Economic Commission for Europe

Conference of European Statisticians

Work Session on Demographic Projections

Belgrade, 25–27 November 2019
Item No. 9 of the provisional agenda
Assumptions on mortality

Modelling frontier mortality using Bayesian generalised additive models[1]

Note by Jason Hilton[1], Erengul Dodd[1], Jonathan J. Forster[2], Peter W.F Smith[1],
[1] Centre for Population Change, University of Southampton,
[2] Department of Statistics, University of Warwick

Summary

Mortality rates differ across countries and years, and the country with the lowest mortality has historically tended to change over time. Following the classical *Science* paper by Oeppen and Vaupel (2002), a hypothetical mortality 'frontier' can be defined, representing the lowest set of mortality rates possible for each year. It is expected that change in this frontier reflects global technological and medical advances, which may display a more stable trend over time than the patterns in mortality improvement displayed by any particular country. This paper presents a method that aims to take advantage of this assumed stability for forecasting purposes.

Adapting the model of Hilton et al. (2019), we set out a model that jointly estimates this frontier mortality as well mortality rates for individual countries. Generalised additive models are used to estimate a smooth set of baseline frontier mortality rates and mortality improvements, and country-level mortality is modelled as a set of smooth, positive deviations from this, forcing the mortality estimates for individual countries to lie above the frontier.

This model is fitted to data for a selection of countries from the Human Mortality Database using Bayesian methods. The efficacy of the model in forecasting over a 10-year horizon is compared to a similar model fitted to each country separately.

I.   Introduction

1. 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. Best-practice life expectancy, defined as lowest value of life expectancy globally, has shown sustained increases over many decades (Oeppen and Vaupel 2002), and furthermore national life expectancies in different states appear to be converging (Wilson 2001). The extent to which we can expect these trends to continue in the long term is subject to debate. However, as highlighted by Oeppen and Vaupel (2002), previous predicted limits to life expectancy have been surpassed not long after they were proposed. 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. In contrast to the social, political and economic influences on mortality decline, the frontier is influenced to a greater extent by the pace of increase in medical knowledge and technological advances, which is assumed to be more consistent. As noted by Bijak (2004), Torri and Vaupel (2012) and M. D. 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. Furthermore, estimating the location and evolution of the global mortality frontier may help policy-makers understand where there are opportunities to improve mortality. 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.

II.   Model of Mortality

2. There are various different approaches to the modelling of mortality, of which Booth and Tickle (2008) provides an extensive review. Mortality is the demographic component most amenable to forecasting; unlike migration and fertility, both the age pattern of the rates and the direction of change has remained steady over a very long time horizon. A few key approaches to mortality forecasting are highlighted in this section. One strand of the literature is based on the idea of reducing the dimensionality of the problem by identifying leading principle components of the matrix of log-mortality rates and using these for forecasting. The seminal paper in this area is Ronald D Lee and Carter (1992). Their method decomposes the centered log-mortality rates into a time index describing the overall rate of mortality decline and a vector of age-specific factors describing the rate of decline of each age-specific rate relative to this index, so that that $\log(m_{xt}) = a_x + b_x k_t$. The vectors $\mathbf{b} = (b_0,\ b_1,\ \ldots,\ b_X)$ and $\mathbf{k_t} = (k_1,\ k_2,\ \ldots,\ k_T)$ correspond to the first principal component of the centred log-rate matrix, and can therefore be estimated using singular value decomposition. Since only the index κ_t varies over time, the forecasting problem is much simplified. Typically, simple time series models suffice for κ_t, and in particular the random walk with drift has been found to perform well. A wide range of extensions of the Lee-Carter model have been proposed, a testament to the simplicity and efficacy of the model (e.g. R D Lee and Tuljapurkar (1994), Booth et al. (2006), Li, Lee, and Gerland (2013)). R. J. Hyndman and Ullah (2007) provide an extension of the Lee-Carter model from within the functional data analysis framework, allowing for more than one principal component to be employed in forecasting, and for the smoothing of the age-profile of mortality decline.

3. From a different perspective, Currie, Durban, and Eilers (2004) employ 2-dimensional penalised B-splines to capture log-mortality rates, allowing considerable flexibility in the shape of the mortality surface. Forecasting is possible through the interpretation of the

smoothing penalisation of basis function coefficients as a time series model, allowing basis function coefficients for new periods to be generated. Also employing penalised B-splines, Hilton et al. (2019) fit generalised additive models in order to capture smooth age, age-specific improvement, and cohort components together with a period effect capturing deviations from the linear trend (for which roughness is deemed appropriate). Taking a more general view, A. J. G. Cairns et al. (2009) describe a family of models in which log-mortality is considered as a sum of terms of age, period and cohort effects, possibly including interactions. This family includes the Lee-Carter model and the model of Currie, Durban, and Eilers (2004) as special cases.

4. As well as attempting to model mortality directly, one can attempt to specify a model that describes how the mortality frontier evolves, and describe how far behind this frontier each individual country is. Bijak (2004) provide fertility and mortality forecasts for 27 European countries using a mortality model based on the assumption that frontier life expectancy increases linearly, and that individual countries converge exponentially toward the frontier with different rates of convergence for males and females. Similarly, Torri and Vaupel (2012) model both frontier life expectancy and the gap between such life expectancy and that of individual countries. The gap is modelled using a logarithm transform to ensure countries always remain below the frontier, and various time-series models are applied to the gap, including the discrete geometric Brownian motion and the discrete geometric mean-reverting process. M. D. Pascariu, Canudas-Romo, and Vaupel (2018) present a 'two-gap' mortality model, which considers both the gap between the female frontier life expectancy and the equivalent value for any particular country, and the gap between female and male life expectancy in that country, allowing for coherence both between and within countries. M. P. Bergeron-Boucher et al. (2018) are concerned with the gap between male and female mortality, and provide a model which constructs a forecast of female mortality, and then separately forecast male-female mortality ratios. These papers provide ample evidence of the potential efficacy of thinking about mortality forecasting in terms of a mortality frontier. The model presented in this paper differs from these approaches in that it attempts to estimates a smooth frontier mortality profile at the level of age-specific rates, based on all available data, and jointly estimates positive deviations from this frontier in a Bayesian hierarchical framework.

III.   Model Specification

5. 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 logarithm of mortality rates are considered as a smooth function of age and cohort, together with smooth age-specific improvement factors and non-smoothed period effects. Smooth terms are modelled using penalised B-splines (Wood 2006). The model proposed in this paper extends this approach to provide for the inclusion of a mortality frontier. For the sake of simplicity, cohort effects included in the model of Hilton et al. (2019) are jettisoned in order to simplify the development of the model, and an extension of the model could allow their re-inclusion.

6. Starting from the likelihood, age-specific death counts D_{xt} are given a negative binomial distribution, with a parameter $\exp(\phi)$ determining the degree of over-dispersion relative to the Poisson:

$$D_{xt} \sim \text{Negative Binomial}(m_{xt}R_{xt}, \exp(\phi)).$$

7. The log mortality rate $\log(m_{xt})$ is then 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 $s_\gamma^c(x)$ 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

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

8. 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 $h(x,t,c) = s_\delta^c(x)t$. Thus, the term $s_\gamma^c(x)$ can be interpreted as the level of deviation from the frontier at time $t = 0$, and the $s_\delta^c(x)$ 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 $s_\lambda^c(x)t^2$ rectifies this problem, so that $h(x,t,c) = s_\delta^c(x)t + s_\lambda^c(x)t^2$.

9. 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. Two particular special cases may be important. Firstly, we might allow for variations in the pace and direction of mortality change by incorporating the bi-variate form of Lee and Carter (1992), so that $h(x,t,c) = s_\delta^c(x)k_{tc}$. In this case, we would no longer include the period term κ_{tc}, as its function would be subsumed by the new k_{tc} term. The usual Lee-Carter constraints would be required to ensure identifiability. Secondly, an even greater degree of flexibility might be provided by including a two-dimensional spline term $h(x,t,c) = s_\eta^c(x,t)$, in the spirit of the model of Currie, Durban, and Eilers (2004). Again, constraints would be required for in order identify such effects.

10. All smooth terms are modelled using penalised B-splines (Wood (2006)). The term $s_\gamma^c(x)$ is treated slightly differently however; its coefficients are constrained to be positive, ensuring that the smooth term as a whole is positive everywhere, as all elements of the basis function are positive. As with other terms, the coefficient matrix has a smoothness prior applied penalising first differences in the age direction (Currie, Durban, and Eilers 2004), but also double exponential random effect priors. The later prior pulls country-specific deviations toward zero, in effect ensuring that the frontier remains close to the lowest observed mortality rates at each age.

11. 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. In order to ensure that the overall time-trends are captured in the other model parameters, the κ term is constrained so that it sums to zero, and contains no linear or quadratic components. The random walk prior is adjusted to account for these constraints in a similar way to Hilton et al. (2019). In the examples that follow, period effects of different countries are considered independent, although the prior correlation structure could be specified in greater detail, allowing different levels of correlation between countries, or accounting for geographical or social-cultural factors that might induce correlation between mortality rates across countries.

IV.   Data and Estimation

12. 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; future work could plausibly consider modelling males jointly by extending the 'double-gap' life-expectancy model of M. D. Pascariu, Canudas-Romo, and Vaupel (2018) to a mortality rate context. Infant mortality and centenarians were excluded, although extending the model to incorporate these age groups should be possible. 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.

13. The frontier and country-specific elements of the models were fitted jointly using the `stan` Bayesian modelling software (Stan Development Team 2019). Each model run consisted of four chains, each consisting of 8000 iterations, with the first half of each chain used to optimise the relevant sampling parameters and discarded, and additionally the remaining samples where thinned by a factor of two, to reduce memory usage. Diagnostic measures suggested that each chain had converged to the target distribution.

V.   Results

14. 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. In other words, although the force of mortality will generally lie above the frontier, random variation in realised death counts could result in observed rates that lie below it. Thus, the empirical mortality frontier is distinct from the 'true' mortality frontier that we are trying to model. Younger ages are more likely to display this effect, because mortality is much lower at these ages, and so the effect of negative binomial uncertainty on observed log-rates is far greater.

15. It should also be noted that unlike the country-specific deviations, the period effect for particular years $\kappa_t c$ may be negative, and in some cases this may result in modelled mortality rates that lie below the frontier. Given that the scale of the period effects is generally small relative to the deviations, this will only occur for countries that are already very close to the frontier, and is not deemed to be a significant shortcoming in the model specification.

16. 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.
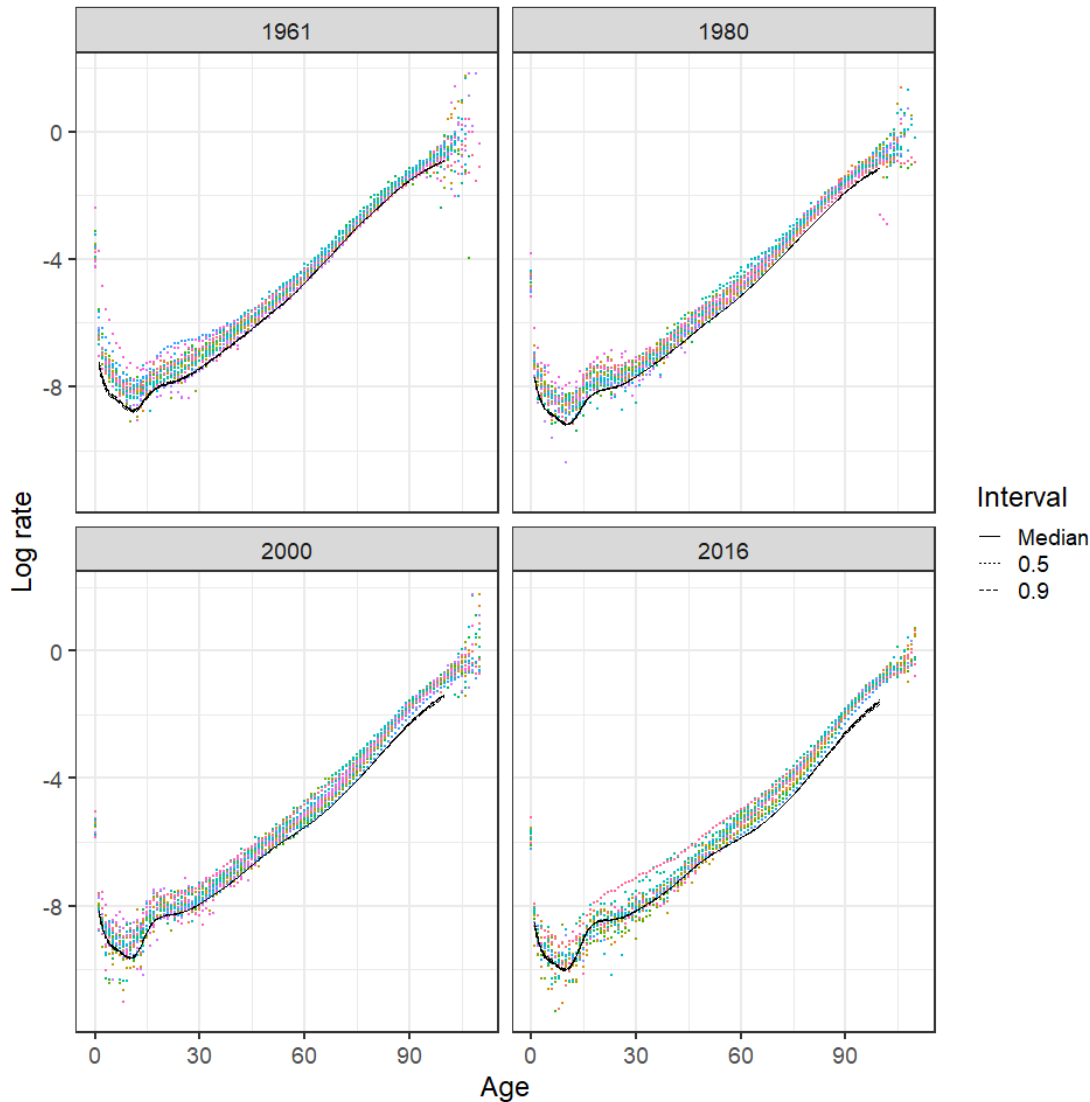
17. A key question is how effectively the model can fit observed data and predict future trends in mortality. For illustrative purposes, we display posterior distributions for particular age-specific rates across time for England and Wales in Figure 2. Empirical rates are plotted as red dots, while the 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. There is some evidence that our forecasts are overly optimistic about the extent to which mortality for England and Wales will decline towards the frontier around age 70; here the last few observations fall outside the predictive interval.

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

Log Rate Posterior for selected ages vs Empirical
England and Wales
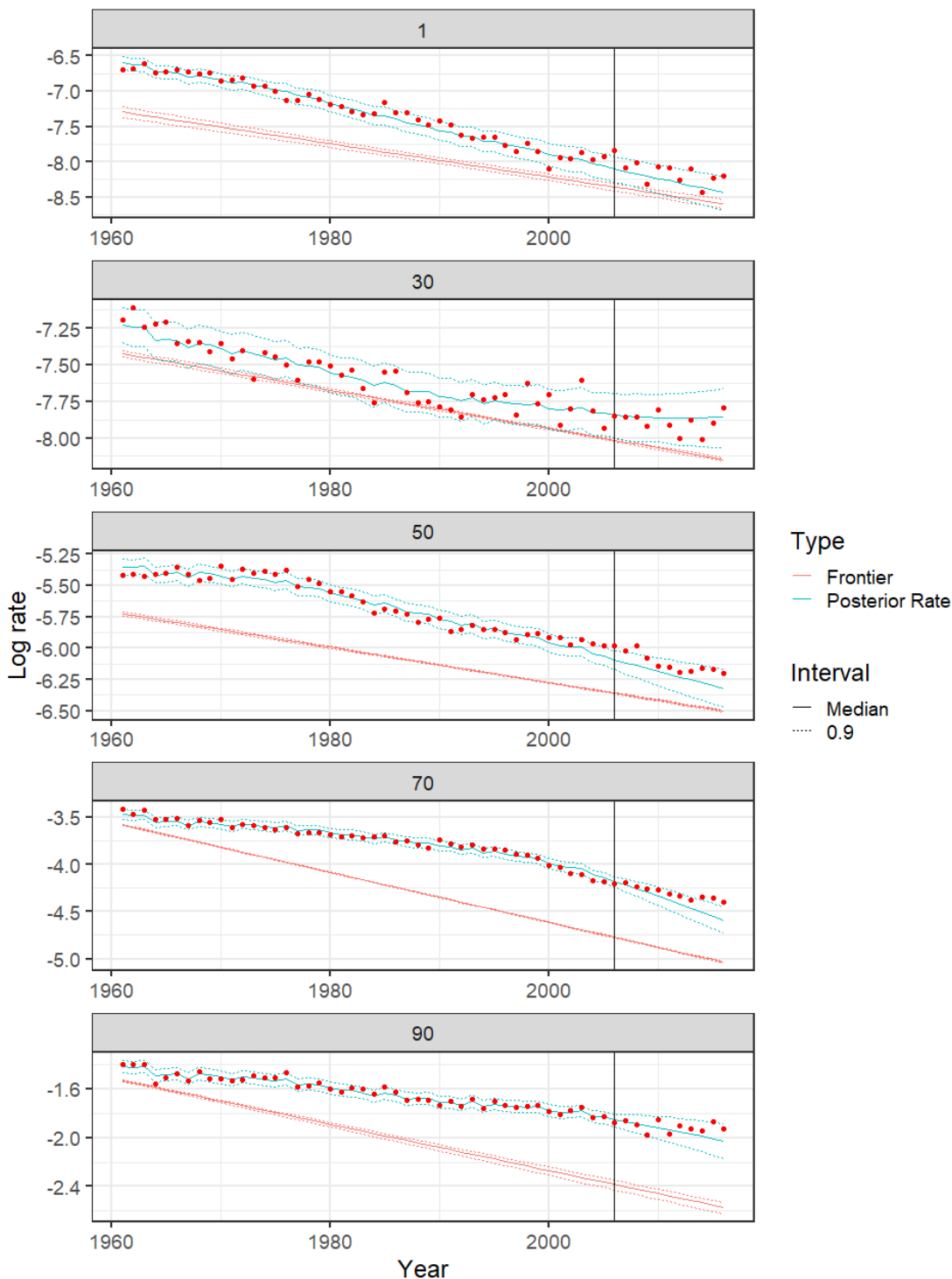
18. Of course, a more thorough examination of the model is needed to decide its efficacy. Extensive plots for all countries can be found in the supplemental material. It is evident that for the quadratic model in particular, some countries display unrealistic forecasts at particular ages; the cause and potential remedy to this issue is discussed in Section 6. For the purposes of formal assessment, root-mean squared error (RMSE) and empirical coverage (the proportion of observations falling within the posterior interval of a given probability) were calculated over the forecast period 2007-2016 for all countries. RMSE was calculated using the mean of the posterior rate for each forecast year and age as the relevant point estimates. One goal of the assessment is to provide evidence that including information about the frontier is useful for forecasting. To this end a series of models were fitted to each country independently which included only smooth age, age-specific improvement, and period terms:

$$\log(m_{xt}) = s_\mu(x) + s_\beta(x)t + \kappa_t.$$

19. Thus, we can compare the forecast performance of the model in which country forecasts are independent (labelled 'Independent' in subsequent plots) with variants of the frontier model we are proposing. Specifically, we investigate two different choices of the $h(x, t, c)$ function determining the change in country mortality relative to the frontier:

$$h_1(x, t, c) = \quad s_\delta(x)t$$
$$h_2(x, t, c) = \quad s_\delta(x)t + s_\lambda(x)t^2.$$

20. These are referred to as the *linear* and *quadratic* models respectively. To give a clear idea of the whether these variants are doing better than the comparator independence model, Figure 3 displays the difference between RMSE for the variants and the independence models for each country. If this value is negative (to the left of the axis at zero in the chart), it indicates that the variant model performs better. If it is positive, the reverse is true. The assessment reveals that for 13 of the 19 countries, the quadratic model has lower a RMSE over the forecast period than the independent model. For the linear model, the results are closer: it is preferred by this metric over the independence model in 11 of 19 cases.

Figure 3: Difference between RMSE of frontier model variants and a similar model fitted independent to each country
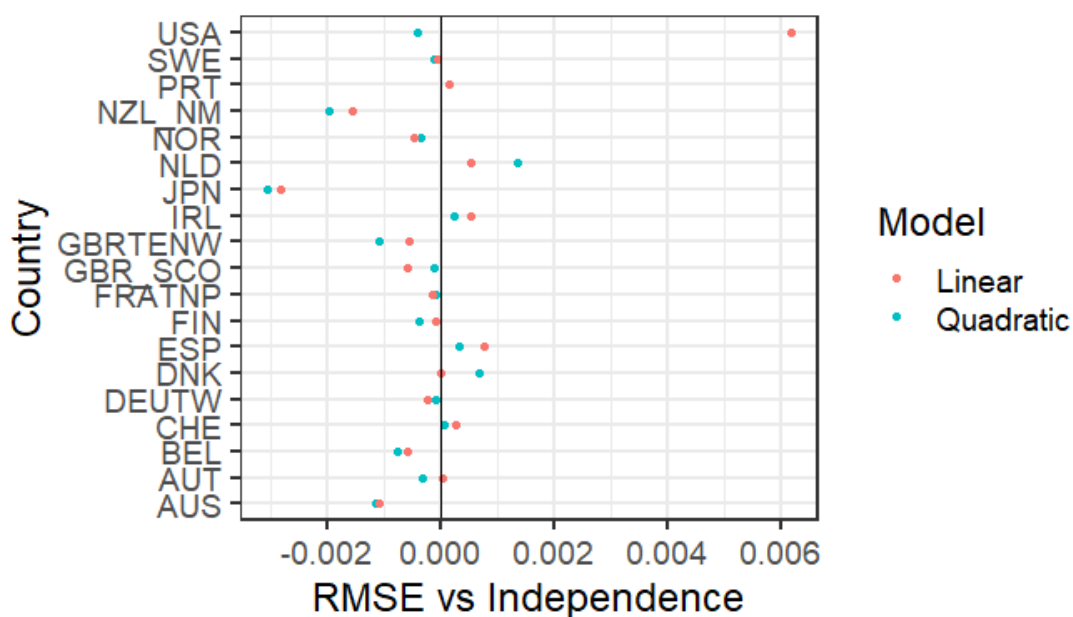
21. The accuracy of point estimates are not the only relevant area of assessment. Quantification of uncertainty in forecasts is important in managing longevity risk, and so the extent to which observations fall within forecast intervals is also important. Figure 4 provides the proportion of observations that fall within the central 90% predictive probability interval. Ideally, this value should approach 90%, indicating forecast uncertainty appears well calibrated. However, given that for each country we only observe one correlated set of rates (over the period 2006-2016), this proportion does not correspond exactly with the frequentist interpretation of coverage, which relies on independent replications of the same experiment. Therefore, we must not over-interpret the reported empirical coverage statistics. In general, the results are encouraging. A majority of all models have empirical coverages ranging between 80% and 95% for the 90% interval. The quadratic model has 5 observations with coverages below 80%, compared to 5 for the independent model and 7 for the linear variant. The USA, Denmark and Spain appear to have patterns of recent mortality decline which are difficult to capture for all models. The quadratic model appears to be the better performing model overall based on these metrics, although it appears to perform particularly badly for both RMSE and coverage in the case of the Netherlands.

Figure 4: Proportion of observations falling within 90% predictive interval for the independent model, and linear and quadratic variants of the frontier model.
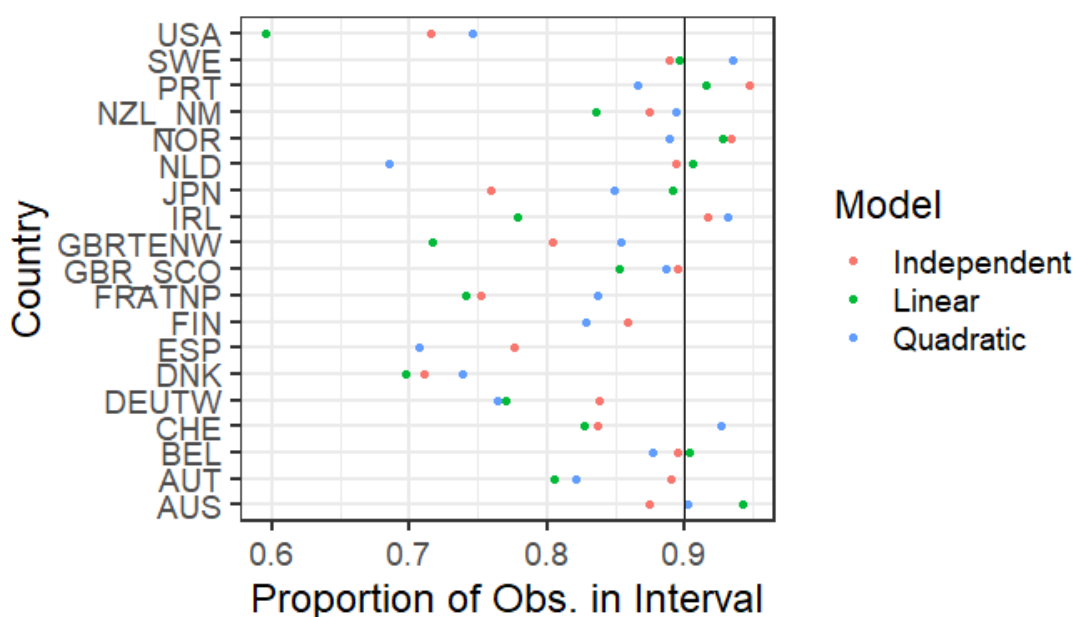
VI.   Conclusion

22. This paper has set out a model of mortality that estimates the evolution of frontier mortality as a set of smooth rates, and then considers individual countries as deviations from this profile. Frontier mortality is constrained to lie below the modelled force of mortality for all individual countries, but the prior specification ensures that it remains close to best-performing countries by penalising the magnitude of the individual country deviations. Estimates of frontier mortality and the extent of particular country deviations from this standard may provide useful benchmarking information to public bodies. 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 for a large group of countries, perhaps particularly by multinational bodies with access to harmonised data from a variety of sources.

23. Some limitations and areas for future investigation can be identified. Firstly, a longer time horizon may be required to accurate assess the usefulness of the model. Mortality forecasts are typically used to compute cohort life expectancies, which require considerably longer forecasts than have been provided here. Secondly, forecasts for females only were produced in the examples above. Extending the approach described to multiple sexes using a 'double-gap' model, as employed by M. D. Pascariu, Canudas-Romo, and Vaupel (2018) for life expectancy, may have some utility. Thirdly, at present simple linear and quadratic terms were chosen to describe the evolution of country specific deviations from the frontier. These may not be the best choices for this element of the model. In particular, over longer time horizons, the quadratic model may predict unrealistic divergences from the frontier at some ages in countries where recent stagnation in mortality rates have been observed, leading in some cases to predicted increases in mortality. Section 3 set out two possible alternative models based on Ronald D Lee and Carter (1992) and Currie, Durban, and Eilers (2004). Specifying priors on the time-varying elements of these models that favour mean-reversion will help to ensure forecast means do not diverge from the frontier over the long-term. Finally, a comparison between frontier models and those that provide for convergence towards a mean trend might be investigated; it may be that such models produce similar conclusions, or that one or another is more efficacious.

VII.   References

Bergeron-Boucher, Marie Pier, Vladimir Canudas-Romo, Marius Pascariu, and Rune Lindahl-Jacobsen. 2018. "Modeling and forecasting sex differences in mortality: a sex-ratio approach." *Genus* 74 (1). Genus. doi:10.1186/s41118-018-0044-8.

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

Booth, Heather, and Leonie Tickle. 2008. "Mortality Modelling and Forecasting: A Review of Methods." *Annals of Acturial Science* 3 (1-2): 3–43.

Booth, Heather, Rob Hyndman, Leonie Tickle, and Piet de Jong. 2006. "Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions." *Demographic Research* 15 (9): 289–310. doi:10.4054/DemRes.2006.15.9.

Cairns, Andrew J. G., David Blake, Kevin Dowd, Guy D. Coughlan, David Epstein, Alen Ong, Igor Balevich, D Coughlan, David Epstein, and Alen Ong. 2009. "A Quantitative Comparison of Stochastic Mortality Models Using Data from England and Wales and the United States." *North American Actuarial Journal* 13 (March): 1–35. doi:10.1080/10920277.2009.10597538.

Currie, Iain D, Maria Durban, and Paul H C Eilers. 2004. "Smoothing and forecasting mortality rates." *Statistical Modelling* 4: 279–98.

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.

Hyndman, Rob J., and Md. Shahid Ullah. 2007. "Robust forecasting of mortality and fertility rates: A functional data approach." *Computational Statistics and Data Analysis* 51 (10): 4942–56. doi:10.1016/j.csda.2006.07.028.

Lee, R D, and S Tuljapurkar. 1994. "Stochastic population forecasts for the United States: beyond high, medium, and low." *Journal of the American Statistical Association* 89 (428): 1, 175–89. http://www.jstor.org/stable/10.2307/2290980.

Lee, Ronald D, and Lawrence R Carter. 1992. "Modeling and Forecasting U.S Mortality." *Journal of the American Statistical Association* 87 (419). American Statistical Association: pp. 659–71. http://www.jstor.org/stable/2290201.

Li, Nan, Ronald Lee, and Patrick Gerland. 2013. "Extending the Lee-Carter method to model the rotation of age patterns of mortality decline for long-term projection." *Demography* 50 (6): 2037–51. doi:10.1007/s13524-013-0232-2.Extending.

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.

Stan Development Team. 2019. "Stan Modeling Language Users Guide and Reference Manual." http://mc-stan.org/index.html.

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.