Slovak University of Technology in Bratislava
Faculty of Mechanical Engineering
15th Conference on Applied Mathematics APLIMAT 2016
**Proceedings**

# SMOOTHING OF MORTALITY CURVE: THE USAGE OF MOVING AVERAGES AND GOMPERTZ-MAKEHAM FUNCTION IN THE CASE OF THE CZECH REPUBLIC

**DOTLAČILOVÁ Petra (CZ), ŠIMPACH Ondřej (CZ)**

**Abstract.** Our paper discusses the possibilities of levelling the age-specific death rates of the Czech population using different lengths of moving averages and the Gompertz-Makeham (G-M) function. Moving average is a technique that belongs between the simpler one, G-M function between the most used approaches for levelling of mortality curves at higher ages. These approaches are very useful in the conditions of the Czech Republic. The lengths of moving averages were used according to Czech literature and empirical studies and finally set to 3, 9 and 19 values with different weights. The best results of smoothing by moving averages is achieved with the use of moving averages of 19 values, on the other hand we lose most initial observations. The G-M functions were estimated in STATA software and results are different e.g. in comparison with 19 values length moving averages. Results from this study will be used for future research to implement this technique into extrapolation script that will be able to smooth the values and then extrapolate them to the highest ages (110 years or above).

**Keywords:** age-specific death rates, modelling of mortality, moving averages, Gompertz-Makeham function

*Mathematics Subject Classification:* Primary 90C30; Secondary 62H12.

## 1 Introduction

Ageing of population is very often discussed topic in last days. For our future it means, that the proportion of persons at higher ages will continue to increase. One cause of ageing population is decrease in mortality. But the empirical values of age-specific death rates show fluctuations. They are noticeable mainly at higher ages. That is the reason why these values are smoothed. At higher ages (60 years and higher), where mortality has more natural character, it is possible to use some of available functions (e.g. Gompertz-Makeham, Kanistö, Coale-Kisker, Thatcher and others). On the other hand moving averages could be used for smoothing of death rates at lower ages.

The aim of this paper is at first the application of moving averages of different length on data about mortality of the Czech population, separately for males and for females. In literature there are commonly used moving averages of length 3, 9 and 19, (but there are authors who use moving averages of another length). Obtained results for every type of moving average will be compared with the empirical values of mortality. Secondly, we estimate parameters of the Gompertz-Makeham function on the basis of the same data as in the case of moving averages and we extrapolate mortality curves up to the age 110 years. At the end we discuss mutual consequences between extreme values arising from empirical data matrix and smoothed values arising from moving average models and Gompertz-Makeham functions.

## 2 Theoretical background and literature review

Mortality is closely related to the health status of population and also it is connected with very often discussed topic of ageing population. Due to the gradual improvement in mortality of population we can observe the extending of human life (see e.g. study by Fiala [5] or Dotlačilová, Langhamrová, Šimpach [3]). At the same time the births rate declined, which results in an increase of the proportion of people living in higher ages and also in the population ageing (see e.g. study by Šimpach [12] or Šimpach, Pechrová [14]). The extension of length of life is influenced mainly by an improvement in medicine. At first, there was a significant improvement in a care of infants, which caused the decrease of infant mortality. Later, it began to improve mortality even at the advanced ages (see e.g. study by Thatcher, Kanistö, Vaupel [15]). One reason for this evolution is higher level of health care. Another equally important reason may be higher interest in a healthy lifestyle and also better environment. Given this population development there grow the interest in capturing of mortality at older ages as accurately as possible. But an important thing is that mortality at older ages is often influenced by systematic and random errors (Gavrilov, Gavrilova [7], Šimpach, Dotlačilová, Langhamrová [13]). To obtain the most accurate development should be these errors corrected by using suitable modelling approach of mortality. One option is mechanical levelling method (using moving averages), which allows to smooth mortality curve mainly at lower ages. Another possibility is the use of an analytic function (e.g. Gompertz-Makeham function; see paper by Gompertz [8], Makeham [9] and application e.g. in study by Šimpach [11]), which is especially useful at higher ages (from 60 years higher). In our paper we apply moving averages of different length (3, 9 and 19 according to literature review; Boleslawski, Tabeau [1], Dotlačilová, Langhamrová, Šimpach [3] or Gavin, Haberman, Verrall [6]) on data about mortality of the Czech population, separately for males and for females. Results for every type of smoothing we compare with the empirical values of Czech mortality and we will discuss about their suitability.

## 3 Materials and Methods

In demographic analysis we use the modelling of age-specific death rates particularly for the purpose of life insurance. Life tables describe the dependence of mortality on the age. The basic characteristics for their calculation are therefore age-specific death rates. In the case of life tables for one calendar year *t* are age-specific death rates calculated by the formula

$$m_{x,t} = \frac{M_{x,t}}{\bar{S}_{x,t}}, \tag{1}$$

where $M_{x,t}$ is the number of deaths at complete age $x$ and $\bar{S}_{x,t}$ is the number of persons at age $x$, which is estimated as the mid-year number of the population. Gompertz [8], later followed by

Makeham [9] and others, found that the age-specific death rates develops (at higher ages) according to an exponential law. These rates are presented in Figure 1 for males (left) and for females (right).

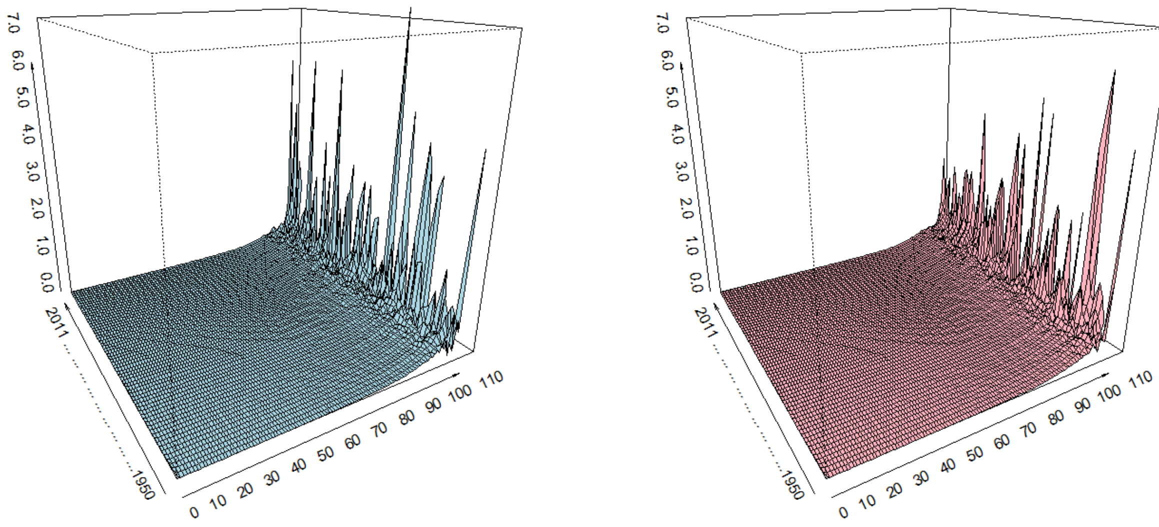


Fig. 1. Empirical values of $m_{x,t}$ in the Czech Republic for males (left) and females (right).
Source: Human mortality database (HMD), authors' construction.

In this paper we use three types of moving averages (with length of 3, 9 and 19). Moving average with the length of 3 values we compute as

$$\tilde{m}_x^{(3)} = \frac{m_{x-1} + m_x + m_{x+1}}{3} \quad \text{for} \quad x \in \langle 3; z-1 \rangle, \tag{2}$$

where $m_x$ is an empirical value of age-specific death rate and $z$ labels the age in which we still have correctly calculated value of age-specific death rate. Moving average with length of 3 values cannot be used for balancing the lowest ages (0, 1 and 2 years), because mortality in these ages has different character, caused by infant and neonatal mortality. The disadvantage of this type of smoothing is that it tends to overestimate death rates especially at older ages. Moving average with the length of 9 values we compute as

$$\begin{aligned}\tilde{m}_x^{(9)} &= 0{,}2 \times m_x + 0{,}16 \times (m_{x-1} + m_{x+1}) + 0{,}12 \times (m_{x-2} + m_{x+2}) \\ &\quad + 0{,}08 \times (m_{x-3} + m_{x+3}) + 0{,}04 \times (m_{x-4} + m_{x+4})\end{aligned} \qquad \text{for} \quad x \in \langle 6; z-4 \rangle. \tag{3}$$

In this type of moving average there are weights designed with basic idea that the value of smoothed death rate has assigned the highest weight and then the other weights gradually decrease. In the case of smoothing by 9 values we also see problem with overestimation of death rates at older ages. The advantage of this type of average is that it significantly removes random deviations at higher ages. On the other hand, the main disadvantage is that this type of average also removes systematic fluctuations in mortality which occurs mainly at young age. Finally, moving average with the length of 19 values we compute as

$$\tilde{m}_x^{(19)} = 0{,}2 \times m_x + 0{,}1824 \times (m_{x-1} + m_{x+1}) + 0{,}1392 \times (m_{x-2} + m_{x+2})$$
$$+ 0{,}0848 \times (m_{x-3} + m_{x+3}) + 0{,}0336 \times (m_{x-4} + m_{x+4}) + 0{,}0128 \times (m_{x-6} + m_{x+6})$$
$$- 0{,}0144 \times (m_{x-7} + m_{x+7}) - 0{,}0096 \times (m_{x-8} + m_{x+8}) - 0{,}0032 \times (m_{x-9} + m_{x+9})$$
$$\text{for} \quad x \in \langle 6; z-9 \rangle . \qquad (4)$$

Into this type of moving average were some weights included with a negative sign, which also allows to eliminate overestimation of actual level of mortality (see e.g. Fiala [5], or Dotlačilová, Langhamrová, Šimpach [3]).

But already mentioned methods for mechanical smoothing have their drawbacks. The first one is that by using moving averages we cannot perform an extrapolation of mortality curve up to the highest ages (e.g. 110 years of life in developed countries; Thatcher, Kanistö, Vaupel [15]). The second one leads to loss of marginal observations - using moving averages is not possible to smooth all values of age-specific death rates. Nowadays we can use several existing models which are used for smoothing and for estimating of unknown parameters. It is possible to obtain them by using the professional software (e.g. STATA in our case). Among the most famous is included Gompertz-Makeham function (see Gompertz [8], Makeham [9], and the application e.g. by Šimpach [11] or Dotlačilová, Šimpach, Langhamrová [4]). Let us write the Gompertz-Makeham function as

$$\mu(x) = a + b.e^{c.x} \qquad (5)$$

where $\mu_x$ 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 still the same with increasing age. For the intensity of mortality is true the equation

$$\mu\left(x + \frac{1}{2}\right) \doteq m_x \qquad \text{for} \qquad x = 1,2,3,\ldots,85 . \qquad (6)$$

Software used in our analysis is STATA with our own programmed script, RStudio (R Development Core Team [10]) and MS Excel 2013, because their usage in the analysis of demographic data is effective. Presentation possibilities of 3D perspective charts are prepared according to Charpentier, Dutang [2] methodology for actuarial presentations.

## 4 Empirical results

### 4.1 Moving averages

For our calculations we used data about mortality of males and females in the Czech Republic from 1950 to 2011 published by the Human Mortality Database (HMD). The following charts gradually show the development of smoothed age-specific death rates with moving averages of the length 3, 9, and 19 values. From graphical representation of empirical data (see Figure 1 in Materials and Methods part) it is clear that the development of death rates is relatively fluent up to the age of 90 years. In higher ages we can see very significant fluctuations. For this reason it is important to smooth mortality curves. Figure 2 shows the smoothing of empirical values of mortality using 3-values long moving average. It is mechanical method of levelling. Its disadvantage is that there is a loss of initial and the last value. If we compare the obtained results with empirical values it is possible to find that in the highest ages was reduced the level of mortality in comparison with the original values. At the same time moving average does not provide optimally levelled values. It is also evident that from about 80 years there occurs faster growth of mortality when we apply 3-values long moving average. This also means that there is an overestimation of empirical mortality values.

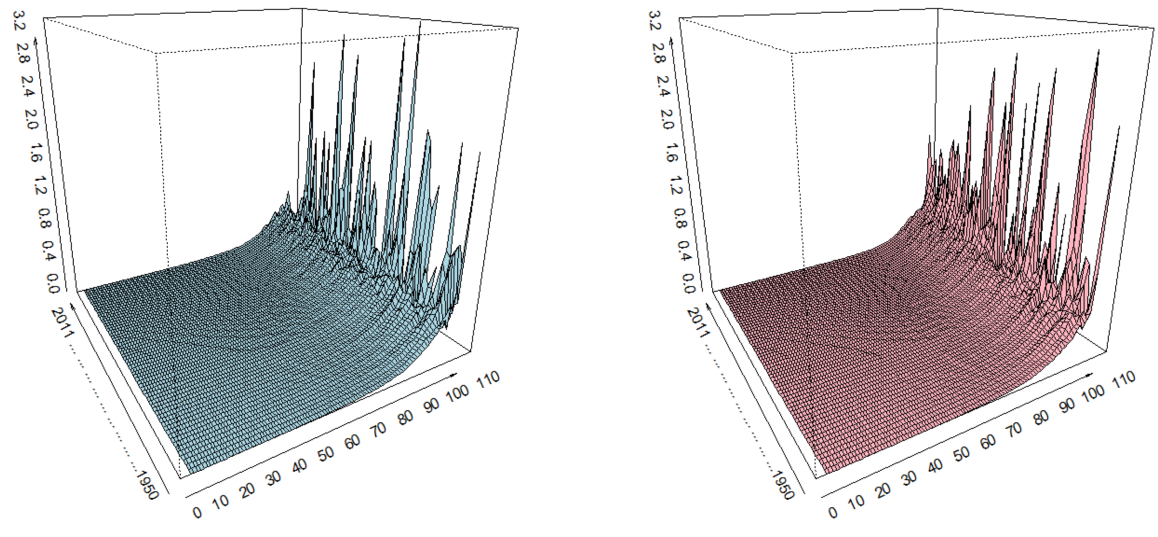


Fig. 2. Smoothing of $m_{x,t}$ using 3-values long moving average for males (left) and females (right). Source: Human mortality database (HMD), authors' construction.

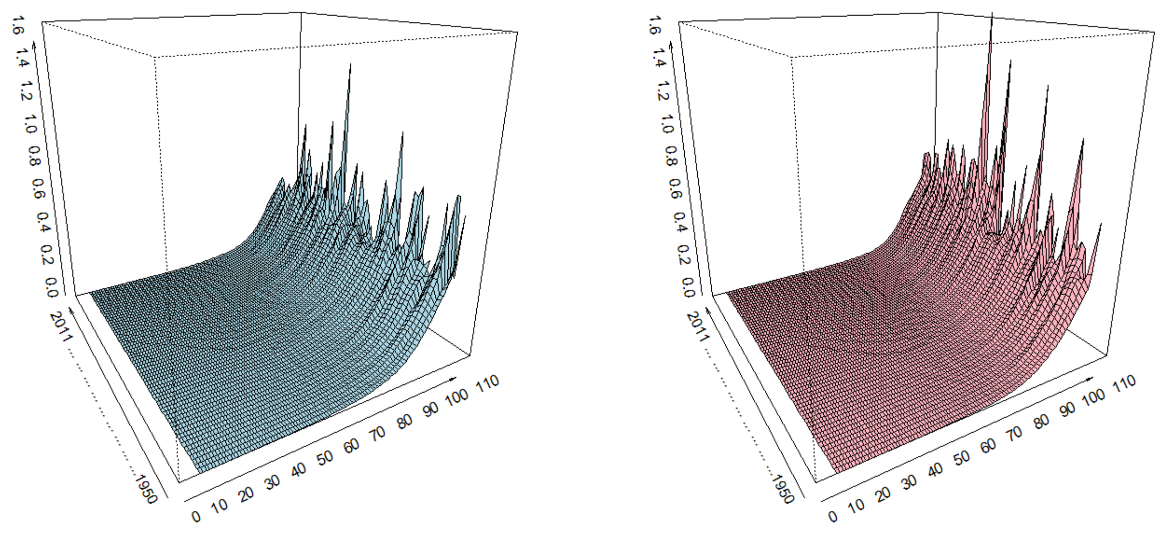


Fig. 3. Smoothing of $m_{x,t}$ using 9-values long moving average for males (left) and females (right). Source: Human mortality database (HMD), authors' construction.

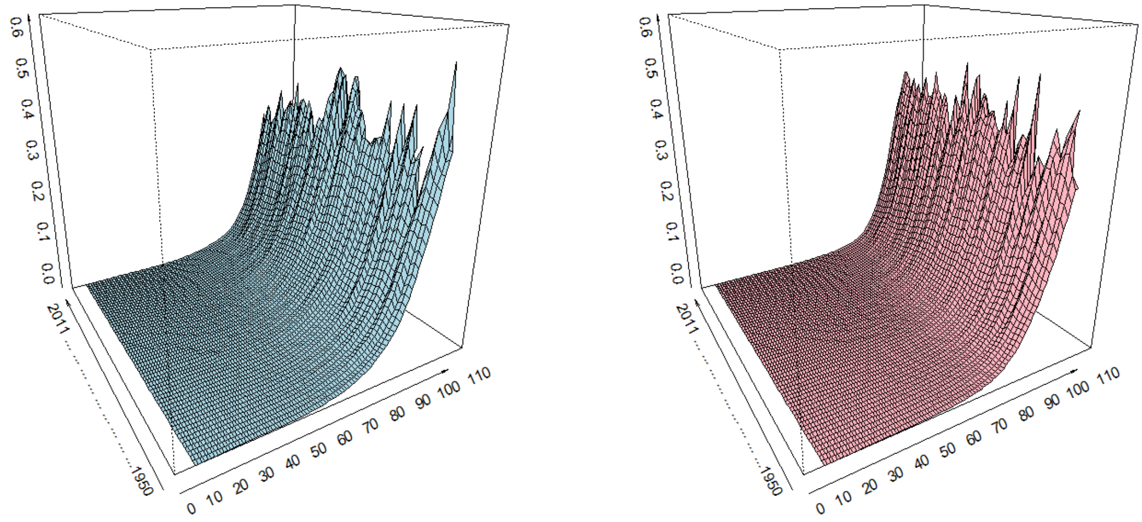


Fig. 4. Smoothing of $m_{x,t}$ using 19-values long moving average for males (left) and females (right).
Source: Human mortality database (HMD), authors' construction.

In the Figure 3 is shown smoothing with 9-values long moving average. At the first sight it is clear that this levelling provides smoother mortality curves. Also together with the previous group of charts is obvious the disadvantage of this mechanical method of smoothing - we lost initial and last values. If we compare these obtained results during the entire age range, we can see lower values of mortality at the highest ages (in comparison with the previous type of smoothing). It is also obvious faster increase in the level of mortality, which can be observed from 70th years of life. Differences between obtained values from using 9-values long moving average and the empirical ones are also underestimation of mortality level. With an increasing age there is more underestimated actual empirical mortality. Final results of smoothing by 9-values long moving average for males and for females in the Czech Republic are shown in Figure 4.

If we compare this smoothing with empirical values of death rates, we find that this moving average smooths fluctuations at the highest ages in the best way. This technique has unfortunately the greatest loss of observation. When we compare the evolution of mortality throughout life, we find that at older ages is the most rapid increase in mortality. This is reflected from about 70 years. If we compare smoothed and empirical values, we can conclude that this technique of 19-values long weighted moving average provides mortality development that it is the closest to real values by mortality law (especially in the highest age).

### 4.2 Gompertz-Makeham function

On the fifth group of charts (see Figure 5) is shown the smoothing of age-specific death rates which we obtained by the Gompertz-Makeham function. The advantage of this smoothing is especially that the selected analytical function can be also used for subsequent extrapolation of death rates and then we can obtain an idea about the evolution of mortality up to the highest ages (i.e. the evolution is modelled using particular function). The Gompertz-Makeham function is more appropriate in comparison with the moving averages in advanced ages (60 years and above). The reason is that data about the number of *x*-years old living and number of *x*-years old deceased at these ages is often quite inaccurate. There is also disrupted the basic assumption of a uniform distribution of deaths in one-year age intervals (which is the key attribute for the usage of different types / length of moving averages).

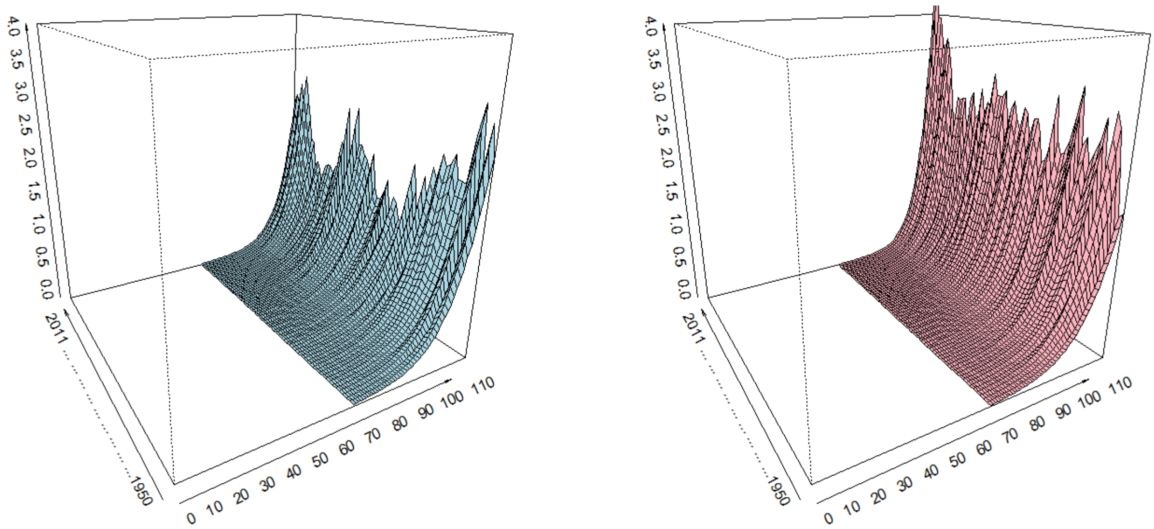


Fig. 5. Smoothing of $m_{x,t}$ using Gompertz-Makeham function for males (left) and females (right).
Source: Human mortality database (HMD), authors' construction.

## 5 Discussion and Conclusion

The aim of this paper was to show the mechanical method of smoothing (by moving averages of different lengths and weights) and its application to the empirical values of mortality in the Czech population. Together with this, we applied an analytical Gompertz-Makeham function, which is currently the most famous used one, and compared the obtained results. Based on our conclusions we can say that better results are obtained for using moving average of 19-values length (see also Figure 6 – Figure 12 in Appendix of this paper). On the other hand the worst results provides smoothing by moving average of 3-values length, where values of mortality are significantly higher in the highest ages in comparison to the actual level of mortality. If we use the technique of smoothing by moving averages it is important to realize, that this approach has its advantages and disadvantages. The advantages include the simplicity of this method. The disadvantages include loss of initial and last observations and the impossibility of extrapolation mortality curves to the highest ages.

The advantage of smoothing by Gompertz-Makeham function was mainly that the estimated parameters could be also used for extrapolation of death rates up to the highest ages. Due to this issue we could obtain an idea about the evolution of mortality up to 110 years. Our script, which we used, was programmed in STATA software and it was prepared for the Gompertz-Makeham function according to literature review.

There are several possibilities for future research. The mechanical method of smoothing age-specific death rates is not universally applicable to all populations in the world. Based on the empirical data from the Human Mortality Database (MHD) we could try this smoothing technique apply to the other populations and determine whether it will or will not provide similar conclusions (see e.g. study by Vogt, Vaupel [16]). (Especially in the case of countries which recorded great variability in mortality data in the past is a chance that even moving average with the length of 19 values would not provide good results).

It is also possible to prepare computing script in RStudio software (R Development Core Team [10]) that will automatically perform technique of moving average' smoothing (see e.g. paper by Gavin, Haberman, Verrall [6] and Charpentier, Dutang [2] (for the case of actuarial insurance modelling by automatized script). This script could be also saved as the original functional package. One of the last opportunities for future research is a combination of this smoothing approach using moving averages with any other analytic function and again try to integrate this combination of multiple approaches into a universal script in RStudio software (see e.g. paper by Šimpach [12] for the case of age-specific fertility rates modeling in R).


## Acknowledgement

The authors of the paper gratefully acknowledge to the Czech Science Foundation project, which is registered under No. P402/12/G097 DYME – Dynamic Models in Economics, for supporting this scientific paper.


## References


[1] Boleslawski, L., Tabeau, E. (2001). Comparing Theoretical Age Patterns of Mortality Beyond the Age of 80. In: Tabeau, E.; van den Berg Jeths, A.; Heathcote, Ch. (eds.): Forecasting Mortality in Developed Countries: Insights from a Statistical, Demographic and Epidemiological Perspective. Springer: Netherlands, pp. 127-155.

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

[3] Dotlačilová, P., Langhamrová, J., Šimpach, O. (2012). Selected logistic models used for levelling and extrapolate mortality curves and their application to the population of the EU countries. Forum Statisticum Slovacum, vol. 8, no. 7, pp. 21-25.

[4] Dotlačilová, P., Šimpach, O., Langhamrová, J. (2014). DERAS Versus MS EXCEL Solver in Levelling the Life Expectancy at Birth. APLIMAT 2014: 13th Conference on Applied Mathematics. Bratislava: Publishing House of STU, p. 108–114. ISBN 978-80-227-4140-8.

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

[6] Gavin J., Haberman S., Verrall R. (1993). Moving weighted average graduation using kernel estimation. Insurance: Mathematics & Economics, vol. 12, pp. 113-126.

[7] Gavrilov, L. A., Gavrilova, N. S. (2011). 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. (1825). 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, vol. 115, pp. 513-585.

[9] Makeham, W. M. (1860). On the Law of Mortality and the Construction of Annuity Tables. The Assurance Magazine, and Journal of the Institute of Actuaries, vol. 8, pp. 301-310.

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

[11] Šimpach, O. (2012). Faster convergence for estimates of parameters of Gompertz-Makeham function using available methods in solver MS Excel 2010. Proceedings of the 30th International Conference on Mathematical Methods in Economics, PTS I and II, pp. 870-874.

[12] Šimpach, O. (2015). Fertility of Czech Females Could Be Lower than Expected: Trends in Future Development of Age-Specific Fertility Rates up to the Year 2050. Statistika, vol. 95, no. 1, pp. 19-35.

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

[14] Šimpach, O., Pechrová, M. (2015). Development of the Czech Farmers' Age Structure and the Consequences for Subsidy Policy. Agris on-line Papers in Economics and Informatics, vol. VII, no. 3, pp. 57-69. ISSN 1804-1930.

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

[16] Vogt, T. C., Vaupel, J. W. (2015). The importance of regional availability of health care for old age survival - Findings from German reunification. Population Health Metrics, vol. 13, no. 26, article no. 6, pp. 1-8.

## Appendix

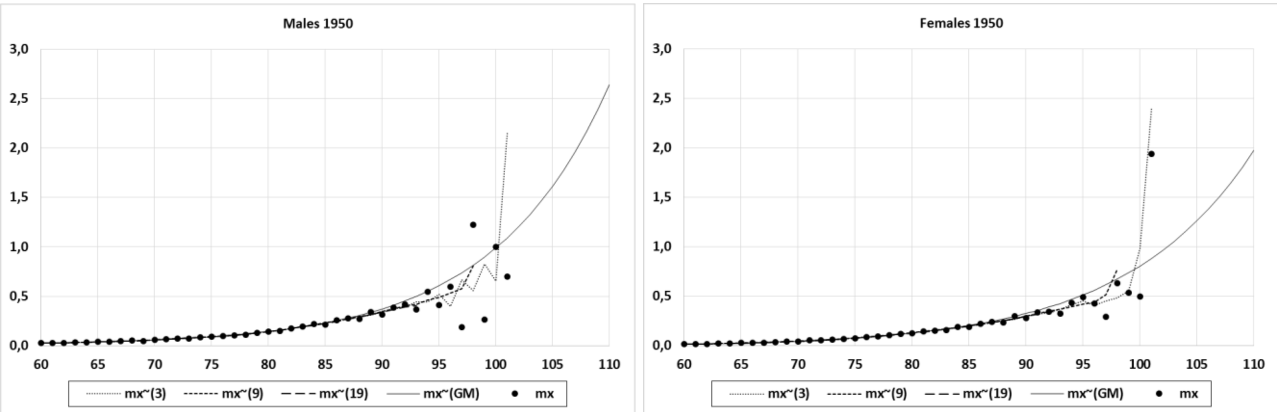


Fig. 6. Smoothing of $m_{x,t}$ using 3, 9 and 19-values long moving average and using Gompertz-Makeham function for males (left) and females (right) in 1950. Source: authors' construction.

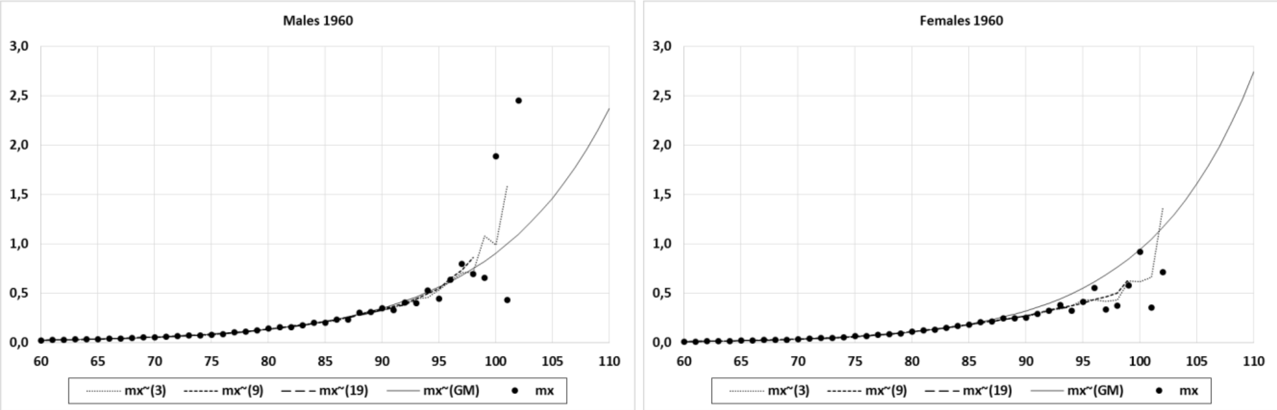


Fig. 7. Smoothing of $m_{x,t}$ using 3, 9 and 19-values long moving average and using Gompertz-Makeham function for males (left) and females (right) in 1960. Source: authors' construction.

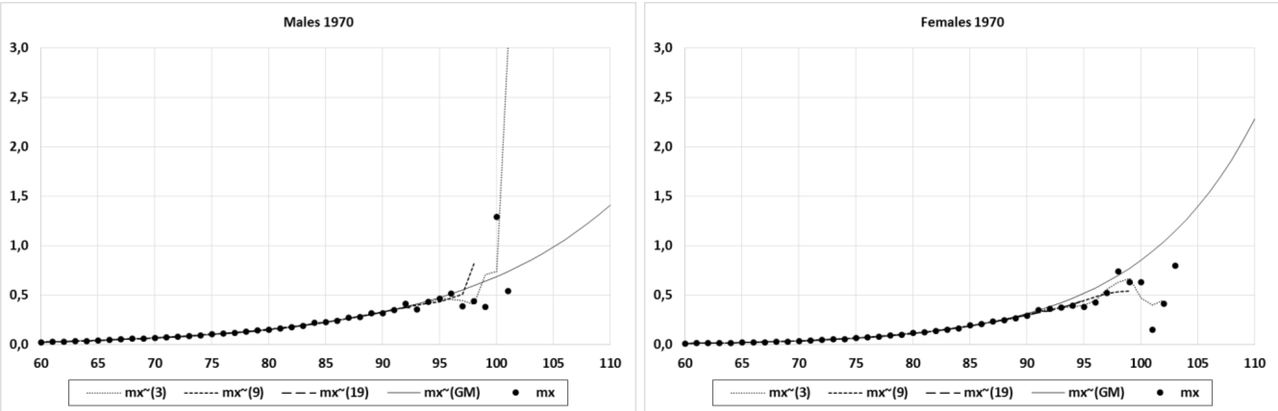


Fig. 8. Smoothing of $m_{x,t}$ using 3, 9 and 19-values long moving average and using Gompertz-Makeham function for males (left) and females (right) in 1970. Source: authors' construction.

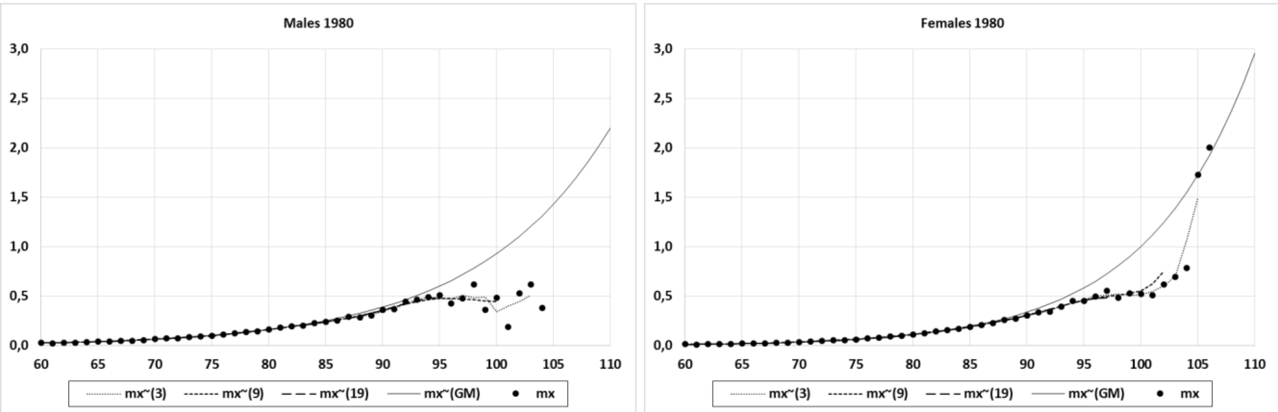


Fig. 9. Smoothing of $m_{x,t}$ using 3, 9 and 19-values long moving average and using Gompertz-Makeham function for males (left) and females (right) in 1980. Source: authors' construction.

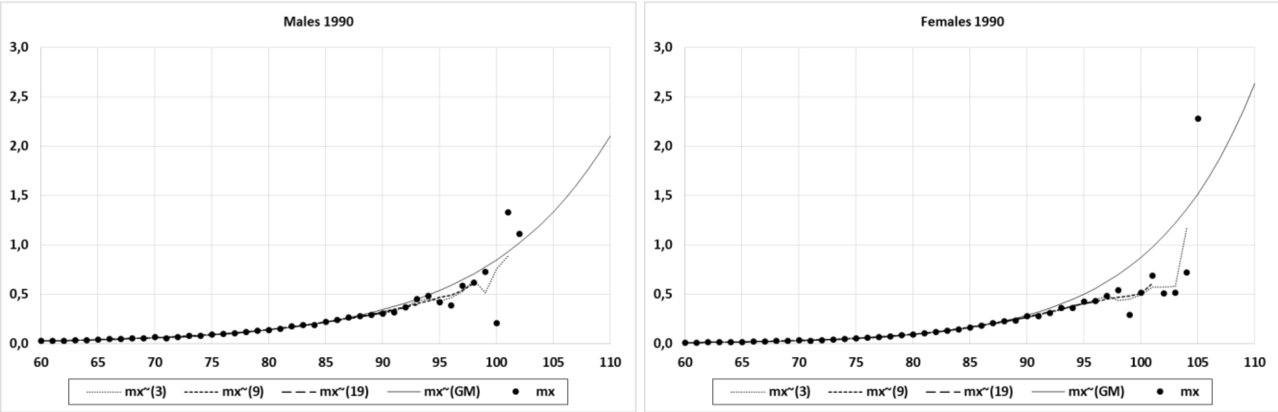


Fig. 10. Smoothing of $m_{x,t}$ using 3, 9 and 19-values long moving average and using Gompertz-Makeham function for males (left) and females (right) in 1990. Source: authors' construction.

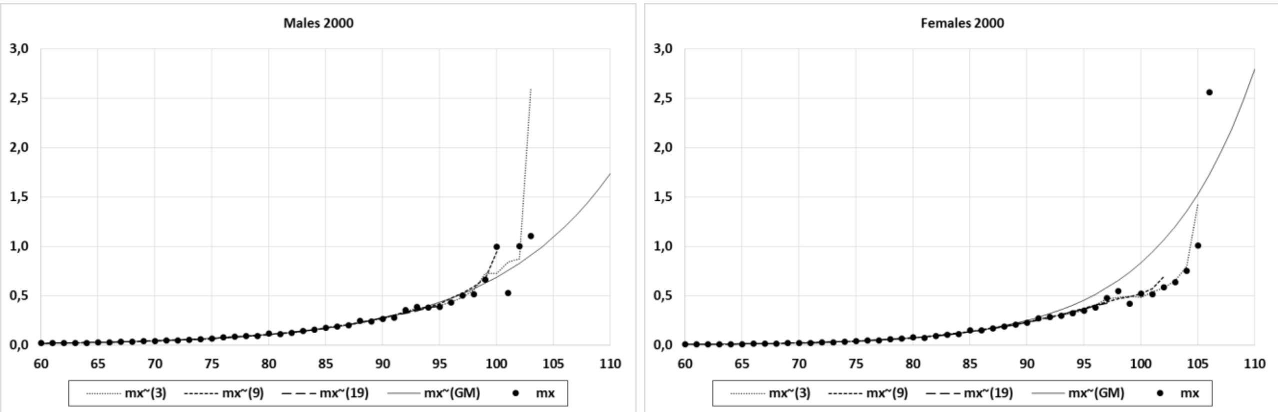


Fig. 11. Smoothing of $m_{x,t}$ using 3, 9 and 19-values long moving average and using Gompertz-Makeham function for males (left) and females (right) in 2000. Source: authors' construction.

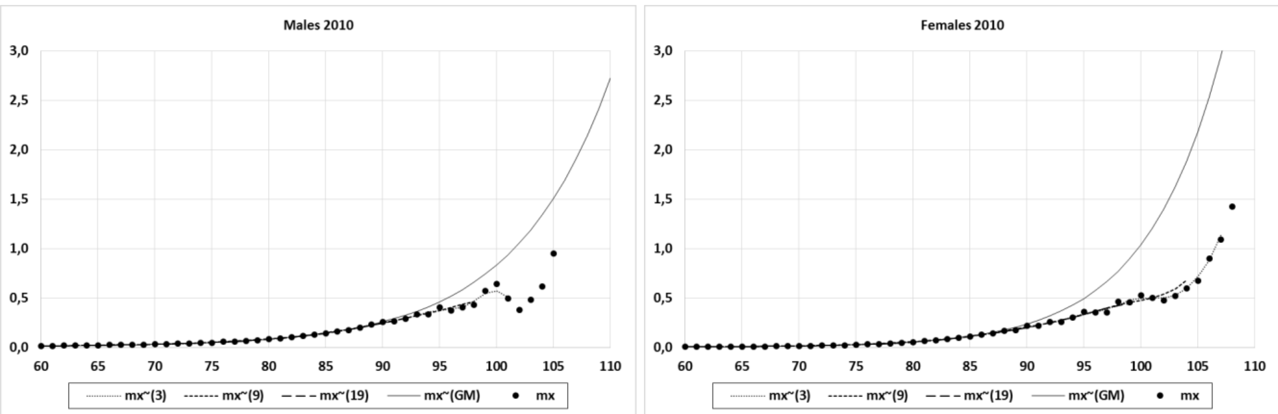


Fig. 12. Smoothing of $m_{x,t}$ using 3, 9 and 19-values long moving average and using Gompertz-Makeham function for males (left) and females (right) in 2010. Source: authors' construction.

**Current address**

**Petra Dotlačilová, Ing.**
University of Economics in Prague, Faculty of Informatics and Statistics
W. Churchill sq. 4, 130 67 Prague 3, Czech Republic
Tel. number: +420 224 095 273, e-mail: petra.dotlacilova@vse.cz

**Ondřej Šimpach, Ing.**
University of Economics in Prague, Faculty of Informatics and Statistics
W. Churchill sq. 4, 130 67 Prague 3, Czech Republic
Tel. number: +420 224 095 273, e-mail: ondrej.simpach@vse.cz