Modelling frontier mortality using Bayesian generalised additive models

Jason Hilton, Erengul Dodd, Jonathan J. Forster, Peter W.F Smith

1 Introduction

Modelling and forecasting mortality is a vital function for government bodies that produce official statistics. Population projections and life expectancy calculations depend on their production, and in turn these influence policy on public pensions, health spending, and planning. Official projections may gain from utilising data from across a range of countries (see, for example Raftery et al. (2013)), as this greater depth of mortality experience may reveal the long-term pattern in mortality more clearly than any single country alone. Additionally, best-practice life expectancy, defined as lowest value of life expectancy globally, has shown sustained increases over many decades (Oeppen and Vaupel 2002). While individual countries may show acceleration and deceleration in their rate of decline, the behaviour of the mortality 'frontier' is suggested to be more regular. As noted by Bijak (2004), Torri and Vaupel (2012) and Pascariu, Canudas-Romo, and Vaupel (2018), this regularity has utility in forecasting, to the extent that we expect advances in health behaviour and medical technology to keep pace with past experience. This paper employs the Bayesian generalised additive mortality model of Hilton et al. (2019) to estimate frontier mortality rates and project them forward at the long run rate of decline, modelling individual country mortality schedules as deviations from this frontier experience.

2 Model Specification

The model presented in this paper employs Generalised Additive Models (GAMs) (Wood 2006) to capture both the frontier mortality surface and deviations from it. GAMs model target quantities as sums of smooth functions of covariates, with identifying constraints ensuring such smooths are distinguishable. Hilton et al. (2019) describe a model for mortality forecasting using GAMs. The model proposed in this paper extends this approach to provide for the inclusion of a mortality frontier. The log mortality rate $\log(m_{xt})$ is modelled as a sum of frontier mortality term $f(x,t)$, a country specific term $g^{+}(x,t,c)$ that is constrained to be positive (ensuring that all country rates lie above the frontier), and a period effect k_{tc}. For the frontier term, smooth functions of age are used to capture the overall pattern of frontier log-mortality $s_{\mu}(x)$ and the age-specific pattern of mortality improvement factors $s_{\beta}(x)$, assuming that frontier mortality declines linearly. The country-specific term is considered to be a product of a smooth positive term $s^{c}_{\gamma}(x)$ describing age-specific deviations from the frontier, and an additional term $\exp(h(x,t,c))$ which describes changes in this deviation over time. The exponent in this factor ensures that the overall country specific term remains positive

$$\begin{aligned} \log(m_{xtc}) &= f(x,t) + g^{+}(x,t,c) + \kappa_{tc} \\ f(x,t) &= s_{\mu}(x) + s_{\beta}(x)t \\ g^{+}(x,t,c) &= s^{c}_{\gamma}(x)\exp(h(x,t,c)). \end{aligned} \tag{1}$$

The function $h(x,t,c)$ describing changes at the level of individual countries can potentially take a number of different forms. As a starting point, we consider $h(x,t,c)$ to comprise a single smooth age term interacting with time $h(x,t,c) = s^{c}_{\delta}(x)t$. Thus, the term $s^{c}_{\gamma}(x)$ can be interpreted as the level of deviation from the frontier at time $t = 0$, and the $s^{c}_{\delta}(x)$ term controls the rate of decline or increase of this deviation. The pace of change with respect to time slows as the term $g^{+}(x,t,c)$ tends to zero, so that country specific rates approach the frontier only asymptotically. However, this model assumes that particular age-specific mortality rates either converge to or diverge from the frontier for particular countries; the direction of change cannot reverse. The introduction of a quadratic term $s^{c}_{\lambda}(x)t^2$ rectifies this problem, so that $h(x,t,c) = s^{c}_{\delta}(x)t + s^{c}_{\lambda}(x)t^2$. More varied patterns of deviations from the frontier can be considered by allowing more flexibility in the specification of $h()$. Any number of combinations of age, period and even cohort terms may be included, as long as these are sufficiently constrained so that the other terms in the model are identifiable. All smooth terms are modelled using penalised B-splines (Wood (2006)), while the prior distribution on the basis function coefficients of $s^{c}_{\gamma}(x)$ pull the country-specific deviations toward zero, in effect ensuring that the frontier remains close to the lowest observed mortality rates at each age. The period effect k_{tc} is a country specific random walk capturing year-to-year random variation in mortality caused by factors such as flu and temperature variations.

3 Data and Results

The Human Mortality Database (Human Mortality Database 2019) was used to obtain age-specific death and exposure data for 19 developed countries with reasonably large populations and for which data is available for at least the period 1961 onward. Only female data are used in this instance and infant mortality and centenarians were excluded. Data from 1961-2006 is used to fit the three models: the linear and quadratic variants of the proposed model and comparator model where each country is fitted independently. Data from 2007-2016 held back for purposes of assessment. Data from 2007-2016 held back for purposes of assessment.

In this section, model results are presented for the quadratic model variant. Starting with the frontier model, Figure 1 shows the posterior distribution of the frontier surface defined by $s_{\mu}(x) + s_{\beta}(x)t$ at selected years. These distribution are plotted together with corresponding empirical log rates for the 19 countries included in the estimation processes. Each country is displayed in a different colour, although distinguishing individual country's observation is not important for interpretation of the chart. The frontier estimates lie below but close to the vast majority of observed rates. At younger ages, some observations lie beyond the frontier. This is to be expected, as the estimated frontier is supposed to represent the lower limit of the central rate $m_{x,t}$, but it does not account for the additional negative binomial uncertainty in deaths. The final panel in Figure 1 is a forecast for 2016. Again observations for the majority of the age range appear consistent with our interpretation of the frontier, although it is possible that decline in the frontier for young adults aged 20-30 is slightly under-estimated by the model.

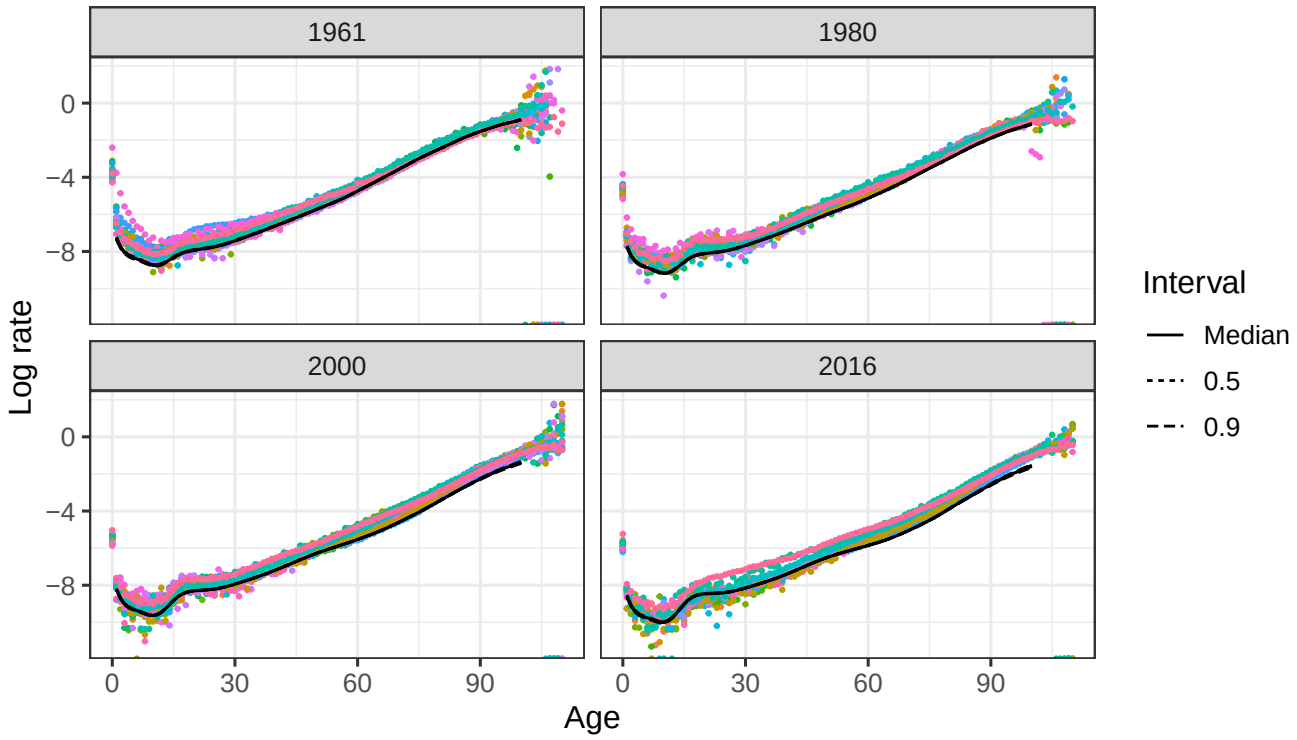

Figure 1: Posterior distribution of frontier mortality, selected years. Plotted data points represent all observations in a given year; colours denote countries.

A key question is how effectively the model can fit observed data and predict future trends in mortality. For illustrative purposes, we display posterior distributions for particular age-specific rates across time for England and Wales in Figure 2. Empirical rates are plotted as red dots, while the beginning of the forecast period is indicated by a black horizontal line. The posterior mean for each age-specific rate lies above frontier mortality boundary. Most empirical observations lie within the 90% credible interval, both over the fitting period and for the forecasts, indicating the model does a reasonable job at capturing our uncertainty about the data.

4 Discussion

Estimates of frontier mortality and the extent of particular country deviations from this standard may provide useful benchmarking information to public bodies. Additionally, although not presented here, the model was fitted jointly to 19 countries, and its performance in short-term forecasting is compared to a similar model without a frontier component in which each country was modelled independently. The frontier model was found to perform better in terms of the accuracy of its central forecasts than the independence model over a 10-year time horizon. These findings suggest that a frontier model has potential for use in forecasting mortality.

Future investigation will conduct forecasts over a longer time horizon and extend the approach to multiple sexes using a 'double-gap' model, as employed by Pascariu, Canudas-Romo, and Vaupel (2018) for life expectancy. Finally, more flexible models to describe the evolution of country specific deviations from the frontier will also be explored.

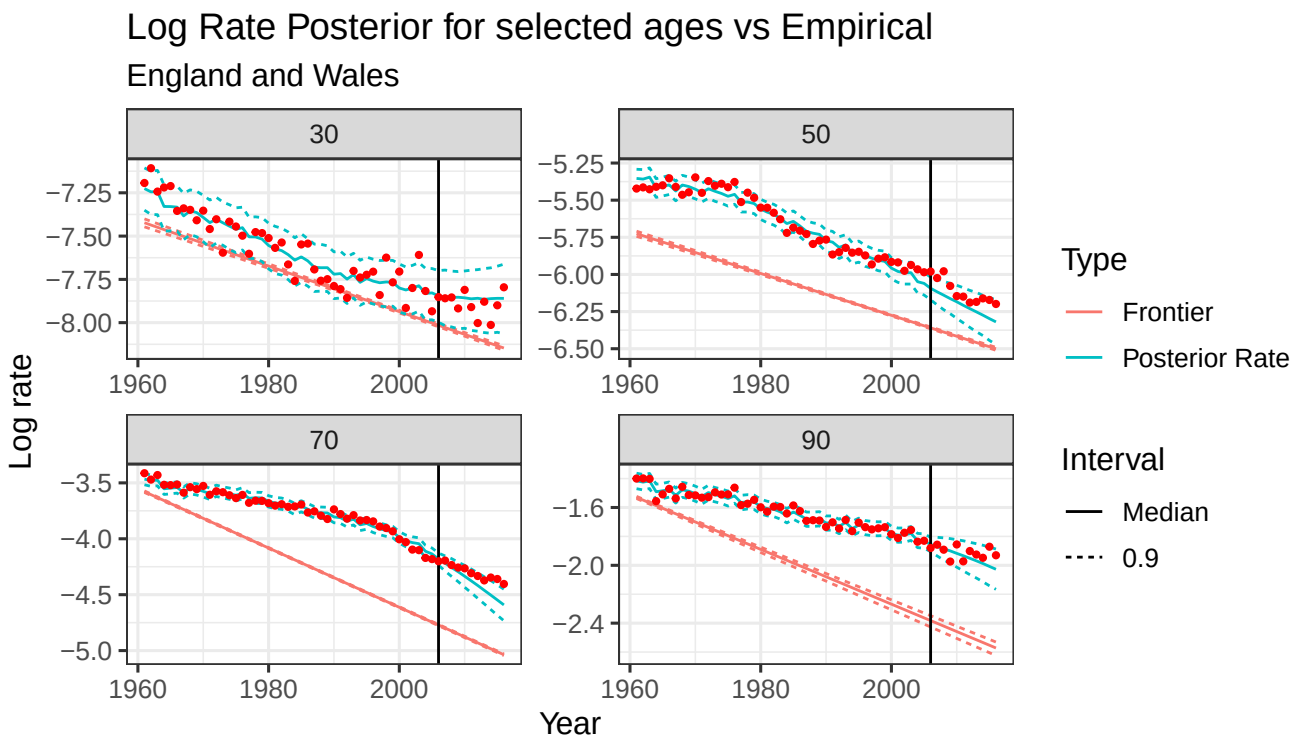

Figure 2: Posterior predictive distribution of log-mortality rates for selected ages, England and Wales

References

Bijak, Jakub. 2004. "Fertility and mortality scenarios for 27 european countries, 2002-2052." CEFMR Working Paper. Warsaw: Central European Forum for Migration Research.

Hilton, Jason, Erengul Dodd, Jon Forster, and Peter W F Smith. 2019. "Projecting UK Mortality using Bayesian Generalised Additive Models." *Journal of the Royal Statistical Society. Series C* 68 (1): 29–49. doi:https://doi.org/10.1111/rssc.12299.

Human Mortality Database. 2019. "Human Mortality Database." University of California, Berkeley (USA),; Max Planck Institute for Demographic Research (Germany). http://www.mortality.org/cgi-bin/hmd.

Oeppen, Jim, and James W Vaupel. 2002. "Broken Limits to Life Expectancy." *Science* 296 (5570): 1029–31. doi:10.1126/science.1069675.

Pascariu, Marius D., Vladimir Canudas-Romo, and James W. Vaupel. 2018. "The double-gap life expectancy forecasting model." *Insurance: Mathematics and Economics* 78 (2018). Elsevier B.V.: 339–50. doi:10.1016/j.insmatheco.2017.09.011.

Raftery, Adrian E., Jennifer L. Chunn, Patrick Gerland, and Hana Ševčíková. 2013. "Bayesian Probabilistic Projections of Life Expectancy for All Countries." *Demography* 50 (3): 777–801. doi:10.1007/s13524-012-0193-x.

Torri, Tiziana, and James W. Vaupel. 2012. "Forecasting life expectancy in an international context." *International Journal of Forecasting* 28 (2). Elsevier B.V.: 519–31. doi:10.1016/j.ijforecast.2011.01.009.

Wood, Simon N. 2006. *Generalised Additive Models: An Introduction with R*. Boca Raton: Chapman; Hall / CRC Press.