x_{j,n+1}^{kl} = 0,\ k, l = 1, 2, ..., n, k \neq l; j = 0, 1, ..., n, j \neq l, \end{array} \tag{5}$$

$$\begin{array}{l} x_{ij}^{kl} \geq 0,\ i, j = 0, 1, ..., n+1,\ i \neq j,\ \ k, l = 1, 2, ..., n,\ k \neq l, \\ y_{kl} \geq 0\,,\ \text{integers},\ \ k, l = 1, 2, ..., n,\ k \neq l. \end{array} \tag{6}$$

The objective (1) corresponds to the sum of the evaluations of all the arcs in the solution, i.e. the total length of all routes. Equations (2) assure the vehicle will leave the location that it will visit. With respect to equations (3), amount of goods being transported from k to l entering node j leaves this node. Inequalities (4) disable exceeding the capacity of the vehicle transporting goods between nodes i and j. Equations (5) assure that both the total flow from the source 0 to node k and the total flow from node l to the sink $n+1$ are equal to the total requirement q_{kl}. All other flows from the source and to the sink are set to 0.

Example 2. *Application of the model* (1) - (6) *to the example introduced above leads to the objective value equal to 260 and to the following values of variables:*

$$y_{12} = y_{13} = y_{24} = y_{34} = 1, \quad y_{31} = y_{43} = 2.$$

Hence, two routes are generated:

route A: 2 $\rightarrow$ 4 $\rightarrow$ 3 $\rightarrow$ 1 $\rightarrow$ 2 of the length 140

and route B: 1 $\rightarrow$ 3 $\rightarrow$ 4 $\rightarrow$ 1 of the length 120.

Optimal solution:

Route A			Route B		
Node	Pickup	Delivery	Node	Pickup	Delivery
2	$q_{24} = 7$ $q_{23} = 5$		1	$q_{14} = 5$ $q_{13} = 7$	
4		$q_{24} = 7$	3		$q_{13} = 7$
3	$q_{31} = 5$ $q_{32} = 6$	$q_{23} = 5$	4	$q_{41} = 8$	$q_{14} = 5$
1		$q_{31} = 5$	1		$q_{41} = 8$
2		$q_{32} = 6$			

Note: Values of matrix Y provide information about the arcs that are included in the optimal routes. Generation of routes based on this information need not to be unique. In addition it is necessary to determine the depot for each route. Selection of the depot has to enable flows to be realized. If node 4 is selected as the depot for the route A, it will not be possible to realize the requirement $q_{23} = 5$. Thus, it is possible we are not able to complete routes with depots on the base of the optimal matrix Y.

3 Routes generation model

The model is based on the assumption there are proposed routes satisfying all conditions of transportation, depots in all proposed routes are determined. The goal is to select routes satisfying requirements given by matrix Q and minimizing the total length of all routes derived from matrix C. A number of vehicles realizing the transport will be determined as well.

Let us assume S routes including arcs of distribution network, the length of each route is denoted $d_s (s = 1, 2, \ldots, S)$. Parameter $a_{ij}^{kl}(s)$ equals 1 if goods transported on the route s from node k to node l uses arc (i, j), 0 otherwise.

Example 3. *Parameters $a_{ij}^{kl}(s)$ for the route A: 2 $\rightarrow$ 4 $\rightarrow$ 3 $\rightarrow$ 1 $\rightarrow$ 2 presented in previous chapter are defined in the following table:*

(i,j) / (k,l)	2,4	2,3	4,3	4,1	3,1	3,2
(1,2)						1
(3,1)				1	1	1
(2,4)	1	1				
(4,3)		1	1	1		

In the model, the following variables are used:

$y_s \geq 0$, integer – number of vehicles on the route $s(s = 1, 2, \ldots, S)$,

$x_{kl}^s \geq 0$, amount of goods (part of the amount q_{kl}) transported on the route $s(k, l = 1, 2, ..., n,\ k \neq l;\ s = 1, 2, ..., S)$.

Mathematical model is:

$$z = \sum_{s=1}^{S} y_s d_s \rightarrow \ \min \tag{7}$$

$$\sum_{s=1}^{S} x_{kl}^s = q_{kl},\ k, l = 1, 2, ...n, k \neq l, \tag{8}$$

$$\sum_{k,l} a_{ij}^{kl} x_{kl}^s \leq V y_s,\ \ i, j = 1, 2, ..., n, i \neq j,\ \ s = 1, 2, ..., S, \tag{9}$$

$$x_{kl}^s \geq 0, k, l = 1, 2, ..., n,\ k \neq l,\ \ y_s \geq 0\,,\ \text{integer},\ \ s = 1, 2, ..., S. \tag{10}$$

The objective function (7) corresponds to the total length of all routes. Equations (8) assure transport of required amount of goods q_{kl} from node k to node l. Inequalities (9) disable exceeding the capacity of the vehicle transporting goods on the route s between nodes i and j. The key issue in this mathematical model is generation of the routes and their number. The same routes with the different depots are considered as different routes in this model.

References

[1] Savelsbergh, M., P., M., Sol., M.: The general pickup and delivery problem. *Transportation Res.* 29 , 1995,17-29.

[2] Hang, X., Zhi-Long, Ch., Rajagopal, S., Arunapuram, S.: Solving a practical pickup and delivery problem. *Transportation science* 2003, vol.37, no.3, 347-364.

[3] Pelikán, Jan: *Diskrétní modely v operačním výzkumu.* Professional Publishing. 2001. ISBN 80-86419-17-7.

Oleksandr Perekhozhuk, Thomas Glauben

Institute of Agricultural Development in Central and Eastern Europe
Theodor-Lieser-Straße 2, D-06120 Halle (Saale), Germany

e-mail: perekhozhuk@iamo.de, glauben@iamo.de

Michael Grings

Institute of Agricultural and Nutritional Sciences
Martin Luther University Halle-Wittenberg
Halle (Saale), Germany

e-mail: michael.grings@landw.uni-halle.de

An Econometric Approach to Measuring the Degree of Oligopsony Power and Conjectural Variations in the Ukrainian Milk Processing Industry

Abstract

There has been a recent increase in the number of New Empirical Industrial Organization (NEIO) studies that measure and test for the degree of market power on agricultural and food markets in developed market economies and that have largely focused on the USA and European food sectors. But imperfect competition seems especially prevalent in the food sectors of transition countries. A technique for assessing the degree of oligopoly market power, based on contributions by Bresnahan (1982) and Lau (1982), is used here to measure the degree of oligopsony power in food processing. As an example, we provide an application to the Ukrainian milk processing industry. For this purpose, a structural econometric model is estimated on the basis of monthly data on the industry level. Estimation of the market structure models did not produce any evidence suggesting the exercise of market power by the milk-processing industry in the estimation period from January 1996 to December 2003. However, it may be appropriate to conduct similar analyses on a regional level, since the concentration of milk processing plants and the structure of agricultural farms in the regions of Ukraine are quite different.

Keywords: *New empirical industrial organization (NEIO), oligopsony, milk processing industry, Ukraine.*

1 Introduction

The analysis of the market structure and pricing behaviour on the agricultural and food markets should be given more attention by agricultural economists, because recent studies suggest that the agricultural markets are more typically oligopsonistic than competitive (McCorriston, 2002). In the last years there has been a recent increase in the number of New Empirical Industrial Organization (NEIO) studies that concentrate on the measurement of and test for the degree of market power in food processing industry as well as in other industries[1]. Moreover, up to now the NEIO studies are typically conducted for industries in developed market economies and have largely focused on the USA and European food sectors, thus ignoring market conditions of the food industry sectors in the transition countries that are potentially different from the situation in the developed market economies. The structure of the food processing industry and the pricing system for agricultural products undergoes fundamental changes during the transition from the planned to a market economy.

The collapse of the placeSoviet Union and the start of transition towards a market economy brought radical changes to all branches of the Ukrainian economy. As in the other successor states of the Soviet Union, transition in Ukraine was accompanied by a severe economic crisis in the 1990s, which also affected agriculture and the food economy. In the dairy sector, the volume of raw milk delivered fell between 1991 and 1999 by around 80%, which meant that at the end of the 1990s the operating rate of dairies was only just above 10%.

At regional and local levels, enterprises in the dairy sector were also subject to direct and indirect interventions by the authorities. With the aim of increasing the operating rates of dairies in their own regions, some regional authorities imposed temporary restrictions on interregional trade in raw milk. In many cases there is the suspicion that the reason for administrative interventions by regional authorities has been to ensure the survival of dairies as employers and tax payers (Baker and Protchenko, 1999).

Although one might assume that the low operating rate triggered sharp competition between the dairies for raw milk, industry concentration in the dairy sector in some regions suggests, on the other hand, a strong market position for the milk processors with the exertion of market power vis-à-vis the raw milk producers.

In the past some dairies have achieved an oligopsonistic market position—at least on a regional scale—on the raw milk market by the restriction or prevention of competition due to interventions by the authorities. This would support the supposition that dairies have been using market power vis-à-vis the raw milk producers. Moreover, inquiries by the State Antimonopoly Committee in 2002 revealed that in some Ukrainian administrative districts there were price arrangements on the raw milk market between milk-processing enterprises.

Initially, therefore, it is not clear what competitive situation the country's dairies found themselves in vis-à-vis the raw milk suppliers—i.e. the agricultural enterprises and small family farms—in the 1990s and which still may exist today. The objective of this study is to use an empirical market structure model based on the NEIO approach to measure the degree of oligopsony market power in the Ukrainian milk processing industry.

2 Structural model of oligopsony market power

Assume that the milk processing industry has n enterprises ($i = 1, ..., n$) producing a homogeneous product (y) and requiring two factors, raw milk (m) and other nonagricultural inputs

[1] For a list of empirical studies on industries with monopoly power, see Bresnahan (1989: 1051) and Carlton and Perloff (1999: 263). For an overview of structural model estimates for the agricultural and food markets, see Sexton and Lavoie (2001), and Wohlgenant (2001).

(z). The production function of the ith milk processing enterprise is

$$y_i = f_i(m_i, z_i) \tag{1}$$

where y_i is quantity of the milk product produced by the ith milk processing enterprise, m_i is the quantity of raw milk bought by the ith milk processing enterprise, and z_i is the quantity of nonagricultural inputs used by that enterprise.

It is assumed that milk producers are price takers in their output market. The inverse supply function for raw milk faced by the ith milk processing enterprise can be represented by

$$W_M = g(M, \mathbf{S}) \tag{2}$$

where W_M is the price of raw milk, M is the total of raw milk purchased by all enterprises of the milk processing industry such that $M = \sum_i m_i$, and S is a vector of supply shifters.

Let the objective of milk processors be to maximize their profit. Given the production function (1) and the supply function of raw milk (2), the profit equation for the ith milk processing enterprise may be defined as

$$\pi_i = P\, f_i(m_i, z_i) \; - \; W_M\, g\,(M, \mathbf{S}) \; - \; W_Z\, z_i, \tag{3}$$

where π_i is profit earned by the ith milk processing enterprise, P is the output price of the milk processing industry, W_M and W_Z are prices of raw milk and other factors, respectively.

If the market for nonagricultural inputs is perfectly competitive, then the first order condition for profit maximization with respect to raw milk that allows for imperfect competition in this market is:

$$W_M + s_i \frac{(1+\lambda_i)}{\varepsilon_{WM}} W_M = \frac{\partial f_i\,(m_i, z_i)}{\partial m_i} P \tag{4}$$

where $\lambda_i = \partial \sum_{j \neq i} m_j \Big/ \partial m_i$ is the conjectural variation of the ith milk processing enterprise, $\varepsilon_{WM} = (\partial M / \partial W_M)\, (W_M / M)$ is the price elasticity of supply of raw milk, $s_i = m_i / M$ is the market share of the ith milk processing enterprise.

Multiplying the left and right hand side of equation (4) by s_i and summing over milk processing enterprises yields the share weighted industry expression, where M and Z correspondent to m_i and z_i on the industry level[2].

$$W_M \cdot \left(1 + \frac{\Theta}{\varepsilon_{WM}}\right) = \frac{\partial f\,(M, Z)}{\partial M} P \tag{5}$$

where Θ is a parameter indexing the degree of oligopsony power. If $\Theta = 0$, then the market for raw milk is perfectly competitive. If $\Theta = 1$, then the market for raw milk is monopsonistic or

[2] We assume that the aggregation of value marginal products over milk processing enterprises also yields consistent aggregates M and Z . The implications for functional forms representing the production technology are not pursued further.

the dairies act like a monopsony (cartel) and the marginal factor cost and the value marginal product should be equated. Intermediate values of Θ imply the presence of an oligopsonistic market structure, in which case the interpretation of the first-order condition is that the 'perceived' marginal factor cost equals the aggregate value marginal product of raw milk.

However, Θ does not explicitly account for market concentration as represented by the Herfindahl-Hirschman coefficient (H). But equation (5) as an aggregate over all enterprises of the milk processing industry can be re-written as

$$W_M \cdot \left(1 + \frac{H\left[1+\lambda\right]}{\varepsilon_{WM}}\right) = \frac{\partial f\left(M,Z\right)}{\partial M} P, \tag{6}$$

Where λ is the industry weighted average of the conjectural variations, $\lambda = \sum_i s_i^2 \lambda_i \big/ \sum_i s_i^2$. The market for raw milk is almost perfectly competitive, if the Herfindahl-Hirschman index is close to zero or the conjectural variation coefficient close to minus one or the price elasticity of supply approaches infinity. If the conjectural variation coefficient approaches zero, then a Cournot oligopsony is obtained. If $\lambda = 0$ and $H = 1$, then the downstream milk processor is a pure monopsonist (Chen and Lent, 1992).

Using industry data over time the degree of oligopsony power Θ as well as the conjectural variation coefficient λ, reflecting the degree of collusion in the milk processing industry, can be estimated.

3 Econometric application of market structure model

According to Bresnahan (1982) and Lau (1982) for identification of market power the inverse supply function (2) (a) must be at least of the second degree in M, (b) must be non-separable and (c) has no constant elasticity with respect to M. This supply function can be approximated by a second order Taylor series[3].

Supposing that the supply function for aggregate supply of raw milk from agricultural enterprises and household plots can be defined as[4]:

$$\ln M = \beta_0 + \sum_j \beta_j \ln W_j + \beta_C \ln C + \delta_T T + 0.5\ \delta_{TT} T^2 + \sum_{jT} \delta_{jT} \ln W_j T + \delta_{CT} \ln C\ T, \tag{7}$$

where W_j ($j = M, D, B, F$) is respectively the price at which milk is supplied (W_M), the direct marketing price for milk that is sold directly to consumers(W_D)[5], the price received for beef cattle (W_B) and the price of mixed feeds (W_F). C is the number of dairy cows as quasi-fixed factors and T is a linear time trend to account for autonomous change (technical change and other unaccounted for factors affecting short-run supply response over time).

[3] Cf. Christensen, Jorgenson and Lau (1973).

[4] Note that equation (7) can be interpreted as a truncated second-order approximation to a general logarithmic farm milk supply function. More general supply functions such as translog would be more appropriate here, but the limited number of monthly observations made it impossible to use such a specification. The supply function in equation (7) satisfies the three conditions for identification of market power.

[5] During the transition the market share of milk sold directly to consumers rapidly increased from 0,1% in 1990 to 21,2% in 2000. This had a significant impact on the milk supply in Ukraine.

Solving equation (7) for W_M and differentiating with respect to M yields the following expression for the marginal effect of the raw milk quantity on milk prices:

$$\frac{\partial g(\cdot)}{\partial M} = \frac{W_M}{(\beta_M + \delta_{MT}T) \ \ M}, \tag{8}$$

where $\beta_M + \delta_{MT}T = \varepsilon_{WM}$ is the own price elasticity of raw milk supply.

We concentrate on the most important factors of the production process in terms of cost components and assume that the milk processing industry uses only four factors, including raw milk (M), labor (L), capital (K) and energy (E) to produce the aggregate output (Y). Using a linear time trend (T) as proxy for technical change in the milk processing industry we assume the production function to be translog[6].

$$\ln Y = \ln \alpha_0 + \sum_{k=1}^{4} \alpha_k \ln X_k + \frac{1}{2} \sum_{k=1}^{4} \sum_{l=1}^{4} \alpha_{kl} \ln X_k \ln X_l + \gamma_T T + \frac{1}{2} \gamma_{TT} T^2 + \sum_{k=1}^{4} \gamma_{kT} \ln X_k T. \tag{9}$$

where $\alpha_{kl} = \alpha_{lk}$ $(k \neq l)$ and $k,\ l = M, L, K, E$. The translog production function is nonhomothetic and imposes no restriction on the technology.

The marginal product is defined as the partial derivative of the translog production function (9) and is given by

$$\frac{\partial Y}{\partial M} = \left(\alpha_M + \sum_{l=1}^{4} \alpha_{Ml} \ln X_l + \gamma_{MT} T \right) \frac{Y}{M}, \tag{10}$$

Using equations (8) and (10), the first-order condition that allows for imperfect competition on the raw milk market is determined as follows

$$W_M = \left(\alpha_M + \sum_{l=1}^{4} \alpha_{Ml} \ln X_l + \gamma_{MT} T \right) \frac{Y}{M} P \bigg/ \left(1 + \frac{\Theta}{\beta_M + \delta_{MT} T} \right). \tag{11}$$

Since $\Theta = H\,[1 + \lambda]$ this expression can be substituted in equation (11) so that the parameter of the conjectural variations λ can be estimated, if the value of H is inserted in the equation. To account for the seasonality in our monthly data eleven monthly dummy variables are added to equations (7) and (11). To allow for the existence of random shocks additive disturbance terms are added to equations (7) and (11) which are assumed to have zero mean, constant variance, and to be independently and normally distributed.

[6] Cf. Christensen, Jorgenson and Lau (1973).

4 Estimation Results and Specification Testing

Since equations (6) and (9) represent a nonlinear simultaneous equation system, they were estimated using nonlinear three-stage least squares (NL3SLS)[7]. The quantity of raw milk (M) and the price of raw milk (W_M) were designated as endogenous. Three alternative specifications of the market structure models were considered (Table 1). The first model represents competitive market conditions. The second model includes the parameter of market power (Θ) in the milk processing industry. Therefore estimation of Θ amounts to estimating a combination of the parameter of conjectural variation (λ) and H. On the other hand, using data for H[8], it is possible to estimate the parameter of conjectural variation (λ). This is the purpose of the third model.

All exogenous variables in the system were used as instruments. The estimations were carried out using the Shazam Econometrics Software (Shazam, 2004). The statistics and parameter estimates for market structure models are reported in Table 1. The asymptotic t-ratios indicate that most parameters (with the exception of some parameters of seasonal dummy variables) are significant at least at the 5 percent significance level.

For the milk supply equation, 16 of the 24 parameters yield t-statistics[9] indicating statistical significance at the 1% level or less. Moreover, the parameters of the supply functions are very robust to change in the model specification. In fact, most parameters are almost identical and change only to the second decimal. The parameters β_M and δ_{MT} are highly significant at any reasonable level of significance. The time trend enters interactively with supply-side exogenous variables, so that the supply curve rotates in successive time periods. A Wald test of the joint hypothesis that the coefficients of the five time trend interaction terms were collectively zero is rejected with a Wald χ^2 statistic of 286,36 at the 1 percent significance level.

The estimates of the parameters of the equation for the first order condition that measure the degree of the market power and the parameter of conjectural variation in the milk processing industry and their implications are of primary interest. The results of the estimated parameters of market power are generally compatible with economic theory. The market structure model allowing for market power establishes that the estimate of Θ is close to zero and insignificant. While the negative value of Θ is not theoretically possible, it ranges from -0,0133 to 0,0013 in the 95% confidence interval. With a Wald χ^2 statistic of 1,89, the hypothesis that the milk processing industry is a price-taker in the raw milk market is not rejected even at the 10 percent level. The hypothesis of monopsonistic behavior ($H_0 : \Theta = 1$) is also tested and rejected at the 1 percent level.

With regard to the market structure at the national level, this finding is probably plausible, as the degree of concentration of milk-processing enterprises at the national level is low[10].

[7] For nonlinear three-stage least squares see Amemiya (1977).

[8] Monthly data on concentration in the milk processing industry are unavailable. Therefore, in the regression equations an estimate for H was used which is based on information given by Bojarunets (2002) for the most recent years of our estimation period. According to Bojarunets, in 2002 the largest milk processing enterprise had a market share of 12 %. In 2001 the largest four enterprises had a share of 28 %, whereas the largest ten controlled about 50 % of the market. The rest is shared by about 350-400 enterprises. Based on this information we used a constant average value of H of 0.03 as an estimate. Using a constant is consistent with the observation that the market structure did not change considerable during the estimation period.

[9] All test statistics and standard errors reported in this article are asymptotic.

[10] Cf. Bojarunets (2002).

Parameter	Competition			Market power			Conjectural variation		
	Coeff.	Stand. Error	t-Ratio	Coeff.	Stand. Error	t-Ratio	Coeff.	Stand. Error	t-Ratio
	Milk supply								
β_M	-1,25	0,32	-3,97	-1,27	0,21	-6,01	-1,24	0,21	-5,86
β_D	-0,21	0,19	-1,12	-0,18	0,17	-1,04	-0,24	0,17	-1,38
β_B	-0,14	0,40	-0,36	-0,15	0,40	-0,38	-0,17	0,40	-0,42
β_F	-1,56	0,20	-7,90	-1,59	0,17	-9,62	-1,53	0,17	-9,23
β_C	4,34	1,43	3,03	4,42	1,14	3,87	4,34	1,13	3,83
δ_T	0,48	0,24	2,01	0,48	0,20	2,42	0,50	0,20	2,51
δ_{TT}	0,00	0,00	-5,37	0,00	0,00	-7,26	0,00	0,00	-7,21
δ_{MT}	0,03	0,01	5,65	0,03	0,00	6,01	0,03	0,00	5,86
δ_{DT}	0,00	0,00	0,30	0,00	0,00	0,27	0,00	0,00	0,37
δ_{BT}	0,01	0,01	1,25	0,01	0,01	1,27	0,01	0,01	1,38
δ_{FT}	0,02	0,01	4,80	0,03	0,00	6,26	0,02	0,00	5,90
δ_{CT}	-0,08	0,03	-3,15	-0,08	0,02	-3,97	-0,08	0,02	-4,04
ς_2	0,03	0,04	0,76	0,03	0,04	0,79	0,03	0,04	0,79
ς_3	0,38	0,04	10,29	0,38	0,04	10,46	0,38	0,04	10,39
ς_4	0,57	0,04	14,75	0,57	0,04	14,83	0,57	0,04	14,68
ς_5	1,09	0,05	23,60	1,10	0,05	24,17	1,09	0,05	23,92
ς_6	1,30	0,06	23,28	1,31	0,05	24,21	1,29	0,05	23,93
ς_7	1,21	0,06	20,87	1,22	0,06	21,90	1,20	0,06	21,63
ς_8	1,09	0,06	19,03	1,10	0,06	19,86	1,09	0,06	19,59
ς_9	0,91	0,05	17,38	0,92	0,05	17,67	0,90	0,05	17,39
ς_{10}	0,70	0,04	15,83	0,70	0,04	15,98	0,69	0,04	15,70
ς_{11}	0,29	0,04	7,26	0,30	0,04	7,46	0,29	0,04	7,24
ς_{12}	0,23	0,05	4,39	0,24	0,04	5,40	0,24	0,04	5,40
β_0	-11,11	12,30	-0,90	-11,68	10,33	-1,13	-11,03	10,26	-1,08

Para-meter	Competition			Market power			Conjectural variation		
	Coeff.	Stand. Error	t-Ratio	Coeff.	Stand. Error	t-Ratio	Coeff.	Stand. Error	t-Ratio
	First order condition (FOC)								
Θ	-	-	-	-0,01	0,00	-1,37	-	-	-
λ	-	-	-	-	-	-	-1,17	0,11	-10,30
α_M	32,13	6,27	5,13	35,75	8,49	4,21	35,50	6,75	5,26
α_{MM}	1,56	0,15	10,38	1,42	0,21	6,93	1,43	0,16	8,74
α_{ML}	-2,83	0,64	-4,43	-3,25	0,87	-3,75	-3,21	0,69	-4,65
α_{MK}	-0,58	0,21	-2,71	-0,35	0,30	-1,17	-0,37	0,24	-1,55
α_{ME}	-0,94	0,12	-7,58	-0,86	0,16	-5,25	-0,87	0,13	-6,69
γ_{MT}	-0,01	0,00	-4,16	-0,01	0,00	-2,40	-0,01	0,00	-3,02
ξ_2	21,69	16,98	1,28	26,43	21,48	1,23	25,54	17,14	1,49
ξ_3	14,64	18,42	0,80	25,07	23,66	1,06	24,73	18,87	1,31
ξ_4	18,27	19,28	0,95	32,39	25,24	1,28	31,98	20,12	1,59
ξ_5	-52,14	24,34	-2,14	-33,48	31,71	-1,06	-33,00	25,29	-1,30
ξ_6	-60,88	26,01	-2,34	-38,59	34,23	-1,13	-38,18	27,30	-1,40
ξ_7	-47,12	25,03	-1,88	-22,54	33,48	-0,67	-22,53	26,70	-0,84
ξ_8	-47,84	24,49	-1,95	-2,47	39,40	-0,06	-3,67	31,45	-0,12
ξ_9	-31,03	22,99	-1,35	-31,01	32,19	-0,96	-28,42	25,44	-1,12
ξ_{10}	-2,48	20,45	-0,12	5,40	26,27	0,21	6,26	20,93	0,30
ξ_{11}	40,46	17,11	2,37	46,47	21,68	2,14	46,38	17,29	2,68
ξ_{12}	108,10	17,46	6,19	116,27	22,34	5,20	115,59	17,81	6,49

Table 1: Results of NL3SLS Estimation of Market Structure Models

However, the information about the degree of market concentration is not reflected in this model. Using this information the model has been modified in order to estimate the conjectural variations parameter λ. The estimation results of the third market structure model show that λ amounts to about minus one, which can be interpreted as evidence for perfect competition in this sector over the estimation period from January 1996 to December 2003. While the estimate of Θ for the market power model is not statistically significant, the estimate of λ it is highly significant at any reasonable level of significance. In view of the low operating rate and concentration of the milk processing industry the estimation results of both parameters are plausible.

To compare the estimated market structure models Table 2 lists some coefficients of statistical inference of NL3SLS estimation.

	Competition		Market Power		Conjectural variation	
	Supply	FOC	Supply	FOC	Supply	FOC
R2	0,9811	0,9748	0,9811	0,9617	0,9810	0,9640
DW	1,0658	1,0897	1,0543	1,5729	1,0745	1,5266
Objective Value	1,6588		1,3010		1,6025	

Table 2: Statistical Inference of NL3SLS Estimation

The fit of each system is quite good. In the case of the raw milk supply equation, the R-square between observed and predicted values is 0,98, while that for the first-order condition is 0,97 and 0,96, respectively. The values of the Durbin-Watson statistic lie in the inconclusive range, where it is not possible to make a decision about the hypothesis of the existence of autocorrelation. It is also clear that they are close to the lower bound where the hypothesis of the existence of autocorrelation cannot be rejected. Further, the minimized values of the objective function in the NL3SLS estimation are used to compare the market structure models. According to this criterion the market structure model is preferred.

Since not all individual parameters of the market structure models can be given a ready economic interpretation, we calculated the elasticities and the rate of autonomous change $\Delta = \partial \ln M/\partial T$ for the raw milk supply equation and the production elasticity of raw milk, which can be inferred from the first order condition. These are evaluated at the sample mean and presented in Table 3.

The own and cross price elasticities of raw milk supply are less than one in absolute terms, they have the expected signs and are compatible with economic theory. The estimated own price elasticity of raw milk supply (ε_{WM}) is 0,13. While for the competition model it is not statistically significant, it is highly significant at any reasonable level of significance for the market power model. The sign structure of the cross-elasticities of raw milk supply is of considerable interest. From the perspective of the milk producers, the raw milk delivered to the milk processing industry is a substitute for the raw milk that is sold directly to consumers and a complement for beef cattle. The price elasticity of mixed feeds is negative and statistically significant at least at the 1 percent level.

	Competition			Market Power			Conjectural variation		
	Coeff.	Stand. Error	t-Ratio	Coeff.	Stand. Error	t-Ratio	Coeff.	Stand. Error	t-Ratio
ε_{WM}	0,13	0,16	0,82	0,13	0,02	5,88	0,12	0,02	5,78
ε_{WD}	-0,15	0,13	-1,15	-0,13	0,12	-1,10	-0,17	0,12	-1,45
ε_{WB}	0,21	0,15	1,41	0,20	0,15	1,33	0,21	0,15	1,41
ε_{WF}	-0,36	0,14	-2,68	-0,35	0,13	-2,76	-0,36	0,13	-2,81
ε_{MC}	0,34	0,59	0,58	0,41	0,58	0,70	0,27	0,58	0,46
Δ	0,01	0,00	3,35	0,01	0,00	4,32	0,01	0,00	4,25
ε_{YM}	1,99	0,08	24,87	1,92	0,11	18,10	1,92	0,08	22,71

Table 3: Elasticities and Rate of Autonomous Change

A well-behaved supply function must be homogeneous of degree zero in prices. A Wald test of the hypothesis that the own and cross price elasticities of raw milk supply evaluated at the sample mean add up to zero is not rejected with a Waldχ^2 statistic of 1,19 even at the 25 percent level.

The variable number of milking cows (C) has a positive but statistically insignificant impact on the raw milk supply. The rate of autonomous change in the raw milk supply evaluated at the sample mean is significant at any reasonable level of significance and amounts to 18,3 percent annually. This number is extraordinarily high and must be considered an overestimation, even if one assumes that a large fraction of this number is due to unaccounted for variables besides technical change[11].

Furthermore, the estimated production elasticity ε_{YM} is unexpectedly large. This result may be due, in part, to the fact that only some of the parameters of the complete translog production function have been estimated. This suggests that in this particular case, the translog production function (9) should be added as an additional equation to the market structure model.

5 Summary and Conclusions

The objective of this paper has been to measure the degree of oligopsony power and conjectural variations for the Ukrainian milk processing industry. For this purpose, three structural econometric models are estimated on the basis of monthly data on the industry level. The estimation of the market structure models did not produce any evidence suggesting the exercise of market power by the milk-processing industry in the estimation period from January 1996 to December 2003. However, it may be appropriate to conduct similar analyses on a regional level, since the concentration of milk processing enterprises and the structure of agricultural farms in the

[11] The observed average annual rate of change in milk supply over the sample period is 6,8 %. The change in exogenous variables explains an annual rate of change in milk supply of –3,9 %, which supports an autonomous rate of change of only 10,7 %.

regions of Ukraine are quite different. While our estimate of the Herfindahl-Hirschman index in the Ukrainian milk processing industry suggests that concentration is low at the national level, on the regional level there is evidence for higher concentration. Additional data for the plant level show that in 8 out of 25 regions the Herfindahl-Hirschman coefficient is larger than 0,200. Hence, it would be desirable to apply the structural econometric model also to regional data and to measure the market power on a regional market level. The authors hope that this can be achieved in further analyses.

References

[1] Amemiya, T. (1977). 'The Maximum Likelihood Estimator and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model.' *Econometrica* 45: 955-968.

[2] Bojarunets, A. (2002). 'Market with a high fat content.' *Companion* 13. [In Russian].

[3] Bresnahan, T. F. (1982). The oligopoly solution concept is identified. *Economics Letters* 10: 87-92.

[4] Bresnahan, T. F. (1989). Empirical Studies of Industries with Market Power. In: R. Schmalensee and R. D. Willig (ed.): *Handbook of Industrial Organization* 2, 1011-1057. Amsterdam: Elsevier Science B.V.

[5] Carlton, D. W. and J. M. Perloff (1999). *Modern Industrial Organization*. 3rd ed. Reading (MA): Addison-Wesley.

[6] Chen, Z., and R. Lent (1992). 'Supply Analysis in an Oligopsony Model.' *American Journal of Agricultural Economics* 74: 971-979.

[7] Christensen, L. R., D.W. Jorgenson and L. J. Lau (1973). 'Transcendental Logarithmic Productions Frontiers.' *Review of Economics and Statistics* 55: 28-45.

[8] Lau, L. J. (1982). 'On identifying the degree of competitiveness from industry price and output data.' *Economics Letters* 10: 93-99.

[9] SAS (1985): *SAS Institute*, Version 5, Cary, New Carolina.

[10] Sexton, R. J. and N. Lavoie (2001). 'Food Processing and Distribution: an Industrial Organization Approach.' In: B. Gardner and G. Rausser (ed.): *Handbook of Agricultural Economics* 1, 863-932. Amsterdam: Elsevier Science B.V.

[11] Shazam (2004): *Shazam Econometrics Software*. User's Reference Manual. Version 10. Northwest Econometrics, Ltd. Vancouver, B.C.

[12] State Committee of Statistics of Ukraine: *Statistical Yearbook of Ukraine*. Various issues. [In Ukrainian].

[13] State Committee of Statistics of Ukraine: *Industrial products of Ukraine*. Various issues. [In Ukrainian].

[14] State Committee of Statistics of Ukraine: *Industrial producer price indices*. Various issues. [In Ukrainian].

[15] State Committee of Statistics of Ukraine: *Various published and unpublished statistical reports*. [In Ukrainian].

[16] Wohlgenant M. K. (2001): Marketing Margins: Empirical Analysis. In: B. Gardner and G. Rausser (ed.): *Handbook of Agricultural Economics* 1, 933-970. Amsterdam etc.: Elsevier Science B.V.

Radomír Perzina, Jaroslav Ramík

School of Business Administration in Karviná

Silesian University in Opava

University square 1934/3, Opava

e-mail: perzina,@opf.slu.cz, ramik@opf.slu.cz

Multicriteria Decision Problems in Fuzzy Environment

Abstract

The main goal of every economic agent is to make a good decision, especially in economic environment with many investment alternatives and evaluation criteria. In this paper a new decision model based on analytic network process (ANP) is proposed. Comparing to classical approach it is modified for solving the decision making problem with fuzzy pair-wise comparisons and a feedback between the criteria. The evaluation of the weights of criteria, the variants as well as the feedback between the criteria is based on the data given in pair-wise comparison matrices with fuzzy elements. Extended arithmetic operations with fuzzy numbers are proposed as well as ordering fuzzy relations to compare fuzzy outcomes. An illustrating numerical example is presented to clarify the methodology. The solution is compared with the same problem solved by the classical analytic hierarchy process (AHP), i.e. with non-fuzzy evaluations in the pair-wise comparisons and without the feedback. All calculations are performed in Microsoft Excel add-in software named FVK that was developed for solving the proposed model. Comparing to other software products for solving multicriteria problems, FVK is free, able to work with fuzzy data, allows easy manipulation with data and utilizes capabilities of widespread spreadsheet Microsoft Excel.

Keywords: *analytic hierarchy process, analytic network process, multi-criteria decision making, pair-wise comparisons, feedback, fuzzy*

1 Introduction

When applying Analytic Hierarchy Process (AHP) in decision making one usually meets two difficulties: when evaluating pair-wise comparisons on the nine point scale we do not incorporate uncertainty or when decision criteria are not independent as they should be. In this paper these difficulties are solved by a proposal of the new method which incorporates uncertainty using pair-wise comparisons by triangular fuzzy numbers, and takes into account interdependences between criteria.

The first difficulty is solved by fuzzy evaluations: instead of saying e.g. "with respect to criterion C element A is 2 times more preferable to element B" we say "element A is possibly 2 times more preferable to element B", where "possibly 2" is expressed by a triangular fuzzy number. In some real decision situations, dependency of the decision criteria occur quite frequently, e.g. the criterion price is naturally influenced by the quality criterion. Here, the dependency is modeled by a feedback matrix, which expresses the grades of influence of the individual criteria on the other criteria.

The interface between hierarchies, multiple objectives and fuzzy sets have been investigated by the author of AHP T.L. Saaty [6]. Later on, Laarhoven and Pedrycz [9] extended AHP to fuzzy pairwise comparisons. Saaty extended AHP to a more general process with feedback called Analytic Network Process (ANP) [7], [8]. In this paper we extend the approaches from [1], [2], [8] to the case of feedbacks between the decision criteria as it was specified in [5], moreover we also supply an illustrating realistic example to demonstrate the proposed method, documented by the outputs from Microsoft Excel add-in FVK that was developed for solving the proposed model.

2 Multi-criteria decisions

In Analytic hierarchy process (AHP) we consider a three-level hierarchical decision system: On the first level we consider a decision goal G, on the second level, we have n independent *evaluation criteria*: $C_1, C_2, \ldots, C_n$, such that $\sum_{i=1}^{n} w(C_i) = 1$, where $w(C_i) > 0,\ i = 1, 2, \ldots, n,\ w(C_i)$ is a positive real number—weight, or, relative importance of criterion C_i subject to the goal G. On the third level m variants (alternatives) of the decision outcomes $V_1, V_2, \ldots, V_m$ are considered such that again $\sum_{r=1}^{m} w(V_r, C_i) = 1$, where $w(V_r, C_i)$ is a non-negative real number—an evaluation (weight) of V_r subject to the criterion $C_i,\ i = 1, 2, \ldots, n$. This system is characterized by the *supermatrix* $\mathbf{W}$, see [8]:

$$\mathbf{W} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{W_{21}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{W_{32}} & \mathbf{I} \end{bmatrix}, \tag{1}$$

where $\mathbf{W}_{21}$ is the $n \times 1$ matrix (weighing vector of the criteria), i.e.

$$\mathbf{W}_{21} = \begin{bmatrix} w(C_1) \\ \vdots \\ w(C_n) \end{bmatrix}, \tag{2}$$

and $\mathbf{W}_{32}$ is the $m \times n$ matrix:

$$\mathbf{W}_{32} = \begin{bmatrix} w(C_1, V_1) & \cdots & w(C_n, V_1) \\ \vdots & \ddots & \vdots \\ w(C_1, V_m) & \cdots & w(C_n, V_m) \end{bmatrix}. \tag{3}$$

The columns of this matrix are evaluations of variants by the criteria, $\mathbf{I}$ is the unit matrix. $\mathbf{W}$ is a *column-stochastic matrix*, i.e. the sums of columns are equal to one. Then the limit matrix $\mathbf{W}^{\infty}$ (see [3]) exists and we can calculate the resulting *priority vector of weights of* *the variants* $\mathbf{Z}$ which is given by formula (4). The variants can be ordered according to these priorities.

$$\mathbf{Z} = \mathbf{W}_{32}\mathbf{W}_{21} \tag{4}$$

In real decision systems with 3 levels there exist typical interdependences among individual elements of the decision hierarchy e.g. criteria or variants. Consider now the dependences among the criteria. This system is then given by the supermatrix $\mathbf{W}$:

$$\mathbf{W} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{W_{21}} & \mathbf{W_{22}} & \mathbf{0} \\ \mathbf{0} & \mathbf{W_{32}} & \mathbf{I} \end{bmatrix}, \tag{5}$$

where the interdependences of the criteria are characterized by $n \times n$ matrix $\mathbf{W}_{22}$:

$$\mathbf{W}_{22} = \begin{bmatrix} w(C_1, C_1) & \cdots & w(C_n, C_1) \\ \vdots & \ddots & \vdots \\ w(C_1, C_n) & \cdots & w(C_n, C_n) \end{bmatrix}.$$

In general, matrix (5) is not column-stochastic, hence the limiting matrix does not exist. Stochasticity of this matrix can be saved by additional normalization. Then there exists a limiting matrix $\mathbf{W}^{\infty}$ and the vector of weights $\mathbf{Z}$ can be calculated by formula (6), see [5].

$$\mathbf{Z} = \mathbf{W}_{32}\left(\mathbf{I} - \mathbf{W}_{22}\right)^{-1}\mathbf{W}_{21} \tag{6}$$

As the matrix $\mathbf{W}_{22}$ is close to the zero matrix, as the dependences among the criteria are usually weak, it can be approximately substituted by the first 4 terms of Taylor's expansion and we get:

$$\mathbf{Z} = \mathbf{W}_{32}(\mathbf{I} + \mathbf{W}_{22} + \mathbf{W}_{22}^2 + \mathbf{W}_{22}^3)\mathbf{W}_{21}. \tag{7}$$

3 Fuzzy numbers and fuzzy matrices

In practice it is sometimes more convenient for the decision maker to express his/her evaluation in words of natural language saying e.g. "possibly 3", "approximately 4" or "about 5". Similarly, he/she could use the evaluations as "A is possibly weak preferable to B", etc. It is advantageous to express these evaluations by fuzzy sets of the real numbers, e.g. triangular fuzzy numbers. A *triangular fuzzy number* a is defined by a triple of real numbers, i.e. $a = (a^L; a^M; a^U)$, where a^L is the *Lower number*, a^M is the *Middle number* and a^U is the *Upper number*, $a^L \leq a^M \leq a^U$. If $a^L = a^M = a^U$, then a is the *crisp number* (*non-fuzzy number*). In order to distinguish fuzzy and non-fuzzy numbers we shall denote the fuzzy numbers, vectors and matrices by the tilde, e.g. $\tilde{a} = (a^L; a^M; a^U)$. It is known that the arithmetic operations $+, -, *$ and $/$ can be extended to fuzzy numbers by the Extension principle, see e.g. [2].

If all elements of an $m \times n$ matrix $\mathbf{A}$ are triangular fuzzy numbers then we call $\mathbf{A}$ the *triangular fuzzy matrix* and this matrix is composed of the triples of real numbers. Particularly, if $\tilde{A}$ is a pair-wise comparison matrix, we assume that it is reciprocal and there are units on the diagonal.

4 Decision making process

The proposed decision support method of finding the best variant, or ordering of all the variants can be described by an algorithm that we describe in this section:

4.1 Calculate the triangular fuzzy weights

We have to calculate the triangular fuzzy weights as evaluations of the relative importance of the criteria, evaluations of the feedback of the criteria and evaluations of the variants according to the individual criteria. We assume that there exists a fuzzy vectors of triangular fuzzy weights $\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_n$, $\tilde{w}_i = (w_i^L; w_i^M; w_i^U)$, $i = 1, 2, \ldots, n$, which can be calculated by the following formula (see [5]):

$$\tilde{w}_k = (w_k^L; w_k^M; w_k^U), \; k = 1, 2, \ldots, n, \tag{8}$$

where

$$w_k^S = \frac{\left(\prod_{j=1}^n a_{kj}^S\right)^{1/n}}{\sum_{i=1}^n \left(\prod_{j=1}^n a_{ij}^M\right)^{1/n}}, S \in \{L, M, U\}. \tag{9}$$

In [2], the method of calculating triangular fuzzy weights by (9) from the triangular fuzzy pair-wise comparison matrix is called the *logarithmic least squares method*. This method can be used both for calculating the triangular fuzzy weights as relative importance of the individual criteria and also for eliciting relative triangular fuzzy values of the criteria for the individual variants out of the pair-wise comparison matrices and also for calculating feedback impacts of some criteria on the other criteria.

4.2 Calculate the aggregating triangular fuzzy evaluations of the variants

Now we calculate the synthesis - the aggregated triangular fuzzy values of the individual variants by formula (6), eventually, the approximate formula (7), applied for triangular fuzzy matrices:

$$\tilde{\mathbf{Z}} = \tilde{\mathbf{W}}_{\mathbf{32}} \left(\mathbf{I} \tilde{-} \tilde{\mathbf{W}}_{\mathbf{22}}\right)^{\mathbf{-1}} \tilde{\mathbf{W}}_{\mathbf{21}},$$

$$\tilde{\mathbf{Z}} = \tilde{\mathbf{W}}_{\mathbf{32}}(\mathbf{I} \tilde{+} \tilde{\mathbf{W}}_{\mathbf{22}} \tilde{+} \tilde{\mathbf{W}}_{\mathbf{22}}^{\mathbf{2}} \tilde{+} \tilde{\mathbf{W}}_{\mathbf{22}}^{\mathbf{3}}) \tilde{\mathbf{W}}_{\mathbf{21}}.$$

Here, for addition, subtraction and multiplication of triangular fuzzy numbers we use the fuzzy arithmetic operations defined in [5].

4.3 Find the best variant

The simplest method for ranking a set of triangular fuzzy numbers in (7*) is the *center of gravity method*. This method is based on computing the x-th coordinates of the center of gravity of each triangle given by the corresponding membership functions of $\tilde{z}_i$, $i = 1,2,\ldots,\text{n}$. Evidently, it holds

$$x_i^g = \frac{z_i^L + z_i^M + z_i^U}{3}. \tag{10}$$

By (10) the variants can be ordered from the best to the worst. There exist more sophisticated methods for ranking fuzzy numbers, see e.g. [4], for a comprehensive review of comparison methods see [2]. In the SW tool FVK described in the next section, we apply two more methods which are based on α-cuts of the fuzzy variants being ranked.

Particularly, let $\tilde{z}$ be a fuzzy alternative, i.e. fuzzy number,$\alpha \in (0, 1]$ be a preselected aspiration level. Then the *α-cut of* $\tilde{z}$, $[\tilde{z}]_\alpha$, is a set of all elements x with the membership value at least α, i.e. $[\tilde{z}]\alpha = \{x|\mu_{\tilde{z}}(x) \geq \alpha$.

Let $\tilde{z}_1$and $\tilde{z}_2$ be two fuzzy variants,$\alpha \in (0, 1]$. We say that $\tilde{z}_1$ is *R-dominated* by $\tilde{z}_2$ at the level α if

$\sup[\tilde{z}_1]_\alpha \leq \sup[\tilde{z}_2]_\alpha$. Alternatively, we also say that $\tilde{z}_2$ *R-dominates* $\tilde{z}_1$ at the level α.

If $\tilde{z} = (z^L, z^M, z^U)$ is a triangular fuzzy number, then

$\sup[\tilde{z}_1]_\alpha = z_1^U - \alpha(z_1^U - z_1^M),\ \sup[\tilde{z}_2]_\alpha = z_2^U - \alpha(z_2^U - z_2^M),$

as can be easily verified.

We say that $\tilde{z}_1$ is *L-dominated* by $\tilde{z}_2$ at the level α if

$\inf[\tilde{z}_1]_\alpha \leq \inf[\tilde{z}_2]_\alpha$. Alternatively, we also say that $\tilde{z}_2$ *L-dominates* $\tilde{z}_1$ at the level α.

Again, if $\tilde{z} = (z^L, z^M, z^U)$, then

$\inf[\tilde{z}_1]_\alpha = z_1^L + \alpha(z_1^M - z_1^L),\ \inf[\tilde{z}_2]_\alpha = z_2^L + \alpha(z_2^M - z_2^L).$

Applying R and/or L domination we can easily rank all the variants (at the level α).

5 Case Study

Here we analyze a decision making situation buying an "optimal" refrigerator with 3 decision criteria and 3 variants. The goal of this realistic decision situation is to find the best variant from 3 pre-selected ones according to 3 criteria: e.g. price, design and efficiency. First, we apply the proposed fuzzy ANP algorithm, then classical AHP and finally classical ANP. All calculations are performed in Microsoft Excel add-in software named FVK that was developed for solving the proposed model.

5.1 Importance of the criteria

First we express the importance of the criteria that is given by the pair-wise comparison matrix **C**:

C =	Crit 1			Crit 2			Crit 3		
Crit 1	1	1	1	2	4	6	1	2	2
Crit 2	0,2	0,25	0,5	1	1	1	1/4	1/2	1
Crit 3	0,5	0,5	1	1	2	4	1	1	1

By formula (9) we calculate the corresponding triangular fuzzy weights, i.e. the relative fuzzy importance of the individual criteria that are given in matrix $\mathbf{W}_{21}$:

W21 =

0,360	0,571	0,616	- C1 (Price)
0,105	0,143	0,227	- C2 (Design)
0,227	0,286	0,454	- C3 (Efficiency)

5.2 Evaluation of variants

Next step is to make fuzzy evaluations of the variants according to the individual criteria that are given by the following 3 pair-wise comparison matrices **A1**, **A2**, **A3**:

A1 =	Var 1			Var 2			Var 3		
Var 1	1	1	1	1/3	1/2	1	1/3	1/3	1/2
Var 2	1	2	3	1	1	1	1/2	1/2	1
Var 3	2	3	3	1	2	2	1	1	1

A2 =	Var 1			Var 2			Var 3		
Var 1	1	1	1	1/4	1/3	1/2	2	2	4
Var 2	2	3	4	1	1	1	1/4	1/2	1
Var 3	0,25	0,5	0,5	1	2	4	1	1	1

A3 =	Var 1			Var 2			Var 3		
Var 1	1	1	1	4	6	8	3	3	6
Var 2	0,167	0,2	0,25	1	1	1	1/3	1/2	1
Var 3	0,2	0,333	0,333	1	2	3	1	1	1

The corresponding fuzzy matrix $\mathbf{W}_{32}$ of fuzzy weights is calculated by (9) as

W32 =

0,143	0,163	0,236	0,263	0,289	0,417	0,602	0,648	0,817	- Variant 1
0,236	0,297	0,428	0,263	0,379	0,526	0,100	0,122	0,166	- Variant 2
0,374	0,540	0,540	0,209	0,331	0,417	0,154	0,230	0,263	- Variant 3

5.3 Feedback between criteria

In order to evaluate fuzzy feedback between the criteria we apply again pair-wise comparison method, then we obtain the following 3 pair-wise comparison matrices **B1**, **B2**, **B3**:

B1 =	Crit 2			Crit 3		
Crit 2	1	1	1	1	2	3
Crit 3	0,333	0,5	1	1	1	1

B2 =	Crit 1			Crit 3		
Crit 1	1	1	1	1/6	1/3	1/2
Crit 3	2	3	5	1	1	1

B3 =	Crit 1			Crit 2		
Crit 1	1	1	1	2	3	6
Crit 2	0,2	0,333	0,5	1	1	1

By using (9), we obtain the fuzzy feedback matrix $\mathbf{W}_{22}$:

W22 =

0,000	0,000	0,000	0,194	0,250	0,306	0,612	0,750	0,968	- C1 (Price)
0,471	0,667	0,816	0,000	0,000	0,000	0,194	0,250	0,306	- C2 (Design)
0,272	0,333	0,471	0,612	0,750	0,968	0,000	0,000	0,000	- C3 (Efficiency)

Z =	L	M	U
Var 1	0,182	0,346	0,870
Var 2	0,112	0,264	0,649
Var 3	0,149	0,390	0,733

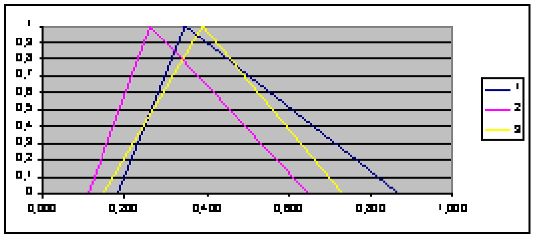


Figure 1: Total evaluation of fuzzy variants

5.4 Total evaluation of the variants

Finally we calculate the synthesis – the aggregated triangular fuzzy values of the individual variants. For this purpose we use the approximate formula (7*), by which we get the matrix of fuzzy weights **Z** for variants. The situation is graphically depicted in Figure 1.

In the last step we rank the evaluations of the above fuzzy variants resulting in the best decision. Here we use ranking methods as described in section 4.3., i.e. Center of gravity, L domination and R domination. For the last two methods level $\alpha = 0.7$ was used. The results are in the following table.

Variant	xgi	Rank		
		Center of gravity	L dominantion	R dominantion
Var 1	0,466	1	2	1
Var 2	0,342	3	3	3
Var 3	0,424	2	1	2

5.5 Classical AHP

Now, we solve the same problem applying classical AHP, i.e. we use non-fuzzy evaluations in the pair-wise comparisons and do not consider the feedback, so the feedback matrix is a zero matrix, i.e. $\mathbf{W}_{22} = \mathbf{0}$. Then we get the matrix of crisp weights **Z** for variants and their rank. The situation is graphically depicted in Figure 2.

5.6 Classical ANP

Finally, we solve the situation again with the same crisp evaluations, however, with a non-zero feedback between the criteria expressed by crisp values (i.e. the classical ANP). Then we get the matrix of crisp weights **Z** for variants and their rank. The situation is graphically depicted in Figure 3.

6 Conclusion

In this paper we have proposed decision model based on ANP for solving the decision making problem with fuzzy pair-wise comparisons and a feedback between the criteria. The evaluation of the weights of criteria, the variants as well as the feedback between the criteria is based

Z =	L	M	U	Rank
Var 1	0,320	0,320	0,320	2
Var 2	0,259	0,259	0,259	3
Var 3	0,421	0,421	0,421	1

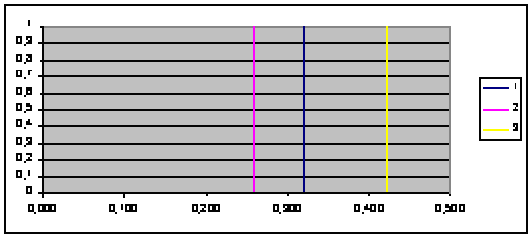


Figure 2: Total evaluation of variants–AHP

Z =	L	M	U	Rank
Var 1	0,346	0,346	0,346	2
Var 2	0,264	0,264	0,264	3
Var 3	0,390	0,390	0,390	1

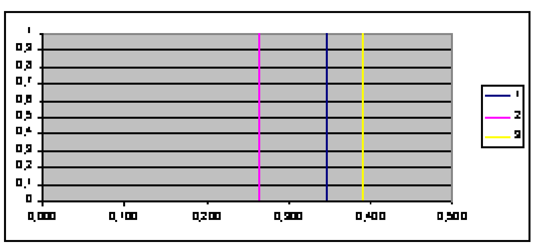


Figure 3: Total evaluation of variants–ANP

on the data given in pair-wise comparison matrices. An illustrating case study has been presented to clarify the methodology. Based on the case study we can conclude that fuzzy evaluation of pair-wise comparisons may be more comfortable and appropriate for decision making. Occurrence of dependences among criteria is more realistic. Dependences among criteria influence the final rank of variants and presence of fuzziness in evaluations change the final rank of variants.

References

[1] Buckley, J.J., Fuzzy hierarchical analysis. *Fuzzy Sets and Systems* 17, 1985, 1, p. 233-247, ISSN 0165-0114.

[2] Chen, S.J., Hwang, C.L. and Hwang, F.P., *Fuzzy multiple attribute decision making.* Lecture Notes in Economics and Math. Syst., Vol. 375, Springer-Verlag, Berlin – Heidelberg 1992, ISBN 3-540-54998-6.

[3] Horn, R. A., Johnson, C. R., *Matrix Analysis*, Cambridge University Press, 1990, ISBN 0521305861.

[4] Ramik, J., Duality in fuzzy linear programming with possibility and necessity relations. *Fuzzy Sets and Systems* 157, 2006, 1, p. 1283-1302, ISSN 0165-0114.

[5] Ramik, J., Perzina, R., Fuzzy ANP – a New Method and Case Study. In *Proceedings of the 24th International Conference Mathematical Methods in Economics 2006*, University of Western Bohemia, 2006, ISBN 80-7043-480-5.

[6] Saaty, T.L., Exploring the interface between hierarchies, multiple objectives and fuzzy sets. *Fuzzy Sets and Systems* 1, 1978, p. 57-68, ISSN 0165-0114.

[7] Saaty, T.L., *Multicriteria decision making - the Analytical Hierarchy Process.* Vol. I., RWS Publications, Pittsburgh, 1991, ISBN.

[8] Saaty, T.L., *Decision Making with Dependence and Feedback – The Analytic Network Process.* RWS Publications, Pittsburgh, 2001, ISBN 0-9620317-9-8.

[9] Van Laarhoven, P.J.M. and Pedrycz, W., A fuzzy extension of Saaty's priority theory. *Fuzzy Sets and Systems* 11, 1983, 4, p. 229-241, ISSN 0165-0114.

This research was partly supported by the grant project of GACR No. 402060431

Pavel Pražák

Faculty of Informatics and Management
University of Hradec Králové
Rokitanského 62, 500 02 Hradec Králové

e-mail: pavel.prazak@uhk.cz

On a Continuous Time Adjustment Cost Model of the Firm

Abstract

The firm can adjust some of its input factors freely in response to changes in demand. The change of other inputs factors requires other costs. By this adjustment cost is meant that in addition to the direct cost of buying new capital there are also cost of installation, cost of reorganizing, cost of retraining etc. The presence of the adjustment cost means that investment decisions can play an active role in the model. The paper is concerned with the model of optimal behavior of the firm facing changes in demand of its output. In the model we consider that the marginal adjustment cost is increasing function of the level of investment and we therefore use a convex adjustment cost function of the level of investment.

1 Introduction

The model considered in this paper is based on models given in Caputo ([2005]) and Hoy et. al ([2001]). Firms are usually exposed to outside changes in demand of its output. It is possible to imagine that the representative firm can react to fluctuation in demand by changing its inputs. At some cases this will involve adjustment cost for some of this input factors - the firm can adjust some of its inputs freely but other factors are quasi-fixed which means that adjustment costs are incurred when they are changed. Assume that there are two inputs to the production process of a firm. The first variable input that can be adjusted freely will be labor L. We will assume that labor can be purchased at the constant unit price of w, $w > 0$. The second input will be capital K. We will assume that capital stocks incurs a constant maintenance cost at the unit price c, $c > 0$. It is assumed that capital is quasi fixed so that the purchase and the installation of it in the rate of investment I involves a cost $gI + C(I)$, where $g > 0$ is the purchase price per unit of capital equipment and $C(I)$ is the adjustment cost. Such adjustment cost are assumed to have the following properties:

- If there is no investment there are no costs, therefore we can consider

$$C(0) = C'(0) = 0.$$

- Both the investment and the disinvestment is expensive. We consider that the adjustment cost $C(I)$ is increasing function of the rate of investment I for $I > 0$ and it is decreasing function for $I < 0$. It means that $C'(I) > 0$ for positive investment $I > 0$ and $C'(I) < 0$ for negative investment $I < 0$. Therefore it is possible to write

$$\text{sign } C'(I) = \text{sign } I. \tag{1}$$

- The costs includes e.g. cost of installing equipment, training workers to use it etc. This all takes time. For example if the firm wants to realize a given installation project in only half the time, then the installation cost are more than doubled - the process of reorganizing the work is more complicated, the risk of mistakes is larger and therefore one have to use better control tools, softwares etc. It seems to be reasonable to consider that it is more costly to increase the stock of usable capital quickly than slowly, therefore we consider

  $$C''(I) > 0 \tag{2}$$

  which means that the marginal adjustment cost $C'(I)$ is increasing function.

Suppose that the technology of the representative firm is

$$Q = F(K, L),$$

where F is a strictly concave neoclassical production function, i.e. $F_K > 0$, $F_L > 0$, $F_{KK} < 0$, $F_{LL} < 0$, $F_{KK}F_{LL} - F_{KL}^2 > 0$ and Inada's condition are valid. Let the unit price of the good produced by the firm is constant and given by $p > 0$.

2 The decision problem of the firm

The instantaneous profit $\Pi(t)$ of the firm at time t that is defined as its revenue $pF(K(t), L(t))$ minus its cost $cK(t) + wL(t) + gI(t) + C(I(t))$ can be written as

$$\Pi(t) = pF(K(t), L(t)) - cK(t) - wL(t) - gI(t) - C(I(t)). \tag{3}$$

If the capital depreciates at a constant rate δ, $\delta > 0$, proportional to the existing stock, the rate of change of the capital stock is the difference of the rate at which new investment goods are purchased, $I(t)$, and the rate at which the given capital depreciates, $\delta K(t)$. We can summarize this at the following ordinary differential equation

$$\dot{K}(t) = I(t) - \delta K(t),\ K(0) = K_0, \tag{4}$$

where K_0, $K_0 > 0$, is the initial stock of capital.

The decision problem of the firm's managers is to choose the time paths of the capital stock, labor input and investment rate throughout the planning horizon $[0, T]$ to maximize the value of the firm

$$V(K(\cdot), L(\cdot), I(\cdot)) = \int_0^T \Pi(t)e^{-rt}dt, \tag{5}$$

where r, $r > 0$, is the market rate of interest that is assumed to be constant.

The problem can be written as the following optimal control problem

$$\max\{V(K, L, I) |\, K \in M_1, L \in M_2, I \in M_3\} \tag{6}$$

where

$$M_1 = \{K \in PC^1([0,T])|\ \dot{K} = I - \delta K, K(0) = K_0 > 0, K \geq 0\} \tag{7}$$

$$M_2 = \{L \in PC([0,T])|\ L \geq 0\} \tag{8}$$

$$M_3 = PC([0,T]) \tag{9}$$

To solve the given optimal control problem we use the Pontryagin maximum principle, see e.g. Joffe and Tichomirov ([1974]). The state variable is the stock of installed capital K, the control variables are the rate of investment I and the level of employment L :

• The current-value Hamiltonian is

$$\mathcal{H}(K, L, I, q) = pF(K, L) - cK - wL - gI - C(I) + q(I - \delta K), \tag{10}$$

where we disregard the time argument t in $K = K(t)$, $L = L(t)$, $I = I(t)$ and where $q = q(t)$ is the current-value adjoint variable associated with (4).
• The necessary condition to maximize $\mathcal{H}$ are

$$\frac{\partial \mathcal{H}}{\partial L}(K, L, I, q) = pF_L(K, L) - w = 0, \tag{11}$$

$$\frac{\partial \mathcal{H}}{\partial I}(K, L, I, q) = -g - C'(I) + q = 0. \tag{12}$$

The concavity of $\mathcal{H}$ in the variable I means that this equation gives a maximum.
• The current-value adjoint variable satisfy the ordinary differential equation

$$\dot{q}(t) - rq(t) = -\frac{\partial \mathcal{H}}{\partial K}(K, L, I, q) = -pF_K(K, L) + c + \delta q(t) \tag{13}$$

• The boundary conditions are $K(0) = K_0$ and $q(T) = 0$.

In the system of ordinary differential equations (4) and (13) we not assumed much specific function therefore we cannot find an explicit solution. Instead, we give a qualitative analysis and use a phase diagram.

Observation 1: Consider the equation (11). This necessary condition means that the marginal product of labour is equal to the real wage. Let us denote

$$\varphi(K, L, w) = pF_L(K, L) - w$$

It is possible to show that for a given $\bar{K} > 0$ and a given $\bar{w} > 0$ there exists a unique $\bar{L} > 0$ such that $\varphi(\bar{K}, \bar{L}, \bar{w}) = 0$. It means that $L = \Phi(K, w)$. With the help of implicit function theorem it is possible to show that

$$L_K(K, w) = -F_{LK}(K, L)/F_{LL}(K, L) > 0,\ L_w(K, L) = 1/F_{LL}(K, L) < 0,$$

which means that the optimal employment level increases with the capital stock and decreases with the wage. Moreover if we consider that F is a homogeneous function of degree one then its first order partial derivative $F_L(K, L)$ is a homogeneous function of degree zero, cf. Hoy et al. ([2001]) page 523. This characteristic enables that the equation (11) can be written

$$F_L(K, L) = F_L(K/L, 1) = w/p.$$

Let us denote $\Psi(\cdot) = F^{-1}(\cdot, 1)$ then it is possible to write

$$L = \frac{K}{\Psi(w/p)}.$$

For given and constant parameters p and w we can put $\Omega = 1/\Psi(w/p)$ and write

$$L = \Omega K.$$

Observation 2: Consider the equation (12), which can be rewritten as

$$q(t) = g + C'(I(t)),\ t \in [0, T]. \tag{14}$$

The right-hand side of this equation present the sum of the price of one unit of capital and the marginal adjustment cost. It means that the right-hand side of (14) is the marginal cost of increasing the capital stock. Since (14) is a necessary condition for optimality, $q(t)$ represents the value to the firm of having one more unit of installed capital at the instant t.

Observation 3: To interpret (13) we need to solve it for optimal $q(t)$, $t \in [0, T]$. Notice that the equation (13) can be written as

$$\dot{q}(t) - (\delta + r)q(t) = -pF_K(K, L) + c,\ q(T) = 0.$$

Multiplying both sides of this equation by the factor $e^{-(r+\delta)t}$ we get

$$(q(t)e^{-(r+\delta)t})^{\cdot} = (-pF_K(K, L) + c)e^{-(r+\delta)t}.$$

If we use the terminal condition $q(T) = 0$ and integrate the last equation between t and T we finally get

$$q(t) = \int_t^T (pF_K(K(s), L(s)) - c)e^{-(r+\delta)(s-t)}ds. \tag{15}$$

The interpretation of $q(t)$ is now that it is the present value of expected future addition to the firm's net cash flow of an additional unit of capital installed at time t.

Observation 4: We show that it is possible to derive an investment function from equation (12). First notice that due the assumption (1) the the following assertions are valid

$$\begin{aligned} I(t) > 0 \quad &\text{for} \quad q(t) > g, \\ I(t) = 0 \quad &\text{for} \quad q(t) = g, \\ I(t) < 0 \quad &\text{for} \quad q(t) < g. \end{aligned} \tag{16}$$

Since in (2) we consider $C''(I) > 0$ then $C'(I)$ is increasing function of the variable I. Hence it is also one to one function and there exists an inverse function $M(\cdot) = (C'(\cdot))^{-1}$. Now it is possible to rewrite equation (12) as

$$I(t) = M(q(t)),$$

where $M'(q(t)) = 1/C''(I(t)) > 0$. It follows that the optimal investment is an increasing function of the adjoint variable q. Since $I(t) = 0$ for $q(t) = g$, we have $M(g) = 0$.

Observation 5: If we derivative equation (12) we get the relation $\dot{q}(t) = C''(I(t))\dot{I}(t)$. It is possible to substitute this relation to equation (13) and get

$$-pF_K(K(t), L(t)) + c + \delta(g + C'(I(t))) = C''(I(t))\dot{I}(t).$$

Since the terminal condition $q(T) = 0$ means that $I(T) = M(0)$, cf. observation4, and we consider (2) we finally gain

$$\dot{I}(t) = \frac{1}{C''(I(t))}[(\delta + r)(g + C'(I)) - pF_K(K(t), L(t)) + c],\ I(T) = M(0). \tag{17}$$

3 On a qualitative solution

Let us consider the system of differential equation (4), (17). To simplify reasoning we use a model of adjustment cost in the form

$$C(I) = \frac{1}{2}I^2$$

and in addition consider that the production function F exhibits *constant return to scale*, cf. observation 1 and notice that $F_K(K, L) = F_K(K/L, 1) = F_K(\Omega, 1)$ is a constant. With this assumption we gain the following system of ordinary differential eqautions

$$\dot{K} = I - \delta K,\ K(0) = K_0, \tag{18}$$

$$\dot{I} = (\delta + r)(g + I) - pF_K(K, L(K)) + c,\ I(T) = -g, \tag{19}$$

where the stock of installed capital K is a state variable and the rate of investment I is a control variable. Let (K°, I°) denote an arbitrary equilibrium of (18),(19). It means that

$$\begin{aligned} I^\circ - \delta K^\circ &= 0, \\ (\delta + r)(g + I^\circ) - pF_K(K^\circ, L(K^\circ)) + c &= 0. \end{aligned}$$

If we linearize the system (18),(19) around (K°, I°) we find its characteristic roots are $\lambda_1 = -\delta < 0$ and $\lambda_2 = \delta + r > 0$. Both roots are real and of opposite signs. Hence (K°, I°) is a saddle point equilibrium.

To draw a phase portrait we begin with drawing isoclines. If $\dot{K} = 0$ in (18) then $I = \delta K$ and if $\dot{I} = 0$ in (19) then $I = (pF_K(\omega, 1) - c)/(\delta + r) - c = I^\circ$. Therefore $K^\circ = I^\circ/\delta$. It is evident that $\dot{K} > 0$ iff $I > \delta K$ and $\dot{K} < 0$ iff $I < \delta K$. It is also evident that $\dot{I} > 0$ iff $I > I^\circ$ and $\dot{I} < 0$ iff $I < I^\circ$. If we summarize all these observations we can draw fig. 1.

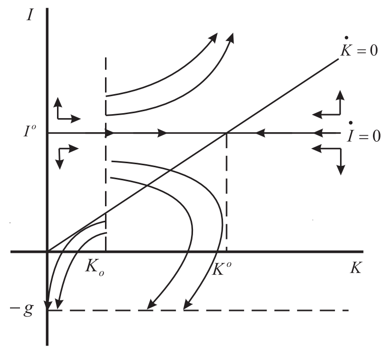


Figure 1: Phase diagram for the investment problem given by the system (18),(19).

4 Conclusion

The main aim of this paper was to present quite a general adjustment cost model and to give interpretations of necessary conditions that we gained from the application of Pontryagin maximum principle. We consider that we will continue to understand models of this type a that we will be able to better specify the possible optimal solution and therefore the optimal investment too. We suppose that there could be a connection with the adjoint variable q used in the present paper and the variable called Tobin's q. It could be also useful to complete the model with a comparative statics.

References

[2005] Caputo R. C.: *Foundations of Dynamic Economic Analysis*, Cambridge University Press 2005

[2001] Hoy M. et al.: *Mathematics for Economics*, MIT Press 2001

[1974] Joffe A. D., Tichomirov V. M.: *Teorija ekstremalnych zadač* (in Russian), Nauka, Moscow 1974

Qu Chenxgin

Department of Operations Management
ESSEC Business School
Avenue Bernard Hirsch, B.P. 105, 95021, Cergy Pontoise, France

e-mail: qu@essec.fr

Loch Christoph

Department of Technology and Operations Management
INSEAD
Boulevard de Constance, 77305, Fontainebleau, France

Kavadias Stylianos

Department of Operations Management
Georgia Tech University
800 west Peachtree Street NW, Atlanta, Georgia 30308-0520, USA

Product Positioning in a Two-dimensional Market Space

Abstract

This paper examines the optimal product portfolio positioning for a monopolist firm in a market where consumers exhibit vertical differentiation for product quality and horizontal differentiation for product feature. In summary, the model shows how portfolio composition decisions depend on the product cost structure and the consumer preferences.

Keywords: *Product Positioning, Portfolio Design, Product Line Design, Development Intensive Products*

1 Introduction

Products may be defined by multiple attributes. We put such attributes into two categories: horizontal differentiation dimension and vertical differentiation dimension. In product design terminology, the horizontal differentiation dimension represents the "feature" choices of product design, whereas the vertical dimension captures the product performance "quality" (also referred to as the main driver of consumer *willingness-to-pay*). For MP3 players, the vertical quality dimension would be memory size (all customers agree that bigger memory size is better), and the horizontal feature dimension would be color, shape, and other "taste" design attributes (customers have different preference on this dimension and do not agree on which color or shape is the best). This paper examines the optimal portfolio for a consumer market with both vertically and horizontally differentiated preferences.

2 Model Setup

In this section, we define the major concepts of our model: the product characteristics; the market space structures; and the firm's objective function.

Product: We group a product's design attributes in two distinct categories: the generic features f and the performance quality q. For example, in a desktop computer, generic features might be the color or design of the tower box, the mouse used, or the arrangement of the keyboard. The features' influence on the costs is negligible; they represent design adjustments that respond to varying customer tastes and do not make the product overall "better" or "worse". The performance quality might be the RAM or ROM memory, or the CPU speed. Quality decisions have an important impact on design and production costs.

The total cost of each product comprises two major components: the fixed investment during design and manufacturing system development, and the per unit variable manufacturing cost. Both types of costs are quadratic functions of quality and independent of the feature. Therefore, designing a product of quality q and producing k units of it costs $C(q, f) = ksq^2 + Sq^2$. Similarly, designing another product with a different feature f but same quality costs another $C(q, f) = ksq^2 + Sq^2$.

The design feature f does not appear in the formula and has no impact on cost, but it is needed in every product. s and S are two unrelated fixed parameters to represent the rate of increase of the marginal design and production cost in quality, respectively.

Among the product attributes, there are attributes that impact variable or fixed cost and attributes that do not. Krishan and Zhu (2006) address the problem of two dimensions both impacting cost. We target specifically the problem of one dimension with impact on cost and one without. This is not to say that feature is costless. Introducing a product with a feature incurs a cost dependant on quality value, regardless of the feature value. It is impossible to introduce a product with feature only.

Market Space: The market space has two dimensions, which reflect consumer preferences about the product attributes. Customers are heterogeneous in their feature tastes and in their valuation of quality. The horizontal dimension captures the ideal feature of the product desired by each consumer. Each consumer has some ideal feature $a \in [0, 1]$ and suffers a quadratic utility loss if the product deviates from the ideal feature. The vertical dimension represents the consumers' valuation of product performance quality, denoted by $b \in [0, 1]$. All consumers agree that more quality is better, but they are heterogeneous in the value they place on additional quality. We can see the feature range from 0 to 1 as color ranges from very light to very dark. This is just a way to denote different choices on this dimension, and it is not a notion of more or less and does not imply any rank ordered preferences. For example, a product with a feature value of 0.7 does not have more features or superior performance than a product with a feature value of 0.5, but only a darker color. In contrast, a product with a quality value of 0.7 does have higher quality than a product with a quality value of 0.5.

In summary, our model of the consumer market space is described by the set $M = \{(a, b) : a \in [0, 1], b \in [0, 1]\}$. Different consumers have different valuations of the product features and qualities, which are distributed according to a known joint probability distribution. We assume the density function $m(a, b)$ over the two-dimensional market space M to be uniform: $m(a, b) = 1, \forall a \in [0, 1],\ \forall b \in [0, 1]$. We later relax this assumption. In Section 7, we examine how a higher concentration of customers in some region of the market space influences the optimal portfolio.

The consumer with preference a in the horizontal dimension and valuation b along the vertical dimension enjoys utility $U(p, q, f; a, b) = bq - e(a - f)^2 - p$ from buying the product (p, q, f). The first part, $b\ q$, captures the ordered nature of consumer preferences with respect to the performance quality; consumer utility increases both in quality and in the valuation of quality.

The second term, $(a-f)^2$, represents the utility loss due to the deviation of the product feature from her/his ideal preference. Changing this term from quadratic to linear does not make the analysis easier and results in a non-smooth demand curve. The scalar e is the marginal utility loss from a unit of deviation. We assume that the marginal utility loss is common across all consumers and products. This assumption, again, allows us to isolate the cost effects on product positioning. Should the customers at the high end become choosier (e increases with b), all demand curves would have sharper slopes; but, this does not qualitatively alter the key results.

Objective function: A monopolistic firm maximizes profit with respect to the size of the product portfolio, and the quality, feature, and price of each product within the portfolio.

3 Optimal Positioning of One Product

We start with the simplest case of one product in the market space, which sets the stage for the more-complex setting of multiple products. Since the distribution of customer valuations is uniform, the product demand equals the market space area where consumers enjoy non-negative utility. Define the curve $L = \{(a, b) \in M\, | U(p, q, f; a, b) = 0\}$as the *indifference curve*, i.e. all consumers who are indifferent to purchasing the product *(p, q, f)*. By solving$U(p, q, f; a, b) = 0$, we have $b = \frac{p+e(a-f)^2}{q}$. The consumers with profiles (a, b) above L, i.e. $b > \frac{p+e(a-f)^2}{q}$, purchase the product.

The following graphs show the potential ways in which the product demand area A fits into the market space M. A is defined by the curve L.

There are three distinct possibilities of how L intersects the side boundaries of M: (i) they do not intersect; (ii) they intersect on both sides; and (iii) they intersect on only one side. Define $MB_1 = \{(a, b) \in M : a = 0\}$, and $MB_2 = \{(a, b) \in M : a = 1\}$. Then, the three cases are described as: $(i) L \bigcap MB_1 = \phi$ and $L \bigcap MB_2 = \phi$, $(ii) L \bigcap MB_1 \neq \phi$ and $L \bigcap MB_2 \neq \phi$ and $(iii) L \bigcap MB_1 \neq \phi$ and $L \bigcap MB_2 = \phi$.

The right-hand picture of Figure 1 cannot be an optimal product positioning: if we shift the grey area A along the horizontal axis, while keeping price and quality constant, the shaded area inside the market space, or demand, increases without further cost increase until we reach Case 1 or Case 2. The increase in demand increases the profit. Hence, Case 3 is always dominated.

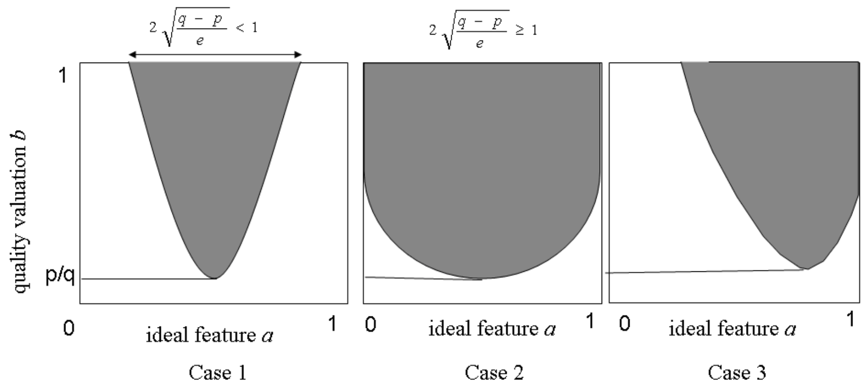


Figure 1: One product in the two-dimensional market

The profit functions in Case 1 and Case 2 are:

$$\pi_1 = (p - sq^2) * \int_{f-\sqrt{\frac{q-p}{e}}}^{f+\sqrt{\frac{q-p}{e}}} (1 - \frac{p + e(f-a)^2}{q})da - Sq^2 = \frac{4}{3}(p - sq^2)\sqrt{\frac{q-p}{e}}\frac{q-p}{q} - S.q^2 \quad (1)$$

$$\pi_2 = (p - sq^2) * \int_0^1 (1 - \frac{p + e(f-a)^2}{q})da - Sq^2 = -\frac{1}{3}(e - 3ef - 3q + 3p + 3ef^2)(p - sq^2)/q - Sq^2 \quad (2)$$

Now we characterize the optimal positioning of the single product.

Lemma 1: In the optimal positioning, $f^* = 1/2$.

By Lemma 1, we can substitute $f^* = 1/2$ into the profit function (1) and then optimize with respect to price and quality. The optimization problem becomes:

$$Max\pi_1(p,q) = \frac{4}{3}(p - sq^2)\sqrt{\frac{q-p}{e}}\frac{q-p}{q} - S.q^2 \text{ if } 2\sqrt{\frac{q-p}{e}} \leq 1 \quad (3)$$

$$Max\pi_2(p,q) = -\frac{(p - sq^2)(e/4 + 3p - 3q)}{3q} - Sq^2 \text{ if } 2\sqrt{\frac{q-p}{e}} \geq 1; \quad (4)$$

Subject to $q > p > 0$

It is straightforward to show that the derivatives of the function to the right and left of the boundary$2\sqrt{\frac{q-p}{e}} = 1$ are the same. Thus, the profit function is differentiable. The feasible regions are defined by the two boundaries $p = 0$ and $q = p$. The profit approaches 0 as the variables approach these boundaries. Therefore, only an interior solution of (3) can provide a positive profit. When we solve the first order conditions of $-\frac{(p-sq^2)(e/4+3p-3q)}{3q} - Sq^2$, we obtain at most four real solutions, regardless of whether the solutions fall into the feasible region ($2\sqrt{\frac{q-p}{e}} \leq 1, q > p > 0$). The first order conditions of $\frac{4}{3}(p - sq^2)\sqrt{\frac{q-p}{e}}\frac{q-p}{q} - S.q^2$ lead to at most six real solutions. Thus, for any given value of the parameters *e, s* and *S*, we can determine the global optimum. However, we are unable to determine the parameter range under which a certain solution is the global optimum, because the 10 solutions have complex functional forms. We can characterize the optimal positioning of the single product in the two special cases, with respect to their cost structure: only when the fixed cost or only the variable cost is present.

When the volume-independent fixed cost is much greater than the production cost, such as for movies and software, we assume that the total cost equals the fixed cost, or *s=0.* We adopt the terminology of *development intensive products,* or DIPs (Krishnan and Zhu 2006).

Lemma 2: For *development intensive products,* the optimal positioning of a single product is:

$p^* = \frac{216}{15625eS^2}$, $q^* = \frac{108}{3125eS^2}$ when $eS \geq 0.288$,

$p^* = \mu_1(e, S)$, $q^* = \nu_1(e, S)$ when $eS \leq 0.288$ (the exact forms of $\mu_1(e, S), \nu_1(e, S)$ are provided upon request. It is too long to be included in the paper)

Another special case is when the volume-independent fixed cost is much smaller than the production cost; we assume that the total cost equals the variable manufacturing cost (Moorthy 1988), a case for trivial design changes. We label such products *production intensive products.*

Lemma 3: For products without fixed cost (S=0), the optimal positioning of a single product is:

$p^* = \frac{15}{64s}$, $q^* = \frac{3}{8s}$ when $es \geq 0.5625$

$p^* = \frac{4-es+4\sqrt{1+es}}{36s}$, $q^* = \frac{1+\sqrt{1+es}}{6s}$ otherwise.

References

[1] Adner, R., D. Levinthal. 2001. Demand Heterogeneity and Technology Evolution: Implications for Product and Process Innovation. *Management Science* 47(5), 611-628.

[2] Cremer, H., J. Thisse. 1991. Location Models of Horizontal Differentiation: a Special Case of Vertical Differentiation Models. *The Journal of Industrial Economics*(39), 383-391.

[3] Economides, N. 1993. Quality Variations in the Circular Model of Variety-differentiated Products. *Regional Science and Urban Economics*(23), 235-257.

[4] Ferreira, R., J. Thisse. 1995. Horizontal and Vertical Differentiation: the Launhardt Model. *International Journal of Industrial Organization*(14), 485-506.

[5] Green, P., A. Krieger. 1985. Models and Heuristics for Product Line. *Marketing Science* 4(1), 1-19.

[6] Hansen, P., B. Jaumard., C. Meyer and J. Thisse. 1998. New Algorithms for Product Positioning. *European Journal of Operational Research*(104), 154-174.

[7] Heeb, R. 2002. Optimal Differentiation. *Econometric Society*. NA-INSEAD conference.

[8] Hotelling, H. 1929. Stability in Competition. *Economic Journal* (39), 41-57.

[9] Krishnan, V., W. Zhu. 2005. Designing a Family of Development-Intensive Products. Forthcoming, *Management Science*.

[10] Lancaster, K. 1998. Chapter 1: "Markets and Product Variety Management", Ho T., and Tang C. (eds.), Product Variety Management: Research Advances, Netherlands, Kluwer Academic Publishers 1-18.

[11] Moorthy, K. 1988. Product and Price Competition in a Duopoly. *Marketing Science* 7(2),

Jiří Rozkovec

Department of Ecomomic Statictics
Faculty of Economics
Technical University of Liberec
Studentská 2, 461 17 Liberec

e-mail: jiri.rozkovec@tul.cz

Mobile Phone Keyboard Layout Optimization

Abstract

The article deals with the mobile phone character set layout optimization. We try to find the layout which on the average minimizes the number of presses of the keys on the phone keyboard within writing a short message service (SMS). MS Access database was used for all the simulations which we programmed and STATGRAPHICS Centurion XV for statistical calculations. The experiment was devided into three phases. At first, we generated random layouts and compared them. We found among them significant differencies from our point of view. At the second part we introduced relative frequencies of all the considered alphabetic characters into the model (they were calculated from the samples) and tried to improve the best found layout. At the third part we added also single-step transition probabilities of the characters to eliminate setting of the character pairs with the highest transition probabilities on the same key. To simplify the problem, the SMS samples used within this article are Czech texts without diacritical signs just like it is usual in SMS communication in present.

1 Introduction

Nearly everybody has used in his life any mobile phone and sent any SMS. Usually, the layout of alphabetic characters starts with ‘a’ which is set on key ‘2’ and ends with ‘z’ on key ‘9’. The question is whether just this layout enables to write messages at minimal number of presses of keys on average. The motivation comes from a comparison with a typewriter or PC keyboards. Their keys are not organised in alphabetical order, we know so called QWERTZ or QWERTY layouts, which are related with characters’ frequencies anyway.

The current simplified mobile phone keys layout is in Table 1.

		Key									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**0**
Position	**1**	.	A	D	G	J	M	P	T	W	“ “
	2	,	B	E	H	K	N	Q	U	X	
	3	’	C	F	I	L	O	R	V	Y	
	4	?						S		Z	
	5	!									

Table 1

Here we could suppose that this layout is not probably the optimal one, because on the first positions we find, for instance, characters like 'G', 'W', which are very rare in Czech language. Whereas, vowels like 'E', 'I', 'O' are on the second or the third positions of their keys. How to optimize that ? Is it possible to reduce on average the number of presses ? Of course, there is a lot of potential combinations how to assign the characters to the keys, so we decided to use a statistical approach based on random simulations of the layouts and their comparison by analysis-of-variance (ANOVA).

2 Random experiments

At first, to get the data we generated 5.000 random layouts [1] and prepared 10 SMS samples. The average length of the message was 146 characters. Then we calculated the number of presses of the keys needed for writing of all the samples by every generated layout. It is clear that the longer the SMS is, the higher number of presses it needs. So, it is necessary to devide every result for every sample by the sample length to get the average number of presses per one character. For every layout we added these 10 averages (we have 10 SMS samples) together and ordered the layouts by this average ascending. We remark here that for the simulations and the data manipulation we used MS Access database where the data are stored in tables, for statistical analyses we used STATGRAPHICS Centurion XV.

From the database table with the ordered layouts we chose following five layouts: the first one (with the smallest number of pressed keys, i.e. the best one), the last one (i.e. the worst one), the middle one and the layouts which correspond with the lower quartile and the upper quartile in the sense of the number of pressed keys – see Table 2. There they are ordered ascending – the layout with the Layout_Id 4854 is the best one, the layout with Layout_Id 4481 is the worst one.

We compared these layouts by One-way ANOVA (F-ratio = 108.42, p-value = 0.00). Thus, we rejected the null hypothesis (5% confidence level) that the layouts have the same average number of pressed keys for all the samples. Furthermore, each two layouts had statistically significant difference [2].

Layout_Id	**The average number of presses per one character for each sample SMS (10 samples)**									
4854	1,60	1,60	1,64	1,62	1,60	1,54	1,56	1,56	1,64	1,55
3379	1,92	1,97	1,90	1,85	1,71	1,82	1,87	1,85	1,87	1,82
3691	1,90	1,78	1,97	1,95	2,01	1,95	1,91	1,93	2,01	1,98
2184	2,02	1,90	2,10	1,97	2,23	1,92	1,86	2,01	2,20	1,95
4481	2,36	2,27	2,37	2,25	2,38	2,25	2,29	2,29	2,23	2,45

Table 2

However, the layouts mentioned above were arranged randomly. We tried to improve the results by additional information. The information is relative frequencies of all the characters within all the samples. It is clear that the most frequent characters should be set on the first positions of the keys to reduce the number of presses. The relative frequencies are in Table 3.

Character	” ”	e	a	o	s	I	n	t	u	r	
Frequency	17,5%	9,4%	7,7%	6,7%	4,7%	4,6%	4,3%	4,2%	4,0%	3,6%	
Character	d	j	m	k	c	v	l	p	h	z	b
Frequency	3,3%	2,8%	2,8%	2,7%	2,6%	2,5%	2,5%	2,5%	2,3%	2,2%	2,1%
Character	.	,	y	f	?	x	g	!	q	w	
Frequency	1,8%	1,4%	1,2%	0,3%	0,2%	0,1%	0,1%	0,1%	0,0%	0,0%	

Table 3

In terms of these values we arranged following layout (Layout_Id = 33333) – Table 4:

		Key									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**0**
Position	**1**	.	E	A	O	S	I	N	T	U	“ “
	2	,	R	D	J	M	K	C	V	L	
	3	’	P	H	Z	B	Y	F	X	G	
	4	?						Q		W	
	5	!									

Table 4

Let us remark here that five vowels occupy the first positions (‘E’, ‘A’, ‘O’, ‘I’, ‘U’). It corresponds with our introductory presumption that the most frequent characters should be set just like this.

Then we evaluated the number of presses of all the samples for this layout 33333 (Table 5) and compared it with the five layouts stated above by One-way ANOVA. Again, there is a significant difference between this layout and the selected five layouts, (F-ratio = 165,32, p-value = 0,00). So, we can say that the incorporated additional information (relative frequencies) improved the result and the layout obtained by this process is better than the other ones.

Layout_Id	The average number of presses per one character for each sample SMS (10 samples)									
4854	1,60	1,60	1,64	1,62	1,60	1,54	1,56	1,56	1,64	1,55
3379	1,92	1,97	1,90	1,85	1,71	1,82	1,87	1,85	1,87	1,82
3691	1,90	1,78	1,97	1,95	2,01	1,95	1,91	1,93	2,01	1,98
2184	2,02	1,90	2,10	1,97	2,23	1,92	1,86	2,01	2,20	1,95
4481	2,36	2,27	2,37	2,25	2,38	2,25	2,29	2,29	2,23	2,45
33333	1,50	1,44	1,50	1,42	1,49	1,44	1,43	1,40	1,55	1,48

Table 5

Using the relative frequencies, the model provided better results definitely. The next modification of the model takes into account transition probabilites [3]. For example, the word ‘eden’ -

it brings complications to our 'layout problem' because there are character pairs in this word from one key ('3') of the typical keyboard (Table 1). They are 'ed' and 'de'. Another such word, for instance, is 'abandom'. The pairs are 'ab', 'ba' (key '2'), 'om' (key '6'). Typing such words we have to wait for the cursor to type 'd' after 'e', 'd' after 'e','o' after 'm' etc. So we should not put together such pairs with a high occurrence like 'ne', 'se', 'na' on the same key. For illustration, the most frequent pairs are in Table 6.

Pair	ne	se	na	ho	te	de	es	pa
Frequency	6,5%	6,2%	5,5%	5,2%	4,8%	4,8%	4,5%	4,5%

Table 6

Therefore we evaluated the model again, but with the penalization of the most frequent pairs – we counted one click more into the number of clicks, if we had to use the same key for typing of two following characters. Again we selected five particular layouts like in Table 2, results are in Table 7.

Layout	**The average number of presses per one character for each sample SMS (10 samples)**									
4854	1,69	1,66	1,71	1,68	1,67	1,56	1,62	1,59	1,73	1,61
3991	1,91	1,86	1,86	1,82	1,99	1,98	1,86	1,78	2,05	1,99
757	1,98	1,91	2,07	1,99	2,02	2,12	1,99	1,93	1,82	2,13
3197	2,02	2,01	2,09	2,02	2,20	2,11	2,07	1,94	2,14	2,16
1711	2,41	2,30	2,43	2,23	2,43	2,51	2,35	2,29	2,25	2,52
33333	1,54	1,48	1,58	1,49	1,55	1,47	1,49	1,44	1,63	1,55

Table 7

Further, we added the layout from Table 4 and according the results of One-way ANOVA there is a significant difference between this layout and the selected five layouts and layout number 33333 is the best one (F-ratio = 139,45, p-value = 0,00).

3 Conclusions

To summarize our experiments, we can say that the results stated above depend obviously on the used samples. We exemplified that various layouts differ significantly as we presupposed. Also there was proved that the relative frequencies of the characters bring better results. The introduction of the transitions probabilities improves the results as well, but not significantly. Based on this, we can say that the 'alphabetical' layout of current mobile phones is not optimal, we found better one.

References

[1] DLOUHÝ, M. *Simulace pro ekonomy.* Oeconomica, Praha, 2005. ISBN 80-245-0973-3

[2] HEBÁK, P. *Vícerozměrné statistické metody 1.* Informatorium, Praha, 2007. ISBN 978-80-7333-056-9

[3] LACHOUT, P., PRÁŠKOVÁ, Z. *Základy náhodných procesů.* Karolinum, Praha, 2005. ISBN 80-7184-688-0

Pavla Řehořová, Miroslav Žižka

Department of Business Administration
Faculty of Economics
Technical University of Liberec
Studentská 2, 461 17 Liberec

e-mail: miroslav.zizka@tul.cz, pavla.rehorova@tul.cz

Factor Analysis of Economically Depressed Areas

Abstract

The aim of this article was to realize the analysis of economically depressed areas (EDAs) causes within Liberec region. The contribution may be divided into three parts. In the first part there is described the methodology of EDAs determination valid for year 2007 and simultaneously there is realized comparison with the previous methodology used in years 2004–2006. In the second part there are expressed in numbers the values of Pearson correlation coefficients for purpose of findings of individual partial indicators interdependence. On basis of this analysis we may submit strength cross-correlation of used indicators. On that account it was approached to simplification of original indicators set into five factors groups, explaining causes of economic backwardness. We talk about these factors: stage carriage service, business activity, education, civil amenities and technical infrastructure. In the third part there were determined factor scores, expressing the rate of individual factors activity within monitored territorial units. On basis of this analysis we may divide the original set of monitored territorial units into several groups, according to prevailing causes of economic backwardness. The factor analysis results enable to aim better the measures of economic policy and thereby to increase the efficiency of public support. Economically depressed areas don't form the homogenous complex from the view of individual factors. The most questionable areas are Českolipsko, Mimoňsko, Semilsko and Frýdlantsko. The factor analysis becomes then the strong subsidiary tool of regional policy.

Keywords: *Region; Factor analysis; Factor score; Stage carriage service; Business activities; Education system; Civic amenities; Technical infrastructure; Economically Depressed Areas (EDAs)*

1 Introduction

One of the economic policy targets is to take notice of administrative units balanced development and to equalize the differences among regions. The regions, whose development is necessary to support with regard to the balanced state development, are mentioned in "Regional Development Strategy of the Czech Republic" [7]. This strategy was elaborated by Ministry for Regional Development and it was ratified by government decree of the Czech Republic No. 560/2006. These regions were defined on basis of four indicators: unemployment rate, tax income, businessmen and purchasing power.

The defined regions include territorial areas of 21 districts and of municipalities with extended powers with the total area equal to 29.4% of the territory of the Czech Republic and with 31.9% of the total population of the Czech Republic [7, p. 31].

Except regions with concentrated state aid (RCSA), the self-governing regions have according to the law [1] possibility to define in their strategic documents other problematic areas with regard to the balanced development of its territorial district. Generally we talk about smaller territorial districts on the level of administrative districts of municipalities that may be hardly affected by the analysis on the whole country rank.

On basis of the analysis published in article [10] it was detected that in those 125 administrative units live 30.9% Czech inhabitants. As for smaller territorial districts there are no accessible indicators used by RCSA valuation (see above), self-governing regions then use different methodologies for socio-economic level measurement of these troubled microregions—so it practically eliminates the interregional comparison. According to the performed analysis of regional strategic documents—development policies—it was detected in sum 30 unique indicators used for EDAs determination (this marking on regional level is most frequent, but by far not the only one). We may claim that self-governing regions most often use these following indicators: unemployment rate (11 regions); population density (9 regions); tax income, i. e. tax yield (6 regions); businessmen (6 regions); share of employment within primary sector (5 regions) and degree of technical and water-economy infrastructure (5 regions). Number of indicators used in region for purpose of EDAs determination varies from one to eleven. Consequently comparison of methodologies is not only difficult, but also we may ourselves logically put a question if the indicators were determined in the right manner and if they really reflect the level of individual regions. Factor analysis gives answers to these questions; it was realized by Liberec region EDAs in this article.

2 Methodology of EDAs determination within Liberec region

Methodology for EDAs determination used by Liberec region came through quite interesting development. In the development policy of Liberec region for years 2004–2006 [11] there were individual municipalities valuated by these six indicators: unemployment rate, tax yield, intensity of business activities, share of economically actives within services, share of population without school-leaving exam and population density. On basis of point evaluation of individual indicators it was subsequently determined the sequence of municipalities. EDAs were logically then merged into six backward areas.

In connection with new programming period the methodology was changed in year 2007 in two main aspects. Number of indicators increased on eleven and instead of municipalities, as a basic territorial district, "generel units" were selected [5]. By this term we think about subregional units involving a municipality with basic facilities (i. e. school, post, health centre, registry office) and its nearest stream territory. Used indicators and their weights are mentioned in table 1. Three indicators have aggregate character. It means that they consist of several partial indicators. Concretely we talk about indicators like stage carriage service, technical infrastructure and civil amenities.

Absolute indicators of individual generel units were calculated to average value of the indicator within Liberec region. From originated coefficients there were calculated weighted arithmetical averages. Resulting values were arranged uplink alphabetically and generel units were divided into three groups—economically mean areas (9); economically depressed areas—EDAs (7) and other (i. e. economically powerful) areas.

Abbreviation	**Indicator**	**Weight**
BYT	Intensity of housing construction	7
VZD	Index of intelligence	8
EZ	Index of economical burden	10
PP	Number of jobs	14
NEZ	Unemployment rate	15
POD	Intensity of business activity	15
DAN	Tax incomes	7
ST	Stage carriage service on Wednesday	8
SO	Stage carriage service on Saturday	5
KAN	Technical infrastructure—canalization PL	1
VOD	Technical infrastructure—water main	1
PL	Technical infrastructure—gas main	1
ZZ	Civil amenities—doctors office	2
SKO	Civil amenities—school existence	5
HUS	Population density	1

Table 1: Indicators used for EDAs determination until year 2007

In economically mean areas and in EDAs there live 169,085 inhabitants of the region (i. e. 39.41% of total number). As compared with the previous methodology, it occurred to enlargement of the area and population that are considered as backwards. Concretely the surface of these regions increased about 670 km^2 and number of inhabitants about 83,886 persons (see article [9]). We may then put a question: Is that true that during three passed years the economic level of Liberec region has become worse so dramatically?

Interesting is also a selection of new evaluative criteria. In study [8, p. 10-12] there was namely referred to the strength cross-correlation of original six indicators. On basis of factor analysis it was found out that from original set of six indicators there were the key ones only three of them—unemployment rate, business activity and intelligence of population. If the Liberec regional authority had made in year 2004 valuation of municipalities only by these three indicators then the result would have been practically the same like in case of all six indicators—from total number of 216 municipalities within Liberec region there were 96 of them determined as EDAs. By using only three indicators there would have replaced categories of EDAs and economically powerful area by 15 municipalities. The Spearmans rank correlation coefficient was 0.7922.

In case of new methodology there paradoxically spectra of applied results spread, though the analysis of original methodology showed the uselessly complicated model. On this account authors of this article returned to this theme and made the factor analysis of newly used indicators.

3 Factor analysis

The factor analysis procedure is designed to extract a number of common factors from a set of many quantitative variables. In many situations, a small number of common factors may be able to represent a large percentage of the variability in the original variables. The ability to express the covariances amongst the variables in terms of a small number of meaningful factors often leads to important insights about the data being analyzed [3]. All computations were calculated by means of statistic packet STATGRAPHICS Centurion XV.

First step of factor analysis is the calculation of correlation among individual indicators. Next step follows—data reduction. The aim is to keep nature and character of original indicators by reduction of their number for purpose of multivariate data analysis simplification [4, p. 101].

For protection of interdependence of individual partial indicators there were in the first phase figured out the values of Pearsons correlation coefficients. The correlation coefficients on significant level $\alpha = 5\%$ are in correlation matrix (table 2) marked in bold type.

Ind.	BYT	DAN	E	HUB	KAN	NEZ	PL	POD	PP	BKO	BO	BT	VOD	VZD	ZZ
BYT	X	0.68	-0.22	-0.06	0.24	-0.26	0.31	0.66	0.33	0.14	0.21	0.03	0.08	0.23	-0.01
DAN	0.68	X	-0.28	0.28	0.61	-0.05	0.48	0.47	0.64	0.00	0.28	0.22	0.16	0.38	0.01
E	-0.22	-0.28	X	-0.17	-0.27	-0.04	-0.32	-0.11	-0.47	-0.14	-0.32	-0.33	-0.22	0.07	-0.16
HUB	-0.06	0.28	-0.17	X	0.37	-0.22	0.61	0.11	0.28	0.09	0.82	0.78	0.14	0.46	-0.06
KAN	0.24	0.61	-0.27	0.37	X	0.08	0.66	-0.03	0.43	-0.04	0.17	0.34	0.15	0.24	-0.15
NEZ	-0.26	-0.05	-0.04	-0.22	0.08	X	-0.09	-0.46	-0.11	0.05	-0.10	-0.12	-0.11	-0.84	-0.13
PL	0.31	0.48	-0.32	0.61	0.66	-0.09	X	0.31	0.42	0.03	0.38	0,47	0.23	0.32	-0.12
POD	0.66	0.47	-0.11	0.11	-0.03	-0.46	0.31	X	0.33	0.08	0.27	0.04	-0.03	0.60	0.26
PP	0.33	0.64	-0.47	0.28	0.43	-0.11	0.42	0.33	X	0.03	0.28	0.28	0.16	0.34	0.15
BKO	0.14	0.00	-0.14	0.09	-0.04	0.05	0.03	0.08	0.03	X	0.62	0.42	0.05	-0.06	0.21
BO	0.21	0.28	-0.32	0.82	0.17	-0.10	0.38	0.27	0.28	0.62	X	0.37	0.01	0.19	0.14
BT	0.03	0.22	-0.33	0.78	0.34	-0.12	0.47	0.04	0.28	0.42	0.37	X	0.12	0.28	0.02
VOD	0.08	0.16	-0.22	0.14	0.15	-0.11	0.23	-0.03	0.16	0.05	0.01	0.12	X	0.23	-0.06
VZD	0.23	0.38	0.07	0.46	0.24	-0.84	0.32	0.60	0.34	-0.06	0.19	0.28	0.23	X	-0.01
ZZ	-0.01	0.01	-0.16	-0.06	-0.15	-0.13	-0.12	0,26	0.15	0.21	0.14	0.02	-0.06	-0.01	X

Table 2: Correlations

The table 2 shows Pearson product moment correlations between each pair of indicators. These correlation coefficients range between -1 and $+1$ and measure the strength of the linear relationship between the indicators.

From the correlation matrix there is evident relatively strength dependence between intensity of housing construction and tax incomes. From this results we may then presume that housing construction supports business activities, thereby number of jobs increases and rate of unemployment decreases. It is reflected in growing tax incomes of municipalities. Farther it is obvious that housing construction is quite logically linked with the technical infrastructure situation (gas connection of houses) and index of intelligence. We may interpret this result also in this way: new flats are built in those municipalities where basic infrastructural and demographic conditions are established for that.

The correlation analysis showed then that tax incomes of municipalities depend not only on housing construction but also on population density, technical infrastructure situation, intensity of business activities, number of jobs and inhabitants education. Negative correlation was founded out between tax incomes and index of economical burden, i. e. share of inhabitants in unproductive age. If we look at correlation links of indicator “economical burden” with other indicators we may then claim that high share of inhabitants in unproductive age (above all pensioners) is localized in municipalities with low number of jobs, bad technical infrastruction situation (absence of canalization and gas main) and bad stage carriage service.

In connectedness with tax incomes it also means that EDAs are characterized by insignificant technical infrastructure, bad stage carriage service, unfavourable demographic situation and low number of jobs.

Quite interesting are correlation links of index of intelligence and other indicators. It was found out that with growing education of population, tax incomes of municipalities increases, rate of unemployment decreases, business activities are getting better and also number of jobs then increases. Simultaneously it is clear that more educated inhabitants are localized in areas with higher population density, and vice versa—less educated inhabitants live in distant areas of Liberec region.

Factor Number	**Eigenvalue**	**Percent of Variance**	**Cumulative Percentage**
1	4.5926	30.617	30.617
2	2.1450	14.300	44.917
3	1.7609	11.739	56.657
4	1.5652	10.434	67.091
5	1.0568	7.045	74.136
6	0.9336	6.224	80.360
7	0.6553	4.369	84.729
8	0.5301	3.534	88.263
9	0.4812	3.208	91.471
10	0.4074	2.716	94.187
11	0.2862	1.908	96.095
12	0.2087	1.392	97.486
13	0.1776	1.184	98.671
14	0.1532	1.021	99.692
15	0.0462	0.308	100.000

Table 3: Eigenvalues of the Factors

Data from correlation matrix provide relatively clear idea of data structure, nevertheless high number of indicators and mutual links among them make more difficult the interpretation of all results.

There are expressed in numbers eigenvectors of source matrix in the second phase. The purpose of the analysis is to obtain a small number of factors which account for most of the variability in the 15 variables. In this case, 5 factors have been extracted, since 5 factors had eigenvalues grater than 1.0. Together they account for 74.14% of the variability in the original data (see the table 3).

It means that we may reduce those 15 original partial indicators to only 5 factors. In the third phase there are determined factor weights, explaining the link among factors and original indicators. They represent the most important information; the interpretation of factors is based on them. It is true that higher factors describe always less and less described variabilities in data.

The table 4 shows the equations which estimate the common factors after rotation has been performed. Rotation is performed in order to simplify the explanation of the factors. The values of the variables are standardized by subtracting their means and dividing by their standard deviations. It also shows the estimated communalities, which can be interpreted as estimating the proportion of the variability in each variable attributable to the extracted factors.

	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**
BYT	-0.0621	**0.7987**	0.1799	0.1452	-0.1199
DAN	0.1555	**0.8340**	0.0889	-0.1390	0.0708
EZ	-0.2308	-0.3991	0.2834	-0.2879	-0.5516
HUS	**0.8170**	0.0609	0.2816	-0.2610	0.1653
KAN	0.3050	0.5174	-0.1545	-0.4847	0.2745
NEZ	-0.0458	-0.0364	-0.8737	-0.1327	-0.0535
PL	0.4607	0.5383	0.0852	-0.3124	0.2238
POD	0.0006	**0.6122**	0.5474	0.3189	-0.1889
PP	0.1678	**0.6757**	0.0733	0.0644	0.3647
SKO	0.4962	0.0225	-0.1717	**0.5615**	-0.1090
SO	**0.8774**	0.2129	0.0364	0.2695	-0.0761
ST	**0.9622**	0.0696	0.0508	0.0211	0.1247
VOD	0.0133	0.0072	0.1731	-0.0392	**0.8131**
VZD	0.2089	0.2652	**0.8236**	-0.1669	0.1301
ZZ	-0.0092	0.0452	0.0912	**0.7429**	0.1223

Table 4: Factor Loading Matrix After Varimax Rotation

The gained factors are groups of indicators recording the similar variability, at the same time the individual factors should be interindependent [6, p. 61].

4 Interpretation of factors

4.1 Factor of stage carriage service

The first factor explains 30.62% of total variability of parameter's set. It is characteristic by high burden of these indicators: ST(0.96), SO(0.88) and HUS(0.82). These indicators characterize stage carriage service during working days and on Saturday, and population density. From the previous analysis it ensued that these indicators are related to each other. Municipalities located in distant areas of the region have at the same time also bad stage carriage service, thereby the country is depopulated and socio-economic level of municipalities is getting worse.

Also factor scores, expressing the rate of individual factors activity in monitored territorial units, were calculated. The highest values of factor scores are achieved above all in Podještědí (Český Dub, Osečná, Sychrov, Příšovice), in Českolipsko (Krompach, Dubá, Žandov) and in a part of Semilsko (Rovensko pod Troskami, Benecko, Vysoké nad Jizerou). It means that there is the worst situation in these generel units. On the contrary, the lowest values were found out above all in bigger cities (Česká Lípa, Jablonec nad Nisou, Liberec) and in Tanvaldsko (Tanvald, Desná, Železný Brod).

4.2 Factor of business activity

The second factor explains 14.30% of total variability of parameter's set. It is filled particularly by the indicators like DAN(0.83), BYT(0.80), PP(0.68) and POD(0.61); these indicators are connected with entrepreneurship and jobs creation. The level of business activity reflects then itself in tax incomes of municipalities.

According to factor scores there is the worst situation above all in Semilsko and Jilemnicko (Horní Branná, Jesenný, Roztoky u Jilemnice, Studenec) and in a part of Podještědí (Jenišovice). On the contrary, the best situation from this view is in Podkrkonoší (Harrachov, Rokytnice nad Jizerou, Benecko) and in surroundings of bigger cities (Liberec, Turnov, Nový Bor).

Factor of education system

The third factor explains 11.74% of total variability of parameter's set. It is characteristic practically by only one indicator—VZD(0.82)—index of intelligence, expressing the average achieved educational level by population older than 15 years of age. The worst situation here is in Českolipsko and Mimoňsko (Dubá, Kravaře, Ralsko) and in Frýdlantsko (Nové Město p. S., Raspenava, Višňová). As quite troublefree is a situation in Jablonecko (Jablonec nad Nisou, Pěnčín, Jenišovice, Zásada), Turnovsko (Malá Skála, Sychrov) and in regional capital Liberec.

Factor of civil amenities

The fourth factor explains 10.43% of total variability of parameter's set. It is saturated above all by these indicators: ZZ (0.74) and SKO(0.56), characterizing the facilities (schools, health centres) of municipalities. We may call it like a factor of civil amenities. According to the factor scores level there is the worst situation surprisingly in big cities (Česká Lípa, Jablonec nad Nisou, Liberec) and in municipalities in Českolipsko (Doksy, Dubá, Ralsko, Zákupy). We may try to explain this relatively bad situation in big cities by high number of inhabitants per one facility. On the contrary, the best situation is in Podještědí (Rynoltice, Osečná, Křižany), Jablonecko (Josefův Důl, Janov nad Nisou) and in Podkrkonoší (Harrachov, Desná, Jablonec nad Jizerou).

Factor of technical infrastructure

The last significant factor explains 7.05% of total variability of parameter's set. It is saturated above all by indicator: technical water-main infrastructure VOD(0.81). Essentially weaker is weight of technical canalization infrastructure KAN(0.27) and technical gas-main infrastructure PL(0.22). The worst situation is in municipalities in Semilsko (Poniklá, Benecko) and in several generel units, creating not continuous geographical unit (Velké Hamry, Raspenava, Křižany, Jablonec nad Jizerou, Dubá). On the contrary, quite good situation is in bigger cities (Česká Lípa, Nový Bor, Stráž pod Ralskem) and in several isolated generel units (Příšovice, Rynoltice, Zahrádky).

5 Conclusion

From the study arose that we may characterize an economical level of Liberec region by means of five basic factors, replacing original large set of indicators. These basic factors are stage carriage service, business activity, education system, civil amenities and technical infrastructure. Simultaneously we may claim that factor analysis of indicators included in the new methodology for EDAs determination is to a great extent in accordance with previous findings published in [8]. Enlargement of used indicators spectra from 6 original to 15 present thus did not bring any fundamental change relating to realized causes of economical backwardness of municipalities within Liberec region. In this regard the new model is uselessly complicated.

On basis of the performed analysis we may at the same time give a recommendation for regional policy of Liberec region. In term of stage carriage service the regional authority should attend to improvement of public transport situation, above all in Podještědí and in parts of Českolipsko and Semilsko. Projects for business activities support should be focused above all on areas in Semilsko, Jilemnicko and Podještědí. With intelligence of population there is the biggest problem in Českolipsko, Mimoňsko and Frýdlantsko. Facilities of civil amenities are insignificant in Českolipsko and in big county cities. Technical infrastructure appears as problematic above all in Semilsko.

At the conclusion we may claim that factor analysis results enable us to aim better the measures of economic policy and thereby to increase the efficiency of public support. Economically depressed areas dont form the homogenous complex from the view of individual factors. In partial aspects some territorial units even show the above-average results, though as a whole they are without question backwardness (it is caused by deep below-average results from the view of rest factors). The generel unit Rynoltice may be one example—it has good facilities of technical infrastructure and civil amenities, but on the other side it has also inconvenient educational structure of population and low intensity of business activities. It is evident that supportive measures concentrated for improvement of the first two factors would have gone astray in this territorial unit. On the contrary, projects focused on education, adoption of ICT [2, p. 414], business and public transport should help in a principle way to eliminate the causes of economic backward in this territorial unit. The factor analysis becomes then the strong subsidiary tool of regional policy.

This article was realized under the state subsidy of the Czech Republic within the

Ministry for Regional Development research and development project "Innovation Approach to Analysis of Disparities on Regional Level" No. WD-30-07-1.

Bibliography

[1] *Act No. 248/2000 Coll. on Promotion of Regional Development*

[2] ANTLOVÁ, K. Strategic Use of ICT in Small Businesses. In *Conference E-Activity and Leading Technologies.* Porto: IASK, 2007, pp. 414 – 418. ISBN 978-972-99397-5-4

[3] *Factor Analysis.* [On-line manuals]. Herndon: StatPoint, 2005 [cit. 2008-02-13]. Available from `www.statgraphics.com`

[4] MELOUN, M., MILITKÝ, J., HILL, M. *Počítačová analýza vícerozměrných dat v příkladech.* 1. vyd. Praha: Academia, 2005. ISBN 80-200-1335-0

[5] *Metodika vymezení hospodářsky slabých oblastí Libereckého kraje.* Liberec: Krajský úřad Libereckého kraje, 2007.

[6] MULÍČEK, O. *Faktorová analýza – příklad Brna.* [online]. Brno: Masarykova univerzita v Brně, 2006. [cit. 2007-02-20]. Available from `http://everest.natur.cuni.cz/akce/segregace/ publikace/Mulicek.pdf`

[7] *Regional Development Strategy of the Czech Republic for the Period 2007 - 2013.* [on-line]. Prague: Ministry for Regional Development, 2006 [cit. 2007-02-08]. Available from `http://www.mmr.cz/strategie-regionalniho-rozvoje-cr-pro-obdobi-2007-2013`

[8] RYDVALOVÁ, P., ŽIŽKA, M. *Klíčové faktory problematického vývoje regionů v České republice.* 1. vyd. Liberec: VÚTS, 2007, 54 pgs. ISBN 978-80-903865-5-6

[9] RYDVALOVÁ, P., ŽIŽKA, M. Metodika hodnocení hospodářsky slabých oblastí v ČR. In *Sborník příspěvků z mezinárodní konference Hradecké ekonomické dny 2006.* Hradec Králové: Gaudeamus, 2006, pp. 405 – 410. ISBN 80–7041–895-8

[10] RYDVALOVÁ, P., ŽIŽKA, M. Ukazatele charakterizující hospodářskou úroveň regionů. In *Sborník příspěvků z konference HRADECKÉ EKONOMICKÉ DNY 2008.* Hradec Králové: Gaudeamus, 2008, 7 pgs. (CD-ROM). ISBN 978-80-7041-202-2

[11] VÍT, V., et al. Program rozvoje Libereckého kraje 2004 – 2006. Liberec: Krajský úřad Libereckého kraje, 2004.

Karel Sladký, Milan Sitař

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4, 182 08 Praha 8

e-mail: sladky@utia.cas.cz, milan.sitar@tiscali.cz

Risk-Sensitive and Mean Variance Optimality in Markov Decision Processes

Abstract

In this note, we compare two approaches for handling risk-variability features arising in discrete-time Markov decision processes: models with exponential utility functions and mean variance optimality models. Computational approaches for finding optimal decision with respect to the optimality criteria mentioned above are presented and analytical results showing connections between the above optimality criteria are discussed. **Keywords:** *Markov decision chains, exponential utility functions, certainty equivalent, expectation and variance of cumulative rewards, mean variance optimality, asymptotic behaviour*

1 Introduction

The usual optimization criteria examined in the literature on optimization of Markov reward processes, e.g. total discounted or mean reward, may be quite insufficient to characterize the problem from the point of the decision maker. To this end it is necessary to select more sophisticated criteria that reflect also the variability-risk features of the problem.

Perhaps the best known approach how to handle such problems stems from the classical work of Markowitz [19] on mean variance selection rules for the portfolio selection problem. Following the mean variance selection rule, the investor selects from among a given set of investment alternatives only investments with a higher mean and lower variance than a member of the given set.

The mean variance selection rule can also be employed in Markovian decision models. Following this approach along with the total reward or long run average expected return (i.e. the mean reward per transition) we consider total or average variance of the (long run) cumulative rewards. For details see [11, 12, 14, 15, 16, 18, 29, 30], the review paper by White [32], and also recent results of the present authors [22, 24, 25, 26, 27]. It is important to notice that in many of the above papers the long run average "variance" is considered only with respect to one-stage reward variance and not to variance of cumulative rewards; hence it is more appropriate to speak about "average variability" instead of "average variance." As it was shown on a number of numerical examples in [22], optimal solutions based on "average variability" are mostly different of optimal solutions based on precisely calculated "average variance."

Another possible approach how to attack the variability-risk features arising in Markovian decision problems is to consider, instead of linear objective functions, exponential utility functions.

Recall that exponential utility functions are the most widely used non-linear utility functions, cf. [6], and only linear and exponential functions are separable and hence appropriate for sequential decisions. Furthermore, Kirkwood [17] shows that in most cases an appropriately chosen exponential utility function is a very good approximation for general utility function. In [10] Howard demonstrates importance of exponential functions for treatment of a wide range of individual and risk preferences. The research of Markov decision processes with exponential objective functions, called risk-sensitive Markov decision processes, was initiated in the seminal paper by Howard and Matheson [9] and followed by many other researchers in recent years (see e.g. [3, 4, 5, 13, 23, 28]).

In this note we focus attention on risk-sensitive optimality criteria (i.e. the case when expectation of the stream of rewards generated by the Markov processes is evaluated by an exponential utility function) and their connections with mean-variance optimality (i.e. the case when a suitable combination of the expected total reward and its variance, usually considered per transition, is selected as a reasonable optimality criterion).

It is well known from the literature (see e.g. [31]) that for an exponential utility function, say $u^\gamma(\cdot)$, i.e. utility function with constant risk sensitivity $\gamma \in \mathbb{R}$, the utility assigned to the (random) reward ξ is given by

$$u^\gamma(\xi) := \begin{cases} \operatorname{sign}(\gamma)\, \exp(\gamma\xi) & \text{if } \gamma \neq 0 \\ \xi & \text{for } \gamma = 0. \end{cases} \tag{1}$$

Obviously $u^\gamma(\cdot)$ is continuous and strictly increasing. Moreover, if $\gamma > 0$ then $u^\gamma(\xi) = \exp(\gamma\xi)$ is convex and the decision maker is risk seeking. On the other hand if $\gamma < 0$ then $u^\gamma(\xi) = -\exp(\gamma\xi)$ is concave and the decision maker is risk averse.

The following facts are useful in the sequel:

1. For $U^{(\gamma)}(\xi) := \mathsf{E}\exp(\gamma\xi)$ the Taylor expansion around $\gamma = 0$ reads (in what follows E is reserved for expectation)

$$U^{(\gamma)}(\xi) = 1 + \mathsf{E}\sum_{k=1}^{\infty} \frac{(\gamma\xi)^k}{k!} = 1 + \sum_{k=1}^{\infty} \frac{\gamma^k}{k!} \cdot \mathsf{E}\,\xi^k. \tag{2}$$

Observe that in (2) the first (resp. second) term of the Taylor expansion is equal to $\gamma\mathsf{E}\,\xi$ (resp. $\frac{1}{2}(\gamma^2)\mathsf{E}\,\xi^2$). In particular, if for random variables ξ, ζ with $\mathsf{E}\,\xi = \mathsf{E}\,\zeta$ it holds $\mathsf{E}\,\xi^2 < \mathsf{E}\,\zeta^2$ (or equivalently $\operatorname{var}\xi < \operatorname{var}\zeta$) then there exists $\gamma_0 > 0$ such that $U^{(\gamma)}(\xi) < U^{(\gamma)}(\zeta)$ for any $\gamma \in (-\gamma_0, \gamma_0)$.

2. For $Z(\xi)$, the certainty equivalent of the (random) variable ξ, given by the condition $u^\gamma(Z(\xi)) = \mathsf{E}[u^\gamma(\xi)]$), we immediately get

$$Z(\xi) = \begin{cases} \frac{1}{\gamma}\ln\{\mathsf{E}\,[\exp(\gamma\xi)]\} & \text{if } \gamma \neq 0 \\ \mathsf{E}\,[\xi] & \text{for } \gamma = 0. \end{cases} \tag{3}$$

Observe that if ξ is constant then $Z(\xi) = \xi$, if ξ is nonconstant then by Jensen's inequality

$$\begin{array}{lll} Z(\xi) > \mathsf{E}\,\xi & (\text{if } \gamma > 0, & \text{the risk seeking case}) \\ Z(\xi) < \mathsf{E}\,\xi & (\text{if } \gamma < 0, & \text{the risk averse case}) \\ Z(\xi) = \mathsf{E}\,\xi & (\text{if } \gamma = 0, & \text{the risk neutral case}) \end{array}$$

3. Finally, recall that exponential utility function considered in (1) is separable what is very important for sequential decision problems, i.e. $u^\gamma(\xi^{(1)} + \xi^{(2)}) = \text{sign}(\gamma)\, u^\gamma(\xi^{(1)}) \cdot u^\gamma(\xi^{(2)})$.

4. In economic models (see e.g. [1], [31]) we usually assume that the utility function $u(\cdot)$ is increasing (i.e. $u'(\cdot) > 0$), concave (i.e. $u''(\cdot) < 0$, what is fulfilled in (1) for $\gamma < 0$) with $u(0) = 0$ and $u'(0) < \infty$ (so called the Inada condition).

Since a linear transformation of the utility function $u^\gamma(\xi)$ preserves the original preferences (cf. [1],[31]) we shall also consider the utility functions

$$\bar{u}^\gamma(x) = 1 - \exp(\gamma x), \quad \text{where } \gamma < 0 \quad (\text{the risk averse case}) \tag{4}$$

$$\tilde{u}^\gamma(x) = \exp(\gamma x) - 1, \quad \text{where } \gamma > 0 \quad (\text{the risk seeking case}) \tag{5}$$

and the function $\bar{u}^\gamma(x)$ satisfies all above conditions imposed on a utility function in economy theory. Observe that the Taylor expansions of $\bar{u}^\gamma(x)$ and of $\tilde{u}^\gamma(x)$ read

$$\bar{u}^\gamma(x) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{|\gamma|^k}{k!} \cdot x^k, \quad \text{where } \gamma < 0, \qquad \tilde{u}^\gamma(x) = \sum_{k=1}^{\infty} \frac{\gamma^k}{k!} \cdot x^k, \quad \text{where } \gamma > 0 \tag{6}$$

and if $x = \xi$ is a random variable for the expected utilities we have

$$\bar{U}^\gamma(\xi) := \mathsf{E}\,\bar{u}^\gamma(\xi) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{|\gamma|^k}{k!} \cdot \mathsf{E}\,\xi^k, \qquad \tilde{U}^\gamma(\xi) := \mathsf{E}\,\tilde{u}^\gamma(\xi) = \sum_{k=1}^{\infty} \frac{\gamma^k}{k!} \cdot \mathsf{E}\,\xi^k. \tag{7}$$

In this note we focus attention on properties of the expected utility and the corresponding certainty equivalents if the stream of obtained rewards is evaluated by exponential utility functions and their connections with more classical mean-variance optimality criteria.

2 Notation and Preliminaries

Consider a Markov decision chain $X = \{X_n, n = 0, 1, \ldots\}$ with finite state space $\mathcal{I} = \{1, \ldots, N\}$, finite set $\mathcal{A}_i = \{1, 2, \ldots, K_i\}$ of possible decisions (actions) in state $i \in \mathcal{I}$ and the following transition and reward structure (we assume that in state i action $a \in \mathcal{A}_i$ is selected):

$$\begin{array}{rl} p_{ij}^a: & \text{transition probability from } i \to j \ \ (i, j \in \mathcal{I}), \\ r_{ij}: & \text{one-stage reward for a transition from } i \to j, \\ r_i^a: & \text{expected value of the one-stage rewards incurred in state } i, \\ r_i^{(2),a}: & \text{second moment of the one-stage rewards incurred in state } i. \end{array}$$

Obviously, $r_i^a = \sum_{j \in \mathcal{I}} p_{ij}^a \cdot r_{ij}$, $r_i^{(2),a} = \sum_{j \in \mathcal{I}} p_{ij}^a \cdot [r_{ij}]^2$ and hence the corresponding one-stage reward variance $\sigma_i^{2,a} = r_i^{(2),a} - [r_i^a]^2$.

Policy controlling the chain is a rule how to select actions in each state. In this note, we restrict on stationary policies, i.e. the rules selecting actions only with respect to the current state of the Markov chain X. Then a policy, say π, is determined by some decision vector f whose ith element $f_i \in \mathcal{A}_i$ identifies the action taken if the chain X is in state $X_n = i$; hence also the transition probability matrix $\boldsymbol{P}(f)$ of the Markov decision chain. Observe that the ith row of $\boldsymbol{P}(f)$ has elements $p_{i1}^{f_i}, \ldots, p_{iN}^{f_i}$ and that $\boldsymbol{P}^*(f) = \lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} [\boldsymbol{P}(f)]^k$ exists. In what follows, $\boldsymbol{R} = [r_{ij}]$ is the transition reward matrix, i.e. $\boldsymbol{R}$ is an $N \times N$ matrix of one-stage rewards. Similarly, $\boldsymbol{r}(f)$ is the (column) vector of one-stage expected rewards with elements $r_1^{f_1}, \ldots, r_N^{f_N}$.

Let elements of the vectors $\boldsymbol{R}^\pi(n)$, $\boldsymbol{S}^\pi(n)$ and $\boldsymbol{V}^\pi(n)$ denote the first moment, the second moment and the variance of the (random) total reward $\xi_i^{(n)}(\pi)$ respectively received in the n next transitions of the considered Markov chain X if policy $\pi \sim (f)$ is followed, given the initial state $X_0 = i$. In what follows we sometimes abbreviate $\xi_i^{(n)}(\pi)$ by $\xi_i^{(n)}$ or by $\xi^{(n)}$ if the dependence on policy $\pi \sim (f)$ or initial state $X_0 = i$ is obvious.

More precisely, for the elements of $\boldsymbol{R}^\pi(n)$, $\boldsymbol{S}^\pi(n)$ and $\boldsymbol{V}^\pi(n)$ we have

$$R_i^\pi(n) = \mathsf{E}_i^\pi[\xi^{(n)}], \quad S_i^\pi(n) = \mathsf{E}_i^\pi[\xi^{(n)}]^2, \quad V_i^\pi(n) = \boldsymbol{\sigma}_i^{2,\pi}[\xi^{(n)}]$$

where $\xi^{(n)} = \sum_{k=0}^{n-1} r_{X_k, X_{k+1}}$ and E_i^π, $\boldsymbol{\sigma}_i^{2,\pi}$ are standard symbols for expectation and variance if policy π is selected and $X_0 = i$. Moreover, if $m < n$ we can write $\xi_{X_0}^{(n)} = \xi_{X_0}^{(m)} + \xi_{X_m}^{(m,n)}$, where $\xi_{X_m}^{(m,n)} = \sum_{k=m}^{n-1} r_{X_k, X_{k+1}}$ is reserved for the (random) reward obtained from the mth up to the nth transition.

Recall that

$$R_i^\pi(n+1) = r_i^{f_i} + \sum_{j\in\mathcal{I}} p_{ij}^{f_i} R_j^\pi(n) \tag{8}$$

or in vector notation

$$\boldsymbol{R}^\pi(n+1) = \boldsymbol{r}(f) + \boldsymbol{P}(f)\cdot \boldsymbol{R}^\pi(n). \tag{9}$$

Similarly, if the chain starts in state i and policy $\pi \sim (f)$ is followed then from (2), (3) for $\xi = \xi^{(n)}$ for the expected utility $U_i^\pi(\gamma, n)$, the certainty equivalent $Z_i^\pi(\gamma, n)$ and its mean value $J_i^\pi(\gamma, n)$ we have

$$U_i^\pi(\gamma, n) := \mathsf{E}_i^\pi[\exp(\gamma \xi^{(n)})] = \mathsf{E}_i^\pi \exp[\gamma\,(r_{i,X_1} + \xi_{X_1}^{(1,n)})], \tag{10}$$

$$Z_i^\pi(\gamma, n) := \frac{1}{\gamma} \ln\{\mathsf{E}_i^\pi\,[\exp(\gamma\xi^{(n)})]\} \qquad \text{for } \gamma \neq 0, \tag{11}$$

$$J_i^\pi(\gamma, n) := \lim_{n\to\infty} \frac{1}{n} Z_i^\pi(\gamma, n) \tag{12}$$

and hence for the expectation of the utility functions $\bar{u}^\gamma(\xi^{(n)})$ and $\tilde{u}^\gamma(\xi^{(n)})$ we have (cf. (1.7))

$$\bar{U}_i^\pi(\gamma, n) := 1 - U_i^\pi(\gamma, n), \qquad \tilde{U}_i^\pi(\gamma, n) := U_i^\pi(\gamma, n) - 1. \tag{13}$$

In what follows let $\boldsymbol{U}^\pi(\gamma, n)$, resp. $\boldsymbol{Z}^\pi(\gamma, n)$, be the vector of expected utilities, resp. certainty equivalents, with elements $U_i^\pi(\gamma, n)$, resp. $Z_i^\pi(\gamma, n)$.

Conditioning in (10) on X_1, since policy $\pi \sim (f)$ is stationary, from (10) we immediately get the recurrence formula

$$U_i^\pi(\gamma, n+1) \;=\; \sum_{j\in\mathcal{I}} p_{ij}^{f_i} \cdot \mathrm{e}^{\gamma r_{ij}} \cdot U_j^\pi(\gamma, n) = \sum_{j\in\mathcal{I}} q_{ij}^{f_i} \cdot U_j^\pi(\gamma, n) \quad \text{with} \;\; U_i^\pi(\gamma, 0) = 1 \;\; \text{or} \tag{14}$$

in vector notation

$$\boldsymbol{U}^\pi(\gamma, n+1) \;=\; \boldsymbol{Q}(f) \cdot \boldsymbol{U}^\pi(\gamma, n) \qquad \text{with} \;\; \boldsymbol{U}^\pi(\gamma, n) = \boldsymbol{e}, \tag{15}$$

where $\boldsymbol{Q}(f) = [q_{ij}^{f_i}]$ with $q_{ij}^{f_i} := p_{ij}^{f_i} \cdot \mathrm{e}^{\gamma r_{ij}}$.

Observe that $\boldsymbol{Q}(f)$ is a nonnegative matrix, and by the Perron–Frobenius theorem (cf. [7]) the spectral radius $\rho(f)$ of $\boldsymbol{Q}(f)$ is equal to the maximum positive eigenvalue of $\boldsymbol{Q}(f)$. Moreover, if $\boldsymbol{Q}(f)$ is irreducible (i.e. if and only if $\boldsymbol{P}(f)$ is irreducible) the corresponding (right) eigenvector $\boldsymbol{v}(f)$ can be selected strictly positive, i.e.

$$\rho(f)\, \boldsymbol{v}(f) = \boldsymbol{Q}(f) \cdot \boldsymbol{v}(f) \qquad \text{with} \quad \boldsymbol{v}(f) > 0. \tag{16}$$

Moreover, under the above irreducibility condition it can be shown (cf. e.g. [9], [28]) that there exists decision vector $f^* \in \mathcal{A}$ such that

$$\boldsymbol{Q}(f) \cdot \boldsymbol{v}(f^*) \;\;\le\;\; \rho(f^*)\, \boldsymbol{v}(f^*) = \boldsymbol{Q}(f^*) \cdot \boldsymbol{v}(f^*), \tag{17}$$

$$\rho(f) \;\;\le\;\; \rho(f^*) \equiv \rho^* \qquad \text{for all} \;\; f \in \mathcal{A}. \tag{18}$$

In words, $\rho(f^*) \equiv \rho^*$ is the maximum possible eigenvalue of $\boldsymbol{Q}(f)$ over all $f \in \mathcal{A}$.

Throughout this note we make the following assumptions.

AS 1. For any stationary policy $\pi \sim (f)$, the transition probability matrix $\boldsymbol{P}(f)$ is irreducible (i.e. all states are communicating) and aperiodic, i.e. $\boldsymbol{P}(f)$ is ergodic (all states are recurrent and aperiodic).

Observe that under AS 1 the matrix $\boldsymbol{Q}(f)$ is irreducible for any $f \in \mathcal{A}$.

AS 2. Transition rewards are nonnegative and nonvanishing, i.e. $r_{ij} \ge 0$ for all $i, j \in \mathcal{I}$ and a strict inequality holds at least for one pair i, j.

Observe that under AS 2 all one-stage expected rewards $r_i(\cdot)$ are nonnegative.

3 Reward Variance and Expected Utility

Since for any integers $m < n$ $\;[\xi^{(n)}]^2 = [\xi^{(m)}]^2 + 2 \cdot \xi^{(m)} \cdot \xi^{(m,n)} + [\xi^{(m,n)}]^2$ we get

$$\mathsf{E}_i^\pi[\xi^{(n)}]^2 = \mathsf{E}_i^\pi[\xi^{(m)}]^2 + 2 \cdot \mathsf{E}_i^\pi[\xi^{(m)} \cdot \xi^{(m,n)}] + \mathsf{E}_i^\pi[\xi^{(m,n)}]^2. \tag{19}$$

In particular, for $m = 1$, $n := n+1$ if policy $\pi \sim (f)$ is followed we get for the second moment of the random reward $\xi^{(n)}$:

$$S_i^\pi(n+1) = \sum_{j\in\mathcal{I}} p_{ij}^{f_i} \cdot \{[r_{ij}]^2 + 2 \cdot r_{ij} \cdot R_j^\pi(n)\} + \sum_{j\in\mathcal{I}} p_{ij}^{f_i} \cdot S_j^\pi(n). \tag{20}$$

Since the variance $V_i(\cdot) = S_i(\cdot) - [R_i(\cdot)]^2$ from (20) we arrive at

$$V_i^\pi(n+1) = \sum_{j\in\mathcal{I}} p_{ij}^{f_i} \cdot \{[r_{ij} + R_j^\pi(n)]^2\} - [R_i^\pi(n+1)]^2 + \sum_{j\in\mathcal{I}} p_{ij}^{f_i} \cdot V_j^\pi(n). \tag{21}$$

From the literature (see e.g. [8, 20, 21] it is well known that under AS 1 there exist vector $\boldsymbol{w}^\pi$ (with elements w_j^π), constant vector $\boldsymbol{g}^\pi$ and vector $\boldsymbol{\varepsilon}(n)$ (where all elements of $\boldsymbol{\varepsilon}(n)$ converge to zero geometrically) such that

$$\boldsymbol{R}^\pi(n) = \boldsymbol{g}^\pi \cdot n + \boldsymbol{w}^\pi + \boldsymbol{\varepsilon}(n) \Rightarrow \lim_{n\to\infty} n^{-1}\boldsymbol{R}^\pi(n) = \boldsymbol{g}^\pi = \boldsymbol{P}^*(f)\cdot\boldsymbol{r}(f). \tag{22}$$

The constant vector $\boldsymbol{g}^\pi$ (with elements g^π) along with vector $\boldsymbol{w}^\pi$ are uniquely determined by

$$\boldsymbol{w}^\pi + \boldsymbol{g}^\pi \quad = \quad \boldsymbol{r}(f) + \boldsymbol{P}(f)\cdot\boldsymbol{w}^\pi, \quad \boldsymbol{P}^*(f)\cdot\boldsymbol{w}^\pi = \mathbf{0}. \tag{23}$$

By using relations (21), (22) and (23) in a number of steps we arrive at (for details see [24, 26, 25, 27]):

$$\boldsymbol{V}^\pi(n+1) = \boldsymbol{s}(\pi) + \boldsymbol{P}(f)\cdot\boldsymbol{V}^\pi(n) + \boldsymbol{\varepsilon}^{(1)}(n) \tag{24}$$

where for elements of the vector $\boldsymbol{s}(\pi)$ we have

$$s_i(\pi) = \sum_{j\in\mathcal{I}} p_{ij}^{f_i} \cdot \{[r_{ij} + w_j^\pi]^2\} - [g^\pi + w_i^\pi]^2 \tag{25}$$

$$= \sum_{j\in\mathcal{I}} p_{ij}^{f_i} \cdot \{[r_{ij} - g^\pi + w_j^\pi]^2\} - [w_i^\pi]^2 \tag{26}$$

and elements of the vector $\boldsymbol{\varepsilon}^{(1)}(n)$ converge to zero geometrically.

In analogy with (22), (23) we can conclude that there exists vector $\boldsymbol{w}^{(2),\pi}$ along with a constant vector $\boldsymbol{g}^{(2),\pi}$ uniquely determined by

$$\boldsymbol{w}^{(2),\pi} + \boldsymbol{g}^{(2),\pi} = \boldsymbol{s}(\pi) + \boldsymbol{P}(f)\cdot\boldsymbol{w}^{(2),\pi}, \quad \boldsymbol{P}^*(f)\cdot\boldsymbol{w}^{(2),\pi} = \mathbf{0} \tag{27}$$

such that

$$\boldsymbol{V}^\pi(n) = \boldsymbol{g}^{(2),\pi}\cdot n + \boldsymbol{w}^{(2),\pi} + \boldsymbol{\varepsilon}(n) \implies \boldsymbol{g}^{(2),\pi} = \lim_{n\to\infty}\frac{\boldsymbol{V}^\pi(n)}{n} = \boldsymbol{P}^*(f)\cdot\boldsymbol{s}(\pi). \tag{28}$$

Moreover, from (23), (24), (28) we can also conclude after some algebra (observe that $\boldsymbol{P}^*(f)\cdot\boldsymbol{P}(f)\cdot\boldsymbol{w}^{(2),\pi} = \boldsymbol{P}^*(f)\cdot\boldsymbol{w}^{(2),\pi}$, cf. [24] for details) that

$$\boldsymbol{g}^{(2),\pi} = \bar{\boldsymbol{g}}^{(2),\pi} + 2\cdot\boldsymbol{P}^*(f)\cdot\tilde{\boldsymbol{r}}(f,\pi) \tag{29}$$

where ($[\cdot]_{\mathrm{sq}}$ denotes that elements of the vector are squared)

$\boldsymbol{r}^{(2)}(f)$ is a column vector with elements $r_i^{(2),f_i} = \sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot[r_{ij}]^2$,

$\tilde{\boldsymbol{r}}(f,\pi)$ is a column vector with elements $\sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot r_{ij}\cdot w_j^\pi$,

$[\boldsymbol{g}^\pi]_{\mathrm{sq}}$ is a constant vector with elements $[g^\pi]^2$,

$\tilde{\boldsymbol{g}}^{(2),\pi} = \boldsymbol{P}^*(f) \cdot \boldsymbol{r}^{(2)}(f)$ is a constant vector with elements $\tilde{g}^{(2),\pi}$, and

$\bar{\boldsymbol{g}}^{(2),\pi} = \tilde{\boldsymbol{g}}^{(2),\pi} - [\boldsymbol{g}^\pi]_{\rm sq}$ is a constant vector with elements $\bar{g}^{(2),\pi}$.

Obviously, $\tilde{g}^{(2),\pi}$ averages expected values of the second moments of one-stage rewards, $\bar{g}^{(2),\pi}$ denotes the average "one-stage reward variance" considered with respect to the mean reward g^π instead of the one-stage expected reward $r_i^{f_i}$ in state $i \in \mathcal{I}$, and the last term in (29) expresses the Markov dependence that occurs if the total variance of cumulative rewards is considered.

On the other hand, on iterating (14) we get if (stationary) policy $\pi \sim (f^*)$ is followed

$$\boldsymbol{U}^\pi(\gamma, n) = (\boldsymbol{Q}(f^*))^n \cdot \boldsymbol{e}. \tag{30}$$

Since under AS 1 the Perron eigenvector $\boldsymbol{v}(f^*)$ is strictly positive, there exist numbers $\alpha_1 < \alpha_2$ such that $\alpha_1 \cdot \boldsymbol{v}(f^*) \le \boldsymbol{e} \le \alpha_2 \cdot \boldsymbol{v}(f^*)$ and hence

$$\alpha_1 \cdot (\rho(f^*))^n \cdot \boldsymbol{v}(f^*) \le \boldsymbol{U}^\pi(\gamma, n) \le \alpha_2 \cdot (\rho(f^*))^n \cdot \boldsymbol{v}(f^*). \tag{31}$$

From (31) we can see that the asymptotic behaviour of $\boldsymbol{U}^\pi(\gamma, n)$ heavily depends on $\rho(f^*)$, and the growth rate of each $U_i^\pi(\gamma, n)$ is the same and equal to $\rho(f^*)$.

On examining $\boldsymbol{Q}(f)$ we can easily conclude that:

If $\gamma < 0$ then $\rho(f^*) < 1$ for all $f \in \mathcal{A}$, hence $\lim_{n\to\infty} [\boldsymbol{Q}(f)]^n = \boldsymbol{0}$ and $U_i^\pi(\gamma, n) \to 0$ for all $i \in \mathcal{I}$ as $n \to \infty$ and the convergence is geometrical.

For $\gamma = 0$ we have $\boldsymbol{Q}(f) = \boldsymbol{P}(f)$, the spectral radius of $\boldsymbol{Q}(f)$ equals one, and the corresponding right eigenvector $\boldsymbol{v}(f^*)$ is a constant vector. Then $\boldsymbol{U}^\pi(\gamma, n) \to \boldsymbol{P}^*(f) \cdot \boldsymbol{e}$, a constant vector.

If $\gamma > 0$ then $\rho(f^*) > 1$ for $f^* \in \mathcal{A}$, hence elements of $[\boldsymbol{Q}(f^*)]^n$ go to infinity and also $U_i^{\pi^*}(\gamma, n) \to \infty$ for all $i \in \mathcal{I}$ as $n \to \infty$.

Moreover, by (2) we can expect that for γ sufficiently close to null the growth of $U_i^\pi(\gamma, n)$ will be dominated by the first and second moment of $\xi^{(n)}$ occurring in (22).

Illustrative Example.

Consider a controlled Markov reward chain with 5 states and only three possible actions in state 1 (in the remaining states no option is possible). Hence only three transition probability matrices, say $\boldsymbol{P}(f^{(1)})$, $\boldsymbol{P}(f^{(2)})$, $\boldsymbol{P}(f^{(3)})$, are available that along with the reward matrix $\boldsymbol{R}$ fully characterize the transition and reward structures of the considered Markov reward chain. Observe that stationary policies $\pi^{(2)} \sim (f^{(2)})$, $\pi^{(3)} \sim (f^{(3)})$ identify a constant sequence of one-stage rewards. In particular, we have

$$\boldsymbol{P}(f^{(1)}) = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0 & 0.5 \end{bmatrix}, \qquad \boldsymbol{R} = \begin{bmatrix} 0 & 1 & 0 & 0.5 & 0.48 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 0.48 & 0 & 0 & 0 & 0.48 \end{bmatrix},$$

$$
\boldsymbol{P}(f^{(2)}) = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0 & 0.5 \end{bmatrix}, \qquad \boldsymbol{P}(f^{(3)}) = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0 & 0.5 \end{bmatrix}.
$$

Obviously, if the chain starts in state 1 and action 2 (resp. 3) is selected then $r_{X_k,X_{k+1}} \equiv 0.5$ (resp. $r_{X_k,X_{k+1}} \equiv 0.48$) for all $k = 0, 1, \ldots$ and the chain visits only the states $1, 4$ (resp. $1, 5$). Hence for the expected reward, the second moment and the variance respectively, we have $R_1(k) = 0.5k$, $S_1(k) = (0.5k)^2$, $V_1(k) \equiv 0$ (resp. $R_1(k) = 0.48k$, $S_1(k) = (0.48k)^2$, $V_1(k) \equiv 0$). On the contrary if the chain starts in state 1 and action 1 is selected on inspecting the chain we can see that the chain visits only the states $1, 2, 3$ and that the sequence of received rewards obeys a binomial distribution with parameter $p = 0.5$, and hence again $R_1(k) = 0.5k$, however $V_1(k) = k\,0.5(1 - 0.5) = 0.25k$.

The same results can be obtained by using the general formulas for expected reward and variance of the Markov reward chains as it is shown in the further text.

Observe that by (22), (23) for

$$
\begin{aligned}
&\pi^{(1)} \sim (f^{(1)}), \quad g^{\pi} = 0.5, \quad \tilde{g}^{(2),\pi} = 0.5, \quad &&\text{hence } g^{(2),\pi} = 0.25, \\
&\pi^{(2)} \sim (f^{(2)}), \quad g^{\pi} = 0.5, \quad \tilde{g}^{(2),\pi} = 0.25, \quad &&\text{hence } g^{(2),\pi} = 0, \\
&\pi^{(3)} \sim (f^{(3)}), \quad g^{\pi} = 0.48, \quad \tilde{g}^{(2),\pi} = (0.48)^2, \quad &&\text{hence } g^{(2),\pi} = 0.
\end{aligned}
$$

Of course, following policy $\pi^{(2)} \sim (f^{(2)})$ we get maximum possible mean reward and null variance, this policy is the best choice. However, the second best policy can be either $\pi^{(3)} \sim (f^{(3)})$ guaranteeing mean reward slightly less than then maximum one and the null variance or policy $\pi^{(1)} \sim (f^{(1)})$ giving the maximum mean reward, but the variance comparable with the mean reward. Considering optimality in accordance the weighted optimality criterion $\alpha g^{\pi} - (1 - \alpha) g^{(2),\pi}$ (where $\alpha \in [0, 1]$) it depends on the decision maker option how to selected the weighting coefficient α and prefer either policy $\pi^{(1)} \sim (f^{(1)})$ or policy $\pi^{(3)} \sim (f^{(3)})$. Since for $\alpha_0 = \frac{25}{27}$ it holds $\alpha_0\, g^{\pi_1} - (1 - \alpha)\, g^{(2),\pi_1} = \alpha_0\, g^{\pi_3}$, if $\alpha \in [\alpha_0, 1]$ we prefer policy $\pi^{(1)}$ above policy $\pi^{(3)}$; if $\alpha < \alpha_0$ we consider policy $\pi^{(3)}$ as the second best.

Employing the risk-sensitive model and supposing that the chain starts in state 1 it is sufficient to consider the following matrices:

$$
\widetilde{\boldsymbol{Q}}(f^{(1)}) = \begin{bmatrix} 0 & \frac{1}{2} \cdot \mathrm{e}^{\gamma} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \cdot \mathrm{e}^{\gamma} \\ \frac{1}{2} \cdot \mathrm{e}^{\gamma} & \frac{1}{2} & 0 \end{bmatrix}, \quad \widetilde{\boldsymbol{Q}}(f^{(2)}) = \begin{bmatrix} 0 & \mathrm{e}^{\gamma 0.5} \\ \frac{1}{2}\mathrm{e}^{\gamma 0.5} & \frac{1}{2}\mathrm{e}^{\gamma 0.5} \end{bmatrix}, \quad \widetilde{\boldsymbol{Q}}(f^{(3)}) = \begin{bmatrix} 0 & \mathrm{e}^{\gamma 0.48} \\ \frac{1}{2}\mathrm{e}^{\gamma 0.48} & \frac{1}{2}\mathrm{e}^{\gamma 0.48} \end{bmatrix}
$$

The right Perron eigenvector of each above matrix is a unit vector of appropriate dimension and for the spectral radii we get $\tilde{\rho}(f^{(1)}) = 0.5\,(\mathrm{e}^{\gamma} + 1)$, $\quad \tilde{\rho}(f^{(2)}) = \mathrm{e}^{\gamma\,0.5}$, $\quad \tilde{\rho}(f^{(3)}) = \mathrm{e}^{\gamma\,0.48}$. Taking into account maximum growth rate of the exponential utility function (1.7), and hence the maximal growth rate of the expected utility $\bar{U}^{\gamma}(\xi) = 1 - U^{(\gamma)}(\xi)$ (since the considered risk aversion coefficient γ is negative), obtained for the minimum possible Perron eigenvalue of the matrix $\widetilde{\boldsymbol{Q}}(\cdot)$, we get that $\tilde{\rho}(f^{(2)}) < \tilde{\rho}(f^{(3)})$ for $\gamma < 0$ implying that policy $\pi^{(2)} \sim (f^{(2)})$ is "better" than $\pi^{(3)} \sim (f^{(3)})$ with respect to the considered risk-sensitive criterion. Similarly, on comparing $\tilde{\rho}(f^{(3)}) = \mathrm{e}^{\gamma\, 0.48}$ and $\tilde{\rho}(f^{(1)}) = 0.5\, (\mathrm{e}^{\gamma} + 1)$ we can decide whether $\pi^{(3)} \sim (f^{(3)})$ or $\pi^{(1)} \sim (f^{(1)})$ is the second best policy for the considered the value of the risk aversion coefficient γ.

Acknowledgement. This research was supported by the Czech Science Foundation under Grants 402/08/0107 and 402/07/1113.

References

[1] J. von Neumann and O. Morgenstern: Theory of Games and Economic Behaviour. Third Edition, Princeton Univ. Press, Princeton, NJ 1953.

[2] Ch. Barz: Risk-Averse Capacity Control in Revenue Management. Springer, Berlin–Heidelberg 2007.

[3] T. Bielecki, D. Hernández-Hernández, and S. R. Pliska: Risk-sensitive control of finite state Markov chains in discrete time, with application to portfolio management. Math. Methods Oper. Res. *50* (1999), 167–188.

[4] R. Cavazos-Cadena and R. Montes-de-Oca: The value iteration algorithm in risk-sensitive average Markov decision chains with finite state space. Math. Oper. Res. *28* (2003), 752–756.

[5] R. Cavazos-Cadena: Solution to the risk-sensitive average cost optimality equation in a class of Markov decision processes with finite state space. Math. Methods Oper. Res. *57* (2003), 253–285.

[6] J. L. Corner and P. D. Corner: Characterization of decision in decision analysis practice. J. Oper. Res. Soc. *46* (1995), 304–314.

[7] F. R. Gantmakher: The Theory of Matrices. Chelsea, London 1959.

[8] R. A. Howard: Dynamic Programming and Markov Processes. MIT Press, Cambridge, Mass. 1960.

[9] R. A. Howard and J. Matheson: Risk-sensitive Markov decision processes. Manag. Sci. *23* (1972), 356–369.

[10] R. A. Howard: Decision analysis: Practice and promise. Manag. Sci. *34* (1988), 679–695.

[11] S. C. Jaquette: Markov decision processes with a new optimality criterion: small interest rates. Ann. Math. Statist. *43* (1972), 1894–1901.

[12] S. C. Jaquette: Markov decision processes with a new optimality criterion: discrete time. Ann. Statist. *1* (1973), 496–505.

[13] S. C. Jaquette: A utility criterion for Markov decision processes. Manag. Sci. *23* (1976), 43–49.

[14] J. Filar, L. C. M. Kallenberg, and H.-M. Lee: Variance penalized Markov decision processes. Mathem. Oper. Research *14* (1989), 147–161.

[15] Ying Huang and L. C. M. Kallenberg: On finding optimal policies for Markov decision chains: a unifying framework for mean-variance-tradeoffs. Mathem. Oper. Research *19* (1994), 434–448.

[16] H. Kawai: A variance minimization problem for a Markov decision process. European J. Oper. Research *31* (1987), 140–145. Springer, Berlin 2004, pp. 43–66.

[17] C. W. Kirkwood: Approximating risk aversion in decision analysis applications. Decision Analysis *1* (2004), 51–67.

[18] P. Mandl: On the variance in controlled Markov chains. Kybernetika *7* (1971), 1–12.

[19] H. Markowitz: Portfolio Selection – Efficient Diversification of Investments. Wiley, New York 1959.

[20] M. L. Puterman: Markov Decision Processes – Discrete Stochastic Dynamic Programming. Wiley, New York 1994.

[21] S. M. Ross: Applied Probability Models with Optimization Applications. Holden–Day, San Francisco, Calif. 1970.

[22] M. Sitař: Mean-Variance Optimality in Markov Decision Processes. Doctoral Thesis, Charles University, Prague 2006.

[23] K. Sladký: On dynamic programming recursions for multiplicative Markov decision chains. Math Programming Study *6* (1976), 216–226.

[24] K. Sladký and M. Sitař: Optimal solutions for undiscounted variance penalized Markov decision chains. In: Dynamic Stochastic Optimization (LNEMS 532, K. Marti, Y. Ermoliev, and G. Pflug, eds.), Springer, Berlin 2004, pp. 43–66.

[25] K. Sladký and M. Sitař: Algorithmic procedures for mean variance optimality in Markov decision chains. In: Operations Research Proceedings 2005 Springer, Berlin 2006, pp. 799–804.

[26] K. Sladký and M. Sitař: Mean–variance optimality in Markov decision chains. In: Mathematical Methods in Economics 2005, Gaudeamus, Hradec Králové 2005.

[27] K. Sladký: On mean reward variance in semi-Markov processes. Math. Methods Oper. Res. *62* (2005), 387–397.

[28] K. Sladký: Growth rates and average optimality in risk-sensitive Markov decision chains. Kybernetika *44* (2008), 206–217.

[29] M. J. Sobel: The variance of discounted Markov decision processes. J. Appl. Probab. *19* (1982), 794–802.

[30] M. J. Sobel: Maximal mean/standard deviation ratio in an undiscounted MDP. Oper. Research Lett. *4* (1985), 157–159.

[31] A. Takayama: Analytical Methods in Economics. Harvester Wheatsheaf, Hertfordshire 1994.

[32] D. J. White: Mean, variance and probability criteria in finite Markov decision processes: A review. J. Optim. Theory Appl. *6* (1988), 1–29.

Olga Subanova

Finance Academy under the Government of the Russian Federation
49, Leningradsky prospect, Moscow, 125468, Russia

e-mail: osubanova@mail.ru

Design of Optimal Spending Rule for the University Endowment

Abstract

University endowments are an important source of revenue which allocate to student financial aid, faculty chairs or professorships, scientific research and etc. The basic concept of an endowment is a fund of money created by donor's stipulation that requires the gift to be invested in perpetuity or for a specified term. Endowments typically grow over time through a combination of donations and investment returns. But it may be different year-by-year and universities have established various rules for spending their endowment earnings that smooth the effects of charitable giving resizing, high or low investment returns and operational management costs. I present a mathematical model for optimal endowment spending based on hybrid rule segregates the endowment into three parts: the original endowment, the stabilization fund and the "service" part for management costs. The model is developed for Russian universities which operate their endowments through asset management companies. **Keywords:** *Endowment management, educational finance, analytical hierarchical process*

1 Introduction

The basic concept of an endowment is a fund of money created by donor's stipulation that requires the gift to be invested in perpetuity or for a specified term. The fund is invested and a determined percent of the portfolio is used for the purposes restricted by the endowment gift or, if it is unrestricted, by the intent of the university administration. Thus, the endowment becomes the gift that lives forever and helps universities maintain affordability by providing revenue to subsidize the operating budget [1].

An endowment is not a singular entity. It's an accumulation of different donations that are put together for investment purposes. It usually consists of many of different funds and most of these funds are restricted meaning that they can only be spent for legally binding purposes that have been specified by the donors. Endowments typically grow over time through a combination of donations and investment returns and decrease in size from payouts for student financial aid, teaching, create professorships, research and innovation, staff and administration salaries and etc. Very few endowments are entirely unrestricted.

In many cases, three notions that probably come to mind when considering endowment management are accumulation of gifts, investment policy and spending rule. Managing the endowment is not a "stand-alone" endeavour; it's a perpetual process (Figure 1).

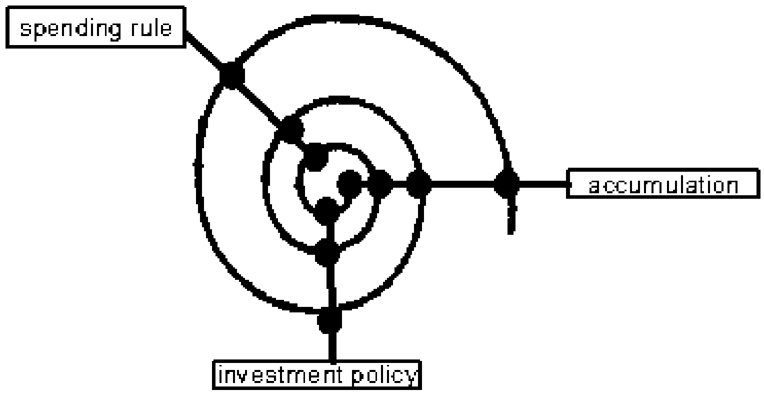


Figure 1: The endowment management process framework.

The accumulation of endowment is, in effect, a form of saving, presumably for expenditure in the future [2]. Endowments are managed for the long-term to strike a balance between money that are spent immediately and the obligation to preserve it for future generations. "Intergenerational equity" means, in the words of Yale economist 1981 Nobel laureate James Tobin, "to protect the interest of the future against the demands of the present". Future generations of students have as much entitlement to the benefits of the endowments as those currently enrolled, and their rights must be protected.

In other words, future generation of students receive at least the same level of support from an endowment as the current generation enjoys. Factors affecting intergenerational equity are: gift flows, returns and inflation, asset allocation (a target percentage (or percentage range) in stocks, bonds, cash and other asset classes that is consistent with the purpose for which the endowments are invested) and spending policy/spending rate (what sum should be spent per year and how this amount should change as circumstances change (e.g. unusually high or low investment returns, or new gifts received).

Each university adopts its own strategy and spending models to maximize its endowment's value to support both current activities and future needs. As Merton (1991) noted two universities with similar objectives and endowments can have very different optimal portfolios and expenditure patterns if their non-endowment sources of cash flow are different. Examples of such sources include public and private-sector grants, patents, gifts, university business income and etc. [3]. Taking account of those assets can cause the optimal spending.

Table 1 summarized popular spending approaches. According different analytical studies, nearly three-quarters of all American colleges and universities target their endowment spending at about 5% of a three-year rolling average of total endowment market value.

In Russia all of universities invest their endowments through asset management companies (it's against the law to manage with their own staff) that is why they don't have an enigma of asset

#	Name	Description
1	Ad Hoc	Spending rule varies from year-to-year based on the needs
2	Income based	Spend all current income
3	Inflation based	Increase spending each year based on rate of inflation
4	Banded approach	Last year's spending plus an inflation rate, but bound by ranges (e.g. no more than 6% nor less than 3%)
5	Pure asset-based	Pre-specified percentage of moving average of market value – typically 5% of a three year moving average of market values
6	Spending reserve	Segregation of 5% -10% of market value in separate account, invested in 90 day treasury bills. Reserve is drawn down when endowment performance is less than policy target.
7	Hybrid rule (e.g. "Yale rule", "Stanford rule")	"Yale rule": 70% of the allowable spending in the prior fiscal year adjusted for inflation plus 30% of 5% of the endowment's current market value. Spending is a weighted average of a inflation rule and a pure asset-based rule. Typically, the weight placed on the inflation part is between 60 – 70%, and the weight on the asset- based part is therefore 40-30%.

Table 1: Endowment spending approaches

allocation and disclaim responsibility of the day-to-day endowment investment management. Obviously, contract with professionals means additional costs of managing which reduce the endowment's value, especially in difficult years when return may be negative (commissions and management fees are paid in any cases).

2 Method

When considering a judgment such as "what is more important, money that are spent immediately for current student financial aid, research, scholarships and etc. or money that are re-invested to preserve the endowment's value for the future" one of the most significant problem is to define the level of spending. Unfortunately, the spending level for present must be set before the real return is known. The proposed approach based on hybrid rule and can be explained by segregates the endowment into three parts: the original endowment, the stabilization fund and the "service" part for management costs.

The stabilization fund as Mehrling (2005) defined is as a kind of bank account with overdraft privileges, but one that has the same return as the endowment. On average, the account is zero, but when there is a period of abnormally high returns, the account becomes positive, and when there is a period of abnormally low returns, the accounts becomes negative [3]. The "service" part is a sum which includes various commissions and management fees. This amount can be estimated at the beginning of the year.

The model used is of the following form: $SPt = \alpha * ENDOWMENTt - 1 + \beta * STABFUNDt - 1 + SERVt$

The variables are defined as follows: SPt is current spending from endowment;

$ENDOWMENTt - 1$ is an endowment wealth at the end of the previous fiscal period;

$STABFUNDt - 1$ is the stabilization fund balance at the end of the previous fiscal period;

$SERVt$ is the "service" part for current period;

both α and β are spending rates which set up by endowment managers.

As Mehrling remarked [3, p.17]: "...setting α somewhat below the expected real return and β about twice α universities can establish a reasonable balance between spending and real fund growth uncertainty". Commonfund's 2004 Benchmark Study reported results of a survey of 650 senior investment or financial professionals at educational institutions about their objectives in managing endowments. 54% of respondents said their primary concern was to "provide a consistent and growing stream of income". 26% of respondents chose "maximize intergenerational equity" as their top concern and 11% chose "smooth variations" [4].

This model can be modified over time though α and β. The endowment managers, according to their opinion, setting α and β via quantitatively assess the importance of the various criteria such as endowment wealth, the nominal rate of return, endowment management costs, university needs (tuition fees, scholarships, research and grants, etc.), market situation, gifts flow and many others. In unusually favorable financial times, decision-makers may be able to make upward adjustments in endowment spending (for example, $\alpha > 0.05$), while in unusually unfavorable times they may need to estimate $\beta = 0$. So far the objective of the research is not to determine the correct spending rates α and β, but to explore what criteria and success factors are needed to estimate it.

The major research method was the analytical hierarchy process (AHP). The AHP is a methodology for breaking down a complex and unstructured situation into its component parts, then arranging those parts (or variables) into a hierarchical order. The AHP is based on the assignment of numerical values for subjective judgments on the relative importance of each variable, then synthesizing the judgments to determine which variables have the highest priority.

The proposed AHP model incorporated following criteria (Table 2).

Abbreviation	Definition
C1	Return
C2	Endowment size
C3	University needs
C4	Endowment management costs

Table 2: The proposed criteria of the AHP model

Each criterion is described with a certain number of sub-criteria. The quantity of these sub-criteria (within each criterion) does not have to be the same (Table 3). The proposed hierarchy was established by the survey results.

Sub-criteria abbreviation		Definition	Success Factors
C1	SC11	Asset allocation	SF11 Level of risk tolerance SF12 Market situation SF21 Potential corporate donors SF22 Potential private donors SF23 An average donation's amount SF31 Tuition fees SF32 Scholarships SF33 Research SF34 Grants SF35 University business income SF36 Patents
	SC12	Time period of return of invested funds	
	SC13	Inflation	
C2	SC21	Gifts flow	
	SC22	Endowment per student	
	SC23	Quantity of non-restricted endowments	
C3	SC31	Balanced budget	
	SC32	Non-endowment sources of cash-flow	
CC4	SC41	Investment costs	
	SC42	Ancillary services costs	

Table 3: The proposed sub-criteria and success factors

	"α"	"β"	Priority	Rank
"α"	1	6	0,86	1
"β"	1/6	1	0,14	2

Table 4: The results of pair-wise comparisons for the α and β

First, the variables α and β are compared. The importance of α and β is shown in table 4.

In step two, the priority of criteria for α and β are defined. "University needs" was the most important criterion in estimation of α and β. The results of pair-wise comparisons are listed in tables 5,6 and the results of composite weights are shown in table 7.

In next step the priority of sub-criteria are defined. "Asset allocation" was the most important criterion in estimation of "return". The results of pair-wise comparisons are listed in table 8.

What criterion is more important for estimate α?	Return	Endowment size	University needs	End – t Costs	Priority
Return	1	3	1/5	1/5	0,12
Endowment size	1/3	1	1/5	5	0,13
University needs	5	5	1	7	0,64
Endowment management costs	5	1/5	1/7	1	0,11

Table 5: Ranked results of pair-wise comparisons for "α"

What criterion is more important for estimate β?	Return	Endowment size	University needs	End – t Costs	Priority
Return	1		3	5	0,33
Endowment size	2	1	3	5	0,38
University needs	1/3	1/3	1	5	0,23
Endowment management costs	1/5	1/5	1/5	1	0,06

Table 6: Ranked results of pair-wise comparisons for "β"

	"α" (0,86)	"β" (0,14)	Composite Priority	Rank
Return	0,12	0,33	0,15	3
Endowment size	0,13	0,38	0,17	2
University needs	0,64	0,23	0,58	1
Endowment management costs	0,11	0,06	0,10	4

Table 7: The results of composite weights for "α", "β"

3 Conclusion

It is very difficult to make a universal model (a model which could be suitable for a large number of universities). From the AHP – design survey, an analysis of criteria needed to successfully manage university endowments identified and prioritizes. This result could be accepted as a start point in a process of estimate the optimal spending rule.

References

[1] BEAIRD, S.; HAYES, W. Building an endowment: what, why and how, National Catholic Educational association, Washington, StateDC, 1999, p. 83. ISBN 1-555833-222-7

[2] HANSMANN, H. Why universities have endowments? Journal of Legal Studies, vol. XIX, January 1990, p. 3-42

[3] MERTON, R. Optimal investment strategies for university endowment fund, National Bureau of Economic Research, Cambridge, StateMA, 1991, p.44

[4] MEHRLING, P., GOLDSTEIN, P., SEDLACEK, V. Endowment spending: goals, rates and rules, Forum Futures 2005, Exploring the Future of Higher Education

What criterion is more important for estimate "Return"?	SC11	SC12	SC13	Priority	Weighted priority
Asset allocation (SC11)	1	5	3	0,64	0,096
Time period of return of invested funds (SC12)	1/5	1	1/3	0,10	0,015
Inflation (SC13)	1/3	3	1	0,26	0,039

Table 8: The results of pair-wise comparisons for the importance of sub-criteria of "return"

What criterion is more important for estimate "Endowment size"?	SC21	SC22	SC23	Priority	Weighted priority
Gift flow (SC21)	1	1/3	1/5	0,10	0,017
Endowment per student (SC22)	3	1	1/3	0,29	0,049
Quantity of non-restricted endowment (SC23)	5	3	1	0,61	0,103

Table 9: The results of pair-wise comparisons for the importance of sub-criteria of "Endowment size"

What criterion is more important for estimate "University needs"?	SC31	SC32	Priority	Weighted priority
Balanced budget (SC31)	1	5	0,83	0,481
Non-endowment sources of cash-flow (SC32)	1/5	1	0,17	0,098

Table 10: The results of pair-wise comparisons for the importance of sub-criteria of "University needs"

What criterion is more important for estimate "Endowment management costs"?	SC41	SC42	Priority	Weighted priority
Investment costs (SC41)	1	3	0,75	0,075
Ancillary services costs (SC42)	1/3	1	0,25	0,025

Table 11: The results of pair-wise comparisons for the importance of sub-criteria of "University needs"

#	Sub-criteria	#	Sub-criteria
1	Balanced budget	6	Endowment per student
2	Quantity of non-restricted endowments	7	Inflation
3	Non-endowment sources of cash-flow	8	Ancillary services costs
4	Asset allocation	9	Gifts flow
5	Investment costs	10	Time period of return of invested funds

Table 12: The importance of sub-criteria needed for estimate spending rule

Tomáš Tichý

Finance department
VŠB – Faculty of Economics
Technical University Ostrava
Sokolská 33, 701 21 Ostrava

e-mail: tomas.tichy@vsb.cz

Alternative Dependency Measures and a Small Currency Portfolio

Abstract

A standard tool to measure the dependency is the coefficient of linear correlation. Although it has several drawbacks (zero correlation is not equivalent to independency, tail dependency and even asymmetry in the dependency are ignored), it is the most popular vehicle to arrive at optimal portfolio composition or calculate the portfolio risk. In this paper, we review several alternative measures of dependency among random evolution of particular variables, such as Spearman's rho, Kendall's tau, Gini's gamma. We also explain the concept of copula functions. Next, we select representative FX rates and examine their mutual dependency. We provide the matrices of dependency measures and the evolution (stability) in time is studied, too. Subsequently, we formulate several optimization tasks to find an optimal minimum risk portfolio. Alternatively, we set several portfolios with the aim to minimize the Value at Risk for a given probability level. In both issues we combine min-var criterion with alternative dependency measures. It is found that none of the alternative approaches significantly overperforms the standard rho.

Keywords: *Dependency measure, linear correlation, rank measure, copula function, optimal portfolio*

1 Introduction

Many issues of financial modelling assume that a dependency among sources of randomness can be measured and/or modelled. We take into account a dependency in evolution, e.g. within portfolio optimization tasks, under both min-variance and min-VaR (Value at Risk) problems, in pricing of options on baskets of assets, in pricing of credit products, such as CDO's, in risk management problems, such as an economic capital calculation.

Respecting the significance of these issues, it is not clear why the coefficient of linear correlation r is still so much popular, despite its general drawbacks. First, it does not measure dependency coherently. Generally, there exists a plenty of real functions f, for which the pair (x, y), where $y = f(x)$, gives zero correlation. Obviously, it is not equivalent to zero dependency. Such a simple case shows that inadequate application of linear correlation coefficient in e.g. risk management issues can lead to fatal results, since for example a portfolio of some structure products, independent at the first sight, can be, in fact, strongly dependent.

Another problem is that the coefficient of linear correlation works well only for random variables following Gaussian distribution. However, it is well known that prices of financial assets, or to be more correct, the returns of them are far from Gaussianity. That is, the true distribution of returns has fatter tails and is skewed either to the left or right.

In this paper we focus on alternative measures of dependency of financial returns. More particularly, we examine basic dependency measures of several FX rates, each with respect to Czech koruna. Next, the coefficient of linear correlation is red by one of the alternative measures in order to detect optimal portfolios under redefined min-var and min-VaR criterions.

We proceed as follows. In the next section, the most common ways to measure the dependency among random returns are reviewed. Next, in Section3 we describe the data set and provide the matrix of dependency measures over the whole period in study (2000–2007). We also examine selected measures for moving time window. Finally, in Section 4 we formulate min-var and min-VaR tasks as based on alternative dependency measures. We conclude the paper by several observed facts.

2 Measures of dependency

Assuming the well defined probability space, i.e. the triplet of feasible states W, the collection of events A and the probability measure P, we can describe the evolution of n random variables $X_1, X_2, \ldots, X_n$ directly by the joint distribution function:

$$F_{X_1,X,\ldots X_n}\left(x_1, x_2, \ldots, x_n\right) = P\left(X_i \leq x_i \left| i \in \{1,2,\ldots,n\}\right.\right). \tag{1}$$

Hence, it is the probability that i-th random variable X_i will be lower or equal to x_i simultaneously for each i. Generally, for a given $i = j$ it converges to the marginal distribution function as x_i for i different from j approaches to infinity:

$$F_{X_j}\left(x_j\right) = P\left(X_j \leq x_j\right) = \lim_{x_i \to \infty} F_{X_1,X,\ldots X_n}\left(x_1, x_2, \ldots, x_n\right), \qquad i \in \{1,2,\ldots,n\} / \{j\}. \tag{2}$$

A general measure of dependency is the copula function, i.e. the projection of the dependency into the unit interval:

$$C : [0,1]^n \to [0,1] \quad \text{on} \quad R^n,\ n = \{2,3,\ldots\}. \tag{3}$$

Due to a Sklar theorem, we can write:

$$F_{X_1,X,\ldots X_n}\left(x_1, x_2, \ldots, x_n\right) = C\left(F_{X_1}\left(x_1\right), F_{X_2}\left(x_2\right), \ldots, F_{X_n}\left(x_n\right)\right), \tag{4}$$

or inversely:

$$C\left(u_1, u_2, \ldots, u_n\right) = F_{X_1,X,\ldots X_n}\left(F_{X_1}^{-1}\left(u_1\right), F_{X_2}^{-1}\left(u_2\right), \ldots, F_{X_n}^{-1}\left(u_n\right)\right). \tag{5}$$

We can interpret formula (4) such that the information given by the joint distribution function is equivalent to the information given mutually by the copula function and particular marginal distribution function. Hence, the copula function provides us with the form of the dependency among the variables, while the marginal distribution functions give us, clearly, the distribution. In order to asses the dependency, it is sufficient to know (estimate) a proper copula function, but to fill the risk model, we need to know more about the particular distribution.

2.1 Linear correlation

The coefficient of linear correlation measures the dependency between two variables, X_i and X_j. It is commonly denoted by ρ_{ij}. It is defined as follows:

$$\rho_{i,j} = \frac{cov\left(X_i, X_j\right)}{\sqrt{var\left(X_i\right) var\left(X_j\right)}}. \tag{6}$$

Here var x denotes the variance and covar x is the covariance:

$$cov\left(X_i, X_j\right) = E\left[X_i X_j\right] - E\left[X_i\right] E\left[X_j\right]. \tag{7}$$

with E denoting the expected value (mean). When we deal with n random variables, n-dimensional covariance and correlation matrix can be built up.

2.2 Kendall tau

Kendall tau τ_K is a measure of concordance, i.e. it is the probability of concordance minus the probability of disconcordance (a simplified version for two realizations of continuous random variables X and Y):

$$\tau_K = 2P\left(\left(x_1 - x_2\right)\left(y_1 - y_2\right) > 0\right) - 1. \tag{8}$$

Similarly to the coefficient of linear correlation, it takes values on the interval [–1,1]. It is related to a two-dimensional copula function as follows:

$$\tau_K\left(C\right) = 4\int_{[0,1]^2} C\left(u, v\right) dC\left(u, v\right) - 1. \tag{9}$$

2.3 Spearman rho

In order to define the Spearman rho ρ_S we need to assume at least three independent realizations of two random variables:

$$\rho_S\left(X, Y\right) = 3\left(P\left(\left(x_1 - x_2\right)\left(y_1 - y_2\right) > 0\right) - P\left(\left(x_1 - x_2\right)\left(y_1 - y_3\right) < 0\right)\right). \tag{10}$$

This measure of dependency can be regarded as a transition form between the coefficient of correlation and Kendall tau, since it describes the dependency between the marginal distribution functions:

$$\rho_S\left(X, Y\right) = \frac{cov\left(F_X, F_Y\right)}{\sqrt{var\left(F_X\right) var\left(F_Y\right)}}. \tag{11}$$

It is also related to a copula function. For a two-dimensional example we get:

$$\rho_S\left(X, Y\right) = 12\int_{[0,1]^2} uv dC\left(u, v\right) - 3. \tag{12}$$

2.4 Ginni's gamma

Ginni's gamma, γ_G, is the measure of concordance of ranking of two variables:

$$\gamma_G(X,Y) = 2E\left(|F_X + F_Y - 1| - |F_X - F_Y|\right). \tag{13}$$

This can be rewritten in terms of copula function as follows:

$$\gamma_G(X,Y) = 2\int_{[0,1]^2} \left(|u_1 + v - 1| - |u - v|\right) dC(u,v). \tag{14}$$

2.5 Blomqvist beta

Blomqvist beta, β_B, determines the correlation of medians:

$$\begin{aligned} \beta_B(X,Y) &= P\left((X-\tilde{x})(Y-\tilde{y}) > 0\right) - P\left((X-\tilde{x})(Y-\tilde{y}) < 0\right) \\ &= 4H(\tilde{x},\tilde{y}) - 1 \end{aligned} \tag{15}$$

with tilde used to denote the median. This can be rewritten in terms of copula function as follows:

$$\beta_B(X,Y) = 4C\left(\tfrac{1}{2},\tfrac{1}{2}\right) - 1. \tag{16}$$

2.6 Tail dependece

We distinguish lower (left) tail dependency:

$$\lambda_L(X,Y) = \lim_{u\to 0^+} P\left(X < F_X^{-1}(u) \,\middle|\, Y < F_Y^{-1}(u)\right) \tag{17}$$

and upper (right) tail dependency:

$$\lambda_U(X,Y) = \lim_{u\to 1} P\left(X > F_X^{-1}(u) \,\middle|\, Y > F_Y^{-1}(u)\right). \tag{18}$$

Obviously, it stressed the dependency among extreme observations.

1. **Data set**

We consider a data set of daily FX rates for EUR, GBP, HUF, PLN, SKK, and USD with respect to CZK as published by the Czech National Bank, i.e. markets quotes usually taken at 2 p.m. The time series starts on January 1, 2000 and ends on December 31, 2007. Thus, the length of the series of log-returns we handle with is 2014. In Table 1 we present basic statistical parameters – mean, standard deviation, variance, skewness and kurtosis – for daily returns (per annum, if applicable).

In order to simplify the construction of the portfolio and to concentrate our attention on the non-linearity and non-normality, we normalize the vectors of returns to get zero mean and

	EUR	GBP	HUF	PLN	SKK	USD
Mean	0.0378	0.0578	0.0375	0.0189	0.009	0.0849
St.Dev	0.0525	0.0817	0.0784	0.1002	0.058	0.1055
Var	0.0028	0.0067	0.0062	0.01	0.0034	0.0111
Skew	0.2969	0.4115	0.7898	0.5327	0.0646	0.1223
Kurt	7.4324	5.3196	9.7477	12.2096	7.6179	4.0475

Table 1: Statistical parameters for daily returns

unit variance. Clearly, it will have any effect neither on the values of skewness and kurtosis, nor the correlation matrix.

We can see that the mean (drift of the series) varies substantially between –1% (SKK) and –9% (USD). The standard deviation of two FX rates is around 5% (SKK, EUR), another two are close to 8% (GBP, HUF) and the last two are slightly more than 10% (PLN, USD). Except the SKK rate, the skewness is significantly negative, the highest is for HUF (–0.8). By contrast, the highest kurtosis can be observed for the PLN rate (12), while the USD is not very far from the Gaussian. When testing if the distribution can be regarded to be Gaussian, several tests of Jarque-Bera type can be used. Here, the hypothesis of normality must be strongly rejected for all FX rates.

Next, we calculate the matrix of linear correlation coefficients, see Table 2. Inspecting the results, we can see that all correlations are between 0.25 and 0.66. Thus, the dependence is positive and not very high. The highest results give us the pairs (GBP, USD) and (EUR, SKK). Relatively high is also (EUR, GBP). By contrast, the other Central European FX rates (HUF and PLN) exhibit lower correlations.

	EUR	GBP	HUF	PLN	SKK	USD
EUR	1.	0.54	0.45	0.25	0.62	0.41
GBP	0.54	1.	0.26	0.36	0.38	0.66
HUF	0.45	0.26	1.	0.47	0.51	0.18
PLN	0.25	0.36	0.47	1.	0.35	0.38
SKK	0.62	0.38	0.51	0.35	1.	0.22
USD	0.41	0.66	0.18	0.38	0.22	1.

Table 2: Correlation matrix of daily returns

In order to examine the dependency among particular FX rates, we also estimate the matrix for Spearman rho and Kendall tau, see Table 3 bellow. It is obvious, that Kendall tau gives us the lowest levels of dependency – there is no pair of FX rates, for which it is above 0.5.

Since the dependency of returns can be floating in time, we have examined the linear correlation, Spearman rho and also Kendal tau for one year intervals assuming a moving time window, day-by-day. The results are depicted in particular charts in Appendix A. The first conclusion that arises is that none of the measures of dependency is more stable than any

	EUR	GBP	HUF	PLN	SKK	USD
EUR	1.	0.49	0.47	0.2	0.58	0.33
GBP	0.49	1.	0.26	0.32	0.32	0.61
HUF	0.47	0.26	1.	0.45	0.53	0.15
PLN	0.2	0.32	0.45	1.	0.34	0.35
SKK	0.58	0.32	0.53	0.34	1.	0.16
USD	0.33	0.61	0.15	0.35	0.16	1.

	EUR	GBP	HUF	PLN	SKK	USD
EUR	1.	0.35	0.34	0.14	0.43	0.23
GBP	0.35	1.	0.18	0.22	0.22	0.44
HUF	0.34	0.18	1.	0.31	0.38	0.1
PLN	0.14	0.22	0.31	1.	0.24	0.24
SKK	0.43	0.22	0.38	0.24	1.	0.11
USD	0.23	0.44	0.1	0.24	0.11	1.

Table 3: Spearman (left) and Kendall (right) dependency matrix of daily returns

other. The relative distance among them is generally kept at the same level. Next, for several pairs of FX rates, the measures move between zero (no dependency) and one (absolutely positive dependency) for the time period in study. Sometimes they also get slightly negative values.

3 Potential issues

Within many risk management issues it is required to deal with a covariance matrix. However, to model the dependency more properly, we can re the coefficient of linear correlation in

$$cov\left(X,Y\right)=\sqrt{var\left(X\right)}\sqrt{var\left(Y\right)}\rho\left(X,Y\right) \tag{19}$$

by e.g. Spearman rho:

$$cov_S\left(X,Y\right)=\sqrt{var\left(X\right)}\sqrt{var\left(Y\right)}\rho_S\left(X,Y\right), \tag{20}$$

or Kendall tau:

$$cov_K\left(X,Y\right)=\sqrt{var\left(X\right)}\sqrt{var\left(Y\right)}\tau_K\left(X,Y\right). \tag{21}$$

For a risk averse investor it can be optimal to minimize the risk of the portfolio (**Task 1**):

$$\text{var}\left(\Pi\right)\rightarrow\min,\quad \text{var}\left(\Pi\right)=x'\cdot C\cdot x. \tag{a}$$

$$s.t.\quad x'1=1, \tag{b}$$

where C is the covariance matrix due to (21) and vector x specifies the fractions of the overall wealth invested into particular assets, that is weights (the composition of this vector should be subject to additional constraints according to a feasibility of particular positions). Moreover, benchmark return can be specified, which gives us

$$s.t.\quad x'\mu=r_P, \tag{c}$$

with vector of expected returns of particular assets μ.

However, we can set up the covariance matrix either due to (20), C_S, or (21), C_K, and potentially add additional constraints on the return of the portfolio, i.e. to provide the same return, r, that might be obtained when Task 1 was followed, Task 1a:

$$var\left(\Pi\right)\rightarrow\min,\quad var\left(\Pi\right)=x'\cdot C_S\cdot x. \tag{a}$$

$$s.t. \quad x'1 = 1, \tag{b}$$
$$s.t. \quad x'\mu = r. \tag{c}$$

We can get similar specification of the task for C_K. In this way we restructure the portfolio to respect the concordance measures.

Another popular problem is to manage the left tail risk using VaR measure, i.e. minimize the loss which can be incurred with a given probability α (**Task 2**):

$$VaR(\Pi, \alpha) \to \min, \qquad \begin{array}{l} VaR(\Pi, \alpha) = -E[r_P] - \sigma_P \cdot \Phi^{-1}(\alpha) \\ \sigma_P = \sqrt{x' \cdot C \cdot x} \end{array} \tag{a}$$
$$s.t. \quad x'1 = 1. \tag{b}$$

Here, Φ^{-1} denotes the inverse to the probability distribution function of the portfolio. Also Task 2 can be reconsidered to incorporate the other dependency measures, i.e. in order to calculate the standard deviation of the portfolio return we replace C by either C_S or C_K.

4 Conclusions

A dependency among particular risk factors is a crucial input factor into many risk management models. In this paper we have briefly reviewed the most important alternatives to the coefficient of linear correlation. In particular, two concordance measures – Spearman rho and Kendall tau – were calculated and analyzed for time series of six distinct FX rates, each with respect to Czech koruna. Finally, two classical optimization problems (min-var and min-VaR) were redefined to incorporate alternative dependency measures.[1]

These redefined tasks of financial optimization can be used e.g. to test if there is some space for improvement in financial risk measuring approaches. Since the alternative measures of dependency take generally lower values under the data considered here, the redefined VaR should be lower too. Notwithstanding, when the number of failures of the model were examined, we received no significant differences when compared to the standard approach.

[1] CHERUBINI, G., LUCIANO, E., VECCHIATO, W. *Copula Methods in Finance.* Wiley 2004.

[2] JONDEAU, E., POON, S.-H., ROCKINGER, M. *Financial Modeling Under Non-Gaussian Distributions*, Springer Verlag, 2006.

[3] MEUCCI, A. *Risk and Asset Allocation.* Springer-Verlag, 2005.

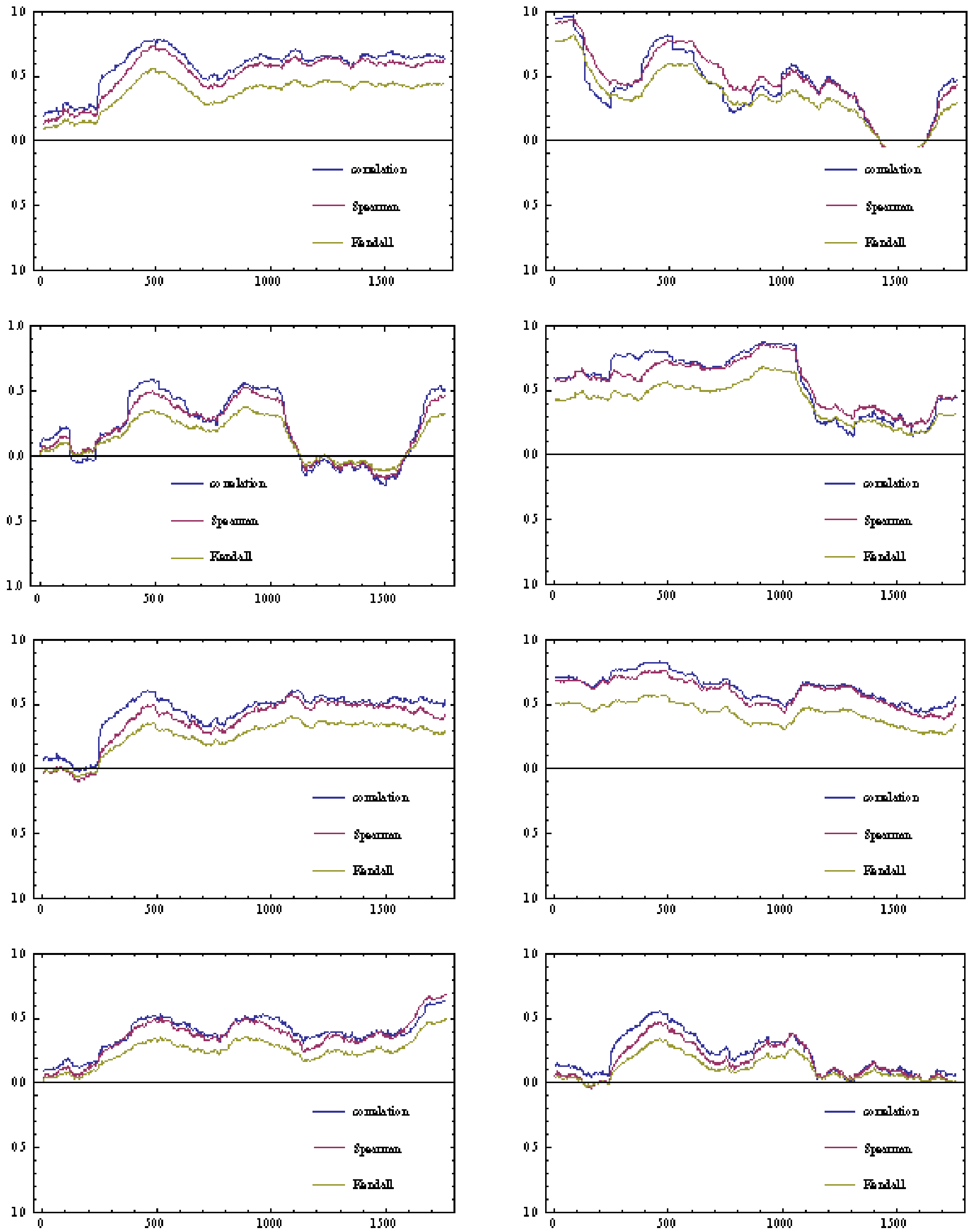
[4] NELSEN, R.B. *An Introduction to Copulas.* 2nd ed. Springer, 2006.

[5] TSAY, R.S. *Analysis of Financial Time Series*, 2nd Edition. John Wiley&Sons, 2005.

[6] TICHÝ, T. Modeling of FX sensitive portfolio, In: *Future of the Banking after the Year 2000 in the World and in the Czech Republic*, Slezská Univerzita, Karviná, 2007.

[1] Due to a limited number of pages, particular results will be available on the author's webpages.

Appendix A Evolution of dependency measures in time – daily returns over one year, daily moving time window, from left to right and from top to bottom: (EUR, GBP), (EUR, HUF), (EUR, PLN), (EUR, SKK), (EUR, USD), (GBP, USD), (PLN, SKK), and (SKK, USD)



Hana Tomášková

Faculty of Informatics and Management
University of Hradec Králové
Rokitanského 62, 50003 Hradec Králové

e-mail: tomahana@atlas.cz

Eigenvalue Problem for Parametrized Circulant Matrices

Abstract

The eigenproblem for parametrized matrices in max-plus algebra was recently investigated by J. Plavka. Three types of matrices were considered in the parametrized eigenproblem: with 1 parameter in the first column, with the same parameter in the first k columns, and with k different parameters in the first k columns.

In the contribution, the parametrized eigenproblem for circulant matrices is studied in the general case. Arbitrary parameters at any diagonals of the matrix are allowed. Parametrized formula for the eigenvalue is presented and conditions are given for using particular parameter values in the formula. On the other hand, it has been shown that the eigenvectors of a parametrized circulant matrix can have a rather complex form even for matrix with one parameter.

1 Introduction

In many applications the models use operations of maximum and minimum together with further arithmetical operations. The so-called max-plus algebra is useful for investigation of discrete events systems and the sequence of states in discrete time corresponds to powers of matrices in max-plus algebra. A typical application of discrete events systems are production lines, where every machine must wait with starting a new operation until the operations on other machines are completed. The eigenproblem for max-plus matrices describes the steady state of the system and, therefore, it has been intensively studied by many authors, see e.g. [1, 2, 3, 7, 9, 10]. The theoretical results were found and polynomial algorithms were proposed for computing the eigenvalue and eigenvectors of a max-plus matrix.

For special types of matrices such as circulant or Monge matrix, the computation can often be performed in the simpler way than in the general case, hence the investigation of special cases is important from the computational point of view. In real situations the exact values of the inputs in the model are not known. Small changes of input values can influence the eigenvalue or eigenvectors. This observation leads to parametrized eigenproblem which was recently investigated in [4, 5, 6, 8].

In this paper the eigenproblem for max-plus matrices with parameters is studied, when the input matrix is circulant. The eigenvalues of circulant parametrized matrix are described in section 5. It is show in the last section that the computation of parametrized eigenvectors is a rather difficult problem, even if the input matrix is circulant.

2 Max-plus algebra

By a max-plus algebra we understand a triple $(R, \oplus, \otimes)$, where R is the set of all real numbers and $\oplus, \otimes$ are binary operations on R defined as

$$a \oplus b = \max(a, b), \quad a \otimes b = a + b$$

for all $a, b \in R$. The two operations are extended to matrices and vectors in the same way as in conventional linear algebra.

For matrices $A, B \in R^{n \times n}$ the matrix product $A \otimes B = C$ is defined by formula

$$c_{ij} = \bigoplus_{k=1}^{n} a_{ik} \otimes b_{kj} = \max_{k=1,\ldots,n} (a_{ik} + b_{kj})$$

for $i, j \in n$.

The kth power of $A \in R^{n \times n}$ is denoted by $A^{(k)}$ and defined by recursion on $k = 2, 3, \ldots$

$$A^{(k)} = A \otimes A^{(k-1)}.$$

3 Eigenproblem

The *eigenproblem* in max-algebra is formulated as follows. Given $A \in R^{n \times n}$, find $x \in R^n$ and $\lambda \in R$ satisfying

$$A \otimes x = \lambda \otimes x.$$

The value λ and the vector x fulfilling the above equation are called the eigenvalue and the eigenvector of the matrix A.

This problem was treated by several authors during the sixties, c.g. [1, 2], survey of the results concerning this and similar eigenproblems can be found in [9, 10].

In [7], the following notation was introduced. Let $N = \{1, 2, \ldots, n\}$ and let C_n be the set of all cyclic permutations defined on nonempty subsets of N. For a cyclic permutation $\sigma = (i_1, i_2, \ldots, i_l) \in C_n$ and for $A \in R^{n \times n}$, we denote l, the length of σ by $l(\sigma)$ and define the weight $w_A(\sigma)$ and the mean weight $\mu_A(\sigma)^{1/l(\sigma)}$ of σ

$$w_A(\sigma) = a_{i_1 i_2} \otimes a_{i_2 i_3} \otimes \cdots \otimes a_{i_l i_1}, \quad \mu_A(\sigma)^{1/l(\sigma)}, \quad \lambda(A) = \bigoplus_{\sigma \in C_n} \mu_A(\sigma).$$

The power of μ_A is computed in the sense of max-plus algebra, and it is equal to multiplication by $1/l(\sigma)$ in the standard notation. The following lemma has been proved in [1].

Lemma 1. *Let $A \in R^{n \times n}$. Then $\lambda(A)$ is the unique eigenvalue of A.*

In other words, the eigenvalue $\lambda(A)$ is equal to the maximal mean weight of a cycle in the corresponding digraph $G(A)$.

4 Circulant matrices

In this section, we will consider special form of matrices in max-plus algebra, called circulant matrices. Matrix A of type $n \times n$ is circulant if

$$a_{ij} \;=\; a_{i'j'}$$

whenever

$$i - i' \;=\; j - j' \pmod{n}.$$

Hence, matrix A is fully determined by its inputs in the first row, denoted as $a_1, a_2, \ldots, a_n$. (a_1 is the common value of all diagonal inputs, and similarly, each a_i is the common value of all inputs on a line parallel to the matrix diagonal).

$$\begin{pmatrix} a_1 & a_2 & a_3 & \ldots & a_n \\ a_n & a_1 & a_2 & \ldots & a_{n-1} \\ a_{n-1} & a_n & a_1 & \ldots & a_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_2 & a_3 & a_4 & \ldots & a_1 \end{pmatrix}$$

The following two lemmas have been proved in [7].

Lemma 2. *If A and B are circulant matrix the same type of $n \times n$, then $A \otimes B$ is also the circulant matrix type of $n \times n$.*

Lemma 3. *For a circulant matrix A the eigenvalue $\lambda(A)$ is equal to the maximal value of a_i, for i from 1 to n.*

As a consequence of Lemma 3, the computation of the eigenvalue of a circulant matrix $A \in R^{n\times n}$ has the complexity $O(n)$, in contrast with the complexity $O(n^2)$ in general case [1].

5 Eigenvalue of parameterized circulant matrices

The parametrized eigenproblem for different types of parameterized matrices was investigated in [4, 5, 6, 8].

In this paper we considered three types of parametrized matrices: a matrix with one parameter p in the first diagonal, the k identical parameters on some diagonals and k arbitrary parameters on diagonals.

In the first type of parameterized circulant matrix the parameter p is added to all inputs a_1 on the main diagonal.

$$A = \begin{pmatrix} a_1 + p & a_2 & a_3 & \ldots & a_n \\ a_n & a_1 + p & a_2 & \ldots & a_{n-1} \\ a_{n-1} & a_n & a_1 + p & \ldots & a_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_2 & a_3 & a_4 & \ldots & a_1 + p \end{pmatrix}$$

Theorem 6. *If A is a circulant matrix with one parameter p added in first diagonal and if we denote $a_m = \max\{a_i;\ i \in N - \{1\}\}$ then*

$$\lambda(A) = \begin{cases} a_m & \text{for } p < a_m - a_1 \\ a_1 + p & \text{for } p \geq a_m - a_1. \end{cases} \tag{1}$$

Proof. The assertion (1) follows directly from Lemma 3. $\square$

Note 1. *For the matrix with one parameter p added in the rth diagonal we denote $a_m = \max\{a_i;\ i \in N - \{r\}\}$ and we get analogous formula*

$$\lambda(A) = \begin{cases} a_m & \text{for } p < a_m - a_r \\ a_r + p & \text{for } p \geq a_m - a_r. \end{cases} \tag{2}$$

Example 4. *We consider a matrix with one parameter in the first diagonal*

$$A = \begin{pmatrix} 3+p & 2 & 5 \\ 5 & 3+p & 2 \\ 2 & 5 & 3+p \end{pmatrix}.$$

In this example $a_m = \max\{a_i; i \in N - \{1\}\} = \max\{5, 2\} = 5$, $a_m - a_1 = 5 - 3 = 2$

$$\lambda(A) = \begin{cases} 5 & \text{for } p < 5 - 3 \\ 3 + p & \text{for } p \geq 5 - 3. \end{cases} \tag{3}$$

The values of λ it dependence on parameter p are shown on Figure 1.

Theorem 7. *If A is a circulant matrix with the k identical parameters added in diagonals whose indices belong to a set $K \subseteq N$, $|K| = k$, we denote $a_m = \max\{a_i;\ i \in N - K\}$ and $a_r = \max\{a_i;\ i \in K\}$ and we get analogous formula*

$$\lambda(A) = \begin{cases} a_m & \text{for } p < a_m - a_r \\ a_r + p & \text{for } p \geq a_m - a_r. \end{cases} \tag{4}$$

Proof. The assertion follows from Lemma 3. $\square$

Theorem 8. *If A is a circulant matrix with the k arbitrary parameters added in diagonals whose indices belong to a set $K \subseteq N$, $|K| = k$, we denote $a_m = \max\{a_i;\ i \in N - K\}$ and for $g \in K$ we get*

$$\lambda(A) = \begin{cases} a_m & \text{if } (\forall h \in K)\ p_h < a_m - a_h \\ a_g + p_g & \text{if } p_g \geq a_m - a_g \text{ and } (\forall h \in K - \{g\})\ p_g \geq p_h + a_h - a_g. \end{cases} \tag{5}$$

Proof. The assertion follows from Lemma 3. $\square$

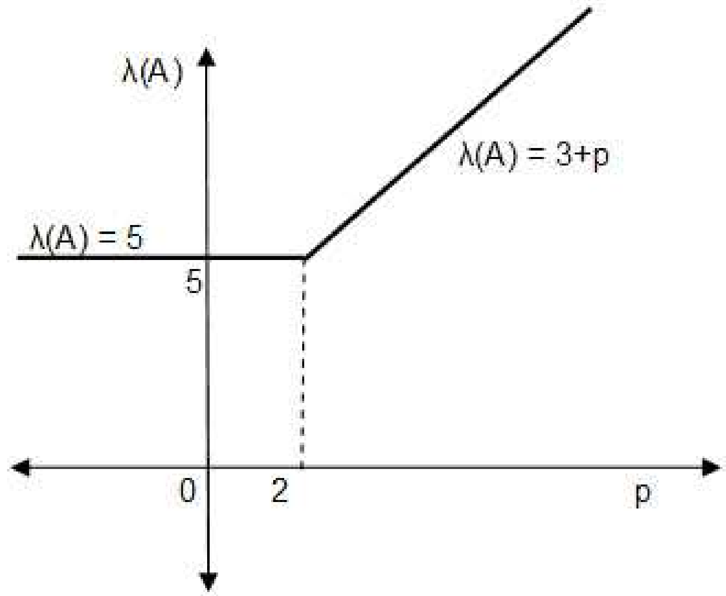


Figure 1: Eigenvalue of matrix with one parameter

Example 5. *We considere the matrix A with different parameters p, q*

$$A = \begin{pmatrix} 3+p & 2+q & 5 \\ 5 & 3+p & 2+q \\ 2+q & 5 & 3+p \end{pmatrix}.$$

In this example $K \subseteq N$, $|K| = 2$, we denote $a_m = \max\big\{a_i;\, i \in N - \{2,3\}\big\} = \max\{5\} = 5$ and for $p, q \in K$ we get

$$\lambda(A) = \begin{cases} 5 & \text{if } p < 5-3 \text{ and } q < 5-2 \\ 3+p & \text{if } p \geq 5-3 \text{ and } p \geq g+2-3 \\ 2+q & \text{if } q \geq 5-2 \text{ and } q \geq p+3-2. \end{cases} \tag{6}$$

The values of λ it dependence on two parameters p and q are shown on Figure 2.

6 Eigenvector of parameterized circulant matrices

In the general case the computation of the eigenvectors of a given matrix A requires computation of powers $A, A^{(2)}, A^{(3)}, \ldots, A^{(n)}$. If A is circulant, the powers are circulant too, but still the computation of the eigenvectors of parameterized circulant matrices is more difficult then computing the eigenvalue. Even in the case of one parameter, the eigenvector has different

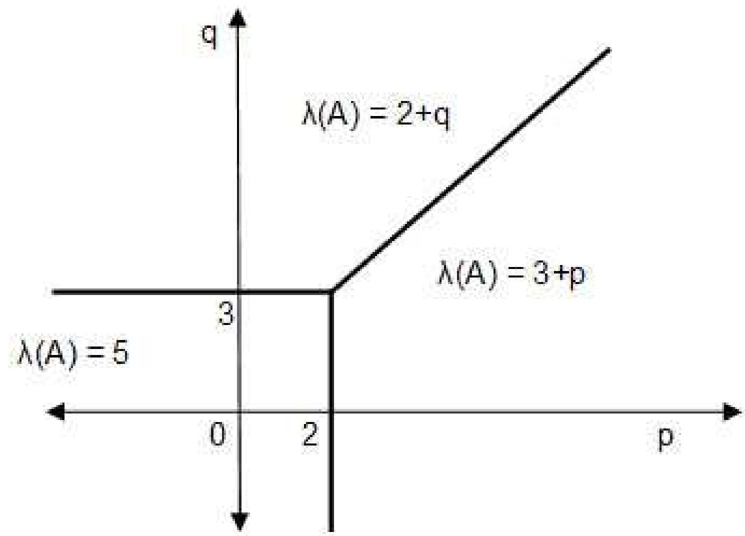


Figure 2: Eigenvalue of parametrized matrix

forms on more than two intervals, differently from the eigenvalue. Next example indicates the complexity of the problem.

Example 6. *We consider the matrix from Example 4 with one parameter in the first diagonal*

$$A = \begin{pmatrix} 3+p & 2 & 5 \\ 5 & 3+p & 2 \\ 2 & 5 & 3+p \end{pmatrix}.$$

We have seen in Example 4 that the eigenvalue of A has two different forms: in the interval $(-\infty, 2\rangle$ the eigenvalue is equal to 5, and in the interval $\langle 2, \infty)$ the eigenvalue is $3+p$.

In this example, during the computation of inputs of the power matrix $A^{(2)}$ we shall find as much as 3 new critical values of parameter p. For the computation we use all rows of matrix A but only the first column of A. We know from the Lemma 2 that the result will be a circulant matrix, hence it is sufficient to compute only the first column.

$$\begin{pmatrix} 3+p & 2 & 5 \\ 5 & 3+p & 2 \\ 2 & 5 & 3+p \end{pmatrix} \otimes \begin{pmatrix} 3+p \\ 5 \\ 2 \end{pmatrix}$$

$$a_{11} = \max\{6+2p; 7; 7\} = \begin{cases} 6+2p < 7 & \text{if } p < 1/2 \\ 6+2p \geq 7 & \text{if } p \geq 1/2 \end{cases}$$

$$a_{21} = \max\{8+p; 8+p; 4\} = \begin{cases} 8+p; < 4 & \text{if } p < -4 \\ 8+p; \geq 4 & \text{if } p \geq -4 \end{cases}$$

$$a_{31} = \max\{5+p; 10; 5+p\} = \begin{cases} 5+p < 10 & \text{if } p < 5 \\ 5+p \geq 10 & \text{if } p \geq 5 \end{cases}$$

$$A_1^{(2)} = \left| \begin{array}{c|c|c|c|c} (-\infty, -4\rangle & \langle -4, 1/2\rangle & \langle 1/2, 2\rangle & \langle 2, 5\rangle & \langle 5, \infty\ ) \\ 7 & 7 & 6+2p & 6+2p & 6+2p \\ 4 & 8+p & 8+p & 8+p & 8+p \\ 10 & 10 & 10 & 10 & 5+p \end{array} \right| .$$

In addition to the original critical value 2 the formula for $A_1^{(2)}$ continues new critical values $-4, 1/2, 5$.

Similarly, we compute $A_1^{(3)}$

$$A_1^{(3)} = \left| \begin{array}{c|c|c|c|c} (-\infty, -4\rangle & \langle -4, 1/2\rangle & \langle 1/2, 2\rangle & \langle 2, 5\rangle & \langle 5, \infty\ ) \\ 15 & 15 & 15 & 9+3p & 9+3p \\ 12 & 12 & 11+2p & 11+2p & 11+2p \\ 9 & 13+p & 13+p & 13+p & 8+2p \end{array} \right| .$$

In the computation of $A_1^{(3)}$ no new critical values have appeared, but this can happen for different input matrix.

References

[1] R. A. Cuninghame-Green, *Minimax algebra*, Lecture Notes in Econ. and Math. Systems 166, Springer-Verlag, Berlin, 1979.

[2] 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.

[3] M. Gavalec, J. Plavka, *Eigenproblem in extremal algebras*, Proc. of the 9th Int. Symp. Oper. Res. '07, Nova Gorica, Slovenia (2007), 15–21.

[4] M. Gavalec, J. Plavka, *Fast algorithm for extremal biparametric eigenproblem*, Acta Electrotechnica et Informatica **7** (2007), 23–27.

[5] M. Gavalec, J. Plavka, *Biparametric eigenproblem in max-plus algebra*, Abstracts of the 11th Inter. Conf. on Oper. Res. KOI 2006, Pula (Croatia) (2006), p. 13.

[6] M. Gavalec, J. Plavka, *Multiparametric eigenproblem in max-plus algebra*, Abstracts of the 22nd Europ. Conf. on Oper. Res. EURO 2007, Prague (2007), p. 201.

[7] J. Plavka, *Eigenproblem for circulant matrices in max-algebra*, Optimization 50 (2001), 477-483.

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

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

[10] U. Zimmermann, *Linear and Combinatorial Optimization in Ordered Algebraic Structure*, Ann. Discrete Math., No 10, North Holland, Amsterdam, 1981.

Jaromír Tonner, Osvald Vašíček[1]

Department of Applied Mathematics and Computer Science
Faculty of Economics and Administration of Masaryk University
Lipová 41a, 602 00 Brno

e-mail: osvald@econ.muni.cz, jtonner@tiscali.cz

States Estimation in Baseline New Keynesian DSGE Model: Kalman or Bootstrap Filter?

Abstract

The aim of this article is to compare the ability of estimation and filtering methods. Under assumptions of normality and linear structure, Kalman filter seems to be more accurate in comparison with Bootstrap filter that cooperates with empirical distributions. But we can show that 'chances' are approximately similar if we set for Bootstrap filter the same variances of structural shocks as we obtain from Kalman smoothed structural shocks. A construction of Kalman filter enables to flexibly change estimated covariance matrix of observation vector estimation error during filtering and smoothing procedure. But this 'advantage' is not implemented in Bootstrap filter since the bootstrap filter measurement error variances and structural shocks variances must be constant during filtering and smoothing procedure. The results will be shown on the DSGE baseline new Keynesian model of Czech economy. The model will be filtered and smoothed by above mentioned filters and a comparison study will be carried out. Practical experiences will be used for time - varying parameters estimation of DSGE models by Bootstrap filter since time - varying models are nowadays modern and very popular tool for analyzing changes in economic environment.

1 System

The solution of DSGE models is usually written in the state-space form

$$\begin{aligned} x_{t+1} &= A_t x_t + B_t u_t + R_t v_t \qquad v_t \sim N(0,Q) \\ y_t &= C_t x_t + D_t u_t + e_t \qquad e_t \sim N(0,H) \end{aligned}$$

where $x_0 \sim N(\mu_0, \Sigma_0)$ is known and v_t and e_t are independent of past and current states, inputs and outputs. Let the data set $D_t = \{y_0, u_0, \ldots, y_t, u_t\}$ is known and denote $x_{t|k} = x_t | D_k$. N denotes number of observations.

[1]This work was supported by MŠMT project Research centers 1M0524, and funding of specific research at ESF MU.

In general predicted, filtered and smoothed estimates of x_t are given by

$$\begin{aligned}
p(x_t|D_t) &= \frac{p(y_t|x_t, u_t)}{p(y_t|D_{t-1}, u_t)} p(x_t|D_{t-1}) \\
p(x_{t+1}|D_t) &= \int p(x_{t+1}|x_t, u_t) p(x_t|D_t) dx_t \\
p(x_t|D_N) &= p(x_t|D_t) \int p(x_{t+1}|x_t, u_t) \frac{p(x_{t+1}|D_N)}{p(x_{t+1}|D_t)} dx_{t+1}.
\end{aligned}$$

1.1 Bootstrap filter

Conditional probability density function (CPDF) can be represented by an M-dimensional sample set $x^{(i)}$. Suppose we know the prior empirical CPDF of the initial state $p_M(x_{0|-1}) = \sum_{i=1}^{M} w_{0|-1}^{(i)} \delta(x - x_{0|-1}^{(i)})$. Then the empirical densities of $x_{t|t-1}, x_{t|t}, x_{t|M}$, or more simply of $x_{t|t'}, t' = t-1, t, M$

$$p_M(x_{t|t'}) = \sum_{i=1}^{M} w_{t|t'}^{i} \delta(x - x_{t|t'}^{i})$$

are given by Bootstrap algorithm, if

$$\begin{aligned}
x_{t|t}^{i} &= x_{t|t-1}^{i} \\
w_{t|t}^{i} &= \frac{p(y_t|x_{t|t-1}^{i}, u_t) w_{t|t-1}^{i}}{\sum_{j=1}^{M} p(y_t|x_{t|t-1}^{j}, u_t) w_{t|t-1}^{j}} \\
x_{t+1|t}^{i} &= A_t x_{t|t}^{i} + B_t u_t + R_t v_t^{i} \\
w_{t+1|t}^{i} &= w_{t|t}^{i} \\
x_{t|N}^{i} &= x_{t|t}^{i} \\
w_{t|N}^{i} &= \frac{w_{t|t}^{i} \sum_{j=1}^{M} w_{t+1|N}^{j} p(x_{t+1|N}^{j}|x_{t|N}^{i}, u_t)}{\sum_{k=1}^{M} w_{t|t}^{k} \sum_{j=1}^{M} w_{t+1|N}^{j} p(x_{t+1|N}^{j}|x_{t|N}^{k}, u_t)}.
\end{aligned}$$

The predicted, filtered and smoothed estimates of the state can be obtained as a weighted sample mean

$$\mu_{t|t'} = \sum_{i=1}^{M} w_{t|t'}^{i} x_t^{i}.$$

1.2 Kalman filter

The predicted, filtered and smoothed estimates

$$x_{t|t'} \sim N(\mu_{t|t'}, \Sigma_{t|t'})$$

are given by

$$
\begin{aligned}
\mu_{t|t} &= \mu_{t|t-1} + K_t(y_t - C_t\mu_{t|t-1} - D_t u_t) \\
\Sigma_{t|t} &= \Sigma_{t|t-1} - K_t C_t \Sigma_{t|t-1} \\
\mu_{t+1|t} &= A_t\mu_{t|t} + B_t u_t \\
\Sigma_{t+1|t} &= A_t\Sigma_{t|t}A_t' + R_t Q R_t' \\
K_t &= \Sigma_{t|t-1}C_t'(C_t\Sigma_{t|t-1}C_t' + H)^{-1} \\
\mu_{t|N} &= \mu_{t|t} + F_t(\mu_{t+1|N} - \mu_{t+1|t}) \\
\Sigma_{t|N} &= \Sigma_{t|t} + F_t(\Sigma_{t+1|t} - \Sigma_{t+1|N})F_t' \\
F_t &= \Sigma_{t|t}A_t'\Sigma_{t+1|t}^{-1}.
\end{aligned}
$$

A more detailed description of relations in section 1 can be seen in [5] and [8].

2 Model

Household needs to max $\mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t d_t \left\{ \log(c_{jt} - hc_{jt-1}) + v\log(\frac{m_{jt}}{p_t}) - \varphi_t \psi \frac{l_{jt}^{1+\gamma}}{1+\gamma} \right\}$ under five conditions:

- $\log d_t = \rho_d \log d_{t-1} + \sigma_d \epsilon_{d,t}$ where $\epsilon_{d,t} \sim \mathrm{N}(0,1)$,
- $\log \varphi_t = \rho_\varphi \log \varphi_{t-1} + \sigma_\varphi \epsilon_{\varphi,t}$ where $\epsilon_{\varphi,t} \sim \mathrm{N}(0,1)$,
- $c_{jt} + x_{jt} + \frac{m_{jt}}{p_t} + \frac{b_{jt+1}}{p_t} + \int q_{jt+1,t} a_{jt+1} d\omega_{j,t+1,t} = w_{jt} l_{jt} + (r_t u_{jt} - \mu_t^{-1}\Phi[u_{jt}])k_{jt-1} + \frac{m_{jt-1}}{p_t} + R_{t-1}\frac{b_{jt}}{p_t} + a_{jt} + T_t + F_t$,
- $k_{jt} = (1-\delta)k_{jt-1} + \mu_t \left(1 - V\left[\frac{x_{jt}}{x_{jt-1}}\right]\right) x_{jt}$,
- $\mu_t = \mu_{t-1} e^{\Lambda_\mu + z_{\mu,t}}$ where $z_{\mu,t} = \sigma_\mu \epsilon_{\mu,t}$ and $\epsilon_{\mu,t} \sim \mathrm{N}(0,1)$

where d_t is a shock to intertemporal preference, c_{jt} consumption of the j-th household, $\frac{m_{jt}}{p_t}$ real money of the j-th household, l_{jt} hours worked of the j-th household, φ_t a labor supply shock, x_{jt} investments of the j-th household, b_{jt} amount of goverment bounds hold by j-th household, a_{jt+1} the amount of those securities that pay one unit of consumption in event $\omega_{j,t+1,t}$ purchased by household j at time t at real price $q_{jt+1,t}$, w_{jt} the real wage of the j-th household, r_t the real rental price of capital, u_t the intensity of use of capital, $(\mu_t)^{-1}\Phi[u_{jt}]$ the physical cost of u_{jt} in resource terms, R_t nominal interest rate, T_t a lump - sum transfer, F_t the profits of the firms in the economy, k_{jt} capital of the j-th houshold, μ_t an investment specific technological shock, $V[\cdot]$ a quadratic adjustment cost function in a law of motion for capital, $z_{\mu,t}$ a shock term in investment specific technological shock.

Labour supplier problem is to $\max_{l_{jt}} w_t l_t^d - \int_0^1 w_{jt} l_{jt} dj$ under condition $l_t^d = \left(\int_0^1 l_{jt}^{\frac{\eta-1}{\eta}} dj \right)^{\frac{\eta}{\eta-1}}$. By the zero profit condition, we get $l_{jt} = \left(\frac{w_{jt}}{w_t}\right)^{-\eta} l_t^d$ and finally $l_t = v_t^w l_t^d$ where w_t is a market real wage, l_t^d the aggregate labor demand, l_{jt} labor demand of the j-th household, l_t the aggregate labor supply of households, v_t^w the aggregate loss of efficiency induced by wage dispersion.

Calvo's wages setting means to adjust the relevant part of the lagrangian for the household: $\max_{w_{jt}} \mathrm{E}_t \sum_{\tau=0}^{\infty}(\beta\theta_w)^\tau \left\{-d_t\varphi_t\psi\frac{l_{jt+\tau}^{1+\gamma}}{1+\gamma} + \lambda_{jt+\tau}\prod_{s=1}^{\tau}\frac{\Pi_{t+s-1}^{\chi w}}{\Pi_{t+s}}w_{jt}l_{jt+\tau}\right\}$ subject to: $l_{jt+\tau} = \left(\prod_{s=1}^{\tau}\frac{\Pi_{t+s-1}^{\chi w}}{\Pi_{t+s}}\frac{w_{jt}}{w_{t+\tau}}\right)^{-\eta} l_{t+\tau}^d$. Because of complete markets, we can drop jth index. So the real price index evolves $1 = \theta_w\left(\frac{\Pi_{t-1}^{\chi w}}{\Pi_t}\right)^{1-\eta}\left(\frac{\frac{w_{t-1}}{z_{t-1}}}{\frac{w_t}{z_t}}\frac{z_{t-1}}{z_t}\right)^{1-\eta} + (1-\theta_w)(\Pi_t^{w*})^{1-\eta}$ where λ_{jt} is the lagrangian multiplier associated with the budget constraint, Π_t inflation rate, Π_t^{w*} equilibrium wage inflation rate, z_t the fixed cost of production.

The final goods producer must $\max_{y_{it}} p_t y_t^d - \int_0^1 p_{it}y_{it}di$ under condition $y_t^d = \left(\int_0^1 y_{it}^{\frac{\epsilon-1}{\epsilon}}di\right)^{\frac{\epsilon}{\epsilon-1}}$. By the zero profit condition, we get $y_{it} = \left(\frac{p_{it}}{p_t}\right)^{-\epsilon} y_t^d$
where p_t is the final good price, y_t^d the aggregate demand of final goods, p_{it} the price of goods produced by the i-th firm, y_{it} amount of intermediate goods produced by the i-th firm.

Intermediate good producers two-stages problem is to $\min_{l_{it}^d, k_{it-1}} w_t l_{it}^d + r_t k_{it-1}$ subject to their supply curve:

- $y_{it} = A_t k_{it-1}^{\alpha}(l_{it}^d)^{1-\alpha} - \phi z_t$ with
- $A_t = A_{t-1}e^{\Lambda_A + z_{A,t}}$ where $z_{A,t} = \sigma_A\epsilon_{A,t}$ and $\epsilon_{A,t}$ ~N$(0,1)$,
- $z_t = z_{t-1}e^{\Lambda_z + z_{z,t}}$ where $z_{z,t} = \frac{z_{A,t}+\alpha z_{\mu,t}}{1-\alpha}$, $\Lambda_{z,t} = \frac{\Lambda_{A,t}+\alpha\Lambda_{\mu,t}}{1-\alpha}$.

where y_{it} is a production of ith firm at time t, A_t an aggregate technological progress, k_{it-1} the capital rented by the firm, l_{it}^d the amount of the 'packed' labor imput rented by the firm, $z_{A,t}$ a shock term in technological progress, $z_{z,t}$ a shock term in law of the fixed cost of production. So the firms rent inputs to satisfy their static minimization problem:

- $\frac{u_t k_{t-1}}{l_t^d} = \frac{\alpha}{1-\alpha}\frac{w_t}{r_t}$,
- $mc_t = \left(\frac{1}{1-\alpha}\right)^{1-\alpha}\left(\frac{1}{\alpha}\right)^{\alpha}\frac{w_t^{1-\alpha}r_t^{\alpha}}{A_t}$.

where mc_t a marginal cost.

Calvo's prices are given by $\max_{p_{it}} \mathrm{E}_t\sum_{\tau=0}^{\infty}(\beta\theta_p)^\tau\frac{\lambda_{t+\tau}}{\lambda_t}\left\{\left(\prod_{s=1}^{\tau}\Pi_{t+s-1}^{\chi}\frac{p_{it}}{p_{t+\tau}} - mc_{t+\tau}\right)y_{it+\tau}\right\}$ subject to $y_{it+\tau} = \left(\prod_{s=1}^{\tau}\Pi_{t+s-1}^{\chi}\frac{p_{it}}{p_{t+\tau}}\right)^{-\epsilon} y_{t+\tau}^d$. Because of complete markets, we can drop ith index. Consequently, the real wage index evolves: $1 = \theta_p\left(\frac{\Pi_{t-1}^{\chi}}{\Pi_t}\right)^{1-\epsilon} + (1-\theta_p)\Pi_t^{*1-\epsilon}$ where Π_t is the inflation rate, Π_t^* the equilibrium inflation rate.

Markets clear if $y_t^d = \frac{A_t(u_t k_{t-1})^{\alpha}(l_t^d)^{1-\alpha} - \phi z_t}{v_t^p}$ and $y_t^d = c_t + x_t + \mu^{-1} a[u_t] k_{t-1}$ where v_t^p the aggregate loss of efficiency induced by price dispersion.

Government problem is $\frac{R_t}{R} = \left(\frac{R_{t-1}}{R}\right)^{\gamma_R} \left(\left(\frac{\Pi_t}{\Pi}\right)^{\gamma_\Pi} \left(\frac{\frac{y_t^d}{y_{t-1}^d}}{\Lambda_{y^d}} \right) \right)^{1-\gamma_R} e^{m_t}$

with $m_t = \sigma_m \epsilon_{m,t}$ where $\epsilon_{m,t} \sim$N$(0,1)$.
R_t denotes the nominal interest rate, R denotes the steady-state gross return of capital, Π the inflation target, Λ_{y^d} the steady state gross growth rate of y, m_t a random shock to monetary policy.
A description of the fully derived model can be found in [1].

3 Results

Section 2 describes a model solving procedure which embodies in optimization where the representative consumer/firm chooses optimal sequence of variables from time $t = 0$ till ∞ to maximize utility subject to the given conditions. Unfortunately the resulting system of first order conditions does not have to have an analytic, closed-form solution in general. Then it is necessary to use numerical methods to approximate its behavior. The most easiest and popular is the log-linearization method. It consists of three steps: making variables stationary (because most of them are growing in average), looking for the steady-states by the standard technique of differential equations system solution and finally log-linearization of the equilibrium conditions. After this procedure the system attains the form which can be written in the sense of Uhlig (Uhlig, 2004) or Sd̈erlind (Söderlind, 2003) and we can find the state-space representation. Then comes an estimation of a model that embodies in determination of unknown parameters from given data set. A calibration of parameters is very popular. It is, simply said, setting of parameters from an economic theory and/or by expert's pleasure. Once the parameters are set, model states estimation can be carried out.

Let the model from section 2 be written in the state space form and parametrized. Let the observed variables be the output gap yd_t, nominal interest rate gap R_t, inflation gap π_t and real wages gap w_t. The model state estimation can be carried out either by Kalman or Bootstrap filter. Observed variables and states can be filtered and smoothed in the same quality, if the setting of shocks variances is equivalent for both of the methods (see Fig. 1).

4 Conclusion

After many experiments and computational exercises we can draw tho following conclusions. By equivalent setting of shocks variances for Kalman smoother and for Bootstrap filter we can reach for successful filtered and smoothed observed variables and states. Because of flexibly changing estimated covariance matrix of observation vector estimation error, Kalman smoother enables us to smooth observed variables vector with unlimited accuracy but at the expense of overrun of set standard deviations of model shocks. This is not possible in case of Bootstrap filter, because the algorithm fails if the covariance matrix of observation vector estimation error is set very tightly. Much more, in Bootstrap filter algorithm we often must not use smoothing run of filter because of matrix R_t in state-transition equation. But even over this handicap, Bootstrap filter can give satisfying results and can be used for estimation of time-varying parameters in nonlinear, nongaussian systems (see [7]).

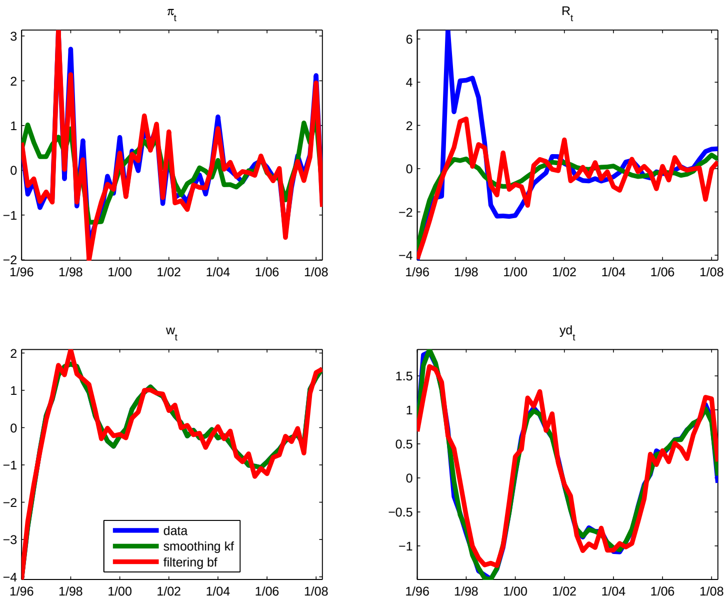


Figure 1: Filtered and/or smoothed observed variables

References

[1] Fernández-Villaverde, J., Rubio-Ramírez, J. F.: How structural are structural parameters? Working paper 13166, National Bureau of Economic Research, 2007.

[2] Hansen, G. D.: Indivisible labor and the business cycle. Journal of Monetary Economics 16, 309-327 (1985)

[3] Ireland, P. N.: A method for taking models to the data. Journal of Economic Dynamics and Control 28, 1205-1226 (2004)

[4] Juillard, M.: Dynare manual. Awww.cepremap.cnrs.fr/dynare/download/manual

[5] Maley, J.: Lectures Notes on the Theory, Calibration & Estimation of Dynamic Stochastic General Equilibrium Models. Departure of Economics, University of Glasgow, 2004.

[6] Söderlind, Paul.: Lectures Notes for Monetary Policy (PhD course at UNISGN). Departure of Economics, University of St. Gallen and CEPR, 2003

[7] Tonner, J., Vašíček, O., Štecha, J., Havlena, V.: Estimation of time - variable parameters of macroeconomic model with rational expectations. In IFAC Symposium, Computional Economics & Financial and Industrial Systems, 201-206 (2007)

[8] Trnka, P.: Matlab toolbox for Sequential Monte Carlo Methods (SMCtool). Czech Technical University in Prague, Faculty of Electrical Engineering, Department of Control Engineering, 2004.

$$x^{(i)}(t|t-1) \sim p(x(t)|D^{t-1}), \quad i = 1, \ldots, M.$$

The samples can be considered as an empirical CPDF

$$p_N(x(t)|D^{t-1}) = \sum_{i=1}^{M} w^{(i)}(t|t-1)\delta(x(t) - x^{(i)}(t|t-1))$$

with uniform weights

$$w^{(i)}(t|t-1) = \frac{1}{M}, i = 1, \ldots, M.$$

Then the knowledge of the state described by $p(x(t)|D^{t-1})$ can be updated by data update step

$$p_N(x(t)|D^{t}) = \sum_{i=1}^{M} w^{(i)}(t|t)\delta(x(t) - x^{(i)}(t|t))$$

where $x^{(i)}(t|t) = x^{(i)}(t|t-1)$ and $w^{(i)}(t|t) = \frac{p(y(t)|x(i)(t|t-1),u(t))}{\sum_{j=1}^{M} p(y(t)|x(j)(t|t-1),u(t))}$ for $i = 1, \ldots, M$ and by time update step by importance resampling of $p_M(x(t)|D^t)$ with additive noice $v(t)$, i.e.

$$x^{(i)}(t+1|t) = f(x^{(j)}(t|t), u(t)) + v^{(i)}(t)$$

where $x^{(j)}(t|t)$ is a sample drawn from $p_M(x(t)|D^t)$ and $v^{(i)}(t)$ is sample drawn from $p(v(t))$. The predicted, filtered and smoothed estimates of the state can be obtained as a weighted sample mean

$$\hat{x}_{MS}(t|t') = \sum_{i=1}^{M} w^{(i)}(t|t')x^{(i)}(t).$$

More detailed description can be found in Štecha and Havlena, 1995 or Havlena and Štecha, 1998 or Trnka, 2004.

4.0.1 The General Case: Nontriviality.

We assume that H is (A_∞, B_∞)-subquadratic at infinity, for some constant symmetric matrices A_∞ and B_∞, with $B_\infty - A_\infty$ positive definite. Set:

$$\gamma : \;=\; \text{smallest eigenvalue of } \; B_\infty - A_\infty \tag{1}$$

$$\lambda : \;=\; \text{largest negative eigenvalue of } \; J\frac{d}{dt} + A_\infty \; . \tag{2}$$

Theorem 9 tells us that if $\lambda + \gamma < 0$, the boundary-value problem:

$$\begin{aligned} \dot{x} &= JH'(x) \\ x(0) &= x(T) \end{aligned} \tag{3}$$

has at least one solution $\overline{x}$, which is found by minimizing the dual action functional:

$$\psi(u) = \int_o^T \left[\frac{1}{2} \left(\Lambda_o^{-1} u, u \right) + N^*(-u) \right] dt \tag{4}$$

on the range of Λ, which is a subspace $R(\Lambda)_L^2$ with finite codimension. Here

$$N(x) := H(x) - \frac{1}{2} \left(A_\infty x, x \right) \tag{5}$$

is a convex function, and

$$N(x) \le \frac{1}{2} \left(\left(B_\infty - A_\infty \right) x, x \right) + c \quad \forall x \ . \tag{6}$$

Proposition 3. *Assume $H'(0) = 0$ and $H(0) = 0$. Set:*

$$\delta := \liminf_{x \to 0} 2N(x) \left\| x \right\|^{-2} \ . \tag{7}$$

If $\gamma < -\lambda < \delta$, the solution $\overline{u}$ is non-zero:

$$\overline{x}(t) \ne 0 \quad \forall t \ . \tag{8}$$

Proof. Condition (7) means that, for every $\delta' > \delta$, there is some $\varepsilon > 0$ such that

$$\|x\| \le \varepsilon \Rightarrow N(x) \le \frac{\delta'}{2} \|x\|^2 \ . \tag{9}$$

It is an exercise in convex analysis, into which we shall not go, to show that this implies that there is an $\eta > 0$ such that

$$f \|x\| \le \eta \Rightarrow N^*(y) \le \frac{1}{2\delta'} \|y\|^2 \ . \tag{10}$$

Since u_1 is a smooth function, we will have $\|hu_1\|_\infty \le \eta$ for h small enough, and inequality (10) will hold, yielding thereby:

$$\psi(hu_1) \le \frac{h^2}{2} \frac{1}{\lambda} \|u_1\|_2^2 + \frac{h^2}{2} \frac{1}{\delta'} \|u_1\|^2 \ . \tag{11}$$

If we choose δ' close enough to δ, the quantity $\left(\frac{1}{\lambda} + \frac{1}{\delta'} \right)$ will be negative, and we end up with

$$\psi(hu_1) < 0 \qquad \text{for} \quad h \ne 0 \quad \text{small} \ . \tag{12}$$

On the other hand, we check directly that $\psi(0) = 0$. This shows that 0 cannot be a minimizer of ψ, not even a local one. So $\overline{u} \ne 0$ and $\overline{u} \ne \Lambda_o^{-1}(0) = 0$. □ □

Figure 2: This is the caption of the figure displaying a white eagle and a white horse on a snow field

Corollary 1. *Assume H is C^2 and (a_∞, b_∞)-subquadratic at infinity. Let $\xi_1, \dots, \xi_N$ be the equilibria, that is, the solutions of $H'(\xi) = 0$. Denote by ω_k the smallest eigenvalue of $H''(\xi_k)$, and set:*

$$\omega := \text{Min}\,\{\omega_1, \dots, \omega_k\}\ . \tag{13}$$

If:

$$\frac{T}{2\pi} b_\infty < -E\left[-\frac{T}{2\pi} a_\infty\right] < \frac{T}{2\pi}\omega \tag{14}$$

then minimization of ψ yields a non-constant T-periodic solution $\overline{x}$.

We recall once more that by the integer part $E[\alpha]$ of $\alpha \in \mathbb{R}$, we mean the $a \in \mathbb{Z}$ such that $a < \alpha \le a + 1$. For instance, if we take $a_\infty = 0$, Corollary 2 tells us that $\overline{x}$ exists and is non-constant provided that:

$$\frac{T}{2\pi} b_\infty < 1 < \frac{T}{2\pi} \tag{15}$$

or

$$T \in \left(\frac{2\pi}{\omega}, \frac{2\pi}{b_\infty}\right)\ . \tag{16}$$

Proof. The spectrum of Λ is $\frac{2\pi}{T}\mathbb{Z} + a_\infty$. The largest negative eigenvalue λ is given by $\frac{2\pi}{T}k_o + a_\infty$, where

$$\frac{2\pi}{T}k_o + a_\infty < 0 \le \frac{2\pi}{T}(k_o + 1) + a_\infty\ . \tag{17}$$

Hence:

$$k_o = E\left[-\frac{T}{2\pi} a_\infty\right]\ . \tag{18}$$

The condition $\gamma < -\lambda < \delta$ now becomes:

$$b_\infty - a_\infty < -\frac{2\pi}{T}k_o - a_\infty < \omega - a_\infty \tag{19}$$

which is precisely condition (14). □ □

Lemma 4. *Assume that H is C^2 on $\mathbb{R}^{2n} \setminus \{0\}$ and that $H''(x)$ is non-degenerate for any $x \ne 0$. Then any local minimizer $\widetilde{x}$ of ψ has minimal period T.*

Table 1: This is the example table taken out of The TEXbook, p. 246

Year	*World population*
8000 B.C.	*5,000,000*
50 A.D.	*200,000,000*
1650 A.D.	*500,000,000*
1945 A.D.	*2,300,000,000*
1980 A.D.	*4,400,000,000*

Proof. We know that $\widetilde{x}$, or $\widetilde{x} + \xi$ for some constant $\xi \in \mathbb{R}^{2n}$, is a T-periodic solution of the Hamiltonian system:

$$\dot{x} = JH'(x) \ . \tag{20}$$

There is no loss of generality in taking $\xi = 0$. So $\psi(x) \geq \psi(\widetilde{x})$ for all $\widetilde{x}$ in some neighbourhood of x in $W^{1,2}\left(\mathbb{R}/T\mathbb{Z}; \mathbb{R}^{2n}\right)$.

But this index is precisely the index $i_T(\widetilde{x})$ of the T-periodic solution $\widetilde{x}$ over the interval $(0,T)$, as defined in Sect. 2.6. So

$$i_T(\widetilde{x}) = 0 \ . \tag{21}$$

Now if $\widetilde{x}$ has a lower period, T/k say, we would have, by Corollary 31:

$$i_T(\widetilde{x}) = i_{kT/k}(\widetilde{x}) \geq k i_{T/k}(\widetilde{x}) + k - 1 \geq k - 1 \geq 1 \ . \tag{22}$$

This would contradict (21), and thus cannot happen. □ □

Notes and Comments. The results in this section are a refined version of [**?**]; the minimality result of Proposition 14 was the first of its kind.

To understand the nontriviality conditions, such as the one in formula (16), one may think of a one-parameter family x_T, $T \in \left(2\pi\omega^{-1}, 2\pi b_\infty^{-1}\right)$ of periodic solutions, $x_T(0) = x_T(T)$, with x_T going away to infinity when $T \to 2\pi\omega^{-1}$, which is the period of the linearized system at 0.

Theorem 9 (Ghoussoub-Preiss)**.** *Assume $H(t,x)$ is $(0,\varepsilon)$-subquadratic at infinity for all $\varepsilon > 0$, and T-periodic in t*

$$H(t,\cdot) \qquad \text{is convex} \quad \forall t \tag{23}$$

$$H(\cdot,x) \qquad \text{is} \quad T\text{−periodic} \quad \forall x \tag{24}$$

$$H(t,x) \geq n\left(\|x\|\right) \qquad \text{with} \quad n(s)s^{-1} \to \infty \quad \text{as} \quad s \to \infty \tag{25}$$

$$\forall \varepsilon > 0 \ , \quad \exists c \ : \ H(t,x) \leq \frac{\varepsilon}{2}\|x\|^2 + c \ . \tag{26}$$

Assume also that H is C^2, and $H''(t,x)$ is positive definite everywhere. Then there is a sequence x_k, $k \in \mathbb{N}$, of kT-periodic solutions of the system

$$\dot{x} = JH'(t,x) \tag{27}$$

such that, for every $k \in \mathbb{N}$, there is some $p_o \in \mathbb{N}$ with:

$$p \geq p_o \Rightarrow x_{pk} \neq x_k \ . \tag{28}$$

$\square$

Example 7 (External forcing)**.** *Consider the system:*

$$\dot{x} = JH'(x) + f(t) \tag{29}$$

where the Hamiltonian H is $(0, b_\infty)$-subquadratic, and the forcing term is a distribution on the circle:

$$f = \frac{d}{dt}F + f_o \qquad \text{with} \quad F \in L^2\left(\mathbb{R}/T\mathbb{Z}; \mathbb{R}^{2n}\right) \ , \tag{30}$$

where $f_o := T^{-1}\int_o^T f(t)dt$. For instance,

$$f(t) = \sum_{k \in \mathbb{N}} \delta_k \xi \ , \tag{31}$$

where δ_k is the Dirac mass at $t = k$ and $\xi \in \mathbb{R}^{2n}$ is a constant, fits the prescription. This means that the system $\dot{x} = JH'(x)$ is being excited by a series of identical shocks at interval T.

Definition 3. *Let $A_\infty(t)$ and $B_\infty(t)$ be symmetric operators in $\mathbb{R}^{2n}$, depending continuously on $t \in [0,T]$, such that $A_\infty(t) \leq B_\infty(t)$ for all t.*

A Borelian function $H : [0,T] \times \mathbb{R}^{2n} \to \mathbb{R}$ is called (A_∞, B_∞)-subquadratic at infinity if there exists a function $N(t,x)$ such that:

$$H(t,x) = \frac{1}{2}\left(A_\infty(t)x, x\right) + N(t,x) \tag{32}$$

$$\forall t \ , \quad N(t,x) \qquad \text{is convex with respect to} \quad x \tag{33}$$

$$N(t,x) \geq n\left(\|x\|\right) \qquad \text{with} \quad n(s)s^{-1} \to +\infty \quad \text{as} \quad s \to +\infty \tag{34}$$

$$\exists c \in \mathbb{R} \ : \quad H(t,x) \leq \frac{1}{2}\left(B_\infty(t)x, x\right) + c \quad \forall x \ . \tag{35}$$

If $A_\infty(t) = a_\infty I$ and $B_\infty(t) = b_\infty I$, with $a_\infty \leq b_\infty \in \mathbb{R}$, we shall say that H is (a_∞, b_∞)-subquadratic at infinity. As an example, the function $\|x\|^\alpha$, with $1 \leq \alpha < 2$, is $(0, \varepsilon)$-subquadratic at infinity for every $\varepsilon > 0$. Similarly, the Hamiltonian

$$H(t,x) = \frac{1}{2}k\|k\|^2 + \|x\|^\alpha \tag{36}$$

is $(k, k+\varepsilon)$-subquadratic for every $\varepsilon > 0$. Note that, if $k < 0$, it is not convex.

Notes and Comments. The first results on subharmonics were obtained by Rabinowitz in [**?**], who showed the existence of infinitely many subharmonics both in the subquadratic and superquadratic case, with suitable growth conditions on H'. Again the duality approach enabled Clarke and Ekeland in [**?**] to treat the same problem in the convex-subquadratic case, with growth conditions on H only.

Recently, Michalek and Tarantello (see [**?**] and [**?**]) have obtained lower bound on the number of subharmonics of period kT, based on symmetry considerations and on pinching estimates, as in Sect. 5.2 of this article.

Osvald Vašíček and Adam Remo[1]

Department of Applied Mathematics and Computer Science
Faculty of Economics and Administration of Masaryk University
Lipová 41a, 602 00 Brno

e-mail: osvald@econ.muni.cz, adamremo@seznam.cz

Monetary Policy with Commitment: Case of the Czech National Bank[2]

Abstract

This article estimates a New Keynesian small open economy model of the Czech economy. The model incorporates household's habit formation in consumption and monopolistic competition with staggered prices in sector of domestic producers and importers. The prices of domestic goods and imports are modeled in Calvo style. The monetary policy is not formulated as Taylor-type rule as commonly used, because the preferences of central bank and sensitivity of macroeconomic variables to nominal interest rate are mixed within the parameters in Taylor rule. In this article, the central bank is rather treated as another optimizing agent. This approach enables us to estimate the weights (representing preferences) that the Czech National Bank puts on inflation targeting, output stabilization and nominal interest rate smoothing. The article uses solution for optimal commitment policy proposed by R. Dennis. Estimates of the model's parameters are obtained by Bayesian technique with Kalman filter and Metropolis-Hastings algorithm. The estimates of the weights in central bank's loss function indicate that the Czech National Bank pays only a little attention to output stabilization in comparison to inflation targeting. This is in accordance with proclaimed policy of the Czech National Bank. Next, the estimated weight of the interest rate smoothing is quite substantial which reflects the carefulness of the Czech National Bank in setting nominal interest rates in recent years.

Keywords: *central bank's preferences, monetary policy, optimal commitment policy, small open economy, New Keynesian, DSGE, Bayesian estimation*

1 Introduction

This article estimates a New Keynesian model of a small open economy developed by Gali and Monacelli in [7]. It treats the central bank as another optimizing agent instead of incorporating the Taylor-type rule into the model. Such specification of the central bank's monetary policy enables estimation of the weights the Czech National Bank (CNB) puts on its objectives. The case with the Taylor-type rule is considered by Musil and Vašíček in [10].

[1] We thank Miroslav Hloušek and Jan Čapek for their helpful advices and comments. We appreciate the access to the METACentrum computing facilities provided under the research intent MSM6383917201.

[2] This work was supported by MŠMT projet Research Centers 1M0524, and funding of specific research at ESF MU.

The article uses the solution algorithm for optimal commitment monetary policy developed by Dennis in [3]. Similar estimates in case of the central bank of Canada, New Zealand, and Australia were done by Kam, Lees and Liu in [9]. They considered discretionary policy instead of commitment.

2 Model

The used model of a small open economy is the one developed by Gali and Monacelli in [7]. This model includes one real rigidity (the household's habit formation in consumption), and two nominal rigidities (the monopolistic competition with Calvo style staggered prices in sector of domestic producers and importers). The model before introducing the central bank's behaviour into consists of eleven equations:

$$c_t - hc_{t-1} = E_t(c_{t+1} - hc_t) - \frac{1-h}{\sigma}(r_t - E_t\pi_{t+1}) \ , \tag{1a}$$

$$\begin{aligned}\pi_{H,t} = \beta E_t(\pi_{H,t+1} - \delta_H \pi_{H,t}) + \delta_H \pi_{H,t-1} + \\ +\lambda_H \left[\phi y_t - (1+\phi)a_t + \alpha s_t + \frac{\sigma}{1-h}(c_t - hc_{t-1})\right] + \lambda_H \varepsilon_t^H \ ,\end{aligned} \tag{1b}$$

$$\pi_{F,t} = \beta E_t(\pi_{F,t+1} - \delta_F \pi_{F,t}) + \delta_F \pi_{F,t-1} + \lambda_F[q_t - (1-\alpha)s_t] + \lambda_F \varepsilon_t^F \ , \tag{1c}$$

$$E_t(q_{t+1} - q_t) = (r_t - E_t\pi_{t+1}) - (r_t^* - E_t\pi_{t+1}^*) + \varepsilon_t^q \ , \tag{1d}$$

$$y_t = (1-\alpha)c_t + \alpha\eta q_t + \alpha\eta s_t + \alpha y_t^* \ , \tag{1e}$$

$$\pi_t = (1-\alpha)\pi_{H,t} + \alpha\pi_{F,t} \ , \tag{1f}$$

$$s_t - s_{t-1} = \pi_{F,t} - \pi_{H,t} + \varepsilon_t^s \ , \tag{1g}$$

$$a_t = \rho_a a_{t-1} + \varepsilon_t^a \ , \tag{1h}$$

$$\pi_t^* = a_1 \pi_{t-1}^* + \varepsilon_t^{\pi^*} \ , \tag{1i}$$

$$y_t^* = b_2 y_{t-1}^* + \varepsilon_t^{y^*} \ , \tag{1j}$$

$$r_t^* = c_3 r_{t-1}^* + \varepsilon_t^{r^*} \ . \tag{1k}$$

All variables in these equations are expressed as gap from steady-state. The Equation (1a) is the consumption Euler equation. Parameter $h \in [0,1]$ is measure of habit persistance in consumption, and parameter $\sigma > 0$ is inverse elasticity of intertemporal substitution. The Equation (1b) is the New Keynesian Phillips curve of domestic goods inflation. The parameter $\delta_H \in [0,1]$ is degree of inflation indexation, $\phi > 0$ is inverse elasticity of labour supply, $\alpha \in [0,1]$ is degree of openness of home economy, and $\lambda_H = (1-\beta\theta_H)(1-\theta_H)\theta_H^{-1}$ where β is discount factor[3], and $\theta_H \in [0,1]$ is Calvo parameter. The Equation (1c) is the New Keynesian Phillips curve of imported goods inflation. The interpretation of parameters in this equation is analogous to that in the previous one. The Equation (1d) is the uncovered interest parity condition. The Equation (1e) is the goods-market clearing condition. The parameter $\eta > 0$ is elasticity of substitution between home and foreign goods. The Equation (1h) describes development of technology shock. The three AR processes (1i)–(1k) describe foreign economy. The shocks ε_t^j are gaussian with zero mean and standard deviation σ_j for $j = a,\ s,\ q,\ H,\ F,\ \pi^*,\ y^*,$ and r^*.

[3]This parameter is not estimated but rather is set to a calibrated value 0.99.

The behaviour of the central bank is modeled as optimal commitment policy. The case of Taylor-type rule is considered in [10]. In this article, the central bank sets nominal interest rate to minimize an expected loss

$$E_0 \sum_{t=0}^{\infty} \beta^t [z_t' W z_t + x_t' Q x_t] = E_0 \sum_{t=0}^{\infty} \beta^t [\tilde{\pi}_t^2 + \mu_y y_t^2 + \mu_r (\Delta r_t)^2] \ , \tag{2}$$

subject to

$$A_0 z_t = A_1 z_{t-1} + A_2 E_t z_{t+1} + A_3 x_t + A_4 E_t x_{t+1} + A_5 v_t \ , \tag{3}$$

where $\tilde{\pi}_t = \sum_{i=0}^{3} \pi_{t-i}/4$ is quarterlized gap of annual inflation and $\Delta r_t = r_t - r_{t-1} + \varepsilon_t^r$, where shock ε_t^r is gaussian with zero mean and standard deviation σ_r. Δr_t is targeted change in short-term interest rate. The monetary shock ε_t^r represents central bank's imperfect ability to control nominal interest rate. The parameters $\mu_y,\ \mu_r \in [0, \infty)$ are weights on output stabilization and interest rate smoothing in central bank's decision making, respectively. Both parameters are expressed relatively to the weight on inflation stabilization. The constraint (3) consists of the equation system (1) plus some auxiliary equations.

The solution algorithm for optimal commitment developed by Richard Dennis in [3] is used in this article.

3 Data

The data are obtained from the Czech National Bank, the Czech Statistical Office, the Euro Area Business Cycle Network and the Data Service & Information. We used quarterly data from 1Q1996 to 3Q2007. We did not annualize the data. The used measurements and corresponding model variables are

$\pi_{F,t}$ – Deviation of seasonally adjusted import prices q-o-q inflation from a long run trend. The trend is computed by the HP filter[4].

π_t – Inflation gap is computed as

$$\pi_t = PIE_t - PIETAR_t/4,$$

where PIE is seasonally adjusted CPI q-o-q inflation and PIETAR is y-o-y inflation target.

y_t – Real GDP gap. The gap is obtained from the CNB.

i_t – Gap of 3-month PRIBOR from its long run trend. The trend is computed by the HP filter.

q_t – Real exchange rate CZK/EUR gap. The gap is obtained from the CNB.

s_t – Logarithm of the terms of trade is computed as

$$\log\left(\frac{IPI_t}{EPI_t}\right),$$

[4]Hodrick-Prescott filter.

where IPI is price index of imported goods and EPI is price index of exported goods. s_t is deviation from the long run trend. The trend is computed by the HP filter.

π_t^* – Foreign inflation gap is computed as gap of seasonally adjusted q-o-q EMU CPI inflation from the estimated trend. The trend is computed by the HP filter.

y_t^* – EMU real GDP gap. The gap is obtained from the CNB.

i_t^* – Gap of 3-month EURIBOR from its long run trend. The trend is computed by the HP filter.

4 Results

The estimates of the model parameters are obtained by the Bayesian estimation technique with use of the Metropolis-Hastings algorithm and the Kalman filter. Length of the simulated Markov chain is 1,000,000 draws. The results of estimation are reported in Table 1.

The estimated values of the parameters are in most cases acceptable with regard to structural characteristics of the Czech economy. Nevertheless it is convenient to do some remarks before proceeding to discussion about the estimates of the monetary policy parameters. The estimate of the degree of openness (parameter α) is 0.78. This value is almost twice the magnitude of the calibrated value in [10][5]. The estimated value of the parameter σ_F which stands for standard deviation of the importers' cost-push shock is 11.26. This value is high even if multiplied by estimate of λ_F (see Equation (1c))[6]. An insufficient discriminability of some shocks in the model structure may be cause of this imperfection.

Now lets focus on the estimates of the monetary policy parameters. The estimated weight on output stabilization (parameter μ_y) is 0.09. This value is almost zero and indicates that the CNB is not very interested in output stabilization. The estimated weight on interest rate smoothing (parameter μ_r) is 0.52. This value is quite high in comparison to the estimate of the parameter μ_y. The parameter μ_r may represent carefulness of the CNB in setting nominal interest rate. It is found out that stabilization of inflation is the most important objective of the CNB in pursuing its monetary policy, because the estimates of both parameters (μ_y and μ_r) are significantly lower than one.

5 Conclusion

This article estimates New Keynesian model of a small open economy with optimal commitment monetary policy. The estimate is based on the Czech Republic and the European Economic and Monetary Union (EMU) data. The estimated values of the parameters are in most cases acceptable with regard to structural characteristics of the Czech economy.

The monetary policy specification which is adopted in this article enables estimation of the CNB's "preferences". It is found out that the CNB's attention to output stabilization is negligible in comparison to inflation targeting. The estimate of the weight on interest rate smoothing is 0.6. This weight can be interpreted as meausure of carefulness of the CNB in setting nominal interest rates. The inflation targeting is the most important objective of the CNB, because the estimated weights on the other two objectives are less than one. This result is in accordance with inflation targeting regime the CNB adopted in 1998.

[5]They set $\alpha = 0.4$.

[6]$\lambda_F \sigma_F \doteq 2.4$

References

[1] BINDER, M., H. PESARAN: Multivariate rational expectations models and macroeconomic modelling: a review and some new results. University of Cambridge, Cambridge working papers in economics, working paper 9415, 1995.

[2] CALVO, G.: Staggered Contracts in a Utility-Maximising Framework. Journal of Monetary Economics, 12, ,(1983),pages 383-398.

[3] DENNIS, R.: Optimal policy in rational expectations models: new solution algorithms. Federal Reserve Bank Of San Francisco, working paper 2001-09, 2005.

[4] DENNIS, R.: Inflation targeting under commitment and discretion. Federal Reserve Bank Of San Francisco, Economic Review, 2005, pages 1-13.

[5] DENNIS, R.: Specifying and Estimating New Keynesian Models with Instrument Rules and Optimal Monetary Policies. Federal Reserve Bank of San Francisco, Working Paper 2004-17 (rev. January 2005),(2005).

[6] FISHMAN, G. S.: Monte Carlo: concepts, algorithms and applications. Springer-Verlag New York, Inc., 1995.

[7] GALI, J., T. MONACELLI: Monetary policy and exchange rate volatility in a small open economy. NBER working paper, no. 8905, 2002.

[8] GEWEKE, J.: Contemporary Bayesian Econometrics and Statistics. John Wiley & Sons, Inc., 2005.

[9] KAM, T., K. LEES, P. LIU: Uncovering the hit-list for small inflation targeters: a Bayesian structural analysis. ANU Centre For Applied Macroeconomic Analysis, 2006.

[10] MUSIL, K., O. VAŠÍČEK: Behavior of the Czech Economy: New Open Economy Macroeconomics DSGE Model. CVKS MU, working paper No. 23/2006, Brno, 2006.

[11] PEERSMAN, G., R. STRAUB: Putting the New Keynesian Model to a Test. Ghent University, Working Papers of Faculty of Economics and Business Administration, wp 06/375, Belgium, 2006.

[12] SAWYER, S.: The Metropolitan-Hastings algorithm and extensions. Washington University, 2006

Appendix

	Interpretation	**Prior mean**	**5%**	**95%**	**Post. mean**	**5%**	**95%**
α	degree of openness	0.70	0.52	0.85	0.78	0.71	0.83
h	habit in consumption	0.80	0.61	0.94	0.89	0.86	0.92

continue on next page

Table 1: Estimation Results

	Interpretation	Prior mean	5%	95%	Post. mean	5%	95%
σ	inverse elasticity of intertemporal substitution	0.50	0.22	0.87	0.51	0.35	0.69
ϕ	inverse elasticity of labour supply	2.00	1.46	2.61	2.19	1.63	2.81
η	elasticity of substitution between home and foreign goods	0.60	0.26	1.06	0.59	0.42	0.85
δ_H	degree of inflation indexation in prices of products	0.80	0.61	0.94	0.82	0.65	0.94
δ_F	degree of inflation indexation in prices of imports	0.80	0.61	0.94	0.87	0.73	0.96
θ_H	fraction of non-optimizing producers	0.50	0.34	0.66	0.53	0.43	0.62
θ_F	fraction of non-optimizing importers	0.60	0.43	0.76	0.63	0.56	0.69
a_1	foreign inflation AR(1) parameter	0.70	0.54	0.87	0.67	0.53	0.83
b_2	foreign output AR(1) parameter	0.90	0.74	1.07	0.91	0.85	0.97
c_3	foreign interest rate AR(1) parameter	0.80	0.64	0.97	0.75	0.62	0.89
ρ_a	inertia of technology	0.85	0.66	0.97	0.90	0.84	0.95
μ_y	weight on output stabilization	0.30	0.10	0.58	0.09	0.04	0.16
μ_r	weight on interest rate smoothing	0.60	0.38	0.87	0.52	0.33	0.74
σ_H	std. deviation of producers' cost-push shock	1.00	0.34	1.94	0.86	0.65	1.09
σ_F	std. deviation of importers' cost-push shock	11.00	9.41	12.70	11.26	9.87	12.73
σ_a	std. deviation of technology shock	0.80	0.64	0.97	0.75	0.60	0.92

continue on next page

Table 1: Estimation Results

	Interpretation	**Prior mean**	**5%**	**95%**	**Post. mean**	**5%**	**95%**
σ_q	std. deviation of UIP shock	0.10	0.03	0.19	0.02	0.01	0.04
σ_s	std. deviation of terms of trade shock	0.70	0.54	0.87	0.60	0.47	0.74
$\sigma_{\pi*}$	std. deviation of foreign inflation shock	0.06	0.02	0.12	0.02	0.01	0.04
σ_{y*}	std. deviation of foreign output shock	0.25	0.17	0.34	0.26	0.21	0.30
σ_{r*}	std. deviation of foreign interest rate shock	0.10	0.03	0.19	0.03	0.02	0.04
σ_r	std. deviation of monetary shock	0.20	0.13	0.29	0.09	0.06	0.11

Table 1: Estimation Results

Petr Volf

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4, 182 08 Praha 8

e-mail: volf@utia.cas.cz

Model for Difference of Two Series of Poisson-like Count Data

Abstract

When the discrete count data are analyzed, we are tempted to use the Poisson model. Consequently, we are often facing the problems with insufficient flexibility of Poisson distribution. However, in many instances the variable of the interest is the difference of two count variables. For these cases, so called Skellam distribution is available, derived originally as the difference of two correlated Poissons. The model contains latent variables, which leads quite naturally to the use of Bayes approach and to data augmentation via the Markov Chain Monte Carlo generation. In our contribution we apply the approach to the numbers of serious road accidents. The Skellam distribution is used for the comparison of the situation before and after the introduction of a new "point" system (in 7/2006).

1 Problems with Fit of Poisson Model

The main problem with Poisson model is that it has only one parameter (λ), characterizing both the expectation and variance, so that the flexibility of the model is rather limited. A typical problem encountered here is the "overdispersion", i. e. the case that data shows larger variance than the mean. The case can be solved, explained, modeled, by several methods.

The most common way how to cope with the problem is to consider a random factor as a part of Poisson intensity. Namely, the intensity is now $\lambda = Z \cdot \lambda_0$, where Z is a positive random variable with $EZ = 1$, λ_0 is a baseline intensity. The variable Z is called the heterogeneity, frailty, and represents certain factors which cause the intensity variation and which we are not able to explain more precisely in the present stage of analysis. Thus, if a random variable X has Poisson $(Z \cdot \lambda_0)$ distribution, then $EX = \lambda_0$, $EX^2 = E_z(E(X^2|Z)) = E_z(\lambda_0^2 Z^2 + \lambda_0 Z) = \lambda_0^2 EZ^2 + \lambda_0$, hence

$$\operatorname{var} X = \lambda_0 + \lambda_0^2(EZ^2 - 1) = \lambda_0 + \lambda_0^2 \operatorname{var} Z = \lambda_0(\lambda_0 \operatorname{var} Z + 1).$$

It is seen that the variance is now larger than λ_0 (if var $Z > 0$) and can be adapted to each variation of data. It is also well known that the case with gamma distributed Z leads actually to negative binomial distribution of X.

In frequent cases the data corresponds roughly to the Poisson distribution except at one or several values. In such cases the improvement utilizes the idea of the mixture of distributions, the Poisson distribution is "inflated" by another discrete one. Then, the data are expected to follow the model

$$P(x) = \pi\, P_0(x) + (1-\pi)\, P_1(x),$$

where $0 \le \pi \le 1$, P_0 is the Poisson probability and P_1 gives more probability to certain points. Thus,

$$\begin{aligned} EX &= \pi\, EX_0 + (1-\pi)\, EX_1, \\ EX^2 &= \pi\, EX_0^2 + (1-\pi)\, EX_1^2, \end{aligned}$$

so that $\operatorname{var} X$ could be either larger or even smaller than EX. For instance, let $P_0 \sim$ Poisson(10), P_1 be concentrated at 10: $P_1(X = 10) = 1$, $\pi = 0{,}8$. Then $P(X = x) = e^{-10}\, 10^x/(x!) \cdot 0{,}8$ for $x \neq 10$ and $P(X = 10) = e^{-10}\, 10^{10}/(10!) \cdot 0{,}8{+}0{,}2$. EX remains 10, but the variance is $10 \cdot 0{,}8 = 8$. The same could be expressed by the model $P(x) = C(x) \cdot P_0(x)$, where $C(x)$ is constant at the most of points and supports or weaken the probability of other points.

Convolution of Poisson distribution with another one, which in general leads to the notion of compound point processes or marked point processes. Let us consider here a special case when each event with occurrence given by Poisson(λ) causes (is composed from) a set of other events, their number given (independently) by Poisson(μ). An example – a collision of a particle (event 1) gives a rise to several photons (event 2). The number of photons, X, is then given by $\sum_{i=0}^{Y} Z_i$, where $Y \sim$ Poisson(λ) and $Z_i \sim$ Poisson(μ). Hence, $EX = \lambda \cdot \mu$, while $\operatorname{var} X = \lambda(\mu^2 + \mu) = \lambda\mu(\mu + 1) > EX$.

Quite naturally, in highly heterogeneous cases, i. e. non-proportional, with different development in different groups, the first task is to separate these groups, and after the separation to fit group-specific models. This step of separation can be done for instance by the model-based clustering technique.

Except of overdispersion, the opposite phenomenon of "underdispersion" can sometimes be encountered, too. It can be caused e. g. by ties (dependencies) in data, also the case when extreme values are not observed may lead to the underdispersion effect.

2 Distribution of Difference of Two Poisson Variables

In many cases of the (discrete time) count data analysis we are interested in the difference of two count variables. Examples are frequent, in medicine, economy, in sport statistics (difference of score of a match). For these cases, so called **Skellam distribution** is available, derived originally as the difference of two independent Poissons (Skellam, 1946).

Let us consider two independent Poisson random variables U, V with parameters λ_1, λ_2, respectively. Then the distribution of the difference $Z = V - U$ is the following:

$$P(Z = z) \;=\; e^{-(\lambda_1+\lambda_2)} (\frac{\lambda_2}{\lambda_1})^{z/2} B_{|z|}(2\sqrt{\lambda_1\lambda_2}), \tag{1}$$

where B is the modified Bessel function of the first kind, namely

$$B_r(x) = (\frac{x}{2})^r \sum_{k=0}^{\infty} \frac{(x^2/4)^k}{k!\Gamma(k+r+1)}\,.$$

It is also seen that the same distribution is obtained when $Z = Y - X$, where $X = U + W$, $Y = V + W$, U, V are as before and W is an arbitrary discrete distribution, with finite moments, say, in order to overcome formal problems with the moments existence.

Thus, the model offers one of possibilities how to increase the flexibility of Poisson models. If X, Y, Z are observed, then U, V are latent independent Poisson variables, W characterizes the common part of X and Y, so that the distributions of them is rather arbitrary.

In other words, though we observe Z as $Y - X$, we can treat it as $V - U$, of course only in the case that the Skellam distribution fits to observations of Z. This can be tested by the standard chi-square goodness-of-fit test.

3 Estimation of Parameters

Let the random sample $\{Z_i\}, i = 1, ..., n$ be available, the aim is to find corresponding Poisson random variables U, V such that $Z = V - U$, i.e. their parameters or their representation. In such a simple case (of i.i.d. variables), the easiest method just compares the sample moments.

Namely, let $\bar{Z}, s_z^2$ be the sample mean and variance from the data, then, as $\mathrm{E}Z = \lambda_2 - \lambda_1$, $\mathrm{var}(Z) = \lambda_2 + \lambda_1$, a natural estimates are

$$\lambda_1 = \frac{s_z^2 - \bar{Z}}{2}, \; \lambda_2 = \frac{s_z^2 + \bar{Z}}{2}$$

(notice that with positive probability such estimates can be negative).

3.1 MCMC Computation

In many applications, however, we deal with the time sequence of Skellam-distributed variables, eventually dependent also on other factors, so that the parameters λ_1, λ_2 follow a regression model and can also develop in time. For such cases we have to find another way of identification of variables U, V. Their nature of latent variables leads to the use of Bayes inference and to the method of data augmentation (i.e. artificial generation of representation of U and V), via the Markov chain Monte Carlo (MCMC) procedure. The main advantage is that while the maximum likelihood estimation (MLE) in the distribution (1) is analytically hardly tractable, the conditional likelihood containing U and V, given Z, is quite simple. Namely, let $z_i = v_i - u_i, i = 1, ..., n$ be random sample of observed variables, u_i, v_i latent Poisson variables, then we have

$$f(u, v | z, \lambda_1, \lambda_2) = \prod_{i=1}^{n} e^{-(\lambda_1 + \lambda_2)} \frac{\lambda_1^{u_i} \lambda_2^{v_i}}{u_i! v_i!} \cdot I[z_i = v_i - u_i], \tag{2}$$

where z, u, v denote corresponding vectors $n \times 1$, $I[.]$ is an indicator function. Then the posterior distribution of parameters $\lambda_k, k = 1, 2$ is proportional to (2) times the prior of parameters and the scheme for the Bayes analysis is completed. A natural choice of conjugate priors for λ_k are independent gamma distributions, then the MCMC updating of them uses the Gibbs sampler.

3.2 Updating the Values of U and V

Let us recall here also the Metropolis step used for instance by Karlis and Ntzoufras (2006) for updating the values of latent variables:

Let u_i, v_i be actual values. Then

– if $z_i < 0$, propose v_i^* from Poisson(λ_2), set $u_i^* = v_i^* - z_i$ and accept it with
$p = min\{1, \lambda_1^{(v_i^* - v_i)} \frac{u_i!}{u_i^*!}\}$,

– if $z_i > 0$, propose u_i^* from Poisson(λ_1), set $v_i^* = u_i^* + z_i$ and accept it with
$p = min\{1, \lambda_2^{(u_i^* - u_i)} \frac{v_i!}{v_i^*!}\}$.

3.3 Modified EM Algorithm

EM algorithm is a standard method used in cases of missing (i.e. also latent) data. In our case, the M step (MLE of parameters provided latent data are available) is straightforward:

$$\lambda_1 = \bar{U},\ \lambda_2 = \bar{V}.$$

However, the E step computing the expectation of latent variables, given observed data and parameters, from distribution (2), is rather difficult. Therefore, we consider the variant combining the M-step (as above) with Monte Carlo updating of latent values following the method described in the preceding part.

In the sequel we shall employ either the MCMC procedure or also the variant using the direct computation of parameters via the M-step. In both cases the result consists of the samples of generated values of U and V (representing the distributions of latent variables) and samples of model parameters representing their posterior distributions. The cases with time or regressor-dependent variables and parameters can be incorporated quite easily.

3.4 Prediction of Future Values

Once the model is evaluated, we can use it for the prediction of new data (i.e. under non-changed conditions). In the Bayes scheme it means to construct the predictive distribution

$$p(x_{new}|\boldsymbol{x}) = \int_{\Theta} p_m(x|\theta) g_a(\theta|\boldsymbol{x}) d\theta,$$

where p_m describes the model, θ its parameters and g_a the aposterior distribution of them, which is based on observed data $\boldsymbol{x}$. After MCMC solution, instead of g_a the sample $\{\theta^{(i)}\}$ representing it is available. Hence, instead of integration, the averaging is used. In some cases the average of $p_m(x|\theta^{(i)})$ can be obtained directly, in a tractable form, for each possible value x. However, the sampling approach is preferred usually. Namely, the sample representing the predictive distribution is obtained in such a way that from each $p_m(x|\theta^{(i)})$ one value (or fixed number of values) is generated.

4 Artificial Example

First, let us demonstrate the procedure of solution on simple artificially generated data. We generated Poisson data U, V, n=100, with parameters $\lambda_1 = 5, \lambda_2 = 10$, and set $Z = V - U$. Estimated mean and variance of Z were: 4.7600, 15.1027, hence the moment method estimates of λ_1, λ_2 were 5.1714, 9.9314,

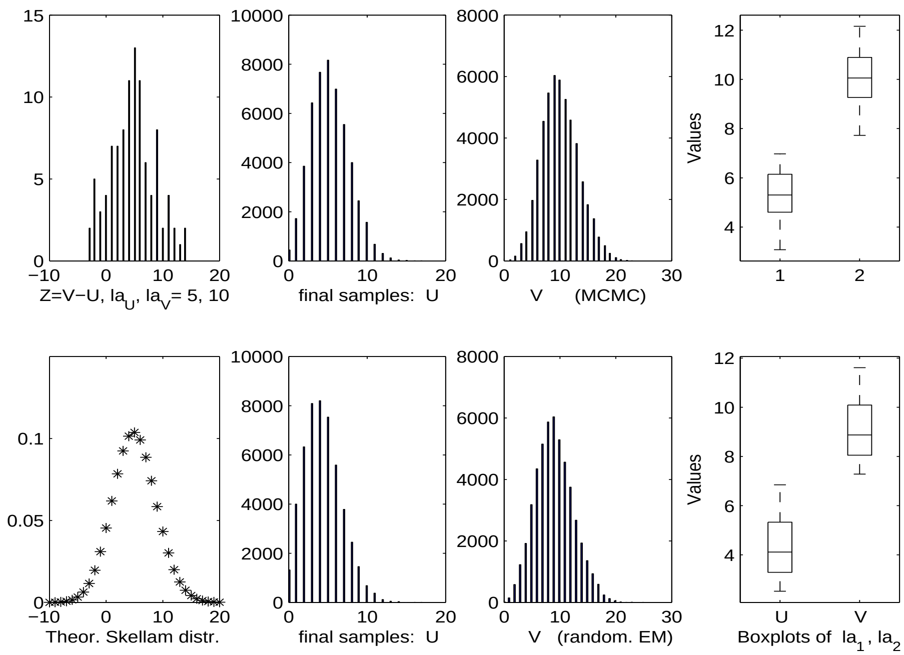


Figure 1: Results of artificial example: samples of latent variables, boxplots of posterior samples of parameters λ_1, λ_2

Then, the MCMC method was used. We selected the same (and rather wide) Gamma priors for both λ-s with parameters $a_0 = 1, b_0 = 10$, i.e. with E=10, var=100. The generation started from randomly uniformly selected integers between 1 and 20 for u and v. 1000 sweeps (iterations of the procedure) were performed, results computed from last 500 of them are displayed in Figure 1. The samples had the following characteristics:

(mean U, var U)= (5.2845, 5.9704), (mean V, var V)= (10.0445, 10.9520).

A variant with direct estimation of parameters (randomized EM algorithm) yielded the following:

(mean U, var U)= (4.3520, 5.6449), (mean V, var V)= (9.1120, 11.4121).

Some other features of solution were noticed, for instance: 1000 iterations quite sufficed – increase of number of iterations did not change results significantly. Increase of n led to narrower posterior distributions of parameters λ, simultaneously they were closer to real values (the consistency).

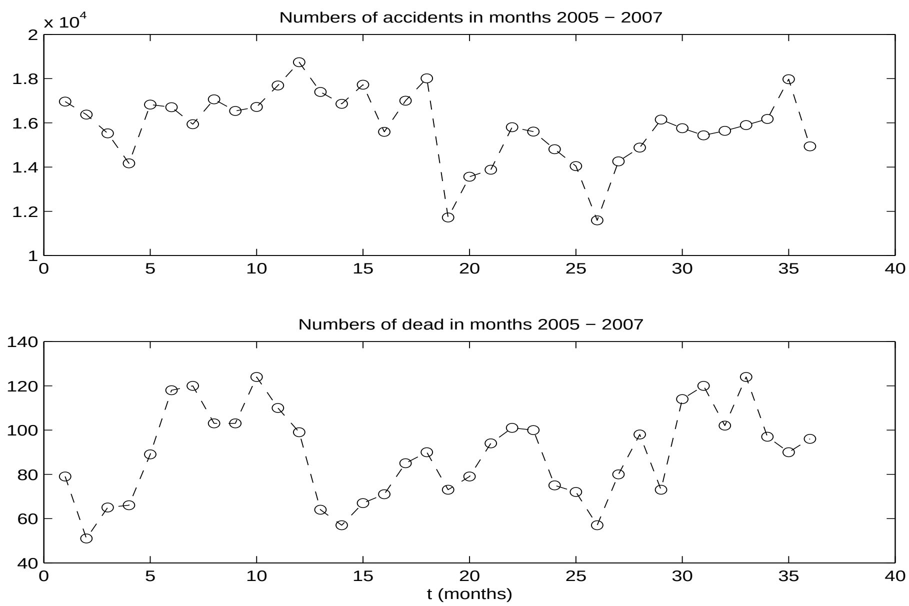


Figure 2: Numbers of accidents (above) and deaths (below), in months 2005–2007

5 Real Data Example

Figure 2 shows the development of monthly numbers of (reported) car accidents in Czech Republic in years 2005–2007, in the upper subplot, and the numbers of their fatal consequences (dead people), below. A new point system together with significant increase of fines and other sanctions was introduced from July, 2006 (here month 19). It is seen that there was certain decline, especially in the numbers of accidents, but the values reached the former level very

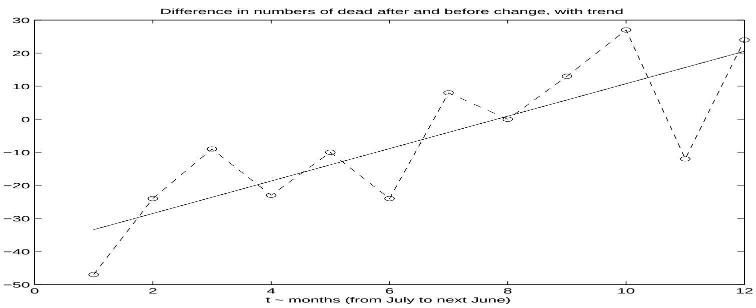


Figure 3: Differences of numbers of deaths, from July 2006–July 2005 to June 2007–June 2006, and fitted linear trend

quickly. The system had just temporal effect and now the situation, especially regarding the serious consequences of accidents, is worse than before.

Here, we analyze the data displayed in Figure 3, namely the differences in numbers of dead, after and before the change of punishment system, in corresponding months. Namely, July 06 – July 05, August 06 – August 05, ... , till June 07 – June 06, so that graph contains 12 such differences. Their linear trend is evident, its estimate (described further) is displayed, too. As such data forms a rather short time series, with no additional factors, we consider the following simple model: Observed data, $Z(t), t = 1, .., 12$, are Skellam variables. It means that they may be expressed as differences $Z(t) = V(t) - U(t)$, where $V(t), U(t)$ are Poisson. Further, we assume that they both have parameters with a linear trend, namely $\lambda_u(t) = a_u + b_u t$, $\lambda_v(t) = a_v + b_v t$, Then, the task is to estimate those four trend parameters. It also means that the mean and variance of $Z(t)$ develop linearly, too, with parameters given by the difference and sum, respectively, of parameters of $V(t)$ and $U(t)$.

As regards the computations, we preferred the randomized EM procedure. We started from $U(t), V(t)$ equal to observed numbers of dead in corresponding months. They evidently have a more complex structure, for instance they contain a seasonal component. It is assumed that this is common for both, and is substracted away. Then, parameters of linear trend, in a framework of Poisson model, were fitted. It means to find maximum likelihood estimates of a and b, i.e. the maximizer of log-likelihood, which is (for variables u, for instance):

$$L = \sum_{t=1}^{12} \left[-a - bt + u(t) \cdot log(a + bt) \right].$$

It was solved with the aid of several iterations of Newton–Raphson type. Then, from estimated parameters, new values of $u(t), v(t)$ were generated by the step described in part 3.2. Such a loop (one sweep) was repeated 1000 times The results displayed further were obtained from last 500 sweeps. Figure 4 shows histograms of samples of all four linear trend parameters. The means of these samples were (in order $\hat{a_u}, \hat{b_u}, \hat{a_v}, \hat{b_v}$) : $105.0160, 2.3203, 66.6882, 7.2287$.

Then the means of trend parameters for variables $Z(t)$ are the differences, namely $\hat{a_z} = -38.3278, \hat{b_z} = 4.9084$, with Bayes credibility intervals, given by 2,5% and 97,5% quantile of

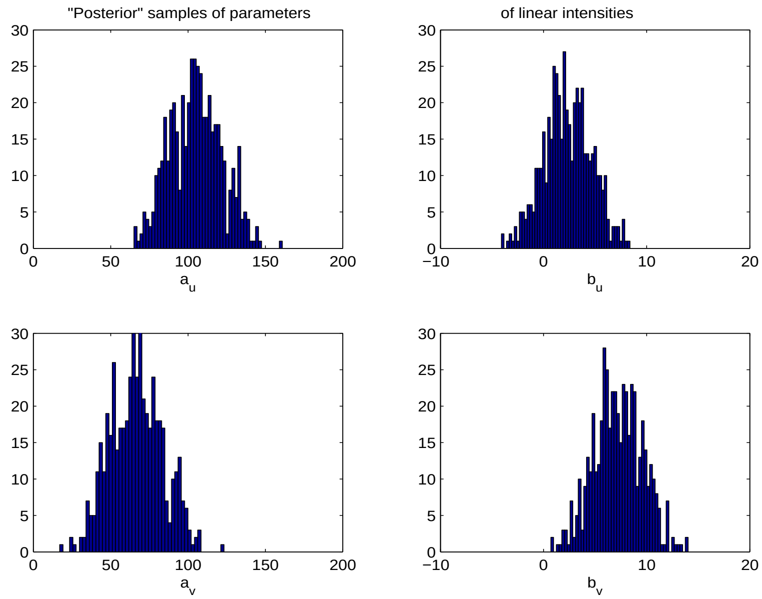


Figure 4: Histograms of approximate posterior samples of trend parameters

posterior sample: (-39.8290, -36.6471) for a_z and (4.6508, 5.1404) for b_z. Hence, Figure 3 contains the trend line with parameters $\hat{a_z}, \hat{b_z}$.

6 Conclusion

We have introduced the distribution for difference of two independent Poisson variables and presented corresponding methods of statistical analysis. What remains to be done is the testing the model fit, at least by the test whether resulting samples of latent components of U, V correspond to Poisson distribution.

Acknowledgement:
The research was supported by the grant of GACR No 402/07/1113.

References

[1] Karlis, D., Ntzoufras, I.: Bayesian analysis of the differences of count data. Statistics in Medicine 25, 1885-1905 (2006).

[2] Skellam, J.G.: The frequency distribution of the difference between two Poisson variates belonging to different populations. Journal of the Royal Statistical Society, Series A, 109, p.296 (1946).

Alicja Wolny-Dominiak

Department of Statistical and Mathematical Methods in Economics
Faculty of Economics
University of Economics in Katowice
ul. 1-Maja 50, 40-287 Katowice

e-mail: woali@ae.katowice.pl

Michał Trzęsiok

Department of Mathematics
Faculty of Management
University of Economics in Katowice
ul. 1-Maja 50, 40-287 Katowice

e-mail: trzesiok@ae.katowice.pl

Monte Carlo Simulation Applied To A‘Priori Rate Making

Abstract

Insurance companies specialising in casualty insurance create their own rating systems for setting fair premiums for every risk for different kinds of insurance portfolios. The rating system is mostly based on the data analysis concerning the number and the value of claims for individuals or groups (classes) of insured people within a given portfolio. Based on a given rating system, the premium for a particular risk is calculated in two stages: *a priori* rating and *a posteriori* rating. In this paper, the process of *a priori* rating is analyzed with the emphasis on the claims distributions used for modelling the rating variables.

The assumption made about the predictors' distribution is crucial for good claims modelling and for the reasonable *a'priori* rate making with GLM. The goal of this paper is to present a simulation procedure which allows to evaluate different variants of *a priori* rating systems which are created using generalized linear models (GLMs) and assume different claims' distributions. The procedure is based on the Monte Carlo method and it results in the distributions of GLM parameters for rating variables, estimation of the average value of claims and the errors of estimation. These results allow to compare different rating systems according to selected model quality measures. In the paper, an empirical example is presented for illustration. In the example we use Gamma distribution and Inverse Gaussian distribution for modelling claims so we receive two rating systems which we compare.

1 Introduction

Insurance companies specialising in casualty insurance create their own rating systems for setting fair premiums for every risk for different kinds of insurance portfolios. The calculation of the pure premium for the overall portfolio of risks should meet two basic conditions: it should ensure that the insurance company will receive premiums at a level adequate to cover the claims and should fairly reflect the probability of an insured event for different groups of customers (i.e. a higher premium for the group of customers where the probability of an event or the sum of claims is higher) [6]. Thus, the development of the rating system involves classifying risks from the inhomogeneous overall portfolio in order to obtain homogeneous sub-portfolios. The homogeneous sub-portfolio is defined as a subset of insurance policies where claims are generated independently and the random variables – the number of claims – are identically distributed within the subset and the random variables – the value of compensation claimed – are identically distributed, too. The same pure premium is assigned to every policy in the homogeneous sub-portfolio [5].

The rating system is mostly based on the data analysis concerning the number and the value of claims for individuals or groups (classes) of insured people within a given portfolio. Based on a given rating system, the premium for a particular risk is calculated in two stages: *a priori* rating and *a posteriori* rating. *A priori* rating relies on the base rate calculation taking into consideration factors which describe the insured person, the specification of the insured object (automobile, property, etc) and the general insurance experience. *A posteriori* rating is the calculation of an additional rate (mostly calculated as a percentage of the base rate) where a number of claims made by the insured person in the past is taken into consideration. Therefore the "bonus-malus" system is mostly used, i.e. the system of discounts and increases in the premium according to an individual's claim experience. In this paper, the process of *a priori* rating is analyzed with the emphasis on the claims distributions used for modelling the rating variables.

As mentioned above, the value of claims in the portfolio depends on different predictors. Treating these predictors as factors (i.e. qualitative random variables) allows us to conduct a statistical data analysis in order to measure the influence of every predictor on the level of claims. Having this done the only problem in the *a'priori* rating that needs to be solved is the estimation of the levels of predictors. Due to their nature, the predictors are called "rating variables". In the automobile insurances, for example, the standard rating variables are: the driver's region, gender, age and the engine capacity. In the paper we present the simulation procedure supporting the choice of distribution for the average claim in the a'priori rate making. The overall *a'priori* rating process is performed using the Generalized Linear Models (GLM) for estimating the influence of the rating variables on the value of claims.

2 Methods used for estimating the rating variables

In the literature there are two groups of methods for estimating the levels of rating variables in the a'priori rate making: minimum bias and maximum likelihood methods. The minimum bias procedures are based on the system of equations balancing the correlations between the variables. The solution of this system of equations is usually biased. Thus, the system is often solved using different methods and the final solution is the one with the minimum bias. In the maximum likelihood procedure we assume the linear dependence between claims and rating variables. As insurance data do not normally meet the assumption of normality, the models widely used for estimating the parameters for every rating variable are the Generalized Linear Models (GLM). The advantage of using these models for modelling claims is that we are able to evaluate the model in terms of goodness of fit, stability or significance of its parameters.

The assumption made about the predictors distribution is crucial for the reasonable a'priori rate making with GLM. The results of many studies showed that the insurance claims distributions have the following properties: they are right-skewed, their values are positive with the variance proportional to the mean [4]. Thus, the claims are mostly modelled by the gamma, lognormal or the Inverse Gaussian distribution. In practice, it is often hard to estimate the distribution of the claims based on the history, so it is necessary to make an assumption about the claims distribution in the portfolio of risks.

3 Generalized Linear Model for estimating the levels of rating variables

In order to present the a'priori rating system based on GLM we use the following notations. Let $X_1, ..., X_n$ be the rating variables. Y denotes the mean value of the claim in the portfolio and $w_1, ..., w_m$ – relativities. The relativities show how to adjust the base premium for the given combination of rating variables values. We assume the model has the linear form: $Y = \sum_{j=1}^{n} \beta_j X_j + \xi$, where $\beta_1, ..., \beta_n$ are the linear regression parameters and ξ is the error. In order to estimate $\beta_1, ..., \beta_n$ using the least squares method we need to meet the assumptions, e.g. the normality of the error $\xi \sim N(0, \sigma^2)$. However, in the actuarial problems the assumption that the mean value of claims is normally distributed or has constant variation is not met. Let us consider the portfolio consisting of insured women and men. In practice, the mean claim for a woman is very often lower than for a man, so the assumption about the normality of the error is met only in sub-portfolios: insured women and, separately, insured men. Thus GLMs are suitable for the a'priori rate making, because the error distribution is assumed to be the one from the extended exponential dispersion family. Since the predictors are 0/1 variables (binary encoded factors), we may interpret the parameters $\beta_1, ..., \beta_n$ as showing the influence of the given feature (predictor) on the claims, e.g. in the automobile insurances parameters may show the influence of the gender, the engine's capacity or the usage time on the average claims' value.

Let us assume that y_i are the realizations of the variable Y which has the distribution from the extended exponential dispersion family. Then the density function has the general form [3]: $f(y_i, \theta_i, \phi) = \exp\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\}$, where ϕ is a constant *scale parameter*, θ_i is some function of $x_{1,i}, ..., x_{n,i}$, e.g. the mean μ_i, for $i = 1, ..., m$denoting the policy number in the portfolio. We also assume that functions θ_i have the linear form $\theta_i(\mu_i) = \sum_j x_{i,j}\beta_j$ (where $x_{1,i}, ..., x_{n,i}$ – independent observations). Switching to the matrix calculus notation, let us denote by X the matrix representing the realizations of predictors X= $\begin{bmatrix} 1 & 0 & 0 \\ 0 & .... & .... \\ ... & 1 & .... \\ 0 & 0 & 1 \end{bmatrix}$ and by B – vector of parameters $[\beta_1, ..., \beta_n]$. Then the model can be written in the form [3]: $\eta(\mu)$ =XB, where the function η is called the linear predictor.

The linear predictor η_i for the ith variable is associated with the expected value μ_i by the *link function* g, which satisfies the conditions of being strongly monotonic and differentiable:

$\eta_i = g_i(\mu_i) = x_i^T \beta$, where x_i denotes the vector corresponding to the ith variable. There is usually one type of link function g used for all the observations.

In order to estimate the parameters $\beta_1, ..., \beta_n$, we use the maximum likelihood technique. Due to the large number of observations and its consequence – high computational complexity, it is, however, necessary to apply some iteration techniques such as the Newton–Raphson method or Fisher's scoring.

The estimated values of parameters $\beta_1, ..., \beta_n$ are usually used for building the so called rating table $T = [t_{ij}]$. If the multiplicative model is assumed, the elements of table T are calculated as: $t_{ij} = \exp(\beta_i + \beta_j)$. Let us assume that risks in the automobile insurances model are classified based on two features: the region and the engine capacity. Then the elements of table T show how much you have to adjust the base premium for the ith region and jth engine capacity.

4 Monte Carlo simulation for the estimation of the average value of claims in the a'priori rate making

The procedure supporting the choice of the underlying distribution for variable Y uses the Monte Carlo technique of generating randomly different values from the given distribution. In order to focus attention, we assume the multiplicative form of the relationship between the rating variables and two features that classify the risks in the portfolio. As a result of the procedure, we obtain a stochastic rating table for every assumed distribution. Analyzing these tables and comparing the distribution selection criteria allows to make a decision which distribution to apply in the final model. The overall simulation procedure may be presented in the following stages (one iteration is described):

STAGE 1 Assume the distribution for the average value of claims (independent variable) and assume the type of a model (multiplicative/additive)

STAGE 2 Select the rating variables (predictors)

STAGE 3 Select the distribution selection criteria

STAGE 4 Select the method for estimating the levels of rating variables (in the a'priori rate making)

STAGE 5 Estimate the levels of rating variables

STAGE 6 Calculate the elements of the rating table $[t_{ij}]$

STAGE 7 Make the prediction of the average value of the claims

STAGE 8 Calculate the predictions errors

The whole simulation procedure produces the stochastic rating table, the prediction of the average value of the claims and the values of the distribution selection criteria. In stage 3 you can choose one of the statistical measures: goodness of fit or the stability of the parameters $\beta_1, ..., \beta_n$ or the stability of the prediction for the dependent variable Y. In the example presented in chapter 5 we use mean squared error (MSE) measuring the goodness of fit of the model to the input data set [2]:

$$MSE = \frac{\sum n_{ij}(E(\hat{Y}_{ij}) - \mu_{ij})^2}{\sum n_{ij}} \to \min$$

In chapter 5 we present an empirical example illustrating the procedure described above. The calculations were conducted using the statistical software R.

5 Empirical example for the automobile insurance dataset – case analysis

In the empirical example we analyze data concerning the claims experience in the automobile insurances (data set taken from [1]). We assume the multiplicative model and two features classifying the risk: a driver's age/marital status and the character of the use of the vehicle. The response variable is the average value of clams weighted by the number of claims. Table 1 presents the labels of variables in the analysed data set.

Variable level's label	Variable level's description	Variable level's label	Variable level's description
X_1^1	Private car, driver's age under 25	X_1^2	Not driving to work
		X_2^2	Driving to work < 10 mil
		X_3^2	Driving to work > 10 mil
X_2^1	Private or business	X_4^2	Occasional driver
		X_5^2	Driving mostly on business
X_3^1	Married male, under 25	X_6^2	Age 18–20
		X_7^2	Age 21–24
X_4^1	Unmarried male, under 25	X_8^2	Age 18
		X_9^2	Age 19–20
		X_{10}^2	Age 21–22
		X_{11}^2	Age 23–24
X_5^1	Female	X_{12}^2	Age under 20
		X_{13}^2	Age 21–24

Table 1: Rating variables

Having defined the rating variables and the notations of their levels in Table 1, we present the data used in the study (Table 2).

The simulation procedure is performed for two types of the average claim distribution: the Gamma distribution with the density function $f(x,k,\theta) = \frac{1}{\theta^k \Gamma(\alpha)} x^{k-1} e^{-\frac{x}{\theta}}$ and the Inverse Gaussian distribution with the density function $f(x,\mu,\lambda) = \sqrt{\frac{\lambda}{2\prod x^3}} \exp(\frac{-\lambda(x-\mu)^2}{2\mu^2 x})$. The two main characteristics of the Gamma distribution are: $E(X) = k\theta$, $Var(X) = \frac{\mu^2}{k}$, while for the Inverse Gaussian distribution: $E(X) = \mu$, $Var(X) = \frac{\mu^3}{\lambda}$.

The method chosen for the *a'priori* rate making was the Genaralized Linear Model (GLM). The analysis based on the Monte Carlo simulation is conducted for every distribution independently. In the Monte Carlo simulation we estimate the parameters $\beta_1, ..., \beta_{18}$ for the rating

Values of claims

	X(1,1)	X(1,2)	X(1,3)	X(1,4)	X(1,5)
X(2,1)	155,4	268,4	227,3	229,3	532,1
X(2,2)	170,1	303,5	392,7	463,4	712,3
X(2,3)	185,8	280,2	320,2	331,1	797,7
X(2,4)	100,6	102,5	146,9	197,1	167,5
X(2,5)	205,5	514,1	445,1	299,7	340,3
X(2,6)	114,7	815,0	445,8	1 619,8	437,5
X(2,7)	243,7	324,5	548,2	250,7	1 509,5
X(2,8)	405,2	835,3	775,9	757,6	700,0
X(2,9)	270,2	531,1	527,1	1 238,2	1 414,9
X(2,10)	320,4	490,8	1 109,9	396,1	771,1
X(2,11)	233,9	351,4	379,7	614,2	680,6
X(2,12)	258,4	287,3	315,1	373,9	529,1
X(2,13)	197,2	247,3	499,2	296,4	356,6

Numbers of claims (weights)

	X(1,1)	X(1,2)	X(1,3)	X(1,4)	X(1,5)
X(2,1)	1 032 596	69 952	7 176	6 531	7 531
X(2,2)	908 551	92 324	12 630	11 138	8 376
X(2,3)	171 145	22 770	2 333	2 275	2 115
X(2,4)	22 509	67 929	7 527	8 865	4 315
X(2,5)	101 962	13 586	1 177	1 214	3 025
X(2,6)	238	1 471	118	119	57
X(2,7)	22 395	7 768	890	682	397
X(2,8)	439	6 876	1 448	1 096	516
X(2,9)	2 406	17 515	1 421	1 112	874
X(2,10)	25 362	16 827	1 756	1 420	950
X(2,11)	37 145	11 345	1 201	981	648
X(2,12)	2 374	17 957	2 447	1 738	900
X(2,13)	50 032	18 679	2 121	1 669	905

Table 2: Values and numbers of claims (weights) for every combination of levels of the rating variables

variables using the link function $g(x) = \ln(x)$ and assuming the error has the Gamma distribution. As a result, we obtain estimators of 16 variables. Two variables X_1^1 and X_1^2 were excluded automatically by the algorithm, because vectors associated with these variables were linearly dependent with other vectors in data matrix X (aliasing). After repeating the iteration 1000 times we obtain the distribution of the parameters for the rating variables. The deciles of the distribution are presented in Table 3 (for Gamma distribution assumed for average value of the claims) and Table 4 (for Inverse Gaussian distribution assumed for average value of the claims).

		X(2,1)	X(2,2)	X(2,3)	X(2,4)	X(2,5)	X(2,6)	X(2,7)	X(2,8)	X(2,9)	X(2,10)	X(2,11)	X(2,12)	X(2,13)	X(1,1)	X(1,2)	X(1,3)	X(1,4)	X(1,5)
Decile	Const.	B(1)	B(2)	B(3)	B(4)	B(5)	B(6)	B(7)	B(8)	B(9)	B(10)	B(11)	B(12)	B(13)	B(14)	B(15)	B(16)	B(17)	B(18)
0%	5,04	0,000	0,100	0,161	-0,774	0,310	0,958	0,390	1,036	0,727	0,690	0,373	0,122	0,158	0,000	0,488	0,699	0,795	1,163
10%	5,04	0,000	0,103	0,166	-0,768	0,319	1,006	0,399	1,059	0,742	0,701	0,381	0,138	0,165	0,000	0,492	0,707	0,803	1,178
20%	5,04	0,000	0,103	0,167	-0,766	0,320	1,016	0,401	1,064	0,745	0,703	0,383	0,140	0,167	0,000	0,493	0,709	0,806	1,180
30%	5,05	0,000	0,104	0,168	-0,765	0,321	1,023	0,403	1,067	0,747	0,705	0,385	0,143	0,168	0,000	0,494	0,710	0,808	1,182
40%	5,05	0,000	0,104	0,168	-0,764	0,322	1,029	0,405	1,070	0,749	0,706	0,386	0,145	0,169	0,000	0,494	0,712	0,809	1,184
50%	5,05	0,000	0,104	0,169	-0,763	0,322	1,035	0,407	1,072	0,751	0,707	0,387	0,146	0,170	0,000	0,495	0,713	0,810	1,186
60%	5,05	0,000	0,105	0,169	-0,762	0,323	1,042	0,408	1,075	0,753	0,708	0,388	0,148	0,171	0,000	0,495	0,715	0,812	1,187
70%	5,05	0,000	0,105	0,170	-0,761	0,324	1,049	0,410	1,077	0,755	0,709	0,389	0,150	0,172	0,000	0,496	0,716	0,813	1,189
80%	5,05	0,000	0,106	0,171	-0,760	0,325	1,056	0,411	1,080	0,757	0,711	0,391	0,152	0,173	0,000	0,496	0,718	0,815	1,190
90%	5,05	0,000	0,106	0,172	-0,759	0,326	1,067	0,413	1,085	0,760	0,713	0,393	0,155	0,175	0,000	0,497	0,720	0,818	1,193
100%	5,05	0,000	0,108	0,176	-0,752	0,332	1,101	0,423	1,102	0,779	0,720	0,401	0,167	0,182	0,000	0,501	0,728	0,825	1,206

Table 3: Distribution of the parameters $\beta_1, ..., \beta_{18}$ for Gamma distribution of claims

We observe that in both tables (3 and 4) the statistical range is relatively small.

As the result of the Monte Carlo simulation procedure we obtain the values of the MSE – distribution selection criterion presented in Table 7.

We observe that the value of the MSE is lower for the Gamma distribution. Thus the *a'priori* rating system should be created based on GLM model assuming Gamma distribution of the claims.

		X(2,1)	X(2,2)	X(2,3)	X(2,4)	X(2,5)	X(2,6)	X(2,7)	X(2,8)	X(2,9)	X(2,10)	X(2,11)	X(2,12)	X(2,13)	X(1,1)	X(1,2)	X(1,3)	X(1,4)	X(1,5)
Decile	**Const.**	**B(1)**	**B(2)**	**B(3)**	**B(4)**	**B(5)**	**B(6)**	**B(7)**	**B(8)**	**B(9)**	**B(10)**	**B(11)**	**B(12)**	**B(13)**	**B(14)**	**B(15)**	**B(16)**	**B(17)**	**B(18)**
0%	5,035	0,000	0,089	0,136	-0,803	0,284	0,652	0,319	0,903	0,622	0,639	0,329	0,030	0,122	0,000	0,466	0,639	0,738	1,097
10%	5,041	0,000	0,098	0,157	-0,778	0,308	0,858	0,373	1,003	0,707	0,680	0,366	0,113	0,150	0,000	0,484	0,685	0,778	1,146
20%	5,043	0,000	0,100	0,161	-0,773	0,312	0,918	0,385	1,023	0,721	0,689	0,374	0,124	0,157	0,000	0,488	0,695	0,788	1,158
30%	5,044	0,000	0,102	0,164	-0,769	0,316	0,961	0,394	1,040	0,732	0,695	0,379	0,133	0,163	0,000	0,490	0,702	0,797	1,169
40%	5,045	0,000	0,104	0,167	-0,766	0,320	0,997	0,401	1,055	0,741	0,701	0,383	0,140	0,167	0,000	0,492	0,708	0,805	1,176
50%	5,046	0,000	0,105	0,169	-0,763	0,322	1,026	0,407	1,067	0,749	0,707	0,388	0,147	0,170	0,000	0,495	0,713	0,811	1,184
60%	5,047	0,000	0,106	0,171	-0,760	0,325	1,060	0,412	1,081	0,757	0,712	0,392	0,153	0,174	0,000	0,497	0,719	0,817	1,192
70%	5,048	0,000	0,107	0,173	-0,757	0,329	1,095	0,419	1,098	0,767	0,718	0,397	0,161	0,178	0,000	0,499	0,724	0,824	1,200
80%	5,049	0,000	0,109	0,177	-0,752	0,333	1,137	0,427	1,114	0,778	0,725	0,403	0,170	0,183	0,000	0,502	0,732	0,833	1,210
90%	5,050	0,000	0,111	0,181	-0,747	0,338	1,193	0,436	1,137	0,793	0,736	0,411	0,180	0,189	0,000	0,505	0,742	0,842	1,224
100%	5,058	0,000	0,127	0,196	-0,726	0,357	1,535	0,475	1,228	0,872	0,769	0,441	0,238	0,219	0,000	0,520	0,782	0,883	1,280

Table 4: Distribution of the parameters $\beta_1, ..., \beta_{18}$ for Inverse Gaussian distribution of claims

	Min					Median					Max				
	X(1,1)	X(1,2)	X(1,3)	X(1,4)	X(1,5)	X(1,1)	X(1,2)	X(1,3)	X(1,4)	X(1,5)	X(1,1)	X(1,2)	X(1,3)	X(1,4)	X(1,5)
X(2,1)	**1,00**	1,63	2,01	2,21	3,20	**1,00**	1,64	2,04	2,25	3,27	**1,00**	1,65	2,07	2,28	3,34
X(2,2)	1,11	1,80	2,22	2,45	3,54	1,11	1,82	2,27	2,50	3,63	1,11	1,84	2,31	2,54	3,72
X(2,3)	1,18	1,92	2,36	2,60	3,76	1,18	1,94	2,42	2,66	3,87	1,19	1,97	2,47	2,72	3,98
X(2,4)	0,46	0,75	0,93	1,02	1,48	0,47	0,76	0,95	1,05	1,53	0,47	0,78	0,98	1,08	1,57
X(2,5)	1,36	2,22	2,74	3,02	4,36	1,38	2,26	2,82	3,10	4,52	1,39	2,30	2,89	3,18	4,66
X(2,6)	2,61	4,25	5,24	5,77	8,33	2,82	4,62	5,74	6,33	9,21	3,01	4,96	6,23	6,86	10,04
X(2,7)	1,48	2,41	2,97	3,27	4,72	1,50	2,46	3,06	3,38	4,91	1,53	2,52	3,16	3,48	5,10
X(2,8)	2,82	4,59	5,67	6,24	9,02	2,92	4,79	5,96	6,57	9,56	3,01	4,97	6,24	6,87	10,05
X(2,9)	2,07	3,37	4,16	4,58	6,62	2,12	3,47	4,32	4,77	6,93	2,18	3,60	4,51	4,97	7,28
X(2,10)	1,99	3,25	4,01	4,41	6,38	2,03	3,32	4,14	4,56	6,64	2,06	3,39	4,26	4,69	6,86
X(2,11)	1,45	2,37	2,92	3,22	4,65	1,47	2,42	3,01	3,31	4,82	1,49	2,46	3,09	3,41	4,98
X(2,12)	1,13	1,84	2,27	2,50	3,61	1,16	1,90	2,36	2,60	3,79	1,18	1,95	2,45	2,70	3,95
X(2,13)	1,17	1,91	2,36	2,59	3,75	1,19	1,94	2,42	2,67	3,88	1,20	1,98	2,48	2,74	4,00

Table 5: Stochastic rating table (min, median, max) for Gamma distribution of claims

	Min					Median					Max				
	X(1,1)	X(1,2)	X(1,3)	X(1,4)	X(1,5)	X(1,1)	X(1,2)	X(1,3)	X(1,4)	X(1,5)	X(1,1)	X(1,2)	X(1,3)	X(1,4)	X(1,5)
X(2,1)	**1,00**	1,59	1,90	2,09	3,00	**1,00**	1,64	2,04	2,25	3,27	**1,00**	1,68	2,19	2,42	3,60
X(2,2)	1,09	1,74	2,07	2,29	3,28	1,11	1,82	2,27	2,50	3,63	1,13	1,91	2,48	2,74	4,08
X(2,3)	1,15	1,83	2,17	2,40	3,43	1,18	1,94	2,42	2,66	3,87	1,22	2,05	2,66	2,94	4,37
X(2,4)	0,45	0,71	0,85	0,94	1,34	0,47	0,76	0,95	1,05	1,52	0,48	0,81	1,06	1,17	1,74
X(2,5)	1,33	2,12	2,52	2,78	3,98	1,38	2,26	2,82	3,10	4,51	1,43	2,41	3,13	3,46	5,14
X(2,6)	1,92	3,06	3,64	4,01	5,75	2,79	4,58	5,69	6,28	9,11	4,64	7,81	10,15	11,22	16,69
X(2,7)	1,38	2,19	2,61	2,88	4,12	1,50	2,46	3,06	3,38	4,90	1,61	2,71	3,52	3,89	5,78
X(2,8)	2,47	3,93	4,68	5,16	7,39	2,91	4,77	5,93	6,54	9,50	3,42	5,75	7,47	8,26	12,28
X(2,9)	1,86	2,97	3,53	3,90	5,58	2,12	3,47	4,31	4,76	6,91	2,39	4,03	5,23	5,78	8,60
X(2,10)	1,90	3,02	3,59	3,96	5,68	2,03	3,33	4,14	4,56	6,62	2,16	3,63	4,71	5,21	7,76
X(2,11)	1,39	2,21	2,63	2,91	4,16	1,47	2,42	3,01	3,32	4,81	1,55	2,61	3,40	3,76	5,59
X(2,12)	1,03	1,64	1,95	2,15	3,09	1,16	1,90	2,36	2,61	3,78	1,27	2,13	2,77	3,07	4,56
X(2,13)	1,13	1,80	2,14	2,36	3,39	1,19	1,95	2,42	2,67	3,87	1,24	2,09	2,72	3,01	4,47

Table 6: Stochastic rating table (min, median, max) for Inverse Gaussian distribution of claims

	MSE
Gamma distribution	1285,84
Inverse Gaussian distribution	1312,04

Table 7: Values of the MSE – distribution selection criterion

6 Summary

The procedure presented in the paper supports the decision about which distribution of claims to choose when calculating the rating table used later for calculating the level of the premium in casualty insurances. The procedure assures that the level of the premium is more likely to be fair and reasonable, better reflecting the dependencies between the rating variables and the sum of the claims in different groups (classes) of insured people. The presented procedure can be applied even if it is hard to estimate the distribution of the claims based on the historical data, e.g. when the insurance company does not have the complete information about the claims experience in a given portfolio, in particular when the available data sets lack or do not isolate certain rating variables. The advantage of the presented approach is that it is very flexible. It offers the possibility of modifying different elements of the analysis, e.g. selecting different types of distributions, changing the rating method or taking into consideration more distribution selection criteria. The weakness of the procedure might appear when introducing too many rating variables to the model, because large number of obtained results may cause the analysis hard to conduct.

Bibliography

[1] BROWN R. J., *Minimum Bias with Generalized Linear Models*, PCAS LXXV, p. 187-268, 1988.

[2] FU L., MONCHER R. B., *Severity Distributions for GLMs: Gamma or Lognormal? Evidence from Monte Carlo Simulations*. Casualty Actuarial Society Discussion Paper Program Casualty Actuarial Society - Arlington, Virginia 2004.

[3] LINDSEY J. K., *Applying Generalized Linear Models*. Springer – Verlag New York, 1997.

[4] McCullagh P., Nelder J. A., *Generalized Linear Models*. Chapman & Hall/CRC, New York, 1999.

[5] OSTASIEWICZ W. (red.), Premiums and insurance risk. *Stochastic modelling*. (in Polish) Wrocław University of Economics Publishing House, Wrocław, 2004.

[6] ŚLIWIŃSKI A., Insurance risk – premium's calculation and optimization. (in Polish) Poltex, Warszawa 2002.

Wei-Bin Zhang

College of Asia Pacific Management
Ritsumeikan Asia Pacific University
Japan

e-mail:wbz1 @apu.ac.jp

Interregional Migration and Economic Growth in a Multi-Region Multi-Group Economy—Spatial Amenity, Productivity, and Agglomeration

Abstract

It is obvious that there are close interactions among interregional migration and economic growth and environment. Nevertheless, economic theories have failed to provided an integrated theory to address these interactions within a compact framework. The purpose of this paper is to examine dynamic interactions among interregional migration, economic development and environmental change on microeconomic foundation. Our approaches are based on some well-established theories in growth theory and regional economics. The model describes multi-regional economic growth with heterogeneous households, and capital accumulation under assumptions of profit maximization, utility maximization, and perfect competition. First, we simulate the national economy of a three-regional economy with homogeneous population. It is demonstrated that when the production functions take on the Cobb-Douglas form and the parameters are properly specified, the dynamics has a unique equilibrium. Then we carry out comparative statics analysis with regard to regions' productivity and amenity levels, the propensity to save, and the population. The simulation results show that any change in any region's amenity or productivity may improve or worsen some regions with regard to some variables. We will also examine the dynamics of a two-region two-group economy.

Keywords: *two-region growth model, capital accumulation, amenity, heterogeneous households*

1 Introduction

Regional dynamics of capital accumulation with profit-maximization and utility maximization are so difficult that few serious attempts have been made. One of reasons for the lacking of theoretical interest is that the traditional approaches to consumer behavior will result in such a complicated dynamic system that the behavior of the system becomes intractable. This paper builds a dynamic one-commodity and multiple-region trade model to examine interdependence between regional trades and national growth. We analyze trade issues within the framework of a simple international macroeconomic growth model with perfect capital mobility. This model is influenced by the neoclassical trade theory with capital accumulation. Since the publication of Oniki and Uzawa's paper on theory of trade and economic growth [1], various trade models with endogenous capital have been proposed [e.g., 2, 3, 4, 5]. It is well known that dynamic-optimization models with capital accumulation are associated with analytical difficulties. To avoid these difficulties, this study applies an alternative approach to consumer behavior. This paper is organized as follows. Section 2 defines the multi-region model with heterogeneous households and capital accumulation. Section 3 simulates the motion of the 3-region national economy with a homogeneous population. Section 4 examines the effects of changes in different regions' productivity levels.

2 The multi-region trade model with capital accumulation

In describing economic production, we follow the neoclassical trade framework. It is assumed that the regions produce a homogenous commodity. Most aspects of production sectors in our model are similar to the neo-classical one-sector growth model [e.g., 6, 7]. This paper extends a two-regional model by Zhang [8] to any number of regions. The system consists of J regions, indexed by $j = 1, ...J$. Perfect competition is assumed to prevail in good markets both within each region and between the regions, and commodities are traded without any barriers such as transport costs or tariffs. The labor markets are perfectly competitive within each region and among regions. Let prices be measured in terms of the commodity and the price of the commodity be unity. We assume that the national population is classified into 2 groups, indexed by q.

A group q in region j is indexed by (j, q). We introduce

$$Q^* \equiv \{(j, q) \,|\, j = 1, \dots, J, q = 1, 2\}.$$

Let the number of group q in region j be $N_{jq}(t)$. The aggregated labor force, $N_j(t)$, of region j is given by

$$N_j(t) = \sum_{q=1}^{Q} z_q N_{jq}(t), \ (j, q) \in Q^*, \tag{1}$$

where z_q are the level of human capital of group q. In this initial stage, we assume human capital to be constant. We denote wage rates and interest rate by $w_{jq}(t)$, $(j, q) \in Q^*$, and $r(t)$.

Let $\bar{N}_q$ L_j stand for, respectively, the fixed population of group q and the fixed (residential) area of region j. The neoclassical production functions are given by, $F_j(K_j(t), N_j(t))$, $j = 1, \dots J$, where $K_j(t)$ and $N_j(t)$ are the levels of capital stocks employed by region j's production sector. We have

$$f_j(t) = f_j(k_j(t)), \ f_j(t) \equiv \frac{F_j(t)}{N_j(t)}, k_j(t) \equiv \frac{K_j(t)}{N_j(t)}.$$

Markets are competitive; thus labor and capital earn their marginal products, and firms earn zero profits. For any individual firm $r(t)$ and $w_{jq}(t)$ are given. Region j's production sector chooses two variables, $K_j(t)$ and $N_j(t)$, to maximize its profit. The marginal conditions are given by

$$r + \delta_{kj} = f_j'(k_j), \ w_{jq}(t) = z_q \left(f_j(k_j) - k_j f_j'(k_j)\right), (j, q) \in Q^*, \tag{2}$$

where δ_{kj} are the depreciation rate of physical capital in region j .

Each worker may get income from land ownership, capital ownership and wages. In order to define incomes, it is necessary to determine land ownership structure. It can be seen that land properties may be distributed in multiple ways under various institutions. Let $\bar{k}_{jq}(t)$ stand for the per capita wealth of group q in region j . A consumer q of region j obtains income

$$y_{jq}(t) = r(t)\,\bar{k}_{jq}(t) + w_{jq}(t), (j,q) \in Q^* \tag{3}$$

from the interest payment, $r\bar{k}_{jq}$, and the wage payment, w_{jq} . We call y_{jq} the current income in the sense that it comes from the consumer's wage income and consumer's current earnings from ownership of wealth. The income that the consumer is using for consuming, saving, or transferring is not necessarily equal to the current income because the consumer can sell wealth to pay, for instance, the current consumption if the current income is not sufficient for buying food and touring the country. Retired people may live not only on the interest payment but also have to spend some of their wealth. The total value of wealth that a consumer of group j can sell to purchase goods and to save is equal to $\bar{k}_{jq}(t)$. Here, we assume that selling and buying wealth can be conducted instantaneously without any transaction cost. The disposable income is equal to

$$\hat{y}_{jq}(t) = y_{jq}(t) + \bar{k}_{jq}(t)\ , (j,q) \in Q^* . \tag{4}$$

The disposable income is used for saving and consumption. A consumer distributes the total available budget among lot size, $l_{jq}(t)$, saving, $s_{jq}(t)$, consumption of goods, $c_{jq}(t)$. The budget constraints are

$$R_{jq}(t)\,l_{jq}(t) + c_{jq}(t) + s_{jq}(t) = \hat{y}_{jq}(t) = (1 + r(t))\,\bar{k}_{jq}(t) + w_{jq}(t)\ . \tag{5}$$

where $R_j(t)$is region j's land rent. Equation (5) means that housing, consumption and saving exhaust the consumer's disposable personal income.

We assume that utility level, $U_j(t)$, that the consumers obtain is dependent on the housing, $l_{jq}(t)$, the consumption level of commodity, $c_{jq}(t)$, and saving $s_{jq}(t)$. We specify $U_{jq}(t)$, $(j,q) \in Q^*$, as follows

$$U_{jq}(t) = \theta_{jq} l_{jq}^{\eta_{0q}}(t)\, c_{jq}^{\xi_{0q}}(t)\, s_{jq}^{\lambda_{0q}}(t)\ , \eta_{0q}\,, \xi_{0q}\,, \lambda_{0q} > 0\,, \tag{6}$$

in which $\eta_{0q}\,, \xi_{0q}$, and λ_{0q} are the elasticities of utility of a typical person of group q with regard to lot size, commodity and saving. We call $\eta_{0q}\,, \xi_{0q}$, and λ_{0q} propensities to consume lot size, to consume goods, and to hold wealth (save), respectively. Maximizing $U_j(t)$ subject to (5) yields

$$l_{jq}(t)\,R_j(t) = \eta_q \hat{y}_{jq}(t)\,,\ c_{jq}(t) = \xi_q \hat{y}_{jq}(t)\,,\ s_{jq}(t) = \lambda_q \hat{y}_{jq}(t)\,,\ (j,q) \in Q^*, \tag{7}$$

$$\eta_q \equiv \frac{\eta_{0q}}{\rho_q}\,, \xi_q \equiv \frac{\xi_{0q}}{\rho_q}\,, \lambda_q \equiv \frac{\lambda_{0q}}{\rho_q}\,, \rho_q \equiv \eta_{0q} + \xi_{0q} + \lambda_{0q}.$$

According to the definitions of $s_{jq}(t)$, the wealth accumulation of household $(j,q) \in Q^*$ is given by

$$\dot{\bar{k}}_{jq}(t) = s_{jq}(t) - \bar{k}_{jq}(t)\,, (j,q) \in Q^*\,. \tag{8}$$

As households are assumed to be freely mobile among the regions, the utility level of people of the same group should be equal, irrespective of in which region they live, i.e.

$$U_{jq}(t) = U_{mq}(t), j, m = 1, ... J\,, q = 1, 2\,. \tag{9}$$

The total capital stocks, $K(t)$, is equal to the total wealth owned by the population. That is

$$K(t) = \sum_{j=1}^{J} K_j(t) = \sum_{j=1}^{J}\sum_{q=1}^{2} \bar{k}_{jq}(t)\, N_{jq}. \tag{10}$$

The national production is equal to the national consumption and net savings. That is

$$C(t) + S(t) - K(t) + \sum_{j=1}^{J} \delta_{kj} K_j(t) = F(t)\,, \tag{11}$$

where

$$C(t) \equiv \sum_{j=1}^{J}\sum_{q=1}^{2} c_{jq}(t)\, N_{jq}\,, S(t) \equiv \sum_{j=1}^{J}\sum_{q=1}^{2} s_{jq}(t)\, N_{jq}\,, F(t) \equiv \sum_{j=1}^{J} F_j(t)\ .$$

The assumption that labor force and land are fully employed is represented by

$$\sum_{j=1}^{J} N_{jq}(t) = \bar{N}_q, \sum_{q=1}^{2} l_{jq}(t)\, N_{jq}(t) = L_j, j = 1, ...J\,. \tag{12}$$

We have thus built the model which explains the endogenous capital accumulation and regional capital and labor distribution in the national economy in which all the markets perfectly competitive.

3 Simulate a 3-region model with a homogenous population

We are first concerned with a three-region model with a homogeneous population Hence, we omit index q. We also require $\eta + \xi + \lambda = 1$ without affecting our analysis. We specify the production

$$F_j(t) = A_j K_j^{\alpha_j}(t)\, N_j^{\beta_j}(t)\ , \alpha_j + \beta_j = 1, \alpha_j\,, \beta_j > 0\,, j = 1\,, \cdots, J\,,$$

where A_j is region j's productivity. From $f_j = A_j k_j^{\alpha_j}$, where $k_j \equiv K_j/N_j$,and equations (2), we have

$$k_j = \phi_j\left(k_1\right) \equiv \left(\frac{\alpha_1 A_1 k_1^{-\beta_1} - \delta_j}{\alpha_j A_j}\right)^{-1/\beta_j}, j = 2, \cdots, J.$$

By equations (2) and (4)

$$w_j = \bar{\phi}_j \equiv A_j \beta_j \phi_j^{\alpha_j}\left(k_1\right), \hat{y}_j = \tilde{f}\left(k_1\right) \bar{k}_j + A_j \beta_j \phi_j^{\alpha_j}\left(k_1\right), j = 1, \cdots, J, \tag{13}$$

where we use $\phi_1\left(k_1\right) = k_1$ and $\tilde{f}\left(k_1\right) = \alpha_1 A_1 k_1^{-\beta_1} + \delta_1$. Substitute $c_j = \xi \hat{y}_j$ and $s_j = \lambda \hat{y}_j$ in equations (7) and $l_j = L_j / N_j$ into the utility function

$$U_j(t) = \theta_j L_j^{\eta} \xi^{\xi} \lambda^{\lambda} N_j^{-\eta} \hat{y}_j^{\xi+\lambda}, j = 1, \cdots, J.$$

Inserting the above equations into equations (9), we get

$$n_j\left(t\right) = \tilde{\theta}_j \left(\frac{\hat{y}_j}{\hat{y}_1}\right)^{\bar{\eta}}, j = 2, \cdots, J, \tag{14}$$

where

$$n_j\left(t\right) \equiv \frac{N_j\left(t\right)}{N_1\left(t\right)}, \tilde{\theta}_j \equiv \left(\frac{\theta_j L_j^{\eta}}{\theta_1 L_1^{\eta}}\right)^{1/\eta}, \bar{\eta} \equiv \frac{\xi + \lambda}{\eta}.$$

We can rewrite equation (10) as

$$\sum_{j=1}^{J} k_j\left(t\right) N_j\left(t\right) = \sum_{j=1}^{J} \bar{k}_j\left(t\right) N_j\left(t\right).$$

Insert $k_j = \phi_j\left(k_1\right)$ into the above equation

$$\bar{k}_1\left(t\right) = k_1 + \sum_{j=2}^{J} n_j \left(\phi_j\left(k_1\right) - \bar{k}_j\right). \tag{15}$$

Insert equations (14) into equation (15)

$$\bar{k}_1\left(t\right) = k_1 + \frac{1}{\hat{y}_1^{\bar{\eta}}} \sum_{j=2}^{J} \tilde{\theta}_j \hat{y}_j^{\bar{\eta}} \left(\phi_j\left(k_1\right) - \bar{k}_j\right). \tag{16}$$

Solve equation (13) for $\hat{y}_1$ with regard to $\bar{k}_1$

$$\bar{k}_1 = \frac{\hat{y}_1 - \bar{\phi}_1\left(k_1\right)}{\tilde{f}\left(k_1\right)}. \tag{17}$$

Equaling both right-hand sides of equations (16) and (17), we have

$$\hat{y}_1^{1+\bar{\eta}} - \hat{f}\left(k_1\right) \hat{y}_1^{\bar{\eta}} - \bar{f}\left(k_1, \left\{\bar{k}\left(t\right)\right\}\right) = 0, \tag{18}$$

where

$$\hat{f}\left(k_1\right) \equiv \bar{\phi}_1\left(k_1\right) + k_1 \tilde{f}\left(k_1\right) = A_1 k_1^{\alpha_1} + \delta_1 k_1 > 0\,,$$

$$\bar{f}\left(k_1, \left\{\bar{k}\left(t\right)\right\}\right) \equiv \tilde{f}\left(k_1\right) \sum_{j=2}^{J} \tilde{\theta}_j \hat{y}_j^{\bar{\eta}} \left(\phi_j\left(k_1\right) - \bar{k}_j\right) =$$

$$= \left(\alpha_1 A_1 k_1^{-\beta_1} + \delta_1\right) \sum_{j=2}^{J} \tilde{\theta}_j \hat{y}_j^{\bar{\eta}} \left(\phi_j\left(k_1\right) - \bar{k}_j\right)\,. \tag{19}$$

It should be noted that the functions $\hat{f}$ and $\bar{f}$ are only dependent on J variables, k_1 and $\left\{\bar{k}\left(t\right)\right\}$. This is important as it allows us to express the dynamics in a J-dimensional differential equations system.

In the remainder of this study, we limit our study to the case of $\bar{\eta} = 2$. As $\xi + \lambda + \eta = 1$, we have $\eta = 1/3$. We specify this value as it enables us explicitly and uniquely solve equation (18) as follows

$$\hat{y}_1\left(k_1, \left\{\bar{k}_j\left(t\right)\right\}\right) = \Lambda_0\left(k_1, \left\{\bar{k}\left(t\right)\right\}\right) = \frac{\hat{f}}{3} + \frac{\sqrt[3]{2}\hat{f}^2}{3\tilde{F}} + \frac{\tilde{F}}{3\sqrt[3]{2}}\,, \tag{20}$$

where

$$\tilde{F}\left(k_1, \left\{\bar{k}_j\left(t\right)\right\}\right) \equiv \left[2\hat{f}^3 + 27\bar{f} + \sqrt{27\bar{f}\left(4\hat{f}^3 + 27\bar{f}\right)}\right]^{1/3}\,.$$

Substitute equation (20), $\bar{\phi}_1\left(k_1\right) = \alpha_1 A_1 k_1^{-\beta_1}$ and $\tilde{f}\left(k_1\right) = \alpha_1 A_1 k_1^{-\beta_1} + \delta_1$ into equation (16)

$$\bar{k}_1 = \Lambda\left(k_1, \left\{\bar{k}\left(t\right)\right\}\right) = \frac{\hat{f} + \sqrt[3]{2}\hat{f}^2/\tilde{F} + \tilde{F}/\sqrt[3]{2}}{3\alpha_1 A_1 k_1^{-\beta_1} + 3\delta_1} - \frac{\alpha_1 A_1}{\alpha_1 A_1 + \delta_1 k_1^{\beta_1}}\,. \tag{21}$$

According to the above analysis and equations (8), (21) and (13), we have

$$\dot{k}_1 = \Lambda_1\left(k_1, \left\{\bar{k}_j\right\}\right) =$$

$$= \frac{\left(\alpha_1 A_1 k_1^{-\beta_1} + \delta_1\right)\lambda - 1}{3\left(\alpha_1 A_1 k_1^{-\beta_1} + \delta_1\right)} \left(\hat{f} + \frac{\sqrt[3]{2}\hat{f}^2}{\tilde{F}} + \frac{\tilde{F}}{\sqrt[3]{2}}\right) + \frac{\alpha_1 A_1}{\alpha_1 A_1 + \delta_1 k_1^{\beta_1}}, \tag{22}$$

$$\dot{\bar{k}}_j = \Lambda_j\left(k_1,\ \bar{k}_j\right) = \left(\alpha_1 \lambda A_1 k_1^{-\beta_j} - \bar{\lambda}\right)\bar{k}_j + \lambda A_j \beta_j \phi_j^{\alpha_j}\left(k_1\right),\ j = 2, \cdots, J. \tag{23}$$

Taking derivatives of equation (21) with respect to t yields

$$\dot{\bar{k}}_1 = \frac{\partial \Lambda}{\partial k_1}\dot{k}_1 + \sum_{j=2}^{J} \Lambda_j \frac{\partial \Lambda}{\partial \bar{k}_j}. \tag{24}$$

Equaling two sizes of equations (22) and (24) yields

$$\dot{k}_1 = \frac{\Lambda_1\left(k_1, \{\bar{k}_j\}\right) - \sum_{j=2}^{J} \Lambda_j\left(k_1, \{\bar{k}_j\}\right) \partial\Lambda/\partial\bar{k}_j}{\partial\Lambda/\partial k_1}. \tag{25}$$

An equilibrium point is given by setting $\dot{k}_1 = 0$ and $\dot{k}_j = 0$ in equations (22) and (23). That is

$$\frac{\left(\alpha_1 A_1 k_1^{-\beta_1} + \delta_1\right)\lambda - 1}{3\left(\alpha_1 A_1 k_1^{-\beta_1} + \delta_1\right)}\left(\hat{f}(k_1) + \frac{\sqrt[3]{2}\hat{f}^2(k_1)}{\tilde{F}\left(k_1, \{\bar{k}_j\}\right)} + \frac{\tilde{F}\left(k_1, \{\bar{k}_j\}\right)}{\sqrt[3]{2}}\right) + \frac{\alpha_1 A_1}{\alpha_1 A_1 + \delta_1 k_1^{\beta_1}} = 0\,, \tag{26}$$

$$\left(\alpha_1 \lambda A_1 k_1^{-\beta_j} - \bar{\lambda}\right)\bar{k}_j + \lambda A_j \beta_j \phi_j^{\alpha_j}(k_1) = 0\,, j = 2, 3\,. \tag{27}$$

From equations (27), we solve

$$\bar{k}_j = \frac{-\lambda A_j \beta_j \phi_j^{\alpha_j}(k_1)}{\alpha_1 \lambda A_1 k_1^{-\beta_j} - \bar{\lambda}}\,, j = 2, 3\,. \tag{28}$$

Substitute equations (28) into equations (13) and (19)

$$\hat{y}_j(k_1) = \frac{-\lambda A_j \beta_j \tilde{f}(k_1)\phi_j^{\alpha_j}(k_1)}{\alpha_1 \lambda_j A_1 k_1^{-\beta_j} - \bar{\lambda}} + A_j \beta_j \phi_j^{\alpha_j}(k_1)\ , j = 2, 3\,,$$

$$\bar{f}(k_1) = \left(\alpha_1 A_1 k_1^{-\beta_1} + \delta_1\right)\sum_{j=2}^{J} \tilde{\theta}_j\left(\phi_j(k_1) - \bar{k}_j\right)\left[\frac{-\lambda A_j \beta_j \tilde{f}(k_1)\phi_j^{\alpha_j}(k_1)}{\alpha_1 \lambda A_1 k_1^{-\beta_j} - \bar{\lambda}} + A_j \beta_j \phi_j^{\alpha_j}(k_1)\right]^{\bar{\eta}}. \tag{29}$$

Substituting equations (19) and (29) into the definition of $\tilde{F}\left(k_1, \{\bar{k}_j(t)\}\right)$, we solve $\tilde{F}$ as a function of k_1

$$\tilde{F}(k_1) = \left[2\hat{f}^3(k_1) + 27\bar{f}(k_1) + \sqrt{27\bar{f}(k_1)\left(4\hat{f}^3(k_1) + 27\bar{f}(k_1)\right)}\right]^{1/3}. \tag{30}$$

Insert equation (30) to equation (26)

$$\Omega(k_1) \equiv \frac{\left(\alpha_1 A_1 k_1^{-\beta_1} + \delta_1\right)\lambda - 1}{3\left(\alpha_1 A_1 k_1^{-\beta_1} + \delta_1\right)}\left(\hat{f}(k_1) + \frac{\sqrt[3]{2}\hat{f}^2(k_1)}{\tilde{F}(k_1)} + \frac{\tilde{F}(k_1)}{\sqrt[3]{2}}\right) + \frac{\alpha_1 A_1}{\alpha_1 A_1 + \delta_1 k_1^{\beta_1}} = 0\,. \tag{31}$$

To simulate the model, we specify the parameter values as follows

$$\begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 1.2 \\ 1 \end{pmatrix}, \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} = \begin{pmatrix} 4 \\ 3.8 \\ 4.3 \end{pmatrix}, \begin{pmatrix} \eta \\ \xi \\ \lambda \end{pmatrix} = \begin{pmatrix} 1/3 \\ 2/9 \\ 4/9 \end{pmatrix},$$

$$\begin{pmatrix} L_1 \\ L_2 \\ L_3 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ 3.5 \end{pmatrix}, \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 0.3 \\ 0.3 \\ 0.3 \end{pmatrix}, \begin{pmatrix} \delta_{k1} \\ \delta_{k2} \\ \delta_{k3} \end{pmatrix} = \begin{pmatrix} 0.15 \\ 0.14 \\ 0.13 \end{pmatrix}, \tag{32}$$

and $N = 10$. The equation $\Omega(k_1) = 0$ has a unique meaningful solution given: $k_1 = 1.118$. From $k_1 = 1.118$ and Eqs. (28), we get the equilibrium values of $\bar{k}_2$ and $\bar{k}_3$. We get the equilibrium values of all the other variables. We list the simulation results as follows

$$F = 12.73, K = 9.06, r = 0.266,$$

$$\begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} = \begin{pmatrix} 1.551 \\ 1.116 \\ 0.852 \end{pmatrix}, \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix} = \begin{pmatrix} 1.117 \\ 0.786 \\ 0.586 \end{pmatrix}, \begin{pmatrix} N_1 \\ N_2 \\ N_3 \end{pmatrix} = \begin{pmatrix} 4.93 \\ 2.92 \\ 2.15 \end{pmatrix},$$

$$\begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix} = \begin{pmatrix} 7.643 \\ 3.257 \\ 1.835 \end{pmatrix}, \begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} = \begin{pmatrix} 2.719 \\ 1.159 \\ 0.653 \end{pmatrix},$$

$$\begin{pmatrix} \bar{k}_1 \\ \bar{k}_2 \\ \bar{k}_3 \end{pmatrix} = \begin{pmatrix} 1.104 \\ 0.794 \\ 0.606 \end{pmatrix}, \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} 1.086 \\ 0.781 \\ 0.596 \end{pmatrix}, \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 0.552 \\ 0.397 \\ 0.303 \end{pmatrix},$$

$$\begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix} = \begin{pmatrix} 0.609 \\ 1.371 \\ 1.625 \end{pmatrix}, \begin{pmatrix} R_1 \\ R_2 \\ R_3 \end{pmatrix} = \begin{pmatrix} 1.360 \\ 0.435 \\ 0.280 \end{pmatrix}. \tag{33}$$

The output, wealth, population, and wage income distribution among regions are given by

$$\begin{pmatrix} \hat{F}_1 \\ \hat{F}_2 \\ \hat{F}_3 \end{pmatrix} = \begin{pmatrix} 60.0\% \\ 25.6\% \\ 14.4\% \end{pmatrix}, \begin{pmatrix} \hat{K}_1 \\ \hat{K}_2 \\ \hat{K}_3 \end{pmatrix} = \begin{pmatrix} 60.0\% \\ 25.6\% \\ 14.4\% \end{pmatrix},$$

$$\begin{pmatrix} \hat{N}_1 \\ \hat{N}_2 \\ \hat{N}_3 \end{pmatrix} = \begin{pmatrix} 49.3\% \\ 29.2\% \\ 21.5\% \end{pmatrix}, \begin{pmatrix} \hat{W}_1 \\ \hat{W}_2 \\ \hat{W}_3 \end{pmatrix} = \begin{pmatrix} 60.0\% \\ 25.6\% \\ 14.4\% \end{pmatrix}, \tag{34}$$

where a variable x_j with circumflex, $\hat{x}_j$, denotes region j's share of the corresponding variable in the national economy. The shares of the output, capital stocks, population, and wage income of the advanced region in the national economy are respectively 60.0%, 60.0%, 49.3%, and 60.0%. The remote region's corresponding shares are respectively 14.4%, 14.4%, 21.5%, and 14.4%. As the population is homogeneous and people can move freely and costlessly, the remote region still attracts some people mainly because of its amenity and residential land availability.

We simulate the model with the parameter values specified as in (32) and $k_1(0) = 1.2$, $\bar{k}_2(0) = 0.75$, $\bar{k}_3(0) = 0.6$. The simulation results are plotted in Figure 1.

4 Parameter Changes and Regional Economic Structures

We increase the productivity level A_1 from 1.5 to 1.7. The simulation results are summarized below, where $\bar{\Delta}$ stand for the change rate of the equilibrium value of the variable in percentage.

$$A_1 : 1.5 \Rightarrow 1.7 \Rightarrow \bar{\Delta}F = 17.54, \bar{\Delta}K = 17.67, \bar{\Delta}r = 0.42,$$

$$\begin{pmatrix} \bar{\Delta}f_1 \\ \bar{\Delta}f_2 \\ \bar{\Delta}f_3 \end{pmatrix} = \begin{pmatrix} 19.44 \\ -0.11 \\ -0.11 \end{pmatrix}, \begin{pmatrix} \Delta k_1 \\ \Delta k_2 \\ \Delta k_3 \end{pmatrix} = \begin{pmatrix} 19.12 \\ -0.37 \\ -0.37 \end{pmatrix}, \begin{pmatrix} \bar{\Delta}N_1 \\ \bar{\Delta}N_2 \\ \bar{\Delta}N_3 \end{pmatrix} = \begin{pmatrix} 17.99 \\ -17.48 \\ -17.47 \end{pmatrix},$$

$$\begin{pmatrix} \bar{\Delta}F_1 \\ \bar{\Delta}F_2 \\ \bar{\Delta}F_3 \end{pmatrix} = \begin{pmatrix} 40.93 \\ -17.57 \\ -17.57 \end{pmatrix}, \begin{pmatrix} \bar{\Delta}\bar{k}_1 \\ \bar{\Delta}\bar{k}_2 \\ \bar{\Delta}\bar{k}_3 \end{pmatrix} = \begin{pmatrix} \bar{\Delta}c_1 \\ \bar{\Delta}c_2 \\ \bar{\Delta}c_3 \end{pmatrix} = \begin{pmatrix} 19.58 \\ 0.00 \\ 0.00 \end{pmatrix},$$

$$\begin{pmatrix} \bar{\Delta}w_1 \\ \bar{\Delta}w_2 \\ \bar{\Delta}w_3 \end{pmatrix} = \begin{pmatrix} 19.44 \\ -0.11 \\ -0.11 \end{pmatrix}, \begin{pmatrix} \bar{\Delta}l_1 \\ \bar{\Delta}l_2 \\ \bar{\Delta}l_3 \end{pmatrix} = \begin{pmatrix} -15.25 \\ 21.18 \\ 20.18 \end{pmatrix},$$

$$\begin{pmatrix} \bar{\Delta}R_1 \\ \bar{\Delta}R_2 \\ \bar{\Delta}R_3 \end{pmatrix} = \begin{pmatrix} 41.09 \\ -17.48 \\ -17.48 \end{pmatrix}, \begin{pmatrix} \bar{\Delta}\hat{F}_1 \\ \bar{\Delta}\hat{F}_2 \\ \bar{\Delta}\hat{F}_3 \end{pmatrix} = \begin{pmatrix} 19.90 \\ -29.87 \\ -29.87 \end{pmatrix},$$

$$\begin{pmatrix} \bar{\Delta}E_1 \\ \bar{\Delta}E_2 \\ \bar{\Delta}E_3 \end{pmatrix} = \begin{pmatrix} 1.68 \\ 10.42 \\ -8.71 \end{pmatrix}. \tag{35}$$

We summarize the effects of change in the remote region's productivity in (35).

$$A_3 : 1 \Rightarrow 1.2 \Rightarrow \bar{\Delta}F = 2.25, \bar{\Delta}K = 2.46, \bar{\Delta}r = -1.11,$$

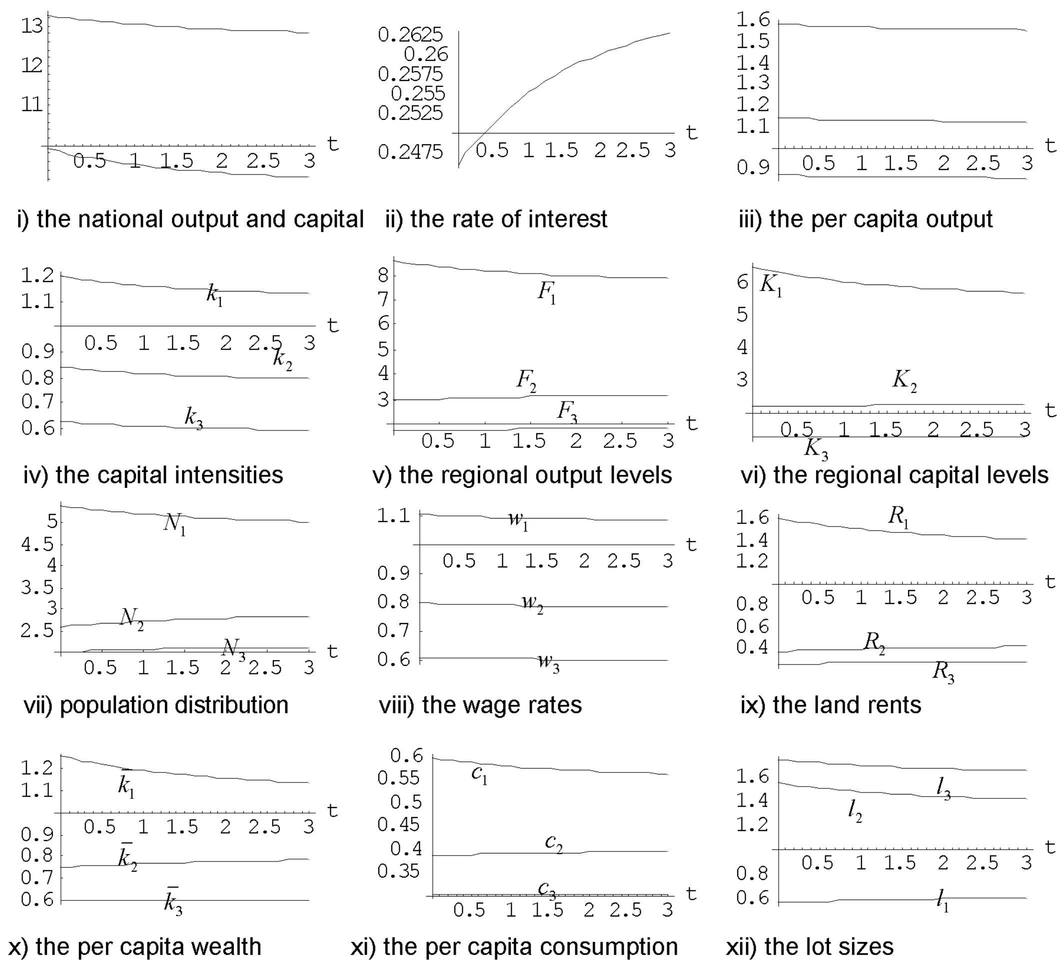


Figure 1: The Motion of the National Economy

$$\begin{pmatrix} \bar{\Delta} f_1 \\ \bar{\Delta} f_2 \\ \bar{\Delta} f_3 \end{pmatrix} = \begin{pmatrix} 0.31 \\ 0.30 \\ 30.1 \end{pmatrix}, \begin{pmatrix} \Delta k_1 \\ \Delta k_2 \\ \Delta k_3 \end{pmatrix} = \begin{pmatrix} 1.03 \\ 1.00 \\ 31.0 \end{pmatrix}, \begin{pmatrix} \bar{\Delta} N_1 \\ \bar{\Delta} N_2 \\ \bar{\Delta} N_3 \end{pmatrix} = \begin{pmatrix} -13.39 \\ -12.41 \\ 47.45 \end{pmatrix},$$

$$\begin{pmatrix} \bar{\Delta} F_1 \\ \bar{\Delta} F_2 \\ \bar{\Delta} F_3 \end{pmatrix} = \begin{pmatrix} -13.13 \\ -12.16 \\ 91.88 \end{pmatrix}, \begin{pmatrix} \bar{\Delta} \bar{k}_1 \\ \bar{\Delta} \bar{k}_2 \\ \bar{\Delta} \bar{k}_3 \end{pmatrix} = \begin{pmatrix} 0.01 \\ 1.02 \\ 31.07 \end{pmatrix}, \begin{pmatrix} \bar{\Delta} w_1 \\ \bar{\Delta} w_2 \\ \bar{\Delta} w_3 \end{pmatrix} = \begin{pmatrix} 0.31 \\ 0.30 \\ 30.13 \end{pmatrix},$$

$$\begin{pmatrix} \bar{\Delta} c_1 \\ \bar{\Delta} c_2 \\ \bar{\Delta} c_3 \end{pmatrix} = \begin{pmatrix} 0.01 \\ 0.57 \\ 30.48 \end{pmatrix}, \begin{pmatrix} \bar{\Delta} l_1 \\ \bar{\Delta} l_2 \\ \bar{\Delta} l_3 \end{pmatrix} = \begin{pmatrix} 15.46 \\ 14.16 \\ -32.18 \end{pmatrix},$$

$$\begin{pmatrix} \bar{\Delta} R_1 \\ \bar{\Delta} R_2 \\ \bar{\Delta} R_3 \end{pmatrix} = \begin{pmatrix} -13.39 \\ -11.91 \\ 92.40 \end{pmatrix}, \begin{pmatrix} \bar{\Delta} \hat{F}_1 \\ \bar{\Delta} \hat{F}_2 \\ \bar{\Delta} \hat{F}_3 \end{pmatrix} = \begin{pmatrix} -15.04 \\ -14.08 \\ 87.65 \end{pmatrix},$$

$$\begin{pmatrix} \bar{\Delta} E_1 \\ \bar{\Delta} E_2 \\ \bar{\Delta} E_3 \end{pmatrix} = \begin{pmatrix} -56.70 \\ -9.86 \\ 95.30 \end{pmatrix}.$$

References

[1] Oniki, H. and Uzawa, H. (1965) Patterns of Trade and Investment in a Dynamic Model of International Trade. *Review of Economic Studies* 32, 15-38.

[2] Eaton, J. (1987) A Dynamic Specific-Factors Model of International Trade. *Review of Economic Studies* 54, 325-38.

[3] Findlay, R. (1984) Growth and Development in Trade Models. In Jones, R.W., Kenen, R.B. (Eds.): *Handbook of International Economics.* Amsterdam: North-Holland.

[4] Ikeda, S. and Ono, Y. (1992) Macroeconomic Dynamics in a Multi-Country Economy - A Dynamic Optimization Approach. *International Economic Review* 33, 629-44.

[5] Zhang, W.B. (2008) *International Trade Theory: Capital, Knowledge, Economic Structure, Money and Prices over Time and Space.* Berlin: Springer.

[6] Burmeister, E. and Dobell, A.R. (1970) *Mathematical Theories of Economic Growth.* London: Collier Macmillan Publishers.

[7] Zhang, W.B. (2005) *Economic Growth Theory.* Hampshire: Ashgate.

[8] Zhang, W.B. (1998) A Two-Region Growth Model - Competition, References, Resources, and Amenities. *Papers in Regional Science* 77, 173-188.

Název:	**Proceedings of 26$^{\text{th}}$ International Conference "Mathematical Methods in Economics 2008"**
Autoři:	Kolektiv autorů
Vydavatel:	Technická univerzita v Liberci
Určeno:	pro vědecké pracovníky, účastníky mezinárodní konference, odbornou veřejnost
Schváleno:	Rektorátem TU v Liberci, č. j. RE115/08
Vyšlo:	v září 2008
Počet stran:	532
Vydání:	první
Tiskárna:	Vysokoškolský podnik, s.r.o., Liberec
Číslo publikace:	55-092-08
ISBN:	978-80-7372-387-3