Modelling mortality rate of the very old

Ján Gogola[1]

Abstract

One of the factors of the economic and financial prosperity of the country is the age structure of the population. The bases for the age population structure calculation are the life tables. The indicator specific mortality rate is usually calculated on the base of the known data concerning the deaths and exposures to risk. Future mortality of the very old presents additional challenge since data quality can be poor at such ages. The greatest problem is lack of data at these age groups for the statistical processing. We consider stochastic mortality – Cairns, Blake and Dowd model, which is particularly well suited at very high ages. Characteristics specific mortality rate was calculated for the age categories 85-110. Results for selected countries (Czech Republic, Slovakia, Poland, Hungary, Austria and Japan) were found and compared.

Keywords: *Gompertz law, mortality rate, stochastic model, smoothing, very high age*

JEL Classification: C22, J11

1. Introduction

In advanced population recently there has been a gradual extension of the life expectancy (Finkelstein, 2005). In the Czech Republic life expectancy currently hovers around 78 years (75 for males/81 for females) (Andel, 2014). It is lower than most of Western countries. In the paper (Langhamrová et al., 2011) there are analysed trends in life expectancy in selected EU countries since 1950 with special attention to the Czech Republic. Another study (Boncz et al.) analyses change in the health status of the population in the central and Eastern Europe (CEE) for years 1990-2010. They showed that life expectancy at birth for males improved in CEE by 4,8 years and in EU-15 by 5,4 years.

Our paper follows on article Jindrová and Slavíček (2012). They deal with the development and the prediction of life expectancy in selected European countries (Czech Republic, Slovakia, Finland, Spain) by applying Lee-Carter model.

We use data on male and female deaths and exposures from the Human Mortality Database (www.mortality.org). The data are available by country for (a) single ages from age 0 to age 110+, and (b) single years (c) males, females and the whole population. The periods for which data are available vary by country. We will report results for six countries (Poland, Czech

[1] Corresponding author: University of Pardubice, Faculty of Economics and Administration, Institute of Mathematics and Quantitative methods, Studentská 84, 532 10 Pardubice, Czech Republic, e-mail: jan.gogola@upce.cz.

Republic, Slovakia, Hungary, Austria, Japan). We consider the restricted age range 40 to 110, the range of greatest interest to providers of pensions and annuities.

Let calendar year t run from exact time t to exact time t+1 and let $d_{x,t}$ be the number of deaths aged x last birthday in calendar year t. We suppose that the data on deaths are arranged in a matrix $\boldsymbol{D} = (d_{x,t})$. In a similar way, the data on exposure are arranged in a matrix $\boldsymbol{A} = (a_{x,t})$ where $a_{x,t}$ is a measure of the average population size aged x last birthday in calendar year t, the so-called *central exposed to risk*. We approximate the initial exposed to risk $e_{x,t} = a_{x,t} + \frac{1}{2} d_{x,t}$ and let $\boldsymbol{E} = (e_{x,t})$ denote the matrix of *initial exposures*. We suppose that $\boldsymbol{D} = (d_{x,t})$ and $\boldsymbol{E} = (e_{x,t})$ are each $n_a \times n_y$ so that we have n_a ages and n_y years.

We denote the *mortality rate* at exact time t for lives with exact age x by $q_{x,t}$. The mortality rate expresses the probability that a life aged x at exact time t dies in the following year, i.e. between t and t +1.

If we treat $d_{x,t}$ as a random variable, $D_{x,t}$, and initial exposure $e_{x,t}$, as fixed, then $D_{x,t}$ has a binomial distribution:

$$D_{x,t} \sim Bi(e_{x,t}; q_{x,t}), \tag{1}$$

which leads us to maximum likelihood estimates of $q_{x,t}$ as:

$$\hat{q}_{x,t} = \frac{d_{x,t}}{e_{x,t}}. \tag{2}$$

A typical life table provides one-year probabilities of death q_x according to age x, of individuals. Crude mortality rates can be derived directly from death and exposure data at each age. However, mortality rates are often averaged over time to smooth the year-to-year variation in crude mortality rates.

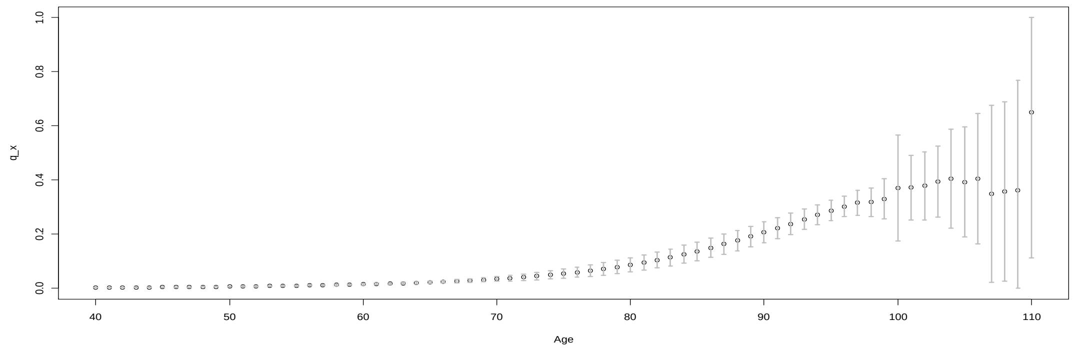

Fig. 1. Crude mortality rate with 95% confidence intervals for population of Poland.

It is generally accepted that immediate post-natal mortality and the “accident hump” aside, mortality rates increase with age.

Let
$$\hat{q}_x = \frac{\sum_t d_{x,t}}{\sum_t e_{x,t}}, \tag{3}$$
be a measure of mortality by age *x* averaged over years *t*.

Figure 1 shows crude mortality rates $\hat{q}_x$ with the corresponding 95% confidence interval for the restricted to ages 40-110 for population of Poland. We can see how the intervals widen steeply for high ages (higher than 100 years).

The modelling of mortality at high ages presents additional challenge to the actuary since data quality can be poor at such ages.

In this paper we firstly apply *generalized linear model* (*GLM*). GLM for individual mortality is typically based around the Bernoulli model, using $\hat{q}_x$ and *logistic* (or *logit*) transformation, which has been widely used in actuarial practice. The logistic mortality law can be written as:
$$\log\left(\frac{q_x}{1-q_x}\right) = \alpha + \beta \cdot x, \tag{4}$$
which is an expression of a simple GLM, where age is the only covariate, and the linear predictor is a linear combination of age and a constant.

One of the benefits of using a logistic link in the model relates to the ease of interpretation. The transformation $\frac{q_x}{1-q_x}$ turns a probability into the equivalent odds in favour of the event (in this case, death). Thus, the odds that an older individual dies within one year increase over that of a younger individual with each year of age by e^{β}.

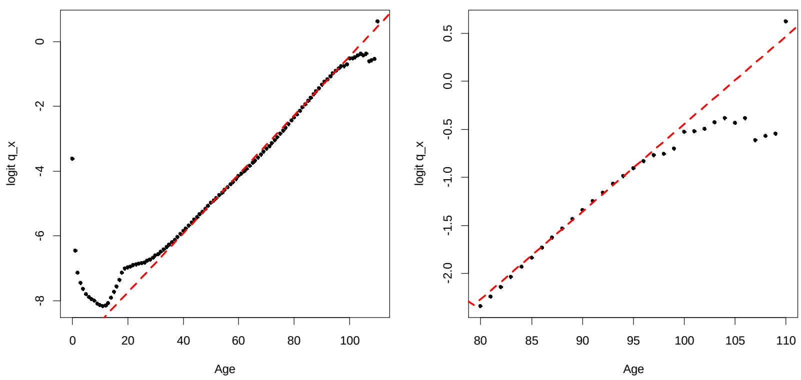

Fig. 2. Logit mortality rates by age for Poland averaged over years; observed (points •) and fitted by GLM (dashed line).

The scatter plot in Figure 2 shows observed logit mortality rates averaged over time. A striking feature of this plot is the fall in mortality rate for ages around 105 years. We do not believe mortality rates fall at high ages, so there must be a problem with data.

However actuaries still need to produce forecast at high ages since the calculation of lifetime annuities and pensions requires mortality rates up to say 105 or 110 years, well above the age at which good-quality data are typically available.

The solution of this problem we see in modelling mortality up to an age where the data are credible. Therefore we apply a model with extrapolation for higher ages built into its structure. For this purpose we use Cairns-Dowd-Blake model (we will refer to it as CBD model):

$$\operatorname{logit} q_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)} \cdot (x - \bar{x}) \tag{5}$$

where $\bar{x}$ is the mean age in the sample range.

2. Results

CBD model is a stochastic model which is designed for modelling longevity risk in pensions and annuities. It exploits the relative simplicity of the mortality curve (i.e. near log linearity). CBD model is built on observation that logit of mortality rates are approximately linear for adults (at age above 40).

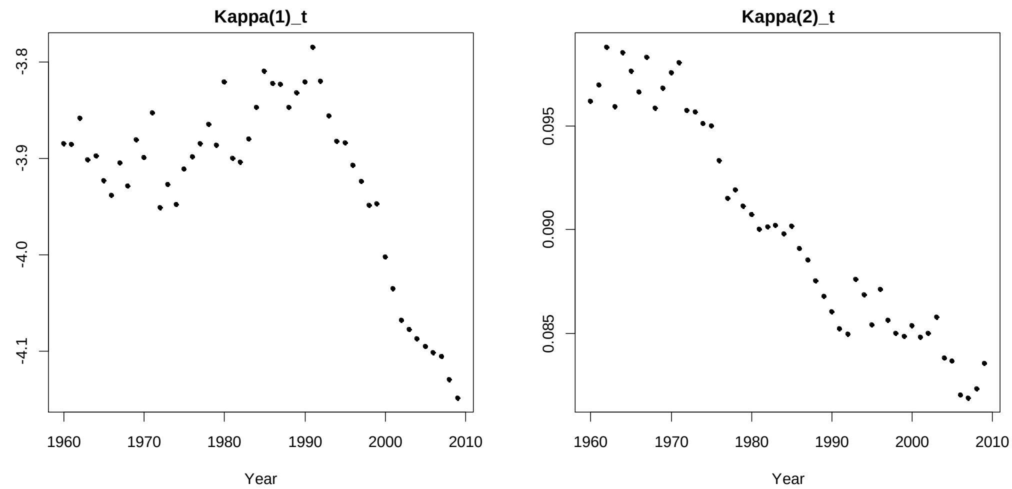

Fig. 3. Values of $\kappa_t^{(1)}$ (left) and $\kappa_t^{(2)}$ (right) fitted to population of Poland.

We can think of CBD model as a collection of Gompertz models (Lee et al.), one model for each year. The coefficients, the intercepts $\kappa_t^{(1)}$ and slopes $\kappa_t^{(2)}$ in the CBD model, with data of Poland restricted to ages 40-85 and years 1960-2009, are shown in Figure 3.

On the left-hand side of Figure 3. we can see how the general level of mortality has fallen the last 25 years, while the right-hand side plot shows the rate of change (slope) of mortality.

We can see why CBD model is so well suited to predict in the age direction. The structure of CBD model allows us to extend the forecast in the age direction. So first we fit CBD model with data in which we have confidence (ages 40-85). Then with the fitted values of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ in place, expression (5) provides fitted and forecasted mortalities for suspected ages 85-110. Extrapolation in age is straightforward with the CBD model, since the age function is a straight line. This extrapolation does not depend in any way on data above age 85. The extrapolation follows the curve established with the data up to age 85.

In Figure 4 we can see that the fitted line follows quite well up to around age 80, the model assumption under-rides the data above age 80. It suggests that CBD model underestimates the mortality rate at high ages.

Conclusion

Figure 3 suggests that it should be possible to forecast $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$. Cairns et al. (2009) used a bivariate random walk with drift for forecasting. With the forecasted values $\hat{\kappa}_t^{(1)}$ and $\hat{\kappa}_t^{(2)}$ in place, expression (5) yields a forecast for the whole life table.

Extrapolative projection models are the most common of the models used for forecasting mortality, but such models will only be reliable if past trends continue. The medical advances can cancel extrapolative projections by changing the trend.

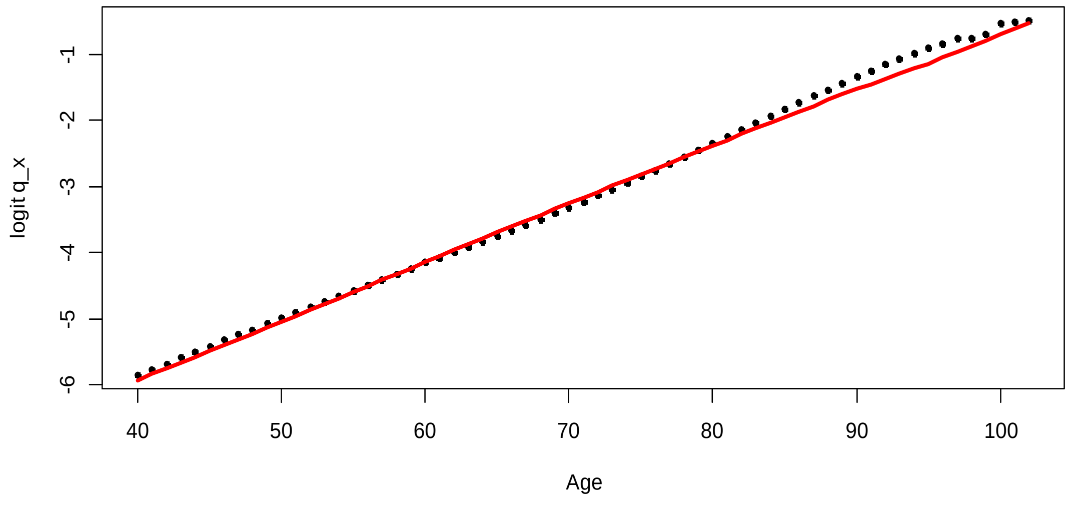

Fig. 4. Logit of mortality rate by age averaged over year for population of Poland: observed (points •) and fitted by CBD model (solid line).

We have applied the CBD model for different countries (Poland, Czech rep., Slovakia, Hungary, Austria, Japan). CBD model shows that mortality rates in all countries have the following features in common.

In the Appendix (Figures 5-9) we have plotted the maximum likelihood estimates for the parameters for each country. While in developed countries (Japan, Austria) the $\kappa_t^{(1)}$ has declining continuously whole the time (from 1960 to 2009), in former socialistic countries we can see declining only over the last 20-25 years. Changes in political system are likely to influence the improvements in population health, as indicated by life expectancy and mortality rates. The changes over time in $\kappa_t^{(1)}$ have been approximately linear. An increase in $\kappa_t^{(2)}$ over last 20 years in the case of former socialistic countries, that is an increase in the steepness of the logit-transformed mortality curve, means that mortality (in logit scale) at younger ages improves more rapidly than at older ages. With the rapid aging of the population, mortality forecasting becomes increasingly important, especially for the insurance and pension industries.

Acknowledgements

This research could be performed due to research grant "ROUTER" EE2.3.30.0058 Development of Research Teams at the University of Pardubice.

Model fitting was done in **R** statistical software, which was also used for all graphs.

References

Andel, R. (2014). Aging in the Czech Republic. *The Gerontologist*, *54*(6), 893-900.

Boncz, I., Vajda, R., Ágoston, I., Endrei, D., & Sebestyén, A. (2014). Changes in the health status of the population of Central and Eastern European countries between 1990 and 2010. *The European Journal of Health Economics*, *15*(1), 137-141.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., & Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal*, *13*(1), 1-35.

Finkelstein, M. S. (2005). Lifesaving explains mortality decline with time. *Mathematical biosciences*, *196*(2), 187-197.

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www. mortaliy.org.

Jindrová, P., & Slavíček, O. (2012). Life expectancy development and prediction for selected European countries. In: *6-th International Scientific Conference Managing and Modelling of Financial Risk Proceedings*, 303-312.

Langhamrová, J., Miskolczi, M., & Langhamrová, J. (2011). Life expectancy trends in CR and EU. *Intenational days of Statistics and Economics*, 298-306.

Lee, S. J., Boscardin, W. J., Kirby, K. A., & Covinsky, K. E. (2014). Individualizing Life Expectancy Estimates for Older Adults Using the Gompertz Law of Human Mortality. *PloS one*, *9*(9), e108540.

Appendix

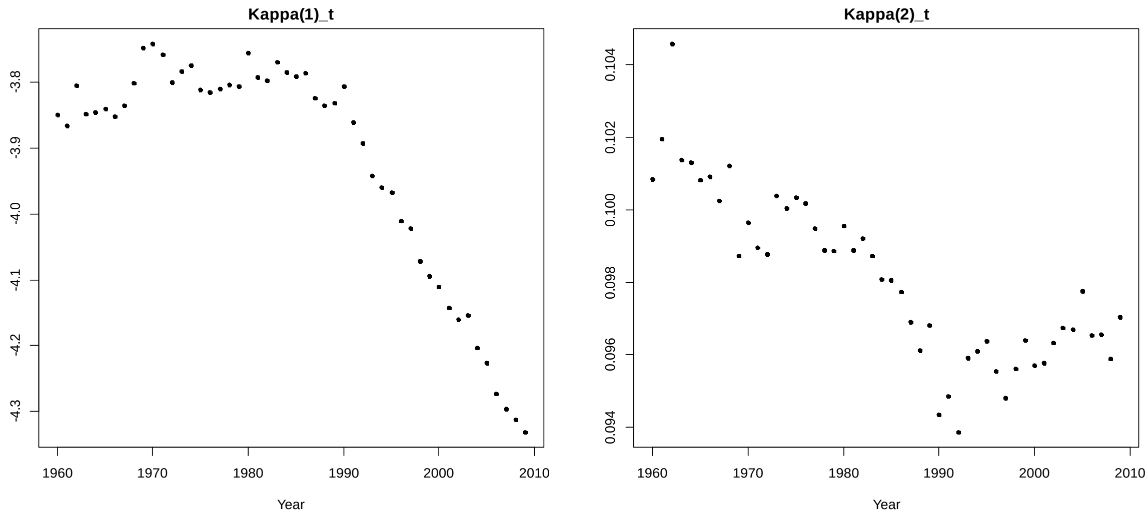

Fig. 5. Values of $\kappa_t^{(1)}$ (left) and $\kappa_t^{(2)}$ (right) fitted to population of the Czech Republic.

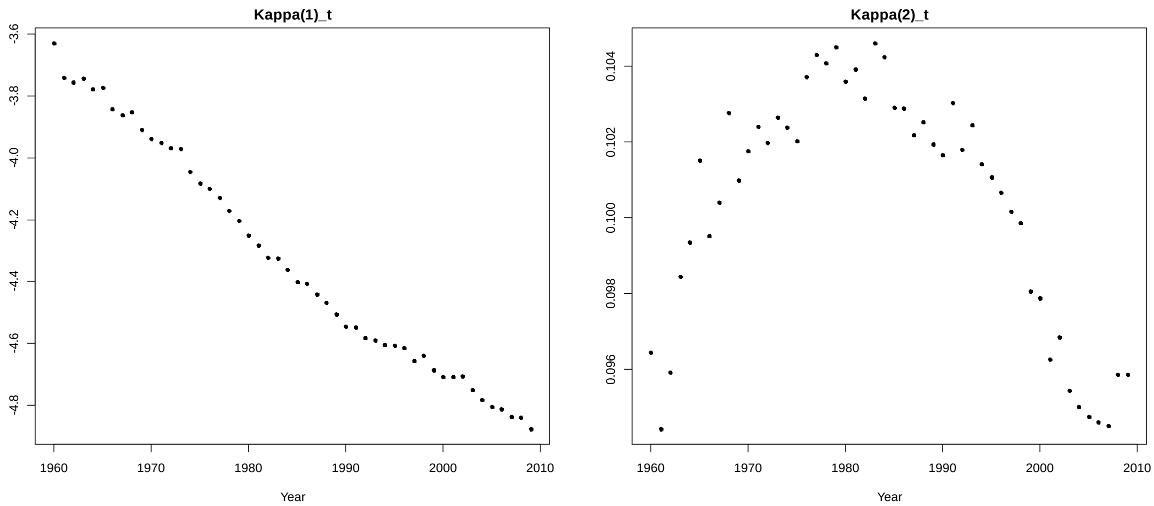

Fig. 6. Values of $\kappa_t^{(1)}$ (left) and $\kappa_t^{(2)}$ (right) fitted to population of Japan.

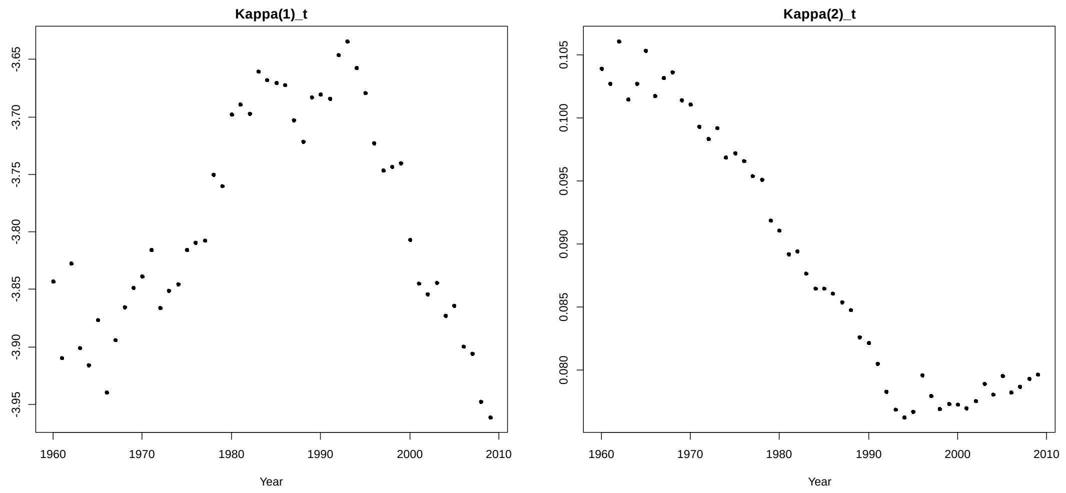

Fig. 7. Values of $\kappa_t^{(1)}$ (left) and $\kappa_t^{(2)}$ (right) fitted to population of Hungary.

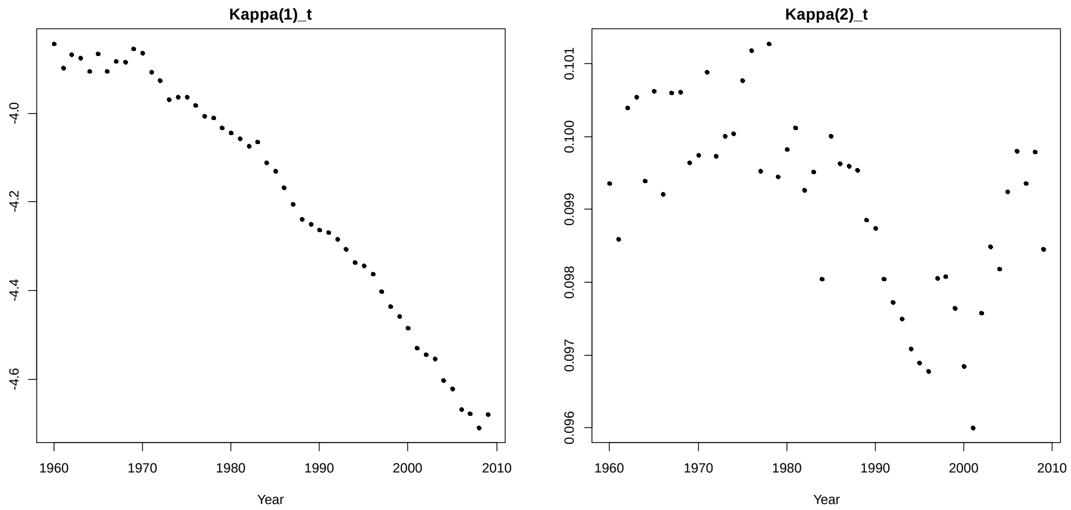

Fig. 8. Values of $\kappa_t^{(1)}$ (left) and $\kappa_t^{(2)}$ (right) fitted to population of Austria.

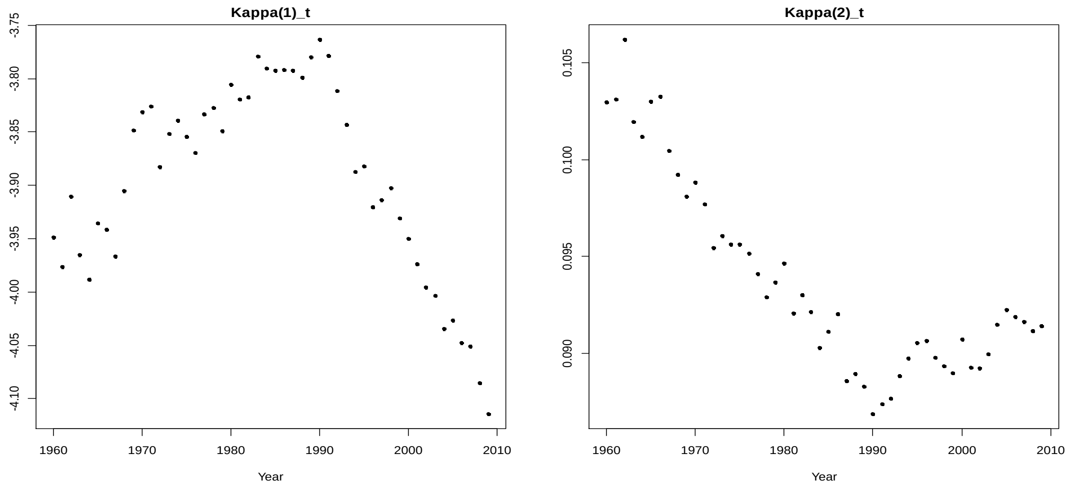

Fig. 9. Values of $\kappa_t^{(1)}$ (left) and $\kappa_t^{(2)}$ (right) fitted to population of Slovakia.

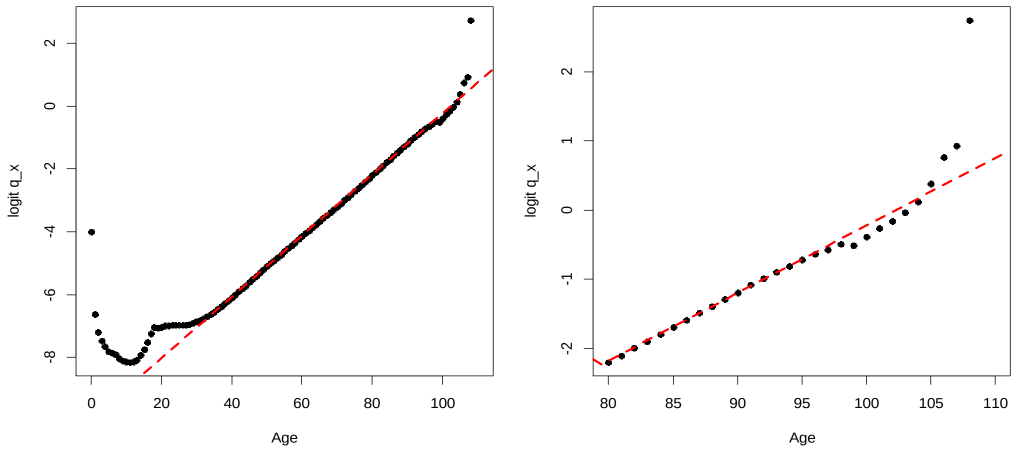

Fig. 10. Logit mortality rates by age for the Czech republic averaged over years; observed (points •) and fitted by GLM (dashed line).

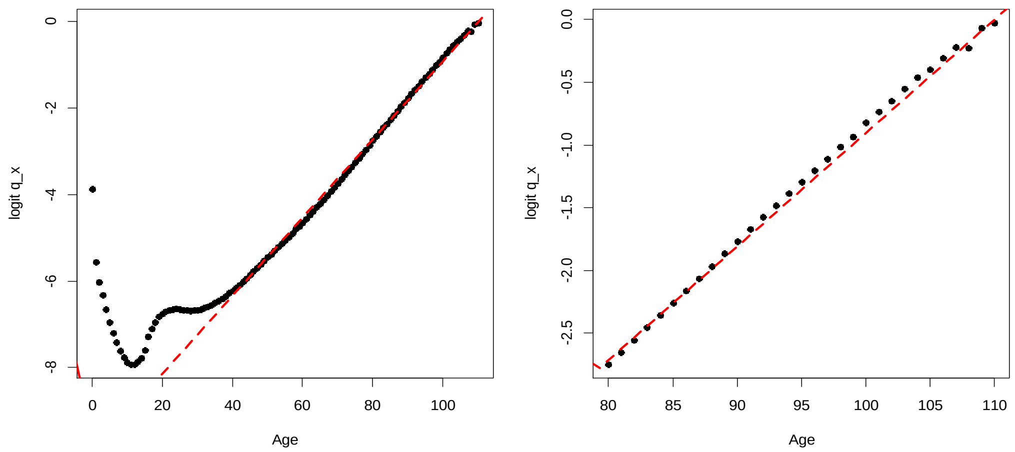

Fig. 11. Logit mortality rates by age for Japan averaged over years; observed (points •) and fitted by GLM (dashed line).

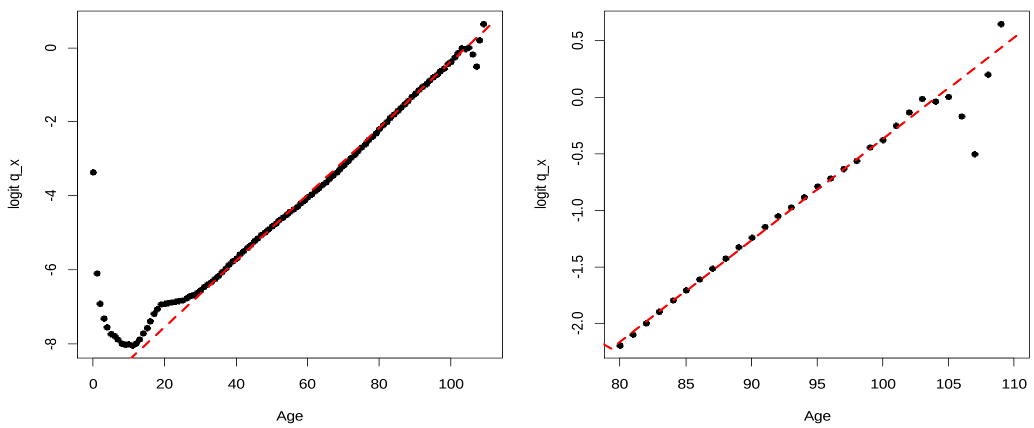

Fig. 12. Logit mortality rates by age for Hungary averaged over years; observed (points •) and fitted by GLM (dashed line).

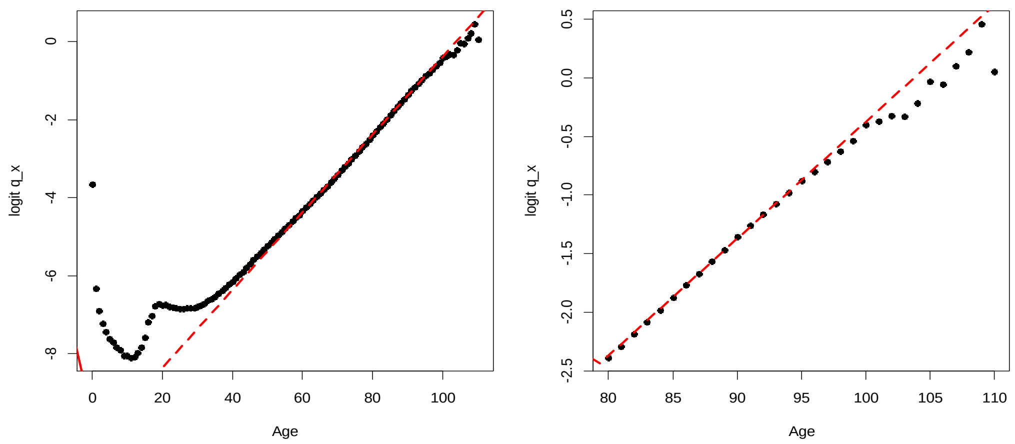

Fig. 13. Logit mortality rates by age for Austria averaged over years; observed (points •) and fitted by GLM (dashed line).

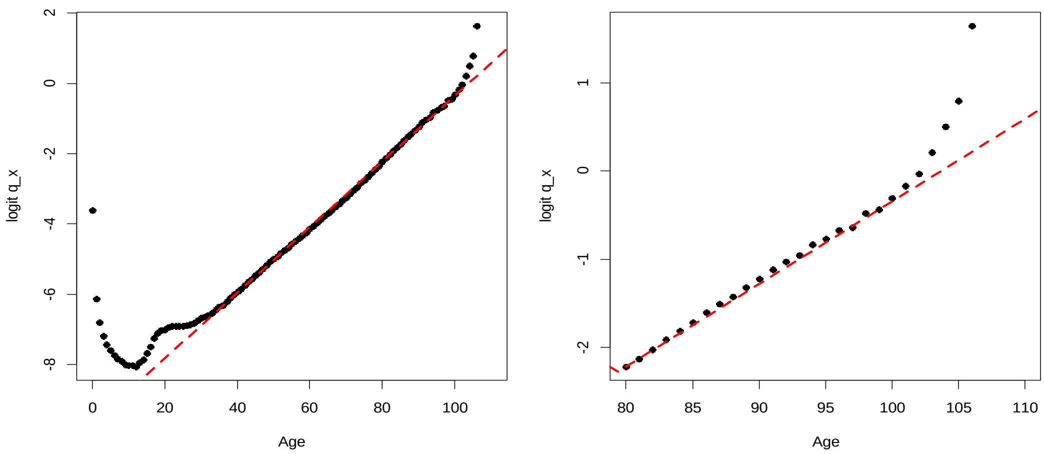

Fig. 14. Logit mortality rates by age for Slovakia averaged over years; observed (points •) and fitted by GLM (dashed line).