18th AMSE APPLICATIONS OF MATHEMATICS AND STATISTICS IN ECONOMICS
International Scientific Conference | Jindřichův Hradec, 2-6 September 2015

MODELING OF HIGH AGE MORTALITY FOR THE CZECH REPUBLIC USING BAYESIAN METHODOLOGY

TOMÁŠKAREL, JAN FOJTÍK, MARTIN MATĚJKA, PAVEL ZIMMERMANN
University of Economics, Prague, Faculty of informatics and statistics,
Department of statistics and probability calculus,
W. Churchill Sq. 4, Prague 3, Czech Republic
email: tomas.karel@vse.cz

Abstract
This article presents one of the possibilities for modeling high age mortality. The most important problem in high age mortality modeling is the lack of data. Bayesian approach helps to solve this problem.In this article we model the high age death rates of mortality of the Czech Republic using empirical bayesian approach. Namely prior distributions are fitted using data from surrounding countries that are reasonably close both geographically as well as economically. This empirical prior information helps to remove one of the most criticized characters of bayesian models - subjective choice of prior distributions. Results proved that using the prior information results in considerable decrease of the credibility intervals of the estimates.

Key words: *High age mortality, Empirical Bayes, Bayesian GLM.*

1. Introduction

Mortality models have been studied and applied in many fields of science such as demography, political sciences, actuarial sciences or statistics in general. Different approaches are usually applied for adult ages and for (very) high ages. For adult ages, models are usually formulated in nonparametric or semiparametric way as the number of observations is usually sufficient for such analysis. For (very) high ages, popular modelsare usually formulated as one dimensional regression functions of age(“mortality laws”) as the number of observations is much lower or even not existing at all and the growth of mortality in lower ages needs to be extrapolated to higher ages. The lack of observations naturally appears in this context, especially when modeling small populations. Although such population certainly have its specifics, it is probably not reasonable in usual tasks to assume that it evolves completely isolated from what is happening in surrounding populations. Hence it seems reasonable to find suitable methods to incorporate information collected from geographically and / or economically close areas which might be considered relevant for the population of our interest. Natural way of doing so is application of Bayesian methods. In this article we apply Bayesian general linear models to fit the high age mortality model for the Czech Republic. Data collected from the so called “Visegrad four” (V4) countries, which are reasonably close both economically as well as geographically to the Czech Republic are used as the prior information for the model.

There are many parametric functions suggested based on variety of assumptions. Firstmodelswere 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'). They can be dated already to the 19th century by Benjamin Gompertz (see (Gompertz, 1825)) who assumed exponential increase of the force of mortality with increasing age. This model waslateronmodified by William Makeham who added an extra parameter to the model.

With more data available in older ages, exponential increase appears to be too fast for some authors and other models were developed. In (Koschin, 1999) a modification of the Gompertz-Makeham model was suggested. 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)). But many other specifications may be found. An overview is provided in (Burcin et al., 2010) or in (Pitacco et al., 2009).

In this article we focus purely on the Kannisto's model as the logistic specification is presently one of the most popular.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). The methodology is however general enough to be used with any of the regression models mentioned above. The methodology of bayessiangeneralized linear modeling was studied e.g. inKoop (2003).Bayesian methodology was previously applied on the problem of micro region mortality modeling in (Jonkers, 2012) or recently in (QU, 2015). To the authors' knowledge the empirical bayesian methodology was not previously applied in the context of high age modeling.Solution to determine a practical approach for small areas was also suggested by (Stephens, 2013).

2. Methodology

As stated above, we assume the logistic specification for the dependence of the force of mortality on age, namely we assume that

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

where $m(x)$ denotes the forceof mortality, x is the age, x_0 is the high age threshold and β_j are the parameters. 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 application of empirical Bayesian approach seems to be reasonable approachto incorporate data from surrounding V4 countries as a prior information into the analysis. Usually the first choice for of the parametric prior distributions in this context are independent normal distributions see e.g. Koop (2003) due to the asymptotic normality of the classical regression estimates of the regression parameters. Namely we use an informative prior distribution in the form of independent normal distributions

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

where μ_j and σ_j are parameters estimated using classical GLM approach on the data collected from the surrounding V4 countries.

Analytical formula for the posterior density does not exist. Therefore Markov chain Monte Carlo (MCMC) simulations are used in order to approximate the posterior densities and its characteristics for both the parameters as well as the predictions. The methodology is described in detail for example in Gelman et al. (2003)orKoop (2003).

3. 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 year 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 deaths counts are presented in (Wilmoth et al., 2012).Available data are presented in Appendix.All data files and scripts used for calculation of the following results are available at:
https://adweb.vse.cz/htcomnet/Handlers/Download.ashx?action=download&file=Home%20drive%2FR_AMSE.rar

4. Results

Calculations presented in this article were performed using the packageData Analysis Using Regression and Multilevel/Hierarchical Models(arm) in statistical freeware R. See the documentation Gelman et al.(2015) for details. As there are only few observations available for the Czech Republic for ages higher than 85 years and further more data are not always of reasonable quality, we tried to incorporate the data available for surrounding V4 countries. These data will be further referred as prior data.

As stated above, independent normal distributions were assumed as the prior distributions for the parameters β_j. Following the empirical bayesian approach, prior data were used to fit the parameters of the prior distributions.

See Figure 1 and Figure 2 for prior and posterior distributions of parameters β_0 and β_1 for females. For males is the situation similar. Bayesian parameter estimates as well as its posterior distributions were calculated using data collected from the Czech Republic.The results are in the Table 1 and Figures3 - 5.

Table 1. Comparison of prior and posterior densities of parameters

Gender	Coefficients	Prior		Posterior	
		Exp.value	Stdev.	Exp.value	Stdev.
Females	β_0	-12,900	0,057	-12,970	0,039
	β_1	0,127	0,001	0,129	0,000

Males	β_0	-10,216	0,086	-10,410	0,058
	β_1	0,099	0,001	0,102	0,001

Source: Authors´ computation in R

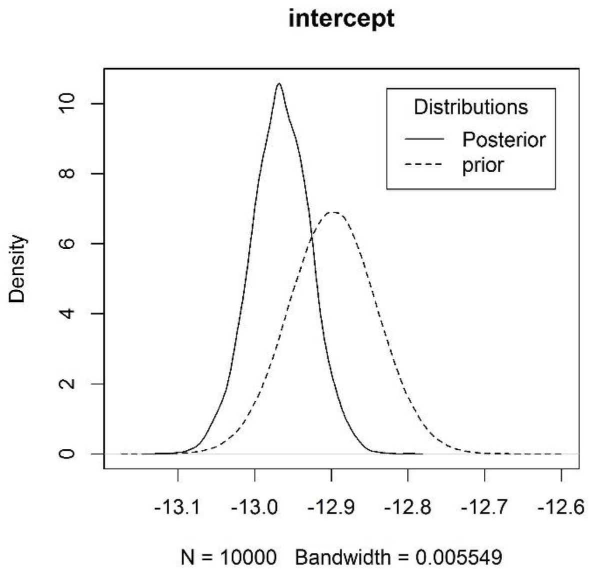

Figure 1.Prior and Posterior densities of coefficient β_0 for females.
Source: Authors´ computation in R.

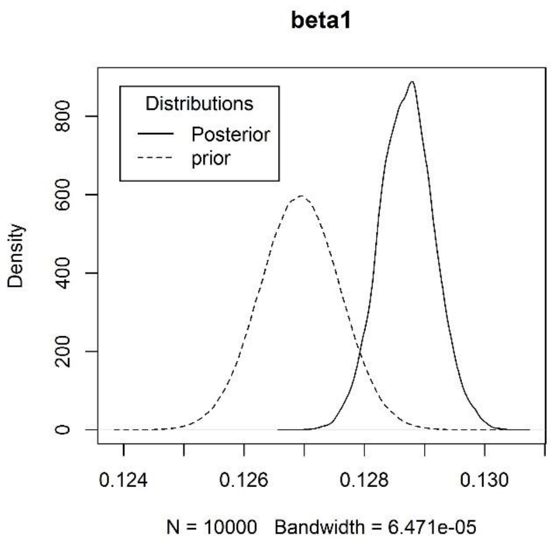

Figure 2.Prior and Posterior densities of coefficient β_1 for females.
Source: Authors´ computation in R

Table 2.Classical GLM estimates for the Czech data.

Gender	Coefficients	Exp. value	Stdev.
Females	β_0	-14,111	0,131
	β_1	0,142	0,002
Males	β_0	-12,749	0,196
	β_1	0,130	0,002

Source: Authors´ computation in R

Table 2 contains classical estimates based purely on the Czech data. It is obvious that the parameter estimates were “pulled” towards the prior estimates. Despite the fact that the prior distribution itself was informative, a massive reduction of variability can still be observed in the posterior distribution. 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 for both females and males decreased by approximately32 %. The comparison of the prior data, prior fit, Czech data, classical fit and bayesian fit is displayed in Figure 3 and 4.

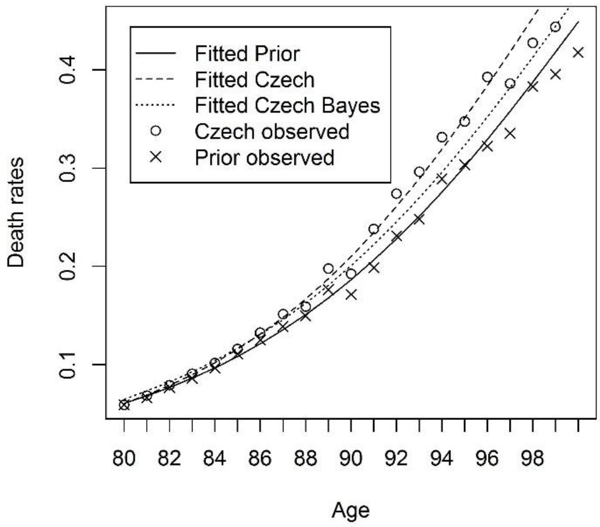

Figure 3.Fitted data using empirical bayesianmethodology for females
Source: Authors´ computation in R

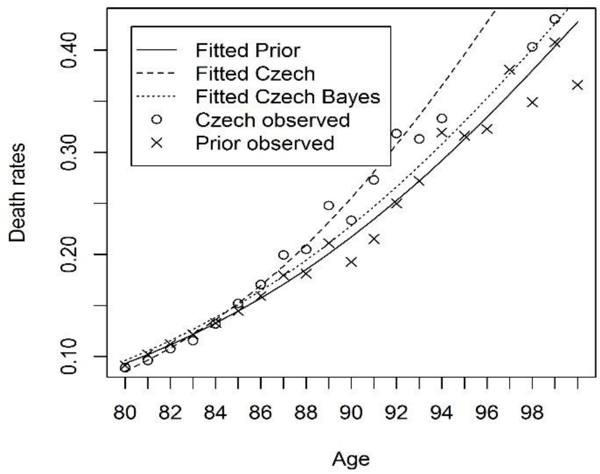

Figure 4. Fitted data using empirical bayessian for males
Source: Authors´ computation in R

Extrapolation to the age 101 – 105 years based on the bayesian fit is then displayed together with the credible intervals basedon posterior densities of parametersin Figure 5.

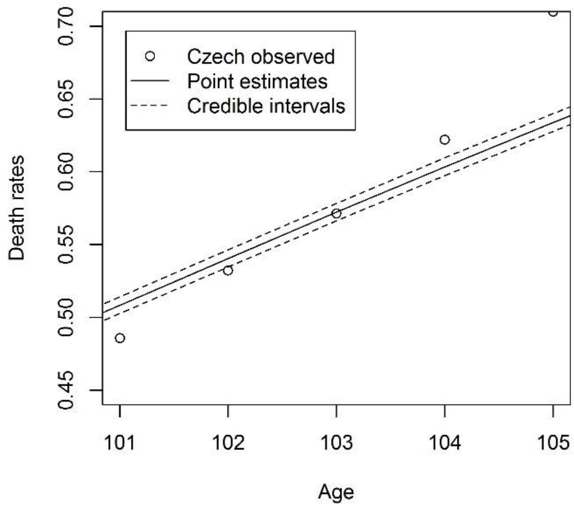

Figure 5. Extrapolation to the age of 101 – 105 years with credible intervals for females
Source: Authors´ computation in R

5. Conclusions

Bayesian methods are a natural way to overcome problem with lack of data, hence it is reasonable to apply these methods in the field of high age 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.

Acknowledgements

The support of the grant IG 410025 "Využitíbayesovskýchmetod pro modelováníúmrtnosti" is gladly acknowledged.

References

1. BEARD, R. E. 2008. Appendix: Note on some mathematical mortality models. In: Ciba Foundation Symposium-The Lifespan of Animals (Colloquia on Ageing), vol 5. pp. 302-311.
2. BURCIN, B., TESÁRKOVÁ, K., Šídlo, L. Nejpoužívanější metody vyrovnávání a extrapolace křivky úmrtnosti a jejich aplikace na českou populaci. Demografie. 2010, vol. 52,pp. 77-89.
3. GELMAN, A. Bayesian Data Analysis. CRC Press. 2003. ISBN 978-1-4398-4095-5.
4. 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
5. 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. 1825. vol. 115. pp. 513-583.
6. JONKER, M. F., et al. 2012. Comparasion of Bayesian Random Effects and Traditional Life Expenctacy Estimation in Small-Areas Application. In American Journal of epidemiology. 2012. vol. 176.1, pp. 929-937.
7. KOOP, G. Bayesian Econometric. John Wiley & Sons. 2003. ISBN 0-470-84567-8.
8. KOSCHIN, F. 1999. Jak vysoká je intenzita úmrtnosti na konci lidského života. Demografie. 1999. vol. 41.2. pp. 105-119.
9. PITACCO, E., et al. 2009. Modelling longevity dynamics for pensions and annuity business. Oxford University Press, 2009. ISBN 978-0-19-954727-2.
10. QU. Y., MEEDEN, G. D., ZHANG, B. 2015. An objective stepwise Bayes approach to small area estimation. In Journal of statistical computation and simulation, 2015, vol. 85, iss. 7, pp.1474-1494.
11. STEPHENS, A. S., et al. 2013. Life expectancy estimation in small administrative areas with non-uniform population sizes: application to Australian New South Wales local government areas. 2013. In BMJ Open. vol. 3, iss. 12.
12. THATCHER, A. R., KANNISTO, V., VAUPEL, J. W. 1998. The force of mortality at ages 80 to 120. Odense University Press. 1998. ISBN 87-7838-381-1.

13. THATCHER, A. R. 1999. The long-term pattern of adult mortality and the highest attained age. In Journal of the Royal Statistical Society: Series A (Statistics in Society),1999,vol. 162, iss. 1, pp. 5-43.
14. WILMOTH, J. R., SHOKOLNIKOV, V., BARBIERI, M. 2012. Human mortality database. 2012. www.mortality.org.

Appendix

Table 3. Count of Deaths 2009

Age	Czech Republic		Hungary		Poland		Slovakia	
	Females	**Males**	**Females**	**Males**	**Females**	**Males**	**Females**	**Males**
85	2646	1470	2575	1410	7888	4047	1166	563
86	2634	1367	2621	1355	7752	3763	1113	530
87	2530	1278	2581	1234	6686	3322	1094	508
88	2089	1010	2225	981	5766	2633	853	396
89	1806	818	2164	940	5321	2354	783	351
90	1005	409	1137	502	3631	1406	457	201
91	718	267	755	301	2710	1010	272	109
92	674	247	699	265	2577	944	234	96
93	664	207	642	226	2375	826	219	93
94	723	207	760	265	2447	809	257	105
95	624	224	670	222	2069	657	226	75
96	471	155	506	145	1576	457	160	69
97	307	90	328	109	1123	352	126	43
98	215	46	241	59	851	202	70	27
99	140	33	151	37	561	161	53	14
100	91	24	99	22	365	87	28	9

Source: The Human Mortality Database and Eurostat.

Table 4. Exposure to Risk 2009

Age	Czech Republic		Hungary		Poland		Slovakia	
	Females	Males	Females	Males	Females	Males	Females	Males
85	22877	9672	22385	9189	73608	28878	8829	3565
86	19896	8015	20677	8134	63498	24238	7808	3106
87	16713	6406	17792	6915	50338	18667	6575	2626
88	13138	4926	14819	5598	39287	14586	4994	1955
89	9141	3300	12114	4646	30975	11088	3834	1543
90	5223	1751	7501	3053	20595	6926	2397	978
91	3018	978	4076	1635	13578	4486	1166	477
92	2461	776	3173	1231	11055	3606	976	387
93	2242	661	2748	1012	9424	2867	887	331
94	2181	621	2754	951	8344	2417	899	322
95	1797	495	2333	779	6722	1964	719	273
96	1200	326	1590	514	4885	1381	480	184
97	795	194	1036	317	3350	897	317	110
98	503	114	644	182	2187	585	203	59
99	315	77	392	103	1413	381	130	36
100	177	40	234	57	873	243	71	22

Source: The Human Mortality Database and Eurostat.