



Modelling and forecasting mortality improvement rates with random effects

Arthur Renshaw¹ · Steven Haberman¹ 

Received: 20 January 2021 / Revised: 23 March 2021 / Accepted: 29 March 2021 / Published online: 6 May 2021
© EAJ Association 2021

Abstract

A common feature in the modelling and extrapolation of the trends in mortality rates over time, based on fitted parametric structures, has tended to involve the treatment of a structured fitted main effects period component (with possibly a cohort component) as a random effects time series. In this paper, we follow the lead of Haberman and Renshaw (Insurance Math Econ 50:309–333, 2012) and other authors in modelling and forecasting mortality improvement rates over time, rather than mortality rates. In this context, we assume linear parametric structures for mortality improvement rates, and we examine the feasibility of modelling the main period effects (and possibly any cohort effects) as a random effect from the outset. We argue that this leads to a more unified approach to model fitting and extrapolation.

Keywords Mortality improvements · Random effects modelling · Hierarchical generalised linear modelling · Age heteroscedasticity · Mortality forecasting

1 Introduction

One of the themes of the recent longevity related academic literature has been the consideration of the modelling of mortality improvement rates (MIR), rather than mortality rates (MR). One of the motivations has been a practical one, as noted by Haberman and Renshaw [6], Denuit and Trufin [4] and Hunt and Villegas [10], inter alia. Many standard life tables used by actuaries (and required by regulators) for annuity pricing or reserving are increasingly based on an assumption about the dynamics of suitably defined mortality improvement rates. Thus, Haberman and Renshaw [6] and Denuit and Trufin [4] specifically mention current actuarial practice in Austria, Belgium, Denmark, Switzerland, UK and US that uses mortality improvement rates as a building block. Further we note that, in the UK, the Continuous Mortality Investigation Bureau (CMI) has recently developed and recommended

✉ Steven Haberman
s.haberman@city.ac.uk

¹ Cass Business School, City, University of London, 106 Bunhill Row, London EC1Y 8TZ, UK

a new mortality projection model based on improvement rates: see CMI [3]. These developments indicate a need for a sound theoretical foundation for the modelling of improvement rates—and this has led to a stream of contributions to the literature. We will refer to these different contributions, as appropriate, in the main body of the paper.

A second motivation for the interest of the academic literature has been the recognition that there may be theoretical and practical advantages in modelling improvement rates. A key issue in modelling mortality dynamics is understanding the dominant downwards time trend that has manifested itself over at least the last 70 years. It is well known in time series work that there are advantages if the underlying process that generates the time trend is time-invariant. One of the common methods in time series analysis of transforming a so-called non-stationary time series into a stationary one is by de-trending the series i.e. taking first differences: see Li et al. [13], Haberman and Renshaw [6, 7], Mitchell et al. [14], Hunt and Villegas [10] and Bohk-Ewald and Rau [1]. This transformation implies that the mortality trend relates to the previous year's mortality rates rather than the trend in a hidden mortality factor, like κ_t in the seminal model of Lee and Carter [11]. As we will show below, the definition of mortality improvement rates is linked to this transformation.

A further point noted by Hunt and Villegas [10] is that by considering mortality improvements from national data sets, inferences can be made about mortality trends in smaller sub-populations, although this requires consideration of longevity basis risk (see [8], and [19] for a further discussion).

We refer the reader to the text by Lee et al. [12] for an account of the general background to the theoretical modelling aspects underpinning this paper (Sects. 4–6), where we develop generalised linear MIR and MR models by the inclusion of random effects. Quoting from the epilogue to this text (p 360) “We [the authors] suspect that there may be many other new extensions waiting to be explored where the ideas underlying this book could be usefully exploited”. We believe that this paper provides such an extension.

All the applications presented in this paper were produced using the computer package R. Outline details are given in “Appendix II”.

Our main contribution is to show that, by attributing random effects to the period and cohort components (or just the period components) of a main effects age-period-cohort structured linear predictor from the outset, it is possible to present a comprehensive self-contained process for modelling and extrapolating mortality improvement rates, which incorporates structured dispersion and an apparent self-selecting time series. We show that this methodology also extends to the modelling and extrapolation of mortality rates provided that the predictor structure is linear. We argue that this methodological framework leads to a more unified approach to model fitting and extrapolation as a result of treating the time element as a random effect from the outset, thereby impacting both fitting and extrapolation stages.

The paper is arranged as follows. In Sect. 2, we investigate the inter-relationship between two alternative measures of mortality improvement rates. In Sect. 3, we focus on a direct and indirect method of modelling mortality improvement rates with specific reference to the linear age-period-cohort structure, and suggest the inclusion

of a random effects period component. In Sects. 4 and 5, we describe a method of fitting respectively Normal-Normal and Poisson-Normal generalised linear mixed models, which incorporate fixed and random effects components, and have an additional provision for the joint modelling of a structured dispersion parameter. In Sect. 6 we indicate how a simple structured time series can also be incorporated into the fitting process. In each of Sects. 4–6, we include an investigation into the potential of the respective methodology by applying it to the recent UK male mortality experience. Sections 7 and 8 provide a discussion and some concluding comments.

2 Mortality improvement statistics

Ideally, in continuous time, age specific mortality improvement rates are quantified by the partial derivative of the log-mortality rate, either with respect to period (calendar-year) or with respect to birth-year (cohort). Given that the former approach is the most commonly used, we assume this to be the case here unless stated otherwise.

Consider a rectangular age-period Lexis plane divided into annual cells supportive of mortality data.

$(d_{xt}, e_{xt}, \omega_{xt})$: age $x = x_1, x_2, \dots, x_k$, period $t = t_1, t_2, \dots, t_n$, and birth-year $s = t - x$, where d_{xt} , reported number of deaths; e_{xt} , matching central exposure to the risk of death; ω_{xt} , prior (0/1) weights to indicate empty/non - empty data cells and denote $m_{x,t}$, the central rate of mortality.

Assuming central rates of mortality throughout, two superficially different mortality improvement rate statistics have been proposed in the literature and are of interest:

$$\text{Statistic I: } y_{xt} = \frac{-\Delta_t m_{x,t}}{(m_{x,t-1} + m_{x,t})/2} = 2 \frac{(1 - m_{x,t}/m_{x,t-1})}{(1 + m_{x,t}/m_{x,t-1})}, \text{ and}$$

Statistic II: $z_{xt} = \Delta_t \log m_{x,t} = \log m_{x,t} - \log m_{x,t-1}$ where Δ_t is the differencing operator and the statistics are computed using the estimate $\hat{m}_{x,t} = d_{xt}/e_{xt}$; see Haberman and Renshaw [6] and Mitchell et al. [14] respectively for applications.

We investigate next the precise nature of the difference between the two statistics. Firstly, we note that while both statistics are discrete representations of the partial derivative of $\log m_{x,t}$, Statistic I is based on the actual partial derivative itself. Secondly, we recall the monotonic increasing characteristic of the log function. Then a comparison of $-\Delta_t m_{x,t}$ with $\Delta_t \log m_{x,t}$, which determines the respective signs of the two statistics, associates respectively positive and negative values with actual mortality improvements. In addition, exploratory scatter plots of the two statistics, using any rectangular age-period UK male mortality data array, are found to exhibit perfect negative correlation which implies an exact connection between the two statistics.

Indeed, a mathematical tractable relationship between the two statistics reads as follows:

$$y_{xt} = 2 \frac{(1 - m_{x,t}/m_{x,t-1})}{(1 + m_{x,t}/m_{x,t-1})} = 2 \frac{(1 - \exp(\Delta_t \log m_{x,t}))}{(1 + \exp(\Delta_t \log m_{x,t}))} = 2 \tanh(-z_{xt}/2).$$

Correcting for the disparity in the sign of z_{xt} the relationship connecting the two statistics reads $y_{xt} = 2 \tanh(z_{xt}/2)$ so that

$$z_{xt} = 2 \tanh^{-1}(y_{xt}/2) = \log\left(\frac{1 + y_{xt}/2}{1 - y_{xt}/2}\right) = y_{xt} + \frac{2}{3}\left(\frac{y_{xt}}{2}\right)^3 + o(y_{xt}^5); \left(\frac{y_{xt}}{2}\right)^2 < 1$$

a convergent power series comprising odd powers only, with the implication that the absolute value of y_{xt} is less than the matching absolute value of z_{xt} and hence the greater accuracy in general. Unless stated otherwise we use Statistic I.

2.1 An application

We utilise the UK male mortality data set, covering the period 1960–2016, ages 0–102, and comprising annual death counts and matching central exposures to the risk of death, as compiled by the Human Mortality Database [9]. By truncating the data at the upper age limit of 102, possible complications arising from the irregular nature of the lost fragments including zero entries are avoided. In addition, the number of any such fragments lost is extremely small, especially for the early calendar years.

One aspect of the computed empirical MIRs concerns the nature of any information provided in respect of patterns in the resulting age profiles. In order to investigate this feature, we have computed and displayed the n -year empirical MIR rolling-averages for a range of values of n . One such set of results for the 5-year MIR rolling averages, expressed as percentages, and centred on the periods 2014(−1)2006 are depicted in the various panels of Fig. 1. (We make frequent use of the notation a(c)b to denote sequences of numbers ranging from a to b at intervals of c). In addition we have fitted and display a smooth-spline curve in each panel (using the R smooth.spline function with a parameter setting of 0.8). We note an identifiable crude pattern in each panel subject to a degree of variation between (annually) adjacent panels. We return to this issue later in Sect. 4.

3 Mortality improvement rates and random effects

We start from the premise, implicit in Hunt and Villegas [10] that, in continuous time, the MIR is quantified by

$$-\frac{\partial}{\partial t} \log m_{x,t}$$

(with the negative sign ensuring that improvement is positive and deterioration negative). We focus on linear parametric predictor structures $\eta_{x,t}$ and specifically the familiar main effects APC (age-period-cohort) structure, so that

$$-\frac{\partial}{\partial t} \log m_{x,t} = \eta_{x,t} = \alpha_x + \iota_{t-x} + \kappa_t \tag{1}$$

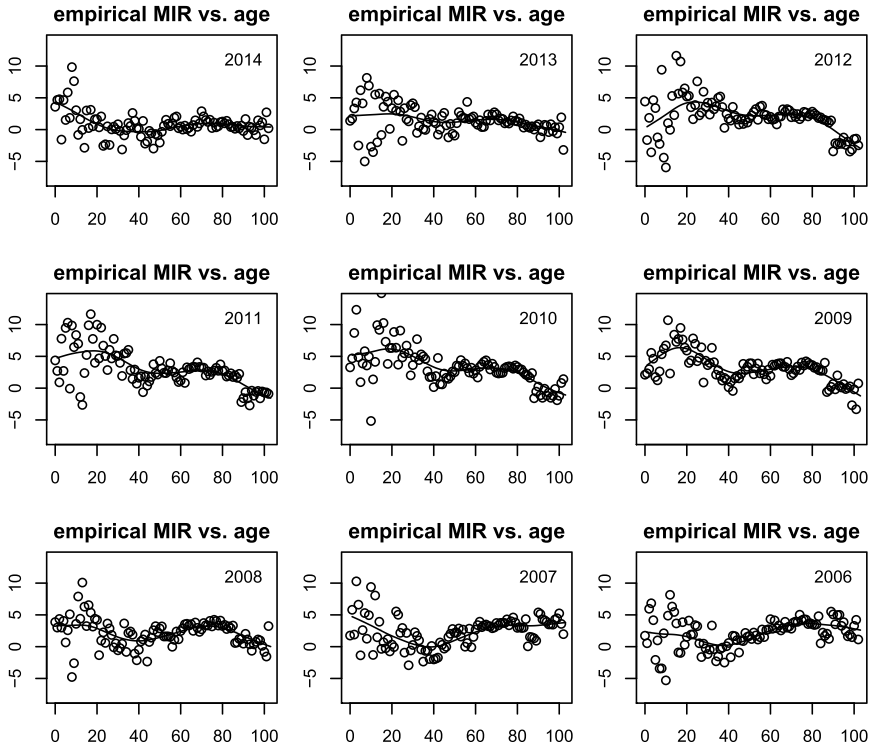


Fig. 1 UK males 2004–2016, ages 0–102. 5-year empirical MIR rolling-averages, centred on 2014(–1)06, with smooth-spline curves. MIR percentage scale; R smooth-spline parameter setting = 0.8

comprising respectively age, birth-year and period main effects: while partial integration gives

$$\log m_{x,t} = \tilde{\eta}_{x,t} = A_x - \alpha_x(t - t_1) - \int_{t_1}^t l_{u-x} du - \int_{t_1}^t \kappa_u du \tag{2}$$

with the $\tilde{\eta}$ corresponding discrete version

$$\log m_{x,t} = \tilde{\eta}_{x,t} = A_x - \alpha_x(t - t_1) - \sum_{u=t_1}^t l_{u-x} - \sum_{u=t_1}^t \kappa_u \tag{3}$$

as listed in Table 1 of Hunt and Villegas [10]. Here $\tilde{\eta}_{x,t}$ denotes the transformed linear predictor and $A_x = \log m_{x,t_1}$ the initial log-mortality rates (or ‘constant’ of integration). Alternatively, Eq. (2) can be rearranged to read as

$$\log m_{x,t} = \eta_{x,t} = \alpha_x + \beta_x(t - \bar{t}) + l_{t-x} + \kappa_t \tag{4}$$

subject to the following transformation and redefinition of symbols

$$\alpha_x \rightarrow A_x - \alpha_x(\bar{i} - t_1) + I_{t_1-x} + K_{t_1}$$

$$\beta_x \rightarrow -\alpha_x; \iota_{t-x} \rightarrow -I_{t-x}; \kappa_t \rightarrow -K_t$$

where I and K denote the respective integrands of ι and κ .

Noting the close relationship between all three structures, MIR may be modelled directly using either of the Sect. 2 statistics as Normal responses [6], here in combination with structure (1), or indirectly using either (3) [10] or (4) (e.g. Richards et al. [17]; CMI [3]) in combination with Poisson responses $\hat{m}_{x,t} = d_{x,t}/e_{x,t}$ and a log-link function.

We note that, subsequent to model fitting, (with the possible exception of the latest UK Continuous Mortality Investigation model), the fixed effects parameter κ_t (and sometimes ι_{t-x}) are treated as random variables to facilitate model extrapolation. Hence we investigate the effect of reformulating the modelling assumptions underpinning (1), (3) and (4) by including the random effects from the outset. To do so, we follow the approach of Lee et al. [12] which is based on hierarchical generalised linear models (HGLMs).

4 Normal linear mixed modelling and MIR

We start by modelling the linear predictor structure (1) treating both κ_t and ι_{t-x} as random effects and using Statistic I of Sect. 2 as responses. In formulating the model matrices which follow, the Lexis plane is scanned along the age axis for each increasing time period in sequence and we attach the suffix $i = (x, t)$ (subject to exchangeability as appropriate), when the need arises to refer to the individual components.

Consider the multivariate normal mixed model

$$y = X\beta + Zv + e; e \sim MVN(O, \Sigma), v \sim MVN(O, \Lambda)$$

subject to independence with $\Sigma = \sigma_e^2 I, \Lambda = \sigma_v^2 I; \tau = (\sigma_e^2, \sigma_v^2) = (\phi, \lambda)$ and focus on the associated augmented linear model

$\begin{pmatrix} y \\ \psi_M \end{pmatrix} = \begin{pmatrix} X & Z \\ O & I \end{pmatrix} \begin{pmatrix} \beta \\ v \end{pmatrix}$, abbreviated to $y_a = T\theta$, with quasi random effects responses $\psi_M = E(v) = 0$ and augmented variance-covariance matrix

$$\Sigma_a = \begin{pmatrix} \Sigma & O \\ O & \Lambda \end{pmatrix}, \Sigma = \text{diag}(\phi_i), \Lambda = \text{diag}(\lambda)$$

and the suffix i is introduced to indicate optional fixed effects variable dispersion.

Specifically, the matrices are designated as follows:

$X, N \times p$ fixed effects design matrix. $\beta, p \times 1$ fixed effects parameters $\alpha_x, Z, N \times q$ random effects design matrix. $v, q \times 1$ matrix of random effects comprising κ_t and ι_{t-x} . $T, (N + q) \times (p + q)$ augmented design matrix. $\theta, (p + q) \times 1$ augmented matrix of fixed parameters and random effects. $y_a, (N + q) \times 1$ matrix of augmented responses comprising either one of the MIR statistics and quasi random effects

responses $\psi_M = E(v)$. Σ_{a^*} $(N + q)$ augmented diagonal matrix of scale parameters. I and O , respective identity and zero matrices of appropriate size where in terms of $n_1 = x_k - x_1 + 1$, $n_2 = t_n - t_1 + 1$.

$$N = n_1(n_2 - 1), p = n_1, q = n_1 + 2(n_2 - 2).$$

The three constraints $\kappa_{\min(t)} = t_{\min(s)} = t_{\min(s) + 1} = 0$ are applied, which are sufficient in number to ensure that the matrix T has full rank.

The model is fitted using the iterative weighted least squares (IWLS) procedure outlined in ‘‘Appendix I’’. Three sets of (Studentised) residuals

$$\frac{\hat{e}_i}{\sqrt{\hat{\phi}_i(1 - h_i)}}, \frac{\hat{t}_i}{\sqrt{\hat{\lambda}_2(1 - h_{Mi})}} \text{ and } \frac{\hat{\kappa}_i}{\sqrt{\hat{\lambda}_1(1 - h_{Mi})}} \tag{5}$$

are generated where

$$\hat{e}_i = y_i - \hat{\mu}_i, \mu_i = E(y_i|v) = X_i\beta + Z_iv$$

subject to the respective constraints

$$\sum_i \hat{e}_i = 0, \sum_s \hat{t}_s = 0 \text{ and } \sum_t \hat{\kappa}_t = 0.$$

We recall that the modelling assumptions imply that the mean values of all three sets of residuals are also zero.

4.1 An application

We make use of the UK male mortality data set, period 1960–2016, ages 0–102. We remark that the setting of the upper age limit (102) avoids any empty data cells for very little loss of other data above the age of 102. Alternatively, the introduction of 0/1 prior weights can be used to marginally extend the upper age limit.

Details of the fitted mixed effects model structure (1), with variable dispersion, are presented in Fig. 2, followed by some associated residual plots in Figs. 3, 4, 5. Thus the upper two panels in Fig. 2 depict the respective first and second moment fixed effects parameter estimates $\hat{\alpha}_x$ and $\hat{\phi}_x$, while the lower two panels depict the respective random effects components \hat{t}_{t-x} and $\hat{\kappa}_t$.

Figure 3, the upper pair of panels in Fig. 4, together with the left hand column of panels in Fig. 5 refer to the first or primary set of residuals (5); with the centre and right hand columns of panels (Fig. 5) referring to the respective sets of cohort (centre) and period random effects residuals. For these sets of residual we find that

$$\sum_i \hat{\epsilon}_i = -0.18, \quad \bar{\hat{\epsilon}} = -0.00003, \quad \sum_s \hat{\iota}_s = -0.3, \quad \bar{\hat{\iota}} = -0.002,$$

$$\sum_t \hat{\kappa}_t = -0.007, \quad \bar{\hat{\kappa}} = -0.0001.$$

We note the following features from the figures: the constant nature of the bands of residuals in Fig. 3, consistent with the successful capture of dispersion effects; the generally satisfactory distribution of positive and negative residuals across the Lexis plane, with no noteworthy gaps (Fig. 4, upper two panels); and the marginal nature of the centre column of Normal and half-Normal plots (Fig. 5) associated with the cohort random effects. Taken as a whole, the residual plots are indicative that the modelling assumptions have been largely met.

We turn next to the replacement of the empirical MIR displayed in Fig. 1 with the corresponding fitted MIR. These are reproduced in the panels of Fig. 6, with

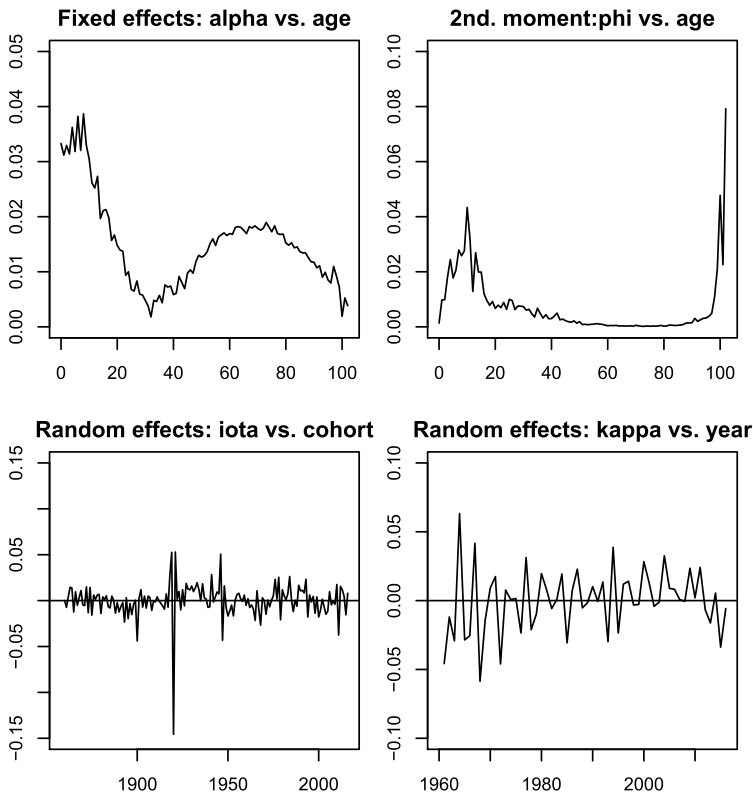


Fig. 2 UK males 1960–2016, ages 0–102, normal—normal mixed modelling, cohort and period random effects, structure (1), variable age dispersion. Upper panels: fixed age effect. Lower panels: component random effects

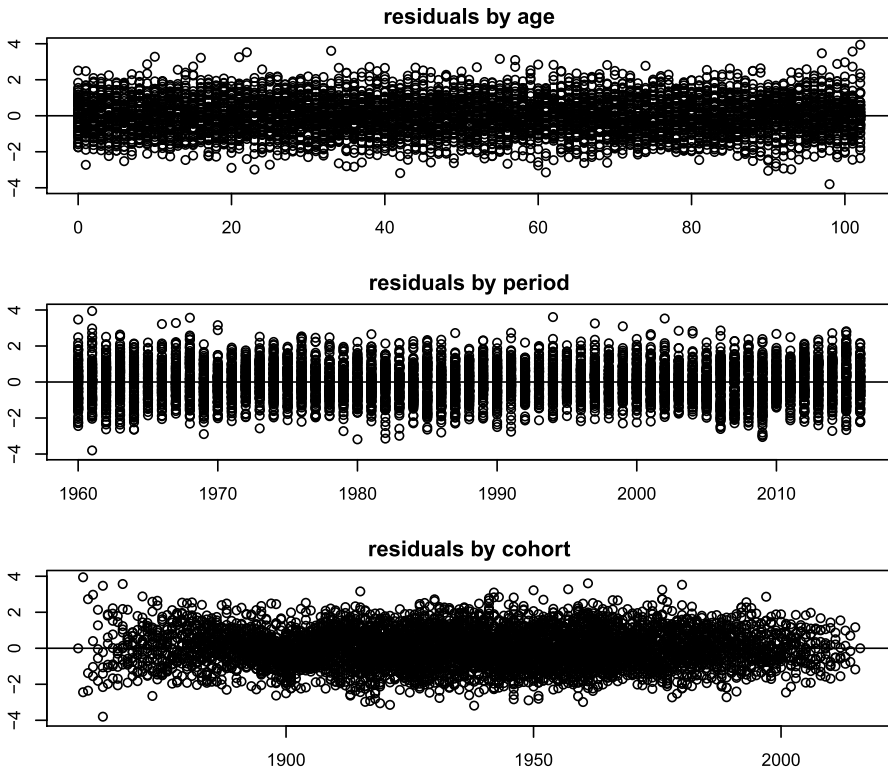


Fig. 3 UK males, 1960–2016, ages 0–102. Normal—normal mixed modelling, cohort and period random effects, structure (1), variable age dispersion. Model residuals plotted respectively against age, period and cohort-year

the exception of the lower right panel for 2006. In addition, we have used the same parameter setting to generate the matching smooth-spline curves, which are duplicated together with that generated for 2006 in the lower right panel. We note the sharpening focus of the patterns in each panel as a consequence. We note also the consistent shape of the resulting spline-smoother curves, which are characterised by two local extremes, subject to a slight drift in their positioning on the age axis. The two lower somewhat detached spline curves (in the lower right panel) are the product of the two most recent years, reflecting the recent decline in mortality improvement rates (a particular phenomenon which is analysed further by [5]). Almost without exception, for all panels, the readings (in terms of improvement rates) are positive.

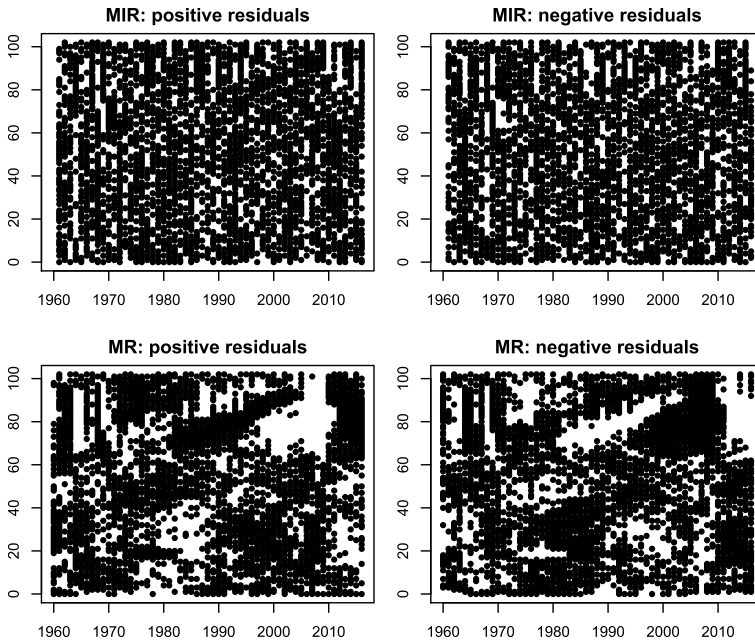


Fig. 4 UK males 1960–2016, ages 0–102. Positive/negative residuals in adjacent Lexis planes. Upper panels: Normal—Normal mixed model (1). Lower panels: Poisson—Normal mixed model (4). Both models with cohort and period random effects, and variable age dispersion

5 Poisson-Normal generalised linear mixed modelling and MR

In this section, we model the log-link linear predictor structure (4), with cohort and period random effects, using a combination of Poisson empirical mortality rates $y_i = \hat{m}_i$ and quasi random effects $\psi_M = E(v) = 0$ to act as responses (dependent variable). Thus we:-

Define an augmented GLM with responses $y_a^t = (y^t, \psi_M^t)^t$ such that

$$E(y|v) = \mu \quad E(\psi_M) = v$$

$$\text{var}(y|v) = \phi V(\mu) \quad \text{var}(\psi_M) = \lambda V_M(\psi_M)$$

with potentially structured dispersion or scale parameters $\tau = (\phi, \lambda)$, respective Poisson and Normal variance functions $V(\mu_i) = \mu_i$, $V_M(\psi_{Mi}) = 1$, with log and identity link functions and augmented linear predictor

$$\eta_a = \begin{pmatrix} \eta \\ \eta_M \end{pmatrix} = \begin{pmatrix} \log \mu \\ v \end{pmatrix} = \begin{pmatrix} X & Z \\ O & I \end{pmatrix} \begin{pmatrix} \beta \\ v \end{pmatrix} = T\theta.$$

The matrix/vector symbols are all as defined in Sect. 4, subject to changes in the details. Thus in the case of (4) with the two random effects components

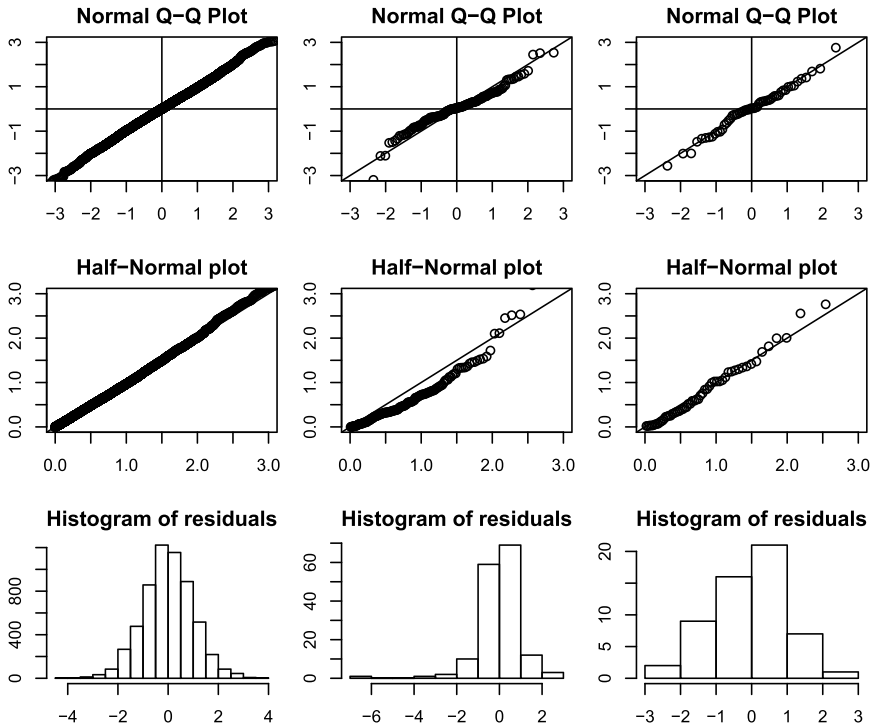


Fig. 5 UK males 1960–2016, ages 0–102. Normal—normal mixed modelling, cohort and period random effects, structure (1), variable age dispersion. Left panels: model residuals. Centre panels: matching cohort effects residuals. Right panels: matching period effects residuals

$$N = n_1 n_2, p = 2n_1, q = n_1 + 2n_2 - 1,$$

while no constraints are necessary to ensure the full rank of the matrix T .

Model fitting requires an adjusted dependent variable $z_a = (z^t, z_M^t)^t$ whose values

$$z_i = \eta_i + \frac{(y_i - \mu_i)}{\mu_i}, z_{Mi} = \psi_{Mi} = 0 \tag{6}$$

reflect the respected log and identity link functions, and a variance covariance matrix

$$\Sigma_a = \begin{pmatrix} \Sigma & O \\ O & \Lambda \end{pmatrix}, \Sigma = \text{diag} \left(\frac{\phi_i}{\mu_i e_i} \right), \Lambda = \text{diag}(\lambda)$$

which reflects the choice of respective variance functions: the suffix applied to ϕ being indicative of fixed effects variable dispersion. Fitting then follows the IWLS procedure as laid out in “Appendix I” subject to the following changes:

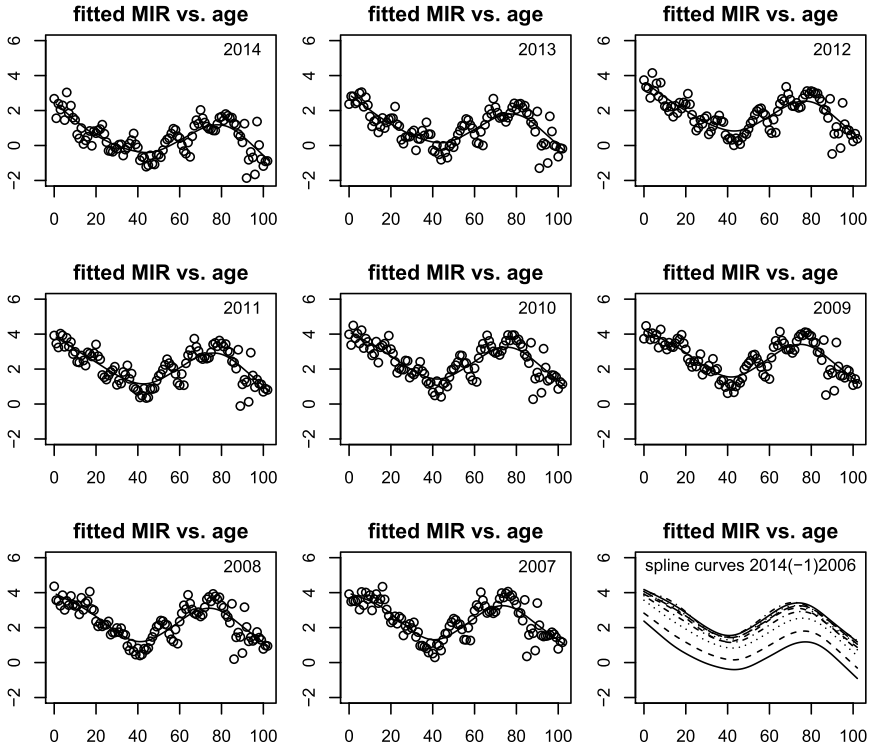


Fig. 6 UK males 1960–2016, ages 0–102. Normal–Normal mixed modelling, cohort and period random effects, structure (1), variable dispersion. 5-year fitted MIR rolling—averages, centred on 2014(–1)07, smoothed—spline curves. Lower right panel: comparison of resulting spline curves (including 2006)

1. Given $\tau = (\phi, \lambda)$

update $\hat{\theta}$ by computing Σ_a^{-1} , z_a and solving $T^t \Sigma_a^{-1} T \hat{\theta} = T^t \Sigma_a^{-1} z_a$.

2. Replace d_i when up-dating the fixed effects dispersion parameter ϕ with either the squared deviance or squared Pearson residuals

$$d_i = 2 \left(y_i \log \left(\frac{y_i}{\mu_i} \right) - (y_i - \mu_i) \right), \quad d_i = \frac{(y_i - \mu_i)^2}{\mu_i}.$$

while the empirical mortality rates provide the starting values for μ . Details of the residuals are as described in Sect. 4.

5.1 An application

We make use of the same UK male mortality data set for the period 1960–2016, ages 0–102, as above. Key details of the mixed effects structure (4) with variable dispersion by age are presented in Fig. 7: with the first moment fixed effects

parameter estimates $\hat{\alpha}_x$ and $\hat{\beta}_x$ displayed in the two upper panels and the two random effects $\hat{\iota}_{t-x}$ and $\hat{\kappa}_t$ in the two centre panels. We note that $\hat{\alpha}_x$ takes the familiar shape of a static life-table' including an 'accident hump'. We also note that the second moment fixed effects parameter $\hat{\phi}_x$, which is not displayed, is not too dissimilar in shape from its counterpart depicted in Fig. 2. In anticipation of the subsequent application of a first order autoregressive integrated ARI(1,1) time series, we depict the matching first differences of the two random effects components in the lower two panels.

Certain residual plots are presented in the two lower panels of Figs. 4 and 8, with the former indicating a lack of randomness in the spatial distribution of positive and negative residuals across the Lexis plane. The pattern of residuals when plotted against age, period and cohort year, result in similar patterns to the ones displayed in Fig. 3 and have not, as a consequence, been reproduced.

The model is extrapolated period-wise by applying an ARI(1,1) time series to the two random effects components as depicted in Fig. 9, together with forecasts and including matching residual plots in the alternative panels. Note that these time series plots replicate the respective details of the two centre panels in Fig. 7,

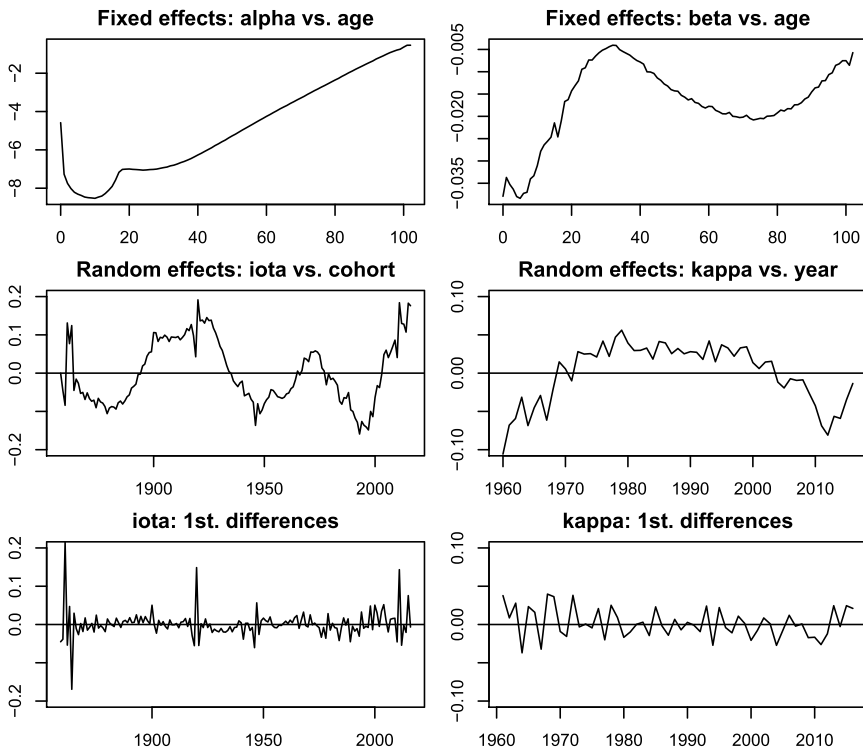


Fig. 7 UK males 1960–2016, ages 0–102. Poisson–Normal mixed model (4). Upper panels: fixed age effects parameters. Centre panels: random effects components. Lower panels: respective random effects first differences

while recalling the roles played by the two lower panels in Fig. 7 in generating the ARI(1,1) forecasts.

On the basis of 1000 time series forecast simulations, we have computed the period-based life expectancies at each simulation, for certain ages, and depict the resulting 0.05, 0.5, 0.95 quantiles in Fig. 10. We have focused on period-based computations as opposed to cohort-based computations since period-based forecasts may be readily calibrated against historical trends based on the raw mortality data stretching back to 1922, which are also featured in Fig. 10. It would appear that the median forecast trends represent an over statement when compared with the most recent trend over the past decade or so, where mortality improvement rates have decelerated (see Fig. 1). Attempts to influence this feature by shortening the modelling period, in order to focus on the more recent mortality trend data, have not proved to be successful.

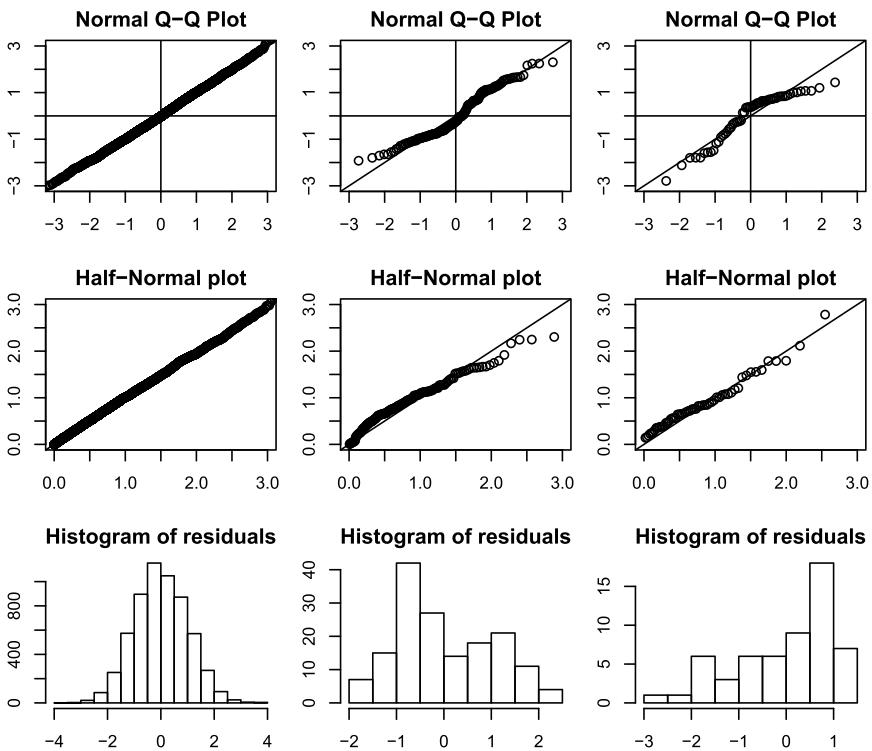


Fig. 8 UK males 1960–2016, ages 0–102. Poisson–Normal mixed modelling, cohort and period random effects, structure (4), variable age dispersion. Left panels: model residuals. Centre panels: matching cohort effects residuals. Right panels: matching period effects residuals

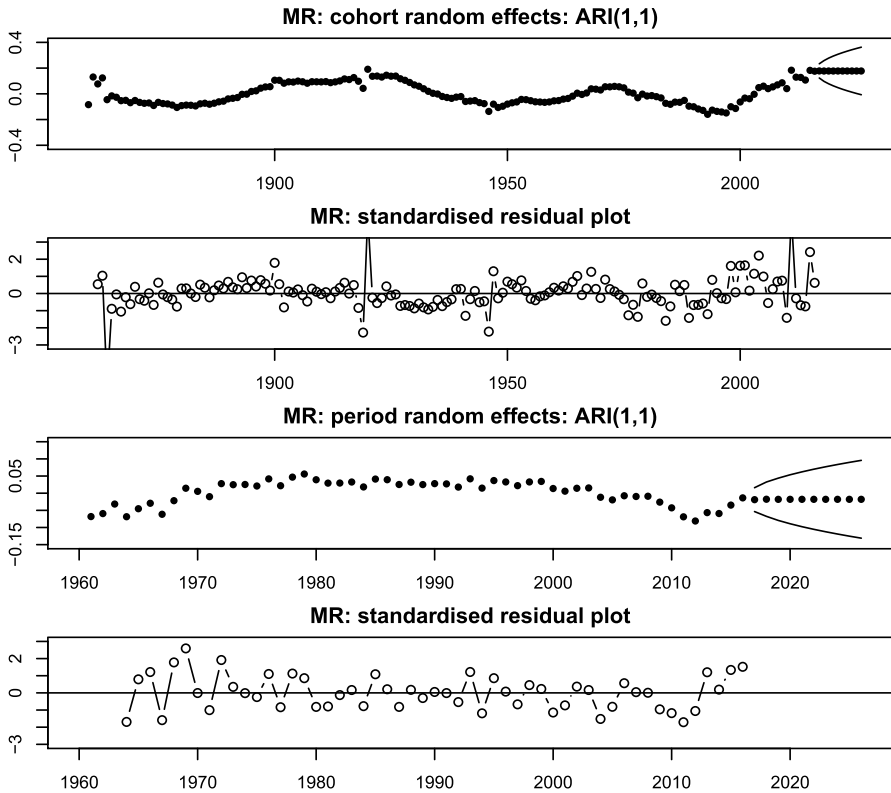


Fig. 9 UK males 1960–2016, ages 0–102. MR model (4): cohort and period ARI(1,1) time series

6 Incorporating an AR(1) times series into the MIR and MR fitting process

When the model structures of Sect. 3 are formulated with only the main period component designated as a random effect, the possibility of simultaneously fitting a single parameter AR(1) period component time series arises. This involves the necessary adjustments to the two design matrices X and Z and composition of the fixed and random effects matrices β and ν coupled with the introduction of a linear transformation of the random effects component

$$\nu = L(\rho)r, \text{ with } r \sim MVN(\mathbf{O}, \Lambda), \Lambda = \text{diag}(\lambda),$$

where

$L(\rho)$ is a $q \times q$ matrix, with parameter ρ , and has $\text{rank}(L(\rho)) = q$.

Specifically for the AR(1) process, under the reverse transformation

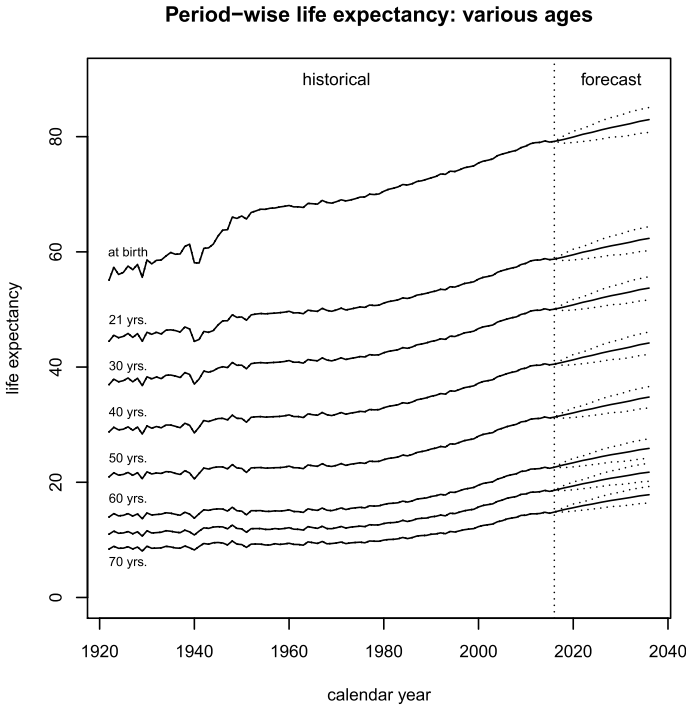


Fig. 10 UK males 1922–2016, ages 0–102. Annual period—wise life expectancies, specific ages. Empirical values with simulated 5%, 50%, 95% quantile forecasts generated using MR model (4)

$$\mathbf{r} = L^{-1}(\rho)\mathbf{v} = (I - K)\mathbf{v}$$

where K has non-zero elements $k_{t+1,t} = \rho$, so that component wise.

$$r_t = v_t - \rho v_{t-1}, t = t_{\min}, t_{\min}+1, \dots, t_n.$$

Comprehensive model fitting then follows by embedding the IWLS fitting process (“Appendix I”) between the following two steps:

1. Given ρ and hence $L(\rho)$ estimate $(\boldsymbol{\beta}, \mathbf{r}, \boldsymbol{\phi}, \boldsymbol{\lambda})$ for the model $\boldsymbol{\eta} = g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}^*r$ where $\mathbf{Z}^* = \mathbf{Z}L(\rho)$. and g is the appropriate identity or log link function.
2. Given $(\boldsymbol{\beta}, \mathbf{r}, \boldsymbol{\phi}, \boldsymbol{\lambda})$ estimate ρ by maximising the adjusted profile likelihood.

Here, with $\mathbf{r} \sim MVN(\mathbf{O}, \boldsymbol{\Lambda}), \boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\lambda})$ and $\mathbf{v} = L(\rho)\mathbf{r}$ we maximise the profile likelihood of $\mathbf{v} \sim MVN(\mathbf{O}, L(\rho)\boldsymbol{\Lambda}L(\rho)^t)$, adjusted to $\mathbf{v} \sim MVN(\mathbf{O}, \boldsymbol{\Lambda})$, as originally defined: with the parameter ρ entering the expression for the adjusted profile likelihood via the response term \mathbf{v} which comprises the linear transformation of r , involving ρ .

We subsequently refer to the above comprehensive self-contained fitting and extrapolation process as Approach A. Alternatively, by setting $L(\rho) = I$, we can retain the IWLS fitting process unchanged and separately fit the single parameter AR(1) time series, subsequently referred to as Approach B.

For (1)

$$N = n_1(n_2 - 1), p = n_1, q = n_1 + 2(n_2 - 2)$$

with the two constraints $\kappa_{\min} = t_{\min} = 0$ applied, and for (3)

$$N = n_1 n_2, p = n_2 + 3(n_1 - 1), q = n_2 - 1$$

with the three constraints $t_{t_1-x_k} = t_{t_n-x_1} = \kappa_{t_1} = 0$ applied.

We note that in the case of (3) care is needed in the construction of matrix X to ensure that the main effects cohort parameters do not automatically take precedence over other fixed effects parameters (which appears to be the case if the R-library *lme4* file is accessed and used in its construction).

6.1 An application

In addition to the direct modelling of MIR by the above process using model (1), we compare results obtained by the indirect modelling of MIR using model (3), which has also been formulated with a single period random effect. We again make use of the UKmale mortality data set, this time with slightly reduced period (1961–2009) and age (1–89) ranges in order to facilitate comparison with certain aspects of the previously reported investigations in Haberman and Renshaw [6]. When fitting (1), we choose the MIR Statistic I of Sect. 2, while noting that the choice does not have a material effect on the ensuing results. When fitting (3), we include the term A_x in the design matrix as a fixed effect to be estimated, as opposed to giving it the status of a predetermined offset. For both models, we allow for the capture of structured dispersion by fixed age effects.

The fitted results obtained using (1) are depicted in Fig. 11, together with one of the resulting residual plots in Fig. 12: the patterns in the remaining residual plots being similar to their counterparts in Fig. 3 and the two upper panels in Fig. 4. Modelling is conducted using Approach A throughout this section, but we have additionally superimposed, in Fig. 11, the equivalent plots obtained using Approach B to modelling (broken lines). We note the consistency of patterns within each panel as is the case when comparing the residual plots (not reproduced here but available from the authors).

Similarly, the fitted results obtained using (3), are presented in Fig. 13 together with one of the resulting residual plots in Fig. 14: the patterns in the remaining residual plots being similar to their counterparts in Fig. 3 and the two lower panels in Fig. 4. We note that the estimated fixed effects term A_x depicted in the upper left panel of Fig. 13 takes on the familiar shape of a static life-table (including an ‘accident hump’), and we have plotted the differences between this estimated age profile and the matching initial empirical log mortality rate profile in the lower left panel of

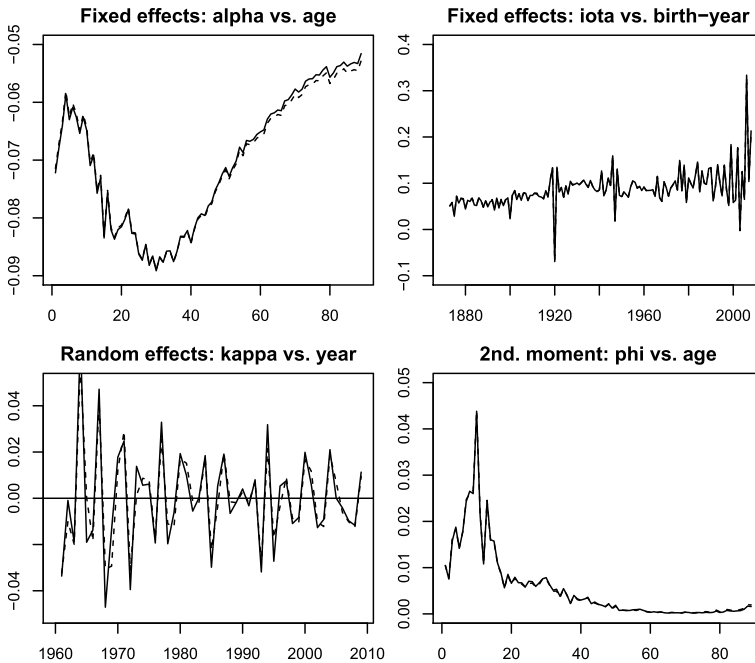


Fig. 11 UK males 1960–2009, ages 1–89. Normal–Normal mixed modelling, period random effects, structure (1), variable dispersion by age. Superimposed Approach A and B (broken-line) results in each panel

Fig. 13 for comparison purposes. We note the pattern of narrowing differences with increasing age. Again, we have superimposed the equivalent plots obtained using Approach B to modelling (broken lines) in all but the lower left panel of Fig. 13 where two superimposed different point plotting symbols are displayed. The mutually compensating displaced patterns in the fixed effects parameter α_x and t_{t-x} plots (upper right, centre left panels) are noteworthy while consistency is preserved in the remaining four panels. Of particular interest is the horizontal trend (around zero) in the respective random effects period component (lower left panel Fig. 11 and centre right panel Fig. 13), which is indicative of the choice of a single parameter AR(1) process. We believe that this is possibly a general feature which is a consequence of the mixed effects model design.

Since the patterns in the matching residual plots are essentially identical under both Approaches A and B to modelling, they have not been replicated. The supporting comparison of residual plots is largely as illustrated in Fig. 7a–d of Haberman and Renshaw [6].

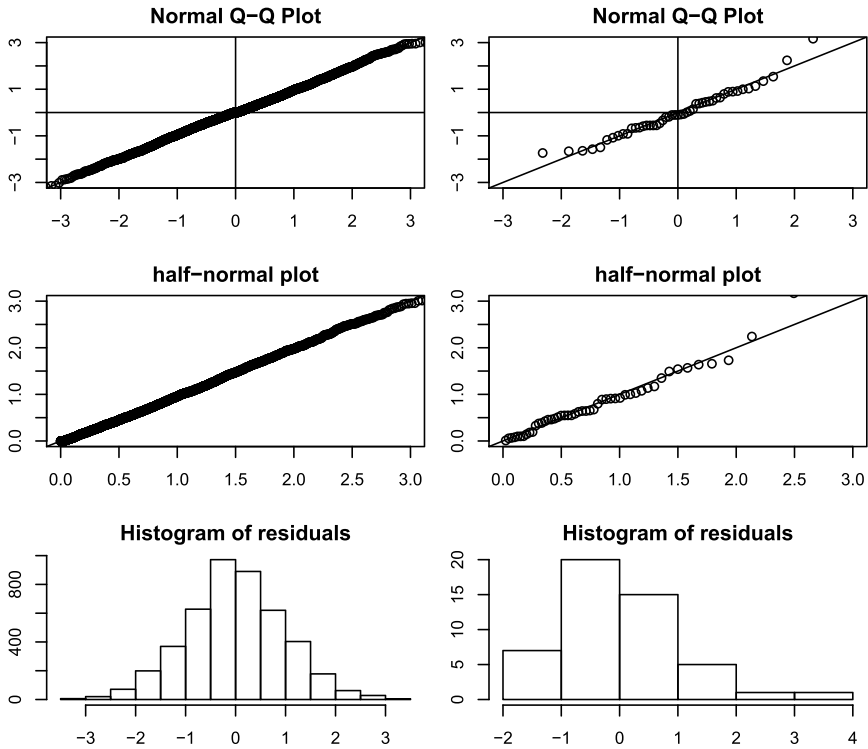


Fig. 12 UK males 1960–2009, ages 1–89. Normal—Normal mixed modelling, period random effects, structure (1), variable age dispersion. Left panels: primary residuals. Right panels: matching period effects residuals

6.2 Conversion of model outputs into model rates or improvement rates

The basic operation used to determine mortality rates from model (1) is that of definite integration, while that used to determine improvement rates from model (3) is that of differentiation. We have conducted each of these operation (in discrete time) to construct Figs. 15 and 16 by superimposing the respective fitted statistic (broken lines) on the corresponding empirical statistic (continuous line). Thus in Fig. 15, we have used the fitted values $\hat{y}_{x,t}$ from model (1) (based on Statistic I of Sect. 2) to compute and depicted the ‘fitted’ log mortality rates

$$\log \tilde{m}_{x,t} = \log m_{x,t_1} + \sum_{u=t_1+1}^t \log \left(\frac{2 - \hat{y}_{x,u}}{2 + \hat{y}_{x,u}} \right), \quad t_1 \leq t \leq t_n$$

for ages $x=40(05)75$. Two cases are displayed using (1) \hat{A}_x from model (3) (broken-lines), and (2) $\log \hat{m}_{x,t_1}$ the initial empirical log rates (dotted lines) to provide the starter log rates $\log m_{x,t_1}$. In Fig. 16, we have performed the reciprocal conversion

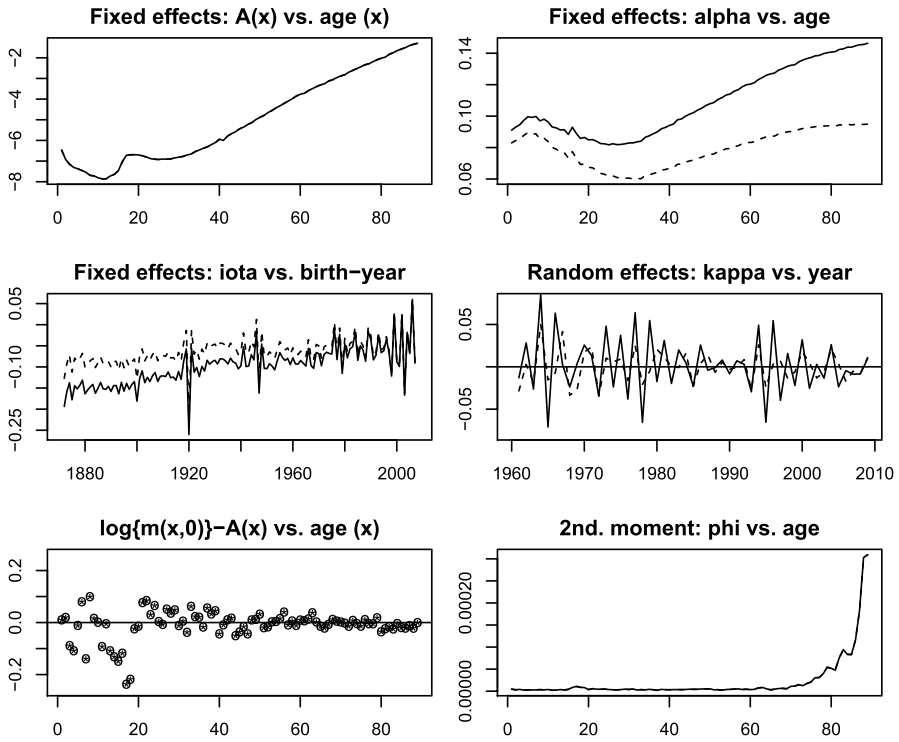


Fig. 13 UK males 1960–2009, ages 1–89. Poisson–Normal mixed modelling, period random effects, structure (3), variable dispersion by age. Superimposed Approach A and B (broken-line) details in five panels. Lower left panel: superimposed initial/estimated log(rate) differences by age

process by using the parameter estimates from models (3) to generate the broken-lines depicting ‘fitted’ improvements in Fig. 16. This issue is discussed further in Sect. 7.

6.3 Model extrapolation

Once the models have been fitted to the data, the method by which mortality improvement rates are extrapolated and converted to rates, using the fixed effects parameter estimates $\hat{\alpha}_x$ and \hat{t}_{t-x} , and random effects time series $\hat{\kappa}_t = \hat{v}_t$, is identical for the two models (1) and (3): only the method of estimating these effects differs.

In the application which follows, when converting from extrapolated improvements to rates, we adopt the procedure described in Sect. 2.6 of Haberman and Renshaw [6]. For extrapolation to ages outside the dataset (referred to as “topping-out” by age in the literature), we follow the period based version of the procedure using a fitted hyperbolic function which is described in Sect. 2.7 of Haberman and Renshaw [7]. In order to facilitate comparison with the earlier results of Haberman and

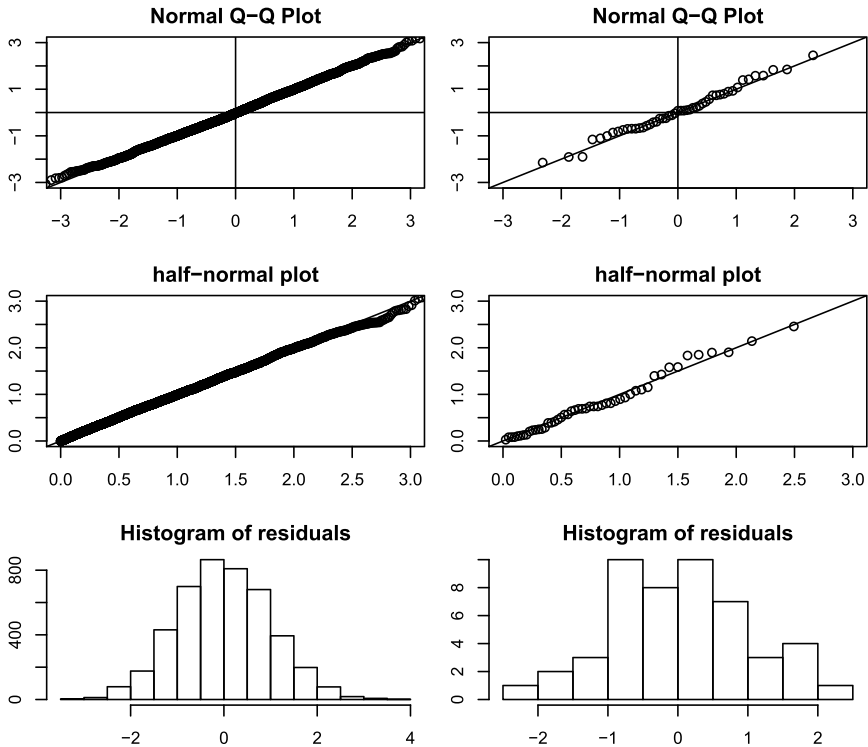


Fig. 14 UK males 1960–2009, ages 1–89. Poisson–Normal mixed modelling, period random effects, structure (3), variable age dispersion. Left panels: primary residuals. Right panels: matching period effects residuals

Renshaw [6], in this application, we have truncated the lower end of the age range to ages 20–89, prior to modelling,

Given the horizontal trend in the random effects period components, coupled with the design feature $E(v_i) = 0$, we extrapolate the models by applying the single parameter $AR(1)$ time series to $\hat{\kappa}_t$. By this method, we have constructed Fig. 17 which depicts extrapolated life expectancy 5%, 50%, 95% quantiles, computed on a cohort basis, and plotted as the individual horizontal lines, each based on 1000 simulations. First, in the two upper panels we show the respective model evolving biennial $t_n = 1995(02)2009$ extrapolations (subject to front-end data deletions shown in descending sequence), thereby using data for 1961–1995, 1961–1997, ..., 1961–2009. Second, in the two lower panels, we show the respective model static $t_n = 2009$ extrapolations having first subjected the data to systematic biennial rear-end truncations 1961(02)1975, thereby using data for the periods 1961–2009, 1963–2009, ..., 1975–2009 (shown in ascending sequence).

We note the similarity of the extrapolations on comparing the different models, together with the material narrowing of the intervals in comparison with the

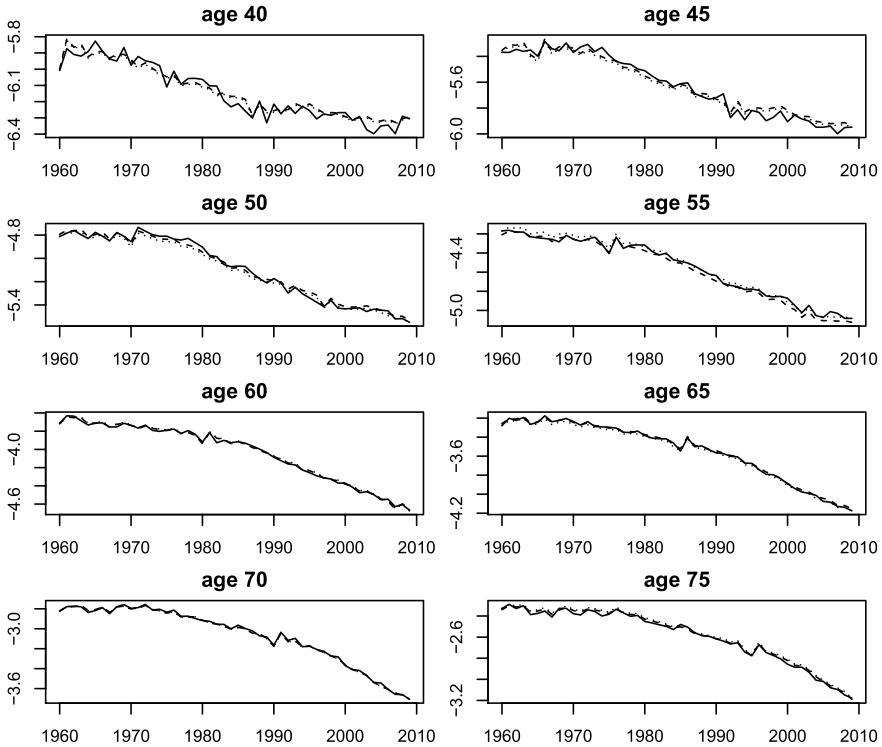


Fig. 15 UK males 1960–2009, ages 1–89. Integration: comparison of empirical $\log(\text{rates})$ (continuous-line) and fitted $\log(\text{rates})$ using model (1) improvements in combination with (i) $A(x)$ from model (3) (broken—lines) and (ii) equivalent initial $\log(\text{rates})$ (dotted-lines), for specific ages

intervals reported in Haberman and Renshaw [6], using a variety of fixed effects model structures.

As further evidence of the horizontal trend about zero in the random effects period component, we have compared the effects of including an additional intercept parameter in the AR(1) time series using Approach B. The results are presented in Table 1, in which we have tabulated the AR(1) parameter estimates, standard errors and p-values, both with and without the intercept parameter included: and note the lack of statistical significance of the intercept parameter throughout.

We now revert to the full 1–89 age range, and an illustration of the 40(05)75 age specific empirical log mortality rates with simulated 5%, 50%, 95% quantile projections using model (1) and Approach A of Sect. 6 is presented in Fig. 18. These projections, based on 2000 simulations, have been generated by sampling the error in the period component time series. As has been widely discussed in the literature (see Lee and Carter [11], for example), this approach is an approximate one and does not take into account the smaller contribution from the uncertainty in the fixed effects parameter estimates. We have adopted this approach

Table 1 UK male mortality experience, various periods as listed in the first column, matching the upper frame in Fig. 11, ages 20–89

Period	Estimate	Std. error	p-value	Estimate	Std. error	p-value
1961–2009	– 0.4454 0.000598	0.1283 0.00281	0.00114 0.833	– 0.4457	0.1296	0.00122
1961–2007	– 0.4502 0.000797	0.1308 0.00291	0.00127 0.786	– 0.4499	0.1319	0.00136
1961–2005	– 0.4493 0.000755	0.1337 0.00303	0.00166 0.805	– 0.4492	0.1347	0.00174
1961–2003	– 0.4483 0.000897	0.1370 0.00315	0.00220 0.777	– 0.4479	0.1383	0.00235
1961–2001	– 0.4523 0.000746	0.1403 0.00327	0.00260 0.821	– 0.4524	0.1411	0.00264
1961–1999	– 0.4606 0.000949	0.1431 0.00337	0.00273 0.780	– 0.4603	0.1442	0.00284
1961–1997	– 0.4623 0.000800	0.1474 0.00354	0.00351 0.823	– 0.4625	0.1480	0.0035
1961–1995	– 0.4740 0.00135	0.1535 0.00372	0.00416 0.720	– 0.4722	0.1546	0.00435

Estimates, standard errors, and p-values for the AR(1) times series with an intercept parameter (second entry), columns 2–4, and without the intercept parameter, columns 5–7

on grounds of simplicity, with the justification that the uncertainty in the fixed effects parameter estimates would only make a marginal difference to the prediction intervals.

7 Discussion

Concerning Figs. 1 and 6, the choice of n when forming the n -year MIR rolling-averages is, in part, somewhat arbitrary. However, as n increases, further investigations (not included here) show that the emerging age patterns in the MIR become sharper in focus and display greater consistency. In addition, the patterns of tightly packed MIR age profiles (lower right panel Fig. 6) are found to extend backwards in time over the best part of the last half century before breaking up. This feature is broadly supported by the equivalent details of Fig. 1 which extends backwards in time much further.

The same mathematical formula connecting the two alternative MIR statistics of Sect. 2 continues to apply when the statistics are redefined from the cohort, as opposed to period perspective, to read as follows:

$$\text{Statistic I: } \tilde{y}_{x,t} = \frac{-\Delta_s m_{x,t}}{(m_{x-1,t-1} + m_{x,t})/2} = 2 \frac{(1 - m_{x,t}/m_{x-1,t-1})}{(1 + m_{x,t}/m_{x-1,t-1})}, \text{ and}$$

$$\text{Statistics II: } \tilde{z}_{x,t} = \Delta_s \log m_{x,t} = \log m_{x,t} - \log m_{x-1,t-1}.$$

See Haberman and Renshaw [7] for application of the former approach.

While noting that Schinzing et al. [18] use a different predictor structure, which is multiplicative bilinear as opposed to our additive linear structure, both

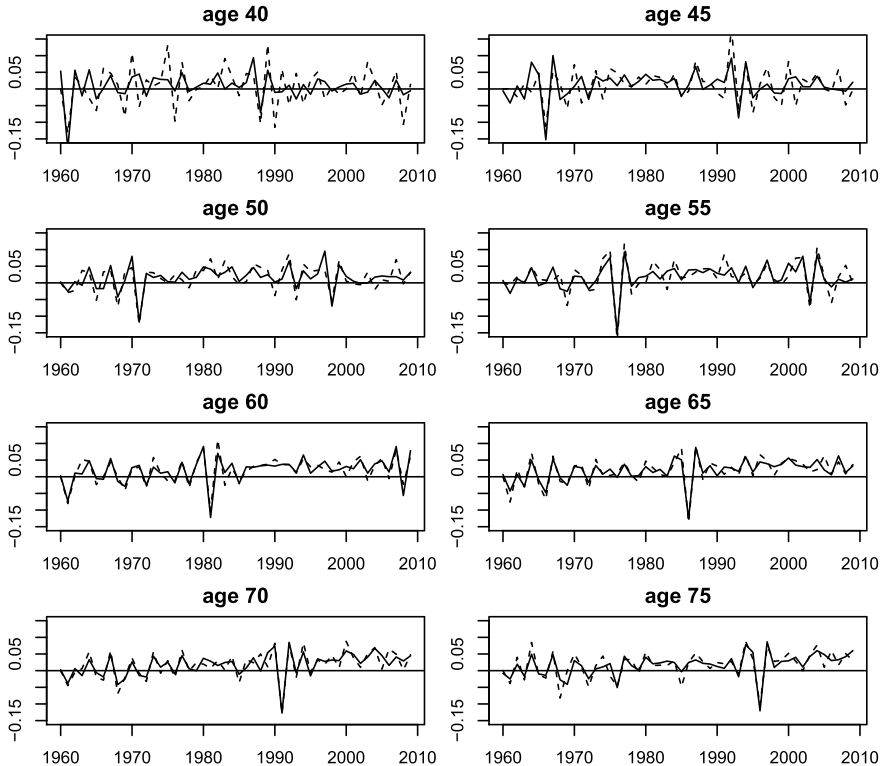


Fig. 16 UK males 1960–2009, ages 1–89. Differentiation: comparison of MIR empirical improvements (continuous line) and fitted improvements based on model (3) parameter estimates (broken-lines), for specific ages

approaches assume Normal MIR random variables. Of particular relevance, we refer to the Centre Right panel of their Fig. 6.1 in which the annual age-aggregated empirical MIR, based on ages 21–100, for periods 1971–2011 are depicted for UK males (and females).

Using our slightly shorter version of the data, and the MIR Statistic II in common with Schinzingler et al. [18], the same results (males only), for the overlapping periods 1971–2009, are reproduced in the Upper Left panel of Fig. 19. In addition, a scatter-plot of the time-adjacent age-aggregated MIR is depicted in the Upper Right panel, to illustrate the degree and nature of the correlation between these two measures, as discussed in Sect. 6.1 of Schinzingler et al. [18]. The equivalent results using the MIR Statistic I are depicted in the two Lower panels of Fig. 19. The comparison of matching upper and lower left panels reveals that the one is the mirror image, in the x-axis, of the other, for reasons given in

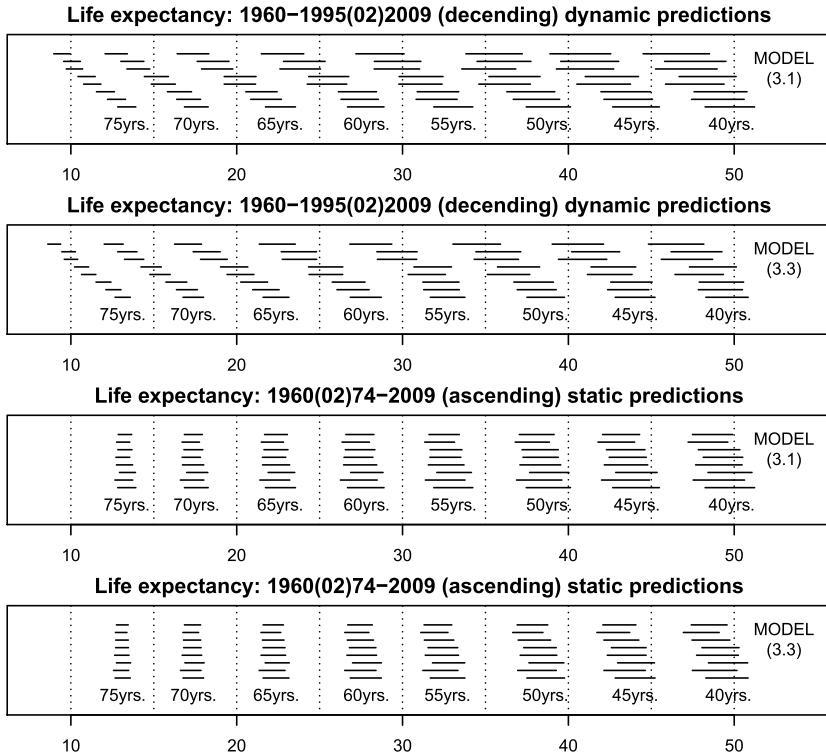


Fig. 17 UK males 1960–2009, ages 20–89. Extrapolated life expectancy 5%, 50%, 95% quartiles (1000 simulations). Upper panels: evolving 1995(02)09 dynamic extrapolations (decending sequence). Lower panels: static 2009 extrapolations subject to front end 1960(02)72 data deletions (ascending sequence). Juxtaposed models (1) and (3)

Sect. 2; while the upwards trend in the pattern in the lower left panel accurately portrays the improvement.

Denuit and Trufin [4] provide a further interesting example of a HGLM with random effects period component, without classifying it as such. Thus, as a means of projecting a regulatory life table, an exponential decline model

$$m_{x,t} = m_{x,t-1} \rho_x \Lambda_t, \quad \rho_x = \beta_x \delta$$

subject to stipulated constraints is formulated, in which the mortality rates $m_{x,t}$ are subjected to a sequence of annual random shocks Λ_t . Thus

$$\hat{m}_{x,t} | \Lambda_t \sim Poi(m_{x,t})$$

with log-link, so that

$$\log m_{x,t} = \log m_{x,t-1} + \log \delta + \log \beta_x + \log \Lambda_t$$

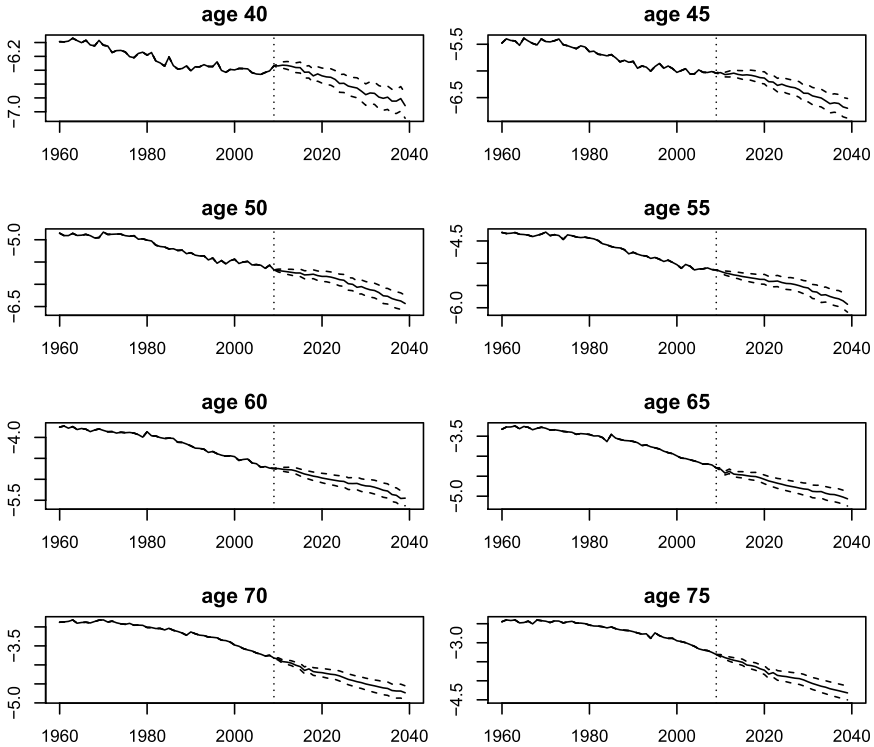


Fig. 18 UK males 1960–2009, ages 1–89. Age specific empirical $\log(\text{rates})$ with simulated 5%, 50%, 95% quantile projections, using model (1)

where the first term on the RHS is treated as an offset $\log \hat{m}_{x,t-i}$, the next two fixed effects terms are estimated, and the final term $v_t = \log \Lambda_t$ is modelled as a normal random effects variable (so that Λ_t has the log-normal distribution); with estimates subject to the stipulated constraints.

The philosophy underpinning this approach to generalised linear modelling with random effects and the associated likelihood theory, which we have followed and which is described in Lee et al. [12], is non-Bayesian.

In addition to providing the detailed methodology on which Sects. 4–6 are based, the methodology of Lee et al. [12] could be extended to the incorporation of smoothing, including smoothing by beta-splines.

We note that there is a long established practice of using the Poisson distribution to model numbers of deaths and hence central mortality rates. This starts historically with a static setting in the construction and graduation of life-tables and moves forward to a dynamic setting as here and in the extensive literature on longevity modelling. As discussed in Section 1 of Hunt and Villegas [10], there would appear to be no such established practice for the direct modelling of central mortality improvement rates. Given our results and in particular the residual plots reported in the applications (and also by

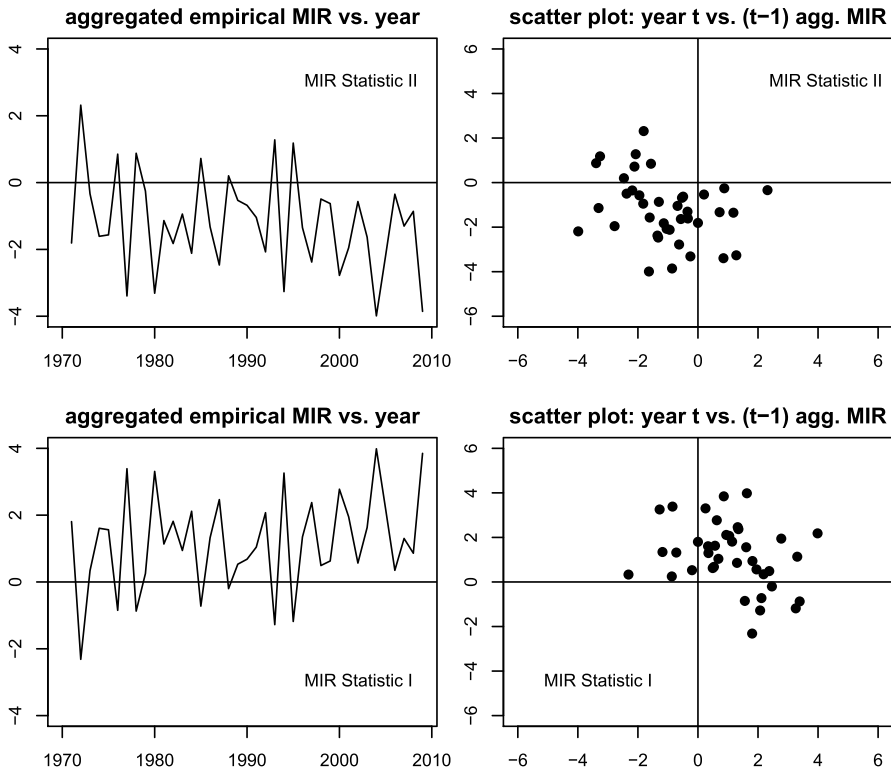


Fig. 19 UK males 1970–2009, ages 21–100. Upper left panel: MIR Type II Statistics aggregated over age for each calendar year. Upper right panel: associated scatter plot of the time adjacent aggregated MIR values. Lower panels: equivalent depictions using Type I Statistics

Haberman and Renshaw [6]), we believe there is adequate evidence for the use of the Gaussian distribution to model mortality improvement rates.

In their Sect. 3.4 of Hunt and Villegas [10] have identified an important inherent difficulty with modelling improvements directly, expressed in terms of a “crude” or “fitted” estimation approach. We conjecture that this manifests itself, in the current modelling framework, as the choice between estimating A_x as part of the fixed effects structure in (3) or treating $\hat{A}_x = \log \hat{m}_{x,t_1}$ as an offset term.

Further, this issue appears to be bound up with the difficulty of reversing the differentiation process which requires a suitable ‘constant’ of integration: a situation reminiscent of the problem of setting boundary values when integrating partial differential equations in mathematical physics. Setting aside model extrapolation, and given that the sequential differencing of rates is implied in the construction of improvements measures, the use of initial rates to reverse the process is a potential concern, and was the motivation behind the construction of Fig. 15 in Sect. 6.

When introducing a main effects cohort term into a dynamic parameterised predictor structure for modelling mortality rates (Renshaw and Haberman [16]), we have naturally assumed that the supporting Lexis plane was divided into unit squares, typically

of size 1 year. We have had the benefit of a data set with sufficient granularity to match this level of detail. In so doing, we have avoided some of the problems that were experienced by Mitchell et al. [14] who have needed to use grouped data and irregular grouped data in some cases: this has led to difficulties in modelling the parameterised cohort components.

We note that the “topping-out” procedure does have a material effect on the values of the projected life expectancies depicted in Fig. 17. We note also that the choice of fixed effects parameter constraints made for the three mixed effects models is not necessarily unique but is sufficient to ensure that the matrix T has full rank.

8 Concluding comments

The comparison of MIR Statistics (Sect. 2) indicates the greater accuracy of Statistic I. However, we have found that the replacement of Statistic I by Statistic II (subject to reversal of sign) does not induce any material consequential differences in our reported results.

By attributing random effects to the period and cohort components or just the period components of a main effects age-period-cohort structured linear predictor from the outset, we have described a comprehensive self-contained process for modelling and forecasting mortality improvement rates, which incorporates structured dispersion and an apparent self-selecting time series. This process is made possible through the implementation of HGLMs with random effects. The methodology also extends to the modelling and forecasting of mortality rates provided that the predictor structure is linear, and, as such, excludes alternatives that involve bilinear decomposition. We argue that this methodological framework leads to a more unified approach to model fitting and forecasting.

For the linear predictor structures under consideration, the choice of the associated time series models, AR(1) for MIR mixed effects modelling and ARI(1,1) for MR mixed effects modelling, follows as a direct consequence of the initial modelling assumptions. The application of these time series models in this context, which are shown to be well supported by the data, may be further generalised by increasing the number of autoregressive terms when this is justified by the practical application.

We note the usefulness of examining the patterns of the random effects residuals as these provide additional critical insight into whether the modelling assumptions are being adhered to.

We note also the usefulness of comparing the empirical and modelled age specific mortality improvement rates. For example, we note that the MIR age profile modelled patterns of Sect. 4 (Fig. 6) are broadly similar to the corresponding empirically generated MIR age profiles of Fig. 1.

There is a long established practice of using the Poisson distribution to model numbers of deaths and hence central mortality rates. Given our detailed modelling results, we believe there is adequate empirical evidence for the use of the Gaussian distribution to model mortality improvement rates.

A word of caution is needed in the use of the Sect. 5 methodology for the long term forecasting of mortality rates and life expectancies in so far as it may not be sufficiently flexible to reflect recently emerging trends in the underlying mortality rates and corresponding life expectancies for certain countries (see, for example, Case and Deaton [2] and Djeundje et al. [5]).

Appendix I

Iterative weighted least squares (IWLS) model fitting procedure:

1. Given $\tau = (\phi, \lambda)$
 update $\hat{\theta}$ by computing Σ_a^{-1} and solving $\mathbf{T}'\Sigma_a^{-1}\mathbf{T}\hat{\theta} = \mathbf{T}'\Sigma_a^{-1}y_a$.
 compute leverages (h_i, h_{Mi}) , diagonal elements of matrix $\mathbf{T}(\mathbf{T}'\Sigma_a^{-1}\mathbf{T})^{-1}\mathbf{T}'\Sigma_a^{-1}$.
2. Given θ

update the components of $\hat{\tau}$ sequentially using the Gamma GLMs-

GLM characteristic	Fixed effects ϕ	Random effects λ
Response	d_i^*	d_{Mi}^*
Mean	$E(d_i^*) = \phi$	$E(d_{Mi}^*) = \lambda$
Variance	$\text{var}(d_i^*) = 2\phi^2$	$\text{var}(d_{Mi}^*) = 2\lambda^2$
Link	$\xi = g(\phi)$	$\xi_M = g_M(\lambda)$
Linear predictor	γ	γ_M
Deviance component	$gamma(d_i^*, \phi)$	$gamma(d_{Mi}^*, \lambda)$
Prior weights	$(1 - h_i)/2$	$(1 - h_{Mi})/2$

where the respective responses

$$d_i^* = \frac{d_i}{1 - h_i}, \quad d_i = (y_i - X_i\hat{\beta} - Z_i\hat{v})^2; \quad d_{Mi}^* = \frac{d_{Mi}}{1 - h_{Mi}}, \quad d_{Mi} = \hat{v}_i^2$$

the deviance components

$$gamma(d_i^*, \phi) = 2\{-\log(d_i^*/\phi) + (d_i^* - \phi)/\phi\}$$

and $i = (x, t)$ identifies matching data cells and matrix components.

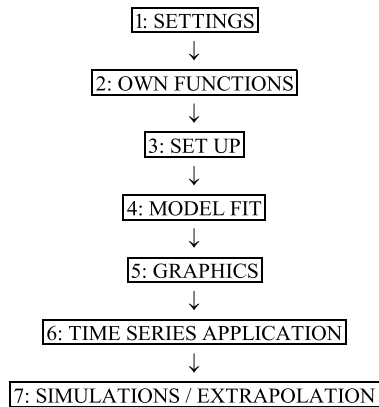
Starting values of the order 0.01 are allocated to the components of $\tau = (\phi, \lambda)$, while termination of the iterative process is controlled by the convergence of the three GLM deviances (and or convergence of the absolute difference in the sequential maximum scale parameter: thus $\max_i |\tau_{i,j} - \tau_{i,j-1}| < tolerance$ for iterations $j = 1, 2, 3, \dots$ and pre-set tolerance). Both the canonical reciprocal link and log link have been found to work

equally well. For uniform dispersion, the linear predictors γ and γ_M are set to be constants, while variable fixed effects dispersion by age, is possible by setting the linear predictors and starting values accordingly. When this is the case, ϕ_i replaces ϕ .

Appendix II

(Outline details of the R-program used in this paper).

The key sections are as follows



1. The settings comprise (1) *scalars* controlling the geometry of the data array, the extent of period-wise extrapolation, the number of simulations, the choice of MIR statistic and model type, (2) *vectors* controlling the choice of ages and quantiles when extrapolating.
2. The collection of own functions includes functions for returning (1) MIR statistics, (2) age-specific period-wise life expectancies, (3) ARI(p,d) time series with forecasts and error forecasts.
3. The set up involves (1) data scanning and (2) *vector* creation of different lengths and types, indexing the annual data cells by age, period and cohort
4. The model fit first requires the construction of *matrices* O , I , X , Z , T , y_a . The R-command library(lme4) aids this. For Poisson-Normal mixed modelling an additional N by 1 matrix *mu*, using empirical mortality rates for starter values, is required to form the adjusted dependent variable, expression (6) The *matrices* Σ_a and θ , together with a diagonal matrix of leverages evolve as part of the iterative fitting procedure (“Appendix I”). Details of the age/period/cohort component effects and residuals are extracted for display.
5. The graphics comprise (1) residual and (2) component fixed/random effects plots.
6. In addition to returning forecasts with root mean square errors, a graphical facility has been included in the ARI own function.

7. For pre-selected ages and future periods, the simulation process results in matrices of simulated (1) log mortality rates and (2) period-wise life expectancies. For MIR modelling, an additional step is applied to convert from improvement rate projections to log mortality rate projections. The number of simulations equates with the number of columns in these matrices, which readily reduce in size on taking quantiles. Graphics are included.

Funding Not applicable.

Availability of data and material The empirical data are publicly available from the Human Mortality Database.

Code availability The code is open source and written in R statistical software. The code is available upon request from the authors.

Compliance with ethical standards

Conflicts of interest We declare that we have no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Bohk-Ewald C, Rau R (2017) Probabilistic mortality forecasting with varying age-specific survival improvements. *Genus* 73:1
2. Case A, Deaton A (2015) Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. *Proc Natl Acad Sci USA* 112:15078–15083
3. Continuous Mortality Investigation (2016) CMI Mortality Projections Model-Working Paper 90. In: Institute and Faculty of Actuaries. <https://www.actuaries.org.uk/learn-and-develop/continuous-mortality-investigation/cmi-working-papers/mortality-projections/cmi-working-papers-90-91-and-93>
4. Denuit M, Trufin J (2016) From regulatory life table to stochastic mortality projections: the exponential decline model. *Insurance Math Econ* 71:295–303
5. Djeundje VB, Haberman S, Bajekal M, Lu J (2020). An analysis of mortality trends in developed countries, focusing on the recent slowdown in mortality improvements. In: Longevity Science Panel working paper. <https://www.longevitypanel.co.uk/>
6. Haberman S, Renshaw AE (2012) Parametric mortality rate modelling and projecting. *Insurance Math Econ* 50:309–333
7. Haberman S, Renshaw AE (2013) Modelling and projecting mortality improvement rates using a cohort perspective. *Insurance Math Econ* 53:150–168
8. Haberman S, Kaishev VK, Millosovich P, Villegas AM, Baxter SD, Gaches AT, Gunlaussen S, Sison M (2014) Longevity basis risk: a methodology for assessing basis risk. Technical Report, Cass Business School and Hymans Robertson LLP. Presented to Institute and Faculty of Actuaries. <https://www.actuaries.org.uk/learn-and-develop/conference-paper-archive/2014>

9. Human Mortality Database (2019) www.mortality.org
10. Hunt A, Villegas AM (2020) Mortality improvement rates: modelling, parameter uncertainty and robustness. In: ARC Centre of Excellence in Population Ageing Research Working Paper 2020/28
11. Lee RD, Carter L (1992) Modelling and forecasting the time series of US mortality projection models. *J Am Stat Assoc* 87:659–671
12. Lee Y, Nelder JA, Pawitan Y (2006) Generalised linear models with random effects. Chapman & Hall
13. Li JS-H, Chan WS, Cheung S-H (2011) Structural changes in the Lee-Carter mortality indexes: detection and implementations. *N Am Actuarial J* 15(1):13–31
14. Mitchell D, Brockett PL, Mendoza-Arriaga R, Muthuraman K (2013) Modelling and forecasting mortality rates. *Insurance Math Econ* 52:275–285
15. R core team (2017) A language and environment for statistical modelling. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
16. Renshaw A, Haberman S (2006) A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics* 38(3):556–570
17. Richards SJ, Currie ID, Kleinow T, Ritchie GP (2017) A Stochastic Implementation of the APCI model for mortality projections. www.longevitas.co.uk
18. Schinzinger E, Denuit MM, Christiansen MC (2016) A multivariate evolutionary credibility model for mortality improvement rates. *Insurance Math Econ* 69:70–81
19. Villegas AM, Haberman S, Kaishev VK, Millosovich P (2017) A comparative study of two population models for the assessment of risk bias in longevity hedges. *ASTIN Bull* 47(3):631–679

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.