COMPARISON OF SELECTED METHODOLOGIES USED FOR SMOOTHING AND EXTRAPOLATION OF MORTALITY CURVE

Petra Dotlačilová

Abstract

During the human life we can observe various fluctuations in mortality. Natural character of mortality begins to prevail at higher ages (approximately 60 years and higher). But especially at very high ages we can observe significant deviations in mortality. That is why it is important to smooth mortality curve for obtaining better results. Due to the aging population it is very important to have the best imagination about development of mortality at the highest ages.

In this paper will be presented methodology which is used for modelling of the intensity of mortality by the Czech Statistical Office. These results will be compared with the methodology which use the modified Gompertz-Makeham function. The main contribution of this paper is the comparing of both mentioned methodologies and then the discussion about suitability of both methods.

For own calculation will be used the data about mortality of population in the Czech Republic since 1960. Results will be published for the beginning of each decade.

Key words: mortality, moving averages, Gompertz-Makeham function, modified Gompertz-Makeham function

JEL Code: J10, J11, J19

Introduction

Aging of population is very often discussed topic in these days. It means higher number of living in higher ages (Kannisto et al., 1994 or Thatcher et al., 1998). It is very important to know that mortality in higher ages (60+) is influenced by systematic and random errors (see Dotlačilová, 2013 or Dotlačilová et al., 2014). It is the reason for modeling of mortality (Dotlačilová, Langhamrová, 2014 or Gavrilov, Gavrilova, 2012). For modeling of mortality in this article will be used the Czech Statistical Office methodology (CZSO, 2016). For comparison will be mentioned the other methodology which use the combination of the Gompertz-Makeham function and its modification.

1 Methodology

1.1 The Czech Statistical Office Methodology

At first we have to calculate age-specific death rates as:

$$m_x = \frac{M_x}{\overline{S}_x}, \tag{1}$$

where M_x is the number of deaths at complete age x and $\overline{S}_x$ is mid-year population at age x.

The input characteristic for the calculation of mortality tables is probability of dying (q_x). It is obtained by using m_x (CZSO, 2016):

$$q_x = 1 - e^{-m_x}, \text{ pro } x=1,2,\ldots,\omega-1, \tag{2}$$

where ω is the lowest age at which nobody is alive.

Probability of dying at age 0 will be calculated like:

$$q_0 = \frac{M_0}{N^V}, \tag{3}$$

where M_0 is the number of deaths at complete age 0, N^V is the number of live births.

The other step is about eliminating of random fluctuations by smoothing probabilities of dying:

$$q_x^{smoothed} = [105 * q_x + 90 * (q_{x-1} + q_{x+1}) + 45 * (q_{x-2} + q_{x+2}) - 30 * (q_{x-3} + q_{x+3})] / 315 \tag{4}$$

where q_x is empirical probability of dying at age x.

Smoothing of probability of dying according to formula (4) is performed from 4 years. It means that the age 0 will not be included into calculations.

After this it is need to calculate the input characteristic for the Gompertz-Makeham function (G-M function). It is natural logaritm of probability of surviving (p_x). The input values of probability of surviving (p_x) are obtained by using smoothed values of probability of dying (q_x):

$$\ln p_x^{smoothed} = \ln(1 - q_x^{smoothed}) \tag{5}$$

Probability of dying will be derived from the G-M function at higher ages (due to small numbers of deaths and living at the highest ages):

$$\ln p_x = a + b.c^x \tag{6}$$

a, b and c are unknown parameters of the G-M function.

The G-M function will be used for extrapolation of mortality curve approximately from age 80.

For the estimation of unknown parameters of the G-M function will be used the King-Hardy methodology. Calculation of initial estimates of parameters is based on three equation (R_1, R_2, R_3). These are derived for three consecutive intervals of the same length (CZSO, 2016).

$$R_1 = \sum_{x=x_0}^{x_0+d-1} \ln p_i^{smoothed} \quad R_2 = \sum_{x=x_0+d}^{x_0+2d-1} \ln p_i^{smoothed} \quad R_3 = \sum_{x=x_0+2d}^{x_0+3d-1} \ln p_i^{smoothed}, \tag{7}$$

where d is the length of age interval (d=8), x_0 is age from which smoothing will be performed (x_0=60).

The unknown parameters of the G-M function characterizing the force of mortality will be calculated according to formulas:

$$c^d = \frac{R_3 - R_2}{R_2 - R_1}, \; c = \sqrt[d]{c^d}, \tag{8}$$

$$b = \frac{(c-1).(R_2 - R_1)}{c^{x_0}.(c^d - 1)^2}, \tag{9}$$

$$a = \left[R_1 - \frac{(R_2 - R_1)}{(c^d - 1)} \right] / d. \tag{10}$$

Parameters of the G-M function are derived from initial estimating equations R_1, R_2, R_3. That is why values of parameters will be negative.

The obtained estimated parameters are substituted into the G-M function. According to this substitution we will obtained modeled values of probability of surviving for age $x \geq 71$ like:

$$r_x = \exp(a + b.c^x) \tag{11}$$

After that it is important to find the age y for which deviation $| p_x^{smoothed} - r_x |$ is minimal. From this age q_x^{GM} (12) will be used (it is unit supplement to r_x). Transition to the extrapolated values is corrected by smoothing values for the age $z = (y-4),\ldots,(y+4)$ like:

$$q_z^{GM} = 1 - \left[(1 - \frac{z-y+5}{10}).p_z^{smoothed} + \frac{z-y+5}{10}.r_z \right] \tag{12}$$

Summary – input probability of dying used in mortality tables:

- age 0 …………… infant mortality,
- ages 1, 2, 3, …….. probability of dying derived from m_x,

- ages from 4 to (y – 5)…smoothed probability of dying,
- from (y – 4) to ω – 1 ….. probability of dying smoothed and extrapolated.

Smoothed values of age-specific death rates will be obtained like:

$$m_x = -\ln(1 - q_x) \tag{13}$$

1.2 Methodology for modeling of the intensity of mortality – combination of the Gompertz-Makeham function and its modification

The first calculation in the second methodology is connected with calculation of age-specific death rates according to formula (1). The second step is smoothing of empirical death rates by different types of moving averages (Fiala, 2005 or Dotlačilová et al., 2014).

At first will be used 3 values moving average:

$$\tilde{m}^{(3)} = \frac{m_{x-1} + m_x + m_{x+1}}{3}, \text{ where } x \in <3; 5>. \tag{14}$$

After that will be used 9 values moving average:

$$\begin{aligned}\tilde{m}^{(9)} &= 0{,}2m_x + 0{,}16(m_{x-1} + m_{x+1}) + 0{,}12(m_{x.-2} + m_{x+2}) \\ &+ 0{,}08(m_{x-3} + m_{x+3}) + 0{,}04(m_{x-4} + m_{x+4}),\end{aligned} \tag{15}$$

where $x \in$ <6; 29>.

At the end 19 values moving average will be used:

$$\begin{aligned}\tilde{m}^{(19)} &= 0{,}2m_x + 0{,}1824(m_{x-1} + m_{x+1}) + 0{,}1392(m_{x-2} + m_{x+2}) \\ &+ 0{,}0848(m_{x-3} + m_{x+3}) + 0{,}0336(m_{x-4} + m_{x+4}) - 0{,}0128(m_{x-6} + m_{x+6}), \\ &- 0{,}0144(m_{x-7} + m_{x+7}) - 0{,}0096(m_{x-8} + m_{x+8}) - 0{,}0032(m_{x-9} + m_{x+9})\end{aligned} \tag{16}$$

for x <30; 59>.

For age 1 and 2 the empirical values of age-specific death rates will be used. At the advanced ages is used the Gompertz–Makeham function (Gompertz, 1825, Makeham, 1860 or Ekonomov, Yarigin, 1989) or the modified Gompertz–Makeham function (mG–M) (23) (see Boleslawski, Tabeau, 2001, Burcin et al., 2010 or Lagerås, 2010):

$$\mu_x = a + bc^x. \tag{17}$$

For the estimation of unknown parameters of the G-M function will be used the King-Hardy methodology. Calculation of initial estimates of parameters is based on three equation (G_1, G_2, and G_3).

$$G_1 = \sum_{x=x_0}^{x_0+k-1} m_x,\ G_2 = \sum_{x=x_0+k}^{x_0+2k-1} m_x,\ G_3 = \sum_{x=x_0+2k}^{x_0+3k-1} m_x \tag{18}$$

where k = 8 and x_0 = 60. We can also express G_1 by using the G-M function (e.g. Fiala, 2005):

$$G_1 = \sum_{x=x_0}^{x_0+k-1}\left(a + b.c^{x+\frac{1}{2}}\right) \quad (19)$$

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

$$c^k = \frac{G_3 - G_2}{G_2 - G_1}. \quad (20)$$

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

$$c = \sqrt[k]{c^k}. \quad (21)$$

For the estimation of the initial values of parameters a and b we have to calculate the value of sub-expression of K_c as

$$K_c = c^{x_0+\frac{1}{2}}.\frac{c^k - 1}{c - 1}.$$

and parameters b and a we obtain as

$$b = \frac{G_2 - G_1}{K_c.(c^k - 1)}, \text{ resp. } a = \frac{G_1 - b.K_c}{k}. \quad (22)$$

At the end calculated parameters a, b and c are substituted into the G-M function (17). For continuation of analysis, let us explain modified Gompertz-Makeham function (mG–M function) (see e.g. Burcin et al., 2010) as

$$\mu_x = a + bc^{x_0+\frac{1}{\gamma}\ln[\gamma(x-x_0)+1]}, \quad (23)$$

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

The last step is the calculation of the weighted squared deviations (WSD), through which we will optimize parameters of the G–M (respectively mG–M) function (Dotlačilová, Langharmrová, 2014 or Fiala, 2005)

$$WSD = \frac{S_{t,x} + S_{t+1,x}}{2.m_x.(1 - m_x)} \cdot \left(m_x - \tilde{m}_x^{(GM)}\right)^2 \text{ for } x \in <60\,; y>, \quad (24)$$

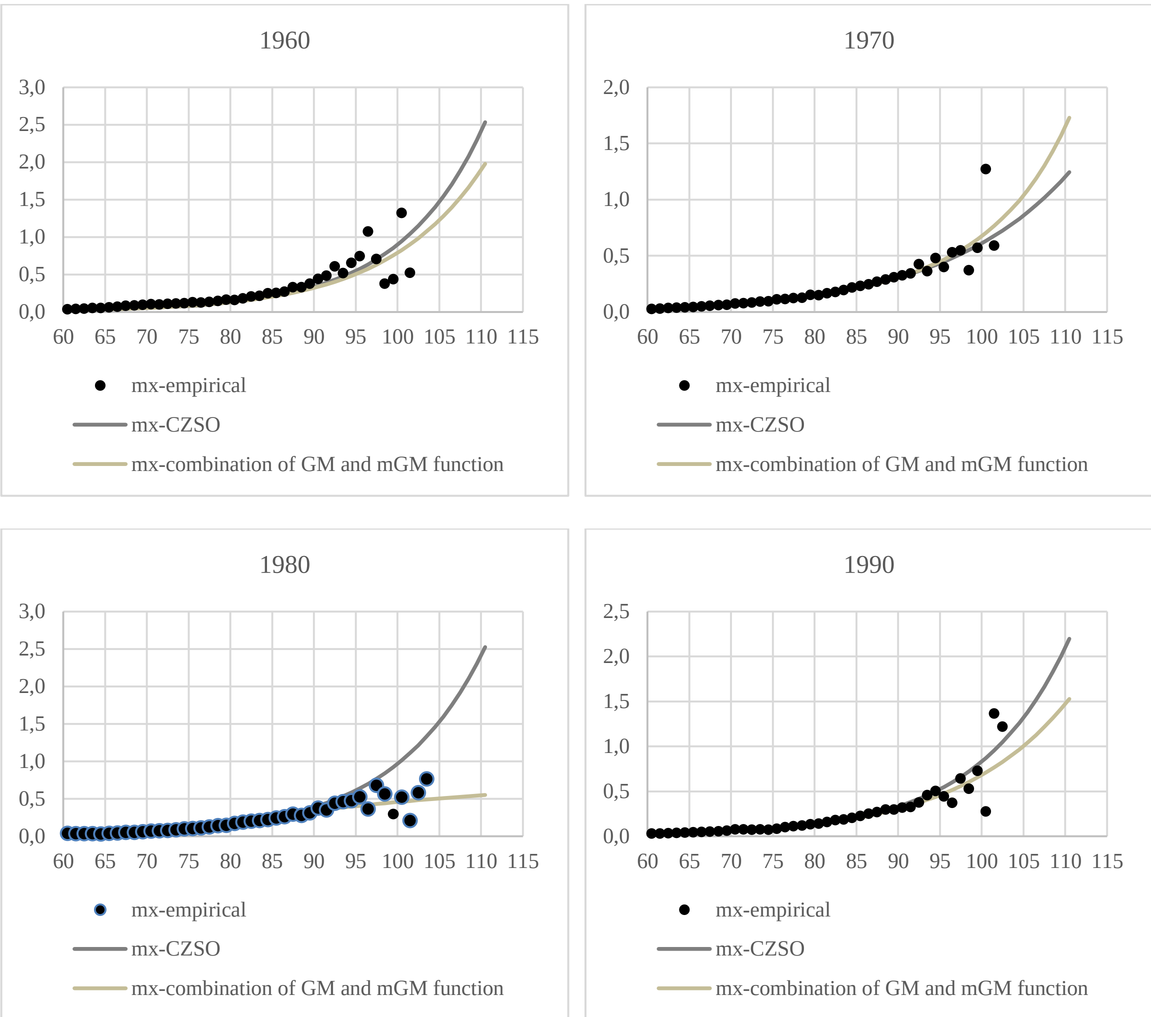
where $S_{t,x}$ is the number of living at age x in year t, S_{t+1},x is the number of living at age x in year t+1, m_x are age-specific death rates, $m_x^{\sim(GM)}$ are smoothed values of age-specific death rates obtained by using G–M or mG–M function, y is the highest age for which we have a non-zero value of m_x. When we optimize parameters using by OLS, it is necessary to create two instrumental sums (S_1 for the Gompertz-Makeham function and S_2 for its modification).

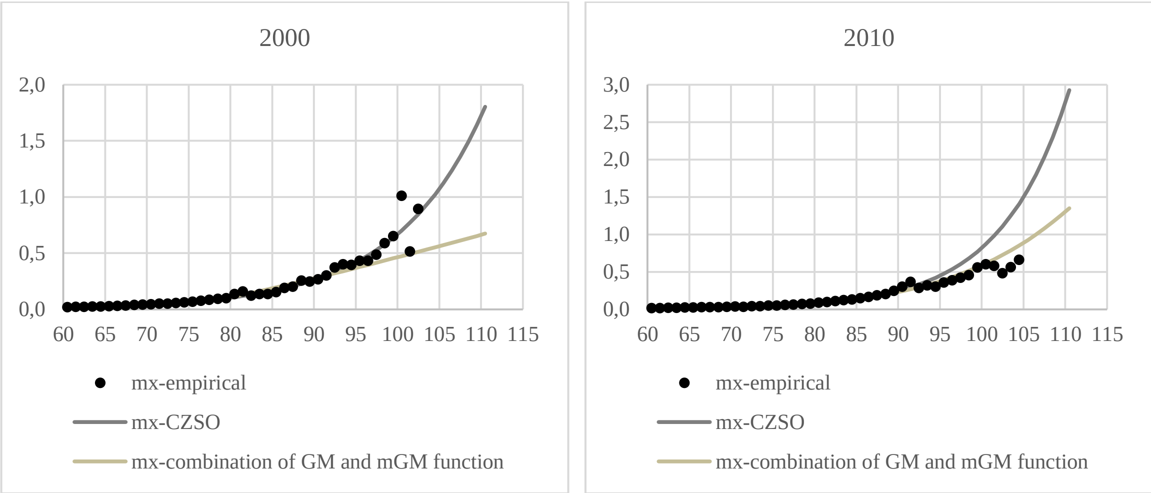
$$S_1 = \sum_{x=60}^{82} \frac{S_{t,x} + S_{t+1,x}}{2.m_x.(1-m_x)} \cdot \left(m_x - \tilde{m}_x^{(GM)} \right)^2 \text{, resp. } S_2 = \sum_{x=83}^{y} \frac{S_{t,x} + S_{t+1,x}}{2.m_x.(1-m_x)} \cdot \left(m_x - \tilde{m}_x^{(GM)} \right)^2 .$$

2 Results

The next part will be dedicated to the presentation of obtaining results according to both methodologies. For own calculations was used the data about mortality of males and females in the Czech Republic since 1960. Results will be published only for the years at the beginning (resp. at the end) of each decade.

Fig. 1: Smoothed values of mortality – CZSO methodology and methodology use for modeling of the intensity of mortality (combination of the Gompertz-Makeham function and its modification – the Czech Republic, males

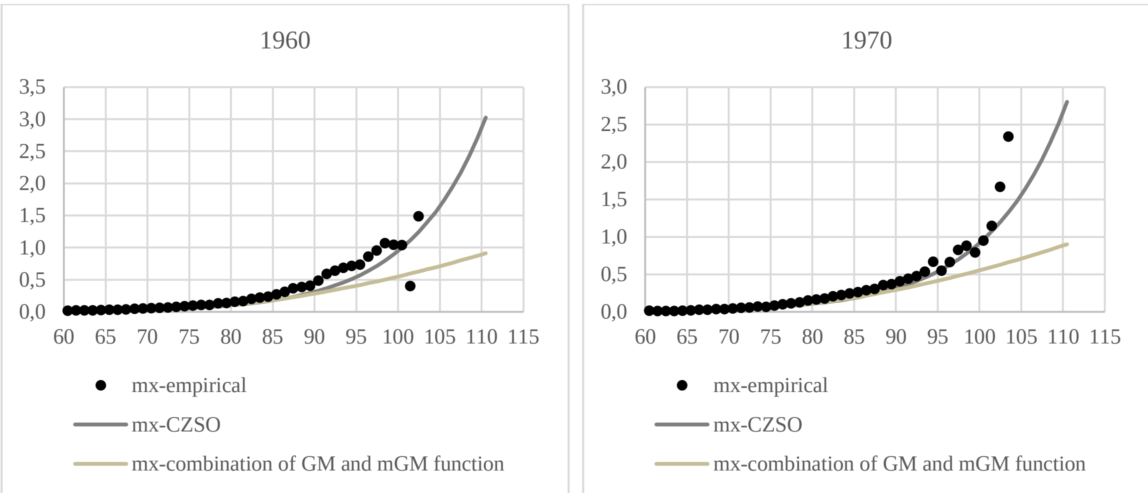


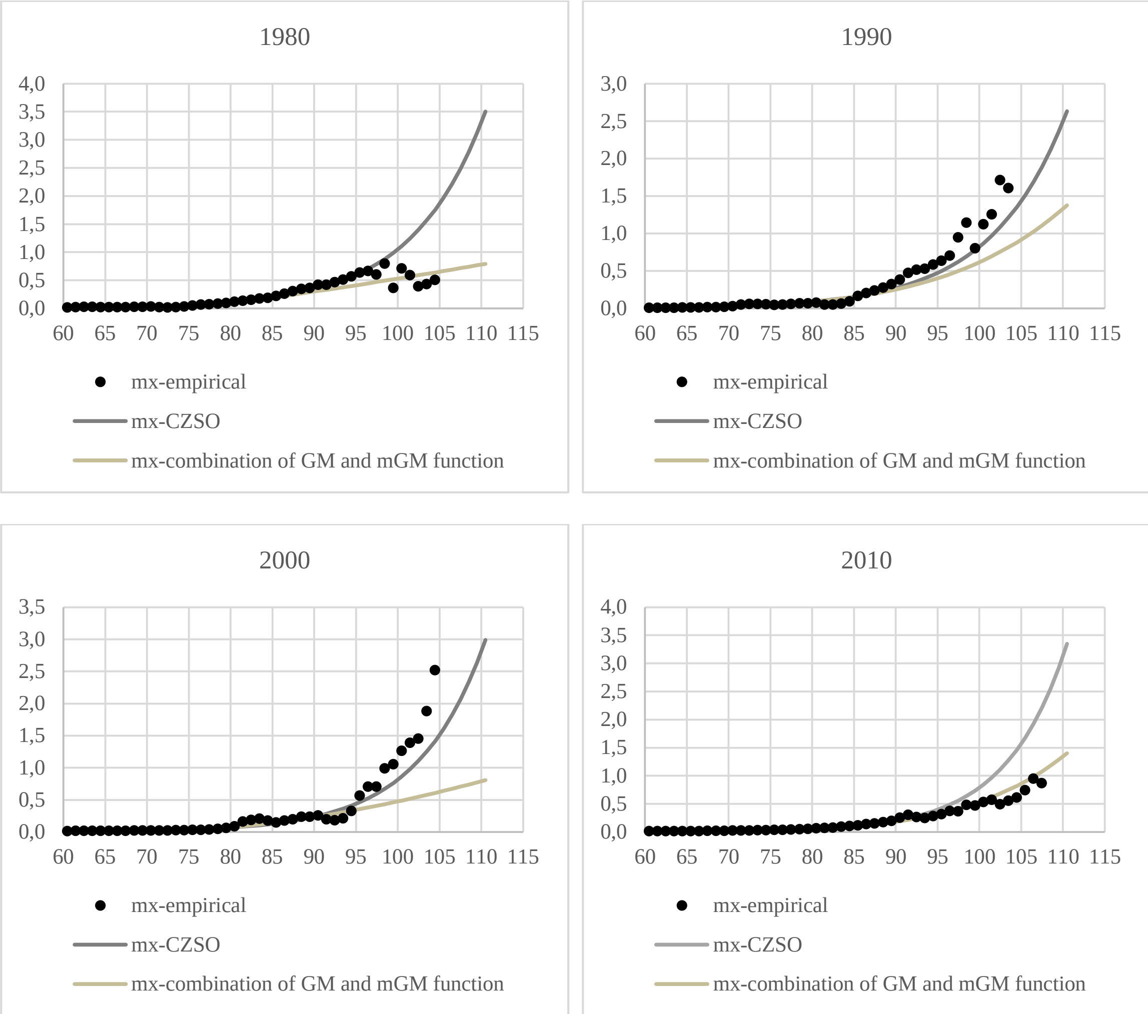


Source: data The Human Mortality Database (HMD, 2014), author's calculations

For each beginning of every decade we find out that second smoothing methodology using the combination of the Gompertz-Makeham function and its modification provides lower values (in comparison with CZSO methodology). On the other hand it is visible different evolution in both curves of mortality. At the beginning of reporting period empirical mortality reaches higher values. It is possible to say that CZSO methodology is more suitable. Towards to the present the second methodology could be better.

Fig. 2: Smoothed values of mortality – CZSO methodology and methodology use for modeling of the intensity of mortality (combination of the Gompertz-Makeham function and its modification – the Czech Republic, females





Source: data The Human Mortality Database (HMD, 2014), author's calculations

We can make similar conclusion for females' population – CZSO methodology provides higher values of mortality (in comparison with the second one). On the other hand it is possible to say that CZSO methodology is more suitable at the beginning of the period. But to the present mothodology using combination of the Gompertz-Makeham function and its modification is better. It is visible especially from the last figur efor 2010.

Conclusion

During the comparison of both mentioned methodologies used for modelling of the intensity of mortality we find out that the second one provides lower values (in comparison with CZSO methodology). It is caused by including the modified Gompertz-Makeham function into the calculation. Modified Gompertz-Makeham function is based on decrease in increase of mortality with an increasing age. It causes lower increase in mortality in higher ages.

For both populations it is noticeable that the second methodology is more suitable towards to the present. This conclusion corresponds with the assumption of decreasing mortality at higher ages.

We can make a conclusion that for the modeling of the intensity of mortality in the Czech Republic is better to use combination of the Gompertz-Makeham function and its modification in these days.

References

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.

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.

CZSO 2016. [cit. 23.04.2016]. Available from WWW: http://www.czso.cz/csu/redakce.nsf/i/umrtnostni_tabulky_metodika

DOTLAČILOVÁ, P. (2013). Changes in the Development of the Normal Length of Life and Life Expectancy in the Czech Population. *International Days of Statistics and Economics* (pp. 357-364). [online] Prague, 19. 09. 2013 – 21. 09. 2013. Slaný: Melandrium. ISBN 978-80-86175-87-4. URL: http://msed.vse.cz/files/2013/144-Dotlacilova-Petra-paper.pdf.

DOTLAČILOVÁ, P., LANGHAMROVÁ, JI. The Influence of Mortality Models for the Expected Future Life-time of Older People. In: *AMSE* [CD ROM]. Jerzmanovice, 27.08.2014 – 31.08.2014. Wroclaw: Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, 2014, s. 65–74. ISBN 978-83-7695-421-9.

DOTLAČILOVÁ, P., ŠIMPACH, O., LANGHAMROVÁ, JI. The Use of Polynomial Functions for Modelling of Mortality at the Advanced Ages. In: *Mathematical Methods in Economics* (MME2014). Olomouc, 10.09.2014 – 12.09.2014. Olomouc: Palacký University in Olomouc, 2014, s. 174–179. ISBN 978-80-244-4208-2. CD ISBN 978-80-244-4209-9.

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.

FIALA, T. (2005). *Výpočty aktuárské demografie v tabulkovém procesoru*. Praha: Vysoká škola ekonomická v Praze. ISBN 80-2450821-4.

GAVRILOV, L.A., GAVRILOVA, N.S. (2012). Mortality measurement at advanced ages: a study of social security admin-istration death master file. *North American actuarial journal*, vol. 15, no. 3, pp. 432 – 447.

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.

HUMAN MORTALITY DATABASE 2014. [cit.18.3.2014]. Available from WWW: www.mortality.org

KANNISTO, V., LAURITSEN, J., THATCHER, A., R., VAUPEL, J., W. (1994). Reductions in mortality at advanced ages - several decades of evidence from 27 countries. *Population and development review*, vol. 20, no. 4, pp. 793-810.

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

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.

THATCHER, R.A., KANISTÖ, V. AND VAUPEL, J.W. *The Force of Mortality at Ages 80 to 120*. Odense University Press, 1998.

Contact

Petra Dotlačilová

University of Economics in Prague (Faculty of Informatics and Statistics, Dep. of Mathematics)

W. Churchill Sq. 4

130 67 Prague 3

Czech Republic

petra.dotlacilova@vse.cz