Distr.: General
1 April 2016

English only

Economic Commission for Europe

Conference of European Statisticians

Joint Eurostat/UNECE Work Session on Demographic Projections
Geneva, 18-20 April 2016
Item 5 of the provisional agenda
Assumptions on mortality

A Comprehensive Framework for Mortality Forecasting

Note by UNIVERSITY OF SOUTHAMPTON, UK[1]

Summary

We propose a comprehensive mortality forecasting framework, which overcomes several of the limitations associated with existing approaches. Firstly, our approach accounts for lack of fit of conventional models by specifying a negative binomial error structure, correctly accounting overdispersion without over-fitting.

Secondly, the facility to impose smoothness in parameter series which vary over age, cohort, and time is integrated into the modelling process using generalised additive models (GAMs). GAMs allow parametric functions and unstructured (but smooth) functions of explanatory variables to appear in the model simultaneously. In particular, GAMs allow us to differentially smooth components, such as cohorts, more aggressively in areas of sparse data for the component concerned.

While GAMs can provide a reasonable fit for the ages where there is adequate data, estimation and extrapolation of mortality rates using a GAM at higher ages is problematic due to high variation in crude rates. At these ages, parametric models can give a more robust fit, enabling a borrowing of strength across age groups. Our forecasting methodology is based on a smooth transition between a GAM at lower ages and a fully parametric model at higher ages.

[1] [Prepared by JONATHAN J FORSTER (J.J.Forster@soton.ac.uk), ERENGUL DODD, JAKUB BIJAK, PETER W F SMITH].

I. Mortality Forecasting: A Brief Review

1. There are many studies on mortality modelling in the literature. Although, theoretical modelling started in 1725 with de Moivre and followed by Gompertz in 1825, the history of stochastic modelling is more recent (for an overview, see Tabeau et al. 2001). The first stochastic model was published by Lee and Carter in 1992. Since then a variety of mortality models have been proposed. More details on such models can be found in Booth and Tickle (2008) who give a comprehensive review on developments in mortality modelling and forecasting since the 1980s.
2. The original Lee-Carter model uses two factors, i.e. age and period, in a bilinear model for mortality rates. Various extensions of the basic Lee-Carter model have been proposed, most notably the introduction of cohort effects (Renshaw and Haberman, 2006). Smooth versions of the Lee-Carter model include approches developed by De Jong and Tickle (2006) and Delwarde et al (2007). Alternatively, linear (rather than bilinear) models with age and period as factors were investigated by Renshaw and Haberman (2003). Alternative linear structures were developed and compared by Cairns et al (2011). Currie et al (2004) proposed modelling mortality as a smooth function in two dimensions (age and time) using P-spline methodology.
3. An alternative to modelling mortality rates is to model mortality improvements. This approach is taken by the Continuous Mortality Investigation and mortality improvements by calendar year are modelled using an age-period-cohort structure (CMI WP49, 2010). Haberman and Renshaw (2012, 2013), Börger and Aleksic (2014) are other studies proposing to model mortality improvements.

II. Proposed Model: Motivation and Description

4. Let μ_{xt} denote central mortality rates at age x in year t, then we consider as the initial model specification

$$\log \frac{\mu_{xt}}{\mu_{x\,t-1}} = \alpha_x + \kappa_t + \gamma_{t-x} \qquad (1)$$

where α_x can be interpreted as a baseline annual mortality improvement at age x, κ_t as the level of mortality improvement in year t and γ_{t-x} represents cohort differences in mortality improvement since cohorts are indexed by year of birth $(t-x)$. This model requires an extra identifiability constraint, which we impose on the cohort effects as $\sum(t-x)\gamma_{t-x} = 0$. Börger and Aleksic advocate the use of this model for projecting mortality, and we also find that it has the required properties of adequately and robustly fitting the observed data.

5. Model (1) is an age-period-cohort model for log-mortality differences (mortality logratios). Note that, here, we represent mortality improvements as logratios, rather than as relative differences, where the model (1) would be expressed as

$$\frac{\mu_{xt} - \mu_{x\,t-1}}{\mu_{x\,t-1}} = \alpha_x + \kappa_t + \gamma_{t-x}\,.$$

6. For all but large mortality rates, differences between $\log\frac{\mu_{xt}}{\mu_{x\,t-1}}$ and $\frac{\mu_{xt}-\mu_{x\,t-1}}{\mu_{x\,t-1}}$ are negligible.
7. Estimation of the model parameters can be based on the Poisson log-likelihood

$$l(\theta) = -\sum_{x,t} E_{xt}\mu_{xt}(\theta) + \sum_{x,t} d_{xt}\log\mu_{xt}(\theta)$$

where θ represents the model parameters ($\alpha_x, \kappa_t, \gamma_{t-x}$), d_{xt} is the observed death count and E_{xt} the central exposed to risk at age x in year t.

8. In terms of mortality rates, model (1) can be expressed as

$$\log\mu_{xt} = \mu_{x0} + \alpha_x t + \kappa_t + \gamma_{t-x} \qquad (2)$$

where there is a straightforward correspondence between the κ_t and γ_{t-x} parameters of models (1) and (2). Later, we present parameter estimates for the parameters of (1), together with μ_{x0}.

9. Note that we have also investigated the fit of bilinear models, such as Lee-Carter-type models to mortality data. Whilst these models can provide a satisfactory fit, they have some undesirable features. In particular their parameter estimates can be sensitive to the range of years used for fitting. There is a lack of robustness of the model parameter estimates (especially the non-stationarity of the period and cohort series) to the time window used for estimation and this, in turn, affects the projections of mortality rates. These concerns do not affect the goodness-of-fit of the models (although we find our models to fit at least as well), however they do present challenges when using the models to coherently project future mortality rates. Another important consideration is that bilinear models are more challenging to estimate efficiently.
10. On the other hand, model (2) is efficient to fit, being simply a generalised linear model. Furthermore its parameter estimates seem to be robust to the time window used to fit the models. However note that under the Poisson model the variance is restricted to be equal to the mean, an assumption which is implausible for a large inhomogeneous population. A more flexible model would be a negative binomial model where the log-likelihood is

$$l(\theta,\mathrm{a}) = \sum_{x,t} \mathrm{a}\,\log\left(\frac{\mathrm{a}}{E_{xt}\mu_{xt}(\theta)+\mathrm{a}}\right) + \sum_{x,t} d_{xt}\log\left(\frac{E_{xt}\mu_{xt}(\theta)}{E_{xt}\mu_{xt}(\theta)+a}\right) + \sum_{x,t}\log\Gamma(\mathrm{a}+d_{xt}) - n\log\Gamma(\mathrm{a})$$

where a is the dispersion parameter such that the variance is $E_{xt}\mu_{xt}(\theta) + (E_{xt}\mu_{xt}(\theta))^2/\mathrm{a}$ and n is the number of positive E_{xt}.

11. Figure 1 presents the maximum likelihood estimates of the model parameters of model (1) under the Poisson distribution (black solid line) and negative binomial distribution (red solid line) for males aged between 1 and 96 using UK data for years 1961-2013.

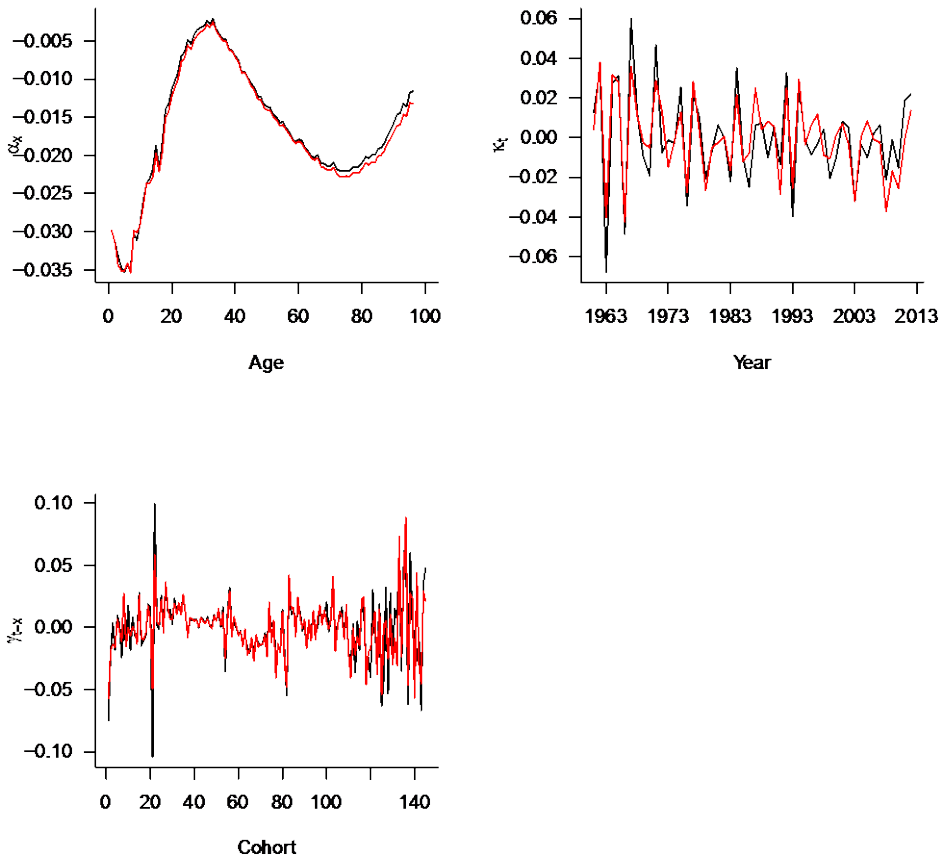


Figure 1. Maximum likelihood estimates of the parameters of model (3) under the Poisson model (black line) and the negative binomial model (red line), data for males 1961–2013

12. Goodness-of-fit of the model can be assessed by comparing its fit to an unstructured (smooth) model fitted using P-spline methodology (Currie et al, 2004) for smoothing observed mortality rates. The fit of our proposed model to the observed data should not be significantly worse. With regard to an assessment of model fit, Figure 2 presents, as a heatmap, the square of Pearson residuals from the current method used by the ONS (based on

the P-splines), and from model (2) under the Poisson distribution. Colour code for all the heatmaps is given in Appendix A.

13. It can be observed from Figure 2 that model (2) fits the data at least as well as the unstructured P-splines. Indeed, by conventional goodness-of-fit measures (residual deviance), model (2) fits better than the P-spline model, even allowing for its increased complexity in terms of the number of degrees of freedom required for parameter estimation. Model (2) seems to do a better job of estimating mortality

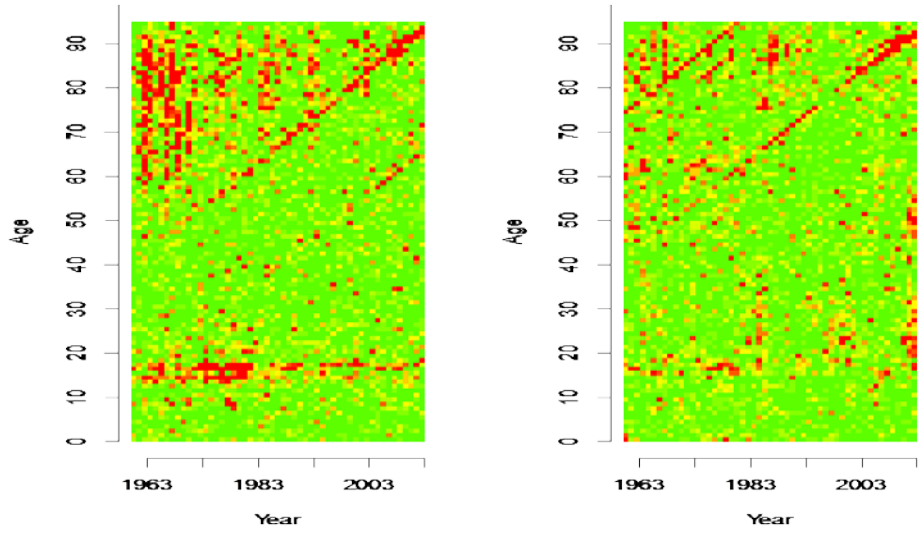


Figure 2. Comparison of residuals: the P-spline estimates (left panel) and model (2) under the Poisson distribution (right panel). For each year and age group the residual is categorised according to its absolute value and plotted with a corresponding colour ranging from green (small residuals) through to red (large residuals).

in the age range 15-20 (at the start of the "accident hump"). Both models have difficulty fitting the 1919 cohort (see Cairns et al, 2016), but arguably this cohort is of limited significance for population projection. Both models, however, fail to fit when assessed by conventional goodness-of-fit measures. Evidence for this is the large number of Pearson residuals with absolute value greater than 3. On the other hand, estimates which allow for overdispersion, either (2), fitted by maximising a negative binomial likelihood, or a P-spline fitted by quasi-likelihood produce standardised residuals within a much more acceptable range (see Figure 3). Therefore we use the negative binomial model to estimate the model parameters for the rest of our analyses.

14. One advantage of the P-spline approach is that it provides estimates of mortality rates which vary smoothly over age and time, as illustrated in

Figure 4, which plots mortality *improvements* for the Poisson model, or Figure 5, which plots the corresponding improvements with a fitting approach which accounts for overdispersion (maximum Poisson quasi-likelihood). For model (2), the maximum likelihood estimates of some of the model parameters, illustrated in Figure 3, do not vary smoothly. As a consequence the estimated mortality rates, presented in Figure 6, are also more irregular than would be desirable. However, this can be easily overcome by adopting an estimation method (penalised likelihood or Bayesian) which penalises roughness in the series of estimates for model (2).

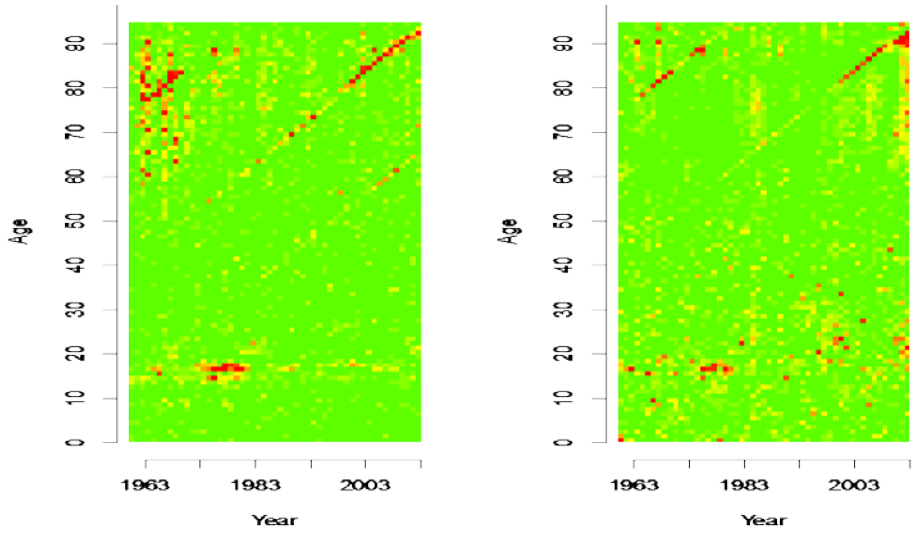


Figure 3. Comparison of residuals: the P-spline approach allowing for overdispersion (left panel) and model (2) under the negative binomial distribution (right panel). For each year and age group the residual is categorised according to its absolute value and plotted with a corresponding colour ranging from green (small residuals) through to red (large residuals).

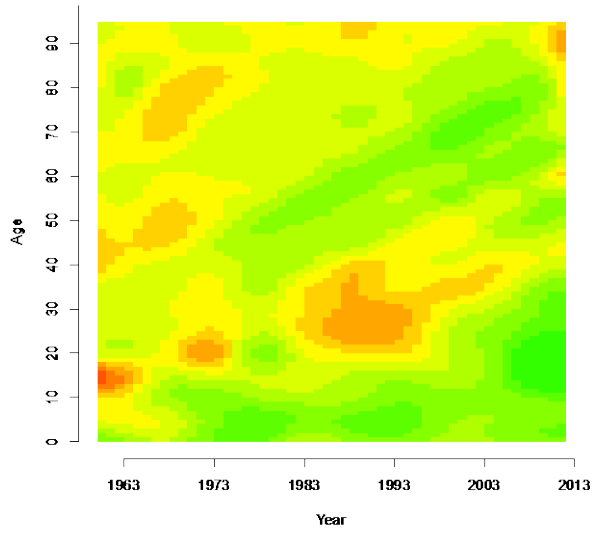


Figure 4. Heatmap of the fitted mortality improvements for the P-spline model. For each year and age group the estimated mortality improvement is categorised according to its absolute value and plotted with a corresponding colour ranging from green (large decrease) through to red (large increase).

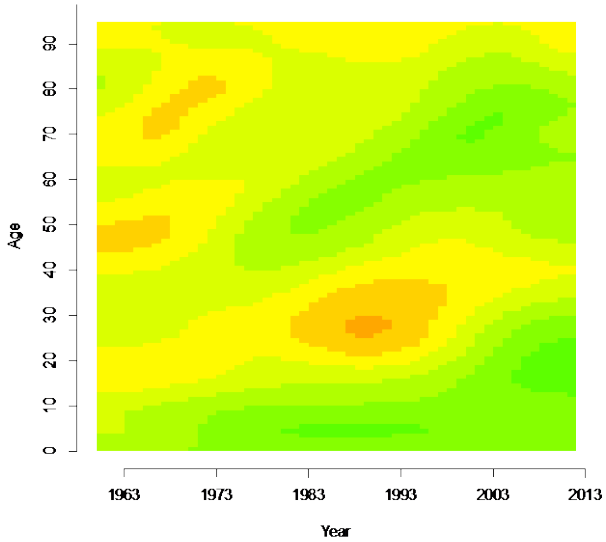


Figure 5. Heatmap of the fitted mortality improvements for the P-spline model allowing for overdispersion (penalized quasi-likelihood method). For each year and age group the estimated mortality improvement is categorised according to its absolute value and plotted with a corresponding colour ranging from green (large decrease) through to red (large increase).

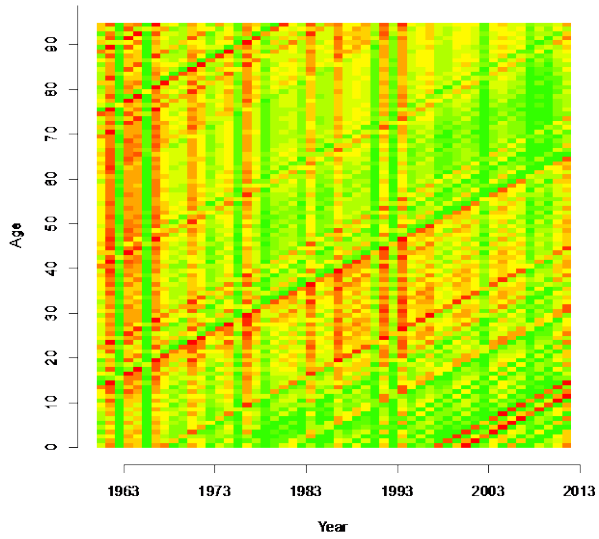


Figure 6. Heatmap of the fitted mortality improvements for model (2). For each year and age group the estimated mortality improvement is categorised according to its absolute value and plotted with a corresponding colour ranging from green (large decrease) through to red (large increase).

15. One possible way of obtaining smoother estimates is to modify (2) to the *generalised additive model*

$$\log \mu_{xt} = s_\mu(x) + s_\alpha(x)t + \kappa_t + s_\gamma(t-x). \qquad (3)$$

where s_μ, s_α and s_γ denote arbitrary smooth functions, which can be estimated by balancing goodness-of-fit to the observed data with smoothness of the corresponding function (Wood, 2006).

16. Figure 7 displays the estimates for smooth model (3) superimposed over the corresponding estimates for model (2). The estimates for model (3) are much more regular and have the desired smoothness, and the fitted mortality rates, displayed in Figure 8, are also smoother. There is an increase in residual deviance, but this is compensated by a corresponding decrease in the effective complexity of the model. Note that we choose not to smooth the period effect κ_t, and this is reflected in the vertical lines prominent in Figure 8.

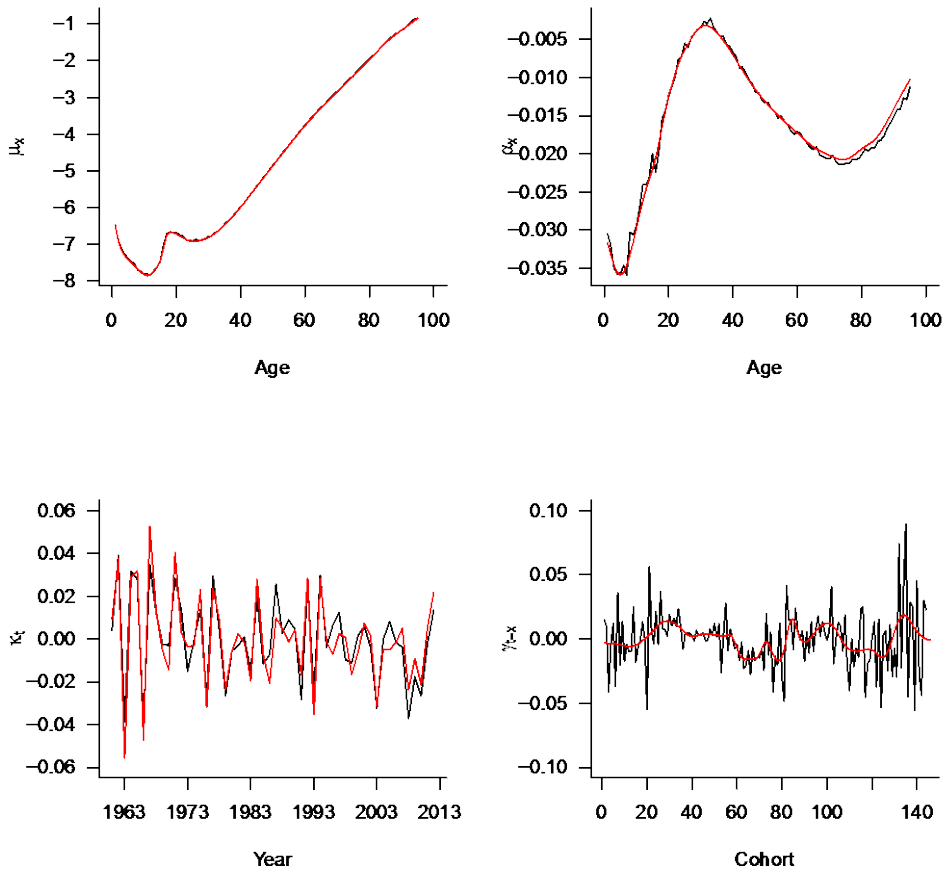


Figure 7. Estimates of the parameters of model (3), data for males 1961-2013, (red lines) superimposed over the corresponding estimates for model (2) (black lines).

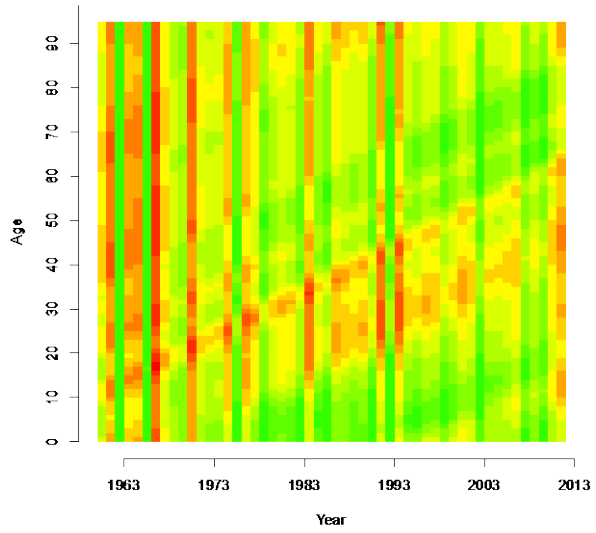


Figure 8. Heatmap of the fitted mortality improvements for model (3). For each year and age group the estimated mortality improvement is categorised according to its absolute value and plotted with a corresponding colour ranging from green (large decrease) through to red (large increase).

17. Providing point projections over any future time horizon is straightforward, based on the parameter estimates of a model such as (3). Such a projection only requires extrapolation of the time effects κ_t for future years *t*, and the cohort effects γ_{t-x} for future birth cohorts. Our research suggests that it is reasonable to set both sets of these future effects to zero (see Figure 7).
18. For the highest ages *x*, for which observed mortality experience is sparse, we recommend that the baseline mortality μ_{x0} and the age-specific mortality differences α_x are estimated by using parametric models, for example a log-linear model or a logistic model, with parameters estimated from the mortality data for the older ages. The resulting log-linear model has the form

$$\log \mu_{xt} = \mu + \mu_X x + (\alpha + \alpha_X x)t + \kappa_t + s_\gamma(t-x) \quad x > x_0 \tag{4}$$

and logistic model has the form

$$\log\left(\frac{\mu_{xt}}{\beta - \mu_{xt}}\right) = \mu + \mu_X x + (\alpha + \alpha_X x)t + \kappa_t + s_\gamma(t-x) \quad x > x_0 \tag{5}$$

where κ_t and $s_\gamma(t-x)$ is the estimates obtained from fitting (3) to the main body of data ($0 < x \leq x_0$). Dodd et al (2016) suggest that for UK mortality data, an optimal age (x_0) at which to make the transition from smooth to linear model for log-mortality is 96 for males and 100 for females, based on 2010-2012 mortality data. For the transition from smooth to logistic model, they suggest age 92 for males, 90 for females.

Note that under the logistic model (5) mortality rates flatten off, converging to a limiting rate β as x tends to infinity. We estimate β as 1.7038 for males and 2.6030 for females, values which we set as constant over time.

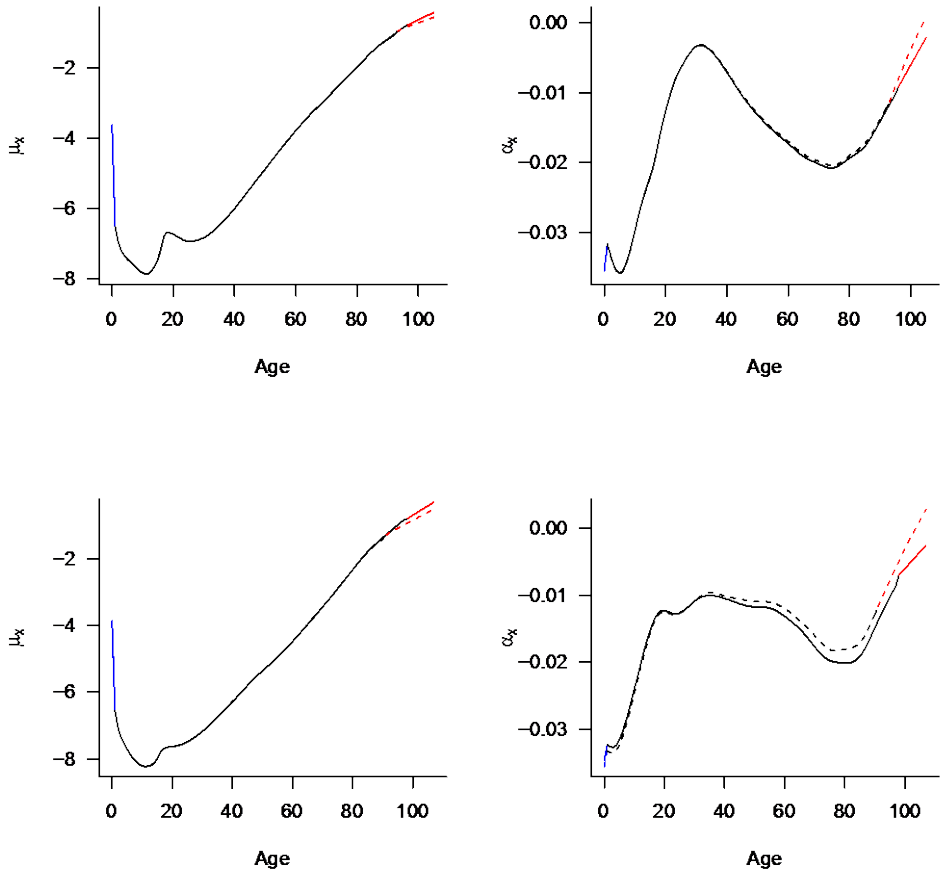


Figure 9. Estimates of the parameters of models (3), (4) and (5), UK, 1961-2013, for males (upper panels; $x_0 = 96$ for log-linear model (solid line) and $x_0 = 92$ for logistic model (dashed line)) and females (lower panels; $x_0 = 100$ for log-linear model (solid line) and $x_0 = 90$ for logistic model (dashed line))

19. The log-linear model therefore has the estimates of the baseline mortality

$$\mu_x = \begin{cases} s_\mu(x) & x \le x_0 \\ \mu + \mu_X x & x > x_0 \end{cases}$$

and mortality improvement

$$\alpha_x = \begin{cases} s_\alpha(x) & x \le x_0 \\ \alpha + \alpha_X x & x > x_0 \end{cases}$$

for both males and females. For the logistic model these estimates are

$$\mu_x = \begin{cases} s_\mu(x) & x \le x_0 \\ \log(\beta\exp(\mu+\mu_X x)/(1+\exp(\mu+\mu_X x))) & x > x_0 \end{cases}$$

$$\alpha_x = \begin{cases} s_\alpha(x) & x \le x_0 \\ \log\left(\dfrac{\beta\exp(\mu+\mu_X x+\alpha+\alpha_X x\ )}{1+\exp(\mu+\mu_X x+\alpha+\alpha_X x\ )}\right) - \log\left(\dfrac{\beta\exp(\mu+\mu_X x\ )}{1+\exp(\mu+\mu_X\ )}\right) & x > x_0 \end{cases}$$

for males and females. Figure 9 presents the estimates of the parameters under the log-linear model (solid lines) and the logistic model (dashed-lines).

20. We also treat infant (age 0) mortality separately. Here, we exclude the period effect κ_t, and fit the model

$$\log \mu_{0t} = \mu_0 + \alpha_0 t + s_\gamma(t-x) \qquad (6)$$

where $s_\gamma(t-x)$ is the estimate obtained from fitting (3) to the main body of data ($0 < x \le x_0$). Observed and fitted infant mortality using (6) are displayed in Figure 10.

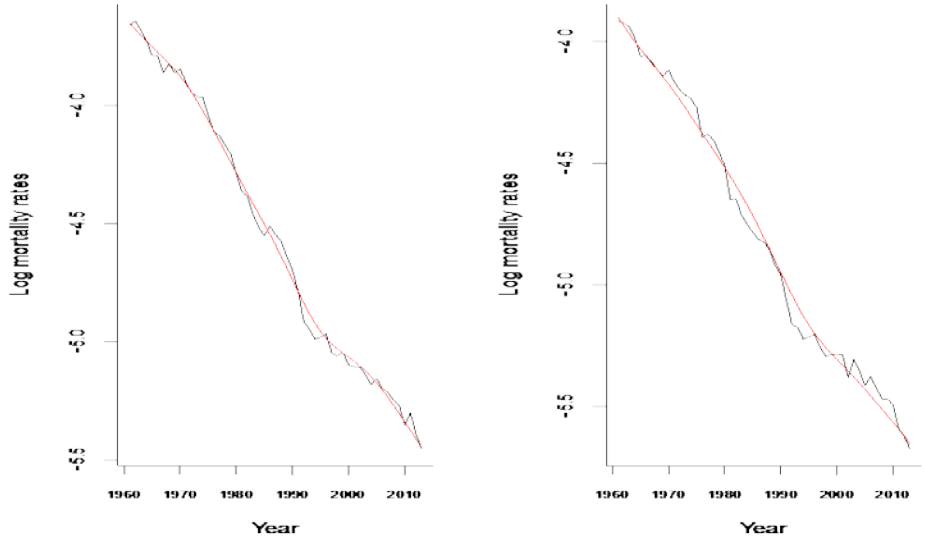


Figure 10. Estimates of infant mortality rates, UK, 1961-2013, for males (left panel) and females (right panel) using model (6; red lines), compared with observed rates (black lines)

III. Projection of mortality rates using the Proposed Method

21. We investigated the robustness of the proposed methodology by exploring the sensitivity of the estimated mortality rates in a later year to changes in the data used to estimate the model. Two different approaches were taken.

In the first, we compared the estimates of 2013 mortality rates and 2013-14 mortality improvements for model (3) fitted on ages $1 \leq x \leq x_0$ using data from 1961-2013, with the equivalent estimates fitted on 1971-2013 and 1981-2013; see Figure 11. Then, we compared the estimates of 2011 mortality rates and 2011-12 mortality using data from 1961-2011, with the equivalent estimates fitted on 1961-2012 and 1961-2013; see Figure 12. This provides a comparison with the approach based on P-splines (also included in the Figures 11 and 12). Our approach seems to be quite robust to the fitting window. Arguably, the P-spline approach is over-sensitive to mortality history in recent years.

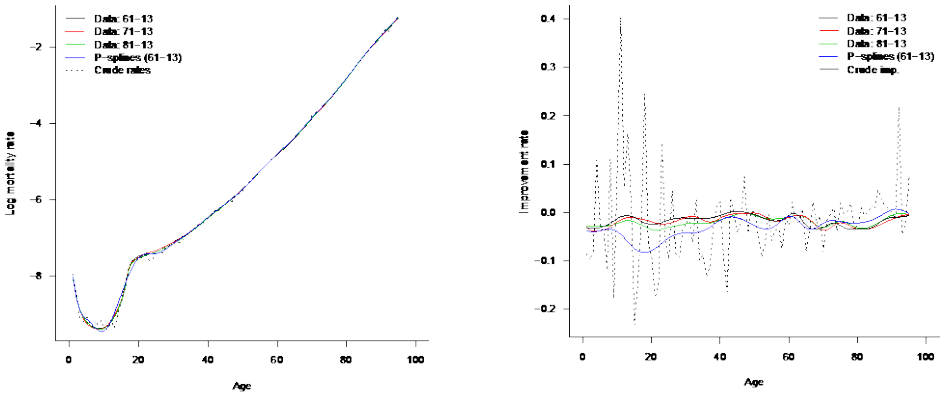


Figure 11. Estimated 2011 mortality rates (left panel), and 2011-12 mortality improvements (right panel) for males, UK, using model (3) and different historical fitting periods.

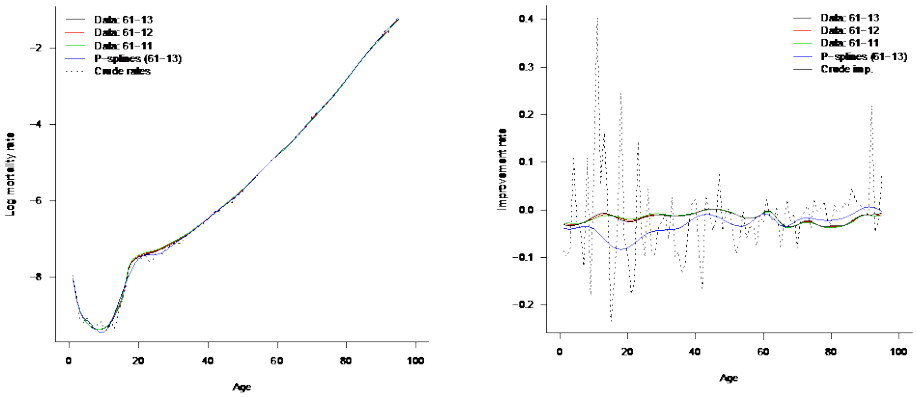


Figure 12. Estimated 2011 mortality rates (left panel), and 2011-12 mortality improvements (right panel) for males, UK, using model (3) and different recent fitting periods

22. Finally, in Figure 13, we present actual projections, for male UK mortality in 2055, based on mortality data from 1961-2013. For comparison, we also present the ONS projections and current mortality rates. There are several features deserving comment. The first is the good agreement between our model and the ONS projections in middle and older ages. In contrast, our methods forecasts a very low level of improvement for young adults. This is driven by the estimates of the α_x parameter being close to zero for ages, x in this range (see upper right panel of Figure 9). Conversely, we forecast a greater improvement than do ONS for children. The main differences between the forecasting methodologies driving these discrepancies is that we allow for a cohort effect consistently in a way in which the current ONS methodology does not. Hence, for example, observed improvements for young adults, over the observed data period, are being attributed, in our modelling, to a cohort effect.

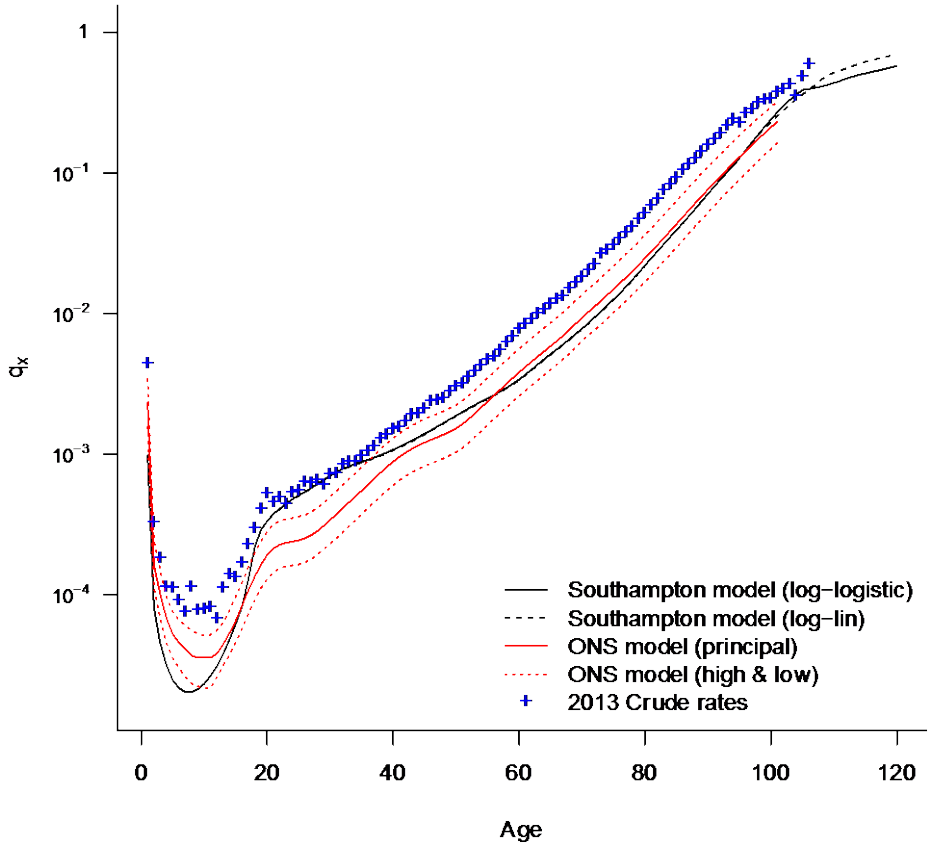


Figure 13. Forecast 2055 mortality rates and 2013 crude mortality for males, UK, together with ONS forecasts. Here the mortality rates have been transformed to the q_x scale (conditional probability of death within 12 months, given survival to age x).

IV. Conclusions

23. Mortality forecasting methods frequently rely on arbitrary and subjective inputs, and make insufficient use of the substantial data available on mortality rates, their improvement rates, and the associated variability.
24. The main features of the proposed approach include:
    - The proposed approach is comprehensive and coherent. There are trade-offs between the fit and robustness of various models that can be used.
    - The proposed models are less reliant on arbitrary assumptions and specific interventions.
    - The models make full use of all available sources of information, with the potential to also include expert opinion.
25. Our investigations (across a wide range of countries) indicate that the prosed model fits the data at least as well as competing models, and often more robustly.
26. Further development work is required. In particular, we propose to investigate extending the model to allow the threshold age x_0 and the limiting mortality rate β to vary over time, rather than being fixed at their optimal 2010-2012 values.
27. We are also developing a fully Bayesian version of the methodology which will allow expert opinion about future mortality rates and/or mortality improvements to be coherently incorporated into forecasts. It will also allow properly calibrated uncertainty intervals to be presented together with the corresponding forecasts.

References

Booth H. and Tickle, L. (2008). Mortality Modelling and Forecasting: A Review of Methods. *Annals of Actuarial Science*, 3, 3–43.

Dodd, E., Forster, J.J., Bijak, J., and Smith, P.W.F. (2016) Smoothing mortality data; producing the English life table. *Available at eprints.soton.ac.uk.*

Börger M, and Aleksic, M-C. (2014). Coherent Projections of Age, Period, and Cohort Dependent Mortality Improvements. Mimeo. Institute for Finance and Actuarial Sciences, University of Ulm.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., and Khalaf-Allah, M. (2011). Mortality density forecasts: an analysis of six stochastic mortality models. *Insurance: Mathematics and Economics*, 48, 355-367.

Cairns, A.J.G., Blake, D., Dowd, K., and Kessler, A. (2016). Phantoms Never Die: Living with Unreliable Mortality Data. Journal of the Royal Statistical Society, A, to appear.

Currie, I., Durban, M. and Eilers, P. (2004). Smoothing and Forecasting Mortality Rates. *Statistical Modelling*, 4, 279–298.

Continuous Mortality Investigation (CMI). (2010). *Working Paper 49 – The CMI Mortality Projections Model, CMI 2010*. Institute and Faculty of Actuaries.

De Jong, P. and Tickle, L. (2006). Extending Lee-Carter mortality forecasting. *Mathematical Population Studies*, 13, 1-18.

Delwarde, A., Denuit, M. and and Eilers, P.H.C. (2007). Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting: A penalized log-likelihood approach. *Statistical Modelling*, 7, 29-48.

Haberman S., and Renshaw, A.E. (2012). Parametric mortality improvement rate modelling and projecting. *Insurance: Mathematics and Economics*, 50, 309–333.

Haberman S., and Renshaw, A.E. (2013). Modelling and projecting mortality improvement rates using a cohort perspective. *Insurance: Mathematics and Economics*, 53, 150–168.

Lee, R. and Carter, L. (1992). Modeling and Forecasting US Mortality. *Journal of the American Statistical Association*, 87, 659–671.

Plat, R. (2011). One-year Value-at-Risk for longevity and mortality. *Insurance: Mathematics and Economics*, 49, 462–470.

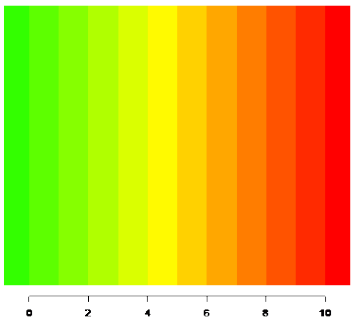
Renshaw, A.E. and Haberman, S. (2003) Lee–Carter mortality forecasting: a parallel generalized linear modelling approach for England and Wales mortality projections. *Applied Statistics*, 52, 119–137.

Renshaw, A.E. and Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, 38, 556-570.

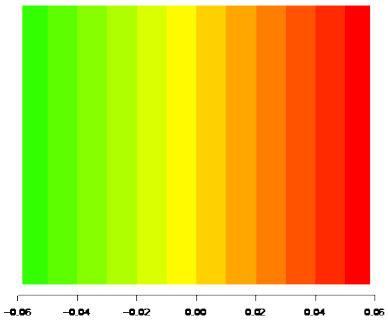
Tabeau, E., van den Berg Jeths, A. and Heathcote C. (2001). *Forecasting Mortality in Developed Countries: Insights from a Statistical, Demographic and Epidemiological Perspective*. Kluwer Academic Publishers.

Wood, S.N. (2006). *Generalized Additive Models: An Introduction with R.* Chapman and Hall/CRC.

Appendix A: Key for Heatmaps



Standardised residuals



Mortality Improvements