BAYESIAN APROACH TO HIGH AGE MODELING

Martin Matějka − Tomáš Karel − Jan Fojtík − Pavel Zimmermann

Abstract

This article describes solution of modeling high age mortality tables for the Czech Republic. Main problem of high age modeling is the lack of data, especially for small areas such as country, region or micro-region. Solution presented in this paper is based on Bayesian approach. This approach, combines information from examined area with information from surrounding areas (as a prior information). Increase of accuracy of the estimates is examined. Results are presented on real data collected from selected countries of the European Union.

Key words: high age mortality, Bayesian approach, prior mortality information, Bayesian GLM

JEL Code: C10, C11, J10

Introduction

Models of mortality became a very popular task in disciplines such as demography, political sciences, actuarial sciences or economy. For high ages models are in most cases formulated in parametric form. Popular models are usually formulated as one dimensional regression functions of age ("mortality laws"). As the number of observations does low or even not exist at all and hence the growth of mortality in lower ages needs to be extrapolated to higher ages. The major problem of modeling mortality for (very) high ages is often associated with the lack of observations, especially for small populations. Every population has its specifics, but it seems reasonable to find suitable methods to incorporate information collected from geographically or economically similar areas. One of the possibilities to apply this approach is to use Bayesian approach. In this paper we apply Bayesian general linear model to fit the high age mortality data for the Czech Republic. Two different priors, informative and non-informative, are tested and compared. The informative prior is based on empirical Bayesian methodology using the information collected from the so called "Visegrad four" (V4) countries. These countries could be assumed reasonably close to the Czech Republic in economical and in geographical way. Non-informative prior is used in situations where no external or other prior information is available and usually some flat prior, e.g.

uniform or normal distribution with (very) high variability, is assumed. The results based on non-informative prior are almost identical as results computed using classical (frequentist) GLM methodology.

1 Laws of mortality

Over the history, many parametric functions were suggested based on variety of assumptions. First models were not always specified only for very high ages and not specifically for extrapolation but rather to describe the growth of mortality with age ('mortality law'). Benjamin Gompertz (see (Gompertz, 1825)) assumed exponential increase of the force of mortality with increasing age. Later on exponential increase was questioned and other models were developed. One of the most popular alternatives to the exponential models are models based on logistic function. Such specification occurs for example in Beard's model (see (Beard, 1959)), Thatcher's model (see (Thatcher, 1999)) or in Kannisto's model (see (Thatcher et al., 1998)). An overview of other specifications is provided in (Burcin, Tesarkova, and Sidlo, 2010) or in (Pitacco et al., 2009).

In this article we focus purely on the logistic specification as it is presently one of the most popular models. It is also used for data extrapolating by one of the most popular world wide data source, the Human mortality database (Wilmoth et al., 2012).

2 Model and assumptions

The logistic specification of the dependence of the force of mortality on age is defined as

$$m(x) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x))} \text{ for } x \geq x_0, \tag{1}$$

where $m(x)$ denotes the force of mortality, x is the age, x_0 is the high age threshold and β_j are the parameters. Formula (1) can also be written as

$$\ln\left(\frac{m(x)}{1-m(x)}\right) = \beta_0 + \beta_1 x \text{ for } x \geq x_0. \tag{2}$$

Therefore this regression model is normally treated as a member of the broad class of generalized linear models. It is then assumed, that the number of deaths D_x has binomial distribution, i.e.

$$D(x) \sim Bi(E(x), m(x)), \tag{2}$$

where $E(x)$ is the exposure. Maximum likelihood method is commonly used to estimate the parameters β_j in the classical analysis.

The main focus of this article is to incorporate data from surrounding V4 countries as a prior information into the analysis. Hence the application of empirical Bayesian approach seems to be reasonable. Namely we use an informative prior distribution in the form of independent normal distributions

$$\beta_j \sim N(\mu_j, \sigma_j), \text{ for } j = 1,2 \tag{3}$$

where μ_j and σ_j are parameters estimated using classical GLM method on the data collected from the surrounding "V4" countries for fitting informative prior. Normal distribution with variability set 100 times higher than in the informative case was used as the non-informative prior.

The most important criticism of the Bayessian approach is the subjectivity of the selected prior distribution. Note the using the empirical Bayesian approach, this problem was largely avoided. The other point of criticism of this approach is its computation cost during calculation of posterior densities of coefficients β_j. In most cases is not possible to express the posterior density analytically and simulations are necessary. The posterior densities $H\left(\overline{\beta};\sigma^2|X\right)$ of the coefficients in the model were computed using the Bayesian formula

$$\begin{aligned} &\text{Posterior} \propto \text{Prior} \times \text{Likelihood} \\ &H\left(\overline{\beta};\sigma^2|X\right) \propto P\left(\underline{\beta};\sigma^2\right) F\left(X|\underline{\beta};\sigma^2\right) \end{aligned} \tag{4}$$

where the prior density $P\left(\underline{\beta};\sigma^2\right)$ is multiplied by the likelihood $F\left(X|\underline{\beta};\sigma^2\right)$ calculated on the dataset X. Likelihood in the model express the probability of prior information under condition of the observed dataset in the area that we are interested in.

Posterior densities in this article were estimated in R using iterative Markov chain Monte Carlo sampling methods. Details of the use Gibbs sampler for simulating posterior densities are described e.g. in (Koop(2003)).

4. Data used in the analysis

Data including number of deaths and exposure to risk have a period character and are collected for 'Visegrad four' countries (Czech Republic, Slovakia, Hungary and Poland). The data are selected for the years 2004-2009 in ages from 80 to 100 year olds. Data are displayed in the appendix. Data comes from two sources providing detailed mortality and population data sets. Specifically the Demography section in Eurostat database and The Human Mortality

Database (HMD). Methods of calculation exposure to risk and missing death counts are presented in (Wilmoth, et al., 2012).

4. Results

Calculations presented in this article were performed using the package Data Analysis Using Regression and Multilevel/Hierarchical Models (arm) in statistical freeware R. See the documentation Gelman et al. (2015) for details. As stated above, two different priors were assumed. Firstly informative prior distributions were assumed for both parameters β_0 and β_1. These informative prior distributions were fitted based on the data collected from the surrounding V4 countries. Posterior distributions were then calculated using the Czech data. The results are presented in Table 1, which contains both the prior estimates as well as the posterior estimates. The corresponding densities are displayed in Figure 1.

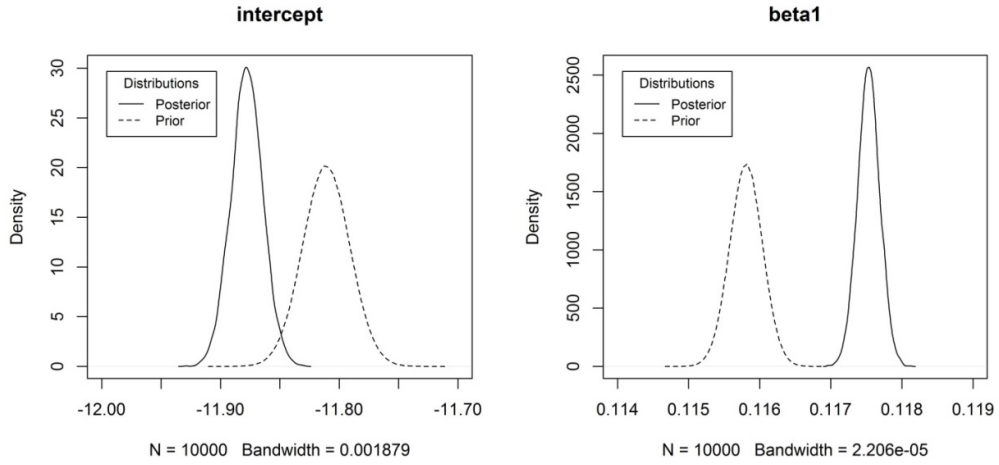
It is obvious that the prior information influences the posterior densities and estimates. The influence of the prior distribution may be measured by comparing the results to the results obtained using only limited prior information, i.e. similar to the situation where no prior data or knowledge is available. The non-informative priors assumed here were again normal distributions with the same prior mean μ_j as in the case of the informative prior. This time, however, the prior standard deviation σ_j was multiplied by 100. The results are also displayed in table Table 1 and in Figure 2.

Tab. 1: Comparison of prior and posterior densities of parameters

		Prior		Posterior	
	Coefficients	Exp. value	Stdev.	Exp. value	Stdev.
Informative	β_0	-11,181	0,01978	-11,880	0,01335
	β_1	0,115	0,00023	0,1175	0,00015
Non-informative	β_0	-11,181	1,9779	-12,9300	0,04384
	β_1	0,115	0,02312	0,1299	0,00051

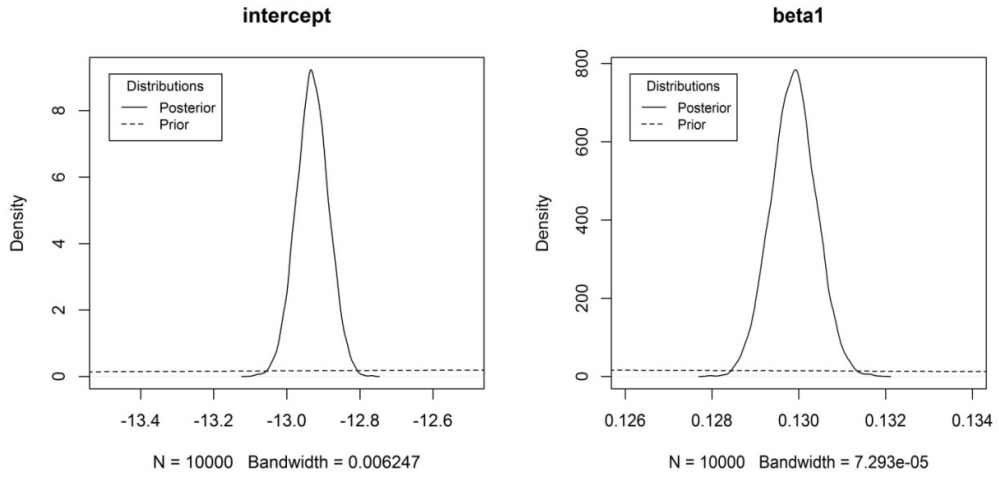
Source: Authors´ computation in R

Fig. 1: Informative prior and posterior densities of coefficient β_0 and β_1



Source: Authors´ computation in R

Fig. 2: Non-informative Prior and Posterior densities of coefficient β_0 and β_1

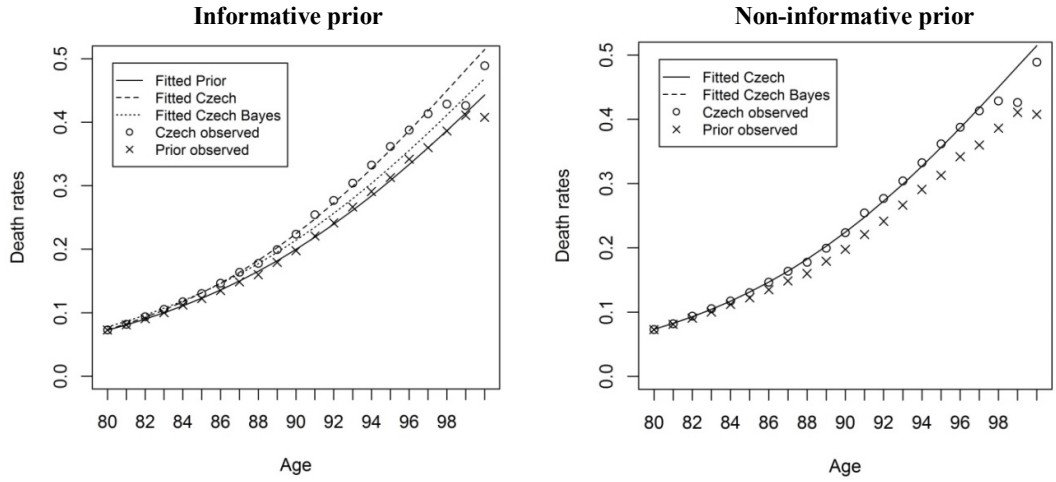


Source: Authors´ computation in R

The comparison of prior and posterior densities of parameters in Figure 1 and Figure 2 presents how much observed data influenced the prior, or the other way around, how much the prior data influenced the estimate based purely on Czech data.

The credible intervals for both β_0 and β_1 parameters, and decreased by approximately 42 %. The comparison of the prior data, prior fit, Czech data, classical fit and Bayesian fit is displayed in Figure 3.

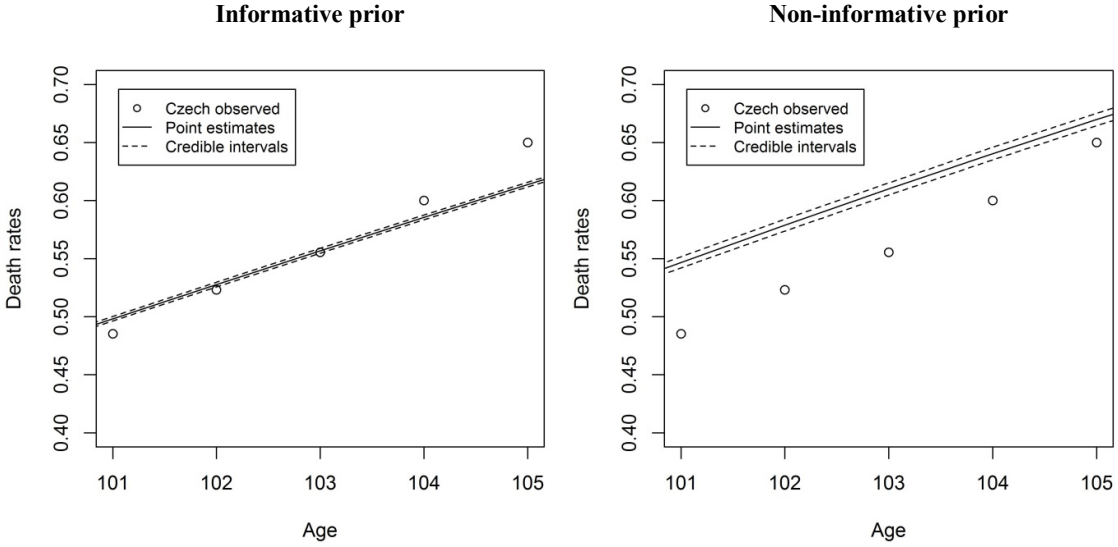
Fig. 3: Fitted data using empirical Bayesian methodology with informative and non-informative prior



Source: Authors´ computation in R

Extrapolation to the age 101 – 105 years based on the Bayesian fit is displayed together with the credible intervals based on posterior densities of parameters in Figure 4.

Fig. 4: Extrapolation to the age of 101 – 105 years with credible intervals using informative and non-informative prior



Source: Authors´ computation in R

Conclusions

Bayesian methods are a natural way to overcome the problem with lack of data, hence it is reasonable to apply these methods in the field of high age mortality models. On one hand, demographic models are typically set up on the country level. On the other hand, high age

data will always be scarce in small areas and it is obvious that areas reasonably similar (both geographically, as well as economically) will also have similar mortalities. Therefore it is natural to base the country level estimates not only on the country of interest data but also on the data collected in the surrounding areas. As can be seen from the results displayed above, the final empirical Bayesian model is then "somewhere between" the classical model based purely on the Czech data and the classical model based purely on the data from the other V4 countries. The decrease of the variability between prior and posterior parameter distribution is rather high. Predictions are based on much more information available for the estimates.

As the empirical Bayesian approach was applied, i.e. the prior distributions were fitted based on observed data, the subjectivity in selecting priors was only limited to the choice of the distribution family, in this case normal family, and on the choice of the prior data, but not on the parameters of the priors itself.

Acknowledgment

The support of the grant IG 410025 "Využití bayesovských metod pro modelování úmrtnosti" is gladly acknowledged.

References

1. Beard, R. E. (2008). Appendix: Note on some mathematical mortality models. *Ciba Foundation Symposium-The Lifespan of Animals (Colloquia on Ageing)*, Vol 5. 302-311.

2. Burcín, B. Tesárková, K., Šídlo, L. (2010) Nejpoužívanější metody vyrovnávání a extrapolace křivky úmrtnosti a jejich aplikace na českou populaci. *Demografie*. vol. 52: 77-89.

3. Gelman, A., et al. (2015). Data Analysis Using Regression and Multilevel/Hierarchical Models, version 1.8-4, 2015. http://cran.r-project.org/web/packages/arm/arm.pdf

4. 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,* 513-583.

5. Koop, G. (2003). Bayesian Econometric. John Wiley & Sons. ISBN 0-470-84567-8

6. Pitacco, E. (2009). *Modelling longevity dynamics for pensions and annuity business*. Oxford: Oxford University Press.

7. Thatcher, A., & Kannisto, V. (1998). *The force of mortality at ages 80 to 120*. Odense, Denmark: Odense University Press.

8. Thatcher, A. (1999). The long-term pattern of adult mortality and the highest attained age. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 5-43.

9. Wilmoth, J. R., Shokolnikov, V., Barbieri, M. 2012. Human mortality database. 2012. www.mortality.org.

Appendix

Tab. 2: Count of Deaths 2004-2009

Age	Czech Republic	Hungary	Poland	Slovakia
80	23232	25990	72791	10853
81	23971	26659	72698	11165
82	24997	26969	72159	11275
83	24974	26414	69586	11113
84	23806	25855	66349	10649
85	21388	22992	59810	9480
86	18335	19932	52359	8015
87	15198	16686	44908	6460
88	12108	13547	38008	4983
89	10387	11800	34034	4321
90	9358	10202	30274	3835
91	8895	9343	27866	3504
92	8122	8773	25174	3160
93	7124	7608	22456	2825
94	5894	6335	18904	2336
95	4471	4773	14838	1724
96	3112	3531	10999	1306
97	2080	2332	7724	862
98	1318	1487	5330	575
99	778	983	3509	357
100	509	619	2090	204

Source: The Human Mortality Database and Eurostat

Tab. 3: Exposure to Risk 2004-2009

Age	Czech Republic	Hungary	Poland	Slovakia
80	318401	326513	1049434	129173
81	293541	299771	946233	118381
82	266779	273232	838981	107349
83	237204	245554	730862	94917
84	203253	215758	620716	81899
85	164050	181097	507028	67278
86	125092	144158	399696	51536
87	92811	111137	309254	38344
88	68157	84718	240667	28375
89	52070	66184	191638	22116
90	41814	52164	154545	17589
91	34952	42322	127913	14287

Age	Czech Republic	Hungary	Poland	Slovakia
92	29353	35293	106445	12023
93	23431	28264	85471	9660
94	17734	21603	65832	7390
95	12365	15248	47723	5233
96	8030	10007	32907	3431
97	5034	6288	21900	2151
98	3076	3829	13994	1325
99	1826	2295	8719	790
100	1041	1326	5368	453

Source: The Human Mortality Database and Eurostat

Contact

Martin Matějka
University of Economics, Prague
náměstí Winstona Churchilla 4, 130 67 Praha 3
martin.matejk@vse.cz

Jan Fojtík
University of Economics, Prague
náměstí Winstona Churchilla 4, 130 67 Praha 3
xfojj00@isis.vse.cz

Tomáš Karel
University of Economics, Prague
náměstí Winstona Churchilla 4, 130 67 Praha 3
tomas.karel@vse.cz

Pavel Zimmermann
University of Economics, Prague
náměstí Winstona Churchilla 4, 130 67 Praha 3
zimmerp@vse.cz