ANALYSIS OF CHANGES IN TRENDS OF FERTILITY TIME SERIES IN THE CZECH REPUBLIC: NORMALITY AND PRINCIPAL COMPONENT APPROACH

Ondřej Šimpach

Abstract

The aim of this paper is to analyse trend changes in long-term time series of age-specific fertility rates of females in the Czech Republic. Database from the Czech Statistical Office (CZSO) is available from 1925 to 2012, which means that state systems and political regimes in our society changed several during this period. We search when the distribution of age-specific fertility rates was normal, i.e. we use series of Jarque-Bera tests of normality for individual time series of age-specific fertility rates. Although it may seem that normality was not observed in the 70s of the last century – (due to significant pro-population arrangements of political regime), it is not true. The time series are not normal during first half of 80s only. Than the medians were analysed using principal components method, (which is implemented in “demography” package for the RStudio software). On the basis of four 20-years long time intervals from 1931–1950 to 1991–2010 we examine the changes in median fertility curves, which were significantly shifted to the right over the time. These results are important for subsequent analysis because for working with demographic data about fertility and constructing e.g. the fertility projection it is necessary to take into account the most recent data only, which are not significantly skewed and influenced by various factors.

Key words: age-specific fertility rates, normality, principal component method

JEL Code: C22, C32, J13

Introduction

Fertility is an important part of the natural population change. Modelling and estimating future fertility is in demography more complicated than e.g. in the case of mortality, because fertility is influenced by several factors (see e.g. Šimpach, Langhamrová, 2014b). The population development and improving the living standard in the country is closely related to postponement of first childbirth to the later age (see e.g. Bílková, 2009 or Šimpach, Pechrová, 2013) and together with the decline of number of live births in total. This decrease is below the level of simple reproduction of the population (2.08 children per 1 female within the reproduction period) in the Czech Republic. The database of Czech age-specific fertility rates recorded many significant changes from 1925 to 2012, because the development of these rates was affected by a wide range of social changes(e.g. by the 2ndWorld War, two parts of the Communist regime, pro-population policies in 70s and massive support of young families with children, and as well as the downturn of this development during the post-revolutionary period).

The aim of this paper isto explore time period when the distribution of age-specific fertility rates was normal using series of Jarque-Bera tests of normality for individual time series of age-specific fertility rates from 1925 to 2012. Our results reject the hypothesis claiming, that this distribution was not normal during the pro-population policies in 70s, and together we confirm, that the distribution of Czech fertility has normal distribution in these days. Fertility distribution loses its kurtosis and positive skewness. It becomes flat, with median almost equal to mean. This important information is necessary for the next analysis of demographic fertility data. Using the principal component approach[1], implemented in RStudio "demography" and "rainbow" package (Hyndman, Shang, 2009, Hyndman, 2012) we analyse the changes in the development of median fertility curves during four 20-years long time intervals, (which we set as 1931–1950, 1951–1970, 1971–1990 and 1991–2010).Last mentioned period has normal distribution and thus on the bases of the found median fertility curve we estimate the parameters of the distribution of this curve, which could be used instead of a_x parameter of the Lee-Carter model for predicting levels of fertility in future population studies.

1 Materials and Methods

The data were obtained from the Czech Statistical Office (CZSO) database. Using the numbers of live-born persons to$x$-year-old mothers in the year t(labelled $N_{x,t}$) and the numbers of mid-year females' x-year-old in time t (labelled $\bar{S}_{x,t}$), we establish the *BASE* (according to Hyndman, 2012) in RStudio software (R Development Core Team, 2008). The *BASE*is filled by age-specific fertility rates, particularly by these ratesin exponential expression as

[1]Principal component and analysis of main explaining factor is an important tool for researchers and therefore is used in many other studies, e.g. by Bartošová, Bína (2012) in the case of monetary poverty and households,Bartošová, Longford (2014) in the case of Czech Republic's income or Bílková (2012) for Czech wages and incomes according to highest level of education attained.

$$BASE \leftarrow f_{x,t} = \frac{N_{x,t}}{\bar{S}_{x,t}}, \text{ respectively } BASE \leftarrow \exp(f_{x,t}) = \exp\left(\frac{N_{x,t}}{\bar{S}_{x,t}}\right). \qquad (1)$$

These *BASEs* of rates are established due to the reason that this structure together with "demography" and "rainbow" package (Hyndman, Shang, 2009, Hyndman, 2012) is used for stochastic fertility modelling by Lee-Carter model (Lee, Carter, 1992, Shang, Hyndman, 2010)

$$f_{x,t} = a_x + b_x \cdot k_t + \varepsilon_{x,t}, \qquad (2)$$

Where a_x are the age-specific fertility profiles independent of time, b_x are the additional age-specific components determine how much the fertility in each age group changes when k_t changes and k_t are the time-varying parameters –fertility indices. $\varepsilon_{x,t}$ is an error term with the classical characteristics of white noise. Using the principal components (PC) method, (which can be acquired using the mentioned packages), it is possible to identify the outlying years of the analysed population. These is graphically displayed based on the results of the PC scores.

Shang, Hyndman (2010) introduced the functional and bivariate bag plots for clear visualization of the outliers in functional data. We use their approach to identify the outliers in the development of age-specific fertility rates. The functional and bivariate bag plots always contain two regions: dark and light grey. The first one contains 50% of all observations. There are also three curves –black curve shows median and 2 dashed lines are 95% confidence intervals. As noted by Shang, Hyndman (2010): "functional curves that are outside the border region are considered to be outliers". These curves are in functional bag plot shown in colour. The bivariate bag plot does not show the median curve, but Tukey depth median (see e.g. Tukey, 1975). Coloured points with year label outside the fence region are outliers.

According to the methodology elaborated by Jarque, Bera (1980) we are able to examine, whether the distribution is normal at the 5% significance level. We test the null hypothesis H_0: there is normal distribution, against the alternative H_1: non H_0. The test criterion

$$JB(TC) = \left(SK^2 + K^2\right) \approx \chi^2(\nu = 2), \qquad (3)$$

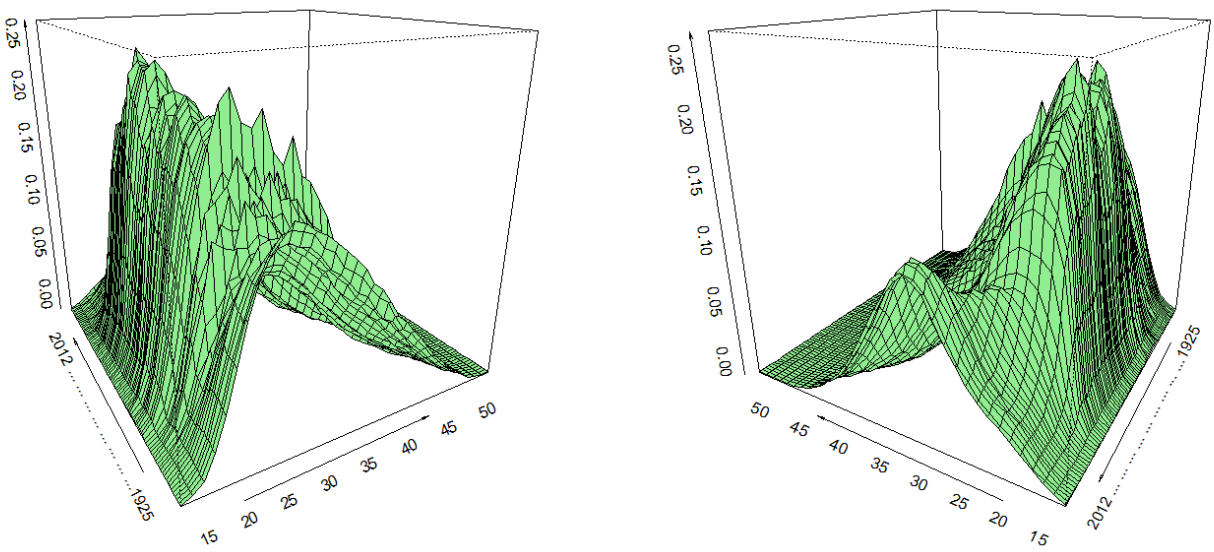
Where *SK* is skewness and *K* is kurtosis, has chi-square distribution with 2 degrees of freedom[2]. We reject the null hypothesis if $JB(TC) \geq \chi_{0,95}(\nu = 2)$.

[2] The presence of normality examined in empirical studies, e.g. also Löster, Pavelka (2013) in the case of clustering in real economic tasks, or Löster (2014), who dealt with the income inequality in the regions of the Czech Republic.

2 Results and Discussion

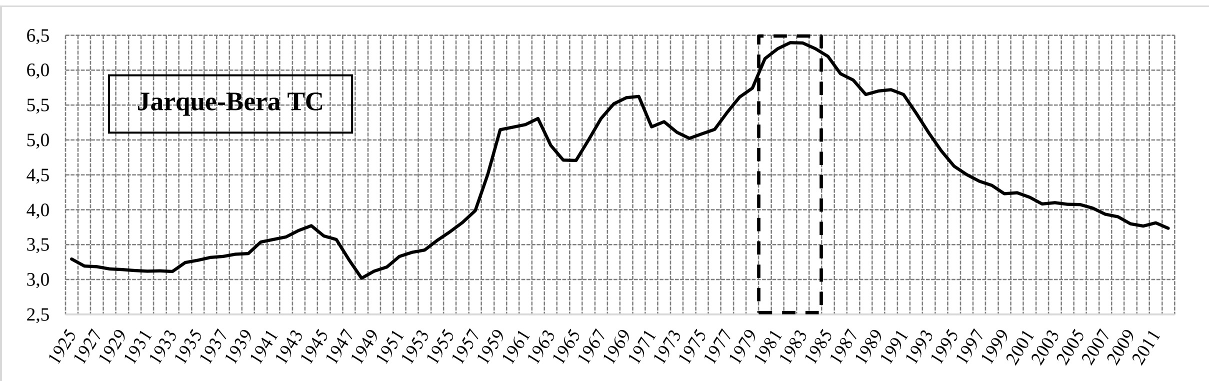
Empirical age-specific fertility rates $f_{x,t}$ of the Czech females in 1925–2012 can be seen in Fig. 1, where is evident the significant change of mode of this distribution to the advanced ages. We look on every single year as an individual time series and verify the null hypothesis of normality. Calculated Jarque-Bera TC(with estimated p-values) are shown in Fig. 2. In the case that the p-value is less than 0.05, the hypothesis of normality is rejected at the 5% significance level. This situation occurred in the years 1980–1985 (dashed line highlighted field).

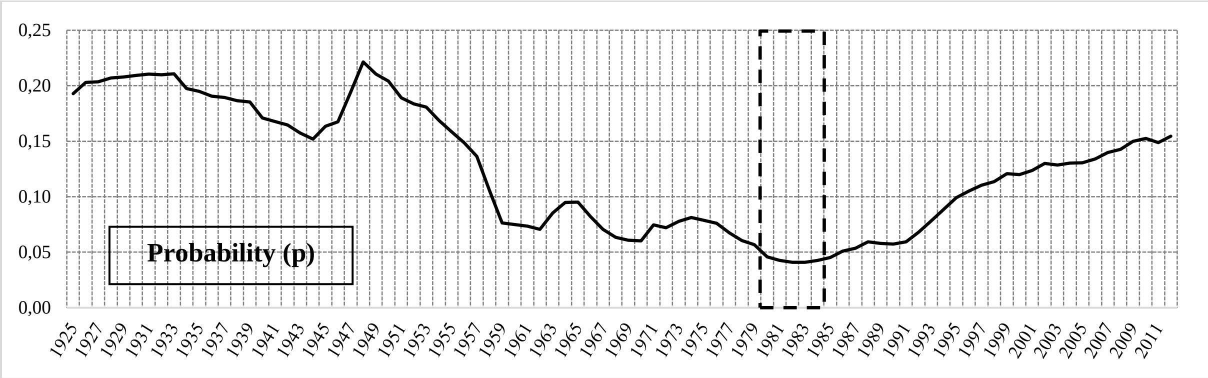
Fig. 1: Age-specific fertility rates of Czech females in 1925–2012 in 3D perspective charts



Source: data CZSO, author's calculation and illustration

Fig. 2: Jarque-Bera normality tests with development of probability values in 1925–2012

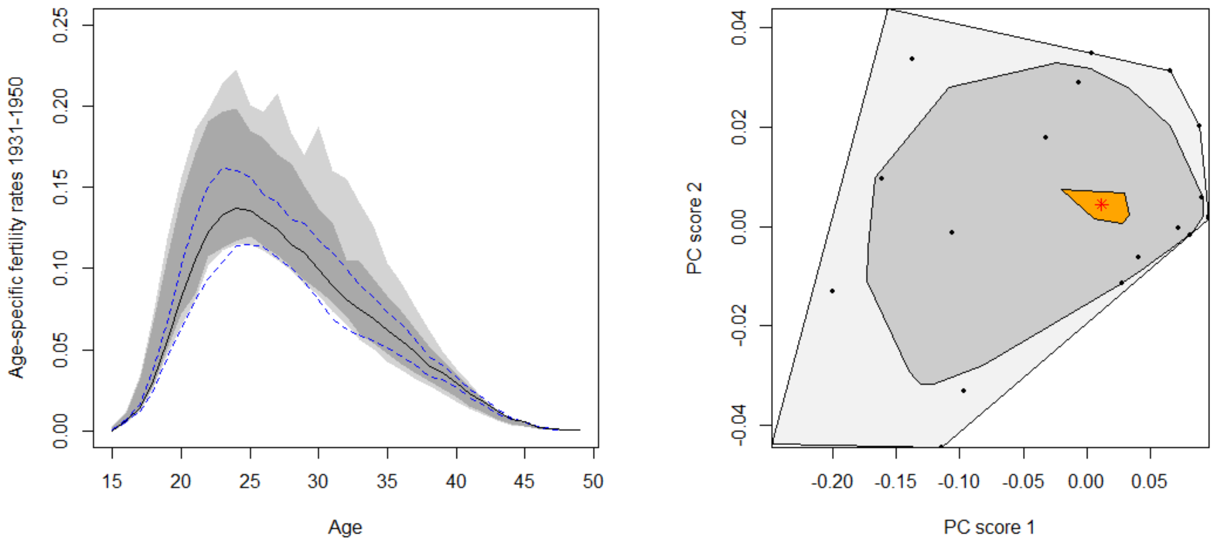


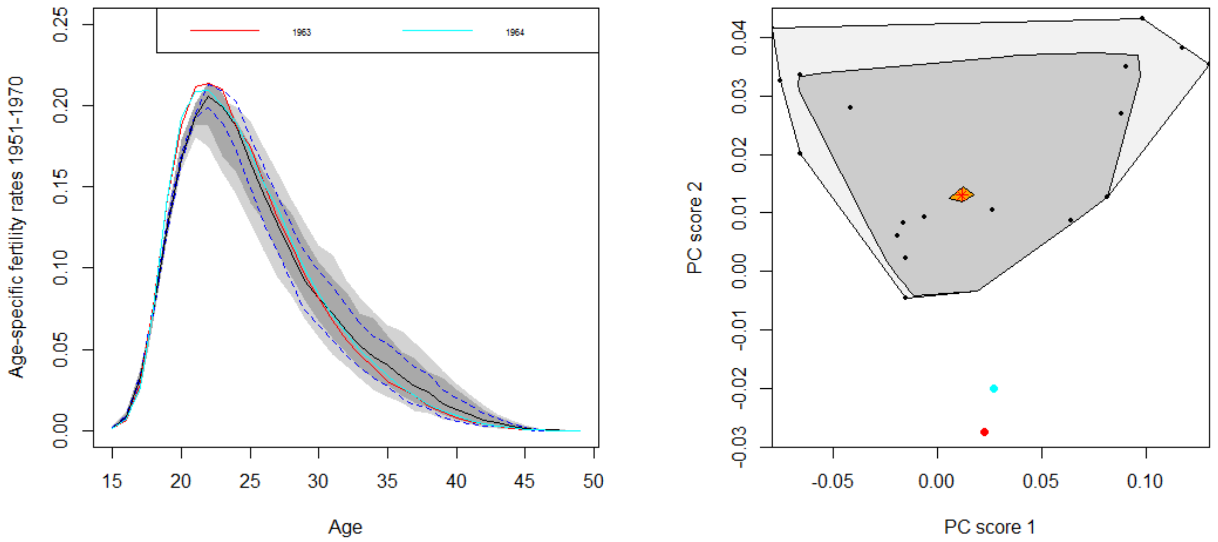


Source: author’s calculation and illustration

Apart from the mentioned period, the distribution of Czech fertility has approximately a normal distribution, for which could be estimated its parameters. Known parameters could help in the projection of age-specific fertility rates to the future because this distribution could be used as an estimate of the average profile of fertility by age independent of time (parameter $a_x$of the Lee-Carter model – see the empirical study and comparison results e.g. by Šimpach, Langhamrová, 2014a or Šimpach, Dotlačilová, Langhamrová, 2013).

Fig. 3: Median of Czech females’ age-specific fertility rates in1931–1950 and 1951–1970

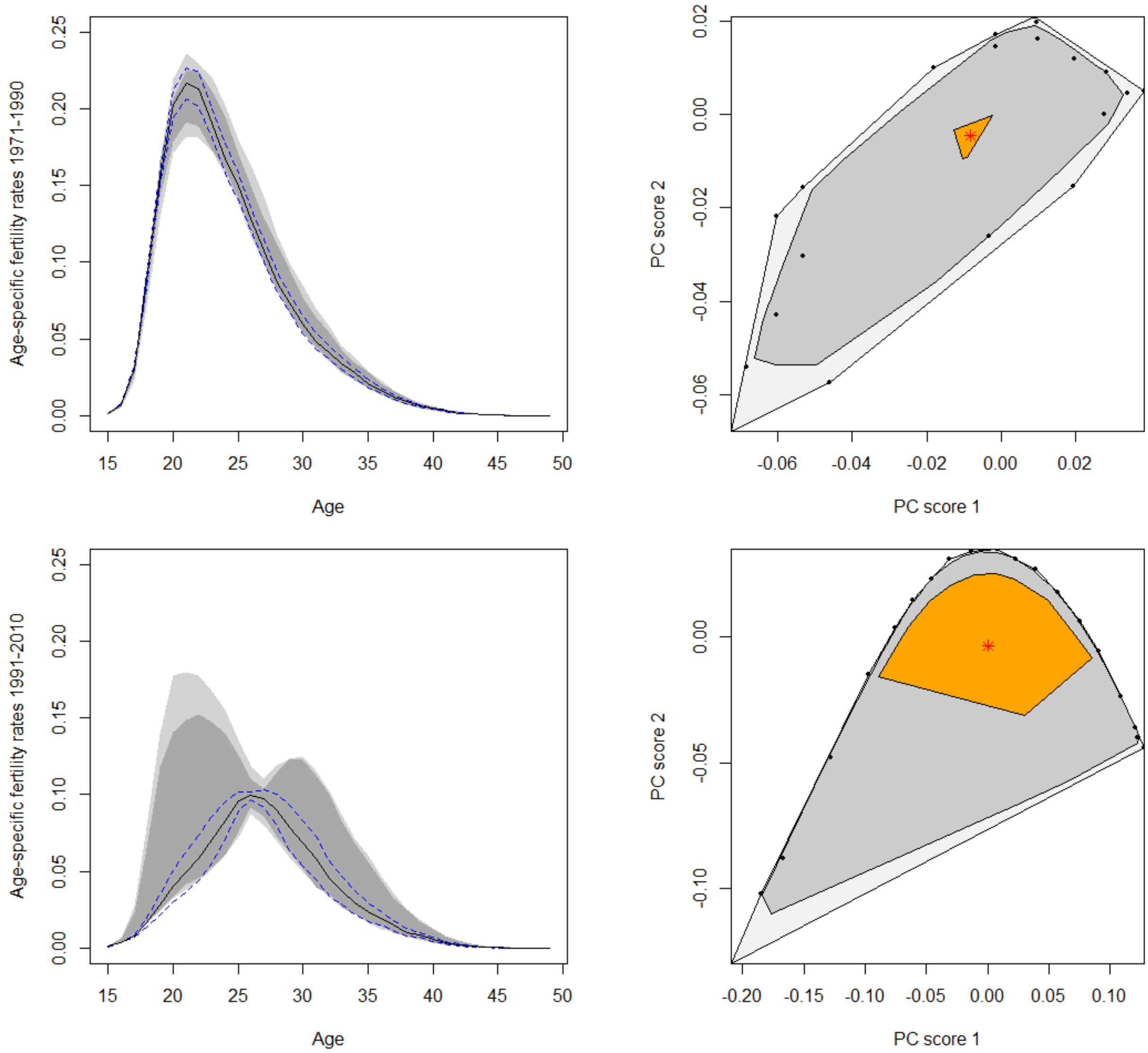




Source: data CZSO, author’s calculation and illustration in RStudio by “rainbow” package

We divide the analysed period on four 20-years long series(1931–1950, 1951–1970, 1971–1990 and 1991–2010). The oldest period (1925–1930) is omitted, as well as the latest years 2011 and 2012. We process these series by principal components method to identify outliers according to Hyndman, Shang, (2009) and Hyndman (2012) and to find median curves. Results can be seen in the Fig. 3 for the case of 1931–1950 and 1951–1970, and in the Fig. 4 for the case of 1971–1990 and 1991–2010. We identified outliers of age-specific fertility rates (years 1963 and 1964)in the analysed period 1951–1970 only, but this remoteness is not significant. Much more significant is the shift in median curves of these rates to the right. The peak of this curve was in age of 27 years within the analysed period 1991–2010 compared to the other three periods, when this peak was in the range of 22–23 years. The years 1980–1985, which have been evaluated as non-normal by Jarque-Bera test, are the part of output in the Fig. 4 (top).We can see that the shape of this distribution is significantly positively skewed and reminds the student’s t-distribution.

Fig. 4: Median of Czech females' age-specific fertility rates in 1971–1990 and 1991–2010



Source: data CZSO, author's calculation and illustration in RStudio by "rainbow" package

The most important for future analysis and modelling of Czech fertility is the output in the Fig. 4 (bottom). Resulting median curve, which was calculated from the interval 1991–2010 is a compromise between mixtures of two normal distributions. The second one normal distribution with 30 years of modal fertility age is a reflection of the latest trend in our society. The last 10 years of age-specific fertility rates could create an estimation of parameter a_x of the Lee-Carter model with parameters

$$\approx \mathrm{N}\left(\mu = 0{,}03741773\ ;\ \sigma^2 = 0{,}00141841\right), \tag{4}$$

which can be used instead of classical average age-specific fertility profile independent of time

$$a_x = \frac{\sum_{t=1}^{T} f_{x,t}}{T}, \tag{5}$$

which is not robust, because it is easily influenced by outliers. Using explicitly specified parameters μ and σ^2based on expert judgment eliminates the possibility of bias of this profile and our predictions could have a more realistic development. Furthermore the prediction can be divided into several time intervals with possibility to select different parameters μ and $\sigma^2$for each interval.

Conclusion

Contrary to expectations, age-specific fertility rates of Czech females were normal at the 5% significance level in all analysed years, except for the period of 1980–1985. Based on the principal components method was found that within the specified 20-years-longtime intervals there have been no unexpected outliers. Positive skewness of the distribution weakened in time and currently there is a relatively flat normal distribution with approximate values of parameters: N(μ = 0,03741773; σ^2 = 0,00141841).Those parameters of Czech fertility can be used for population projection by Lee-Carter model, where instead of the original parameters a_x can be used defined distribution of the fertility profile. This important information is necessary for the next analysis of demographic fertility data. According to Lee-Carter methodology (Lee, Carter, 1992) the parameter $a_x$is not dependent on singular value decomposition of age-and-sex specific matrix of demographic rates and therefore it can be explicitly inserted into the calculation without compromising the correctness of estimation parameters b_x and k_t. It is also possible to change these parameters over time based on additional information and expert opinion.

Acknowledgment

This paper was supported by grant no. P402/12/G097 DYME “Dynamic Models in Economics” of the Czech Science Foundation and also by long term institutional support of research activities by Faculty of Informatics and Statistics, University of Economics, Prague.

References

BARTOŠOVÁ, J., LONGFORD, NICHOLAS, T. (2014).A study of income stability in the Czech Republic by finite mixtures. *Prague Economic Papers*, vol. 23, no. 3, pp. 330-349.

BARTOŠOVÁ, J., BÍNA, V. (2012). Sensitivity of monetary poverty measures on the setting of parameters concerning equalization of household size. In: *Ramík, J., Stavárek, D. (eds.) Proceedings of 30th International Conference Mathematical Methods in Economics 2012*. Karviná: Silesian University, School of Business Administration, pp. 25-30.

BÍLKOVÁ, D. (2009). Pareto Distribution and Wage Models. In: *Aplimat 2009 [CD-ROM]*. Bratislava: Slovak University of Technology, pp. 723–732.

BÍLKOVÁ, D. (2012). Development of wage Distribution of the Czech Republic in Recent Years by Highest Education Attaintment and Forecasts for 2011 and 2012. In: *6th International Days of Statistics and Economics*. Prague: University of Economics, Prague, pp. 162–182.

HYNDMAN, R. J., SHANG, H. L. (2009). Forecasting functional time series. *Journal of the Korean Statistical Society*, 38(3), pp. 199-221 (With discussion).

HYNDMAN, R. J. (2012). *demography: Forecasting mortality, fertility, migration and population data*. R package v. 1.16. URL: <http://robjhyndman.com/software/demography/>

JARQUE, C. M., BERA, A.K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals, *Economics Letters*, vol. 6, no. 3, pp. 255-259.

LEE, R. D., CARTER, L. R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, vol. 87, pp. 659-675.

LÖSTER, T., PAVELKA, T. (2013). Evaluating of the Results of Clustering in Practical Economic Tasks. In: *7th International Days of Statistics and Economics*. Slaný: Melandrium, pp. 804-818.

LÖSTER, T. (2014).Vývoj základních příjmových charakteristik v jednotlivých krajích ČR v letech 2005-2013. In: *Nerovnosť a chudoba v Európskejúnii a na Slovensku*. Košice: Technická univerzita v Košiciach, pp. 120-134.

R DEVELOPMENT CORE TEAM. (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.R-project.org/>

SHANG, H. L., HYNDMAN, R. J. (2010). Exploratory graphics for functional data. *Working paper of the Department of Econometrics and Business Statistics*, Monash University, Clayton, Australia, August 3, 2010, pp. 1-9.

ŠIMPACH, O., PECHROVÁ, M. (2013). Assessing the impact of standard of living on the life expectancy at birth using Vector Autoregressive Model. In: *Mathematical Methods in Economics 2013, PTS I and II*. Jihlava: College of Polytechnics Jihlava, pp. 921–926.

ŠIMPACH, O., DOTLAČILOVÁ, P., LANGHAMROVÁ, J. (2013). Logistic and ARIMA models in the Estimation of Life Expectancy in the Czech Republic. In: *Mathematical Methods in Economics 2013, PTS I and II*. Jihlava: College of Polytechnics Jihlava, pp. 915–920.

ŠIMPACH, O., LANGHAMROVÁ, J. (2014a). Stochastic Modelling of Age-specific Mortality Rates for Demographic Projections: Two Different Approaches. In: *Mathematical Methods in Economics 2014*. Olomouc:Palacký University in Olomouc, pp. 890–895.

ŠIMPACH, O., LANGHAMROVÁ, J. (2014b). Development of Socio-Economic Indicators and Mortality Rates during Ten Years of the CR Membership in the EU. In: *Proceedings of the 2ndInternational Conference on European Integration (ICEI 2014).* Ostrava: VŠB TU, pp. 675–683.

TUKEY, J. W. (1975). Mathematics and the picturing of data. In: *R. D. James, ed., "Proceedings of the International Congress of Mathematicians"*, vol. 2, Canadian mathematical congress, Aug. 21-29, 1974, Vancouver, pp. 523–531.

Contact

Ing. Ondřej Šimpach

University of Economics Prague, Faculty of Informatics and Statistics

W. Churchill sq. 4, 130 67 Prague 3 – Žižkov, Czech Republic

ondrej.simpach@vse.cz