The Use of Polynomial Functions for Modelling of Mortality at the Advanced Ages

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

Abstract. Mortality at the advanced ages becomes more and more important topic for demographers. The extensions of the length of life is influenced mainly by the improvement in medicine. At first, there was a significant improvement in the care of infants, which caused the decrease of infant mortality. Later, it began to improve the mortality even at the advanced ages. One reason for this evolution is just a higher level of health care. Another equally important reason may be more interest in a healthy lifestyle and also a better environment. Given to this evolution, it is more and more important to have the best imagination about how long in average will live not only the youngest persons, but the oldest ones too. From the obtained results it is evident that the level of mortality of younger persons is different in comparison with the oldest ones. Therefore, it is necessary to correct the estimates of mortality at the highest ages. For this correction are used the various types of models. Given to the importance of the most accurate capture of mortality at the advanced ages are still required new models and methods that would provide the best imagination of current trends. The aim of this paper is to introduce the possibility of using of the polynomial functions for levelling and for extrapolation of mortality curves at the advanced ages. The results will be compared with the methodology, which is used by the Czech Statistical Office and with general mortality tables.

Keywords: mortality, life expectancy, polynomial functions and order, life tables

JEL Classification: C61, C63
AMS Classification: 62H12

1 Introduction

The mortality significantly affects the length of human life. Recently there occur the improvement in mortality in all age groups, but it is important to note that during the human life there become the changes in the character of mortality (Arltová, Langhamrová JI, Langhamrová JA [1] or Fiala [6]). The most different mortality trend is especially at the oldest ones (persons aged 60 and more years, see e.g. Šimpach, Dotlačilová, Langhamrová JI [14]). It is also important to realize that the data about the numbers of deaths M_x and the numbers of living (mid-year population) $\overline{S}_x$ are not exactly reliable at the highest ages (Boleslawski, Tabeau [2]) (and they can be influenced by both random and systematic error). Therefore it is important to modify mortality in some way especially at the highest ages (Gavrilov, Gavrilova [7]). There are various methods that can be used for smoothing mortality curves (e.g. model by Kannistö, Thatcher or Coale-Kisker. See e.g. Thatcher, Kanistö and Vaupel [15] or Burcin, Tesárková, Šídlo [3]). Another option is the application of polynomial functions.

We present in this paper the possibility of use of polynomial functions of different orders for the modelling of mortality curves (see e.g. Burcin, Tesárková, Šídlo [3]), and our results will be compared with the methodology of the Czech Statistical Office (see e.g. study by Koschin [9] or Šimpach [13]) and also with the general results of mortality tables. For smoothing and extrapolation of mortality curves we use a polynomial function of 2nd, 3rd and 4th order, depending on the appropriateness and significance. We use the data about the numbers of deaths $M_{x,t}$, the numbers of living $\overline{S}_{x,t}$ from the CZSO database from 1920 to 2010 for males and females in the Czech Republic, while the calculations are carried out for the 10-year time instants t = 1920, 1930, ..., 2010. Empirical, balanced and extrapolated values are presented for selected years in the synoptic charts separately for males and females. To complement our outputs there will be supplemented the values calculated according to the CZSO's methodology, which will be presented in the part called Materials and Methods. With the obtained results we lead the

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

discussion about the suitability of a specific approach together with the recommendations applicable to the Czech population.

2 Materials and Methods

Let us denote the polynomial function of 2nd, 3rd and 4th order respectively as

$$m_{x,t} = \begin{cases} \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t \\ \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \beta_3 x_t^3 + \varepsilon_t \\ \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \beta_3 x_t^3 + \beta_4 x_t^4 + \varepsilon_t \end{cases}, \quad (1)$$

where $m_{x,t}$ are the age-specific mortality rates calculated by ratio

$$m_{x,t} = \frac{M_{x,t}}{\overline{S}_{x,t}}, \quad (2)$$

$\beta_0, \beta_1, \ldots, \beta_4$ are the parameters of models and ε_t is the error term with the characteristics of white noise, x is age where x = <0 ; 110> and t is time where t = <1920 ; 2010> in 10-year time instants. These polynomial functions we use for smoothing of the age-specific mortality rates from 60 to 85 years. For the advanced ages we perform the extrapolation. The results will be subsequently compared with the mortality tables of CZSO. Their calculation can be summarized in a few steps. First, there is performed the smoothing of empirical age-specific mortality rates, followed by the calculation of mortality tables. In lower ages we smooth the age-specific mortality rates by moving averages with the length of 3 (e.g. Šimpach [13] or Šimpach, Dotlačilová, Langhamrová [14]) as

$$\widetilde{m}^{(3)} = \frac{m_{x-1} + m_x + m_{x+1}}{3}, \text{ where } x \in <3 ; 5>, \quad (3)$$

with the length of 9 as

$$\begin{aligned} \widetilde{m}^{(9)} &= 0{,}2m_x + 0{,}16(m_{x-1} + m_{x+1}) + 0{,}12(m_{x-2} + m_{x+2}) \\ &+ 0{,}08(m_{x-3} + m_{x+3}) + 0{,}04(m_{x-4} + m_{x+4}) \end{aligned}, \quad (4)$$

where $x \in$ <6 ; 29> and with the length of 19 as

$$\begin{aligned} \widetilde{m}^{(19)} &= 0{,}2m_x + 0{,}1824(m_{x-1} + m_{x+1}) + 0{,}1392(m_{x-2} + m_{x+2}) \\ &+ 0{,}0848(m_{x-3} + m_{x+3}) + 0{,}0336(m_{x-4} + m_{x+4}) - 0{,}0128(m_{x-6} + m_{x+6}), \\ &- 0{,}0144(m_{x-7} + m_{x+7}) - 0{,}0096(m_{x-8} + m_{x+8}) - 0{,}0032(m_{x-9} + m_{x+9}) \end{aligned} \quad (5)$$

where $x \in$ <30 ; 59>. For age 1 and 2 are used the empirical values of age-specific mortality rates. At the advanced ages is used the Gompertz–Makeham function (G–M) (Gompertz [8], Makeham [11]) or the modified Gompertz–Makeham function (mG–M) for smoothing of the age-specific mortality rates. Let us explain the G–M function (see e.g. Boleslawski, Tabeau [2]) as

$$\mu_x = a + bc^x, \quad (6)$$

where μ_x is the intensity of mortality, x is age and a, b, and c are parameters. The function is based on the assumptions that the increments of the intensity of mortality are constant with increasing age. Given that this assumption is not possible to apply from 85 years higher, therefore it was designed mG–M function. It assumes that the increments of the intensity of mortality decrease with increasing age. For the intensity of mortality at lower ages is true the equation:

$$\mu\left(x + \frac{1}{2}\right) \doteq m_x, \text{ where } x = 1, 2, \ldots, 59. \quad (7)$$

At first there will be used the Solver in MS Office Excel (version 2010 or higher, for more information see e.g. Fiala [5] Šimpach [12]). We have to find the initial estimates of parameters in the first step and we will improve them by using the ordinary least squares method (OLS). Because we need to estimate three parameters, we will need three equations. It is important to choose the beginning of the first interval x_0 (i.e. the age from which will be performed the equalization) and set the length of the interval k. After that, we will calculate the sum of empirical mortality rates in the three intervals

$$G_1 = \sum_{x=x_0}^{x_0+k-1} m_x \,,\; G_2 = \sum_{x=x_0+k}^{x_0+2k-1} m_x \text{ and } G_3 = \sum_{x=x_0+2k}^{x_0+3k-1} m_x \,, \tag{8}$$

where $k = 5$ and $x_0 = 60$. The value of G_1 we can also express using by model parameters and age x (e.g. Fiala [5] or Šimpach [13]) as

$$G_1 = \sum_{x=x_0}^{x_0+k-1} \left(a + b.c^{x+\frac{1}{2}} \right). \tag{9}$$

In the same way, we could expressed G_2 and G_3. By subtracting and dividing of the individual equations we exclude the most of the parameters and we get

$$c^k = \frac{G_3 - G_2}{G_2 - G_1}\,. \tag{10}$$

The value of the parameter c we get as k-(th) square root of the c^k

$$c = \sqrt[k]{c^k}\,. \tag{11}$$

To estimate the initial values of a and b we have to calculate the value of sub-expression of K_c as

$$K_c = c^{x_0 + \frac{1}{2} \cdot \frac{c^k - 1}{c - 1}}\,, \tag{12}$$

and the parameters b and a we obtain as

$$b = \frac{G_2 - G_1}{K_c.\left(c^k - 1\right)},\quad a = \frac{G_1 - b.K_c}{k} \quad \text{respectively.} \tag{13}$$

The last step is the calculation of the weighted squared deviations (*WSD*), through which we will optimize the parameters of G–M (respectively mG–M) model

$$WSD = \frac{S_{t,x} + S_{t+1,x}}{2.m_x.(1 - m_x)} \cdot \left(m_x - \tilde{m}_x^{(GM)} \right)^2 \quad \text{for} \quad x \in <60\,; y>, \tag{14}$$

where $S_{t,x}$ is the number of living at age x in year t, $S_{t+1,x}$ is the number of living at age x in year t+1, m_x are age-specific mortality rates, $m_x^{\sim(GM)}$ are smoothed values of age-specific mortality rates obtained by using G–M or mG–M model, y is the highest age for which we have a non-zero value of m_x. When we optimize the parameters using by the OLS, it is necessary to create two instrumental sums

$$S_1 = \sum_{x=60}^{82} \frac{S_{t,x} + S_{t+1,x}}{2.m_x.(1 - m_x)} \cdot \left(m_x - \tilde{m}_x^{(GM)} \right)^2, \tag{15}$$

respectively

$$S_2 = \sum_{x=83}^{y} \frac{S_{t,x} + S_{t+1,x}}{2.m_x.(1 - m_x)} \cdot \left(m_x - \tilde{m}_x^{(GM)} \right)^2, \tag{16}$$

which will help us to optimize the initial values of parameters. For continuation of analysis, let us explain mG–M function (see e.g. Lagerås [10] or Thatcher et al. [15]) as

$$\mu_x = a + bc^{x_0 + \frac{1}{\gamma}\ln[\gamma(x - x_0) + 1]}\,, \tag{17}$$

where a, b and c are the parameters of the original G–M function, x_0 is the age from which will be performed equalization and γ is a parameter of mG–M function.

3 Results and Discussion

Using the methodology of polynomial functions of various orders we smooth the mortality curves of the Czech males and females at the highest ages. This smoothing is applied to the data matrix of age-and-sex specific mortality rates in 1920, 1930, ..., 2010. Then we compare the obtained results from these functions with the results obtained by the CZSO's methodology. The aim of this paper is primarily to find out how it changes the suitability of various functions and the approaches for specific gender and time period. In the Fig. 1 is shown a matrix of charts with empirical and smoothed values of age-specific mortality rates of males in the Czech Republic. The charts below show only functions, which were based on the statistically significant parameter estimates (see Tab. 1 for males). It is clear that from the obtained results are statistically significant the polynomial functions of the 2nd and 3rd order in all analysed periods. On the other hand the polynomial function of the 4th order is significant for two years only (1960 and 1970). If we compare the various ways of smoothing, we conclude that a polynomial function is suitable for smoothing of mortality curves approximately to 80 years. On the contrary for the higher ages is better suited the polynomial function of the 3rd order. If we examine the suitability of these functions over time, we find that at the beginning of the analysed period was slightly better the polynomial function of the 2nd order, but towards presence is more suitable for our requirements the polynomial function of the 3rd order. If we compare our smoothing by polynomial functions with the results which we calculated by the Czech Statistical Office methodology, we conclude that there has been an undervaluation of actual level of mortality in certain years. Better smoothing provides the polynomial function of the 3rd order (especially at the highest ages).

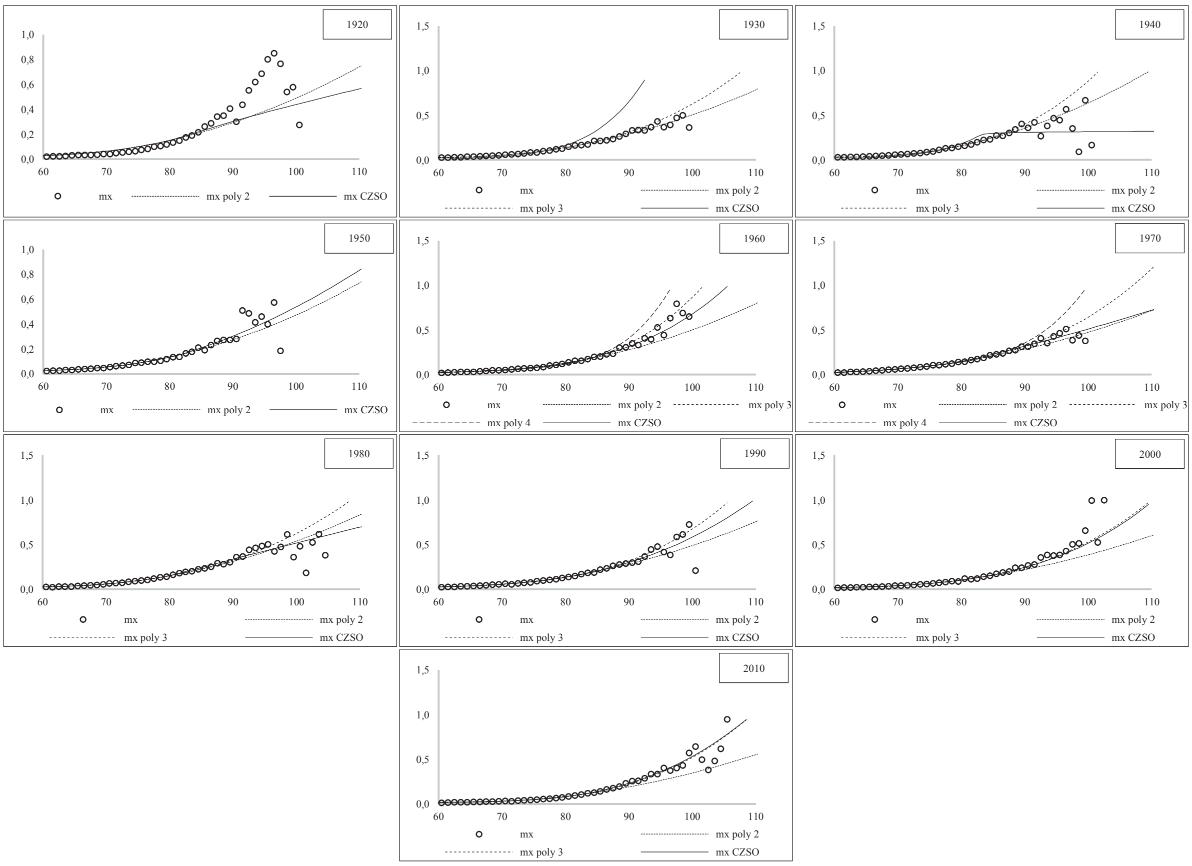

Figure 1 Age-specific mortality rates of Czech males from 1920 to 2010. Source: CZSO, own calculation

In the Fig. 2 is shown a matrix of charts with empirical and smoothed values (for parameters see Tab. 2) of age-specific mortality rates of females. We examine the various types of smoothing of these values and we come to a similar conclusion as in the case of males. The polynomial function of the 4th order is statistically significant for two years only (2000 and 2010). If we compare the smoothed values calculated by polynomial functions of 2nd and 3rd order, we find that these functions provide good estimates approximately to 85 years. From our results it is also evident that the use of polynomial functions of the 2nd and 3rd order is particularly suitable at the beginning of analysed period (approximately until 1950). Since 1960 seems to be the polynomial function of the 3rd order and CZSO's methodology as optimal smoothing approach. For smoothing of values at the advanced ages (over 85) it is more suitable the polynomial function of the 3rd order. When we compare our results calculated by the CZSO's approach, we get the simple conclusion: closer to present, the more are the values of mortality overestimated (especially the oldest ages). Then it is better to use the universal approach of polynomial function of the 3rd order.

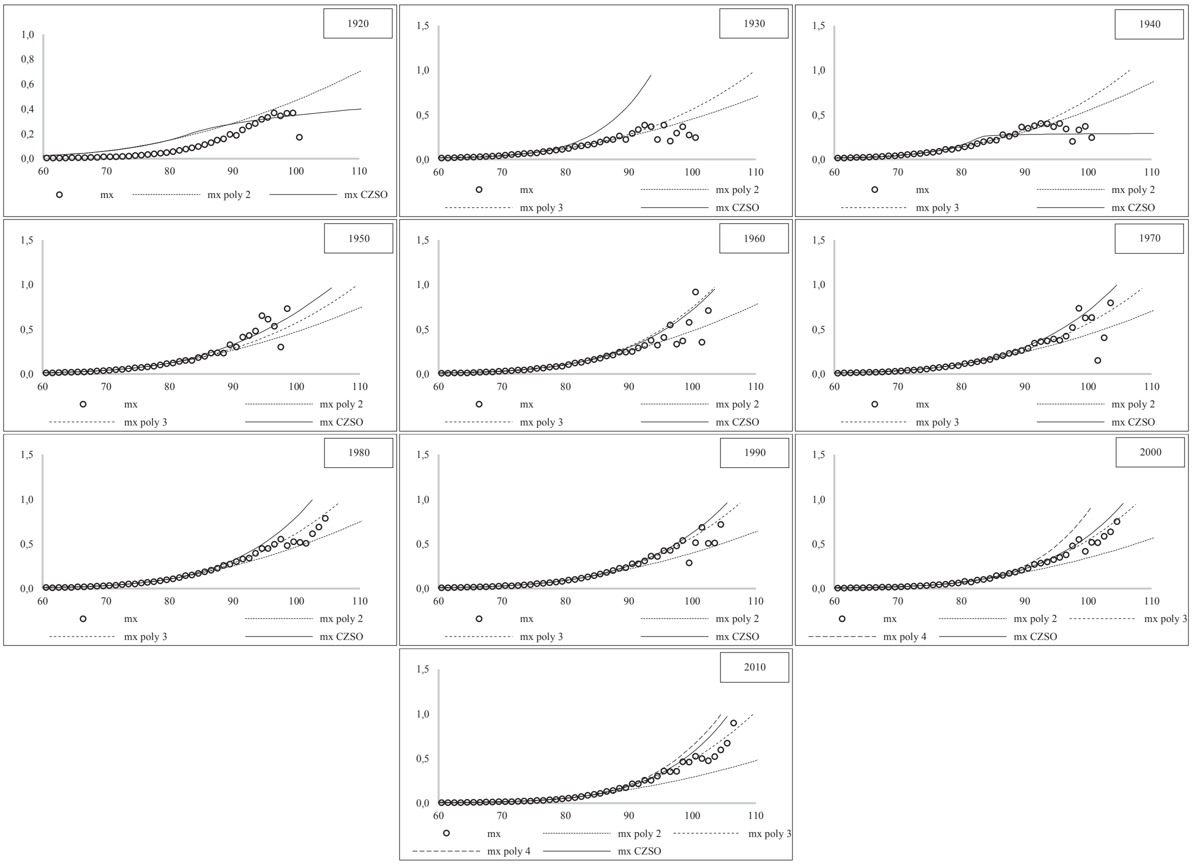

Figure 2 Age-specific mortality rates of Czech females from 1920 to 2010. Source: CZSO, own calculation

	1920	1930		1940		1950	1960			1970		
par.	2nd	2nd	3rd	2nd	3rd	2nd	2nd	3rd	4th	2nd	3rd	4th
β_0	9,604E-1	3,237E-4	-1,431E+0	1,774E+0	1,454E-5	1,116E+0	1,430E+0	2,155E-5	1,378E-6	9,140E-1	-2,825E+0	1,991E+1
β_1	-3,183E-2	-3,966E-2	7,253E-2	-5,556E-2	1,722E-1	-3,577E-2	-4,447E-2	-4,333E-3	-3,780E-4	-3,041E-2	1,263E-1	-1,146E+0
β_2	2,724E-4	1,246E+0	-1,234E-3	4,435E-4	-2,719E-3	2,944E-4	3,530E-4	2,931E-1	3,892E-2	2,612E-4	-1,915E-3	2,465E-2
β_3	x	x	7,161E-6	x	-3,660E+0	x	x	-6,623E+0	-1,778E+0	x	1,000E-5	-2,354E-4
β_4	x	x	x	x	x	x	x	x	3,040E+1	x	x	8,461E-7

	1980		1990		2000		2010	
	2nd	3rd	2nd	3rd	2nd	3rd	2nd	3rd
…	1,237E+0	-6,491E-1	1,178E+0	-2,896E+0	9,667E-1	-2,198E+0	1,019E+0	-2,837E+0
…	-4,002E-2	3,904E-2	-3,757E-2	1,332E-1	-3,073E-2	1,020E-1	-3,159E-2	1,301E-1
…	3,318E-4	-7,656E-4	3,080E-4	-2,063E-3	2,501E-4	-1,592E-3	2,496E-4	-1,994E-3
…	x	5,046E-6	x	1,090E-5	x	8,469E-6	x	1,032E-5
…	x	x	x	x	x	x	x	x

Table 1 Estimated parameters of significant polynomial functions for males. Source: author’s calculations

	1920	1930		1940		1950		1960		1970	
par.	2nd	2nd	3rd	2nd	3rd	2nd	3rd	2nd	3rd	2nd	3rd
β_0	8,000E-1	1,059E+0	-1,341E+0	1,392E+0	-1,383E+0	1,181E+0	-1,012E+0	1,469E+0	-4,366E+0	1,176E+0	-1,529E+0
β_1	-2,744E-2	-3,421E-2	6,641E-2	-4,458E-2	7,176E-2	-3,801E-2	5,393E-2	-4,577E-2	1,988E-1	-3,750E-2	7,590E-2
β_2	2,423E-4	2,825E-4	-1,114E-3	3,628E-4	-1,252E-3	3,105E-4	-9,657E-4	3,603E-4	-3,035E-3	3,028E-4	-1,271E-3
β_3	x	x	6,422E-6	x	7,425E-6	x	5,868E-6	x	1,561E-5	x	7,238E-6
β_4	x	x	x	x	x	x	x	x	x	x	x

	1980		1990		2000			2010		
	2nd	3rd	2nd	3rd	2nd	3rd	4th	2nd	3rd	4th
…	1,339E+0	-2,046E+0	1,141E+0	-2,672E+0	1,093E+0	-3,306E+0	1,676E+1	9,964E-1	-3,389E+0	6,021E+0
…	-4,216E-2	9,977E-2	-3,589E-2	1,240E-1	-3,389E-2	1,505E-1	-9,718E-1	-3,050E-2	1,533E-1	-3,731E-1
…	3,356E-4	-1,635E-3	2,857E-4	-1,934E-3	2,649E-4	-2,295E-3	2,114E-2	2,351E-4	-2,317E-3	8,675E-3
…	x	9,059E-6	x	1,020E-5	x	1,177E-5	-2,047E-4	x	1,173E-5	-8,981E-5
…	x	x	x	x	x	x	7,465E-7	x	x	3,501E-7

Table 2 Estimated parameters of significant polynomial functions for females. Source: author’s calculations

4 Conclusion

From the obtained results it is clear that during the analysed period occurred the changes in the suitability of the various types of functions. This is mainly due to changes in the level of mortality (which is the most relevant at the highest ages). All approaches of smoothing of age-specific mortality rates seem to be problematic at the beginning of the analysed period - mainly due to low reliability of empirical data. If we compare the results calculated by the CZSO's methodology with results of polynomial functions, we conclude that especially at the present the CZSO's methodology provides the estimates which significantly overestimate the actual mortality. On the contrary seems to be more suitable the polynomial function of the 3rd order for the majority of the analysed years. Our results also indicate that its suitability for smoothing of the age-specific mortality rates is approximately from 80 years higher. If we look at the smoothed values which we obtained by applying the polynomial function of the 2nd order, we find that its application should be up to 80 years, no more. The polynomial function of the 2nd order is not suitable for smoothing at the higher ages, because it underestimates the actual values of mortality. Based on the calculated results we can also conclude that neither one of the polynomial functions is universally applicable for smoothing in whole age range and to the subsequent extrapolation of mortality curves to the highest ages. On the contrary it might be appropriate to use a combination of both types of functions for smoothing of mortality curves. Up to the age of 80 years we could applied the polynomial function of the 2nd order and at the higher ages the polynomial function of the 3rd order. This idea is also a challenge for our future research.

Acknowledgements

This paper was supported by the project of Internal Grant Agency of VŠE Praha under the No. IGS F4/24/2013.

References

[1] Arltová, M., Langhamrová, JI, Langhamrová, JA. Development of life expectancy in the Czech Republic in years 1920-2010 with an outlook to 2050. *Prague Economic Papers*, 2013, Vol. 22, no. 1, pp. 125–143.

[2] 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.). Forecasting Mortality in Developed Countries: Insights from a Statistical, Demographic and Epidemiological Perspective*, 2001, pp. 127 – 155.

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

[4] Ekonomov, AL., Yarigin, VN. The age dynamics of mortality and the Gompertz-Makeham law, *Zhurnal Obshchei Biologii*, Vol. 50, Issue 2, MAR-APR 1989, pp. 236–243.

[5] Fiala, T. Mortality in the Age from 20 to 70 Years in the Czech Republic in the Period 1950 to 1996. *Acta Oeconomica Pragensia*, 1999, Vol. 7, no. 4, pp. 35–44.

[6] Fiala, T. Reducing the Consequences of Population Ageing for the Pension System by Raising the Retirement Age in the Czech Republic. In: *The 7th International Days of Statistics and Economics*. Slaný: MELANDRIUM, 2013, pp. 398–407.

[7] Gavrilov, L.A., Gavrilova, N.S. Mortality measurement at advanced ages: a study of social security administration death master file. *North American actuarial journal*, vol. 15, no. 3, pp. 432 – 447.

[8] 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), pp. 513–585.

[9] Koschin, F. Jak vysoká je intenzita úmrtnosti na konci lidského života? *Demografie,* 1999, Vol. 41, no. 2, pp. 105–109.

[10] Lagerås, A.N. Commutation functions under Gompertz–Makeham mortality, *Scandinavian Actuarial Journal*, Vol. 2010, Issue 2, pp. 161–164.

[11] 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), pp. 301–310.

[12] Šimpach, O. Faster convergence for estimates of parameters of Gompertz-Makeham function using available methods in solver MS Excel 2010. In: *Proceedings of the 30th International Conference of Mathematical Methods in Economics*. Opava: Universita Opava, 2012, pp. 870–874.

[13] Šimpach, O. Effective use of MS Excel in education of economic demographers. In: *Efficiency and Responsibility in Education*. Praha: Czech University of Life Sciences in Prague, 2013, pp. 557–564.

[14] Šimpach, O., Dotlačilová, P., Langhamrová, JI. Logistic and ARIMA models in the Estimation of Life Expectancy in the Czech Republic. In: *Proceedings of the 31st International Conference of Mathematical Methods in Economics*. Jihlava: College of Polytechnics Jihlava, 2013, pp. 915–920.

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