

Comparison of Estimation Uncertainty in Lifelong Annuities for Bayesian Single- and Multi-population Old Age Mortality Model¹

Pavel ZIMMERMANN – Jan FOJTÍK – Martin MATĚJKA – Tomáš KAREL*

Abstract

Old age mortality modelling is often associated with lack of reliable data, especially for small populations. We focus on an approach to incorporate information contained in the data from closed populations and study its impact on estimation uncertainty in an old age mortality model. We assume a two-dimensional age cohort extension of the exponential (Gompertz) model. We compare uncertainty of the parameter estimates for two models. The first is a single population model based on data solely from one country. The second is a multi-population model for a sample of populations from the central European region. Bayesian generalized linear model and a hierarchical Bayesian generalized linear model is applied. We quantify the difference in the uncertainty of the estimates of the force of mortality and whole life annuity based on root mean squared error of the predictions for different ages, cohorts and populations.

Keywords: *old age mortality, Bayesian GLM, Bayesian Hierarchical model, multi-population model*

JEL Classification: C10, C11, J10, J11

Introduction

Mortality modelling is a major concern for public policy, social security, and private insurance. Sufficient amount and quality of mortality data is absolutely critical for successful model development, its application, and evaluation. Old age mortality modelling is, however, very often associated with lack of data,

* Pavel ZIMMERMANN – Jan FOJTÍK – Martin MATĚJKA – Tomáš KAREL, University of Economics Prague, Faculty of Informatics and Statistics, Department of Statistics and Probability, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic; e-mail: zimmerp@vse.cz; xfojj00@vse.cz; xmatm00@vse.cz; xkart09@vse.cz

¹ The support of the grant IGS 6/2015 *Využití bayesovských metod pro modelování úmrtnosti* is gladly acknowledged. We would like to thank to Ivana Malá for her endless willingness to help.

incorrect or even missing values. This shortcoming often results in wide confidence intervals of predictions or even misleading predictions. The presented article suggests a method for incorporation of additional data to an old age mortality model and quantifies the impact of this additional data on estimation uncertainty measured by the root mean squared error.

In general, mortality is modelled separately for adult and old ages. A low number of observations for old ages prevents using models with age specific parameters, e.g. Lee and Carter (1992) or its modifications, which are generally popular for adult ages. Therefore, old age models are often formulated as one-dimensional models, so called ‘mortality laws’, with only a few parameters. Gompertz (1825) assumed exponential growth of mortality with increasing age. Makeham (1860) subsequently extended this model with a constant parameter which represented mortality caused by age independent external causes. Later on, with increasing amount of empirical observations in older ages, the exponential growth was challenged by some authors and a slower increase was suggested. Koschin (1999) extended Gompertz-Makeham model with an additional parameter in order to slow exponential growth of mortality. Some authors proposed logistic curve as an alternative. Historically, the first mortality model based on logistic curve assumption was introduced by Perks (1932). In the recent past, the logistic curve was applied by Thatcher (1999) or Heligman and Pollard (1980). A comprehensive summary of old age mortality models is provided by Pitacco, Denuit and Haberman (2009) or Burcin, Tesárková and Šídlo (2010). Several publications focus on dynamics of mortality at old age. A comparative study of mortality trajectories at extremely old ages was published in Gavrilov and Gavrilova (2014). Age specific shifts of the logistic model were investigated by Tesárková Hulíková (2012). Long-term mortality trends with respect to high age mortality were studied for example by Gavrilov and Gavrilova (2011).

We assume a two-dimensional extension of the exponential model and model the cohort-age specific force of mortality at old ages simultaneously in the age and cohort dimension. Our main aim is to compare uncertainty of the force of mortality estimates for two models. The first model is a single population model of the old age mortality fitted based on data solely from one population. The second model is a multi-population model for a sample of populations from the Central European region. We quantify the difference in the uncertainty of the estimates from the two models for five countries from the central European region, measured by mean squared error for both the force of mortality and whole life annuity. The main question is: “To what extent data from surrounding populations reduce the estimation uncertainty.” Ahcan et al. (2013) proposed incorporating information from surrounding countries using numerical optimization method to

obtain a linear combination of death rates from surrounding countries which optimally explains the target population. The ‘mixing’ of base population and the linear combination is then based on credibility approach taking into account different exposures of each country. Resulting ‘credibility data’ are then used to apply standard single population models. We use Bayesian methods. Namely Bayesian generalized linear model (BGLM) and hierarchical Bayesian generalized linear model (HBGLM). Czado, Delwarde and Denuit (2005) investigated Bayesian Poisson log-bilinear mortality predictions for adult ages in contrast to traditionally used Lee-Carter model (Lee and Carter, 1992). Antonio, Bardoutsos and Ouburg (2015) extended the model of Czado to include multiple populations. Forecasts of male period life expectancy at birth for all countries of the world until the year 2100 were also performed using a Bayesian hierarchical model proposed by Raftery et al. (2013). The application of Bayesian random-effects approach for improving mortality predictions of a small population was investigated by Jonker, van Lenthe and Congdon (2012), who concluded that Bayesian Monte Carlo simulations outperform traditionally calculated life expectancies in bias, root mean square error and credible intervals coverage. Our main contribution is the construction of a multi population old age HBGLM model and its application on central European region in order to quantify the reduction of uncertainty in parameter estimates in comparison to a single population model.

The structure of the article is as follows. Section 1 presents underlying data and corresponding notation used. Single-population model together with the likelihood specification, prior and posterior distributions are introduced in Section 2. Section 3 extends the single model to a multi-population hierarchical model and describes corresponding prior distributions of model parameters. Section 4 describes simulation procedures used for sampling from posterior distributions of the hierarchical model. Empirical results are presented in Section 5. Conclusions based on presented results are in last section.

1. Data and Notation

Our scripts for statistical software R, together with sample dataset for Czech Republic are placed at a link specified.² Source of all data used in this paper is The Human Mortality Database (Wilmoth, Shkolnikov and Barbieri, 2012), abbreviated as HMD in further text. The input data are the ‘Death rates’ and ‘Exposure-to-risk’ data in cohorts data section in one year age and one year calendar intervals. Number of deaths is calculated as the multiple of death rate and the

² Scripts produced for this research are placed at:
<https://drive.google.com/file/d/1OqdQfMAVrtvKsU4s1CavyiWugaqhIwqU/view?usp=sharing>.

exposure to risk. Five countries are included in our analysis: Austria, Hungary, former East Germany, Czech Republic and Slovakia. These countries were chosen with respect to geographical, historical and social similarities to the Czech Republic. Poland would also be a suitable candidate, unfortunately, reliable data are not available. The model was fitted for the age range 85 – 99. The inclusion of observations from younger ages would reduce the influence of the old age data which are the main focus of the research. Observations above 99 years of age are either not available at all or its quality is dramatically low. For the age category 85 – 99 and the selected set of countries, cohorts from 1879 to 1913 are available.

Population exposed to risk of death of the cohort c at the age x is denoted $E_{x,c}$, the number of deaths is $D_{x,c}$. The force of mortality is denoted $\mu_{x,c}$. Countries are distinguished by upper right index i . The total number of countries involved is denoted q . In age-cohort format, HMD publishes only deaths rates $m_{x,c}$ and exposure to risk $E_{x,c}$. Therefore cohort number of deaths $D_{x,c}$ is calculated as

$$D_{x,c} = m_{x,c} E_{x,c} \quad (1)$$

Probability that a person at age x of the cohort c will survive to an age $x + 1$ is denoted $p_{x,c}$ and calculated as follows

$$p_{x,c} = \exp \int_0^1 \mu_{x+s,c+s} ds \quad (2)$$

The analysis will take into account not only age-cohort specific death rates but also age-cohort specific whole life annuities. Therefore a whole life annuity for an age x and a cohort c is calculated as

$$\ddot{a}_{x,c} = \sum_{t=x}^{\omega} p_x v^t \quad (3)$$

where

v – denotes discounting factor.

2. Single-population Model and Prior Distributions Specification

As stated above, several curves have been proposed for the dependence of old age mortality on age. In order to keep the number of parameters reasonable, only two-parametric specifications were considered. Exponential and logistic curves are the most popular specifications. As a significant difference between these two curves appears mainly in extrapolations to extreme age above 105 years, which is not the main focus of this article, we limit our analysis to the exponential curve (Gompertz model).

2.1. Likelihood of the Number of Deaths

Consistently with recent research (Brouhns, Denuit and Vermunt, 2002; Antonio, Bardoutsos and Ouburg, 2015; or Czado, Delwarde and Denuit, 2005) we assume that number of deaths in each age x for each cohort c follows a Poisson distribution with expected value proportional to the population exposed to risk of death

$$D_{x,c} = Po(\mu_{x,c} E_{x,c}) \quad (4)$$

The following 4 parametric linear predictor, i.e. the linear function of age and cohort, is assumed for the dependence of the force of mortality on age, cohort, and interaction of age and cohort:

$$\ln(\mu_{x,c}) = \beta_1 + \beta_2 x + \beta_3 c + \beta_4 xc \quad (5)$$

This means the single population model is specified as a Bayesian generalized linear model with a Poisson distribution of the response variable $D_{x,c}$ and logarithmic link function with the offset equal to the logarithm of the exposure $E_{x,c}$.

2.2. Prior Distributions

Prior distribution represents the information available prior to data collection. Multivariate normal prior distribution is assumed for the vector of regression parameters $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)^T$:

$$\beta \sim N_4(\beta_0, \Sigma_0)$$

where β_0 and Σ_0 are so called hyperparameters determining the shape of the prior distribution. Following (Hoff, 2009), we take an approach based on unit information, in which the prior distributions for the parameters are weakly centered around estimates derived from the observed data. A frequentist GLM model with the following schematic specification is constructed in order to set the prior expectation of β_0 :

$$D \sim AGE + COHORT + COHORT : AGE$$

where symbol ‘+’ separates the regression terms and ‘:’ denotes the interactions. Offset $\ln E_{x,c}$ was used. For the prior covariance Σ_0 , a large constant (1000) is selected for the variances on the diagonal, and independence is assumed, i.e. off-diagonal elements are set to 0. High prior variance reflects our uncertainty associated with the choice of prior information.

2.3. Posterior Distribution

Posterior distribution is the distribution that combines the information from both the prior distribution as well as the sample. Inference about parameter estimates is based on posterior density given the mortality data:

$$p(\beta | D) \propto p(D | \beta) p(\beta) \quad (6)$$

As for generalized linear models, an analytical formula for the posterior distribution (6) does not exist in general. We use simulations to create samples from this distribution and perform an empirical analysis. The Metropolis algorithm is applied to generate a dependent sequence of parameter values whose distribution converges to target posterior distribution (6). The Metropolis algorithm is a two-step algorithm in which a new value is first proposed from so called proposal distribution and then either accepted or rejected with an acceptance probability.

A convenient choice for the proposal distribution is some symmetric distribution concentrated around current value of the generated sequence. Often the multivariate normal distribution is used. The basic steps of the Metropolis algorithm are:

For each simulation ('scan') s

1. Generate ('propose') a random value β^* from proposal distribution

$$\beta^* \sim N_4(\beta^{(s)}, uX^T X)$$

where

X – design matrix of the regression function (5),

u – scaling factor set by 'trial and error' in order to keep the acceptance ratio of the parameters in a reasonable range.

This range is generally suggested to be between 20 – 50% (Hoff, 2009). Following (Czado, Delwarde and Denuit, 2005), the scaling factor is updated every 100 simulations in order to keep the acceptance ratio in the required range.

2. Set $\beta^{(s+1)}$ equal to β^* or equal to current value $\beta^{(s)}$ with acceptance probability r defined as

$$r = \frac{p(D | \beta^*) p_0(\beta^*)}{p(D | \beta^{(s)}) p_0(\beta^{(s)})} \quad (7)$$

where

p_0 – prior distribution of the vector of parameters β ,

p – Poisson likelihood.

Further details on Metropolitan algorithm and alternatives can be found e.g. in Gelman, Carlin and Stern (2014).

3. Multi-population Hierarchical Model

In order to use the information from several, to some extent similar populations, a multi-population model is set up. In order to capture the variability between several populations on one hand, and take the advantage of using multiple data sources in order to reduce the uncertainty in estimates, caused by lack of single population observations, on the other hand, a hierarchical structure is assumed. The model is specified as a Bayesian hierarchical generalized linear model.

For multi-population model, we use analogous specifications as for the single population model. Namely, we also assume the exponential increase of the force of mortality with age and the Poisson distribution for a number of deaths. Similar regression function as in (5) is assumed.

In order to keep the complexity of notation reasonable, we use the same notation as for the single population model. As the following two sections are not referencing to Section 2, no confusion should appear.

3.1. Likelihood of the Number of Deaths and Distribution of Parameters β

Analogously to the single population model, we assume that the number of deaths for each age x , cohort c and each population i follows Poisson distribution with expected value proportional to the population exposed to risk of death

$$D_{x,c}^{(i)} = Po(\mu_{x,c}^{(i)} E_{x,c}^{(i)}) \quad (8)$$

Again, we assume the four parametric regression model for the expected force of mortality. This time, however, with population specific parameters:

$$\ln(\mu_{x,c}^{(i)}) = \beta_1^{(i)} + \beta_2^{(i)} x + \beta_3^{(i)} c + \beta_4^{(i)} xc \quad (9)$$

It is assumed that vector of parameters $\beta^{(i)} = (\beta_1^{(i)}, \dots, \beta_4^{(i)})^T$ is a country specific realization of a random variable with multivariate normal distribution

$$\beta \sim N_4(\theta, \Sigma)$$

where θ can be interpreted as a ‘global’ mean of the parameters from which country specific parameters $\beta^{(i)}$ deviate to the extent driven by the covariance matrix Σ . The model specification is, in fact, a specification of a hierarchical generalized linear model with Poisson distribution and the logarithmic link function, i.e. the transformation that links the expected value to the linear predictor. Prior distributions of θ and Σ are specified in accordance with (Gelman, Carlin and Stern, 2014).

3.2. Prior Distribution of the Global Parameters θ and Σ

We assume semi-conjugate multivariate normal prior distribution for the parameter θ with hyperparameters θ_0 and Λ_0 .

In order to set reasonable values of the hyperparameters θ_0 and Λ_0 , a frequentist GLM model with the following schematic specification:

$$D \sim \text{COUNTRY} : \text{AGE} + \text{COHORT} : \text{COUNTRY} + \text{COUNTRY} : \text{COHORT} : \text{AGE}$$

assuming Poisson distribution with logarithmic link function and $\ln E_{x,c}^{(i)}$ as an offset, was fitted. The interactions with the factor variable *COUNTRY* provide country specific estimates of the regression parameters. The hyperparameter θ_0 is then set as the simple average of the country specific parameters and Λ_0 is the empirical covariance matrix of the country specific estimates of these regression parameters multiplied by a large constant (1 000).

3.3. Prior Distribution of the Global Parameter Σ

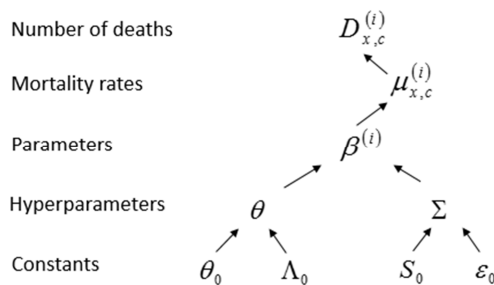
A common choice for the prior distribution of a covariance matrix is the inverse-Wishart distribution with two hyperparameters S_0 and η_0 . In accordance with (Hoff, 2009), we set $S_0 = \Lambda_0$ and in order to set the prior relatively diffuse, we set $\eta_0 = 4 + 2$.

3.4. Summary of the Model

Figure 1 summarizes the structure of the model. The model is built from bottom up to the top of the pyramid. The lowest level consists of parameters determining the distribution of hyperparameters which can be seen in the line above. The middle line represents 4 parameters of the regression model for the expected force of mortality.

Figure 1

Model Construction



Source: Own work.

4. Multi-population Model Posterior Distributions

Inference about the parameter estimates is based on joint posterior density given the mortality data:

$$p(\beta, \theta, \Sigma | D) \propto p(D | \beta, \theta, \Sigma) p(\beta, \theta, \Sigma) \quad (10)$$

We use Markov Chain Monte Carlo simulations to generate samples from this distribution and perform an empirical analysis. Full conditional density of a given parameter is defined as a density of the given parameter conditional on all other parameters involved. For parameters for which analytical form of full conditional density is known, we use Gibbs algorithm. If the full conditional density does not exist in an analytical form, the Metropolis algorithm is applied. The combination of these two algorithms results in the following Metropolis-Gibbs algorithm. Details for each step are in Sections 4.1. – 4.3.

Given current values at scan s of the Markov chain, generate new values at scan $s + 1$ as follows:

1. Sample $\theta^{(s+1)}$ from its full conditional distribution.
2. Sample $\Sigma^{(s+1)}$ from its full conditional distribution.
3. For each country i .

Propose a random value β^* from a proposal distribution specified in Section 4.3.

Set $\beta^{(i,s+1)}$ equal to β^* or equal to current value $\beta^{(i,s)}$ with an acceptance probability specified in Section 4.3.

4.1. Gibbs Sampling for θ

In linear as well as in generalized linear models, full conditional distribution $p(\theta | \beta, \Sigma, D)$ of the mean of the regression parameters does not depend on the distribution assumed for the mortality $p(D | \beta, \theta, \Sigma)$. Assuming that β have a multivariate normal distribution and that the prior distribution for θ is also a normal distribution, the full conditional distribution $p(\theta | \beta, \Sigma, D)$ is the multivariate normal distribution $N_p(\theta_m, \Lambda_m)$, where

$$\Lambda_m = (\Lambda_0^{-1} + q\Sigma^{-1})^{-1} \quad (11)$$

and

$$\theta_m = \Lambda_m (\Lambda_0^{-1} \theta_0 + \Sigma^{-1} \bar{\beta}) \quad (12)$$

where $\bar{\beta} = \beta^{(1)} + \beta^{(2)} + \dots + \beta^{(q)}$ is the vector sum of the vector parameters $\beta^{(i)}$ over all countries involved.

4.2. Gibbs Sampling for Σ

Assuming that β have a multivariate normal distribution and that prior for Σ is the inverse-Wishart distribution, the full conditional distribution $p(\Sigma|\beta, \theta, D)$ is also the inverse-Wishart distribution $IW_p(\eta_0 + q, S_o + S_\theta)$, where

$$S_\theta = \sum_{i=1}^q (\beta^{(i)} - \theta)(\beta^{(i)} - \theta)^T \quad (13)$$

4.3. Metropolis Sampling for β

A new value is in a given scan s first proposed for each country i and then either accepted or rejected. The index of the country i is now dropped for simplicity. Standard proposal probability for the vector of parameters β at a current scan s is a multivariate normal distribution with mean equal to the current value $\beta^{(s)}$ and covariance equal to $v\Sigma$:

$$\beta^* \sim N(\beta^{(s)}, u\Sigma) \quad (14)$$

where u is again a scaling factor analogous to the scaling factor introduced in Section 2.3 for the single population model. It is again set by “trial and error” and updated every 100 simulations in order to keep the acceptance ratio between 20 – 50%. The probability that the proposed value β^* is accepted is at the step s calculated as

$$r = \frac{p(D|\beta^*)p(\beta^*|\theta^{(s)}, \Sigma^{(s)})}{p(D|\beta^{(s)})p(\beta^{(s)}|\theta^{(s)}, \Sigma^{(s)})} \quad (15)$$

A random indicator with alternative distribution with mean equal to r is generated in each scan. If a success is generated, then the proposed value is accepted, i.e. $\beta^{(s+1)} = \beta^*$. Otherwise, $\beta^{(s+1)}$ is set equal to $\beta^{(s)}$.

5. Empirical Results

The estimates of parameters for single and multi-population models are derived from 100 thousand scans. Every 200th scan was saved to reduce the auto-correlation of consecutive scans resulting in 500 simulations. The estimates are based on analyses of data from 34 male cohorts (1879 – 1913) in age category 85 – 99 years for five countries of the Central European region. Whole life annuities are calculated in each simulation for a fixed interest rate of 1%. Estimates from the hierarchical model, showing the diversity of the old age mortality between all

countries considered, as well as the global prediction, common for the whole region, are for illustration displayed for the last cohort $c = 1913$ on Figure 2. The dashed curve represents estimate of $\mu_{x,1913}$ calculated based on estimated values of the expected value θ of the country specific parameters. This curve can be interpreted as a global estimate for the whole region. Country specific curves $\mu_{x,1913}^i$ are then displayed as the lines scattered around this global curve. These curves are calculated based on the estimates of the country specific vector of parameters $\beta^{(i)}$.

For each of the five countries involved in the analysis, the root mean squared error (RMSE) of the estimates of the force of mortality and whole life annuities in every cohort and every age were calculated for both the single and multi-population model. For the single-population model, RMSE is for each age point x , cohort c , and country i , the standard deviation of the simulated estimates:

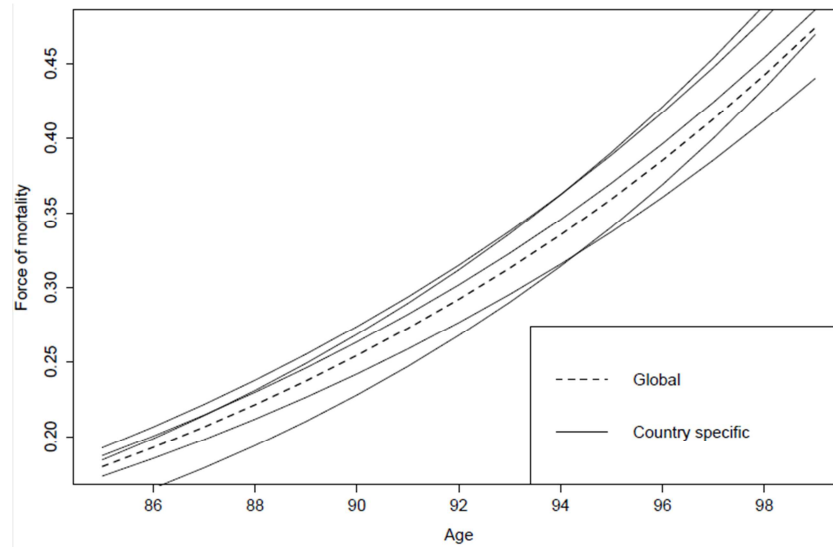
$$RMSE(x, c, i) = \left(\frac{\sum_s (\mu(x, c, i, s) - \bar{\mu}(x, c, i))^2}{n} \right)^{1/2} \quad (16)$$

where

- $\mu(x, c, i, s)$ – estimate of the single population model in a simulation s ,
- $\bar{\mu}(x, c, i)$ – mean of the estimates of the single-population model over all simulations,
- n – number of simulations.

Figure 2

Multi-population Model: Estimates for Cohort $c = 1913$ Calculated Based on Parameters Estimated for Each Country and the Global Parameters



Source: Own work.

For the multi-population model, RMSE is calculated as the square root of the sum of the variance of the estimation over all simulations and squared estimate of the bias. The bias is estimated as the mean difference of the estimates from the multi population model from the mean estimate from the single population model. So the RMSE of the multi population estimate (denoted by additional index m) is calculated as:

$$RMSE(m, x, c, i) = \left(\frac{\sum_s (\mu(m, x, c, i, s) - \bar{\mu}(m, x, c, i))^2}{n} + \left(\frac{\sum_s (\mu(m, x, c, i, s) - \bar{\mu}(x, c, i))}{n} \right)^2 \right)^{1/2} \quad (17)$$

where

- $\mu(m, x, c, i, s)$ – estimate of the multi-population model in a simulation s ,
- $\bar{\mu}(m, x, c, i)$ – mean of the estimates of the multi-population model over all simulations,
- $\bar{\mu}(x, c, i)$ – mean of the estimates of the single-population model over all simulations,
- n – number of simulations.

For each country, the relative difference between RMSE was compared:

$$\frac{RMSE(x, c, i) - RMSE(m, x, c, i)}{RMSE(x, c, i)} \quad (18)$$

Contour plots for the relative difference of RMSE of force of mortality and whole life annuities for each country are displayed in Appendix in Figure 3 and Figure 4 respectively. The contour corresponding to zero difference is marked with the bold white line. Single-population model has higher RMSE (of force of mortality and whole life annuity, respectively) than the multi-population model in areas with a positive difference (lighter areas). In areas with a negative difference (darker areas), the single-population model outperformed the multi-population model. For all countries, there were certain positive and certain negative areas. We may state that:

1. For the Czech Republic, Austria and East Germany relatively large difference occurs in certain areas (see left column of Figures 3 and 4 in Appendix), while as for Slovakia and Hungary, the relative difference is negligible (see right column of Figures 3 and 4 in Appendix). Due to the difference in scale of the relative difference, the scale of the palette is country specific.
2. In general, we may state that multi-population model outperformed the single-population model for lower and middle cohorts and higher and lower ages while as the single-population model outperforms multi-population model for lower cohorts and middle ages.

3. For the highest cohorts, the difference between the two models is generally closer to zero than for the lowest cohorts.

4. Highest positive values of the relative difference in MSE reaches much higher values than the absolute value of the lowest negative difference in MSE. So positive gains in using multi-population models are, on the age and cohort ranges analysed, often higher than negative drawbacks.

Conclusions and Discussion

The main focus of this article is on a comparison of the performance of the single-population and multi-population model. On one hand, including information contained in data from geographically and economically closed populations could reduce variance in old age force of mortality estimates. On the other hand, a bias occurs. A Bayesian hierarchical model is a suitable way to setup a multi-population model. In our application, we set up a common model for the Czech Republic, Austria, Hungary, former East Germany and Slovakia. Resulting uncertainty for single and multi-population model was compared for each country involved. The empirical comparison shows the following main conclusions:

1. It cannot be stated, that multi or single-population model generally outperforms the other one. The difference in RMSE (of both force of mortality as well as whole life annuities) and its sign depend on the values of explanatory variables (age and cohort).

2. For some countries, the relative difference in RMSE of force of mortality as well as the whole life annuity is rather low, and countries for which the difference is high.

3. There are ranges of age and cohorts, common to all countries, for which multi-population model outperforms single-population model and other common ranges, for which single-population model outperforms the multi-population model.

4. In areas in which the multi-population model outperforms the single-population model, the relative difference of RMSE of both the force of mortality and the whole life annuity is generally higher than in areas where single-population model outperforms the multi-population model.

References

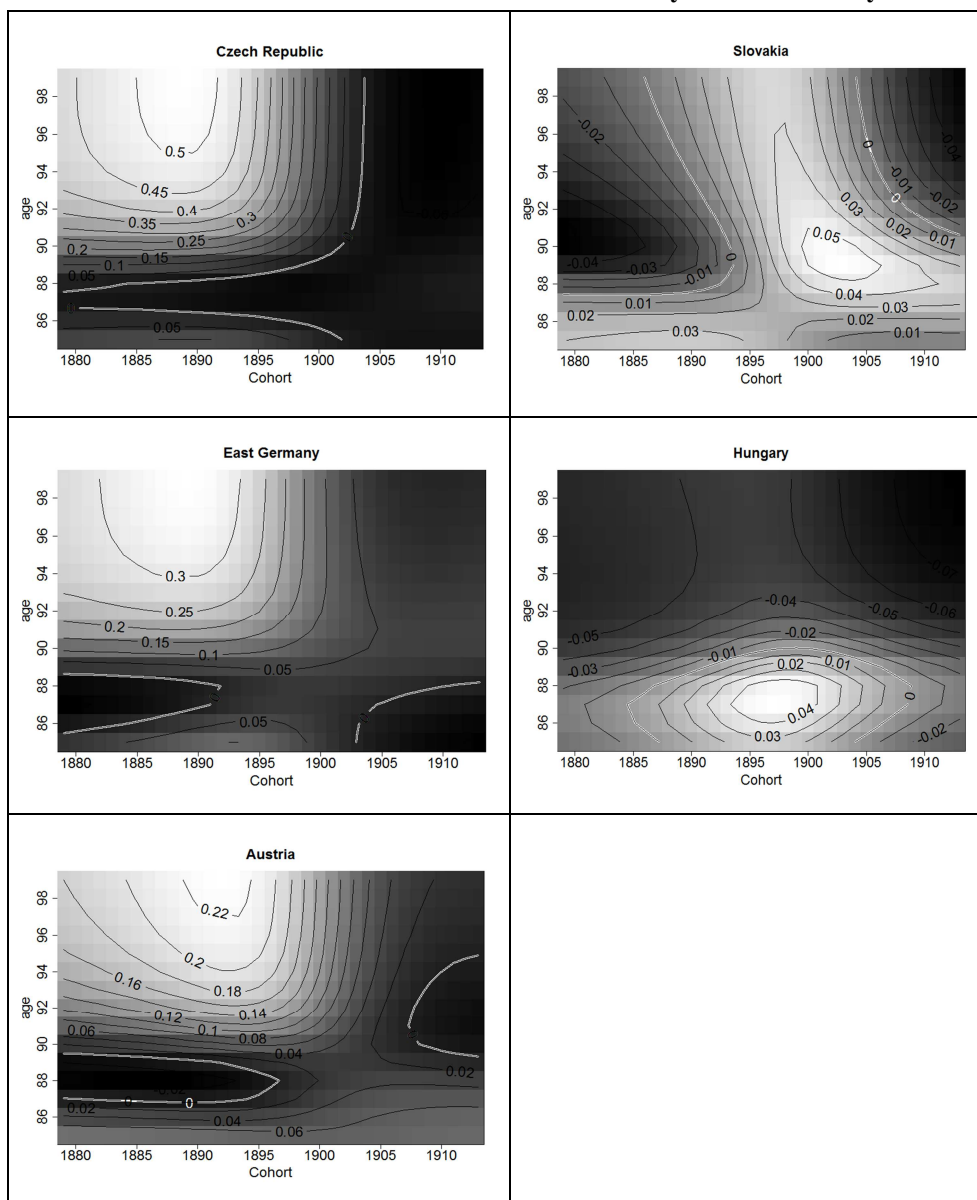
- AHCAN, A. – MEDVED, D. – OLIVIERI, A. – PITACCO, E. (2014): Forecasting Mortality for Small Populations by Mixing Mortality Data. *Insurance: Mathematics and Economics*, 54, 2014, pp. 12 – 27.

- ANTONIO, K. – BARDOUTSOS, A. – OUBURG, W. (2015): Bayesian Poisson Log-bilinear Models for Mortality Projections with Multiple Populations. *European Actuarial Journal*, 5, No. 2, pp. 245 – 281.
- BROUHNS, N. – DENUIT, M. – VERMUNT, J. K. (2002): A Poisson Log-bilinear Approach to the Construction of Projected Lifetables. *Insurance: Mathematics & Economics*, 31, No. 3, pp. 373 – 393.
- BURCIN, B. – TESÁRKOVÁ, K. – ŠÍDLA, L. (2010): Nejpoužívanější metody vyrovnávání a extrapolace křivky úmrtnosti a jejich aplikace na českou populaci. *Demografie*, 52, No. 2, pp. 77 – 89.
- CZADO, C. – DELWARDE, A. – DENUIT, M. (2005): Bayesian Poisson Log-bilinear Mortality Projections. *Insurance: Mathematics and Economics*, 36, No. 3, pp. 260 – 284.
- GAVRILOV, L. A. – GAVRILOVA, N. S. (2011): Mortality Measurement at Advanced Ages: A Study of the Social Security Administration Death Master File. *North American Actuarial Journal*, 15, No. 3, pp. 432 – 447.
- GAVRILOV, L. A. – GAVRILOVA, N. S. (2014): Mortality Trajectories at Extreme Old Ages: A Comparative Study of Different Data Sources on US Old-Age Mortality. *Living to 100 Monograph*. Available at: < <https://www.soa.org/essays-monographs/2017-living-to-100/>>.
- GELMAN, A. – CARLIN, J. B. – STERN, H. S. (2014): *Bayesian Data Analysis*. 2. Boca Raton, FL, USA: Chapman & Hall/CRC.
- 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*, 115, 1825, pp. 513 – 583.
- HELIGMAN, L. – POLLARD, J. H. (1980): The Age Pattern of Mortality. *Journal of the Institute of Actuaries*, 107, No. 1, pp. 49 – 80.
- HOFF, P. D. (2009): *A First Course in Bayesian Statistical Methods*. New York, NY, USA: Springer Science & Business Media.
- JONKER, M. F. – van LENTHE, F. J. – CONGDON, P. D. (2012): Comparison of Bayesian Random Effects and Traditional Life Expectancy Estimation in Small-Areas Application. *American Journal of Epidemiology*, 176, No. 1, pp. 929 – 937.
- KOSCHIN, F. (1999): Jak vysoká je intenzita úmrtnosti na konci lidského života. *Demografie*, 2, No. 2, pp. 105 – 119.
- LEE, R. D. – CARTER, L. R. (1992): Modeling and Forecasting US Mortality. *Journal of the American Statistical Association*, 87, No. 419, pp. 659 – 671.
- MAKEHAM, W. M. (1860): On the Law of Mortality and Construction of Annuity Tables. *The Assurance Magazine and Journal of the Institute of Actuaries*, 8, No. 6, pp. 301 – 310.
- PERKS, W. (1932): On Some Experiments in the Graduation of Mortality Statistics. *Journal of the Institute of Actuaries* (1886 – 1994), 63, No. 1, pp. 12 – 57.
- PITACCO, E. – DENUIT, M. – HABERMAN, S. (2009): *Modelling Longevity Dynamics for Pensions and Annuity Business*. Oxford: Oxford University Press.
- RAFTERY, A. E. – CHUNN, J. L. – GERLAND, P. – ŠEVČÍKOVÁ, H. (2013): Bayesian Probabilistic Projections of Life Expectancy for All Countries. *Demography*, 50, No. 3, pp. 777 – 801.
- TESÁRKOVÁ HULÍKOVÁ, K. (2012): *Selected Methods of Mortality Analysis Focused on Adults and the Old-age*. [Ph. D. Thesis, Ph. D. Dissertation.] Prague: Univerzita Karlova.
- THATCHER, A. R. (1999): The Long-Term Pattern of Adult Mortality and the Highest Attained Age. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162, No. 1, pp. 5 – 43.
- WILMOTH, J. R. – SHKOLNIKOV, V. – BARBIERI, M. (2012): *Human Mortality Database 2012*. Available at: <www.mortality.org>.

Appendix

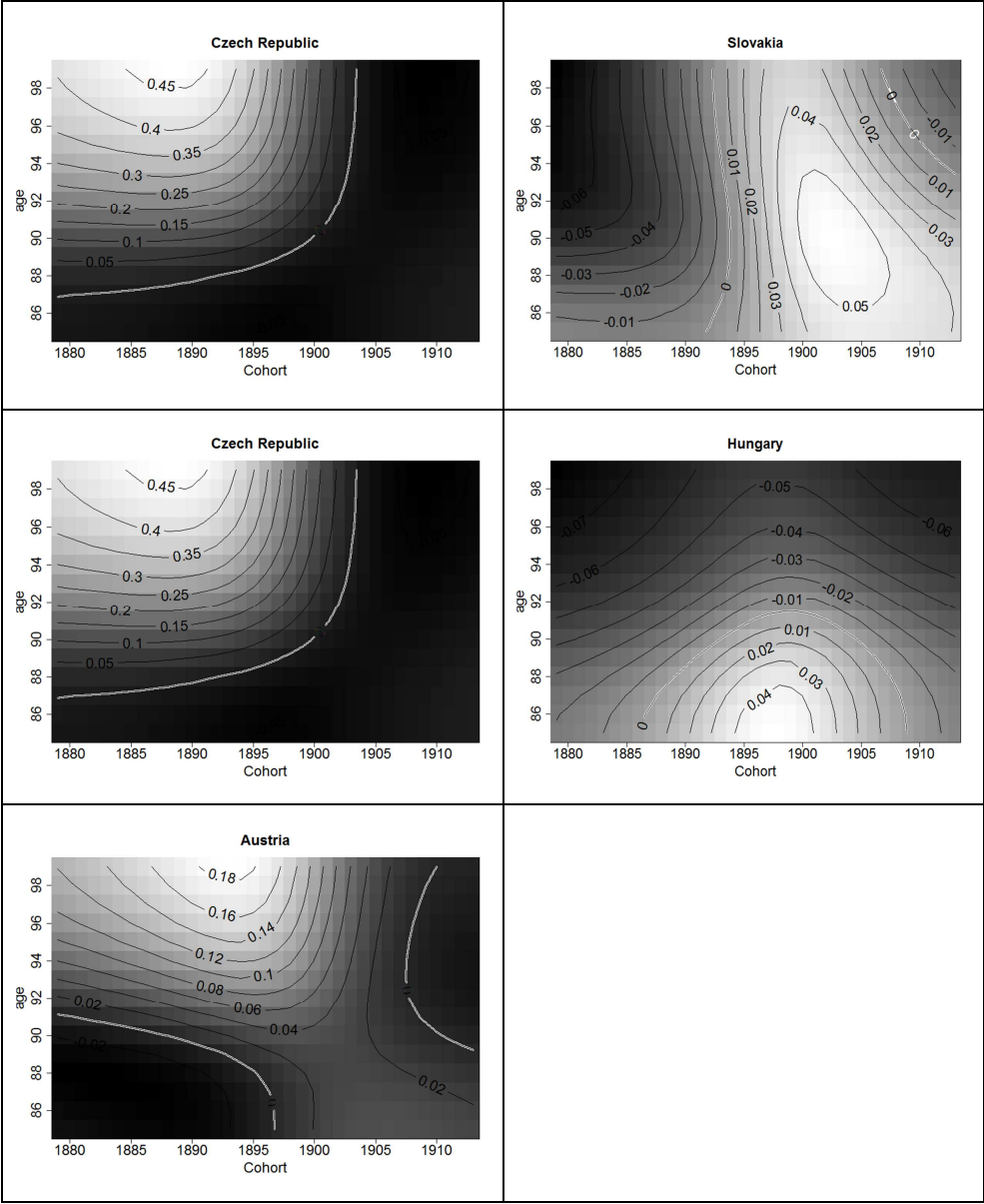
Figure 3

Relative Difference of the RMSE of the Force of Mortality in Each Country



Source: Own work.

Figure 4
Relative Difference of the RMSE of the Whole Life Annuities for Each Country



Source: Own work.