USING REGRESSION SPLINES FOR MORTALITY FUNCTION SMOOTHING

Vlastimil Farkašovský – Emília Zimková

Abstract

As Slovak population ages, the study of mortality patterns and future mortality forecasts and their impact on the social welfare have become an important research issue. In the contribution we propose regression splines to smooth the mortality function of the Slovak citizens. Specifically, we use B-splines and P-splines to solve the problem with variability of the highest age cohorts in situation when there exist insufficient number of historical observations. In the contribution, the Slovak historical mortality data clustered according to age and sex for period 1950-2012 are employed. The results reveal considerable improvement in comparison to the results of previously used techniques. Monotonous limitation ensured non-decreasing course of the mortality function, and despite of the low number of observations in the age group 80 and older the results the results do not lead to a decline in observed mortality rate. The gained research results are useful in further modelling of demographic development.

Key words: mortality function, smoothing, demographic development, B-splines, P-splines

JEL Code: C14, J11.

Introduction

The transition from a pension security system stipulated by the state to a system of pension insurance collectively referred to as the Slovak pension reform belongs to the most important social changes in the modern history of Slovakia. The main impulse for this change were projections of demographic development which due to the population aging pointed to the long-term fiscal unsustainability of the pension system. The Slovak Republic based on these forecasts would overcome one of the biggest demographic shocks among the countries of the European Union. The phenomenon of population aging thus represents one of the greatest challenges of the current research in the area of long-term stability and sustainability of the public finance. The study of mortality patterns and future mortality forecasts as an important chain of the demographic development is therefore essential.

The demographic development can be examined using several models forecasting future demographic trends. Given the importance of forecasting demographic development within these models is necessary to use as precise data characterizing the present demographic development as possible. However, in countries with smaller population, such as the Slovak Republic, given the lack of historical observations in higher age groups, the high variability of mortality occurs. This high volatility which do not reflect the real demographic development for these age groups and may distort further forecasts.

The ambition of this paper is therefore to bring the possibilities for smoothing of mortality function and application of the method of regression splines on the dataset of the Slovak Republic. These smoothed mortality functions can be then used in the models of demographic development since due to lower variability they provide better forecasts of future development. The paper is structured into 4 sections, the first of which is introductory and the last section is a conclusion. The introductory part contains a brief description of a motivation to smoothe mortality function in Slovakia by regression splines. The second part of the paper is devoted to the methodology. The third section brings results and discussion. Finally, the paper includes concluding remarks.

1 Methodology: smoothing mortality function

In this part of the paper we will present the possibilities for smoothing of mortality function. The paper draws data on the number of deaths and the number of people living in the Slovak Republic from the database of Stanford University (Human Mortality Database) since 1950 till 2012. Software R was used for smoothing.

The development of mortality is as we have already mentioned one of the main factors in predicting future demographic development of the population. Since the development of mortality reveals especially in the last age cohorts, due to insufficient number of observations, high degree of variability, we will show in this part the possibilities for mortality function smoothing.

Smoothing creates approximation function, which preserves important patterns of the dataset and at the same time eliminates the noise caused by, for example, lower number of observations in the higher age groups.

Overview of the models of mortality function estimation in higher age groups is shown in the following table:

Tab. 1: An overview of models of mortality function estimation in age group >85 years

Model name	Model	Type of model	Model parameters
Gompertz-Makeham function	$\mu_x = a + b.c^x$	exponential	a, b, c
Heligman-Pollard model	$q_x = \frac{b.e^{a.x}}{1 + b.e^{a.x}}$	logistic	a, b
Thatcher model	$\mu_x = \gamma + \frac{z}{1+z}$ where $z = \alpha.e^{\beta.x}$	logistic	α, β, γ
Kannist model	$\mu_x = \frac{\alpha.e^{\beta.x}}{1 + \alpha.e^{\beta.x}}$	logistic	α, β
Coale-Kisker model	$m_x = e^{a.x^2 + b.x + c}$	linear	a, b, c

Source: Dotlačilová, Langhamrová (2014, p. 72)

Another option for smoothing of mortality function is the use of smoothing with the help of weighted correlation diagrams (scatterplots). This procedure can be found in the work of, for example, Cleveland (1979).

A prominent method used for smoothing of mortality function gradually became smoothing methods using regression splines. Such approach can be found in the work of Hyndman (2013), who states that his model among other things uses weighted approach and limitations of regression splines in predicting smoothing function $f_t(x)$.

Weighted approach treats possible heterogeneity and monotonous limitations[1] for higher age groups lead to better predictions.

Splines belong to the group of interpolation techniques while they fit each element of the dataset into smoothed continuous function consisting of a set of polynomials. When constructing the P-splines we use as a basis for regression B-splines and then we modify the

[1] Limitations providing at a given interval monotonous course of the function. This case concerns limitations, which ensure that the function at a given interval (for age groups above 65 years) is non-decreasing since with increasing age the likelihood of death should not decrease.

value of log-likelihood by penalizing the regression coefficient with the aim to maximize them. B-spline is defined recursively:

$$B_i^k(x) = \left(\frac{x - t_i}{t_{i+k} - t_i}\right) B_i^{k-1}(x) + \left(\frac{t_{i+k+1} - x}{t_{i+k+1} - t_{i+1}}\right) B_{i+1}^{k-1}(x)\,, \tag{1}$$

Where B stands for B-spline, x is a smoothed variable (log mortality), k represents non-negative integer, t denotes vector (non-decreasing sequence of real numbers, where t ≥ k+2) and i represents a spline order.

As we can see from the equation, in order to construct B-splines are crucial nodes, where holds ($x=t_j$). The number of internal nodes is equal to the degree of polynomial of B-spline.

Subsequently, to construct P-splines we use Eilers-Marx (1996) penalization based on finite differentials. It is based on the original idea of O´Sullivana (1986) and its functional form can be written as follows:

$$S(x) = \sum_{i=1}^{m} \left(y_i - \sum_{j=1}^{n} \hat{a}_j B_j(x) \right)^2 + \lambda \sum_{j=k+1}^{n} \left(\Delta^k \hat{a}_j \right)^2 \tag{2}$$

Where λ is a smoothing parameter, S(x) represents a spline function, and $\hat{a}$ is a prediction of parameter α.

As stated by Becker et al. (2009) penalization of B-splines in this way ensures that the coefficient of individual adjacent splines differs minimally.

There are several advantages of the use of P-splines for smoothing of mortality function. Bourbeau and Ouellette (2011) state in particular the fact that P-splines contain good features of many other smoothing techniques mentioned above. Unlike other smoothing techniques P-splines provide quality smoothing even for final values since they do not contain any effects at the end of a time series, which ensures quality smoothing even for higher age groups, whose problems were mentioned in previous text. Also, they work well in case the data misses some values. P-splines also retain moments (averages and variances).

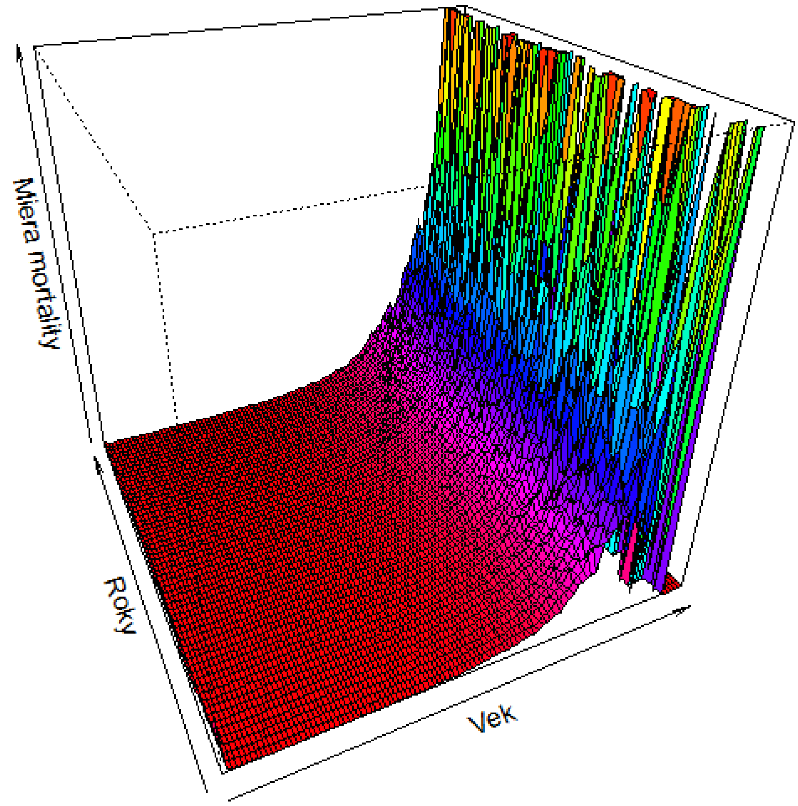
2 Empirical results and discussion

Mortality tables constructed with the help of smoothing function thus thanks to lower variability provide better predictions of future mortality.

Figure 1 depicts real observed mortality rate in the Slovak Republic since 1950 till 2012. Individual figures show exponential course of the mortality function throughout the years and high variability at higher age groups, where we often record lack of observations, or more precisely, their absence. This mainly concerns age group 90 and older.

The three-dimensional figure 1 depicts the development of mortality rate in Slovakia for age cohorts 0 to 110 year old men, for the period 1950-2012. When examining particular age cohort, for example, 25 years old men for period 1950-2012, we observe reduction of mortality rate due to improving health care in Slovakia and other factors which may include increasing education, improved environment and so forth. At the same time, with increasing age of statistical unit, the rate of mortality naturally increases. In figure 1 we can observe that the mortality rate increases exponentially from the 50+ age cohorts, but figure 1 does not depict appropriately the mortality rate in the most monitored, i.e. economically active age groups. Figure 1 also confirms high variability of data on population mortality over the age of 90.

Fig. 1: Mortality rate in the Slovak Republic for period 1950 - 2012 – men aged 0-110



Source: the authors.

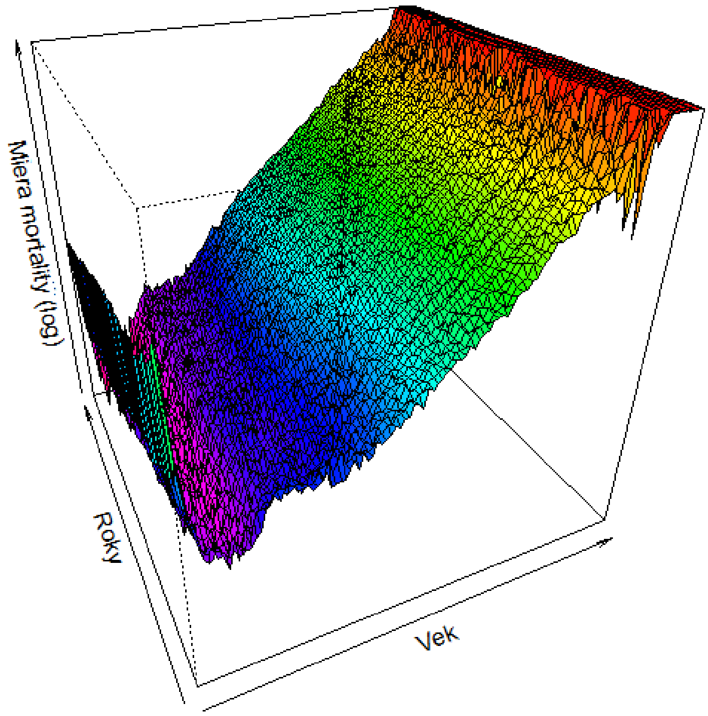
Figure 2 depicts log-transformed data that give us ability to better observe and analyze development of mortality rate in the Slovak Republic since 1950 till 2012 for key age groups. Logarithm shape of mortality function is also used across the spectrum of smoothing and prediction methods of other authors such as Lee-Carter (1993), Hyndman, Ullah (2007), Hyndman (2013), Hatzopoulos, Haberman (2011), Cairns, Black, Dowd et al. (2010) and many others.

In the figure we can clearly observe high rate of infant mortality[2] especially among the oldest observations around year 1950 and a relatively sharp decline towards year 2012. Child mortality[3] unlike infant mortality reaches significantly lower values, however, also in the observed period it considerably decreases. In male population significantly increases mortality rate in age cohorts 15-22 years. Conversely, mortality from about the fortieth year stays at stable values.

[2] Infant mortality = mortality in the first year of age
[3] Child mortality = mortality between the first and fourth year of age

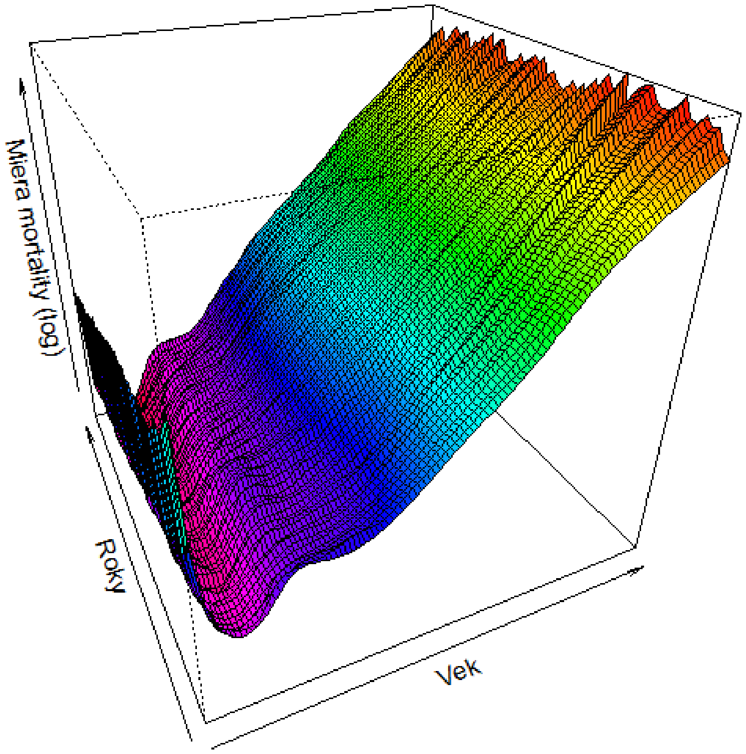
The visualization of log-transformed data allowed to observe high variability of mortality development of age cohorts, while with increasing age, the number of observed statistical units decreases and thus variability increases.

Fig. 2: Logarithmic mortality rate in the Slovak Republic for period 1950 - 2012 – men aged 0-110



Source: the authors.

Fig. 3: Smoothed logarithmic mortality rate in the Slovak Republic for period 1950 - 2012 – men aged 0-110



Source: the authors.

In figure 3 we observe smoothed course of the mortality function of men in the Slovak Republic, while all important data of the dataset about mortality of infants, children, young men, middle-aged men and those aged over 65 years were retained and at the same time noise caused by lower number of observations in higher age groups, more precisely, their absence, was eliminated. Monotonous limitation ensured non-decreasing course of the mortality function, and despite of the low number of observations in the age group 80 and older the results do not lead to a decline in observed mortality rate.

Conclusion

The aim of the paper was to smooth mortality function in the Slovak Republic for one-year age cohorts of male population of the Slovak Republic between 1950-2012 using the latest statistical techniques in the form of regression splines, which compared to previously used techniques better treat terminal data of the statistical series. In addition, the aim of the paper was to find a technique, which would ensure non-decreasing course of the mortality function with increasing age of the statistical units. The paper used untreated as well as log-transformed dataset and proposed possibilities for mortality function smoothing using regression splines. In the last part we smoothed the mortality function for one-year age cohorts and period 1950-2012 in the Slovak Republic. For smoothing we used method of regression P-splines, which are based on B-splines, to which we applied monotonous limitation and penalization maximizing log-likelihood. The use of regression splines compared to previous methods of smoothing appropriately treats smoothing of terminal data of time series, despite the fact that statistical units gradually decrease and the variability of time series increases. Our research on men in the Slovak Republic confirmed increased mortality of infants, men aged 15-22 years and stabilization of growth course of mortality function for men over 40 years. Monotonous limitation ensured non-decreasing course of the mortality function, and despite of the low number of observations in the age group 80 and older the results do not lead to a decline in observed mortality rate.

The research results are useful in further modelling of demographic development, while demographic models constructed with smoothed data can be further applied also for examining long-term sustainability and stability of the pension system for example in the form of variant demographic scenarios to build the model of overlapping generations (OLG – Overlapping Generations Model), or other models of general equilibrium (DSGE Models – Dynamic Stochastic General Equilibrium Models).

Acknowledgment

The support of the grant scheme VEGA 1/0859/16 Dynamics of nonlinear economic processes is gladly acknowledged.

References

BECHER, et al. 2009. Using penalized splines to model age- and season-of-birth-dependent effects of childhood mortality risk factors in rural Burkina Faso. In: *Biometrical Journal*. ISSN 1521-4036 p. 110-122

BOURBEAU, R., & OUELLETTE, N. 2011. Changes in the age at death distribution in four low mortality countries: A nonparametric approach. In: *Demographic research*. ISSN 1435-9871 p. 595-628

CLEVELAND, W.S. 1979. Robust locally weighted regression and smoothing scatterr plots. In *Journal of the American Statistical Association.* ISSN 0162-1459 p. 829-836

DOWD, K., CAIRNS, A.J.G., & BLAKE, D. et al. 2010. Evaluating the goodness of fit of stochastic mortality models. In: *Insurance: Mathematics and Economics*. ISSN 0167-6687 p. 255-265

DOTLAČILOVÁ, P., & LANGHAMROVÁ, J. 2014. The influence of mortality models for the expected future life-time of older people. In *Applications of mathematics and statistics in economy*, Wroclaw : University of Economics in Wroclaw, Vol. 17, 2014. ISBN 978-83-7695-421-9 p. 65-74

EILERS, P.H.C., & MARX, B.D. 1996. Flexible smoothing with B-splines and penalties. In: *Statistical Science*. Beachwood, OH, USA: Institute of Mathematical Statistics, Vol. 10 ISSN 0883-4237 p. 89-121

HATZOPOULOS, P., & HABERMAN, S. 2011. A dynamic parameterization modelling for the age-period-cohort mortality In: *Insurance: Mathematics and Economics*, 49 (2), ISSN 0167-6687 pp 155-174.

HYNDMAN, R. J. 2013. Prospective life tables. In CHARPENTIER, A. et al. 2014. *Computational Actuarial Science with R*. Boca Raton : Chapman & Hall/CRC Press, 2014. ISBN 978-1-4665-9259-9 p. 319-343

HYNDMAN, R. J., & ULLAH, S. 2007. Robust forecasting of mortality and fertility rates: A functional data approach. In *Computational Statistics and Data Analysis* Vol. 51, 2007. ISSN 0167-9473 p. 4942-4956

JURČOVÁ, D. 2005. *Slovník demografických pojmov*. [online]. Bratislava : Výskumné demografické centrum INFOSTAT, 2005. 72 p. [cit. 2014-08-04]. Available at: <http://www.infostat.sk/vdc/pdf/slovnik_2verdd.pdf>

LEE, R. D., & CARTER, L. R. 1992. Modeling and forecasting US mortality. In *Journal of American Statistical Association*, 1992, Vol. 87, p. 659-671. ISSN 0162-1459

O´SULLIVAN, F. 1986. A Statistical Perspective on III-posed Inverse Problems. In: *Statistical Science*. Beachwood, OH, USA: Institute of Mathematical Statistics, Vol. 1, No. 4 ISSN 0883-4237 p. 502-527

Prognóza obyvateľstva SR do roku 2050. [online]. 2002. Bratislava : Výskumné demografické centrum INFOSTAT, 2002. 121 p. [cit. 2013-02-22]. Available at: <http://www.infostat.sk/vdc/pdf/prognoza2050vdc2.pdf>

Prognóza obyvateľstva SR do roku 2060. [online]. 2012. Bratislava : Výskumné demografické centrum INFOSTAT, 2012. 74 p. [cit. 2013-02-22]. Available at: <http://www.infostat.sk/vdc/pdf/obyv452000.pdf>

Contact

Ing. Vlastimil Farkašovský
Matej Bel University in Banská Bystrica, Faculty of Economics
Tajovského 10, 975 90 Banská Bystrica, Slovakia
vlastimil.farkasovsky@umb.sk

doc. Ing. Emília Zimková, PhD.
Matej Bel University in Banská Bystrica, Faculty of Economics
Tajovského 10, 975 90 Banská Bystrica, Slovakia
emilia.zimkova@umb.sk