Statistical analysis of unemployed in regard to the time series models and prediction accuracy for future periods

Štatistická analýza vývoja počtu nezamestnaných v súvislosti s modelmi časových radov a presnosťou predikcie budúcich období

Marek Andrejkovič, Stela Beslerová

Abstract:
The article is appointed to the predictive models. Creation of predictive models is based on the simple linear regression where the predicted number of unemployed is counted from the time based on the time series regression model. We evaluate models by determination index.

Key words:
Unemployment. time series. predictive models.

JEL classification: E24

1 Introduction

Tracking the development of the unemployment rate is currently, due to the turbulent conditions, very important. However predictions of such macroeconomic indicators are somewhat problematic. Usage of prediction models is possible only with respect to other factors that enter into these variables and therefore by simple regression models based on the selected time series is not possible to properly define these procedures. Therefore, in this paper we deal with the problems of time series prediction, which may be affected by external factors, which yet cannot be identified at the time, or their identification is difficult and depends on expert estimates.

Such situation is also represented, for example, by the arrival of the economic crisis. In these times there is a problem in prediction of future development of macroeconomic indicators, as the system of economy and relations between entities varies. For this reason, we highlight the ways that can be used for prediction of time series, despite such external factors.

2 Theoretical assumptions and the current state of the issue

In theory, we encounter several definitions of unemployment. This paper is based on the knowledge published by authors (Mareš, 2002; Hazlitt, 2010, Kennedy, 2012), who dealt with this issue. Unemployment is monitored on the basis of the statistical survey in the population aged 15 to 62 years. In doing so, we divide the population into three main groups, namely:

- employed,
- unemployed,
- others.

On this basis, we can define the group of employed persons and those who at the time do not work directly, but are on leave, strike or are physically indisposed, while having valid employment or carrying on business under a trade license. On the other side there are unemployed people, who are able to work but are not working currently. One of the conditions for inclusion in this group is an active effort to find job by intrinsic activity. This group also includes people under ILO methodology. (Statistical Office of the Slovak Republic, 2011).

To measure unemployment we can use the unemployment rate. The unemployment rate is a calculation where the number of unemployed is divided by the total number of persons included in the labor force, i.e. the number of persons who are able to work. (Martincová, 2005).

In the Slovak Republic we define two ways of determining the number of unemployed. In this case, we describe both methodologies.

The first analyzed is the methodology of the Statistical Office of the Slovak Republic. In this case, the number of unemployed is determined through a sample survey. Statistical Office carried out a survey in 200 cities and towns in which acquires sufficient sample, to be representative, across the whole Slovakia. Subsequently, is these selected cities is performed selection of 10 thousand households, in which survey is performed. This represents half of one percent of all real estate in Slovakia, which should ensure a sufficient sample. Each household thereby can be examined by a maximum of five consecutive quarters, which means that the period must be less than 1.5 of the calendar year. The Statistical Office focuses only on persons older than 15 years. (Statistical Office of the Slovak Republic, 2011).

The second methodology is the methodology of the Ministry of Labor, Social Affairs and Family. This methodology is based on real data monitored through the Offices of Labor, Social Affairs and Family. These authorities know the number of active jobseekers in the Labor offices at employment departments. In this case, the residents who are not registered at labor offices do not enter the system of unemployed and therefore are invisible in this methodology. Thus, there may be a reduction in the rate of unemployment by the formal administrative actions if there is decommissioning of job seekers at labor offices in the case of formal non-fulfillment of registration requirements. Thus, it is possible to positively influence the statistics of unemployment with no real impact on the economy and increase in employment. (Statistical Office of the Slovak Republic, 2011).

For analysis of the development of the unemployed and creation of forecasts several methods and procedures are used. Usually are used regression models that create conditions to define predictions for subsequent periods due to other factors or under the development at the time. It is mostly represented by the use of time-series analysis. We consider the linear regression models. (Tkáč, 2001).

$$y = \beta_0 + \beta_1 \cdot t \tag{1}$$

Another option of statistical analysis is usage of correlation analysis to identify relationships between indicators and individual states. These relationships can then be used to identify the recommendations for the future and also for the creation of regression models, which are not only based on time series, but also on the spatial panel data. In this case, we clearly use the Pearson correlation coefficient that verifies the existence of a linear dependence between variables. For its calculation we use the following procedure mentioned in the publication

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} \tag{2}$$

For the general validity of the mentioned models we use verification of linear model through Fisher's test of significance as well as individual regression coefficients verified through t-test. Accordingly, we can define the null hypothesis, which are at each test verified (falsified). (see Tkáč, 2001)

- for Fisher test of model significance - $H_0 : \beta_i = 0,\ \forall i, i \in \{0,1,...,n\}$

- for t-test of significance of *i* regression coefficient - $H_0 : \beta_i = 0$
- for significance test of correlation coefficient - $H_0 : \rho = 0$

3 Analysis of the number of unemployed

For tracking we picked a uniform methodology to track the number of unemployed under examination determined by the Statistical Office of the Slovak Republic, which is then transmitted to Eurostat statistics. On this basis, it monitors the development of the number of unemployed, while the unemployment rate is not an appropriate indicator, whereas in the Slovak Republic at the time there is no significant difference in the number of workforce and so economic impacts are therefore appropriate to be determined by the number of unemployed, as it is possible to estimate the costs of state budget for social support, etc. directly, without conversion of relative frequencies. Therefore, we focus predictive models just on the number of unemployed and not on the unemployment rate. Of course when comparing several countries it would not be possible to use this indicator, but would be necessary to defined relative rate, i.e. the unemployment rate.

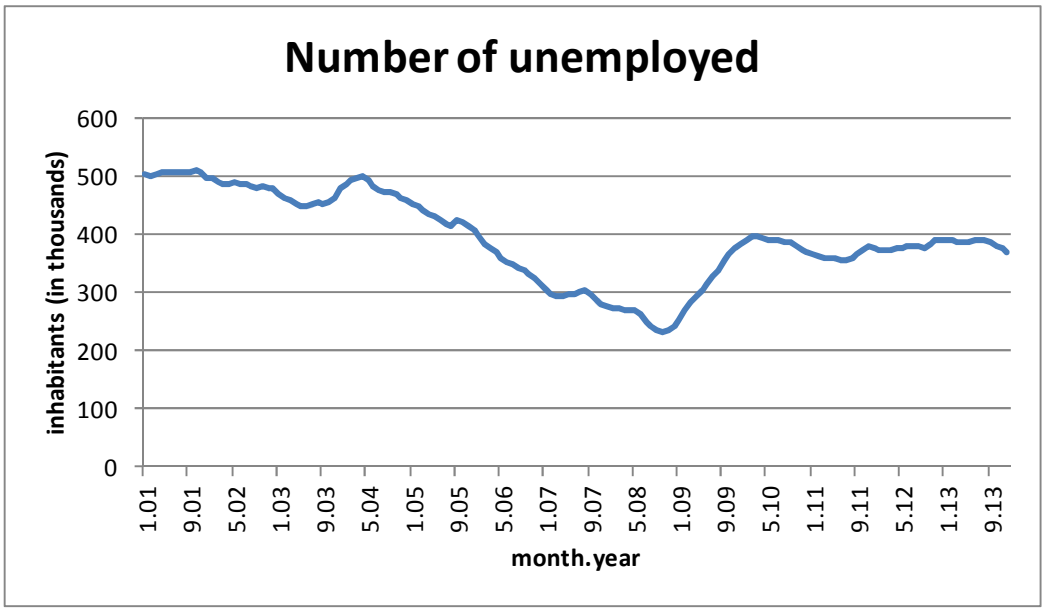

Fig. 1 Development of unemployed
Source: own processing according to EUROSTAT

On this basis, we can follow the development of the number of unemployed in Slovakia from January 2001 until December 2013. During the period, we can see a significant jump in the number of unemployed in the period of 2008 and early 2009. Increase in the number of unemployed stopped at the beginning of 2010.

4 Predictive regression models

In making the prediction models we considered recommendations coming from a theoretical knowledge defined in publications (e.g. Tkáč, 2001). Overall, we are dealing with four prediction models that are used in practice. On this basis, we can define their peer evaluation. The evaluation of these models is realized on the basis of determination index.

Model O1

First prediction model used is the model based on the entire base of the previous data that are used to predict the next time data. Thus, it can be written as follows:

$$y_t = b_{0,t} + b_{1,t} \cdot t \tag{3}$$

Mentioned regression model is created from data base, which includes a plurality of measurements, which can be identified as follows

$$Y_{t-1} = \{y_1; y_2; ...; y_{t-1}\} \tag{4}$$

We are creating a regression model based on all available data from the beginning of the study period. This increases the range of data set and specifies the estimate, but there can be also seen number of historical influences, which currently may not be valid. On this basis, we are at any given time *t* calculated the value of regression coefficients and based on that defined estimate at a given time *t*.

Estimate of regression function in this case was calculated from the measurement no. 73 and thus the previous six years were used as a base. On this basis, we can see the progress of regression estimate, which is based on the assumption of a linear regression model. We can see that changes in development trends are reflected very problematically.

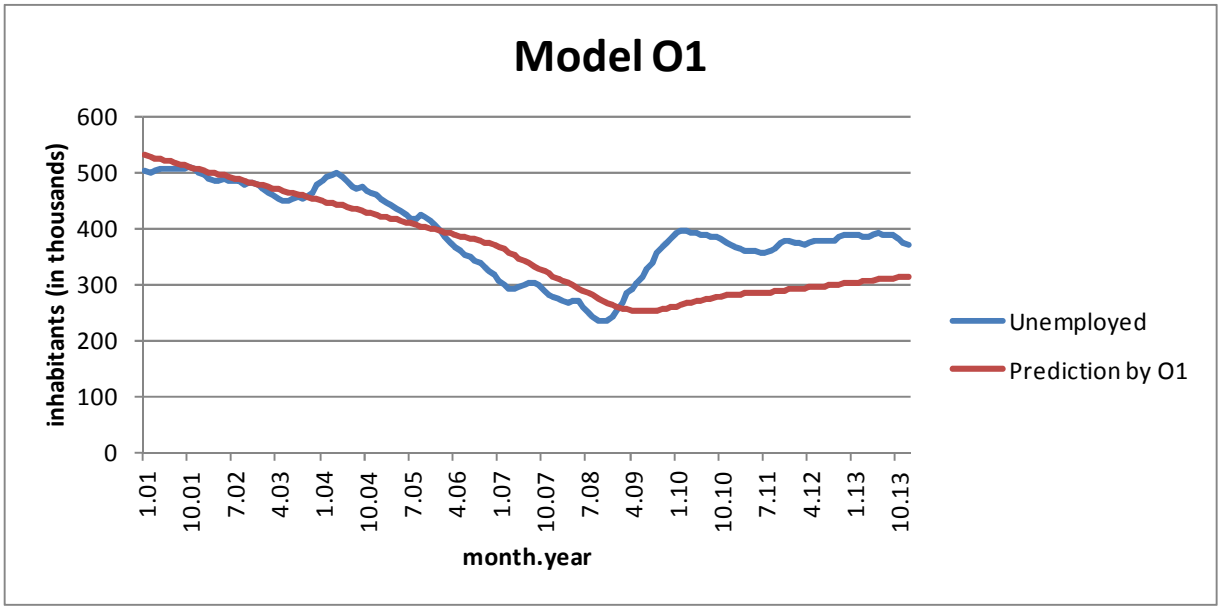

Fig. 2 Use of prediction model O1

Source: own processing

Based on change in development that occurred in 2009, we can see that from that time is given prediction model significantly inaccurate. This was caused by the change represented by economic crisis, which wiped out information content of previous data and since that time there is a distortion of estimates.

Model O2

Second prediction model that we used is the model based on 12 previous data, which are used to predict the next time data

$$Y_{t-1} = \{y_{t-13}; y_{t-12}; ...; y_{t-1}\} \tag{5}$$

Thus we are creating a regression model based on the 12 latest available data from that point. Although there is no increase in the stability of the model with respect to the number of

measurements, but it ensures that the impact of factors is transmitted and historical influences are thus eliminated.

On this basis, we can see the progress of regression estimate, which is based on the assumption of a linear regression model. We can see that changes in development trends are reflected relatively quickly.

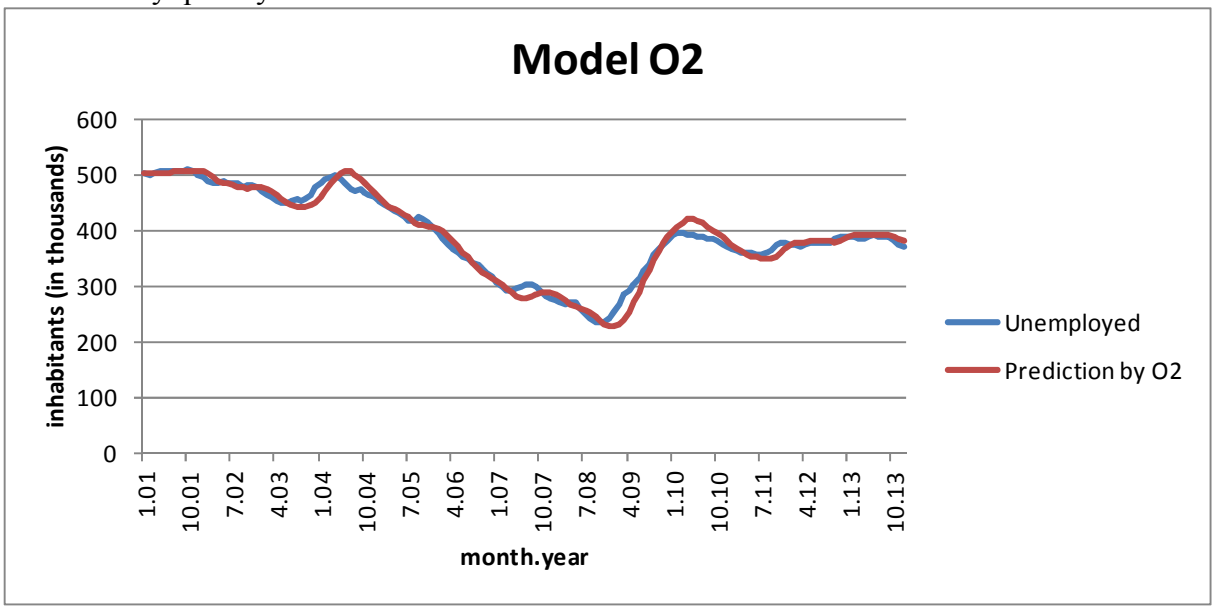

Fig. 3 Use of prediction model O2

Source: own processing

Model O3

The third prediction model that we used is the model created based on moving averages of 12 previous base data used to predict the next time stamp.

$$Y_{t-1} = \{y_{t-13}; y_{t-12}; \ldots; y_{t-1}\} \tag{6}$$

This model is therefore not necessarily the actual regression model, but it can be used especially in cases where there is not a linear development of indicators or in case of indicators that tend to oscillate in certain periods.

On this basis, we can see the progress of regression estimate. We can see that changes in development trends are reflected with a lag, while avoiding the "memory" effect in the data.

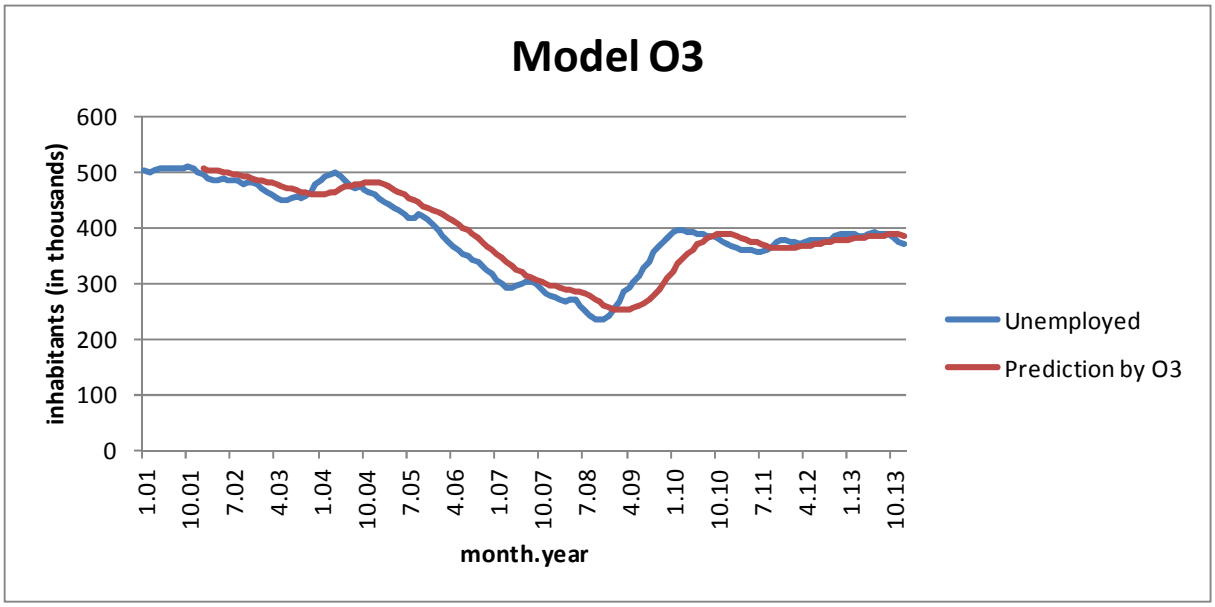

Fig. 4 Use of prediction model O3
Source: own processing

Model O4

Another prediction model that we used is the model based on previous 24 data which are used to predict the next time stamp.

$$Y_{t-1} = \{y_{t-25}; y_{t-24}; \ldots; y_{t-1}\} \tag{7}$$

Thus we are creating a regression model based on the latest 24 available data from that point. Although there is no increase in the stability of the model with respect to the number of measurements, but it ensures that the impact of factors is transmitted and historical influences are thus eliminated.

On this basis, we can see the progress of regression estimate, based on the assumption of a linear regression model. We can see that changes in development trends are reflected relatively quickly.

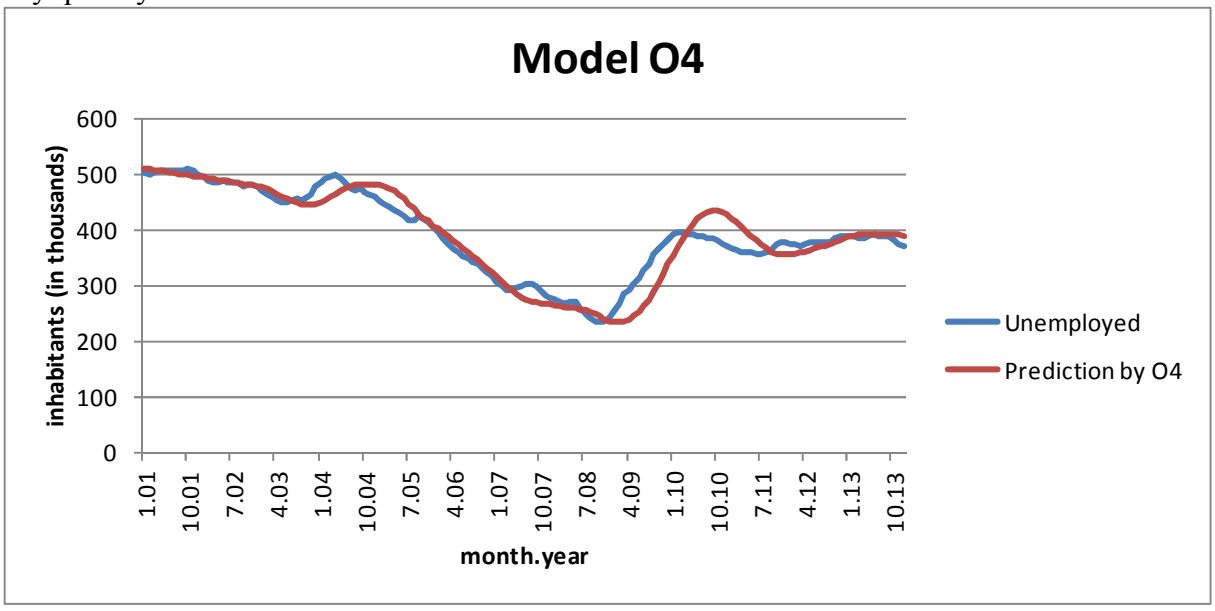

Fig. 5 Use of prediction model O4
Source: own processing

We can see that in this case, there is a greater delay in the change of prediction trends as in the case of the model O2. We can also see that there is some increase of certain influence, which can be seen for example in the turn of 2010 and 2011, where the model significantly overestimates the reality.

5 Evaluation of the models

The evaluation of the models is realized through the index of determination. We may yet see a reconciliation of these models in the following figure. We observe that there is a significant diversion of the first model, which uses all available data base. In this case the value of determination index is significantly lower than in models that track only a relatively short time series.

Overall, the most appropriate model is the one that only tracks 12 most recent data available and from them creates a prediction. In the case of an increase to 24 we can observe decrease in the value of determination index. If you use a moving average, despite the theoretical problems in an overvaluation or undervaluation of the data, the result is relatively good and comparable to the regression models.

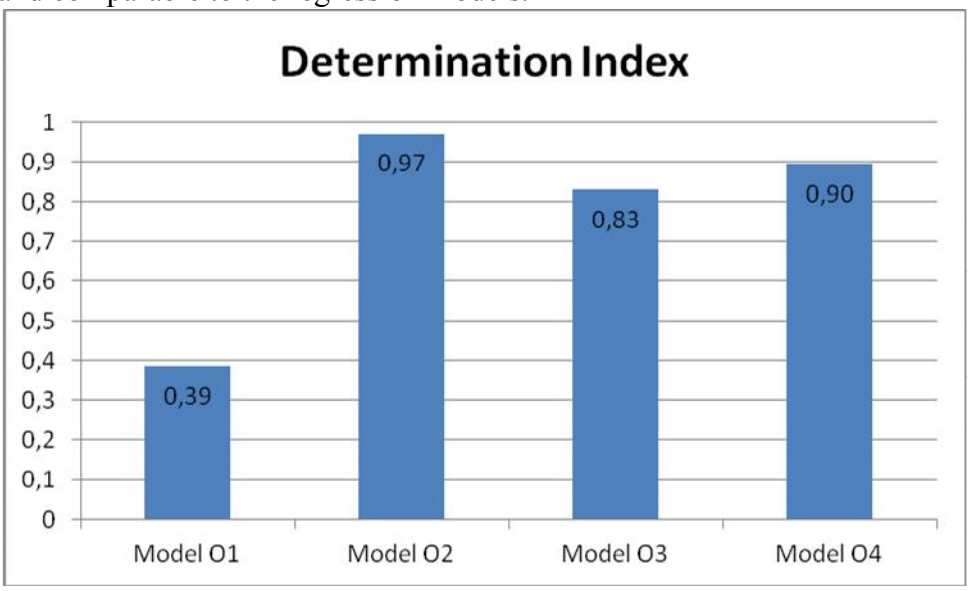

Fig. 6 Determination indices of models
Source: own processing

Overall, on the basis of these results we can conclude that the use of the entire data base is not in favor of the predictive power of analytics and is preferable to use only the last set of data which are not distorted by historical influences which are no longer valid.

6 Conclusion

During the analysis of prediction models in this paper we conclude that the most appropriate model is the formation of the linear regression model of the relatively short time series within a year. In the case of the enlargement of the series, we identified reduction of reliability of the estimate. In the case of the use of long time series from some historical value we recorded significantly worse results. We can therefore say that the use of predictions based on moving regression models is in terms of determination index the most appropriate.

Literature

ALBERT, M. – HAHNEL, R. 2003. *A QUIET Revolution in Welfare Economics.* [online]. London : Public Enterprise Market Economies, 2003. 586 p. [cit. 20.3.2012]. Dostupné na internete: <http://books.zcommunications.org/books/glossary.htm>
BARTO, M. a kol. 2002. *Ekonomické dopady vstupu Slovenska do Európskej únie.* Bratislava : Konzervatívny inštitút Milana Rastislava Štefánika. 2002. 72 s. ISBN – 80-89121-02-0
BAUMOHL, B. 2012. The Secrets of Economic Indicators: Hidden Clues to Future Economic Trends and Investment Opportunities. New York : FT Press, 2012. ISBN 978-0-13-293209-7.
EUROPEAN COMMISION. 2010. A strategy for smart, sustainable and inclusive growth. Brussels. 2010. 34 s.
EUROSTAT. 2014. *Unemployment by sex and age groups – monthly averege*. [online]. 2014. [cit. 2014-01-30]. Dostupné na internete: <http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=une_rt_m&lang=en>
HAZLITT, H. 2010. Economics in one lesson: The shortest and surest way to understand basic economics. Crown Publishing Group, 2010. 220 p. ISBN 0517548232.
KENNEDY, M. M. J. 2012. *Public Finance*. New Delhi : PHI Learning Private Limited, 2012. 352 p. ISBN 9788120345393.
MAREŠ, P. 2002. *Nezaměstnanost jako sociální problém*. Praha: Slon, 2002. 21s. ISBN 80-86429-08-3.
MARTINCOVÁ, M., 2005. *Nezamestnanosť ako makroekonomický problém*. Bratislava: Iura Edition. 2005, ISBN 80-8078-0382.
MIHALIK, J. 2002. *Trh práce a manažment ľudských zdrojov : personálny manažment*. Trenčín : Trenčianska univerzita, 2002. 391 s. ISBN 80-88914-50-7
ŠÚ SR. 2011. *Nezamestnanosť – metodika*. [online] 2011. [Cit. 2012.10.10.] Dostupné na: http://portal.statistics.sk/showdoc.do?docid=1938
TKÁČ, M. 2001. *Štatistické riadenie kvality*. Ekonóm : Bratislava, 2001.

Authors address:

Marek Andrejkovič, Ing. PhD.
University of Economics in Bratislava
Faculty of Business Economics with seat in Košice
Department of Quantitative Methods
Tajovského 13, 041 30 Košice
marek.andrejkovic@euke.sk

Stela Beslerová, Ing.
University of Economics in Bratislava
Faculty of Business Economics with seat in Košice
Department of Quantitative Methods
Tajovského 13, 041 30 Košice
stela.beslerova@euke.sk

Paper originated within the project VEGA 1/0519/12 - Business insurance as an essential part of strategic management during the debt crisis.