Logistic and ARIMA models in the Estimation of Life Expectancy in the Czech Republic

Ondřej Šimpach[1], Petra Dotlačilová[2], Jitka Langhamrová[3]

Abstract. The aim of the presented paper will be the calculation of the estimated life expectancy at birth for males and females in the Czech Republic using selected levelling functions (Gompertz–Makeham, Kanistö and Thatcher) and using ARIMA model and Random Walk with Drift, constructed for the time series modelling. The levelling functions for modelling mortality rate are nonlinear, thus difficult to solve, ARIMA models are in turn based on entirely different principles (stochastic process and backward looking expectations). Subsequently it will be pointed to differences that arise when using both methods and the differences that occur when we compare the estimates that are published by the Czech Statistical Office. Given that the estimates of life expectancy at birth using different approaches are similar, then based on ARIMA model and Random Walk with Drift there will be extrapolated to the near future (2012–2030). Presented alternative estimates will be correlated with predictions which are calculated by the Czech Statistical Office and it will be easier to obtain them, because they are based purely on a statistical approach, which does not require the additional demographic information, which are expensive to obtain.

Keywords: Life Expectancy, ARIMA, Gompertz–Makeham, Kanistö, Thatcher

JEL Classification: C61, C63
AMS Classification: 90C30

1 Introduction

Mortality and its development always has been very interesting topic (not only for demographers). The mortality trend is one of the most important indicators of standard of living (see e.g. Bérenger and Verdier-Chouchane [2]). If people live longer, it means that mortality is going to be better (Arltová et al. [1]). The reason for increasing life expectancy could be e.g. better health care (Jia et al. [15]). The second reason is greater interest in healthy life style. On the other hand the increase in values of life expectancy means population aging (Gavrilov and Gavrilova [11] or Boleslawski and Tabeau [3]). Because more and more people live to the highest ages, it is very important to have the best idea of the trend of mortality at the highest ages. In the previous years it was not so important, because only a few of them live to the highest ages. It is also important to note that data about mortality at the highest ages is unreliable and mortality of oldest persons is different from the younger ones. Therefore it is necessary to use some of the existing models for extrapolating specific mortality rates at the highest ages (see e.g. Burcin et al. [7]).

The aim of the article is to use selected models (Gompertz–Makeham, Thatcher and Kanistö) to smooth the specific mortality rates (see e.g. Gompertz [12], Makeham [17] or Thatcher et al. [19]) and use them to calculate the life expectancy at birth for Czech males and females using available data from 1970 to 2011. Data is published by the Czech Statistical Office (CZSO) and available with the annual frequency. Because the calculation of balanced specific mortality rates (and consequently the life expectancy at birth) are computationally difficult, the values of life expectancy from the life tables will be estimated by an alternative approach – "ex-post" by random walk model with drift and by ARIMA approach (see e.g. Šimpach and Langhamrová [18] or Torri [20]). These estimates will be equivalent to the estimates that provide selected logistic models. Due to the statistically significant estimates of life expectancy there will be performed the extrapolation to subsequent 19 periods, i.e. until 2030. These estimates of the future values will be confronted with the estimates that are published by the Czech Statistical Office as predicted values of life expectancy at birth in low, medium and high variant. It will be shown that the low variant of CZSO population projection best corresponds with the estimates by ARIMA model for both males and females. The estimates of random walk model with drift are for both males and females situ-

[1] University of Economics Prague, Faculty of Informatics and Statistics, Department of Demography, W. Churchill sq. 4, 130 67 Prague 3, ondrej.simpach@vse.cz.
[2] University of Economics Prague, Faculty of Informatics and Statistics, Department of Demography, W. Churchill sq. 4, 130 67 Prague 3, petra.dotlacilova@vse.cz.
[3] University of Economics Prague, Faculty of Informatics and Statistics, Department of Demography, W. Churchill sq. 4, 130 67 Prague 3, langhamj@vse.cz.

ated in the middle of the estimates of low and medium variant of the CZSO population projection. For the analysis it will be first used software DeRaS for levelling mortality rates (Burcin et al. [8]) and consequently the Statgraphics Centurion XVI to calculate the optimized models of random walk with drift and ARIMA models. For calculate the empirical values of life expectancy at birth it will be used a classical approach of calculation of the life tables.

2 Methodology

For the description of mortality development is most often used an indicator known as life expectancy. Nowadays, the development of mortality for the oldest persons (and thus the development of life expectancy) enters to the forefront of research interest (Booth [4]). The extension of life expectancy is still in progress. It is important to note that for the examining of mortality at the highest ages we cannot use the empirical data. It is caused by the different development of mortality trend in the highest age groups compared to younger ones. For the own calculation of development of life expectancy is necessary to smooth the specific mortality rates by some of available models. In the past, the most-used was the Gompertz–Makeham (G–M) model (see e.g. Lagerås [16] or Ekonomov and Yarigin [10]), which can be expressed as

$$\mu_x = a + bc^x \tag{1}$$

where μ_x is the intensity of mortality at age x, x is the age and a, b, and c are parameters. This model is suitable for levelling the specific mortality rates from 60 to 85 years. Due to increased life expectancy in the most populations indicates that the G–M is not the most appropriate model. Nowadays more and more preferred models for levelling mortality rates are logistic models (Gavrilov and Gavrilova [11]). But it is important to know that logistic models are very optimistic (i.e., they provide higher values of life expectancy than other models). This difference is apparent mainly from 60 years. In this paper, we introduce two another models from Thatcher and Kannistö. Due to the recommendations by Boleslawski and Tabeau [3], the Thatcher model we can express as

$$\mu_x = \frac{z}{1+z} + \gamma \tag{2}$$

where $z = \alpha e^{\beta x}$, x is the age and α, β and γ are the parameters of model. Kannistö model we can express as

$$\mu_x = \frac{e^{[\theta_0 + \theta_1 .(x-80)]}}{1+e^{[\theta_0 + \theta_1 .(x-80)]}} \quad \text{for} \quad x \geq 80, \tag{3}$$

where μ_x is the intensity of mortality at age x, θ_0 and θ_1 are the parameters of model, which assumes values > 0. Research papers shows that the Kannistö model is the best suitable for description of mortality at the end of human life (Gavrilov and Gavrilova [11]). Therefore it is suitable especially for the highest ages (i.e. for persons older than 80 years). Life expectancy is obtained as an output indicator from mortality tables. The calculation is carried out in several steps. First, we calculate the specific mortality rates as

$$m_x = \frac{M_x}{\overline{S}_x}, \tag{4}$$

where M_x is the number of deaths at the exact age x and $\overline{S}_x$ is the middle number of living. Between the specific mortality rate and the intensity of mortality is valid the followed relationship

$$m_x \approx \mu(x + \frac{1}{2}). \tag{5}$$

Next, we will calculate the probability of death as

$$q_0 = \frac{M_0}{\alpha N_t^v + (1-\alpha) N_{t-1}^v} \quad \text{for} \quad x = 0, \tag{6}$$

where M_0 is the number of deaths at the age 0, α is the proportion of lower elementary file of deceased and N_t^v, respectively N_{t-1}^v is the number of live births in year t, respectively in year t-1. The probability of death

$$q_x = 1 - p_x \tag{7}$$

is valid for $x > 0$. The calculation of the probability of survival is given by

$$p_0 = 1 - q_0 \quad \text{for} \quad x = 0 \tag{8}$$

and by

$$p_x = e^{-m_x} \quad \text{for} \quad x > 0. \tag{9}$$

Next part of the calculation relates to tabular (i.e. imaginary) population. First, we select the initial number of live births in tabular population: l_0 = 100 000. Based on knowledge of the probability of survival, we are able to calculate the number of survivors in the further exact ages by

$$l_{x+1} = l_x p_x, \tag{10}$$

where l_x is the number of survivors at the exact age x from the default file of 100 000 live births of tabular population. The number of deaths of tabular population is given by

$$d_x = l_x q_x. \tag{11}$$

Next we calculate the number of lived years (L_x), respectively the number of years of remaining life (T_x) as

$$L_0 = l_0 - 0.85 d_0 \quad \text{for} \quad x = 0 \tag{12}$$

(where 0.85 = α is the proportion of lower elementary file of deceased) and

$$L_x = \frac{l_x + l_{x+1}}{2} \quad \text{for} \quad x > 0, \tag{13}$$

respectively

$$T_x = T_{x+1} + L_x. \tag{14}$$

Finally we obtain the life expectancy as

$$e_x = \frac{T_x}{l_x}. \tag{15}$$

In the next part of the calculation it will be used the ARIMA approach by authors Box and Jenkins for modelling of time series, which will be supplemented by the model of random walk with drift. Drift will be optimized using computational system Statgraphics Centurion XVI (version 16.1.11). ARIMA time series modelling is based on analysis of past trend of the time series and does not use an additional information that is in other circumstances required to extrapolate the life expectancy at birth.

3 Results

Based on the data published by the CZSO and using the conventional calculation of life tables there were calculated values of life expectancy at birth for males and females in the Czech Republic from 1970 to 2011. These values have been levelled in DeRaS software using Thatcher, G–M and Kanistö model. Levelled values for males are graphically shown in Fig. 1, levelled values for females are shown in Fig. 3. On the basis of the methodological approach of the authors Box and Jenkins [5] we identified the model ARIMA (1,0,0) without constant for the times series "life expectancy at birth – males" and further the random walk model with drift (see for instance Hughes [13]), where the drift was optimized at the value 0.209533. The estimates of the parameters of the ARIMA (1,0,0) model without constant are given in Tab. 1. For the time series "life expectancy at birth – females" we identified the model ARIMA (0,2,1) without constant (see Tab. 2) and further the random walk model with drift, where the drift was optimized at the value 0.189104. The diagnostic tests of the models indicate, that the non-systematic component of the model is not auto-correlated, is homoscedastic and has normal distribution at the 5% significance level (see for instance Breusch and Godfrey [6], Darnell [9] and Jarque and Bera [14]).

Parameter	Estimate	Stnd. Error	t	P-value
AR(1)	1.00255	0.000612525	1636.75	0.0000

Table 1 ARIMA (1,0,0) model without constant for life expectancy of males. Source: authors' calculations

Parameter	Estimate	Stnd. Error	t	P-value
MA(1)	1.02508	0.00587664	174.432	0.0000

Table 2 ARIMA (0,2,1) model without constant for life expectancy of females. Source: authors' calculations

Using ARIMA models and random walk with drift there were estimated ex-post the values of life expectancy at birth for males in years 1970–2011 (which are graphically illustrated in Fig. 2) and the values of life expectancy at birth for females in years 1970–2011 (which are graphically illustrated in Fig. 4). These levelled values are highly correlated with the values that were calculated by Thatcher, Gompertz–Makeham and Kanistö model. From CZSO population projections from 2009 were acquired the estimates of life expectancy at birth for males and females in low, medium and high variant. Based on ARIMA models without constant and random walk models with drift for males and females there were constructed predictions of life expectancy for the period 2012–2030. These values were graphically compared in Fig. 5 from which it appears that the values predicted by ARIMA models without constant are highly correlated with the values published in low variant of CZSO population projection.

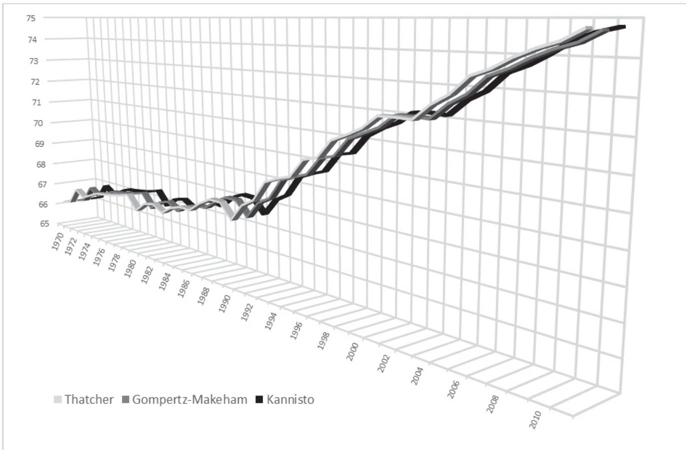

Figure 1 Estimated life expectancy at birth – males from 1970 to 2011 by Thatcher, Gompertz–Makeham and Kanistö model. Source: authors' calculations

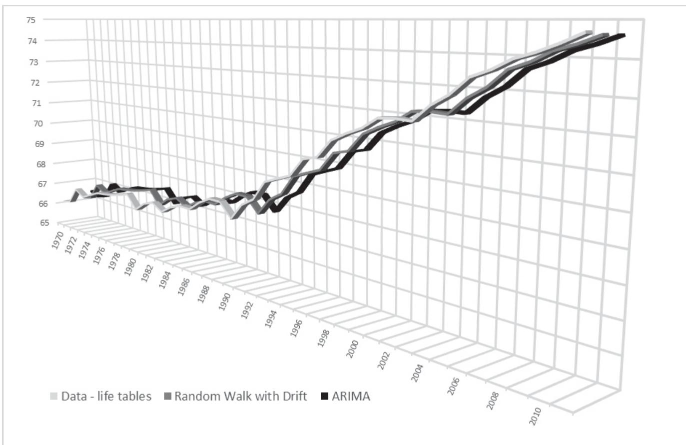

Figure 2 Life expectancy at birth from life tables and estimated life expectancy at birth – males from 1970 to 2011 by random walk model with drift and ARIMA model. Source: CZSO, authors' calculations

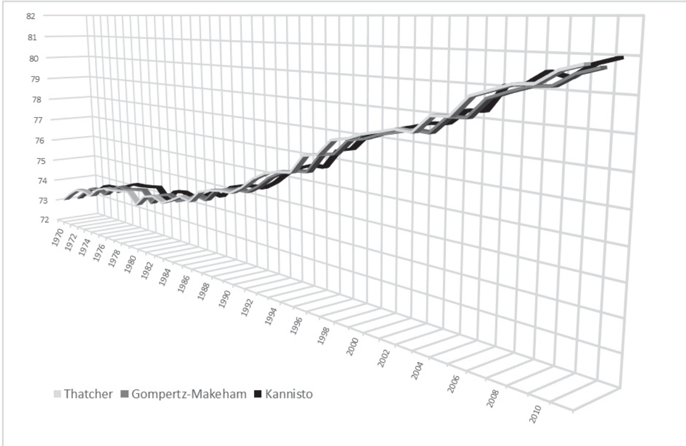

Figure 3 Estimated life expectancy at birth – females from 1970 to 2011 by Thatcher, Gompertz–Makeham and Kanistö model. Source: authors' calculations

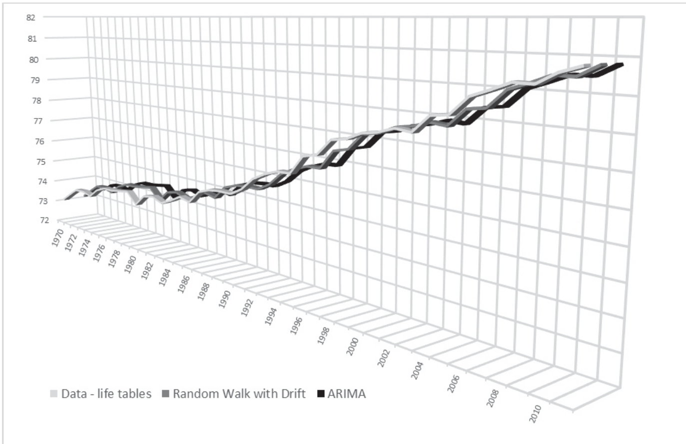

Figure 4 Life expectancy at birth from life tables and estimated life expectancy at birth – females from 1970 to 2011 by random walk model with drift and ARIMA model. Source: CZSO, authors' calculations

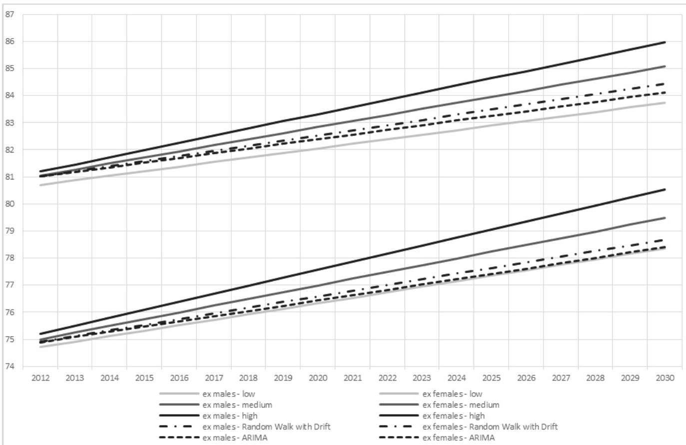

Figure 5 Expected life expectancy at birth from CZSO population projection for males (bottom) and females (top) in low, medium and high variant and an extrapolation of life expectancy at birth by random walk model with drift and ARIMA model for males (bottom) and females (top). Source: CZSO, authors' calculations

The parameters of the Kanistö and Thatcher model were estimated using the nonlinear regression, so their suitability was assessed on the basis of the corrected coefficient of the multiple determination. In the case of Kanistö model, the values were estimated slightly higher than in the case of Thatcher model.

4 Conclusion

The aim of the presented paper was the calculation of life expectancy at birth for males and females in the Czech Republic from 1970 to 2011 using selected levelling functions (Gompertz–Makeham, Kanistö and Thatcher) and using ARIMA model and Random Walk with Drift. It was shown that all approaches provide similar results, but ARIMA models and the random walk models with drift require less computation. Moreover, based on them it is possible to construct the predictions for the future. Predictions by ARIMA and random walk models do not require expensive input data and provide comparable results with low variant by CZSO, which calculation is incomparably more difficult. This provides time and money savings. Therefore, the challenge for future research is to use this approach for constructing cheaper predictions.

Acknowledgements

This paper was supported by project VŠE IGA 24/2013 “The mortality and aging population of the Czech Republic”.

References

[1] Arltová, M., Langhamrová, Jitka, Langhamrová, Jana. Development of life expectancy in the Czech Republic in years 1920-2010 with an outlook to 2050. *Prague Economic Papers*, 2013, roč. 22, č. 1, s. 125–143.

[2] Bérenger, V., Verdier-Chouchane, A.: Multidimensional Measures of Well-Being: Standard of Living and Quality of Life Across Countries. *World Development* **35** (2007), 1259–1276.

[3] Boleslawski, L., Tabeau, E.: Comparing Theoretical Age Patterns of Mortality Beyond the Age of 80. *In: TABEAU, E., VAN DEN BERG JETHS, A. and HEATHCOTE, CH. (eds.) 2001. Forecasting Mortality in Developed Countries: Insights from a Statistical, Demographic and Epidemiological Perspective*, 2001, p. 127 – 155.

[4] Booth, H.: Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting* **22** (2006), 547–581.

[5] Box, G.E.P., Jenkins, G.: *Time series analysis: Forecasting and control*, San Francisco, Holden-Day, 1970.

[6] Breusch, T., Godfrey, L.G.: Data Transformation Tests, *Economic Journal* **96** (1986), 47-58.

[7] Burcin, B., Tesárková, K. a Šídlo, L.: Nejpoužívanější metody vyrovnávání a extrapolace křivky úmrtnosti a jejich aplikace na českou populaci. Demografie **52**, 2010, 77 – 89.

[8] Burcin, B., Hulíková Tesárková, K., Kománek, D.: *DeRaS: software tool for modelling mortality intensities and life table construction*. Charles University in Prague, 2012, Prague. <http://deras.natur.cuni.cz>

[9] Darnell, A.C.: *A Dictionary of Econometrics*, Edward Elgar, 1994.

[10] Ekonomov, AL., Yarigin, VN.: The age dynamics of mortality and the Gompertz-Makeham law, *Zhurnal Obshchei Biologii*, Volume 50, Issue 2, MAR-APR 1989, p. 236-243.

[11] Gavrilov, L.A., Gavrilova, N.S.: Mortality measurement at advanced ages: a study of social security administration death master file. *North American actuarial journal* **15** (3): 432 – 447.

[12] Gompertz, B. On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of the Royal Society of London* **115** (1825): 513–585.

[13] Hughes, B.D.: *Random walks and random environments*, Oxford University Press, 1996.

[14] Jarque, C., Bera, A.: Efficient Tests for Normality, Heteroscedasticity, and Serial Independence of Regression Residuals, *Economics Letters* **6**, (1980) 255 – 259.

[15] Jia, H., Zack, M., Thompson, W., Dube, S.H.: Quality-adjusted life expectancy (QALE) loss due to smoking in the United States, *Quality of Life Research*, February 2013, Volume 22, Issue 1, pp 27-35.

[16] Lagerås, A.N.: Commutation functions under Gompertz–Makeham mortality, *Scandinavian Actuarial Journal*, Volume 2010, Issue 2, 2010, p. 161-164.

[17] Makeham, W.M. On the Law of Mortality and the Construction of Annuity Tables. *The Assurance Magazine, and Journal of the Institute of Actuaries* **8** (1860): 301–310.

[18] Šimpach, O., Langhamrová, J.: Czech Household Computer Facilities as a Reliable Variable in a Life Expectancy Forecast Model up to the Year 2060. Jindřichův Hradec 12.09.2012 – 14.09.2012. *In: IDIMT-2012.* Linz : Trauner, 2012, s. 143–152.

[19] Thatcher, R.A., Kanistö, V. and Vaupel, J.W.: *The Force of Mortality at Ages 80 to 120*. Odense University Press, 1998.

[20] Torri, T.: Coherent mortality forecasting via the aggregate measure of life expectancy. Marrakech 27.09.2009 – 02.10.2009. *In: XXVI International Population Conference. [online]* Paris : IUSSP, 2009, p. 1–20. URL: <http://iussp2009.princeton.edu/papers/92339>