The Challenges of Migration to the European Union for Demographic Modelling

Ondřej Šimpach[1], Marie Pechrová[2]
University of Economics Prague[1], Institute of Agricultural Economics and Information[2]
Faculty of Informatics and Statistics, Department of Statistics and Probability[1]
Winston Churchill sq. 4[1], Mánesova 1453/75[2]
Prague, Czech Republic
e-mail: ondrej.simpach@vse.cz, pechrova.marie@uzei.cz

Abstract
The aim of the paper is to discuss the impacts of recent migration to the European Union on the current modelling tools in demography. The demographic modelling has undergone a long history from component method as a matrix operator or expert estimations and expectations to Lee-Carter models based on principal component analysis and stochastic modelling. However, all models have their pros and cons. Their main weakness is that they are based on historical data. This means that they are not able to take into account unusual situations and demographic shocks. Despite that simplification of the reality is natural for all models, such big change as recent heavy migration to the EU can negatively influence the explanatory power of the models. Migration is a very important demographic component and it cannot be ignored in models or cannot be considered as unchangeable in time. Therefore, paper discusses the migration itself, the models for demographic predictions, their weaknesses and outline possible solutions.

Keywords: *Demographic Predictions, Migration, European Union*

JEL Classification: *C32, J11, J61*

1. Introduction

It is a well-known fact that he European population is rapidly ageing. Situation is although worse in the more developed western European countries, but also the countries in Eastern Europe begin to deal with the same problem. "This implies changing economic and social relations between the generations. In turn this precipitates economic change. In particular, the welfare bill in the future needs to be paid for more dependants by a smaller working population" (Lindh, 2003). Some of the population pressures might be relieved by immigration, but it must not be too high. Otherwise, it could cause the social problems because "with growing disparities between the levels of material wealth in rich and poor countries, migration appears to be an attractive option for inhabitants of less developed countries" (Rowlands, 1999). The increased migration from countries affected by wars and poor economic situation has recently raise many concerns. Beside the policy consequences, the incoming of significant number of people can influence the population structure not only from the nationality point of view, but also the age and gender structure.

Demographic analyses and population projections are typically made under the assumption that future demographic development is deterministic. However, refugees' crisis can represent serious distortions to the models. Also, the longer is the projection horizon, the less probable

is that the assumptions of the models would hold. Therefore, the stochastic approach must be considered. Despite that it enables more precise projections of population, its assumptions might not hold in turbulent development of the reality.

Such big change as recent heavy migration to the EU can negatively influence the explanatory power of the models. Migration is a very important demographic component and it cannot be ignored in models or cannot be considered as unchangeable in time. Therefore, the aim of the paper is to discuss the impacts of recent migration to the European Union on the current modelling tools in demography. The structure of the paper is as follows. Firstly, the methods and the data about migration used are presented. In next section, there is the development of migration to the EU described and the models for demographic predictions evaluated. Last section concludes.

2. Problem Formulation and Methodology

From methodological point of view we distinguish deterministic and stochastic projections. The first one is based on predetermined assumptions. Initially they were done by expert expectations, later by the supporting statistical methods, which were further developed and refined. The expert expectations include estimates based on the recommendations of demographers and experts from the fields of sociology, political science, medicine and law. Among statistical methods belong linear and multiple regressions. Combining these two approaches one can get the expected scenarios, which also serve as an input into the calculation process of the projection itself.

The simplest projection of population can be done by log-linear regression of the population development trend. Mostly used approach for deterministic demographic projections is cohort-component method (e.g. Leslie (1945), Keyfitz (1964) or Bogue et al. (1993)), which can be enriched by the elements of the theory of probability. Cohort-component method is an old algorithm, but it is still popular tool for projection thanks to its usefulness and simplicity. For example a simple Leslie projection matrix requires only the knowledge of age structure and age specific birth and death rates. However, the results are not robust when the population changes are high or affected by other variables. Box-Jenkins approach (Box and Jenkins, (1970)) can be used for forecasting the population too. Pflaumer (1992) showed that forecasting accuracy indicates that the Box-Jenkins method produces population forecasts that are at least as reliable as those done with more traditional demographic methods.

Different approaches are stochastic projections. They are described in detail for example in Bell and Monsell (1991), Lee and Carter (1992), Lee and Tuljapurkar (1994) and Buetner (2002) and applied on the conditions of the CR e.g. by Arltová, (2011), Arlt and Arltová, (2011), and Šimpach and Langhamrová (2014). Bayesian approach for population forecasting was incorporated to Lee-Carter model by Wiśniowski et al. (2015). These projections are based on stochastic modelling of time series of age-specific demographic rates and are complemented by multivariate statistical methods. Great influence on the correct results has a sufficient length of the analysed time series (see e.g. study by Booth et al. (2005)). There are countries with detailed statistics for long period of time. On the other hand, there are also states without them as they either never published them, or gathered them because of political, economic or social circumstances that occur. Other reasons for the incompletion of the data are territorial and political transformations, the establishment of new states and the civil and world wars.

Currently the EU is contending with problems of immigration. The problem of possible datasets distortion should be discussed as many demographic models are not able to adequately explain the migration process. "Population migration involves the relocation of individuals, households or moving groups between geographical locations" (Vitanov and Vitanov, 2016). Our paper analyses the application of the Lee-Carter model (Lee and Carter, (1992); Lee and Tuljapurkar (1994)) on the data about immigrants into the selected European Union countries obtained from Eurostat (2016c) database.

3. Problem Solution

Firstly, the process of the immigration to the EU itself and age-and-sex structure of the immigrants are analysed. Secondly, the urgency of migration issue in V4 countries and Germany is discussed. Thirdly, the possible modifications of Lee-Carter model are outlined in order to make it applicable on the projections of migration processes.

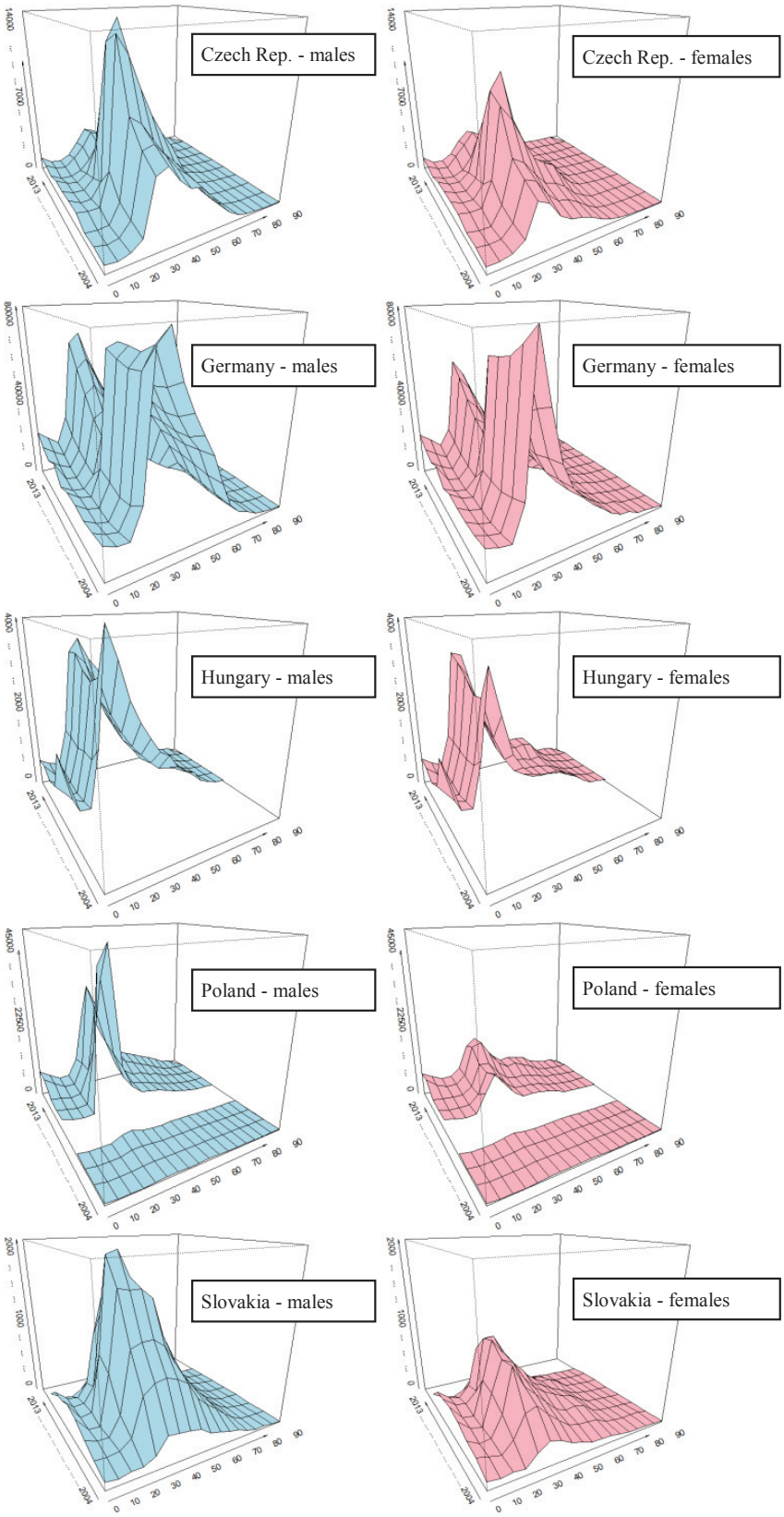
3.1 Development of Immigration to the EU

The EU is believed to attract the immigrants due to its relative economic prosperity and political stability. The number of immigrants continuously increased since 2004. Latest available data from 2013 tells that there were 3.4 million of immigrants. Germany has always been a country with highest total immigration from all EU countries, likely thanks to its good economic situation. It could have been expected that current immigration crisis would affect it the most. On the other hand, recalculating the number of immigrants per country inhabitants, the most of them are in Luxembourg (3.93%). Contrary to that, the share of immigrants on the CR's population is mild. The percentage was around 1% of the total immigration during the period 2005–2013 with exception of years 2007–2008 (1.02% and 1.04%). The shares are even lower in Hungary and Slovakia (under 0.30%). In Poland the share was the highest in 2013 (0.58%). Still, the capacity of the countries might be sufficient.

The most frequented age groups of immigrants are economically productive generations. It can be seen from Figure 1 that the peaks are between 30 to 40 years. Only for females in Poland, the number is equal for all age categories in 2004, 2005. However, later the increase in the number of population around 30 years can be observed. Surprisingly, there are fewer immigrants around 20 years old in Germany and Slovakia, than children younger 10 years. Unfortunately the recent immigration data are available on Eurostat only for year 2013 the latest. Another problem is that many European Union countries have not yet provided data about immigration to Eurostat database, because there are inconsistencies in the measurement methodology and the official system fails in providing and periodically updating the database. Hungary provided the data from 2009 to 2013 only and the Polish population lacks three-year period of 2007–2009. Hungary is very important country in the statistical point of view, because currently the large proportions of immigrants to the Western Europe flow through it.

Talking about recent immigration crisis, we shall therefore analyse the number of asylum applicants, i.e. the persons having submitted an application for international protection or having been included in such application as a family member during the reference period. So far, the problem of immigration is not that urgent in the Czech Republic. As it can be seen from Figure 2, the number of asylum applicants stays rather low in comparison to other EU member states. The share on total EU asylum applicants was the highest in March 2015, when 0.28% of all applied in the Czech Republic. Hence, we may consider the migration to the CR mild and do not have to necessary adjust current population projections or the models of them.

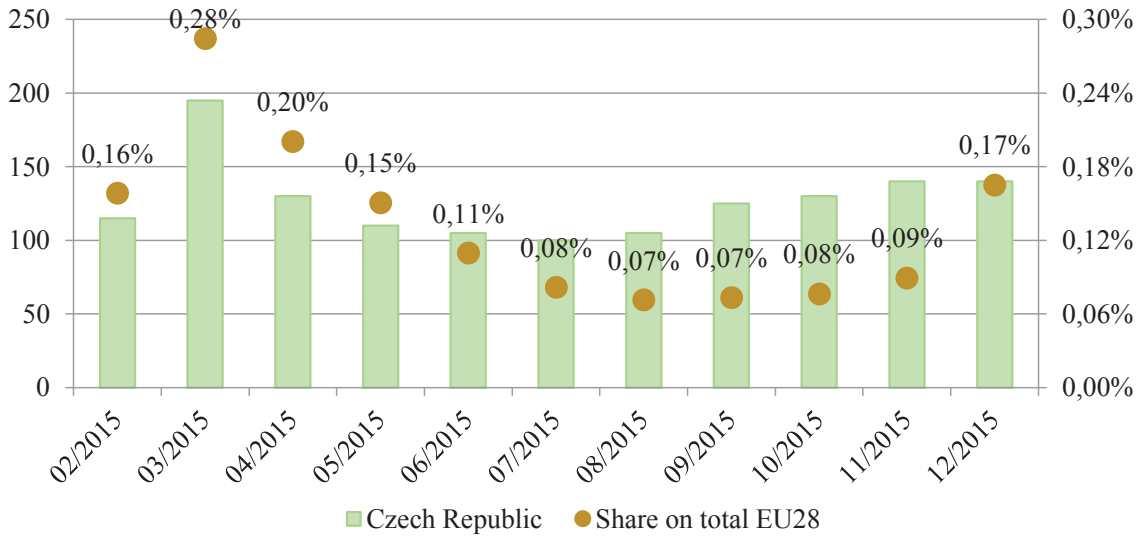
Figure 1: Number of (*x*–*x*+5)-Year Old Immigrants by Sex into Selected EU Countries



Source: Eurostat (2016c); data from 2004 to 2013 with gaps, author's elaboration

Slovakia was in the same favourable situation (the highest share was 0.05% in March 2015). In Poland, the share was higher (0.95% on average from February to September 2015. However, in other EU countries the situation was different. In March 2015, almost half of asylum applicant submitted its applications in Germany. In Hungary, the share achieved even 32.02% in August 2015.

Figure 2: Nr. of Asylum Applicants in the CR (thousands) and Share on Total in EU 28



Source: Eurostat 2016, author's elaboration

3.2 Migration in Population Projection Models

Stochastic projections of the migration process are rarely used. Mostly, the experts expectations based on the economic situation and assumed development of it are used to elaborate immigration and emigration projections. Probably only Lee-Carter model is able to project the migration stochastically. However, while it is extensively used to predict mortality and fertility, its utilization for migration projections is still mild. The challenge for the future is to adjust this model to project the migration realistically.

The basic idea of Lee-Carter model is implemented in decomposition of empirical age-and-sex specific numbers of immigrants in appropriate time periods. The model can be written as

$$I_{x,t}^{M/F} = a_x^{M/F} + b_x^{M/F} \cdot k_t^{M/F} + \varepsilon_{x,t}^{M/F}, \qquad (1)$$

where age x = 0–4, 5–9, ..., 85+ years (according to Eurostat publication methodology), time t = 1, 2, ... T, parameters $a_x^{M/F}$ are the age-specific immigration profiles independent of time, $b_x^{M/F}$ are the additional age-specific component that determines the change in the level of immigration in each age group when the indicator $k_t^{M/F}$ changes, and finally $k_t^{M/F}$ are time-varying parameters – the total immigration indices. $\varepsilon_x^{M/F}$ is a residual element with characteristics of white noise process, where $E(\varepsilon_x) = 0$, $D(\varepsilon_x) = \sigma^2$, $cov(\varepsilon_x ; \varepsilon_x') = 0$ and $\varepsilon_x \approx$ N distribution, letters M and F denote gender. Estimation of parameters $b_x^{M/F}$ and $k_t^{M/F}$ is based on principle of singular value decomposition (SVD) of matrix of age-specific number of immigrants, as presented e.g. on the case of mortality by Bell and Monsell (1991), Lee and Carter (1992) and Hyndman and Ullah (2010). Age-specific number of immigrants $I_x^{M/F}$ at the exact age x and in time t creates $x \times T$ dimensional matrix

$$\mathbf{I} = \mathbf{A} + \mathbf{B}\mathbf{K}^{\mathrm{T}} + \mathbf{E}. \tag{2}$$

Identification of the Lee-Carter migration model is ensured by restrictive conditions

$$\sum_{x=0-4}^{85+} b_x^{M/Ž} = 1 \quad \text{and} \quad \sum_{t=1}^{T} k_t^{M/Ž} = 0. \tag{3}$$

Finally, the simple arithmetic average of age-specific number of immigrants by gender is calculated as

$$a_x^{M/Ž} = \frac{\sum_{t=1}^{T} I_{x,t}^{M/Ž}}{T}. \tag{4}$$

For the prediction of age-specific number of immigrates it is necessary to predict the values of parameter $k_t^{M/F}$ that is mostly done by ARIMA(p,d,q) models.

The first aspect of the Lee-Carter model of migration which has to be considered is that the model needs *a sufficiently long data base* that should be as stable as possible (see e.g. paper by Šimpach et al., (2014)). Instability in the development of matrix of migration time series causes the deflection of the average migration profile $a_x^{M/F}$ and thus bias the estimates of future values. In order to find the main components explaining the trend and previous development the data base must be sufficiently long.

Second issue is that because the migration has never been such discussed topic until recently *not enough attention was paid to these statistics* and most of the European states miss the quality data sets in mutually comparable datasheets.

Third aspect is related to the model itself. "The standard Lee–Carter (LC) model, which uses singular value decomposition, assumes that the errors have a constant variance over all ages" (Koissi and Shapiro, (2006)). However, this does not often hold. Therefore Koissi and Shapiro (2006) suggested fuzzy approach where the errors are viewed as fuzziness of the model structure; hence the homoscedasticity is not an issue. Only when particular European countries, which are affected by high immigration the most, update their databases into uniform and comparable form, the modified approach of the authors Koissi and Shapiro (2006) could be applied as the correction suggested by them is not susceptible to insufficient length of the analysed time series. Of course, everything depends on the data sources, because if there are no data, analysis is not possible.

4. Conclusion

The aim of the paper was to discuss the impacts of recent migration to the European Union on the current modelling tools in demography. Each model for population projection has its advantages and disadvantages, but all have common feature – they are based on historical data. When the time series of migration data is not sufficiently long or some data are missing or are inaccurate, the projections gained by the stochastic modelling might be negatively affected, i.e. biased. In this case, the expert estimations of the future are more reliable. However, this should not prevent the researches to improve current stochastic modelling in terms of migration projections. For other possible models (e.g. gravity model, human capital model or modified gravity model), used for migration forecasting see paper by Greenwood (2005). In our article we focused on Lee-Carter model based on principal component analysis and

stochastic modelling. It shall be further adjusted to be able to predict migration in a feasible way. This is the challenge for our future research.

Acknowledgements

This paper was supported by the Czech Science Foundation project, which is registered under No. P402/12/G097 *DYME – Dynamic Models in Economics*.

References

[1] Arlt, J. and Arltová, M. (2011). Mortality forecasting on the base of Cointegrated Lee-Carter Method. In ASMDA 2011 – Applied Stochastic Models and Data Analysis. Roma: Universita di Roma, pp. 81-87.

[2] Arltová, M. (2011). *Stochastické metody modelování a předpovídání demografických procesů* [habilitation thesis]. Prague: University of Economics Prague, 131 p.

[3] Bell, W.R. and Monsell, B. (1991). Using principal components in time series modelling and forecasting of age-specific mortality rates. In Proceedings of the American Statistical Association, Social Statistics Section, pp. 154–159.

[4] Bogue, D.J. andAnderton, D.L., Arriaga E.E. (1993). Population models, Projections and Estimates. *Readings in Population Research Methodology*, vol. 5.

[5] Booth, H., Tickle, L. and Smith, L. (2005). Evaluation of the variants of the Lee-Carter method of forecasting mortality: a multi-country comparison. *New Zealand Population Review*, vol. 31, iss. 1, pp. 13-34.

[6] Box, G.E.P. and Jenkins, G. (1970). *Time series analysis: Forecasting and control*. San Fran-cisco: Holden-Day.

[7] Buetner, T. (2002). Approaches and experiences in projecting mortality patterns for the oldest old. *North American Actuarial Journal*, vol. 6, iss. 3, pp. 14–25.

[8] Eurostat (2015). *Statistika migrace a migrující populace* [online]. [cit. 2016-03-04]. Available: http://ec.europa.eu/eurostat/statistics-explained/index.php/ Migration_ and_migrant_population_statistics/cs

[9] Eurostat (2016a). *Immigration by five year age group, sex and citizenship [migr_imm1ctz]. Database by themes* [online]. [cit. 2016-03-04]. Available: http://appsso.eurostat.ec.europa.eu/nui/show.do

[10] Eurostat (2016b). *Population on 1 January by five year age group, sex and citizenship [migr_pop1ctz]. Database by themes* [online]. [cit. 2016-03-04]. Available: http://appsso.eurostat.ec.europa.eu/nui/show.do

[11] Eurostat (2016c). *Immigration by five year age group, sex and country of previous residence [migr_imm5prv]. Database by themes* [online]. [cit. 2016-02-29]. Available: http://ec.europa.eu/eurostat/data/database.

[12] Greenwood, M.J. (2005). Modeling migration. *Encyclopedia of Social Measurement*, vol. 2, pp. 725-734.

[13] Keyfitz, N. (1964). The population projection as a matrix operator. *Demography*, Springer-Verlag, vol. 1, iss. 1, pp. 56-73.

[14] Koissi, M.C. and Shapiro, A. F. (2006). Fuzzy formulation of the Lee–Carter model for mortality forecasting. *Insurance: Mathematics and Economics*, vol. 39, iss. 3, pp. 287-309.

[15] Lee, R.D. and Carter, L.R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, vol. 87, pp. 659-675.

[16] Lee, R.D. and Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: beyond high, medium, and low. *Journal of the American Statistical Association*, vol. 89, pp. 1175-1189.

[17] Leslie, P.H. (1945). On the Use of Matrices in Certain Population Mathematics. *Biometrika*, vol. 33, iss. 3, pp. 183-212.

[18] Lindh, T. (2003). Demography as a forecasting tool. *Futures*, vol. 35, pp. 37-48.

[19] Hyndman, R.J. and Ullah, MD. S. (2010). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, vol. 51, iss. 10, pp. 4942-4956.

[20] Pflaumer, P. (1992). Forecasting US population totals with the Box-Jenkins approach. *International Journal of Forecasting*, vol. 8, isso. 3, pp. 329-338.

[21] Rowlands, D. (1999) Domestic Governance and International Migration. *World Development*, vol. 27, iss. 8, pp. 1477-1491.

[22] Šimpach, O., Dotlačilová, P. and Langhamrová, J. (2014). Effect of the Length and Stability of the Time Series on the Results of Stochastic Mortality Projection: An application of the Lee-Carter model. In International work-conference on time series (ITISE 2014). Granada: University of Granada, pp. 1375-1386.

[23] Šimpach, O. and Langhamrová, J. (2014). 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.

[24] Vitanov, N.K. and Vitanov, K.N. (2016). Box model for channels of human migration. *Mathematical Social Sciences*, February 2016, In Press, Accepted Manuscript.

[25] Wiśniowski, A. et al. (2015). Bayesian Population Forecasting: Extending the Lee-Carter Method. *Demography*, vol. 52, iss. 3, pp. 1035-1059.