Application of the Modern Approach to Age-specific Fertility Rates Stochastic Modelling in the Czech Republic

Ondrej Simpach[1]

Abstract. Mortality and fertility represent two basic components of natural population replacement. While the mortality can be modelled based on known conditions (decrease of the probability of death in time and increase of life expectancy, which can be applied to the most of advanced populations in general), in the case of fertility the approach must be slightly different. Changes in the level and distribution of fertility are characterized by social conditions, lifestyle and also by decision makings of individuals. There are studies that try (using the principal components method and stochastic trend) to find a model that explains and predicts the future development of age-specific fertility rates. Our study utilize also this approach.

The aim of this paper is to identify appropriate form of the Lee-Carter model based on the data set of Czech females' age-specific fertility rates. It will be used for modelling a matrix of age-specific fertility rates of Czech females. The predictions of these rates up to the year 2050 will be constructed. It is required that the shape of the distribution of predicted fertility rates follows the current data (which is not respected in some studies). The model also has to respect the trend that arises from past data. Predictions which are output of this paper provide lower values than expectations published by Eurostat. As our projections of age-specific fertility rates of Czech females are predicted by modern and sophisticated system, it is possible to claim that the future would be more pessimistic than Eurostat expects.

Keywords: fertility, Lee-Carter model, demography, stochastic modelling.

JEL classification: C61, C63
AMS classification: 62H12

1 Introduction

The population development and improving the living standard in the country is closely related to postponement of first childbirth to the later age and together the decline of live births in total (see e.g. the paper from Rueda, Rodriguez [13] or Simpach, Pechrova [16]). In many populations this decrease is below the level of simple reproduction of the population (2.08 children per 1 female within the reproduction period). Czech Republic have good database that will allow us to obtain the age-specific fertility rates $f_{x,t}$ of Czech females from 1925 to 2012. During this period, the development of these rates was affected by a wide range of social changes, which also brought the consequences of changes in fertility of Czech females. Difficult to explain and difficult to predict is the behaviour as the result of individual decisions in family planning (Peristera, Kostaki [11]). The level of fertility cannot permanently decrease in the future, because there is its logical value below which the fertility never decreased before, (e.g. the past development of the Czech time series showed that in 1999 there was the total fertility rate 1.13 live birth child per 1 female during her reproductive period and the range of values 1.13–1.18 was in many other cases during the 90s of the last century). Neither this value can permanently grow in the future, because of health point of view there is a maximum possible value that a female cannot exceed (see e.g. Caputo, Nicotra, Gloria-Bottini [3] or Myrskylae, Goldstein, Cheng [10]). The level of fertility varies between

[1]University of Economics Prague, Faculty of Informatics and Statistics, Department of Statistics and Probability, Winston Churchill sq. 4, 130 67 Prague 3 - Zizkov, Czech Republic, ondrej.simpach@vse.cz

its logical lower and upper limits in time, and just as well depends on the shape of the distribution of age-specific rates (Rueda, Rodruguez [13]).

In our paper we estimate the principal components that explain fertility from the multidimensional matrix of age-specific fertility rates (see Hyndman, Booth [4]). This will be performed using the singular value decomposition (SVD) and by Lee-Carter model (Lee, Carter [7]). Estimates are made on the basis of the shortened length of the analysed matrix for the period 1998-2012 (see also Simpach, Langhamrova [15]), because in the case of fertility modelling it makes sense to use only the current and stable database, which is not deflected by past influences. (The shorter is the analysed database of fertility, the more realistic results for the Czech population can be expected). We also calculate the predictions of age-specific fertility rates for the period 2013-2050 on the basis of this model and in the conclusion we compare our results with low variant of fertility projection according to Eurostat. It will be obvious that Eurostat projections are initially overestimated above the level of 2012, on the other hand the projections by Lee-Carter model are initially slightly underestimated bellow the level 2012. We can use these estimates as a lower and upper limit of the potential future fertility in the Czech Republic.

2 Methodology and Data

We use the empirical data from Czech Statistical Office (CZSO) about the number of live-born persons to x-year old mothers in time t ($N_{x,t}$) and the number of mid-year female population x-year old in time t ($S_{x,t}$) for our calculations, where x = 15–49 completed years of life and t = 1925–2012. This allows us to calculate the age-specific fertility rates as

$$f_{x,t} = \frac{N_{x,t}}{S_{x,t}}, \tag{1}$$

and after (× 1,000) we interpret the result as the number of live births per 1,000 x-year old females in time t. In order to be clear, which changes in age-specific fertility rates occurred in the past, their development is shown in Figure 1 as “rainbow” chart. (This technology uses Hyndman, Booth [4] and Hyndman [5]). We can see the changing of maximum values of the age-specific fertility rates in the past and especially since the beginning of 90s of the last century become the trend of postponing childbirth to the later ages and in these days the modal age have exceeded the value of 30 years.

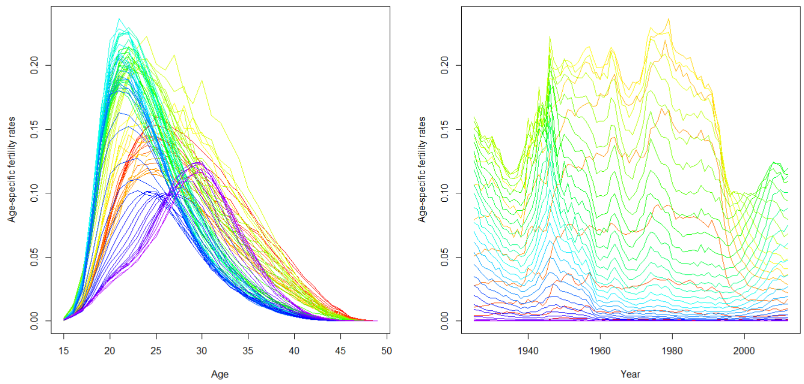

Figure 1 Empirical values of $f_{x,t}$ of the Czech females for the period 1925–2012 in rainbow charts. The left chart represents changes of $f_{x,t}$ over time, the right one represents changes of $f_{x,t}$ within the age groups. *Source: Data CZSO, author's calculations and illustration based on Hyndman [5] approach.*

Used approach based on principal component is, that the empirical values of age-specific fertility rates of Czech females can be decomposed (see Lee, Carter [7] or Lee, Tuljapurkar [8]) as

$$f_{x,t} = a_x + b_x \times k_t + \varepsilon_{x,t} \tag{2}$$

where x = 15–49 years of life, t = 1,2, ..., T, a_x are the age-specific fertility profiles independent of time, b_x are the additional age-specific components determine how much each age group changes when parameter k_t changes and finally k_t are the time-varying parameters - the total fertility indices. ($\varepsilon_{x,t}$ is the error term with characteristics of white noise). The estimation of b_x and k_t is based on Singular

Value Decomposition (SVD) of matrix of age-specific fertility rates, presented e.g. by Bell, Monsell [1], Lee, Carter [7] or Hyndman, Booth [4]. The age-specific fertility rates $f_{x,t}$ at age x and time t according to Lee, Tuljapurkar [8] create 35 × T dimensional matrix

$$\mathbf{F} = \mathbf{A} + \mathbf{B}\mathbf{K}^{\top} + \mathbf{E} \tag{3}$$

and the identification of Lee-Carter model is ensured by

$$\sum_{x=15}^{49} b_x = 1 \qquad \text{and} \qquad \sum_{t=1}^{T} k_t = 0 \tag{4}$$

Finally,

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

is the simple arithmetic average of age-specific fertility rates. For predicting the future age-specific fertility rates it is necessary to forecast the values of parameter k_t only. This forecast is mostly calculated by $ARIMA(p,d,q)$ models with or without drift, (by Box, Jenkins [2] methodology). The values of the parameters a_x and b_x are independent of time and the prediction using the Lee-Carter model is according to Lee, Tuljapurkar [8] purely extrapolative. Our decomposition is applied only to 1998–2012 dataset, because according to the development shown in the Figure 1 (right) is evident, that this shape of distribution is stable reproduction period of Czech females which can continue to the future.

3 Results

In RStudio software (R Development Core Team [12]) we estimate the parameters a_x (age-specific fertility profiles independent in time), b_x (additional age-specific components determine how much each age group changes when k_t changes) and k_t (time-varying parameters - the total fertility indices) for Lee-Carter's model based on data matrix of 1998–2012 only, using the SVD method implemented in the package "demography" (Hyndman [5]). We can see these parameters in the Figure 2.

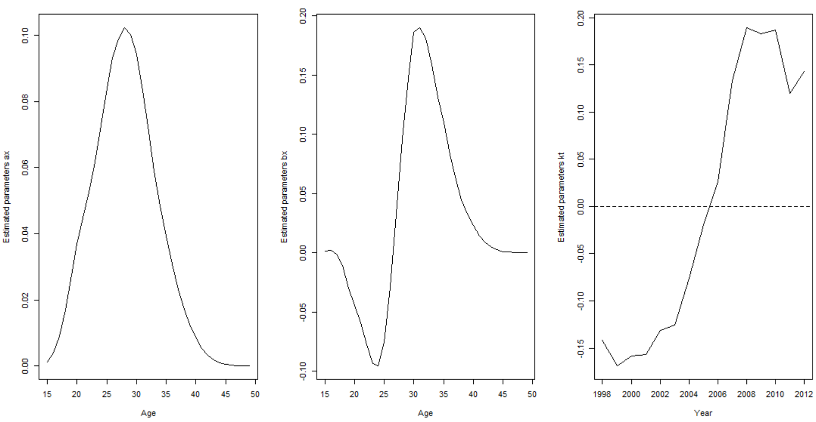

Figure 2 Estimates of age-specific fertility profiles independent in time (parameter a_x, left), the additional age-specific components determine how much each age group changes when k_t changes (parameter b_x, middle) and estimates of the time-varying parameters k_t - the fertility indices (right). *Source: author's calculations and illustration.*

After that we calculated predictions of the parameters k_t up to the year 2050 based on the methodological approach of individual time series modelling ARIMA, (Box, Jenkins [2], Melard, Pasteels [9]) and ran by "forecast" package in RStudio software (Hyndman [5]). Parameters of ARIMA models are displayed in Table 1. It is important to note that the dataset of 1998–2012 is too short for providing forecasts up to the year 2050. Therefore, we use dataset from 1925 to 2012 as a support for our analysis. Given that the resulting predictions of k_t parameters are the same as using both dataset of 1925–2012 and 1998–2012, our results are correct. Realization of predictions of k_t parameters is shown in Figure 3. 95% both-side confidence intervals are slightly closer at the case of model based on 1998–2012 dataset.

Database 1925–2012:	ARIMA (1,1,0)	Database 1998–2012:	ARIMA (1,1,0)
AR(1)	0.3494	AR(1)	0.4782
s.e.	0.1000	s.e.	0.2244
t-stat.	3.4940	t-stat.	2.1310
AIC	-142.98	AIC	-45.690
BIC	-138.05	BIC	-44.410

Table 1 Estimated parameters of two ARIMA models for parameter k_t. We used our knowledge of the shape of the ARIMA model for database 1925–2012 and applied it to the shortened model. Future predictions are comparable despite the fact that the database 1998–2012 is very short for quality prediction. *Source: author's calculations and illustration.*

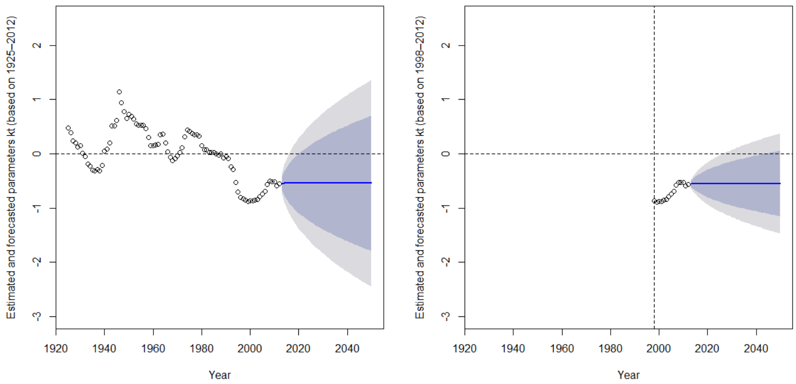

Figure 3 Prediction of the time-varying parameters k_t - fertility indices. On the left side is the model based on data of 1925-2012, on the right side is the model based on data of 1998–2012. There is no significant difference when using different lengths of time series. *Source: author's calculations and illustration.*

Now we evaluate Lee-Carter model on the basis of approach, which is presented by Charpentier, Dutang [6] or Simpach, Dotlacilova, Langhamrova [14]. Using RStudio software we display the Pearson's residues. Lee-Carter's model is evaluated on the basis of the residues by age x and of the residues at time t. The most residues are concentrated around 0, a lot of variability is explained by the estimated model. The Pearson's residues for estimated Lee-Carter model are shown in the Figure 4, where residues by age x dependent on time t are on the left and the residues at time t dependent on age x are on the right.

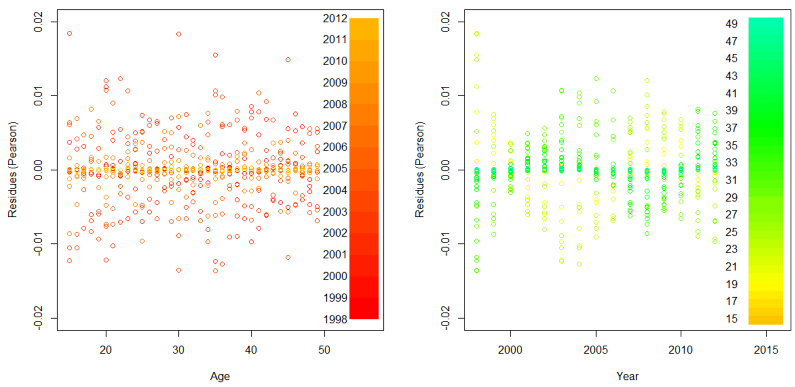

Figure 4 Diagnostic control of the Lee-Carter's model - Pearson's residues for model based on data of 1998–2012. *Source: author's calculations and illustration.*

Based on the estimated parameters $\hat{a}_x$, $\hat{b}_x$ and $\hat{k}_t$ of Lee-Carters model we can estimate the future values of $f_{x,t}$ as

$$f_{x,t} = \hat{a}_x + \hat{b}_x \times \hat{k}_t. \tag{6}$$

Estimated values (left) and the empirical values with the attached estimates (right) of age-specific fertility rates $f_{x,t}$ based on Lee-Carter's model are displayed in 3D perspective chart in the Figure 5 (top). The estimated values based on the expert expectations by Eurostat (left) and then the empirical values with these attached estimates (right) are displayed below.

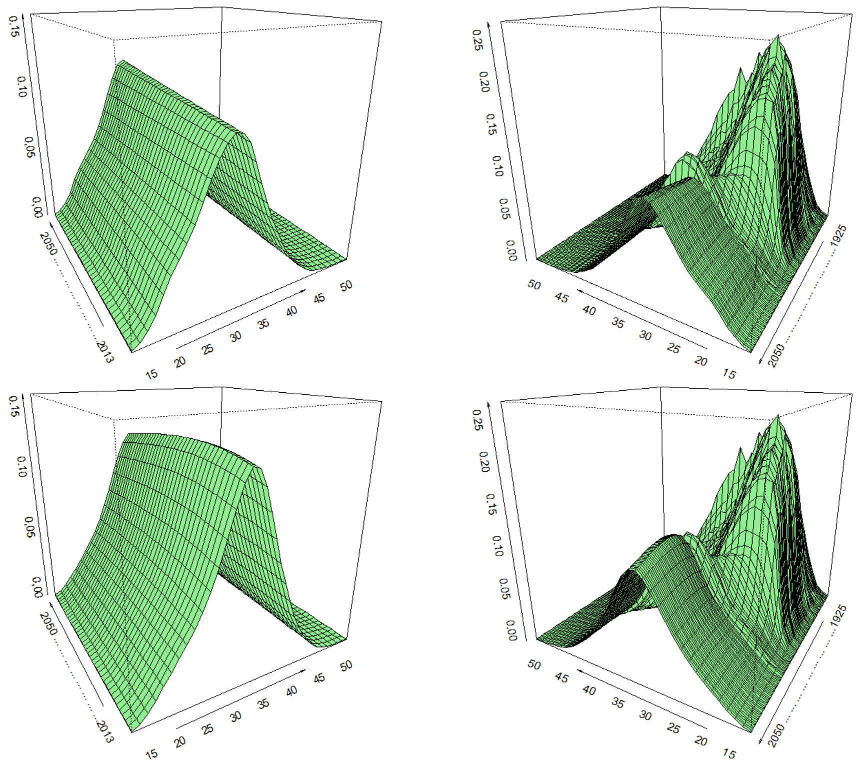

Figure 5 Forecasted values of age-specific fertility rates $f_{x,t}$ of Czech females for the period 2013 to 2050 by Lee-Carter's model based on actual data matrix for the period 1998 to 2012 (top left) and the empirical values of these rates for the period 1925 to 2012 with an forecast attachment (top right). Bottom left are $f_{x,t}$ of Czech females for the period 2013 to 2050 forecasted by Eurostat and bottom right are empirical values with these forecasted values. *Source: author's calculations and illustration, Eurostat database.*

Estimates calculated by the Lee-Carter model are initially underestimated bellow the empirical values of 2012. This is probably due to the deflection of the average fertility profile independent of time (parameter a_x), because shortly before 2000 there was the level of fertility in the Czech Republic significantly lower than today. The forecast assumes a gradual (linear) increase up to the year 2050. On the other hand, the estimates published by Eurostat in its low variant are overestimated over the empirical values of 2012. At a time when the prediction was processed, Eurostat has already calculated with a higher fertility than really occurred. According to Eurostat expectations there is evident an increase of fertility in approximately next 20 years, and later the fertility should start to fall again.

4 Conclusion

The aim of this paper was to estimate the main components explaining the level of fertility in the Czech Republic by the Lee-Carter model and calculate predictions of age-specific fertility rates up to the year 2050. Results were compared to expert estimates by Eurostat in the low variant and there were shown that they are not quite the same. However, if we use estimates of the Lee-Carter model and prediction of Eurostat in the low variant together, we can get a reliable picture of the future level of age-specific fertility rates. Each of these approaches has a different methodology for the estimation, and because the results are not dramatically different, we may take them as a reliable upper and lower limits of the future development of fertility in the Czech Republic. According to the results of prediction by Lee-Carter model and expert expectations by Eurostat is impossible that the level of total fertility rate in the Czech Republic in 2050 would almost reached the simple reproduction of the population.

Acknowledgements

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

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

[2] Box, G. E. P., Jenkins, G.: *Time series analysis: Forecasting and control.* San Francisco, Holden-Day, 537 p., (1970).

[3] Caputo, M., Nicotra, M., Gloria-Bottini, E.: Fertility Transition: Forecast for Demography. *Human biology*, Vol. 80, Iss. 4, pp. 359–376, (2008).

[4] Hyndman, R. J., Booth, H.: Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting*, Vol. 24, Iss. 3, pp. 323-342, (2008).

[5] Hyndman, R. J.: *demography: Forecasting mortality, fertility, migration and population data.* R package v. 1.16, URL: http://robjhyndman.com/software/demography/, (2012).

[6] Charpentier, A., Dutang, Ch.: *L'Actuariat avec R.* [working paper]. Decembre 2012. Paternite-Partage a lindentique 3.0, France de Creative Commons, 215 p., (2012).

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

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

[9] Melard, G., Pasteels, J. M.: Automatic ARIMA Modeling Including Intervention, Using Time Series Expert Software. *International Journal of Forecasting*, Vol. 16, pp. 497–508, (2000).

[10] Myrskylae, M., Goldstein, J. R., Cheng Yen-Hsin., A.: New Cohort Fertility Forecasts for the Developed World: Rises, Falls, and Reversals. *Population and Development Review*, Vol. 39, Iss. 1, pp. 31-56. (2013).

[11] Peristera, P., Kostaki, A.: Modeling fertility in modern populations. *Demographic Research*, Vol. 16, No. 6, pp. 141-194, (2007).

[12] R Development Core Team: *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/, (2008)

[13] Rueda, C., Rodriguez, P.: State space models for estimating and forecasting fertility. *International Journal of Forecasting*, Vol. 26, Iss. 4, pp. 712-724, (2010).

[14] Simpach, O., Dotlacilova, P., Langhamrova, J.: Effect of the Length and Stability of the Time Series on the Results of Stochastic Mortality Projection: An application of the Lee-Carter model. *Proceedings ITISE 2014*, Granada: University of Granada, pp. 1375-1386, (2014).

[15] Simpach, O., Langhamrova, J.: Stochastic Modelling of Age-specific Mortality Rates for Demographic Projections: Two Different Approaches. *Mathematical Methods in Economics 2014*, Olomouc: Palacky University in Olomouc, pp. 890-895, (2014).

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