Fertility of Czech Females Could Be Lower than Expected: Trends in Future Development of Age-Specific Fertility Rates up to the Year 2050

Ondřej Šimpach[1] | *University of Economics, Prague, Czech Republic*

Abstract

Fertility is an essential aspect of reproduction or population replacement of each country. The challenge for demographers is to model fertility and also to estimate its potential future level for the purposes of population projections. In the case of the Czech Republic we have the population projections provided by the Czech Statistical Office (CZSO) with overlooking of the total fertility rate in low, medium and high variant. These estimates despite being based on expert judgments, seem to be too positive compared to the past development of the time series of age-specific fertility rates. The aim of this paper is to assess the situation of fertility in the Czech Republic, to analyse the past development of the time series of age-specific fertility rates using one-dimensional Box-Jenkins models and multidimensional stochastic Lee-Carter approach. Together with found trend in time series and principal components estimated by Lee-Carter's model a forecasts of age-specific fertility rates up to the year 2050 is constructed. These rates are lower than those provided by CZSO in its three variants of the Czech Republic's population projection, and therefore we discuss the causes at the end of the paper. We would like to point out that the potential future development of Czech females fertility could be lower than which are currently expected.

Keywords	***JEL code***
Age-specific fertility rates, ARIMA, Lee-Carter, population projection	*C22, C32, J13*

INTRODUCTION

Mortality and fertility are important parts of the natural population change. Given that the most of populations on Earth started with the dynamic development in recent decades, the standard of living rises and the mortality rates in these countries decline. Due to the fact that the living conditions are better, the forecasting of mortality is not so difficult, because we have common assumptions about the potential

[1] Department of Statistics and Probability, Nám. W. Churchilla 4, 130 67 Prague 3. E-mail: ondrej.simpach@vse.cz, phone: (+420)737665461.

future development, which we simply follow in our predictions (Stauffer, 2002, or Dotlačilová, Šimpach, Langhamrová, 2014). Modelling and estimating future fertility is more complicated, because fertility is influenced by several factors. 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, 2010). This decrease is below the level of simple reproduction of the population (2.08 children per 1 female within the reproduction period) in many populations of developed countries. However, in comparison with mortality, there is still one very important factor that must not be overlooked.

There is a good database in the Czech Republic that will allow us to obtain the age-specific fertility rates of Czech females from 1925 to 2012. During this period, the development of these rates was affected by a wide range of social changes. This was especially the Second World War, the two parts of the consecutive Communist regime, targeted pro-population policies and massive support of young families with a higher number of children, and as well as the downturn of this development during the post-revolutionary period. All these social circumstances brought the consequences of changes in fertility of Czech females, which we are able to justify. It is difficult to explain and to predict as the behaviour is the result of individual decisions in family planning. Neither, the level of fertility can permanently decrease in the future, because there is a value below which the fertility never decreased before. Neither this value can permanently grow in the future, because of health point of view there is a maximum possible value of age that a female cannot exceed (see e.g. Caputo, Nicotra, Gloria-Bottini, 2008, or Myrskylae, Goldstein, Cheng, 2013). The level of fertility varies between its logical lower and upper limits in time, and also depends on the shape of the distribution of age-specific rates. Czech Statistical Office (CZSO) provides regularly updated population projections of the Czech Republic. These projections are constructed by sophisticated cohort-component method, whereby the input attributes and other assumptions are discussed by respected professionals. In the case of the total fertility rate there are currently considered three potential future scenarios, pessimistic (low variant), middle (medium variant) and optimistic (high variant) (CZSO, 2013). Pessimistic scenario consider the same level of total fertility rate in the future as today (1.45 children born to one female during her reproductive period), middle and optimistic consider some increase (see below). The potential future decline is not considered at all, because the past development of the Czech time series showed that e.g. 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. It is important to note that the decline of fertility of Czech females at the end of the last century had been mainly caused by rapid changes in reproductive behaviour – postponing of childbirth to the later ages which is normal in the most of Western European countries today. The sharp fertility decline of younger females was partly compensated (with a delay) by fertility increase of females in higher age groups (Langhamrová, Fiala, 2014). Medium variant of CZSO expect a gradual linear increase up to a value of 1.56, high variant even up to a value 1.61. Is it possible that some of these variations will happen? Can we expect, that females, married couples and partners change their views on the family and this increase will occur? It is possible to read and judge from the population structure of the Czech Republic that the strong generations of 70s are already reproductively exhausted and other strong generation which will be able to significantly revived this situation will not appear within next 20–30 years. The population structure enables to see the development of age-specific fertility rates using a statistical approach and together with the founded trend and the main components explaining fertility levels estimate, how these rates could develop in the future.

The aim of this paper is especially to analyse the past trend in the individual time series of age-specific fertility rates using Box-Jenkins methodology (Box, Jenkins, 1970) with Random Walk models and ARIMA. These models are applied to 35 time series (for age range of 15–49 completed years of life), the periodicity of the time series is annual (1925–2012) with sufficient number of observations. We evaluate the models by diagnostic control (see e.g. Stauffer, 2002) and consequently calculate other predictions of age-specific fertility rates for the period 2013–2050 (different from CZSO approach). The approach of Random Walk models provide one potential prediction, ARIMA slightly different one. At the same time we estimate the principal components that explain fertility from the multidimensional matrix of age-specific fertility rates (see Hyndman, Booth, 2008, or Arltová, 2011). This is performed using the singular value decomposition (SVD), the Lee-Carter model (Lee, Carter, 1992). Estimates are gradually made for the different lengths of the analysed matrix – (I) since 1925 (the beginning of the time series), (II) 1948 (the end of the Second World War, pacification the social situation and the beginning of a new political regime in the country), (III) 1968 (again the restructuring of the society and the beginning of hard normalization), and (IV) 1988 (weakening of the Communist regime in the country and preparing for the new democratic system in our society).[2] Only the results of the model based on data for the period 1925 to 2012 and 1988 to 2012 are presented. It is due to the fact that SDV approach is not appropriate in the case when the multi-dimensional matrix record a wide range of changes in the past and is therefore highly variable. We also calculate the predictions of age-specific fertility rates for the period 2013–2050 on the basis of those two models. All four approaches used for calculation forecasts (Random Walk, ARIMA, LC 1925 and LC 1988) will be compared with each other and with published values of low, medium and high variant of CZSO.

There are other ways to analyse and model fertility in developed populations in order to be able to construct the fertility projections. Peristera and Kostaki (2007) prepared an extensive case study on the United Kingdom, Ireland, France, Greece, Norway, Italy, Denmark, Austria and the United States using the Hadwiger Model, Gamma Model, Beta Model and quadratic Spline Model. Hyndman and Ullah (2010) paid attention to France using robust approach to modelling fertility based on the approach by Lee, Carter (1992). This was used before only on mortality modelling, and its capabilities were extended and used on fertility analysis later. Given that our dataset is suitable for Hyndman and Ullah (2010) approach, (based on studies by Lee and Carter (1992), Lee and Tuljapurkar (1994) and later Rueda, Rodriguez (2010)), we apply this method. The database was also suitable for the application of methodological approach by Box and Jenkins (1970), which is older and which has been used previously in many studies of mortality analysis (Bell, 1997, Stauffer, 2002, or Šimpach, Langhamrová, 2014). It is used in our paper as a comparison of the modern approach of stochastic modelling with principal components and the conventional approach of stochastic modelling with random component. The fertility predictions were calculated in the Czech Republic e.g. by Fiala and Langhamrová (2012), who used a deterministic approach in calculation of the population projections. They calculated with expert judgments of the total fertility rates (available at the CZSO). The age-specific fertility rates were subsequently calculated using the component method.

Using a long series of cross-sectional indicators (fertility rates by age groups) for long-term projections seems, unfortunately, problematic in the Czech Republic – this is the main reason why we do not pay the attention for models LC 1948 and LC 1968. The significant inter-annual fluctuations of fertility rates that do not have the recognizable long-term trend in the cross-sectional point of view are typical for the Czech population (unlike for many developed and Western European countries). It is important for the projection to focus on the relatively recent changes and concentrate on the cohort / generation approach, because the final indicator of fertility in the Czech Republic is long-term stable. (The permanent declining trend begins with generations of females born in 1960 and younger.) This approach was

[2] The various social events did not follow each other exactly at specific 20 years intervals. If we would like to set these dates correctly, we will have to select the 1945, 1968 and 1989. Frequency 20 years between these events was chosen as a compromise variant for the purposes of analysis.

used in article by Myrskylae, Goldstein, Cheng (2013), or in other: Li, Wu (2003), or Morgan, Hagewen (2005). When we analyse the longer time series of variable cross-sectional fertility rates by age groups, we do not improve the projection. The situation is rather the opposite (unlike the natural sciences). This also confirm our results presented at the end of this paper, where the most probable results are provided by the Lee-Carter model for the period 1988–2012, while other models give too high predicted values, which are deflected particularly by high level of fertility of young females between 60s and 80s of the last century and the presented models are not robust for these extreme changes.

Statistically estimated values of the total fertility rates in this article are lower than published values by CZSO. *This paper does not attempt to say in any cases that the estimated future values by CZSO are wrong!* Our predictions were calculated by statistical approach that takes into account the trend and the principal components that make up the main explanatory system. Published values of CZSO take into account the opinions of *demographers* and other experts from the fields of sociology, political science, and medicine. Therefore they are not only enriched by a factor of technological progress, but also affected by subjective thinking that a statistical approach does not have. Another advantage is that statistical approach takes into account in addition the random component (Alders, de Beer, 2004, or Caputo, Nicotra, Gloria-Bottini, 2008). We think that the total fertility rate will be located in 2050 between the level of pessimistic variant of CZSO (1.45) and the results obtained using Random Walk models or ARIMA, (which reached the level of 1.20 in 2050). The values published by CZSO are rather optimistic, the values that provide statistical approaches are rather pessimistic. Let us be more careful in our future expectations and think about whether the Czech Republic has such potential up to the year 2050, in which it could approach their fertility to countries as Belgium, Netherlands or Luxembourg in these days.

1 METHODOLOGY AND DATA

The empirical data from CZSO are used – particularly the number of live-born persons to x-year old mothers in year t ($N_{x,t}$) and the number of midyear female population x-year old in year t ($\overline{S}_{x,t}$), where x = 15–49 completed years of life and t = 1925–2012. This allows us to calculate the age-specific fertility rates as:

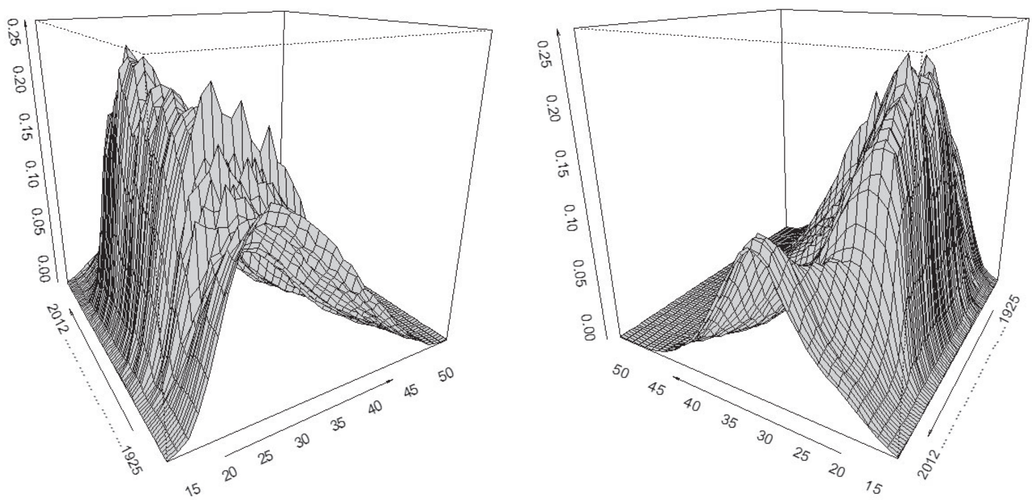
$$f_{x,t} = \frac{N_{x,t}}{\overline{S}_{x,t}}\,, \tag{1}$$

and after (× 1 000) we interpret the result as the number of live births per 1 000 x-year old females in year t. Sum of age-specific fertility rates is the total fertility rate in year t

$$tfr_t = \sum_{x=15}^{49} f_{x,t}\,, \tag{2}$$

which is the sum of live births to one female during her reproductive period. In order to be clear, which changes in age-specific fertility rates occurred in the past, their development is shown in Figure 1 as perspective 3D chart. (This technology uses Charpentier, Dutang, 2012, simple presentation by X-Y chart of these empirical rates shows Figure 8 in Annex). We can see the changing of maximum values of the age-specific fertility rates in the past, as well as the moving of modus. It is especially due to the trend of postponing childbirth to the later ages which emerged at the beginning of 90s of the last century. In these days the modal age have exceeded the value of 30 years.

Figure 1 Empirical data of age-specific fertility rates $f_{t,x}$ of the Czech females for the period 1925 to 2012 in perspective 3D charts. See the significant change of mode of this distribution to the advanced ages



Source: CZSO (2013), author's illustration

If we transform the age-specific fertility rates to the logarithms $f_{x,t} \rightarrow \ln(f_{x,t})$, their variability in time is smaller. We may look on each specific age from 15 to 49 completed years of life as on the individual time series of logarithms of age-specific fertility rates with 88 annual observations (from 1925 to 2012). This approach is generally used by authors to model the logarithms of age-specific mortality rates and for the estimation of the coefficients for declining mortality over time (Stauffer, 2002, or Šimpach, Dotlačilová, Langhamrová, 2014). We do not estimate any coefficients for declining fertility rates over time, but we apply Box-Jenkins methodology on the less variable time series of age-specific fertility rates. When we have a larger number of the analysed time series of similar nature, it is preferable to determine the universal structure of model for all of the considered time series.[3] One of the most frequently used approaches that are chosen as a compromise variant is the Random Walk model (RW) with drift. It is more sophisticated than a simple linear deterministic model (which was used in the past), because it takes into account the random component. We denote the random walk model for the logarithms of age-specific fertility rates as:

$$\ln\left(f_{x,t}\right) = c_x + \ln\left(f_{x,t-1}\right) + \varepsilon_{x,t} \,, \tag{3}$$

[3] In the case that we analyse a small number of time series, it is right that each time series should be analysed with maximum precision (Hyndman et al. 2002). Each obtained model should be tested with a wide range of diagnostic tests and we have to insist that all conditions have been fulfilled to the last detail. For the larger matrices of the time series we are not able to satisfy all assumptions and diagnostic controls, so it is the time to find and choose a compromise structure of model, which satisfies the most of the time series from the matrix (Alders, de Beer, 2004). This issue is devoted e.g. by authors Melard, Pasteels (2000) or Hyndman et al. (2002), who developed the approaches for automatic modelling of time series and automatic forecasting. Nowadays these systems are very sophisticated, but their disadvantage is that the models often include too much of parameters to satisfy the most of the evaluation criteria during the evaluation of model. Very often happens that these parameters are mostly statistically insignificant, or from the logical point of view they do not belong to the model. Their estimated values are also very often incorrect, located in senseless intervals according to statistical theory.

where c_x is drift and $\varepsilon_{x,t}$ is the error term with characteristics of white noise. This formula can be modified according to Box, Jenkins (1970). We get ARIMA (0,1,0) model with drift for age-specific fertility rates as:

$$\ln(f_{x,t}) = \ln(f_{x,0}) + c_x \cdot t_x + \sum_{t=1}^{T} \varepsilon_{x,t}, \tag{4}$$

where $c_x \cdot t_x$ is the deterministic trend. This trend is linear increase / decrease of fertility rates in time. Its most common usage is in the case of modelling age-specific mortality rates, there will be experimentally used in case of fertility. It was further examined by empirical verification which parameters (Auto-Regressive AR or Moving Averages MA) are statistically most important part of the ARIMA model in demographic time series. In the case of modelling mortality is mostly statistically significant component MA(1). Models which contain AR component often do not remove autocorrelation (Melard, Pasteels, 2000). This autocorrelation unfortunately not disappear even if the model includes a drift that often this unpleasant characteristic pulls into itself. Therefore we use component MA and define the model of moving averages without drift according to Box, Jenkins (1970) as:

$$\ln(f_{x,t}) = \varepsilon_{x,t} - \theta_{x,1}\varepsilon_{x,t-1}, \tag{5}$$

with drift respectively as:

$$\ln(f_{x,t}) = c_x + \varepsilon_{x,t} - \theta_{x,1}\varepsilon_{x,t-1}. \tag{6}$$

whereby the provisions of condition $|\theta_{x,1}| < 1$.

The other used approach based on principal component is that the empirical values of age-specific fertility rates can be decomposed (see Lee, Carter, 1992, or Lee, Tuljapurkar, 1994) as:

$$f_{x,t} = a_x + b_x \cdot k_t + \varepsilon_{x,t}, \tag{7}$$

where x = 15–49, 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 k_t changes and finally k_t are the time-varying parameters – the 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 (1991), Lee, Carter (1992), or Hyndman, Ullah (2010). The age-specific fertility rates $f_{x,t}$ at age x and time t create $35 \times T$ dimensional matrix

$$\mathbf{F} = \mathbf{A} + \mathbf{B}\mathbf{K}^{\mathrm{T}} + \mathbf{E}, \tag{8}$$

and the identification of Lee-Carter model is ensured by

$$\sum_{x=15}^{49} b_x = 1 \quad \text{and} \quad \sum_{t=1}^{T} k_t = 0. \tag{9}$$

Finally,

$$a_x = \frac{\sum_{t=1}^{T} f_{x,t}}{T} \tag{10}$$

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 (Box, Jenkins 1970). The values of the parameters a_x and b_x

are independent of time and the prediction using the Lee-Carter model is therefore purely extrapolative (Lee, Tuljapurkar, 1994).

Czech Republic has, unfortunately, very variable development of data of age-specific fertility rates. Therefore we estimate the parameters a_x, b_x and k_t for complete model based on data 1925–2012 (LC 1925), and also for 3 shortened models, based on data 1948–2012 (LC 1948), 1968–2012 (LC 1968) and 1988–2012 (LC 1988). We present the detailed results provided by LC 1925 and LC 1988 model only. Results from complete model LC 1925 provide misleading predictions, because the average age-specific fertility profile a_x is heavily biased by high level of fertility during the period between 50s and 80s. Therefore, the shorter is the analysed database of fertility, the more realistic results for the Czech population can be expected. The second model interpreted also in detail is the LC 1988. Its results are closest to the expected reality. In the case of fertility analysis by Lee-Carter model it is not a priority to analyse the longest time series, but the most stable ones. Our application of cross-sectional fertility rates by age is primarily intended to identify the parameters of changes for particular time periods. It is better for fertility projection to rely on shorter time series with the newest known development, because in the case of fertility this development is significantly affected by decision of people (opposed to mortality, where the development is influenced by mortality law and other factors). Human decision making is currently more social phenomenon than biological.

2 RESULTS

Firstly we universally look at 35 time series of logarithms of age-specific fertility rates as on the random walk process and estimate 35 drifts. Our second used approach is the ARIMA (0,1,1) process with drift. The drift is included into the model, because there is the higher probability that the model will not involve the autocorrelation. This is made even though there is a risk that many drifts in the model will be statistically insignificant (equal to zero). Estimated drifts for random walk models (in logarithms) are shown in Table 1.

Table 1 Estimated drifts (in logarithms) for individual Random Walk models. Each drift was calculated for individual time series of logarithms of age-specific fertility rates in Statgraphics Centurion XVI

Age	**15**	**16**	**17**	**18**	**19**	**20**	**21**
Drift	–0.001585	0.001479	–0.001347	–0.005822	–0.009456	–0.011595	–0.012407
Age	**22**	**23**	**24**	**25**	**26**	**27**	**28**
Drift	–0.013151	–0.012464	–0.011278	–0.009754	–0.007786	–0.005407	–0.003361
Age	**29**	**30**	**31**	**32**	**33**	**34**	**35**
Drift	–0.002307	–0.001229	–0.001165	–0.001894	–0.002166	–0.003306	–0.004398
Age	**36**	**37**	**38**	**39**	**40**	**41**	**42**
Drift	–0.005540	–0.007668	–0.009417	–0.011651	–0.013910	–0.015735	–0.019497
Age	**43**	**44**	**45**	**46**	**47**	**48**	**49**
Drift	–0.021675	–0.025519	–0.031128	–0.022158	–0.024409	–0.021511	–0.030460

Source: Author's calculation

Estimated components MA(1) for all 35 time series of logarithms of age-specific fertility rates are shown in Table 2. It is clear (grey highlighted values), that 9 of 35 parameters are statistically insignificant at 5% significance level. It is not a bad result for situation, when we selected a universal model for all series. Hyndman et al. (2002) have dealt with situations and issues, where the universal form of model and automatic forecasting was based on a much larger number of broken assumptions. We perform the diagnostics of two approaches on autocorrelation tests. The Box-Pierce test (Box, Pierce, 1970) is implemented in an automated process of automatic forecasting system in Statgraphics Centurion XVI. We test the null hypothesis: there is no autocorrelation, and the results for the random walk model with drift are shown in Table 3, the results for the ARIMA (0,1,1) model with drift are in Table 4.

Table 2 Estimated parameters for individual ARIMA (0,1,1) models with drift. Each model was calculated for individual time series of logarithms of age-specific fertility rates in Statgraphics Centurion XVI. Most of drifts are statistically insignificant at the 5% significance level – but the drifts are included due to capture autocorrelation

Age 15	**Est.**	**s.e.**	**t-stat**	**P**	**Age 16**	**Est.**	**s.e.**	**t-stat**	**P**	**Age 17**	**Est.**	**s.e.**	**t-stat**	**P**
MA(1)	0.157	0.118	1.325	0.190	MA(1)	−0.240	0.107	−2.250	0.027	MA(1)	−0.216	0.106	−2.040	0.044
C	−0.002	0.017	−0.113	0.910	C	0.002	0.015	0.110	0.913	C	−0.001	0.015	−0.084	0.934
Age 18	**Est.**	**s.e.**	**t-stat**	**P**	**Age 19**	**Est.**	**s.e.**	**t-stat**	**P**	**Age 20**	**Est.**	**s.e.**	**t-stat**	**P**
MA(1)	−0.593	0.091	−6.492	0.000	MA(1)	−0.498	0.091	−5.493	0.000	MA(1)	−0.304	0.104	−2.940	0.004
C	−0.004	0.014	−0.311	0.756	C	−0.009	0.012	−0.766	0.446	C	−0.011	0.010	−1.129	0.262
Age 21	**Est.**	**s.e.**	**t-stat**	**P**	**Age 22**	**Est.**	**s.e.**	**t-stat**	**P**	**Age 23**	**Est.**	**s.e.**	**t-stat**	**P**
MA(1)	−0.285	0.106	−2.681	0.009	MA(1)	−0.287	0.103	−2.773	0.007	MA(1)	−0.321	0.102	−3.143	0.002
C	−0.012	0.011	−1.126	0.263	C	−0.013	0.009	−1.500	0.137	C	−0.012	0.008	−1.470	0.145
Age 24	**Est.**	**s.e.**	**t-stat**	**P**	**Age 25**	**Est.**	**s.e.**	**t-stat**	**P**	**Age 26**	**Est.**	**s.e.**	**t-stat**	**P**
MA(1)	0.010	0.108	0.093	0.926	MA(1)	−0.187	0.108	−1.736	0.086	MA(1)	−0.211	0.106	−1.997	0.049
C	−0.011	0.008	−1.460	0.148	C	−0.010	0.007	−1.361	0.177	C	−0.008	0.007	−1.127	0.263
Age 27	**Est.**	**s.e.**	**t-stat**	**P**	**Age 28**	**Est.**	**s.e.**	**t-stat**	**P**	**Age 29**	**Est.**	**s.e.**	**t-stat**	**P**
MA(1)	−0.143	0.108	−1.324	0.189	MA(1)	−0.123	0.108	−1.147	0.255	MA(1)	−0.282	0.103	−2.737	0.008
C	−0.005	0.008	−0.659	0.512	C	−0.003	0.008	−0.415	0.679	C	−0.002	0.009	−0.256	0.798
Age 30	**Est.**	**s.e.**	**t-stat**	**P**	**Age 31**	**Est.**	**s.e.**	**t-stat**	**P**	**Age 32**	**Est.**	**s.e.**	**t-stat**	**P**
MA(1)	−0.114	0.108	−1.058	0.293	MA(1)	−0.246	0.105	−2.355	0.021	MA(1)	−0.252	0.104	−2.416	0.018
C	−0.001	0.010	−0.125	0.901	C	−0.001	0,011	−0.105	0.916	C	−0.002	0.011	−0.173	0.863
Age 33	**Est.**	**s.e.**	**t-stat**	**P**	**Age 34**	**Est.**	**s.e.**	**t-stat**	**P**	**Age 35**	**Est.**	**s.e.**	**t-stat**	**P**
MA(1)	−0.249	0.104	−2.405	0.018	MA(1)	−0.281	0.103	−2.720	0.008	MA(1)	−0.281	0.102	−2.748	0.007
C	−0.002	0.012	−0.183	0.855	C	−0.003	0.012	−0.256	0.798	C	−0.004	0.013	−0.316	0.753
Age 36	**Est.**	**s.e.**	**t-stat**	**P**	**Age 37**	**Est.**	**s.e.**	**t-stat**	**P**	**Age 38**	**Est.**	**s.e.**	**t-stat**	**P**
MA(1)	−0.299	0.102	−2.927	0.004	MA(1)	−0.366	0.101	−3.619	0.001	MA(1)	−0.174	0.107	−1.618	0.109
C	−0.005	0.013	−0.420	0.676	C	−0.008	0.014	−0.557	0.579	C	−0.009	0.012	−0.745	0.458
Age 39	**Est.**	**s.e.**	**t-stat**	**P**	**Age 40**	**Est.**	**s.e.**	**t-stat**	**P**	**Age 41**	**Est.**	**s.e.**	**t-stat**	**P**
MA(1)	−0.305	0.103	−2.961	0.004	MA(1)	−0.370	0.100	−3.698	0.000	MA(1)	−0.248	0.107	−2.314	0.023
C	−0.011	0.014	−0.842	0.402	C	−0.014	0.014	−0.975	0.332	C	−0.015	0.016	−0.937	0.351
Age 42	**Est.**	**s.e.**	**t-stat**	**P**	**Age 43**	**Est.**	**s.e.**	**t-stat**	**P**	**Age 44**	**Est.**	**s.e.**	**t-stat**	**P**
MA(1)	−0.076	0.109	−0.694	0.489	MA(1)	0.051	0.108	0.473	0.638	MA(1)	0.083	0.108	0.772	0.443
C	−0.019	0.014	−1.351	0.180	C	−0.022	0.016	−1.347	0.182	C	−0.025	0.018	−1.430	0.156
Age 45	**Est.**	**s.e.**	**t-stat**	**P**	**Age 46**	**Est.**	**s.e.**	**t-stat**	**P**	**Age 47**	**Est.**	**s.e.**	**t-stat**	**P**
MA(1)	0.331	0.102	3.251	0.002	MA(1)	0.343	0.105	3.262	0.002	MA(1)	0.418	0.098	4.264	0.000
C	−0.031	0.018	−1.712	0.091	C	−0.024	0.022	−1.135	0.259	C	−0.024	0.023	−1.074	0.286
Age 48	**Est.**	**s.e.**	**t-stat**	**P**	**Age 49**	**Est.**	**s.e.**	**t-stat**	**P**					
MA(1)	0.610	0.087	7.043	0.000	MA(1)	0.368	0.102	3.615	0.001					
C	−0.021	0.032	−0.676	0.501	C	−0.027	0.022	−1.211	0.229					

Source: Author's calculation

Approach of random walk models with drift is clearly worse after the evaluation by autocorrelation tests. The development system was not well explained in 23 cases from 35, there was left too much unexplained variability and the residues are auto-correlated. Therefore the estimates will be statistically distorted and skewed. The situation is much better in ARIMA (0,1,1) model with drift. Although there are many statistically insignificant drifts at the 5% significance level in this approach, their inclusion into the model fulfilled its goal. Only five models has its residues auto-correlated and therefore there is a real risk of statistical bias only in five cases. The difference between forecasts predicted by relatively bad and by relatively good model will be presented later in Figure 2 and the final comparison of all approaches will be provided in Figure 7.

Table 3 Diagnostic control of individual Random walk models with drift – Box-Pierce serial autocorrelation tests. Null hypothesis: There is no autocorrelation, unfortunately rejected in 23 cases from 35 at 5% significance level (grey highlighted values). This model is not good. (TC = Test Criterion)

Age	**15**	**16**	**17**	**18**	**19**	**20**	**21**
Box-Pierce TC	25.772	23.460	16.528	38.436	81.986	83.879	44.468
P-value	0.262	0.493	0.868	0.031	0.000	0.000	0.007
Age	**22**	**23**	**24**	**25**	**26**	**27**	**28**
Box-Pierce TC	73.289	43.349	19.348	16.990	25.259	17.286	26.033
P-value	0.000	0.009	0.733	0.849	0.392	0.836	0.352
Age	**29**	**30**	**31**	**32**	**33**	**34**	**35**
Box-Pierce TC	40.916	22.742	41.734	45.883	43.472	47.832	45.205
P-value	0.017	0.535	0.014	0.005	0.009	0.003	0.006
Age	**36**	**37**	**38**	**39**	**40**	**41**	**42**
Box-Pierce TC	47.239	48.553	50.099	49.782	56.903	30.147	56.238
P-value	0.003	0.002	0.001	0.002	0.000	0.180	0.000
Age	**43**	**44**	**45**	**46**	**47**	**48**	**49**
Box-Pierce TC	23.729	38.620	54.133	43.958	62.288	22.497	105.053
P-value	0.477	0.030	0.000	0.008	0.000	0.550	0.000

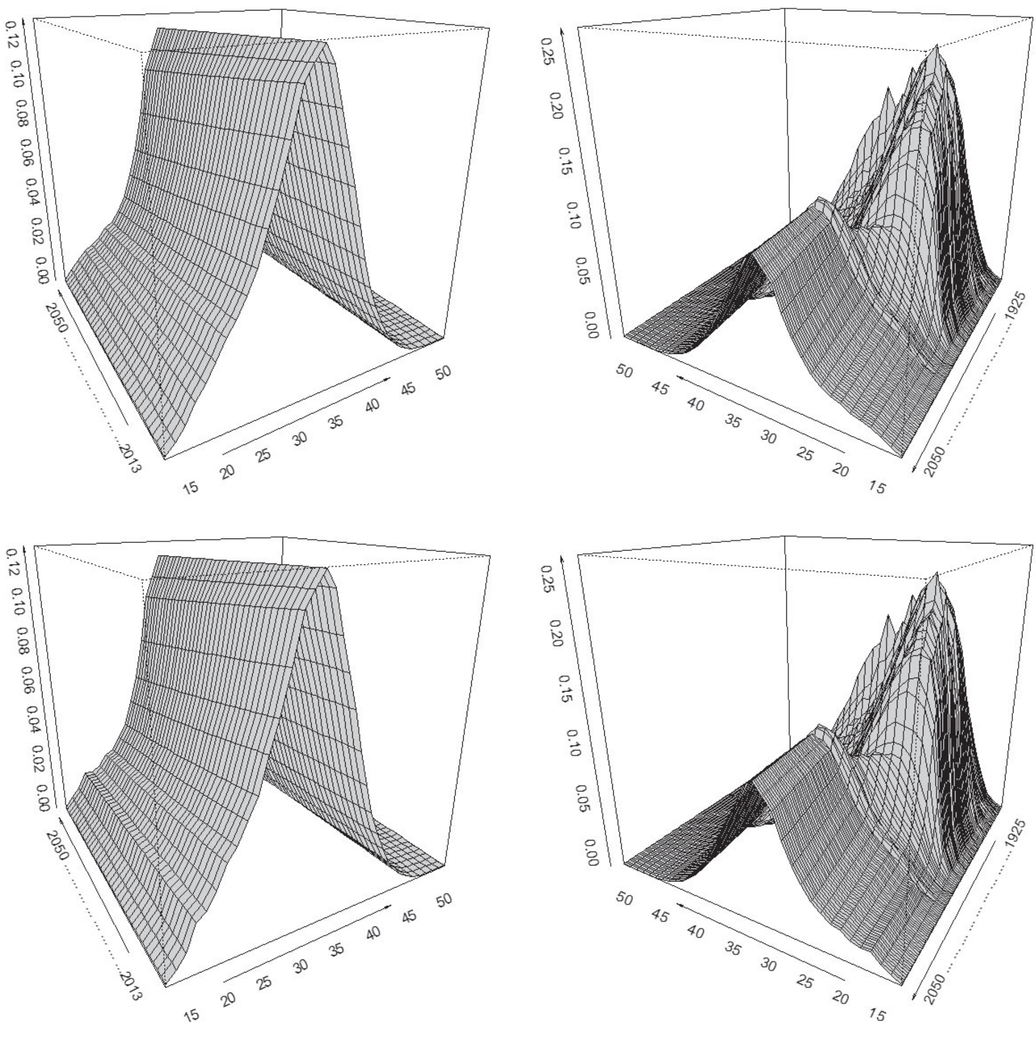
Source: Author's calculation

Table 4 Diagnostic control of individual ARIMA (0,1,1) models with drift – Box-Pierce serial autocorrelation tests. Null hypothesis: There is no autocorrelation, rejected only in 5 cases from 35 at 5% significance level (grey highlighted values). This model is much better than Random walk with drift (Table 3). (TC = Test Criterion)

Age	**15**	**16**	**17**	**18**	**19**	**20**	**21**
Box-Pierce TC	17.859	15.711	12.936	13.367	27.744	39.360	24.183
P-value	0.658	0.867	0.953	0.944	0.226	0.018	0.394
Age	**22**	**23**	**24**	**25**	**26**	**27**	**28**
Box-Pierce TC	45.409	24.697	19.289	13.538	21.157	15.633	24.896
P-value	0.004	0.366	0.684	0.939	0.571	0.871	0.356
Age	**29**	**30**	**31**	**32**	**33**	**34**	**35**
Box-Pierce TC	23.346	20.970	28.974	26.779	23.874	24.790	24.936
P-value	0.441	0.583	0.181	0.265	0.411	0.361	0.354
Age	**36**	**37**	**38**	**39**	**40**	**41**	**42**
Box-Pierce TC	21.860	24.809	35.622	24.346	28.506	19.886	61.300
P-value	0.529	0.360	0.045	0.385	0.197	0.649	0.000
Age	**43**	**44**	**45**	**46**	**47**	**48**	**49**
Box-Pierce TC	23.563	34.705	29.705	19.857	23.006	9.162	52.091
P-value	0.428	0.056	0.158	0.651	0.460	0.995	0.000

Source: Author's calculation

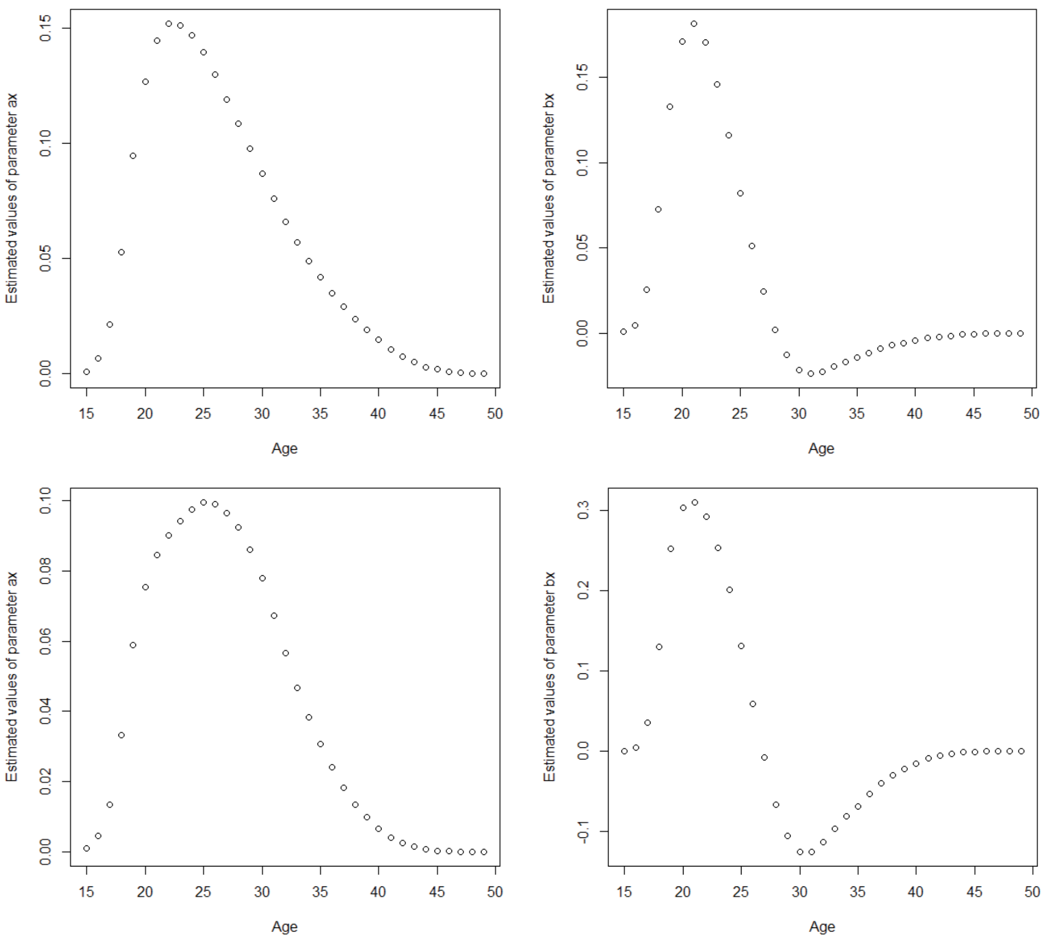
Figure 2 Forecasted values of age-specific fertility rates $f_{x,t}$ of Czech females for the period 2013 to 2050 by Random Walk model with drift (top left) and the empirical values of these rates for the period 1925 to 2012 with attached forecasts (top right). Forecasted values of age-specific fertility rates $f_{x,t}$ by ARIMA (0,1,1) models with drift and the empirical values of these rates with attached forecasts are shown bottom left, bottom right respectively



Source: CZSO (2013), author's construction and illustration

We estimate the parameters $\hat{a}_x$ (age-specific fertility profiles independent in time) and $\hat{b}_x$ (additional age-specific components determine how much each age group changes when k_t changes) for 4 Lee-Carter's models (LC 1925, LC 1948, LC 1968 and LC 1988) using the SVD method implemented in the package "demography" (Hyndman, 2012), which is developed for RStudio. We can see the parameters for LC 1925 and LC 1988 in the Figure 3, from which it is also clear the comparison between the different evolutions of these parameters. The age-specific fertility profiles independent of time ($\hat{a}_x$) are lower in the shortened model (LC 1988), because in the considered period there were already the fertility rates of Czech females lower. Also this profile is deflected to the right (to the highest age groups). Given that the length of the analysed time series is shorter, the variability of the estimated additional age-specific components ($\hat{b}_x$) is higher, especially at the advanced ages. The fertility indices $\hat{k}_t$ (the time-varying parameters) were estimated for the period 1925 to 2012 (LC 1925) and 1988 to 2012 (LC 1988). The estimates are provided in the Figure 4. There were calculated the predictions up to the year 2050 to these estimates based on the methodological approach of ARIMA, (Box, Jenkins, 1970) and ran by "forecast" package in RStudio (Hyndman et al., 2002, Hyndman, 2012). Parameters of ARIMA models are displayed in Table 5.

Figure 3 Comparison of two Lee-Carter's models – The estimates of age-specific fertility profiles independent in time (parameter $\hat{a}_x$, left) and additional age-specific components determine how much each age group changes when $\hat{k}_t$ changes (parameter $\hat{b}_x$, right). Top charts represent the model based on data for the period 1925 to 2012, bottom charts the model based on data for the period 1988 to 2012



Source: Author's construction and illustration

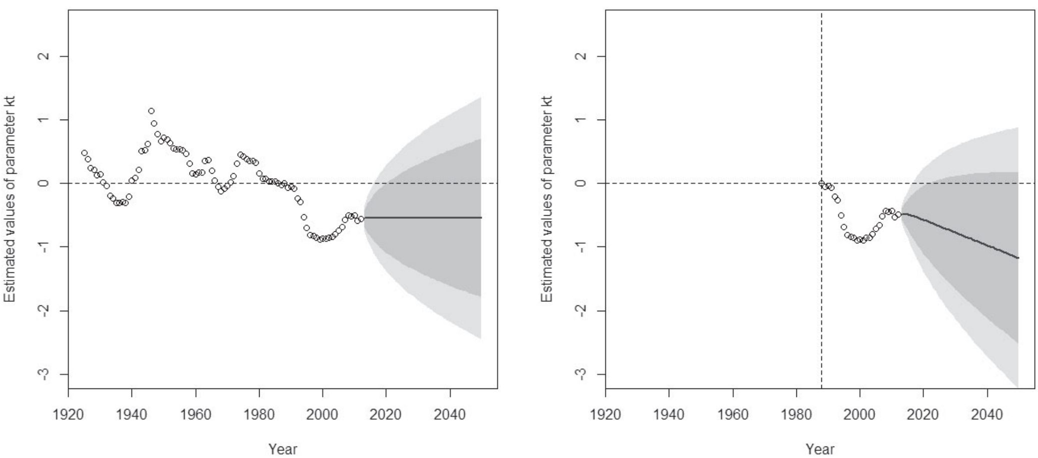
It is clear from these predictions with 95% confidence intervals (which can be seen in Figure 4 too) that the LC 1988 model provides lower values of these estimates (decreasing trend). Confidence intervals are slightly wider at the case of LC 1988 model. Now we evaluate two Lee-Carter models on the basis of approach, which is presented by Charpentier, Dutang, (2012). Using RStudio we display the Pearson's residuals first for the LC 1925 and then for the LC 1988 model. Each model is evaluated on the basis of the residues by age x and of the residues at time t.

Table 5 Estimated parameters of two ARIMA models for parameter $\hat{k}_t$ of two Lee-Carter's models (LC 1925 and LC 1988)

Parameter k_t, Lee-Carter 1925			**Parameter k_t, Lee-Carter 1988**		
ARIMA (1,1,0) without drift			ARIMA (1,1,0) without drift		
Coefficients:			Coefficients:		
	AR(1)	**Drift**		**AR(1)**	**Drift**
	0.3494	x		0.6094	x
s.e.	0.1000	x	s.e.	0.1577	x
[t-stat]	3.4940	x	[t-stat]	3.8643	x
AIC= –142.98	AICc = –142.84	BIC= –138.05	AIC= –52.44	AICc= –51.24	BIC= –48.91

Source: Author's calculation

Figure 4 Comparison of two Lee-Carter's models – The estimates of the time-varying parameters $\hat{k}_t$ – the fertility indices. On the left side is the model based on data for the period 1925 to 2012, on the right side is the model based on data for the period 1988 to 2012



Source: Author's construction and illustration

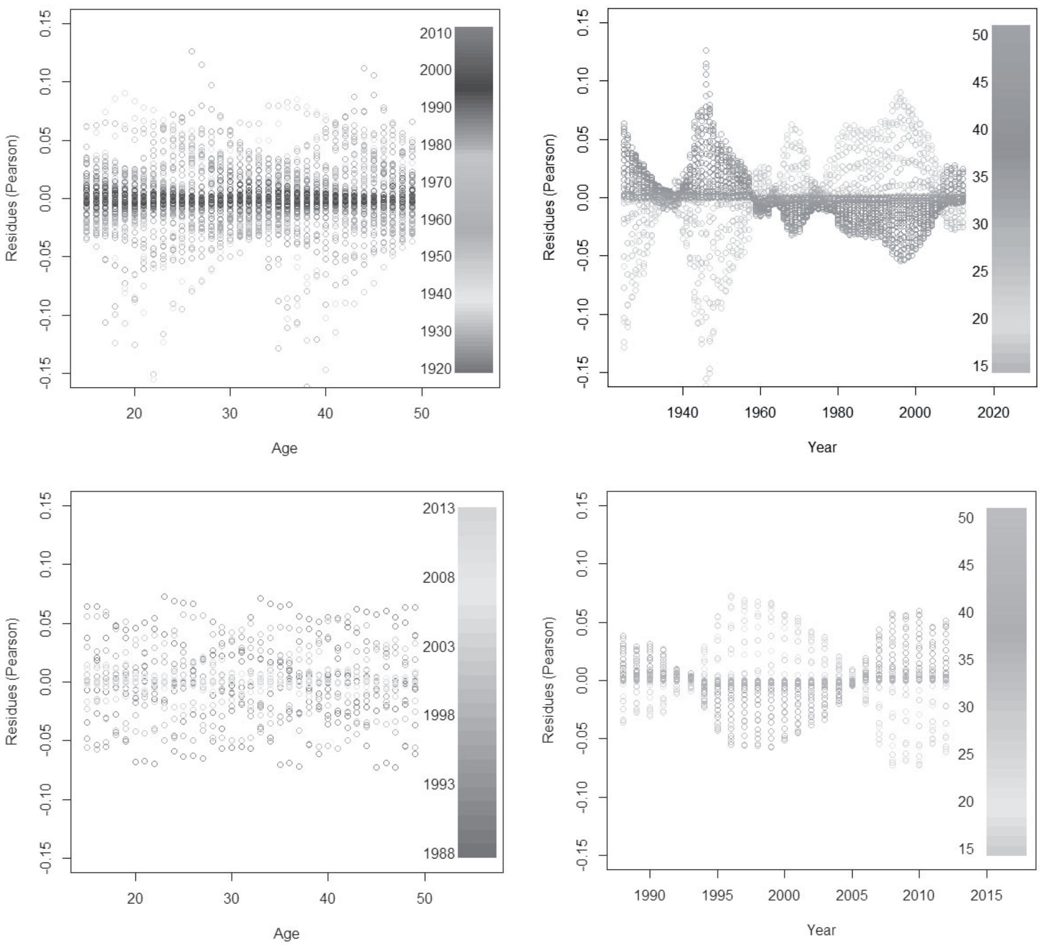
The most residues are concentrated around 0, the more variability is explained by the estimated model. The Pearson's residues for LC 1925 model are shown in the Figure 5 (top), where residues by age x are on the left side and the residues at time t are on the right side. Given that this model also includes the normalisation period, it is understandable that the residues will be much more variable than in the case of shortened model LC 1988. The residues of the shortened model are shown in the Figure 5 (bottom).

Based on the estimated parameters $\hat{a}_x$, $\hat{b}_x$ and $\hat{k}_t$ of two Lee-Carter's models we can estimate the future values of $f_{x,t}$ as:

$$f_{x,t} = \hat{a}_x + \hat{b}_x \cdot \hat{k}_t. \tag{11}$$

Estimated values (left) and the empirical values with the attached estimates (right) of $f_{x,t}$ based on LC 1925 model are displayed in 3D perspective chart in the Figure 6 (top). The estimated values based on the shortened model LC 1988 (left) and then the empirical values with these attached forecasts (right) are displayed below.

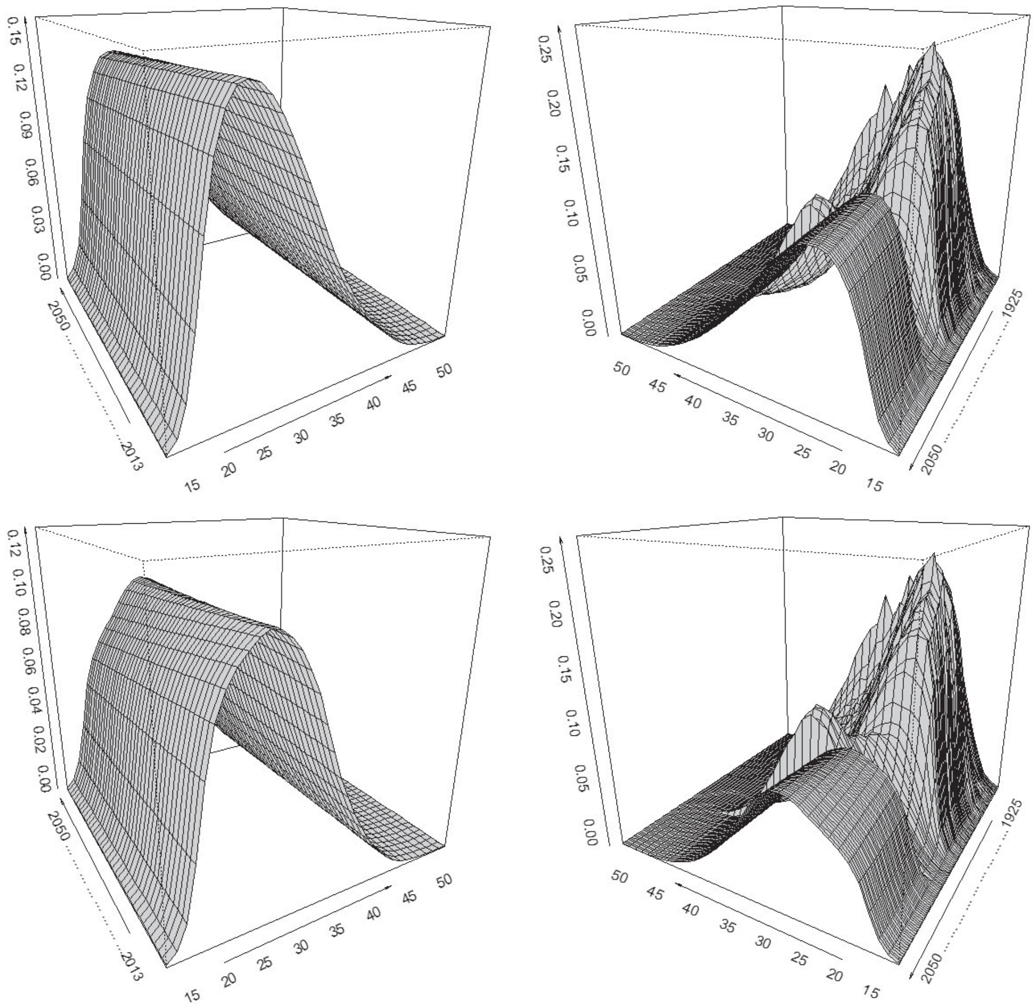
Figure 5 Diagnostic control of the Lee-Carter's model – Pearson's residues (model based on data for the period 1925 to 2012 – top charts, for the period 1988 to 2012 – bottom charts respectively)



Source: Author's construction and illustration

The predicted values by LC 1925 model are unreasonably high (see Figure 6). This inadequacy is caused by non-robustness of Lee-Carter model for the case of fertility, because the prediction of this model is strongly influenced by the average fertility profile independent of time (parameter a_x). The average profile is deflected by high level of fertility of Czech females in the post-war period and during the pro-population policies implemented under the previous regime (Communist party of Czechoslovakia). Excessively high values of age-specific fertility rates ($f_{x,t}$) are not tied up with the empirical data. There is particularly a significant decline in the case of the distribution's mode, which is sharply and vigorously returned to the lower ages. The predicted values $f_{x,t}$ by LC 1988 model are much lower. The parameter a_x is not so much affected by the high fertility rates arising in the period of normalization and the further projection looks more realistic. Unfortunately, even in this case there is not a fluent connection of predictions to the empirical data of $f_{x,t}$, because we can see that the mode of this distribution is not retained at its original level, but also moved back into the lower ages.

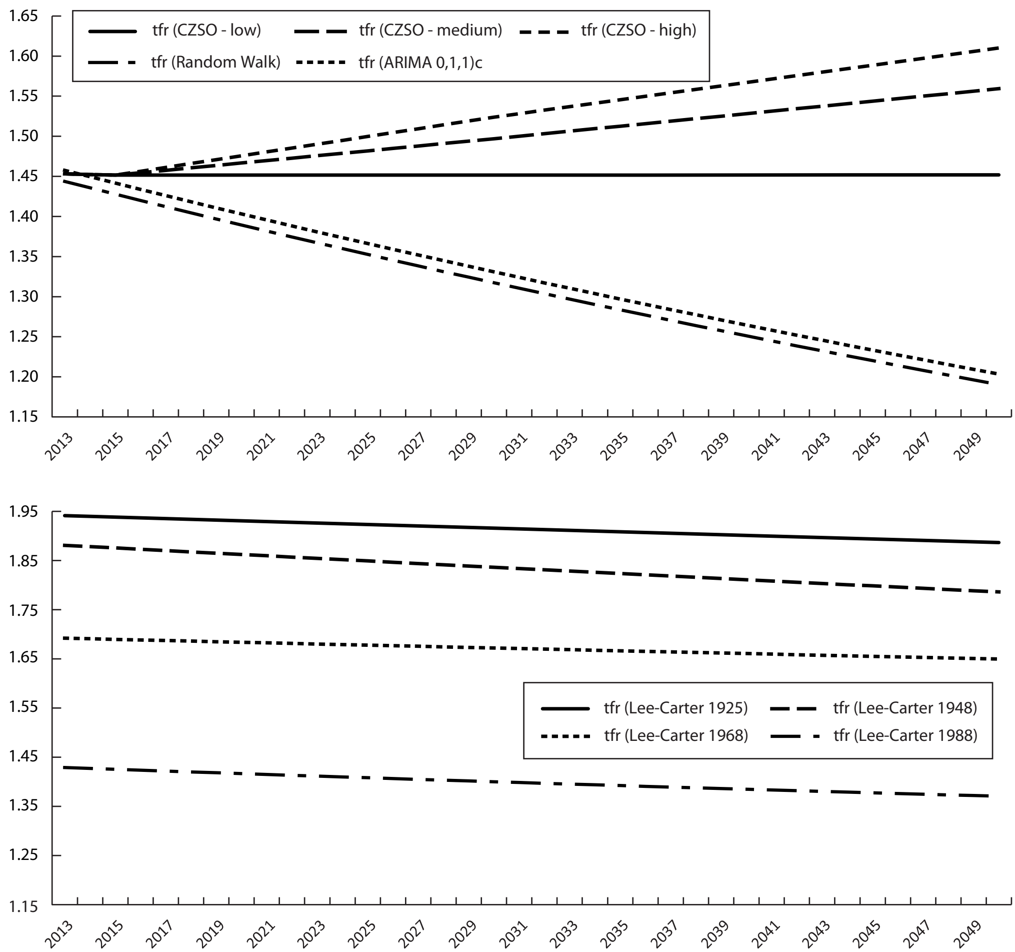
Figure 6 Forecasted values of age-specific fertility rates $f_{x,t}$ of Czech females for the period 2013 to 2050 by Lee-Carter's model based on full data matrix for the period 1925 to 2012 (top left), by Lee-Carter's model based on shortened data matrix for the period 1988 to 2012 (bottom left) respectively, and the empirical values of these rates for the period 1925 to 2012 with attached forecasts based on full Lee-Carter's model (top right), based on shortened model (bottom right) respectively



Source: CZSO (2013), author's construction and illustration

Following section compares the approaches of random walk models with drift, ARIMA (0,1,1) models with drift, LC 1925, LC 1948, LC 1968 and LC 1988 models, (regardless of the fact that models LC 1948 and LC 1968 have not been commented in detail). This evaluation is based on the results of total fertility rates calculated according to formula (2). It is clear from Figure 7 (right) that the models LC 1925, LC 1948 and LC 1968 are useless, because they provide the unrealistic values of total fertility for the situation in the Czech Republic. This fertility development does not follow the current fertility level and is quite skewed. Successive reduction of the input base of the Lee-Carter's model was in this situation quite useless, because the model is not able to respond to the dynamically changing fertility development of Czech females. Models of random walks with drift, ARIMA (0,1,1) models with drift, (and also LC 1988) provide more realistic results (please see Figure 7). Beginnings of prediction start at the current fertility level (1.45) and the predicted values for the future slowly decrease up to the values of 1.190, 1.203 and 1.371 respectively.

Figure 7 Forecast of the total fertility rates for Czech females up to the year 2050 based on prediction by the Czech Statistical Office in low, medium and high scenario and by the prediction by Random Walk model with drift (left chart) and four Lee-Carter's models (one based on data for the period 1925 to 2012, the second one for the period 1948 to 2012, the third for the period 1968 to 2012 and the fourth for the period 1988 to 2012 respectively



Source: CZSO (2013), author's construction and illustration

The prediction provided by ARIMA (0,1,1) model with drift seems to be the most acceptable. Model was relatively positively evaluated and its predicted values are probable. The approach of random walk models with drift, which was unfavourably evaluated, provides not so much different the expected future development. Some of the predicted values of $f_{x,t}$ will be probably skewed due to the consequences of poor model, but the difference should be really negligible in the summary.

We can also see in the Figure 7 the predictions that in their low, medium and high variant publishes the CZSO. Medium and high variant has a growing character, the low one has in all the time of horizon the same level as today. It is really questionable whether the future values of the total fertility rates in the Czech Republic actually rise, fall, or be rather constant at +/– 1.45.

DISCUSSION AND CONCLUSION

The aim of this paper was to construct forecasts of age-specific fertility rates $f_{x,t}$ of Czech females for the period 2013–2050 using different approaches. One of them was the Box-Jenkins methodology for modelling of 35 individual time series $f_{x,t}$ (x = 15–49) on an annual basis of 88 observations (t = 1925–2012). The second approach was the Lee-Carter model for identifying the major components explaining the level of fertility. The sensitivity of this model was to evaluate a total of 4 cases, where we gradually analysed and shortened the different length of $x \times T$ dimensional matrix of $f_{x,t}$. We concluded that the Lee-Carter model should be used for the shortest possible time series development as it is strongly influenced by fluctuations of the past. The average age-specific fertility profiles independent of time are affected by the different shape of the distribution $f_{x,t}$, created during the previous regime and the predicted values are due to this affection largely distorted. More realistic forecasts were provided by Box-Jenkins methodological approach. Looking at the shape of the distribution of predicted $f_{x,t}$, in Figure 2, there are no doubts that they are meaningful. These values were obtained using the found trend of development of each individual time series in the past, while the largest weights are set on the newest values. This implies that the estimated prediction describe the best expected future trend. Annex of the paper (Figure 9 to Figure 12) display all 4 calculated predictions in 5-year intervals using simple X-Y charts.

The values of $f_{x,t}$, which are expected by CZSO are optimistic. *We do not want to say that the expected values of future fertility of Czech females according to CZSO forecasts are wrong.* These expectations are based on expert judgments of professionals from different scientific disciplines and have its reasons. We would like to just point to the fact that the statistical trend, which does not take into account the expert discussed expectation is different – declining. Thus, there is a certain degree of probability that the expected future development of fertility will be rather “low” (pessimistic) variant of the CZSO. The goal for the future research is to find and elaborate a robust approach to modelling fertility in the Czech conditions based on literature review. This approach should not be so much affected by high average of fertility from previous regime. We will use the whole data matrix that is available for a better explanation of random effects.

ACKNOWLEDGEMENT

The author gratefully acknowledge to the Czech Science Foundation project No. P402/12/G097 DYME – “Dynamic Models in Economics” for supporting this paper.

References

ARLTOVÁ, M. *Stochastické metody modelování a předpovídání demografických procesů* [habilitation thesis]. Prague: University of Economics Prague, 2011.

ALDERS, M., DE BEER, J. Assumptions on fertility in stochastic population forecasts. *International statistical review*, 2004, Vol. 72, Iss. 1, pp. 65–79.

BELL, W. R., MONSELL, B. Using principal components in time series modelling and forecasting of age-specific mortality rates. In: *Proceedings of the American Statistical Association, Social Statistics Section*, 1991, pp. 154–159.

BELL, W. R. Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates. *Journal of Official Statistics*, 1997, Vol. 13, No. 3, pp. 279–303.

BOX, G. E. P., JENKINS, G. *Time series analysis: Forecasting and control*. San Francisco: Holden-Day, 1970.

BOX, G. E. P., PIERCE, D. A. Distribution of the Autocorrelations in Autoregressive Moving Average Time Series Models. *Journal of the American Statistical Association*, 1970, Vol. 65, pp. 1509–1526.

CAPUTO, M., NICOTRA, M., GLORIA-BOTTINI, E. Fertility Transition: Forecast for Demography. *Human biology*, 2008, Vol. 80, Iss. 4, pp. 359–376.

CHARPENTIER, A., DUTANG, CH. *L'Actuariat avec R*. [working paper]. Paternite-Partage a lindentique 3.0 France de Creative Commons, Decembre 2012.

CZSO. *Projekce obyvatelstva České republiky do roku 2100* [online]. 2013. [cit.: 09.09.2014] <http://www.czso.cz/csu/2013edicniplan.nsf/publ/4020-13-n_2013>.

DOTLAČILOVÁ, P., ŠIMPACH, O., LANGHAMROVÁ, J. The Use of Polynomial Functions for Modelling of Mortality at the Advanced Ages. In: *Mathematical Methods in Economics 2014*. Olomouc: Palacký University in Olomouc, 2014, p. 174–179.

FIALA, T., LANGHAMROVÁ, J. What Rate of Fertility and Extent of Migration Would Be Needed for Stable Population Development in the Czech Republic in This Century? *Demografie*, 2012, Vol. 54, No. 4, pp. 382–404.

HYNDMAN, R. J., BOOTH, H. Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting*, 2008, Vol. 24, Iss. 3, pp. 323–342.

HYNDMAN, R. J., KOEHLER, A. B., SNYDER, R. D., GROSE, S. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 2002, Vol. 18, Iss. 3, pp. 439–454.

HYNDMAN, R. J., ULLAH, MD. SHAHID Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, 2010, Vol. 51, Iss. 10, pp. 4942–4956.

HYNDMAN, R. J. *Demography: Forecasting mortality, fertility, migration and population data* [online]. 2010, R package v. 1.16. <http://robjhyndman.com/software/demography>.

LANGHAMROVÁ, J., FIALA, T. Změny reprodukčního chování v České republice a jeho důsledky. *Fórum sociální politiky*, 2014, Vol. 8, No. 3, pp. 23–25.

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

LEE, R. D., TULJAPURKAR, S. Stochastic population forecasts for the United States: beyond high, medium, and low. *Journal of the American Statistical Association*, 1994, Vol. 89, pp. 1175–1189.

LI, N., WU, Z. Forecasting cohort incomplete fertility: A method and an application. *Population Studies*, 2003, Vol. 57, No. 3, pp. 303–320.

MELARD, G., PASTEELS, J. M. Automatic ARIMA Modeling Including Intervention, Using Time Series Expert Software. *International Journal of Forecasting*, 2000, Vol. 16, pp. 497–508.

MORGAN, SP., HAGEWEN, K. Is very low fertility inevitable in America? Insights and forecasts from an integrative model of fertility. In: *New Population Problem: Why Families in Developed Countries are Shrinking and What it Means, Penn State University Family Issues Symposia Series*, Mahwah: Lawrence Erlbaum Associates, 2005, pp. 3–28.

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*, 2013, Vol. 39, Iss. 1, pp. 31–56.

PERISTERA, P., KOSTAKI, A. Modeling fertility in modern populations. *Demographic Research*, 2007, Vol. 16, No. 6, pp. 141–194.

RUEDA, C., RODRIGUEZ, P. State space models for estimating and forecasting fertility. *International Journal of Forecasting*, 2010, Vol. 26, Iss. 4, pp. 712–724.

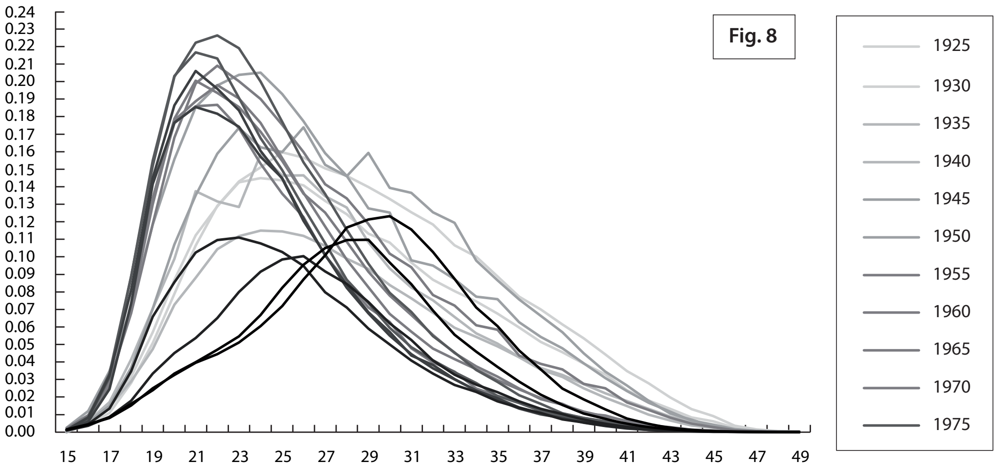
STAUFFER, D. Simple tools for forecasts of population ageing in developed countries based on extrapolations of human mortality, fertility and migration. *Experimental Gerontology*, 2002, Vol. 37, Iss. 8–9, pp. 1131–1136.

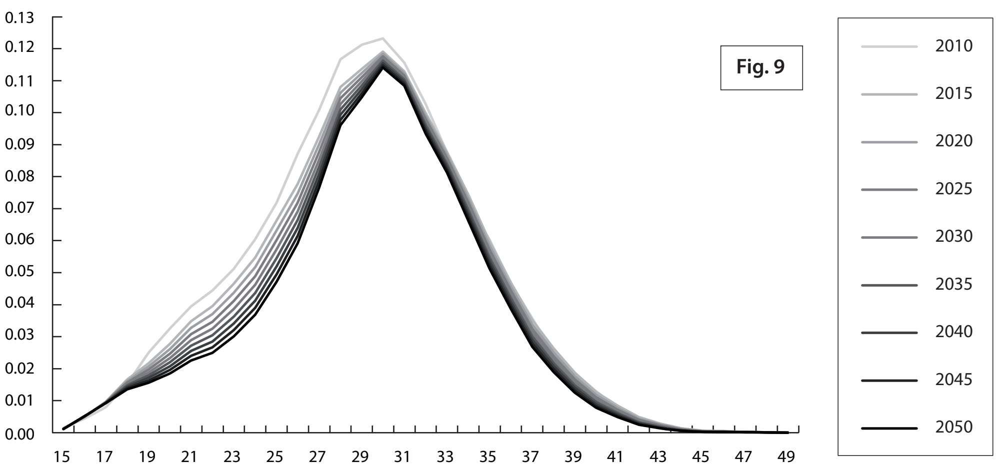
ŠIMPACH, O., DOTLAČILOVÁ, P., LANGHAMROVÁ, J. Effect of the Length and Stability of the Time Series on the Results of Stochastic Mortality Projection: An application of the Lee-Carter model. In: *ITISE 2014*. Granada: University of Granada, 2014, p. 1375–1386.

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

ANNEX

Figure 8–12 Empirical values of age-specific fertility rates $f_{x,t}$ of Czech females in the period 1925–2010 by 5years intervals (Figure 8). Forecasted values of age-specific fertility rates up to the year 2050 in 5years intervals by Random Walk model with drift (Figure 9), by ARIMA (0,1,1) models with drift (Figure 10), by Lee-Carter's model based on full data matrix for the period 1925 to 2012 (Figure 11) and by Lee-Carter's model based on shortened data matrix for the period 1988 to 2012 (Figure 12)





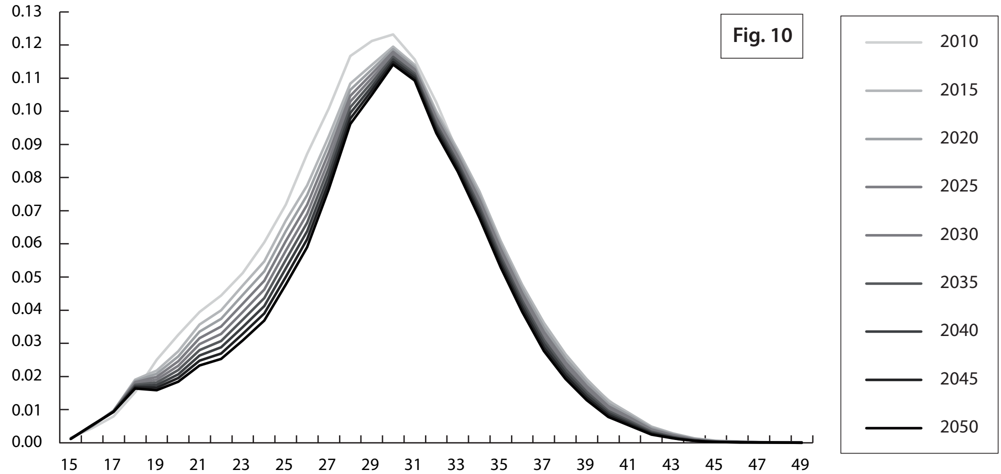


Fig. 10

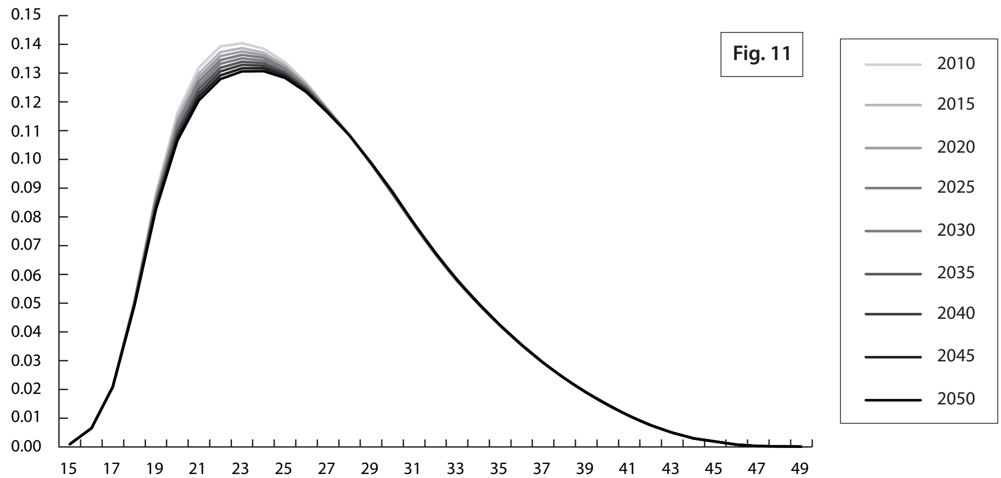


Fig. 11

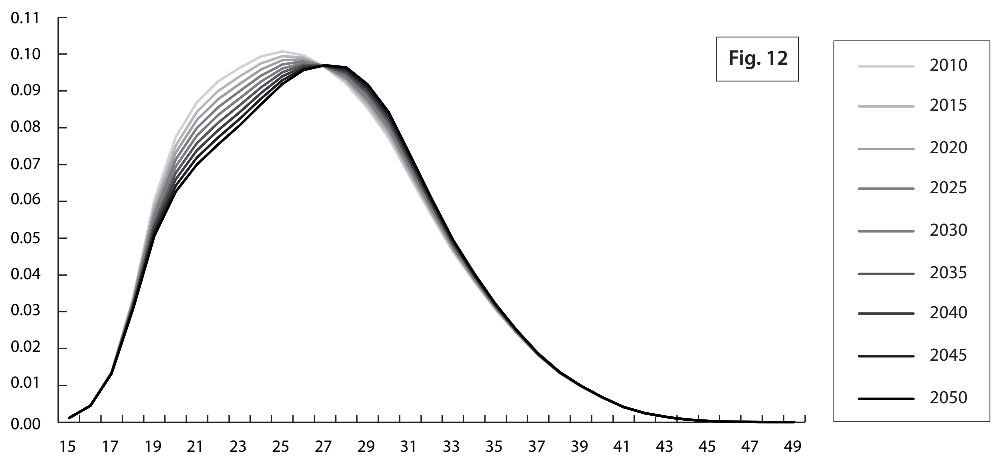


Fig. 12

Source: CZSO (2013), author's construction and illustration