The most frequently used models for extrapolating mortality curves and their application to the Czech population

Petra Dotlačilová

Abstract

Demographers are still trying to find a way of modelling the relationship between mortality and age. So far the Gompertz – Makeham function was for a long time universally used for extrapolating mortality curves. But at the present it is important to develop other models. It is primarily due to the fact that more and more people live to the old age. The second reason is better availability of statistical data. As a result of improving oldest – old mortality and better statistical data many new models have emerged. In this article, some of these models will be introduced and applied to the data on mortality Czech population. The results will be compared with the Czech Statistical Office mortality tables and the mortality tables without extrapolation.

Key words: mortality, models used for extrapolating mortality curves, mortality tables without extrapolation, methodology of the Czech Statistical Office used for calculating mortality tables

JEL Code: C, C1, C10

Introduction

Demographers always emphasized on the issue of estimating the mortality. Often, however, also alluded to the problem of the availability of statistical data. Problematic data was mainly among people over 80 – years. Previously, however, this did not live much aged persons. But right now it is changed. More and more people live to the old age. It also improves the availability of statistical data. In connection with this development, the approach has been to develop more models to estimate the mortality depending on the age. Estimates of the mortality that with increasing age decreases a lot, are dependent on the reliability of empirical data. Generally regarded as critical threshold age of 85 years. It is also important that the highest age there is a significant decrease in the number of deaths. To solve these problems, we can use one of many available models and extrapolated values of mortality up to the highest age.

1 The most frequently used models for extrapolating mortality curves

For a long time the dominant model for describing mortality depending on the age became the Gompertz model. This assumes that mortality intensity changes exponentially with age (ie, with a constant relative proportion), (*Thatcher et al.,* 1998).

Gompertz function: $\mu_x = b.c^x$,

where b a c are parameters a x is age.

Later, improvements Gompertz model William M. Makehamem which model supplemented by a constant expressing the independence of mortality on age. Constant consists mainly of external causes of death, (*Thatcher et al.,* 1998).

Gompertz – Makeham model: $\mu_x = a + b.c^x$,

where x is age and a, b, c are parameters.

Based on empirical data, it was found that the Gompertz - Makeham model is suitable to apply to approximately 80 - years. In later life can no longer be relative mortality increases with age as constant. For this reason the modified Gompertz - Makeham model, which is based on the assumption that the rate of increase in mortality with increasing age decreases (*Koschin*, 1999).

Modified Gompertz – Makeham model: $\mu_x = a + b.c^{x_0 + \frac{1}{\gamma}.\ln[\gamma.(x - x_0) + 1]}$,

where $x > x_0$, x_0 is age, from which the compensation is performed using the modified Gompertz – Makeham model, a, b, c and γ are parameters.

This model is useful for leveling around the age of 85. It is a model that is focused on the highest ages.

The most frequently used models (*Boleslawski & Tabeau*, 2001):

Gompertz – Makeham function

model: $\mu_x = a + b.c^x$, (1)

where x is age and a, b, c are parameters.

Modified Gompertz – Makeham function

model: $\mu_x = a + b.c^{x_0 + \frac{1}{\gamma}.\ln[\gamma.(x - x_0) + 1]}$, (2)

where $x>x_0$, is age, from which the compensation is performed using the modified Gompertz - Makeham model, a, b, c and γ are parametes.

Himes – Preston – Condran

This model is designed primarily for high ages. It is based on two components. The first component is called Himes - Preston - Condran mortality standard. Its design is based on empirical data from 16 - demographically advanced countries in the years 1948 - 1985. The second component is an extension of the mortality standard for the ages from 95 to 115 years (assuming linear growth logit - transformed age - specific mortality rates).

Original model (*Boleslawski & Tabeau*, 2001):

$$\log it(m_x) = \alpha + \beta.x.$$

Model after corection: $m_x = \dfrac{b.e^{a.x}}{1 + b.e^{a.x}}$, (3)

where a, b are estimates of model parameters, x is age a m_x is specific mortality rate.

Heligman – Pollard

This model was originally designed in order to eliminate the lack of the previous models. It contained eight parameters and tried to describe the entire age span. The model consists of three parts. Because previous models focused on mortality at the highest ages, we will consider only the third part of this model this model (ie the one that deals with mortality in old age).

Model (*Boleslawski & Tabeau*, 2001): $q_x = \frac{b.e^{a.x}}{1+b.e^{a.x}}$, (4)

where *a*, *b* are parameters, *x* is age a q_x is probability of death.

Thatcher

Thatcher's model is one of the logistic models because it assumes logistic mortality curve.

Model (*Burcin et al.*, 2010): $\mu_x = \frac{z}{1+z} + \gamma$, (5)

where $z = \alpha.e^{\beta.x}$, α, β a γ are parameters.

Kannistö

Kannistö´s model is a special case of the logistic function where the logit transformation of mortality rates is expressed as a linear function of age.

Model (*Burcin et al.*, 2010): $\mu_x = \frac{e^{[\Theta_0+\Theta_1.(x-80)]}}{1+e^{[\Theta_0+\Theta_1.(x-80)]}}$, pro $x \geq 80$, (6)

where Θ_0, Θ_1 are parameters, which take non-negative values, μ_x is intensity of mortality in age *x*.

Kannistö's model is again specifically designed for the highest ages.

Beard

Beard‘s model is also one of the logistic models.

Model (*Boleslawski & Tabeau*, 2001): $m_x = \frac{b.e^{a.x}}{1+c.e^{a.x}}$, (7)

where *a*, *b*, *c* are parameters and m_x is specific mortality rate in age *x*.

Cubic model

It is an extension of Gompertz - Makeham function.

Model (*Burcin et al.*, 2010): $\mu_x = B.C^x.D^{x^2}.E^{x^3}$. (8)

For the estimation is better to use its logarithmic transformation:

$$\ln \mu_x = \ln B + x.\ln C + x^2.\ln D + x^3.\ln E ,$$

where x is age and B, C, D, E are parameters.

Coale – Kisker

This model is focused on changes in mortality rates between 2 successive ages. It is further assumed that the growth rate of mortality at the highest ages decreases linearly. Based on this assumption was defined variable k_x (*Boleslawski & Tabeau*, 2001):

$$k_x = \ln(\frac{m_x}{m_{x-1}}) , \tag{9}$$

where m_x is specific mortality rate in age x.

From the age 85, it is assumed that k_x is linear:

$$k_x = k_{85} - (x - k_{85}).s , \tag{10}$$

where k_{85} and s are parameters and x is age.

The validity of this model is based on two assumptions. The first is that mortality rates around the age of 85 years must be as reliable and thus we can determine the parameter k_{85} directly from empirical data. The second assumption concerns the mortality rates in the age that is considered to be the highest attainable. Coale and Kisker this age set at 110 years. Determining the rate of mortality in the 110 years of it to estimate the parameters Coale and Kisker set the value of mortality at the age of 110 years to 1.0 for men and 0.8 for women. Both values based on developement of mortality in Sweden (*Boleslawski & Tabeau*, 2001).

The resulting model then essentially corresponds to the (*Boleslawski & Tabeau*, 2001):

$$m_x = e^{a.x^2 + b.x + c} , \tag{11}$$

where a, b a c are parameters.

Denuit – Goderniaux

This model is based on the alignment table function probability of death q_x . Probability of death was calculated based on the knowledge m_x using the relationship $q_x = 1 - e^{-m_x}$.

Model (*Burcin et al.*, 2010): $\ln \hat{q}_x = a + b.x + c.x^2 + \varepsilon_x$, (12)

where $\hat{q}_x$ is the estimate q_x *a*, *b*, *c* are parameters and ε_x is random component.

The model is suitable for ages over 75 years and it is based on two assumptions. The first is that the probability of death at the highest age considered the unit. The second assumption is that the derivative of probability of death is zero.

Polynomial function

The second degree polynomial function (*Boleslawski & Tabeau*, 2001):

$$m_x = a + b.x + c.x^2 , \tag{13}$$

The third degree polynomial function: $m_x = a + b.x + c.x^2 + d.x^3$, (14)

where *x* is age and *a*, *b*, *c*, *d* are parameters.

Weibull

Weibull distribution, in addition to estimating the viability of machines used for modeling mortality. Model had the following form (*Boleslawski & Tabeau*, 2001):

$$m_x = b.x^a , \tag{15}$$

where *x* is age and *a*, *b* are parameters.

Models can be notionally divided into 3 groups. The first group are the models on which the estimated probability of death reach values very close to 1 at relatively low ages (from about 105 to the 115 years). The second group includes models, which also leads to values of probability of deaths very close 1 (1 to the likelihood of reaching the very high ages). The third group includes models for which the likelihood of death only reach values around 0.6, even for high ages (ie ages over 120 years). For the first two groups is assumed unlimited growth flattened values of mortality rates with age (only difference is the speed of increasing at the highest ages). In the third group the unlimited growth is not expected, but assumes a limit approximation to a finite value. The first group includes Coale - Kisker or Gompertz - Makeham model. The second group includes Weibull or modified Gompertz - Makeham model. In the third (the most optimistic) group we find Thatcher or Kannistö model (*Burcin et* *al*.,2010).

It is important to realize that the using of each model influences the level of life expectancy. The lowest life expectancy we get from using the Gompertz - Makeham model. Higher life expectancy is obtained from using logistic models.

Currently, the most commonly used logistic models. The most commonly used model is Kannistö model (also used for leveling of mortality rates in the framework of the Human Mortality Database). Though using logistic models get a higher life expectancy (these models are rather optimistic). On the other hand among pessimistic models belongs eg. Gompertz – Makeham model.

For most models, to estimate the parameters is used the method of least squares (except modified Gompertz - Makeham model, here it is necessary to use the method of nonlinear least squares).

1.1. Application of selected models to the data on mortality Czech population

For practical application, I selected a few of the previously mentioned models: Coale - Kisker, Gompertz - Makeham, modified Gompertz - Makeham, Heligman – Pollard, Kannistö and Thatcher model. The results obtained from using different models will be compared with the results obtained from using the methodology of the Czech Statistical Office (Czech Statistical Office, 2012) and the results of mortality tables without extrapolation. The calculation was based on data from 2009 (Eurostat, 2012).

Tab. 1a: Life expectancy at selected ages after applying selected models for balancing and extrapolating mortality curves

Model	Life expectancy at the exact age - Czech Republic - Men									
	0	15	20	50	65	80	85	90	95	100
Coale - Kisker	**74,2**	59,5	54,7	26,5	15,1	6,6	4,6	3,0	1,8	1,1
Gompertz	**74,2**	59,6	54,7	26,5	15,2	7,1	5,2	3,7	2,6	1,8
Gompertz-Makeham	**74,2**	59,6	54,7	26,5	15,2	6,7	4,7	3,2	2,1	1,4
Heligman-Pollard	**74,2**	59,6	54,7	26,6	15,2	7,2	5,4	3,9	2,9	2,1
Kannistö	**74,4**	59,8	54,9	26,8	15,4	7,2	5,5	4,1	3,1	2,4
Thatcher	**74,2**	59,6	54,8	26,6	15,3	6,8	4,9	3,5	2,6	1,9
Life tables accorging to the Czech Statistical Office methodology	**74,2**	59,6	54,7	26,6	15,3	6,8	4,8	3,4	2,4	1,7
Life tables without extrapolation	**75,0**	60,6	55,7	27,4	15,8	7,2	5,1	3,7	3,3	6,5

Source: http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database

From the table of life expectancy for men in Czech Republic is evident that the lowest life expectancy, is obtained from using the Coale - Kisker model (this applies to the entire age range). On the contrary, the highest values are achieved in Kannistö model. Life expectancy obtained from using the other models are somewhere between these two models. The results obtained from using the Gompertz - Makeham or modified Gompertz - Makeham model are initially very close to those obtained from applying Coale - Kisker model. The situation is changing in 80 years, when the Gompertz - Makeham model is closer to values obtained by Thatcher model. In contrast, the modified Gompertz - Makeham model is still the closest to Coale - Kisker model. If we compare the results obtained from using the Czech Statistical Office methodology, we find out that this methodology provides a somewhat lower life expectancy (this applies especially for lower ages).

If we look at the life tables without extrapolation we find out that we have received a higher life expectancy (compared to the other models). This applies until the age of 85 years, when the higher values of life expectancy are emerging eg Kannistö model. The situation is changed in the age of 95 years when life expectancy derived from mortality tables without extrapolation are again the highest.

Tab. 1b: **Life expectancy at selected ages after applying selected models for balancing and extrapolating mortality curves**

Model	Life expectancy at the exact age - Czech Republic - Women									
	0	15	20	50	65	80	85	90	95	100
Coale - Kisker	**80,1**	65,4	60,5	31,4	18,3	7,5	4,9	2,9	1,5	0,8
Gompertz	**80,3**	65,6	60,7	31,6	18,5	8,0	5,5	3,7	2,3	1,4
Gompertz-Makeham	**80,2**	65,5	60,6	31,5	18,4	7,6	5,1	3,2	1,9	1,1
Heligman-Pollard	**80,3**	65,7	60,7	31,7	18,5	8,1	5,7	3,9	2,6	1,8
Kannistö	**80,4**	65,7	60,8	31,7	18,6	8,2	5,9	4,1	2,9	2,1
Thatcher	**80,3**	65,6	60,7	31,6	18,5	7,8	5,3	3,5	2,4	1,8
Mortality table according to the Czech Statistical Office methodology	**80,3**	65,7	60,7	31,7	18,6	7,8	5,4	3,7	2,5	1,7
Life tables without extrapolation	**81,2**	66,7	61,8	32,6	19,4	8,5	5,9	4,2	3,5	6,1

Source: http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database

From the table of life expectancy for women is evident the same fact as for czech men. This means that the lowest life expectancy is obtained from using Coale - Kisker model (this applies to the entire age range). Similarly, the maximum value is obtained at Kannistö model. Life expectancy according to the other models are somewhere between these two models. The results obtained from using the modified Gompertz - Makeham function is initially very close to those obtained from applying Coale - Kisker model. Change starts at the age of about 90, when the results obtained from the modified Gompertz - Makeham are more different. If we use the Gompertz - Makeham function, the situation is somewhat different. The early results are essentially identical to the life expectancy according to the Thatcher model. Approximately et the age of 80 the situation is changing. Here Gompertz - Makeham model provides a higher life expectancy. According to the Czech Statistical Office methodology we obtain a rather high life expectancy (this is true under the age of 65 years). From 85 years, these values are average among the selected models.

When we compare the results obtained from mortality tables without extrapolation, we can see that the values from mortality tables without extrapolation are higher (this applies to the entire age range). A significant difference in the results obtained can be observed at the age of 100 years. Here, the value of life expectancy from life tables without extrapolation is significantly different from other values .

Conclusion

From the obtained results it can be concluded that some models give us very optimistic results and the others are very pessimistic. Among the most optimistic models can be concluded Kannistö model that provides the highest life expectancy (among selected models). In contrast, the pessimistic models can assign Coale - Kisker, and then also the modified Gompertz - Makeham function. Both of these models provide a lower life expectancy. Opinions on which of the selected models is the best, will probably change in the future. It has to do with further improvements in the development of mortality as well as providing a higher quality of statistical data.

Nowadays demographers give the priority to logistic models (Gavrilov & Gavrilova, 2011). Among the most widely used models belongs Kannistö model, which is used for leveling of mortality curve in the Human Mortality Database.

References

[1] Boleslawski, Lech, Tabeau, Ewa 2001. *Comparing Theoretical Age Patterns of Mortality Beyond the Age of 80*. In: Tabeau, Eva, van den Berg Jeths, A. a

Heathcote, Ch. (eds.) 2001. *Forecasting Mortality in Developed Countries: Insights from a Statistical, Demographic and Epidemiological Perspective*. s. 127 – 155. ISBN 978-0-7923-6833-5.

[2] Burcin, Boris, Tesárková, Klára a Šídlo, Luděk: "Nejpoužívanější metody vyrovnávání a extrapolace křivky úmrtnosti a jejich aplikace na českou populaci.“ *Demografie* 52, 2010: 77 – 89.

[3] ČSÚ 2012. 19. 1. 2012.
<http://www.czso.cz/csu/redakce.nsf/i/umrtnostni_tabulky_metodika>

[4] Eurostat. 4. 11. 2012.
<http://ec.europa.eu/eurostat>

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

[6] Gavrilov, Leonid A., Gavrilova, Natalia S.: "Stárnutí a dlouhověkost: Zákony a prognózy úmrtnosti pro stárnoucí populace.“ *Demografie* 53, 2011: 109 – 128.

[7] Human Mortality Database. 23. 8. 2012.
<www.mortality.org>

[8] Koschin, Felix: "Jak vysoká je intenzita úmrtnosti na konci lidského života?“ *Demografie* 41 (2), 1999: 105 – 109.

[9] Pavlík, Zdeněk, Kalibová, Květa: *Monohojazyčný demografický slovník*. Praha: Česká demografická společnost, 2005

[10] Thatcher, Roger A., Kanistö, Väinö a Vaupel, James W. 1998.: *The Force of Mortality at Ages 80 to 120*. 1998. ISBN 87-7838-381-1.

Contact

Petra Dotlačilová

Vysoká škola ekonomická v Praze

Nám. W. Churchilla 4

130 67 Praha 3

dotlacilova.petra@gmail.com