Insurance: Mathematics and Economics 65 (2015) 156–171

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Multivariate time series modeling, estimation and prediction of mortalities

Erland Ekheden *, Ola Hössjer [1]

Department of Mathematics, Div. of Mathematical Statistics, Stockholm University, Stockholm, SE 106 91, Sweden

ARTICLE INFO

Article history:
Received March 2014
Received in revised form
September 2015
Accepted 24 September 2015
Available online 9 October 2015

MSC:
62M10
62P05

Keywords:
Best linear unbiased predictor
Generalized least squares
Longevity
Mortality prediction
Multivariate time series
Overdispersion

ABSTRACT

We introduce a mixed regression model for mortality data which can be decomposed into a deterministic trend component explained by the covariates age and calendar year, a multivariate Gaussian time series part not explained by the covariates, and binomial risk. Data can be analyzed by means of a simple logistic regression model when the multivariate Gaussian time series component is absent and there is no overdispersion. In this paper we rather allow for overdispersion and the mixed regression model is fitted to mortality data from the United States and Sweden, with the aim to provide prediction and intervals for future mortality and annuity premium, as well as smoothing historical data, using the best linear unbiased predictor. We find that the form of the Gaussian time series has a large impact on the width of the prediction intervals, and it poses some new questions on proper model selection.

1. Introduction

Decreasing mortality rates is not a new phenomenon, it is a trend that has been evident for over a century in the western world and today it is present in practically all countries but the ones worst plagued with civil wars and the AIDS epidemic. "Longevity" is an often used term for this trend, especially when the trend is viewed as an economic risk putting stress on pension plans and health care systems.

Actuaries and demographers have a long tradition of making life tables and models for mortality, trying to determine the death intensity or force of mortality, denoted μ_x at age x. Closely related is the one year death risk

$$Q_x = 1 - \exp\left(\int_0^1 \mu_{x+s} ds\right),$$

for which we use a capital letter (rather than q_x) in order to emphasize its randomness. After age is gender the most important factor for human mortality. Often separate life tables and sets of Q_x are produced for males and females, but for pricing purposes in the EU, a gender neutral life table must be used in order not to discriminate between genders.

In order to quantify dynamic effects, mortality rates and death intensities should be regarded as functions of calendar year t as well. For instance, Osmond (1985) introduced the Age–Period–Cohort model within the medical statistics literature. But the interest in stochastic modeling of mortality first took off with a paper by Lee and Carter (1992) in which a principal components approach of Bozik and Bell (1987) and Bell and Monsell (1991) was modified. Since then a variety of models have been proposed. They differ in several ways, for instance in how age, calendar year and cohort $t - x$ are included as covariates, see Renshaw and Haberman (2003b,c, 2006), Booth and Tickle (2008), Cairns et al. (2008, 2009), Barrieu et al. (2012) and Cairns (2014) for recent overviews with further references.

The richness of proposed models shows that the problem is non-trivial, with a high dimensional data set. There are more than hundred observed age specific mortalities, often gender specific for males and females, collected for over thirty, fifty and even hundred calendar years. Still there are obvious patterns of correlation in data, in that mortality in general increases with age. The improvements of mortalities seem to be non-stationary though,

* Correspondence to: SPP Konsult, Stockholm, SE 105 39, Sweden. Tel.: +46 70 7704685.
E-mail addresses: e.ekheden@gmail.com (E. Ekheden), ola@math.su.se (O. Hössjer).
[1] Tel.: +46 70 6721218.

http://dx.doi.org/10.1016/j.insmatheco.2015.09.013

in that the rates vary over ages and time. On top of this we have random noise, caused by individual variation in a finite population.

When evaluating models, some seem to be too simple. This may either be assessed in an explorative data analysis which may reveal marked patterns in residual plots that signify features of historical data not explained by the model, or formally by some model selection criterion such as BIC (Cairns et al., 2009). Other models seem to be too complex. Even though they fit historical data well, they might be sensitive to variations in data and have less robust forecasts, see Cairns et al. (2009) and Cairns (2014). Bell (1997) showed that a simple model, where the logged death rates constitute a random walk with drift, separately for each age, can sometimes outperform more complex models in terms of forecasting. Bell's work has received relatively little attention in the literature, and there is still work to be done, it seems, in terms of selecting models for mortality and forecasting.

In the Lee–Carter model and many of its successors, it is often taken for granted that either the *observed* death rates $\hat{\mu}_{xt}$ or mortality rates $\hat{Q}_{xt}$ are stochastic processes. It is however seldomly explicitly pointed out that what we observe is a finite population and that the randomness is partially caused by this. Brouhns et al. (2002) used Poisson regression, where instead, the *actual* death rates are non-random, whereas the randomness from the finite population manifests itself in terms of a Poisson distributed number of deaths (see also Wilmoth, 1993; Alho, 2000). This source of variation has been referred to as the Poisson risk (Cairns et al., 2014), and analogously, we speak of the binomial risk if the number of deaths is assumed have a binomial distribution. Both Poisson and binomial risks are examples of diversifiable risks, i.e. they diminish as population grows due to the law of large numbers.

Ekheden and Hössjer (2014) studied the randomness in observed mortality rates $\hat{Q}_{xt}$ more closely, with the aim of getting a better understanding of the underlying processes and to find new means for mortality estimation and prediction. More specifically, the focus in that paper was on variation of logit transformed $\hat{Q}_{xt}$, which were split into three components, systematic risk explained by the covariates, systematic risk not explained by the covariates and binomial risk, due to the finiteness of the population. The second type of systematic risk corresponds to real effects not caught by the model, and the formal test procedure developed by Ekheden and Hössjer revealed that for a small population, it can often be discarded, due to its large binomial risk.

In this paper we propose more specifically a mixed regression model which can be decomposed into a deterministic trend component explained by the covariates, a multivariate Gaussian time series part not explained by the covariates and binomial risk. The main novelty is the time series part, which causes mortality rates to be overdispersed. The mortalities of a binomial model without such a random effect are actually underdispersed, since the variance of a binomial distribution is smaller than its mean. But a mixed binomial distribution is overdispersed, unless the random effect is very small.

The multivariate time series is further divided into three variance components; corresponding to white noise due to for instance a heterogeneous population, period effects due to for instance natural catastrophes or influenzas, and a random walk that incorporates long term departures from the deterministic trend. In absence of this time series, our model reduces to the simple logistic regression that Ekheden and Hössjer (2014) advocated for small countries. Here we include the time series into the analysis and estimate systematic explained risk from logit transformed data by means of an iteratively reweighted version of Aitken's generalized least squares (GLS) estimator (Aitken, 1935) and systematic explained and unexplained risk by means of a best linear unbiased predictor (BLUP, Henderson, 1975, Robinson, 1991). Prediction of future mortality rates, life annuity premiums and life expectancies are also incorporated into our Gaussian framework in a straightforward way that does not require any resampling.

Other overdispersion models include using negative binomial distributions (Delwarde et al., 2007a; Li et al., 2009) and generalized linear models with overdispersed Poisson data (Renshaw and Haberman, 2003b; Djeundje and Currie, 2010) for which parameters can be estimated by extended quasi likelihood methods. A particular feature of our approach is that overdispersion, or unexplained systematic risk, enters as a stochastic process, for which a rather general covariance structure is allowed.

The paper is organized as follows: The mixed regression model is introduced in Sections 2–3, and estimates/prediction of historic/future mortalities in Section 4. Section 5 presents analysis for one Swedish and one US data set, Section 6 contains a discussion and the Appendix, finally, provides mathematical details.

2. Regression model for mortality rates

We study a population of ages $x = x_l, \ldots, x_u$ spanning between lower and upper limits x_l and x_u, during calendar years $t = t_1, \ldots, t_T$, where t_T is the latest year of observations and T is the length of the time window.

Let N_{xt} be the initial exposure-to-risk for individuals of age x alive at the beginning of calendar year t. As in Cossette et al. (2007), we model the number of deaths

$$D_{xt}|Q_{xt} \sim \text{Bin}(N_{xt}, Q_{xt})$$

among these individuals within one year with a binomial distribution, whose death probability or mortality rate Q_{xt} can be estimated as

$$\hat{Q}_{xt} = \frac{D_{xt}}{N_{xt}}. \tag{1}$$

As mentioned in the introduction, it is also possible to assume a Poisson distribution for death counts, with mean proportional to the central exposure-to-risk E_{xt} rather than N_{xt}, see for instance Brouhns et al. (2002, 2005). This is a useful approximation for most ages, but for higher ages, over 80, the Poisson distribution increasingly overestimates the variance, making it less suitable for our purposes. There is no general guideline as to which type of distribution to employ. Currie (2013) used a data set from the UK Office for National Statistics that included some individuals that reached a very high age, and obtained a substantially better fit for the binomial model. On the other hand, Cossette et al. (2007) analyzed population data from the Canadian province of Quebec. They found that the Poisson model, which takes exposure to risk into account, gave a slightly better fit.

Logit transformed mortalities were originally used by Brass (1971) for a one-factor age model. In this paper, we assume a two-factor model

$$\text{logit}Q_{xt} = \log \frac{Q_{xt}}{1 - Q_{xt}} = \alpha_x + \beta_x(t - \tilde{t}) + \varepsilon^s_{xt} \tag{2}$$

with age and calendar years included, whereas cohort effects $t - x$ are not. Similarly as for Generalized Linear Models (Renshaw and Haberman, 2003a), parameters enter linearly in (2), with time as a known covariate. This deterministic period effect is linear in t, with a parametrization centered around calendar year $\tilde{t}$, which can be chosen after convenience, for instance the starting point ($\tilde{t} = t_1$), the mid point ($\tilde{t} = (t_1 + t_T)/2$) or the end point ($\tilde{t} = t_T$) of the chosen time interval, depending on whether the purpose is to fit historical mortalities or to predict future ones. See Section 6 for an extensive discussion of the linear time trend assumption.

The intercepts α_x and slopes β_x represent deterministic age effects, for which we assume a parametrization

$$\alpha_x = \sum_{j=0}^{p_1} a_j \phi_j(x),$$

$$\beta_x = \sum_{j=0}^{p_2} b_j \phi_j(x)$$

in terms of basis functions ϕ_j. In order to capture global agewise effects, one could use polynomials,

$$\phi_j(x) = x^j, \tag{3}$$

whereas single age class indicators

$$\phi_j(x) = 1_{\{x = x_{j+l}\}} \tag{4}$$

will capture age-specific effects, where each age class is assigned a separate intercept and slope parameter, with $p_1 = p_2 = x_u - x_l$, $a_j = \alpha_{x_l+j}$ and $b_j = \beta_{x_l+j}$. A compromise between (3) and (4) is to use B-splines (Eilers and Marx, 1996; Imoto and Konishi, 2003) in order to smooth locally. A first degree B-spline, for instance, has intercept and slope parameters that vary piecewise linearly with age. This corresponds to choosing $m \le (x_u - x_l + 1)$ knot points $x^0 = x_l < x^1 < \cdots x^{m-1} = x_u$, and

$$\phi_j(x) = \begin{cases} \left(1 - (x^j - x)/(x^j - x^{j-1})\right)_+, & x_l \le x \le x^j, \\ \left(1 - (x - x^j)/(x^{j+1} - x^j)\right)_+, & x^j \le x \le x_u, \end{cases} \tag{5}$$

with $z_+ = \max(0, z)$ the positive part function, and $p_1 = p_2 = m - 1$.

The terms ε^s_{xt} are random variables with $E(\varepsilon^s_{xt}) = 0$ that represent unexplained systematic variation, with 's' an acronym for systematic. One possibility is to model ε^s_{xt} in terms of a Gaussian Markov random field (Rue and Held, 2005), since this would facilitate statistical inference. We will use another approach though for which covariance matrices with all $\mathrm{Cov}(\varepsilon^s_{xt}, \varepsilon^s_{x't'})$ are less tractable (more difficult to invert), but on the other hand the parameters have a more straightforward interpretation; in more detail, we assume that $\boldsymbol{\varepsilon}^s_t = \{\varepsilon^s_{tx};\ x_l \le x \le x_u\}$ for all ages at calendar year t is written as a multivariate Gaussian time series

$$\boldsymbol{\varepsilon}^s_t = \boldsymbol{\eta}_t + \boldsymbol{c}\zeta_t + \boldsymbol{d} \sum_{s=1}^{|t-\tilde{t}|} \kappa_{\tilde{t}+\mathrm{sgn}(t-\tilde{t})s}, \tag{6}$$

where $\boldsymbol{c} = \{c_x;\ x_l \le x \le x_u\}$ and $\boldsymbol{d} = \{d_x;\ x_l \le x \le x_u\}$ are deterministic vectors, $\boldsymbol{\eta}_t = \{\eta_{xt};\ x_l \le x \le x_u\}$, ζ_t and κ_t are random, and $\mathrm{sgn}(x)$ equals -1, 0 or 1 depending on whether x is negative, zero or positive.

We will assume that $\{\eta_{xt}\}$, ζ_t and κ_t are independent and normally distributed random variables with mean zero and variances σ^2_η, σ^2_ζ and σ^2_κ. Then the right hand side of (6) involves three components, the first of which is white noise and it is caused, for instance, by a heterogeneous population with a varying mortality rate. The second term represents period effects, such as catastrophes and influenzas, that effect many age classes in a similar way, as specified by $\boldsymbol{c}$. The third term is a two-sided random walk term that incorporates random departures from a linear trend, common to all age classes. This process vanishes at time $\tilde{t}$ (assuming in (6) that $\sum_1^0 = 0$) and starts in two directions from this time point.

As mentioned in Section 1, a number of other methods of mortality estimation have been proposed. Many of those that do not include cohort effects have the form

$$h(Q_{tx}) = \alpha_x + \beta_{1x}\kappa_{1t} + \beta_{2x}\kappa_{2t},$$

where h is a link function, $\{\alpha_x\}$, $\{\beta_{1x}\}$, $\{\beta_{1x}\}$ are age-specific and $\{\kappa_{1t}\}$, $\{\kappa_{2t}\}$ time specific parameters. Some constraints are imposed on these parameters in order to assure their identifiability. Mortalities are predicted using the estimated age-specific parameters, with future time parameters modeled stochastically as a random walk, or more generally an ARIMA time series (Brockwell and Davis, 1991). Cossette et al. (2007) use a link function $h(q) = \log[\log(1/(1-q))]$ and a Lee–Carter parametrization ($\beta_{2x} \equiv 0$). The Cairns–Blake–Dowd model of Cairns et al. (2006) has a logit link function $h(q) = \mathrm{logit}(q)$, and age parameters $\alpha_x \equiv 0$, $\beta_{1x} \equiv 1$, $\beta_{2x} = x$. See also Cairns et al. (2009) for a comparison of a number of models, including LC and CBD. A difference of our curve fitting model (2) and (6) is that time enters as a known covariate, so that linear (rather than bilinear) regression techniques can be used for estimation. But the major advantage of our approach is that historical and future data are both mixed regression models with general covariance structure. This makes it possible to obtain prediction intervals of future mortalities and life expectancies analytically, with future mortality change uncertainties *and* parameter estimation errors accounted for.

3. Mixed regression model for observed mortality rates

Our approach is different from Generalized linear models or Generalized linear mixed models, in that we transform the estimated mortalities $\hat{\boldsymbol{Q}} = \{\hat{Q}_{xt};\ (x,t) \in \Omega\}$ rather than their (conditionally) expected values, for a collection

$$\Omega \subset \{(x,t);\ x_l \le x \le x_u, t_1 \le t \le t_T\}$$

of ages and calendar years. This provides logit transformed mortality (LM) data

$$Y_{xt} = \mathrm{logit}\hat{Q}_{xt} = \log \frac{\hat{Q}_{xt}}{1 - \hat{Q}_{xt}}. \tag{7}$$

The analogous transformations of the true but unknown mortalities $\boldsymbol{Q} = \{Q_{xt};\ (x,t) \in \Omega\}$, are denoted as

$$Y^\infty_{xt} = \mathrm{logit}(Q_{xt})$$

for $(x,t) \in \Omega$, where superscript ∞ signifies a hypothetical population of infinite size with no diversifiable or binomial risk.

This gives a mixed regression model with response variables Y_{xt}, covariates (x,t) and parameters $\boldsymbol{\theta}$. In order to assess how much of the variation in Y_{xt} that is a function of changes in the underlying $\boldsymbol{Q}$, not explained by our model (systematic variation) and how much is due to random noise (binomial risk), we write

$$\begin{aligned} Y_{xt} &= m_{xt} + \varepsilon_{xt} \\ &= m_{xt} + \varepsilon^s_{xt} + \varepsilon^b_{xt}, \end{aligned} \tag{8}$$

as a sum of one part

$$\begin{aligned} m_{xt} &= m_{xt}(\boldsymbol{\theta}) \\ &= E_{\boldsymbol{\theta}}(Y^\infty_{xt}) \\ &= \sum_{j=0}^{p_1} a_j \phi_j(x) + (t - \tilde{t}) \sum_{j=0}^{p_2} b_j \phi_j(x) \end{aligned} \tag{9}$$

explained by a regression model, and another part ε_{xt} not explained by the regression model. We notice that the explained part depends on $p = p_1 + p_2 + 2$ regression parameters

$$\begin{aligned} \boldsymbol{\theta} &= (\theta_1, \ldots, \theta_p) \\ &= (a_0, \ldots, a_{p_1}, b_0, \ldots, b_{p_2})^T. \end{aligned}$$

The unexplained (random) part can further be decomposed into a sum of

$$\varepsilon^s_{xt} = Y^\infty_{xt} - m_{xt}, \tag{10}$$

the unexplained systematic variation (-risk) defined in (6), which by definition satisfies $E(\varepsilon^s_{xt}) = 0$, and

$$\begin{aligned}\varepsilon^b_{xt} &= Y_{xt} - Y^{\infty}_{xt}\\ &= \text{logit}\hat{Q}_{xt} - \text{logit}Q_{xt},\end{aligned} \tag{11}$$

the unexplained random noise, corresponding to binomial risk (with 'b' an acronym for binomial). By means of a second order Taylor expansion of the logit function $g(q) = \text{logit}(q)$ at $q = Q_{xt}$, and from the binomial variance formula $\text{Var}(\hat{Q}_{xt}|Q_{xt}) = Q_{xt}(1 - Q_{xt})/N_{xt}$, we find that

$$\begin{aligned}E(\varepsilon^b_{xt}) &= E\left[\frac{1}{2}g''(Q_{xt})\text{Var}(\hat{Q}_{xt}|Q_{xt})\right] + o\left(\frac{1}{N_{xt}}\right)\\ &= E\left[\frac{Q_{xt} - \frac{1}{2}}{Q_{xt}(1 - Q_{xt})}\right]\cdot\frac{1}{N_{xt}} + o\left(\frac{1}{N_{xt}}\right)\\ &= \int_0^1 \frac{q - \frac{1}{2}}{q(1-q)} f_{Q_{xt}}(q)dq\cdot\frac{1}{N_{xt}} + o\left(\frac{1}{N_{xt}}\right)\\ &= o\left(\frac{1}{\sqrt{N_{xt}}}\right).\end{aligned} \tag{12}$$

In the last step of (12) we used (2) and (8) to rewrite the mortality rate as

$$Q_{xt} = \frac{\exp(m_{xt} + \varepsilon^s_{xt})}{1 + \exp(m_{xt} + \varepsilon^s_{xt})}.$$

Since ε^s_{xt} is normally distributed according to (6), we use tail estimates of the normal distribution to deduce that the density function $f_{Q_{xt}}(q)$ of Q_{xt} is $o(q^c)$ as $q \to 0$, and $o((1-q)^c)$ as $q \to 1$, for any $c > 0$, so that the integral in (12) is finite. This proves the last step of (12), which makes the bias term $E(\varepsilon^b_{xt})$ negligible for large populations, since the standard deviation of ε^b_{xt} is of order $N_{xt}^{-1/2}$. We deduce this by looking at the variance,

$$\begin{aligned}\text{Var}(\varepsilon^b_{xt}) &= \text{Var}\left[E(\varepsilon^b_{xt}|Q_{x,t})\right] + E\left[\text{Var}(\text{logit}\hat{Q}_{xt}|Q_{x,t})\right]\\ &= \text{Var}\left[\frac{Q_{xt} - \frac{1}{2}}{Q_{xt}(1 - Q_{xt})}\right]\cdot\frac{1}{N^2_{xt}}\\ &\quad + E\left[g'(Q_{xt})^2\text{Var}(\hat{Q}_{xt}|Q_{x,t})\right]\\ &= o\left(\frac{1}{N_{xt}}\right) + E\left[\frac{1}{Q_{xt}(1 - Q_{xt})}\right]\cdot\frac{1}{N_{xt}},\end{aligned} \tag{13}$$

where in the second step we used (12) for the first term, and a first order Taylor expansion of $g(q)$ around $q = Q_{xt}$ for the second term, as for the delta method. In the third step we used that the variance term is finite, by means of a similar argument as below (12). Repeating this argument once again, we finally note that $E[1/(Q_{xt}(1 - Q_{xt}))]$, on the last line of (13), is finite as well.

In the Lee–Carter model and many of its extensions, age and period parameters enter bilinearly into the regression function. However, since time enters as a fixed known covariate in terms of a linear time trend in (2), it is possible to rewrite (8) as a multiple linear mixed regression model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{14}$$

where $\boldsymbol{Y} = (Y_{xt};\ (x,t) \in \Omega)^T$ and $\boldsymbol{\varepsilon} = (\varepsilon_{xt};\ (x,t) \in \Omega)^T$ are $n \times 1$ column vectors of observations and errors, $n = T(x_u - x_l + 1)$ is the number of elements of Ω, and $\boldsymbol{X}$ an $n \times p$ design matrix with rows

$$\begin{aligned}&(\phi_0(x), \phi_1(x), \ldots, \phi_{p_1}(x), (t - \tilde{t})\phi_0(x),\\ &\quad (t - \tilde{t})\phi_1(x), \ldots, (t - \tilde{t})\phi_{p_2}(x))\end{aligned}$$

for all $(x,t) \in \Omega$.

4. Estimating and predicting mortality rates

Since Q_{xt} involves unexplained systematic risk, it is random, and therefore it is of interest to estimate $E(Q_{xt})$ and predict Q_{xt} from historical data, as well as predicting future values of $\hat{Q}_{xt}$ and annuity premiums.

4.1. Historical mortality rates

In order to assess historic mortality rates, one option would be to start estimating model parameters from untransformed data D_{xt}. However, the maximum likelihood function is quite complicated with such an approach, since the systematic unexplained risk $\boldsymbol{\varepsilon}^s$ enters as a hidden variable. If the components of $\boldsymbol{\varepsilon}^s$ are independent, we get a generalized (or hierarchical) linear mixed model, for which various approximate estimation algorithms are available (Breslow and Clayton, 1993; Lee et al., 2006), whereas likelihood inference is less tractable when $\boldsymbol{\varepsilon}^s$ involves serial correlation.

We will rather extend the curve fitting approach, reviewed in Section 3 of Bell (1997). The expected mortality rates are functions of the parameters $\boldsymbol{\theta}$ of the multiple linear mixed regression model in (14). We estimate the expected mortality rates by plugging in estimates of $\boldsymbol{\theta}$ from an iteratively reweighted GLS, usually referred to as feasible generalized least squares (FGLS), see for instance Parks (1967) and Hansen (2007). Iteration is needed since the covariance matrix of the unexplained errors $\boldsymbol{\varepsilon} = (\varepsilon_{xt};\ (x,t) \in \Omega)^T$ is unknown and has to be estimated simultaneously with the regression parameters.

For historical data, only observed mortality rates are available, whereas the actual mortality rates are unknown random variables that can be predicted by means of a best linear unbiased predictor (BLUP). In this way we get separate FGLS and BLUP estimates of the expected and actual mortality rates.

We will allow for quite general dependency structures of $\boldsymbol{\varepsilon}$. Its covariance matrix can be written as a sum

$$\boldsymbol{V} = \boldsymbol{V}_{\boldsymbol{\varepsilon}^s} + \boldsymbol{V}_{\boldsymbol{\varepsilon}^b} \tag{15}$$

of two terms, of which the second $\boldsymbol{V}_{\boldsymbol{\varepsilon}^b}$ is the expected covariance matrix of the binomial risk error terms $\boldsymbol{\varepsilon}^b = \{\varepsilon^b_{xt};\ (x,t) \in \Omega\}$. It is shown in the Appendix how its form can be deduced from (13). The first term $\boldsymbol{V}_{\boldsymbol{\varepsilon}^s}$ of (15) is the covariance matrix of the systematic unexplained errors (6). We will assume that $\boldsymbol{c}$ and $\boldsymbol{d}$ are fixed in this equation, so that the only variance parameters that need to be estimated are

$$\boldsymbol{\xi} = (\sigma^2_\eta, \sigma^2_\zeta, \sigma^2_\kappa)^T. \tag{16}$$

Simultaneous maximum likelihood estimation of $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ (Harville, 1977) can be achieved by an iterative GLS procedure for both sets of parameters (Goldstein, 1986). Since this requires fourth moments of $\boldsymbol{\varepsilon}$, we propose a simpler and more robust method that employs logit mortality increments (LMI) data

$$Y^{\text{LMI}}_{xt} = \Delta\text{logit}\hat{Q}_{xt} = \text{logit}\hat{Q}_{xt} - \text{logit}\hat{Q}_{x,t-1} \tag{17}$$

rather than Y_{xt}. It admits a decomposition into explained and unexplained systematic variation and binomial risk, and a multiple linear regression model (28) analogous to (14) with response vector, design matrix and regression parameters $\boldsymbol{Y}^{\text{LMI}}$, $\boldsymbol{X}^{\text{LMI}}$ and $\boldsymbol{\theta}^{\text{LMI}}$. The covariance matrix of the unexplained part of LMI data can be decomposed into two terms

$$\boldsymbol{V}^{\text{LMI}} = \boldsymbol{V}_{\Delta\boldsymbol{\varepsilon}^s} + \boldsymbol{V}_{\Delta\boldsymbol{\varepsilon}^b}, \tag{18}$$

that are the covariance matrices of $\Delta\boldsymbol{\varepsilon}^s = \left(\Delta\varepsilon^s_{xt};\ (x,t) \in \Omega^{\text{LMI}}\right)^T$ and $\Delta\boldsymbol{\varepsilon}^b = (\Delta\varepsilon^b_{xt};\ (x,t) \in \Omega^{\text{LMI}})^T$, with $\Omega^{\text{LMI}} = \{(x,t);\ x_l \le x \le x_u,\ t_2 \le t \le t_T\}$.

An iterative weighted least squares (WLS) method is the used to estimate $\boldsymbol{\xi}$, by fitting an estimate of the covariance function of $\Delta\boldsymbol{\varepsilon}^s$ to data. The WLS approach is not optimal, since the weights do not take dependencies within $\Delta\boldsymbol{\varepsilon}^s$ into account. But when the random walk variance component σ_κ^2 dominates, the WLS approach is rather efficient for LMI transformed data, since $\Delta\boldsymbol{\varepsilon}^s$ has close to independent components for different time points.

We will write $\boldsymbol{V} = \boldsymbol{V}(\boldsymbol{\theta}, \boldsymbol{\xi})$ and $\boldsymbol{V}^{\mathrm{LMI}} = \boldsymbol{V}^{\mathrm{LMI}}(\boldsymbol{\theta}, \boldsymbol{\xi})$, since it follows from (6), (13) and (29) that both covariance matrices are functions of the regression parameters $\boldsymbol{\theta}$ for LM transformed data and the variance parameters $\boldsymbol{\xi}$.

With these preliminaries the procedure for estimating $E(Q_{xt})$ and Q_{xt} for historical data can be summarized by the following iterative scheme:

1. Put $i = 0$. Compute least squares estimates $\hat{\boldsymbol{\theta}}_0 = (\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ and $\hat{\boldsymbol{\theta}}_0^{\mathrm{LMI}}$ in the same way, replacing $\boldsymbol{X}$ and $\boldsymbol{Y}$ by $\boldsymbol{X}^{\mathrm{LMI}}$ and $\boldsymbol{Y}^{\mathrm{LMI}}$.
2. Compute an estimate $\hat{\boldsymbol{\xi}}_i$ of $\boldsymbol{\xi}$ from residuals $\boldsymbol{Y}^{\mathrm{LMI}} - \boldsymbol{X}^{\mathrm{LMI}}\hat{\boldsymbol{\theta}}_i^{\mathrm{LMI}}$ by WLS, as specified in the Appendix.
3. Update the covariance matrix estimates $\hat{\boldsymbol{V}}_i = \boldsymbol{V}(\hat{\boldsymbol{\theta}}_i, \hat{\boldsymbol{\xi}}_i)$ and $\hat{\boldsymbol{V}}_i^{\mathrm{LMI}} = \boldsymbol{V}^{\mathrm{LMI}}(\hat{\boldsymbol{\theta}}_i, \hat{\boldsymbol{\xi}}_i)$.
4. Compute $\hat{\boldsymbol{\theta}}_{i+1} = (\boldsymbol{X}^T\hat{\boldsymbol{V}}_i^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\hat{\boldsymbol{V}}_i^{-1}\boldsymbol{Y}$ and $\hat{\boldsymbol{\theta}}_{i+1}^{\mathrm{LMI}}$ in the same way, replacing $\boldsymbol{X}$, $\hat{\boldsymbol{V}}_i$ and $\boldsymbol{Y}$ by $\boldsymbol{X}^{\mathrm{LMI}}$, $\hat{\boldsymbol{V}}_i^{\mathrm{LMI}}$ and $\boldsymbol{Y}^{\mathrm{LMI}}$.
5. Let $i \to i + 1$.
6. If some parameter estimate has changed by more than tol (a predefined small tolerance number) in the last step, go to step 2.
7. Define the final estimate $\tilde{\boldsymbol{\xi}} = \hat{\boldsymbol{\xi}}_i$ of the unexplained systematic variance parameters, the FGLS estimator $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_i$ of the regression parameters for LM transformed data, and the final estimate $\tilde{\boldsymbol{V}} = \hat{\boldsymbol{V}}_i$ of the covariance matrix of LM transformed data.
8. Estimate the regression function $\boldsymbol{m} = (m_{xt}; (x, t) \in \Omega)^T$ for LM transformed data by $\tilde{\boldsymbol{m}} = \boldsymbol{X}\tilde{\boldsymbol{\theta}}$, and then the expected mortalities by
$$\tilde{Q}_{xt} = \frac{e^{\tilde{m}_{xt}}}{1 + e^{\tilde{m}_{xt}}}.$$
9. Predict the systematic effects $\boldsymbol{m} + \boldsymbol{\varepsilon}^s$ of LM transformed data by a BLUP $\check{\boldsymbol{m}} + \check{\boldsymbol{\varepsilon}}^s = \boldsymbol{X}\check{\boldsymbol{\theta}} + \check{\boldsymbol{\varepsilon}}^s$, where $\check{\boldsymbol{\theta}}$ and $\check{\boldsymbol{\varepsilon}}^s$ solve the system of equations
$$\begin{pmatrix} \boldsymbol{X}^T\tilde{\boldsymbol{V}}_{\boldsymbol{\varepsilon}^b}^{-1}\boldsymbol{X} & \boldsymbol{X}^T\tilde{\boldsymbol{V}}_{\boldsymbol{\varepsilon}^b}^{-1} \\ \tilde{\boldsymbol{V}}_{\boldsymbol{\varepsilon}^b}^{-1}\boldsymbol{X} & \tilde{\boldsymbol{V}}_{\boldsymbol{\varepsilon}^b}^{-1} + \tilde{\boldsymbol{V}}_{\boldsymbol{\varepsilon}^s}^{-1} \end{pmatrix} \begin{pmatrix} \check{\boldsymbol{\theta}} \\ \check{\boldsymbol{\varepsilon}}^s \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^T\tilde{\boldsymbol{V}}_{\boldsymbol{\varepsilon}^b}^{-1}\boldsymbol{Y} \\ \tilde{\boldsymbol{V}}_{\boldsymbol{\varepsilon}^b}^{-1}\boldsymbol{Y} \end{pmatrix}, \tag{19}$$
and $\tilde{\boldsymbol{V}} = \tilde{\boldsymbol{V}}_{\boldsymbol{\varepsilon}^s} + \tilde{\boldsymbol{V}}_{\boldsymbol{\varepsilon}^b}$ is the estimate of (15) from Step 7. Then the smoothed estimated mortalities are
$$\check{Q}_{xt} = \frac{e^{\check{m}_{xt} + \check{\varepsilon}_{xt}^s}}{1 + e^{\check{m}_{xt} + \check{\varepsilon}_{xt}^s}}.$$

4.2. Future mortality rates, life annuity premium and life expectancy

Several authors have suggested prediction methods of future mortality rates, see for instance Denton et al. (2004), Brouhns et al. (2005), Koissi et al. (2006), Li et al. (2009), Cairns et al. (2009), Cairns (2014) and references therein, and Cairns (2013) for applications to hedging. Here we first present a method for predicting the observed mortality rate $\hat{Q}_{xt}$ of a fixed future calendar year $t > t_T$ and age $x_l \le x \le x_u$, and then annuity premiums for a single individual. Our approach is purely analytical and does not require resampling.

The analysis simplifies considerably if the last random walk term of (6) is centered around $\tilde{t} = t_T$. This effectively means that we separate unexplained systematic variation of the past from the future, so that historical mortalities are independent of future mortalities. Together with (12), this implies that
$$E(Y_{xt}|\boldsymbol{Y}) = m_{xt}(\boldsymbol{\theta})$$
would be the predictor of the logit mortality rate $Y_{xt} = \mathrm{logit}\hat{Q}_{xt}$ if all parameters $\boldsymbol{\theta}$ were known, with a prediction error equal to a process error $\varepsilon_{xt} = Y_{xt} - m_{xt}(\boldsymbol{\theta})$ in (8) that has one unexplained systematic and one binomial risk component. Replacing $\boldsymbol{\theta}$ by its estimate $\tilde{\boldsymbol{\theta}}$, we get instead a predictor $m_{xt}(\tilde{\boldsymbol{\theta}})$, and a prediction error
$$Y_{xt} - m_{xt}(\tilde{\boldsymbol{\theta}}) = \varepsilon_{xt} - \left[m_{xt}(\tilde{\boldsymbol{\theta}}) - m_{xt}(\boldsymbol{\theta})\right], \tag{20}$$
that is the difference of the process error ε_{xt} and an estimation error $m_{xt}(\tilde{\boldsymbol{\theta}}) - m_{xt}(\boldsymbol{\theta})$.

In order to find the prediction error variance, we utilize that since historic and future mortalities are independent, the approximately Gaussian process error ε_{xt} is independent of $\tilde{\boldsymbol{\theta}}$, so that the two terms of the right hand side of (20) are independent. If we further assume that $\tilde{\boldsymbol{\theta}}$ is an estimator with an approximately Gaussian distribution, the estimation error will be approximately Gaussian as well, since (9) implies that $m_{xt}(\tilde{\boldsymbol{\theta}}) - m_{xt}(\boldsymbol{\theta}) = \dot{\boldsymbol{m}}_{xt}^T(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is a linear function of $\tilde{\boldsymbol{\theta}}$, with $\dot{\boldsymbol{m}}_{xt}$ a column vector containing the partial derivatives
$$\frac{\partial m_{xt}(\boldsymbol{\theta})}{\partial \theta_i} = \begin{cases} \phi_j(x), & \theta_i = a_j, \\ (t - t_T)\phi_j(x), & \theta_i = b_j, \end{cases}$$
for $i = 1, \ldots, p$. Putting things together, we get a Gaussian predictive distribution
$$Y_{xt} \sim N\left(m_{xt}(\tilde{\boldsymbol{\theta}}), \sigma_{xt}^2\right) \tag{21}$$
of Y_{xt}, with a prediction variance σ_{xt}^2 that is estimated as a sum
$$\tilde{\sigma}_{xt}^2 = \dot{\boldsymbol{m}}_{xt}^T\left(\boldsymbol{X}^T\tilde{\boldsymbol{V}}^{-1}\boldsymbol{X}\right)^{-1}\dot{\boldsymbol{m}}_{xt} + \left(N_{xt}q_{xt}(\tilde{\boldsymbol{\theta}})(1 - q_{xt}(\tilde{\boldsymbol{\theta}}))\right)^{-1} + \tilde{\sigma}_\eta^2 + c_x^2\tilde{\sigma}_\zeta^2 + d_x^2(t - t_T)\tilde{\sigma}_\kappa^2 \tag{22}$$
of estimates of the estimation error variance (using the fact that $\widehat{\mathrm{Var}}(\tilde{\boldsymbol{\theta}}) = (\boldsymbol{X}^T\tilde{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}$), the binomial risk variance and the unexplained systematic risk variance (using estimated variance components $\tilde{\boldsymbol{\xi}} = (\tilde{\sigma}_\eta^2, \tilde{\sigma}_\zeta^2, \tilde{\sigma}_\kappa^2)^T$).

We can easily convert the predictive distribution (21) from a logit to a probability scale. Let $F_{\hat{Q}_{xt}}(z)$ denote the predictive distribution function of the future observed mortality rate at (x, t). By inverting the logit transformation in (20), we obtain from (21) the predictive density function
$$f_{\hat{Q}_{xt}}(z) = F'_{\hat{Q}_{xt}}(z) = \frac{1}{\sqrt{2\pi}\sigma_{xt}z(1 - z)}\exp\left(-\frac{\left[\mathrm{logit}(z) - m_{xt}(\tilde{\boldsymbol{\theta}})\right]^2}{2\sigma_{xt}^2}\right)$$
and α-quantile
$$F_{\hat{Q}_{xt}}^{-1}(\alpha) = \frac{1}{1 + \exp\left[-(m_{xt}(\tilde{\boldsymbol{\theta}}) + \lambda_\alpha\sigma_{xt})\right]}$$
of $\hat{Q}_{xt}$ respectively, where λ_α is the α-quantile of a standard normal random variable.

Mortalities can be used to compute the n-year annuity premium $\hat{\mathcal{P}}_{xt}(n)$. This is the expected annuity during n years of an age x

individual, starting at the beginning of calendar year t. It can be written as a function

$$\begin{aligned}\hat{\mathcal{P}}_{xt}(n) &= P\cdot\sum_{k=0}^{n-1}\left[\prod_{j=0}^{k}(1-\hat{Q}_{x+j,t+j})\right](1+\mathrm{IR})^{-k-1}\\ &= P\cdot\sum_{k=0}^{n-1}\left[\prod_{j=0}^{k}\left(1+e^{Y_{x+j,t+j}}\right)^{-1}\right](1+\mathrm{IR})^{-k-1}\\ &=: f(\boldsymbol{Y}_{xt}),\end{aligned}\tag{23}$$

of the observed logit transformed mortalities $\boldsymbol{Y}_{xt} = (Y_{x+j,t+j})_{j=0}^{n-1}$ within cohort (x,t), with P the yearly annuity, $(1+\mathrm{IR})^{-1}$ the discount factor, corresponding to a yearly interest rate IR, see for instance the Appendix and Brouhns et al. (2002) for details.

It is possible to predict the n-year annuity premium by

$$\begin{aligned}\tilde{\mathcal{P}}_{xt}(n) &= P\cdot\sum_{k=0}^{n-1}\left[\prod_{j=0}^{k}(1-\tilde{Q}_{x+j,t+j})\right](1+\mathrm{IR})^{-k-1}\\ &= P\cdot\sum_{k=0}^{n-1}\left[\prod_{j=0}^{k}\left(1+e^{\tilde{m}_{x+j,t+j}}\right)^{-1}\right](1+\mathrm{IR})^{-k-1}\\ &=: f(\tilde{\boldsymbol{m}}_{xt}),\end{aligned}\tag{24}$$

where $\tilde{m}_{xt} = m_{xt}(\tilde{\boldsymbol{\theta}})$, and $\tilde{\boldsymbol{m}}_{xt} = (\tilde{m}_{x+j,t+j})_{j=0}^{n-1}$ represent estimated expected logit mortalities. It is harder to obtain a closed form expression for the predictive distribution of $\hat{\mathcal{P}}_{xt}(n)$, as it depends nonlinearly on $\boldsymbol{Y}_{xt}$. An approximate normal distribution can be found by means of the multivariate delta method, i.e. a first order Taylor expansion of the nonlinear function $f = f_n$. This gives

$$\hat{\mathcal{P}}_{xt}(n) \sim N(\tilde{\mathcal{P}}_{xt}(n), \sigma^2_{\mathcal{P},xt}),$$

with a predictive variance that can be estimated as

$$\tilde{\sigma}^2_{\mathcal{P},xt} = \sum_{i,j=0}^{n-1}\tilde{\sigma}_{x+i,t+i;x+j,t+j}\frac{\partial f}{\partial\tilde{m}_{x+i,t+i}}\frac{\partial f}{\partial\tilde{m}_{x+j,t+j}},\tag{25}$$

where $\tilde{\sigma}_{xt,x't'}$ is a covariance term analogous to (22). Explicit expressions for $\tilde{\sigma}_{xt,x't'}$ and the partial derivatives $\partial f/\partial\tilde{m}_{x+j,t+j}$ are given in the Appendix.

A notion closely related to (23) is the life expectancy $\hat{\mathring{e}}_{xt}$ of an age x individual at the beginning of calendar year t. It corresponds to putting $P = 1$, $\mathrm{IR} = 0$ and $n = x_{\max} - x$, where $x_{\max}$ is the maximal integer valued age, for instance 110 years. In addition, each term $(1-\hat{Q}_{x+j,t+j})$ in (23) with $j = k$ is replaced by $(1-0.5\hat{Q}_{x+j,t+j})$ when $j = k < n-1$, and by a slightly more complicated expression when $j = k = n-1$. This is based on the assumption that individuals who die during calendar year $t \le t' \le t+n$, do so uniformly over $[t', t'+1]$, see the Appendix for details. Prediction intervals for life expectancies can be derived as above, taking the extra 0.5 factors into account when computing the partial derivatives of f.

5. Data analysis

We will use data sets for Swedish and United States populations in our analysis. Rather than finding a multipopulation model that fits both data sets (Li and Lee, 2005; Cairns et al., 2011), we build a single model separately for each country. A danger with using a single data set is that it contains something specific that one takes to be general. Together, the two countries have a broad range of population sizes and both are popular in the literature, for their economic importance and size (USA) or admittedly good data quality (Sweden). Therefore they constitute a fairly broad range of Western populations. We use data from 1980 to 2011 for Swedish data and from 1980 to 2007 for US data respectively (latest available at time of download), with males and females handled separately. The data comes from the Human Mortality database, see mortality.org for further documentation.

Table 1
Estimates $\tilde{\boldsymbol{\xi}} = (\tilde{\sigma}^2_\eta, \tilde{\sigma}^2_\zeta, \tilde{\sigma}^2_\kappa)$ of the three variance components of unexplained systematic variation for logit mortalities, for US and Swedish men of age 60–90.

Population	$\tilde{\sigma}^2_\eta$	$\tilde{\sigma}^2_\zeta$	$\tilde{\sigma}^2_\kappa$
US m 60–90	0.000127	0.000033	0.000130
SWE m 60–90	0.000300	0.000210	0.000000

5.1. Residual plots

Even with the above derivation of the FGLS estimate and the BLUP, it is instructive to first study the residual plots of

$$\hat{\varepsilon}^{\mathrm{LMI}}_{0,xt} = Y^{\mathrm{LMI}}_{xt} - \hat{m}^{\mathrm{LMI}}_{0,xt},$$

for LMI transformed data, with $\hat{m}^{\mathrm{LMI}}_{0,xt} = m^{\mathrm{LMI}}_{xt}(\hat{\boldsymbol{\theta}}^{\mathrm{LMI}}_0)$ defined in (27) and $\hat{\boldsymbol{\theta}}^{\mathrm{LMI}}_0$ the ordinary least squares estimate in Step 1 of the algorithm in Section 4.1. In Fig. 1 we have plotted these residuals for two age-bands of males from the US population.

5.1.1. Calendar year effects

Calendar effects can be seen as vertical lines in the residual plots. They can be spotted mostly in higher ages, above 60, and a probable cause are such effects as the seasonal influenza, heat waves and cold spells that are known to vary in severity from year to year.

There is a notable exception from the old age only effects, a steep drop for US males in their 30s in 1996–97, the same years as the modern HIV inhibitor medicines reached the markets.

5.2. Estimated and predicted mortality rates

In order to simplify estimation of variance parameters, we will not analyze all age classes as one data set, but rather treat a number of age-bands separately. If this band contains a sufficiently small range of age classes, the noise terms ζ_t and κ_t of the unexplained systematic variation (6) will affect each age by a similar amount. It is then reasonable to assume that $\boldsymbol{c}$ and $\boldsymbol{d}$ have fixed and constant elements $c_x = d_x = 1$. In Section 6 we briefly discuss how to model the whole age-span, using separate and partly overlapping variance components for different age-bands.

Due to space, here we restrict the analysis to US and Swedish males in one age band, 60–90, for which we perform the procedures described in Sections 4.1 and 4.2.

We define our data by setting, for the US

$$\Omega_{US} = \{(x,t);\ 60 \le x \le 90,\ 1980 \le t \le 2007\}$$

and for Sweden

$$\Omega_{SWE} = \{(x,t);\ 60 \le x \le 90,\ 1980 \le t \le 2011\}.$$

Estimates of $\tilde{\boldsymbol{\xi}}$ are found in Table 1. For the Swedish data set no random walk component was detected ($\tilde{\sigma}^2_\kappa = 0$) for the unexplained systematic variation $\{\boldsymbol{\varepsilon}^s_t\}$.

In order to get an overview of the material, the Swedish FGLS estimates of the general mortality development are presented in Fig. 2; both on the logit scale ($\mathrm{logit}\tilde{Q}_{xt}$) and the nominal scale ($\tilde{Q}_{xt}$). It can be seen that the improvements on the logit scale are largest at ages close to 60, but around 80 on the absolute scale.

A closer look at the inference procedure for historical and future data can be found in Figs. 3–4. For ages $x = 60$, 70, 80 and 90

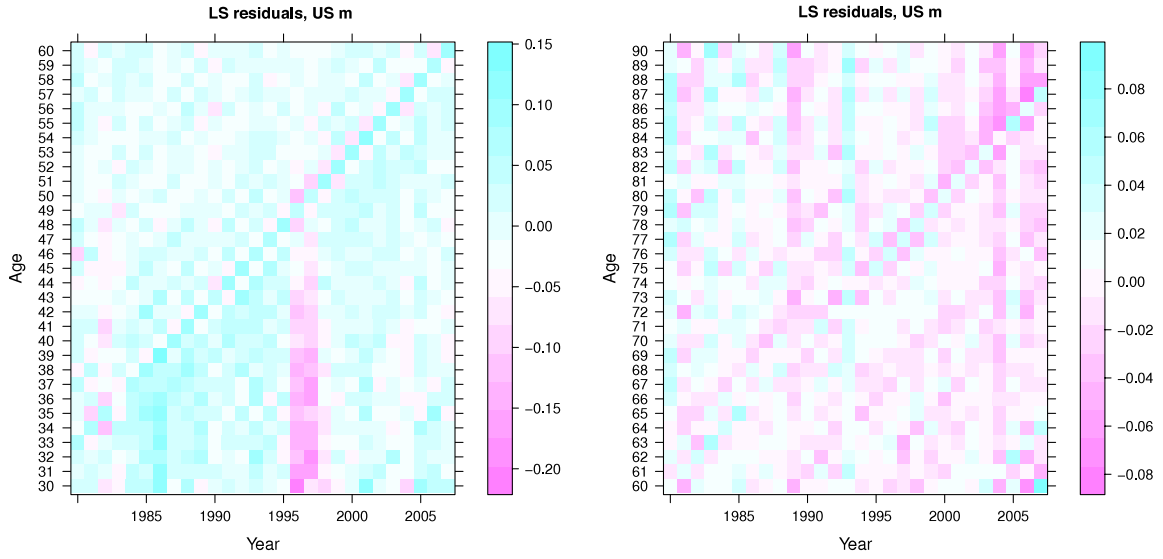

Fig. 1. Residuals of a least squares fit to one year increments of estimated logit mortality rates for US males of ages 30–60 (left) and 60–90 (right).

Fig. 2. Plots, for Swedish data, of $\text{logit}\tilde{Q}_{xt}$ and $\tilde{Q}_{xt}$. The points $(x, t) \in \Omega_{SWE}$ are ordered linearly along the horizontal axis, where the first set of points are for age 60, years 1980–2011, then the rest of the ages 61, . . . , 90 line up from left to right.

we plot observed mortalities $\hat{Q}_{xt}$, FGLS-estimates $\tilde{Q}_{xt}$ and BLUPs $\check{Q}_{xt}$, for $t = 1980, \ldots, 2007/2011$. We also give predictions of $\hat{Q}_{xt}$ for $t = 2008/2012, \ldots, 2036$, and 95% prediction intervals. Since the FGLS estimate fits an average linear trend during the whole time period, the prediction intervals could miss trend changes. For instance, for the US data of Fig. 3 there is a decreased mortality rate during 2002–2007 for $x = 90$ that is not accounted for in the prediction intervals. Recall from Section 4.2 that the random walk component is centered around $\tilde{t} = t_T = 2007$. This implies that the difference between the FGLS estimates and the BLUPs the last few years is due white noise η_{xt} and period effects ζ_t. Since these are both independent over years, neither of them will influence the prediction intervals. In order to have the US prediction intervals for $x = 90$ to start at the last observed value one can either allow for trend changes in the deterministic trend (see Section 6) or force the estimates $\tilde{\sigma}_\eta^2$ and $\tilde{\sigma}_\zeta^2$ of the variances of η_{xt} and ζ_t to be small.

To analyze how well the estimated expected (FGLS) and predicted systematic (BLUP) mortalities fit data, we first introduce their respective logit mortality residuals,

$$\begin{aligned} \tilde{\varepsilon}_{xt} &= Y_{xt} - \tilde{m}_{xt}, \\ \check{\varepsilon}_{xt} &= Y_{xt} - (\check{m}_{xt} + \check{\varepsilon}^s_{xt}), \end{aligned} \tag{26}$$

and then their standardized counterparts

$$\tilde{\varepsilon}^{\text{st}}_{xt} = \tilde{\varepsilon}_{xt} / \sqrt{\widehat{\text{Var}}(\varepsilon^b_{xt})},$$

Table 2
The estimated standard deviation of the standardized FGLS and BLUP residuals $\tilde{\varepsilon}^{\text{st}}_{xt}$ and $\check{\varepsilon}^{\text{st}}_{xt}$, for US and Swedish males of ages 60–90.

Population	$\sqrt{\widehat{\text{Var}}(\tilde{\varepsilon}^{\text{st}}_{xt})}$	$\sqrt{\widehat{\text{Var}}(\check{\varepsilon}^{\text{st}}_{xt})}$
US m 60–90	4.46	0.73
SWE m 60–90	1.44	0.89

$$\check{\varepsilon}^{\text{st}}_{xt} = \check{\varepsilon}_{xt} / \sqrt{\widehat{\text{Var}}(\varepsilon^b_{xt})},$$

where

$$\widehat{\text{Var}}(\varepsilon^b_{xt}) = \frac{1}{N_{xt}\tilde{Q}_{xt}(1 - \tilde{Q}_{xt})}$$

is an estimate, on a logit scale, of the binomial variance (13). If the deterministic part of our model explains all systematic variation, then the standardized FGLS residuals should have a distribution close to a standard normal $N(0, 1)$. The distribution of the standardized BLUP residuals, on the other hand, should be close to a standard normal whenever the mortality rates q_{xt} are accurately predicted.

The values are presented in Table 2. The FGLS estimate leaves some residual variance to explain, since the variance of their residuals is larger by a factor 4.46 (1.44) for US (Swedish) data compared to a model without unexplained systematic variation.

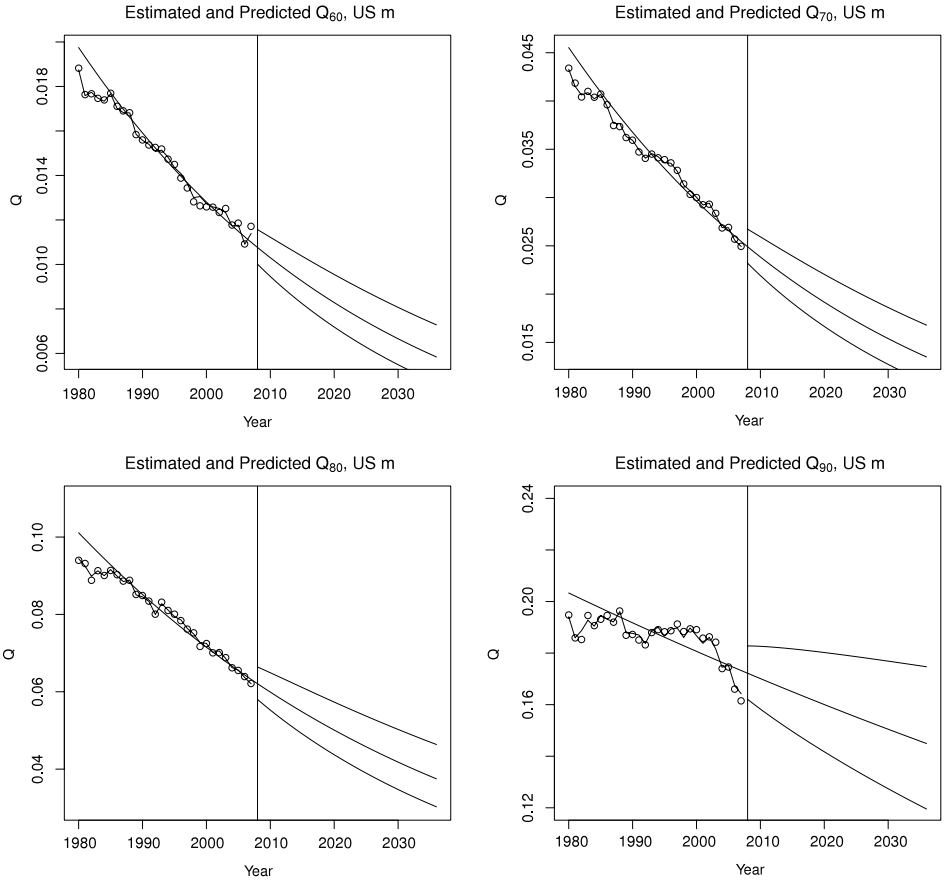

Fig. 3. For US males of ages $x = 60$, 70 80 and 90, plots of observed mortality rates $\hat{Q}_{xt}$, FGLS estimates $\tilde{Q}_{xt}$ and BLUPs $\check{Q}_{xt}$ are shown for $t = 1980, \ldots, 2007$, as well as predictions of future observed mortality rates $\hat{Q}_{xt}$ for $t = 2008, \ldots, 2036$, together with a 95% confidence band.

The BLUP, on the other hand, overfits a bit, explaining all systematic variation, but also part of the binomial variation. To get a graphical view, we plot the residuals in Fig. 5 and order them in a QQ-plot in Fig. 6. Some of the QQ-plots have slopes that differ significantly from 1, in accordance with Table 2. As an unordered collection the residuals seem normal, but the QQ-plots of the FGLS residuals reveal a lighter tail than normal, especially for US data. This is possibly due to the unexplained systematic effects, whose random walk component introduces dependencies between residuals.

We now turn to prediction of future observed mortality rates $\hat{Q}_{xt}$. It is interesting to analyze how the different sources of variation in (22) add to the total prediction variance $\tilde{\sigma}^2_{xt}$. In Figs. 7–8, we plot $\tilde{\sigma}^2_{60,t}$ as well as its three components of estimation error variance, binomial variance and unexplained systematic variance.

For the US population, the estimation error dominates, growing quadratically in t. The unexplained systematic variance increases linearly in t due to the random walk component. The binomial variance is just about 5% of the total variance. To illustrate the impact, we show in Fig. 9 the result of a model fit were the random walk component is excluded from the model. The resulting 95% confidence bands seem to be too narrow.

For Sweden it is the other way around. The binomial variance dominates the other two terms, contributing about 75% of the total prediction variance. The general picture looks the same for higher ages, but due to higher mortality, the binomial variance is lower (up to an age over 95 were it starts to increase again due to the smaller sizes of these cohorts), with a minimal relative variance contribution of about 60% at the age of 90.

Fig. 10 shows (predictions of) the 20-year annuity premiums $\hat{P}_{65,t}(20)$ in (23) for US males of age $x = 65$ at the beginning of 33 consecutive years $t = 1980, \ldots, 2012$. They reflect knowledge of mortalities up to and including 2007, so that the annuities of the first years are known, whereas more recent annuities involve more uncertainties. This can be seen from the 95% prediction intervals that gradually widen as the number of years with unknown mortalities increases.

6. Discussion

We have analyzed logit transformed observed mortality rates by means of a multiple linear mixed model, with three components of variation; binomial risk due to a finite population, systematic risk explained by age and calendar year and systematic risk not explained by these two covariates. In the accompanying paper Ekheden and Hössjer (2014), we applied a variance decomposition of Hössjer (2008) and Hössjer et al. (2009) in order to estimate the explained systematic risk, the unexplained systematic risk and the binomial risk from data. In this way, a test of overdispersion could be defined, which was used to decide whether a simple logistic regression model, without unexplained systematic risk, is sufficient for inference or not. In this paper we include unexplained systematic risk as a multivariate Gaussian time series. We propose an iterative estimation algorithm for historical data, where explained systematic risk is estimated by FGLS, and the total systematic risk by a BLUP. We also obtain closed form expressions for the prediction variance and the prediction interval of life expectancies and future mortalities.

The unexplained systematic risk contains three variance components corresponding to white noise, period effects and a two-sided random walk, which quantifies how much confidence we have in the deterministic long term trend. It is possible to let the strength of period and random walk effects vary with

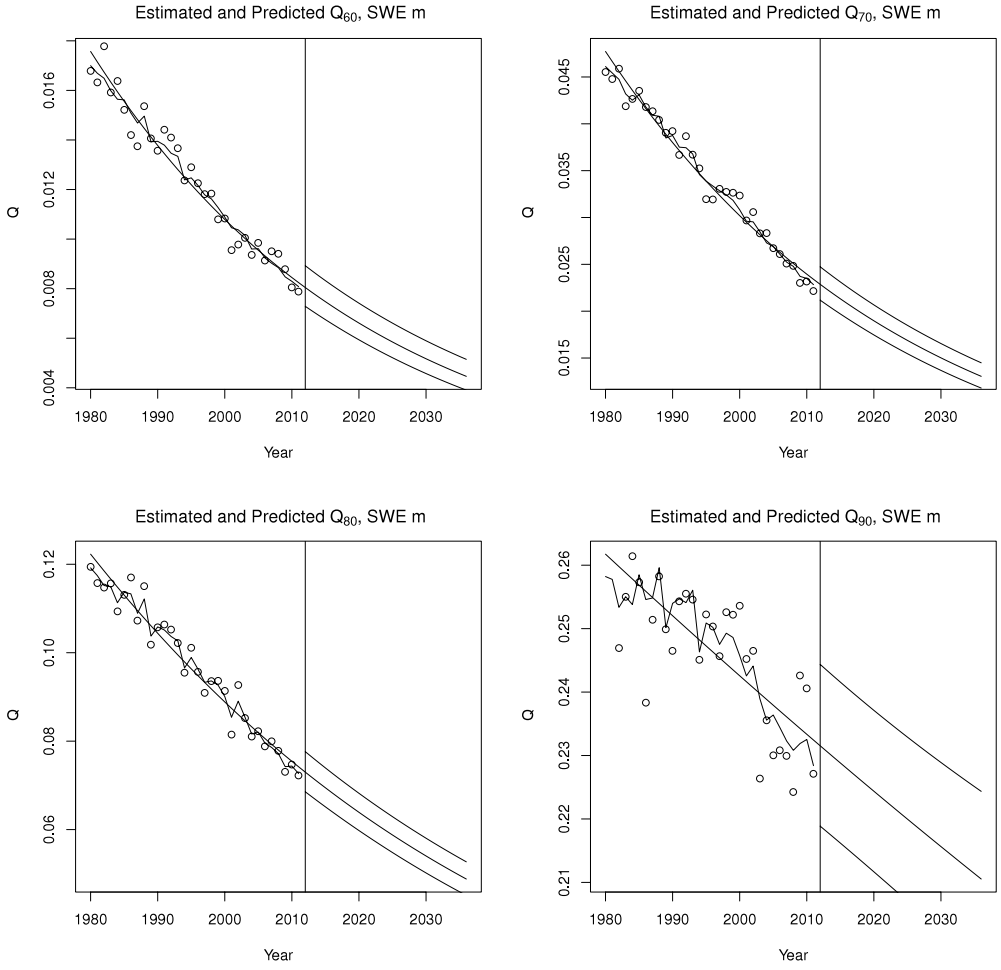

Fig. 4. For Swedish males of ages $x = 60$, 70 80 and 90, plots of observed mortality rates $\hat{Q}_{xt}$, FGLS estimates $\tilde{Q}_{xt}$ and BLUPs $\check{Q}_{xt}$ are shown for $t = 1980, \ldots, 2011$, as well as predictions of future observed mortality rates $\hat{Q}_{xt}$ for $t = 2012, \ldots, 2036$, together with a 95% confidence band.

age by choosing non-uniform weight vectors $\boldsymbol{c}$ and $\boldsymbol{d}$ or to make separate analyses for different age bands. Even so, the mortality development can be quite different in different age bands in a way that is difficult predict. When a new deadly disease enters a population we can have this effect, like HIV that mostly strikes a certain age band. A possible extension is then to introduce more period and random walk components in (6) and weight them together agewise with B-splines for instance. In more detail, define

$$\boldsymbol{\varepsilon}_t^s = \boldsymbol{\eta}_t + \sum_{j=1}^{n_c} \boldsymbol{c}_j \zeta_{jt} + \sum_{j=1}^{n_d} \boldsymbol{d}_j \sum_{s=1}^{|t-\tilde{t}|} \kappa_{j,\tilde{t}+\mathrm{sgn}(t-\tilde{t})s},$$

where $\boldsymbol{\eta}_t = (\eta_{xt};\ x = x_u, \ldots, x_l)^T$ has components $\eta_{xt} \sim N(0, \sigma_\eta^2)$ as before, $\zeta_{jt} \sim N(0, \sigma_{j\zeta}^2)$ and $\kappa_{jt} \sim N(0, \sigma_{j\kappa}^2)$, so that the parameter vector $\boldsymbol{\xi}$ of $\{\boldsymbol{\varepsilon}_t^s\}$ contains $1 + n_c + n_d$ variance components that can be estimated by means of a straightforward extension of the algorithm in the Appendix. If for instance linear splines are used with m knot points as in (5), both for the period and random walk components, then $n_c = n_d = m$, $\boldsymbol{c}_j = \left(\sqrt{\phi_j(x)};\ x = x_l, \ldots, x_u\right)^T$ and $\boldsymbol{d}_j = \left(\sqrt{\phi_j(x)};\ x = x_l, \ldots, x_u\right)^T$ will give a piecewise linear age dependency of the variance of $\boldsymbol{\varepsilon}_t^s$.

The BLUP can be viewed as smoother of historical observed mortality rates, the purpose of which is to remove or suppress the binomial risk part. It automatically adjusts the amount of smoothing to be large for a small country with a high binomial risk, and low for a large country with a small binomial risk. Other smoothing methods have also been suggested in the life insurance or epidemic literature, including two-dimensional penalized B-splines (Currie et al., 2004), penalized B-splines for parameters of the Lee–Carter model and its extensions (Delwarde et al., 2007b; Currie, 2013), generalized additive models and penalized likelihoods (Hall and Friel, 2010) or generalized linear array models and extended quasi likelihoods (Eilers et al., 2008). Guerrero and Silva (2010) use time series methods for the log mortalities, assuming that departures from a linear (or higher order polynomial) trend are fitted by means of penalized least squares, the solution of which has a Kalman filter and a generalized least squares interpretation.

The prediction error can be split into a process and an estimation error. For the latter, it is important to have accurate estimates of the mortality rates Q_{xt_T} at the last observed calendar year t_T, as well as of their linear slopes, on a logit scale. The reason is that future predicted mortalities will be offset by the estimation error at time t_T, and for this reason we chose $\tilde{t} = t_T$ in Section 4.2. This is particularly important for smaller populations with a relatively high binomial risk were the estimation error might be considerable.

We have shown here that a least squares analysis of the observed logit mortalities is enough for a good fit of Swedish data, and it is practically identical to the maximum likelihood fit from a simple logistic regression model (results not shown). Still we performed the full analysis on ages 60–90. No random walk component was found, but white noise and period effects were found for the unexplained systematic mortality. The main advantage is that these add some width to the prediction

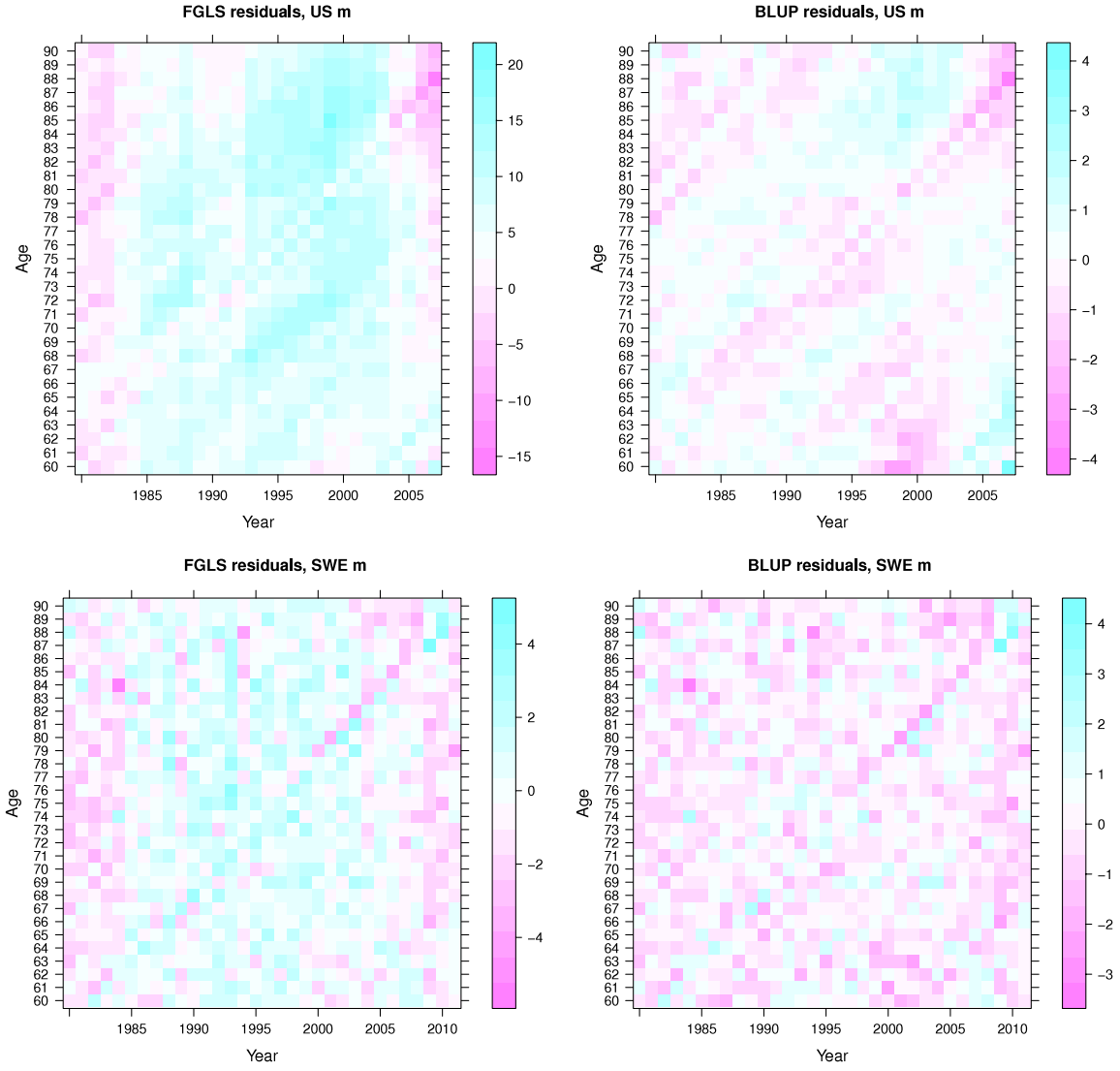

Fig. 5. Standardized FGLS and BLUP residuals $\tilde{\varepsilon}_{xt}^{\mathrm{st}}$ and $\check{\varepsilon}_{xt}^{\mathrm{st}}$ for US males of ages 60–84 and Swedish males of ages 60–84.

intervals when future observed mortalities are forecasted, showing that a logistic regression analysis probably underestimates the uncertainty if it is used for prediction purposes. Even so, the binomial risk dominates the prediction variance.

Interestingly enough, we got quite a different picture when analyzing the US data, age 60–90. Here a random walk component was found. For prediction, it does add to the unexplained systematic risk variance, growing linearly in time. But perhaps a bit unexpected, it adds even more significantly to the estimation error variance, which grows quadratically with time. In total we get confidence bands as wide as or wider than for Sweden, even tough the binomial risk is negligible for the US. This suggests that the prediction intervals for Swedish data are still too narrow, since the size of the random walk component was estimated to zero and this estimate is highly unreliable, because of the high binomial risk. It may therefore be wise to perform a sensitivity analysis, see Fig. 8, and compute prediction intervals for a range of values of the parameter describing the size of the random walk variance component. Another possibility is to use a Bayesian approach, with a prior distribution for all variance parameters. Pedroza (2006) has shown, in the context of the Lee–Carter model, that Bayesian prediction intervals are wider, since more sources of variation are taken into account.

The large influence of the random walk component on the estimation error can also be understood in terms of the FGLS estimate, which does not fit historical data as well as the ordinary least squares solution. On the other hand, if we exclude the random walk component from our model, we get much narrower confidence bands and the FGLS almost coincides with the least squares estimate, effectively meaning that we trust the presently observed rate of decrease in mortality.

The addition of a random walk component in a regression model will have a large impact on the conclusions drawn regarding prediction. In some sense it is to say that we do not trust the long term trend of the regression model, neither for historical (as manifested by a large estimation error) or future (as manifested by a large unexplained systematic risk variance) data. Indeed, how plausible is a model that predicts in essence that a linear trend will go on and on? One can think of factors or events that will change the trend in the future, for example the more and more widespread metabolic syndrome, or if (when) antibiotic resistant bacterias really start to spread and turn today harmless infections back into the deadly diseases they were before the antibiotics were discovered. Looking at historic logit mortality data for Swedish males during 1900–2011, the development for most ages can be described as "piecewise linear", with just two or three knots of which the last one (common to most ages) took place in the early 1980s when the improvements accelerated up to its current speed.

It is possible to extend our model in order to incorporate such change points in at least two ways. Either random structural breaks

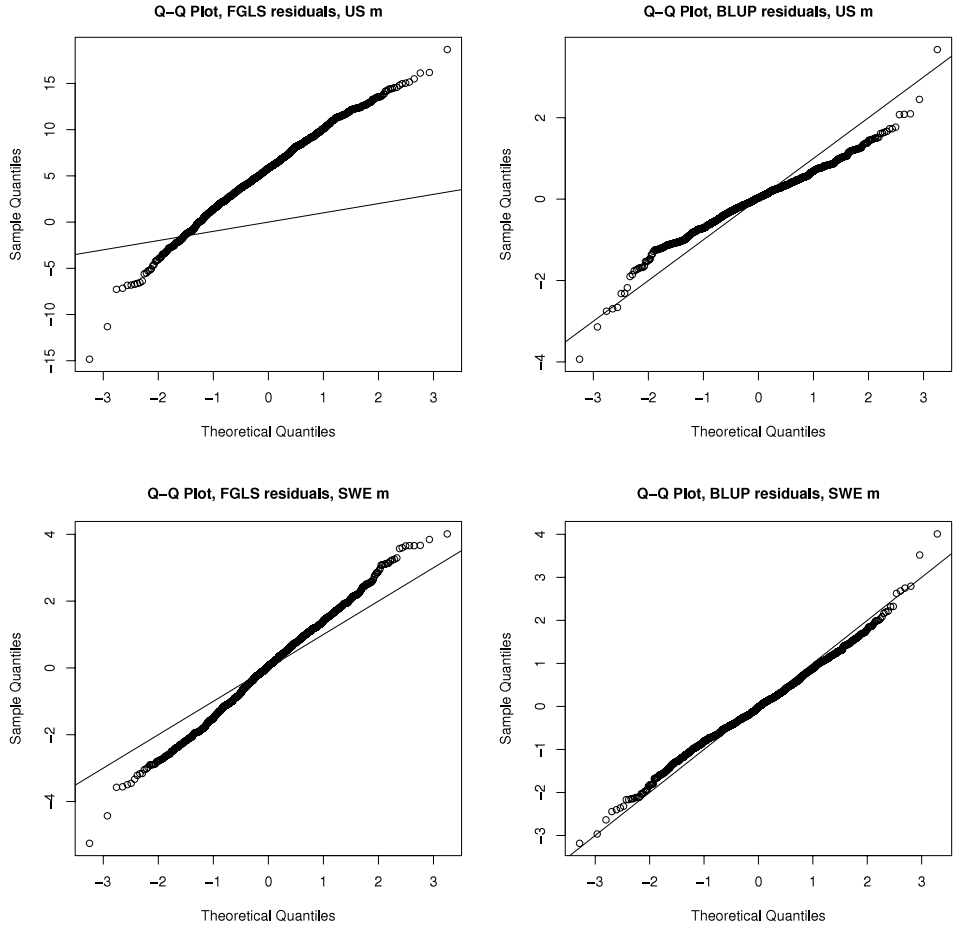

Fig. 6. Quantile–quantile plots of the standardized FGLS and BLUP residuals $\tilde{\varepsilon}^{\mathrm{st}}_{xt}$ and $\check{\varepsilon}^{\mathrm{st}}_{xt}$, for US and Swedish males of ages 60 to 90. A line with slope 1 is added for comparison. Notice the large slope of the upper left plot, which strongly indicates a significant amount of unexplained systematic variation.

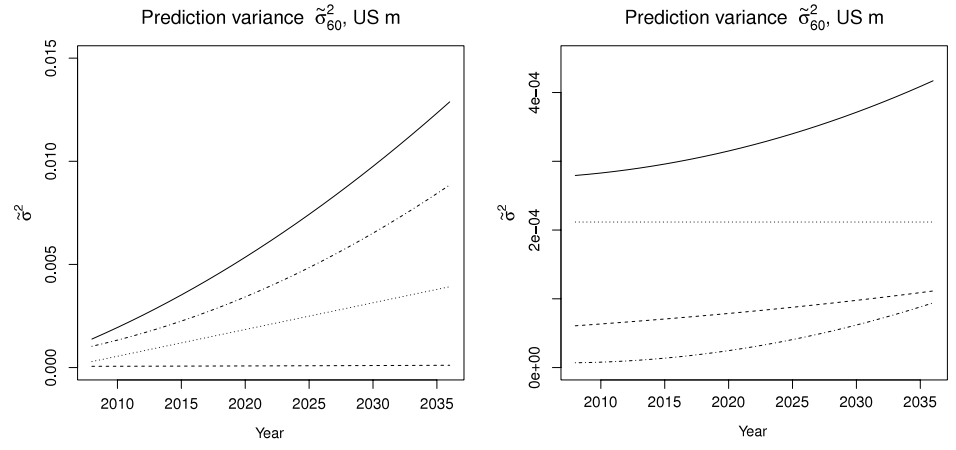

Fig. 7. Left: The solid curve is total prediction variance $\tilde{\sigma}^2_{60,t}$ of the future observed mortality rate for US males of age 60. The three dotted curves represent, from top to bottom, the components of estimation error variance, unexplained systematic variance and binomial variance. Right: Same plot but with the random walk component removed from the model ($\tilde{\sigma}^2_\kappa = 0$). Notice the different scales along the vertical axes.

are built into the drift process $\{\kappa_t\}$ of the unexplained systematic noise in (6). This increases the widths of the prediction intervals, but not very much their locations. Another option is to have a piecewise linear deterministic time trend, and use some model selection criterion, such as F-tests or BIC, in order to estimate the optimal number of change points. This will change the locations of the prediction intervals, but not very much their widths. Such extensions have recently been proposed for the Lee–Carter and Cairns–Blake–Dowd models, see Coelho and Nunes (2011), Li et al. (2011), Sweeting (2011) and van Berkum et al. (2013).

On the other hand, White (2002) argues that a simple linear time trend (2) often performs better than expert opinions. Similar conclusions were drawn in Section 9 of Denton et al. (2004) and in Section 3 of Booth and Tickle (2008). This defies the persisting idea that "improvements cannot continue at this fast rate", and only time will tell if it really is so. If one uses a linear trend, a possible option is to start finding a period during which mortality

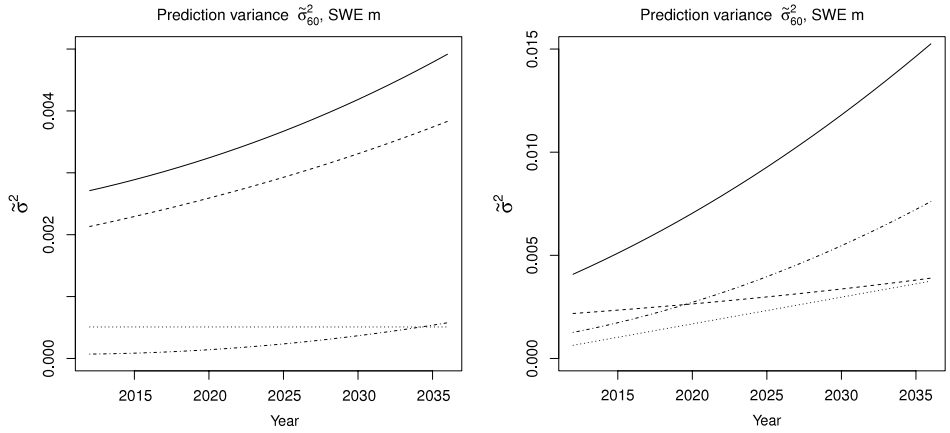

Fig. 8. Left: The solid curve is total prediction variance $\tilde{\sigma}^2_{60,t}$ of the future observed mortality rate for Swedish males of age 60. The three dotted curves represent, from top to bottom, the components of binomial variance, unexplained systematic variance (constant) and estimation error variance. Right: The same plot but when the random walk component is assumed to equal the estimate for the US data. Notice the different scales along the vertical axes.

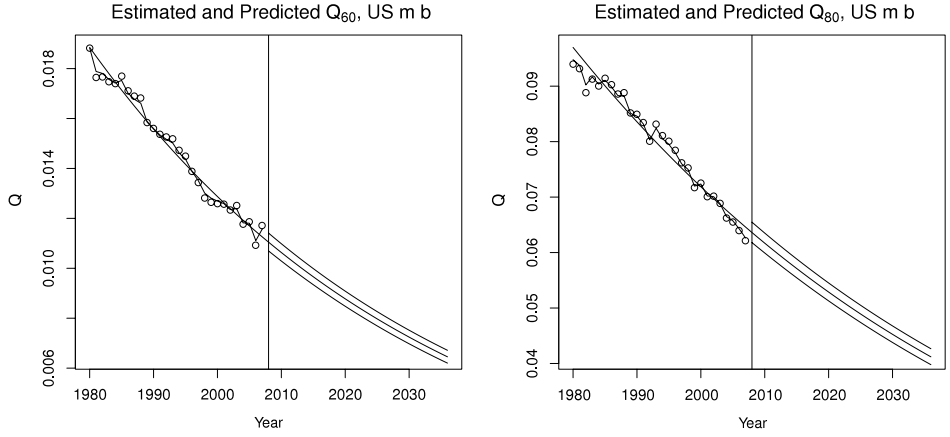

Fig. 9. Adjusted model, without random walk component: US males of ages $x = 60$ and 80, plots of observed mortality rates $\hat{Q}_{xt}$, FGLS estimates $\tilde{Q}_{xt}$ and BLUPs $\check{Q}_{xt}$ are shown for $t = 1980, \ldots, 2007$, as well as predictions of future observed mortality rates $\hat{Q}_{xt}$ for $t = 2008, \ldots, 2036$, together with a 95% confidence band that is much narrower than in Fig. 3.

improvement has not changed much. The R package demography (Hyndman et al., 2014), for instance, has an inbuilt option in its life expectancy function for varying the input period of years. One may also use one of the above mentioned change point detection algorithms in order to find the latest trend break. Then data is analyzed from this time point with a model that has the same linear trend for historic and future periods of time.

Acknowledgments

Ola Hössjer's research was financially supported by the Swedish Research Council, grant nr. 621-2008-4946, and the Gustafsson Foundation for Research in Natural Sciences and Medicine. We also wish to thank one anonymous referee for very useful comments and suggested references, which considerably improved the paper.

Appendix

Mixed regression model for logit mortality increments data. We will define a mixed linear regression model for logit mortality increments data Y^{LMI}_{xt} in (17), for all (x, t) in Ω^{LMI}. It follows from (8)–(11) that

$$\begin{aligned} Y^{\text{LMI}}_{xt} &= m^{\text{LMI}}_{xt}(\boldsymbol{\theta}^{\text{LMI}}) + \varepsilon^{\text{LMI}}_{xt} \\ &= m^{\text{LMI}}_{xt}(\boldsymbol{\theta}^{\text{LMI}}) + \varepsilon^{\text{LMI},s}_{xt} + \varepsilon^{\text{LMI},b}_{xt}, \end{aligned} \tag{27}$$

with mean function, regression parameters, systematic unexplained and binomial error terms given by

$$\begin{aligned} m^{\text{LMI}}_{xt}(\boldsymbol{\theta}^{\text{LMI}}) &= \sum_{j=0}^{p_2} b_j \phi_j(x), \\ \boldsymbol{\theta}^{\text{LMI}} &= (b_0, \ldots, b_{p_2})^T, \\ \varepsilon^{\text{LMI},s}_{xt} &= \varepsilon^{s}_{xt} - \varepsilon^{s}_{x,t-1}, \\ \varepsilon^{\text{LMI},b}_{xt} &= \left(\text{logit}\hat{Q}_{xt} - \text{logit}Q_{xt}\right) - \left(\text{logit}\hat{Q}_{x,t-1} - \text{logit}Q_{x,t-1}\right). \end{aligned}$$

We can write this as a multiple linear regression model

$$\boldsymbol{Y}^{\text{LMI}} = \boldsymbol{X}^{\text{LMI}}\boldsymbol{\theta}^{\text{LMI}} + \boldsymbol{\varepsilon}^{\text{LMI}}, \tag{28}$$

where $\boldsymbol{Y}^{\text{LMI}} = \left(Y^{\text{LMI}}_{xt};\ (x,t) \in \Omega^{\text{LMI}}\right)^T$ and $\boldsymbol{\varepsilon}^{\text{LMI}} = \left(\varepsilon^{\text{LMI}}_{xt};\ (x,t) \in \Omega^{\text{LMI}}\right)^T$ are column vectors of length $n = (T-1)(x_u - x_l + 1)$, and $\boldsymbol{X}^{\text{LMI}}$ a design matrix of dimension $n \times (p_2 + 1)$, with row $(\phi_0(x), \phi_1(x), \ldots, \phi_{p_2}(x))$ corresponding to (x, t).

It follows by a similar argument as in (12) and (13), that the binomial risk variance function satisfies

$$\begin{aligned} \text{Var}(\varepsilon^{\text{LMI},b}_{xt}) &= E\left(\text{Var}(\text{logit}\hat{Q}_{x,t-1}|Q_{x,t-1})\right) + E\left(\text{Var}(\text{logit}\hat{Q}_{xt}|Q_{xt})\right) \\ &\approx E\left(\frac{1}{N_{x,t-1}Q_{x,t-1}(1-Q_{x,t-1})}\right) \\ &\quad + E\left(\frac{1}{N_{xt}Q_{xt}(1-Q_{xt})}\right). \quad \square \end{aligned} \tag{29}$$

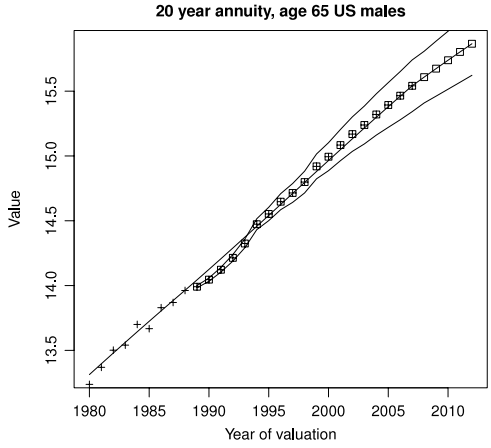

Fig. 10. The graph reflects remaining uncertainties of 20 year annuity premiums for US males in the beginning of year 2008, with mortality data available up to and including year $t_T = 2007$. The crosses are the known annuity premiums $\hat{P}_{65,t}(20)$ in (23) for males aged $x = 65$ at the beginning of years $t = 1980, \ldots, 1988$. The squares are the predicted annuity premiums $\hat{P}_{65,t}(20)$ in (24) for $t = 2008, \ldots, 2012$. The crossed squares are hybrid annuity premiums for the years $t = 1989, \ldots, 2007$ in between. They combine observed mortalities $\hat{Q}_{x+j,t+j}$ the first $r = \min(2008 - t, 20)$ years ($j = 0, \ldots, r-1$) with predicted mortalities $\tilde{Q}_{x+j,t+j}$ the last $20 - r$ years ($j = r, \ldots, 20-1$). The 95% prediction intervals assume a normal predictive distribution, with a predictive variance computed for the last $n = 20 - r$ years, as in (25). The solid line, finally, is a smoothed annuity curve based on FGLS-estimated parameters.

Formulas for $\boldsymbol{V}$ and $\boldsymbol{V}^{\mathrm{LMI}}$. In view of (15) and (18), we need formulas for $\boldsymbol{V}_{\boldsymbol{\varepsilon}^b} = \boldsymbol{V}_{\boldsymbol{\varepsilon}^b}(\boldsymbol{\theta})$, $\boldsymbol{V}_{\Delta\boldsymbol{\varepsilon}^b} = \boldsymbol{V}_{\Delta\boldsymbol{\varepsilon}^b}(\boldsymbol{\theta})$, $\boldsymbol{V}_{\boldsymbol{\varepsilon}^s} = \boldsymbol{V}_{\boldsymbol{\varepsilon}^s}(\boldsymbol{\xi})$ and $\boldsymbol{V}_{\boldsymbol{\varepsilon}^b} = \boldsymbol{V}_{\boldsymbol{\varepsilon}^b}(\boldsymbol{\xi})$.

Starting with the former two matrices, we will first simplify the variance formulas in (13) and (29). It follows from (2), (8) and (10), that

$$Q_{xt} = \frac{e^{m_{xt}+\varepsilon^s_{xt}}}{1+e^{m_{xt}+\varepsilon^s_{xt}}} \Longrightarrow \frac{1}{Q_{xt}(1-Q_{xt})} = \frac{\left(1+e^{m_{xt}+\varepsilon^s_{xt}}\right)^2}{e^{m_{xt}+\varepsilon^s_{xt}}}$$

with $m_{xt} = m_{xt}(\boldsymbol{\theta})$. If the variance components of ε^s_{xt} in (6) are small enough, the approximation

$$E\left(\frac{1}{Q_{xt}(1-Q_{xt})}\right) \approx \frac{(1+e^{m_{xt}})^2}{e^{m_{xt}}}$$

is justified, and we can write the elements of $\boldsymbol{V}_{\boldsymbol{\varepsilon}^b}$ and $\boldsymbol{V}_{\Delta\boldsymbol{\varepsilon}^b}$ as

$$V_{\boldsymbol{\varepsilon}^b,(x,t),(x',t')} = 1_{\{(x,t)=(x',t')\}} \frac{(1+e^{m_{xt}})^2}{N_{xt}e^{m_{xt}}},$$

and

$$\begin{aligned} V_{\Delta\boldsymbol{\varepsilon}^b,(x,t),(x',t')} &= 1_{\{(x,t)=(x',t')\}}\left(\frac{(1+e^{m_{x,t-1}})^2}{N_{x,t-1}e^{m_{x,t-1}}} + \frac{(1+e^{m_{xt}})^2}{N_{xt}e^{m_{xt}}}\right) \\ &\quad - 1_{\{(x,t)=(x',t')+(0,1)\}}\frac{(1+e^{m_{x,t-1}})^2}{N_{x,t-1}e^{m_{x,t-1}}} \\ &\quad - 1_{\{(x,t)=(x',t')-(0,1)\}}\frac{(1+e^{m_{xt}})^2}{N_{xt}e^{m_{xt}}}, \end{aligned} \tag{30}$$

respectively. The elements of $\boldsymbol{V}_{\boldsymbol{\varepsilon}^s}$ have the form

$$\begin{aligned} V_{\boldsymbol{\varepsilon}^s,(x,t),(x',t')} &= 1_{\{(x,t)=(x',t')\}}\sigma^2_\eta + c_x c_{x'} 1_{\{t=t'\}}\sigma^2_\zeta \\ &\quad + d_x d_{x'} 1_{\{(t-\tilde{t})(t'-\tilde{t})>0\}} \min(|t-\tilde{t}|, |t'-\tilde{t}|)\sigma^2_\kappa, \end{aligned}$$

whereas those of $\boldsymbol{V}_{\Delta\epsilon}$ are given by

$$\begin{aligned} V_{\Delta\boldsymbol{\varepsilon}^s,(x,t),(x',t')} &= 21_{\{(x,t)=(x',t')\}}\sigma^2_\eta + 2c_x c_{x'} 1_{\{t=t'\}}\sigma^2_\zeta \\ &\quad + d_x d_{x'} 1_{\{t=t'\}}\sigma^2_\kappa - 1_{\{(x,t)=(x',t')\pm(0,1)\}}\sigma^2_\eta \\ &\quad - c_x c_{x'} 1_{\{t=t'\pm1\}}\sigma^2_\zeta. \quad \square \end{aligned} \tag{31}$$

WLS estimation of $\boldsymbol{\xi}$. We will utilize (30) and (31) in order to estimate $\boldsymbol{\xi}$ in (16), and tacitly assume that data is LMI transformed, so that this superscript is omitted at several places. Denote the elements of $\boldsymbol{V}^{\mathrm{LMI}}$ by $V^{\mathrm{LMI}}_{(x,t),(x',t')}$, and introduce

$$V_{x,x'}(\tau) = V^{\mathrm{LMI}}_{x,x'}(\tau) = \frac{1}{T-1-|\tau|}\sum_{t=t_1+1}^{t_T-|\tau|} V^{\mathrm{LMI}}_{(x,t),(x',t+|\tau|)}, \tag{32}$$

for each lag $\tau = 0, \pm1, \ldots, \pm(T-2)$. It is the covariance function of $\varepsilon_{xt} = \varepsilon^s_{xt} + \varepsilon^b_{xt}$ and $\varepsilon_{x',t+|\tau|} = \varepsilon^s_{x',t+|\tau|} + \varepsilon^b_{x',t+|\tau|}$, averaged over $t_2 \le t \le t_T - |\tau|$. Analogously, we define an average covariance function

$$V_{\Delta\boldsymbol{\varepsilon}^b,x,x'}(\tau) = \frac{1}{T-1-|\tau|}\sum_{t=t_1+1}^{t_T-|\tau|} V_{\Delta\boldsymbol{\varepsilon}^b,(x,t),(x',t+|\tau|)} \tag{33}$$

from $\boldsymbol{V}_{\Delta\boldsymbol{\varepsilon}^b}$ in (30). It follows from (31) that the covariance of $(\Delta\varepsilon^s_{xt}, \Delta\varepsilon^s_{x',t+|\tau|})$ does not depend on t and

$$\begin{aligned} V_{\Delta\boldsymbol{\varepsilon}^s,x,x'}(\tau) &= \frac{1}{T-1-|\tau|}\sum_{t=t_1+1}^{t_T-|\tau|} V_{\Delta\boldsymbol{\varepsilon}^s,(x,t),(x',t+|\tau|)} \\ &= V_{\Delta\boldsymbol{\varepsilon}^s,(x,t),(x',t+|\tau|)} \\ &= 1_{\{\tau=0\}}\left(1_{\{x=x'\}}2\sigma^2_\eta + 2c_x c_{x'}\sigma^2_\zeta + d_x d_{x'}\sigma^2_\kappa\right) \\ &\quad - 1_{\{|\tau|=1\}}\left(1_{\{x=x'\}}\sigma^2_\eta + c_x c_{x'}\sigma^2_\zeta\right). \end{aligned} \tag{34}$$

Combining (18) with (32)–(34), we find that

$$V_{x,x'}(\tau) = V_{\Delta\boldsymbol{\varepsilon}^b,x,x'}(\tau) + V_{\Delta\boldsymbol{\varepsilon}^s,x,x'}(\tau). \tag{35}$$

In the ith iterate of Step 2 of the estimation algorithm in Section 4.1, we first estimate (32) as

$$\hat{V}_{i,x,x'}(\tau) = \frac{1}{T-1-|\tau|}\sum_{t=t_2}^{t_T-|\tau|} \hat{\varepsilon}^{\mathrm{LMI}}_{i,xt}\hat{\varepsilon}^{\mathrm{LMI}}_{i,x,t+|\tau|},$$

where $\hat{\varepsilon}^{\mathrm{LMI}}_{i,xt} = Y^{\mathrm{LMI}}_{xt} - (\boldsymbol{X}^{\mathrm{LMI}}\hat{\boldsymbol{\theta}}^{\mathrm{LMI}}_i)_{xt}$ are the residuals from the generalized least squares fit of $\boldsymbol{\theta}^{\mathrm{LMI}}$ in iteration i. Then we compute estimates

$$\hat{V}_{i,\Delta\boldsymbol{\varepsilon}^b,x,x'}(\tau) = \frac{1}{T-1-|\tau|}\sum_{t=t_2}^{t_T-|\tau|} \hat{V}_{i,\Delta\boldsymbol{\varepsilon}^b,(x,t),(x',t+|\tau|)}$$

of (33), where $\hat{V}_{i,\Delta\boldsymbol{\varepsilon}^b,(x,t),(x',t+|\tau|)}$ is an estimate of $V_{\Delta\boldsymbol{\varepsilon}^b,(x,t),(x',t+|\tau|)}$ in (30), which can be computed in two ways. Either $m_{xt} = m_{xt}(\boldsymbol{\theta})$ in (9) is replaced by $\hat{m}_{i,xt} = m_{xt}(\hat{\boldsymbol{\theta}}_i)$ everywhere, or $(1+e^{m_{xt}})^2/e^{m_{xt}}$ is replaced by $[\hat{Q}_{xt}(1-\hat{Q}_{xt})]^{-1}$ everywhere. The latter version is simpler and does not depend on the iteration number i.

From the last three displayed equations we find that the elements of (34) can be estimated as

$$\hat{V}_{i,\Delta\boldsymbol{\varepsilon}^s,x,x'}(\tau) = \hat{V}_{i,x,x'}(\tau) - \hat{V}_{i,\Delta\boldsymbol{\varepsilon}^b,x,x'}(\tau) \tag{36}$$

in the ith iteration of the algorithm. We will estimate $\boldsymbol{\xi}$ by fitting (36) to (34) by means of a WLS estimator

$$\begin{aligned} \hat{\boldsymbol{\xi}}_i &= (\hat{\sigma}^2_{i\eta}, \hat{\sigma}^2_{i\zeta}, \hat{\sigma}^2_{i\kappa})^T \\ &= \arg\min_{\sigma^2_\eta,\sigma^2_\zeta,\sigma^2_\kappa} \sum_{x_l\le x\le x'\le x_u}\sum_{\tau=0}^{1} W_{xx'}(\tau) \\ &\quad \times \left(\hat{V}_{i,\Delta\boldsymbol{\varepsilon}^s,x,x'}(\tau) - V_{\Delta\boldsymbol{\varepsilon}^s,x,x'}(\tau)\right)^2, \end{aligned} \tag{37}$$

with pre-defined (or inverse variance) non-negative weights $W_{xx'}(\tau)$, and a constrained minimization, so that all three variance components are non-negative.

Since the variance components enter linearly into (34), we can rewrite these equations jointly for all x, x', τ as

$$\boldsymbol{V}_{\Delta\varepsilon^s} = \boldsymbol{Z}\boldsymbol{\xi}, \tag{38}$$

where $\boldsymbol{V}_{\Delta\varepsilon^s} = \left(V_{\Delta\varepsilon^s,x,x'}(\tau);\ x_l \le x \le x' \le x_u,\ \tau = 0, 1\right)^T$ is a column vector of length $m = (x_u - x_l + 1)(x_u - x_l + 2)$ containing all covariances, and $\boldsymbol{Z}$ a design matrix of dimension $m \times 3$, whose row corresponding to (x, x', τ) equals

$$1_{\{\tau=0\}}1_{\{x=x'\}}\left(2, 2c_x^2, d_x^2\right) + 1_{\{\tau=0\}}1_{\{x<x'\}}\left(0, 2c_xc_{x'}, d_xd_{x'}\right) - 1_{\{\tau=1\}}1_{\{x=x'\}}\left(1, c_x^2, 0\right) - 1_{\{\tau=1\}}1_{\{x<x'\}}\left(0, c_xc_{x'}, 0\right).$$

By adding random noise to (38) in terms of estimation errors, we get a multiple linear regression model

$$\hat{\boldsymbol{V}}_{i,\Delta\varepsilon^s} = \boldsymbol{Z}\boldsymbol{\xi} + (\hat{\boldsymbol{V}}_{i,\Delta\varepsilon^s} - \boldsymbol{V}_{\Delta\varepsilon^s}), \tag{39}$$

where $\hat{\boldsymbol{V}}_{i,\Delta\varepsilon^s} = \left(\hat{V}_{i,\Delta\varepsilon^s,x,x'}(\tau);\ x_l \le x \le x' \le x_u,\ \tau = 0, 1\right)^T$ is the observational column vector of length m. The objective function on the right hand side of (37) can be written as a quadratic function

$$(\boldsymbol{\xi} - \tilde{\boldsymbol{\xi}}_i)^T\boldsymbol{A}(\boldsymbol{\xi} - \tilde{\boldsymbol{\xi}}_i) + \text{constant},$$

of $\boldsymbol{\xi}$, where the last, constant term, does not depend on $\boldsymbol{\xi}$,

$$\tilde{\boldsymbol{\xi}}_i = (\boldsymbol{Z}^T\boldsymbol{W}\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{W}\hat{\boldsymbol{V}}_{i,\Delta\varepsilon^s},$$

is the unconstrained WLS solution of (39) with weight matrix $\boldsymbol{W}$, a diagonal matrix containing all $W_{xx'}(\tau)$, and finally $\boldsymbol{A} = \boldsymbol{Z}^T\boldsymbol{W}\boldsymbol{Z}$ is half the Hessian matrix of the objective function.

Because of the constrained minimization in (37), its solution has the form

$$\hat{\boldsymbol{\xi}}_i = \Pi\tilde{\boldsymbol{\xi}}_i \tag{40}$$

where Π is a projection in $\mathbb{R}^3$ onto the region $[0, \infty) \times [0, \infty) \times [0, \infty)$, defined using the scalar product $(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{y}$ between column vectors in $\mathbb{R}^3$.

A more explicit formula for (40) can be found as follows: Introduce $\boldsymbol{e}_i$, the unit vector in $\mathbb{R}^3$ with 1 in position $i \in \{1, 2, 3\}$ and zeros elsewhere. For each binary vector $\boldsymbol{u} = (u_1, u_2, u_3)$ of length 3, we let $\Pi_{\boldsymbol{u}}\boldsymbol{\xi}$, refer to a projection of $\boldsymbol{\xi}$ down to the linear subspace spanned by vectors $\{\boldsymbol{e}_i;\ u_i = 1\}$. More explicitly, we have that $\Pi_{(0,0,0,)}\boldsymbol{\xi} = (0, 0, 0)$ and

$$\Pi_{\boldsymbol{u}}\boldsymbol{\xi} = (\boldsymbol{B}_{\boldsymbol{u}}^T\boldsymbol{A}\boldsymbol{B}_{\boldsymbol{u}})^{-1}\boldsymbol{B}_{\boldsymbol{u}}^T\boldsymbol{A}\boldsymbol{\xi}$$

for $\boldsymbol{u} \neq (0, 0, 0)$, where $\boldsymbol{B}_{\boldsymbol{u}}$ is a design matrix of order $3 \times (\sum_{i=1}^3 u_i)$, with columns all $\boldsymbol{e}_i$ for which $u_i = 1$. Then

$$\Pi\boldsymbol{\xi} = \Pi_{\boldsymbol{u}(\boldsymbol{\xi})}\boldsymbol{\xi},$$

where

$$\boldsymbol{u}(\boldsymbol{\xi}) = \arg\min_{\boldsymbol{u};\,\Pi_{\boldsymbol{u}}\boldsymbol{\xi}\in[0,\infty)^3}(\boldsymbol{\xi} - \Pi_{\boldsymbol{u}}\boldsymbol{\xi})^T\boldsymbol{A}(\boldsymbol{\xi} - \Pi_{\boldsymbol{u}}\boldsymbol{\xi}). \quad \square$$

Annuity premiums and their prediction. We will first motivate the annuity premium formula (23). Let $\tau = \tau_{xt}$ be the minimum of n and the (continuous) remaining life time of a randomly chosen individual of (integer valued) age x at the beginning of calendar year t. The total discounted premium that this individual pays, is nonzero and equal to

$$\mathcal{P}([\tau]) = P \cdot \sum_{k=0}^{[\tau]-1}(1 + \text{IR})^{-k-1},$$

whenever $[\tau]$, the integer part of τ, is positive. With $\bar{F}(r) = P(\tau \ge r)$ the survival function of τ, we can write the expected annuity premium as

$$\begin{aligned}
\hat{\mathcal{P}}_{xt}(n) &= E\left[\mathcal{P}([\tau])\right] \\
&= \sum_{i=1}^n P([\tau] = i)\mathcal{P}(i) \\
&= P \cdot \sum_{i=1}^n P([\tau] = i)\sum_{k=0}^{i-1}(1 + \text{IR})^{-k-1} \\
&= P \cdot \sum_{k=0}^{n-1}(1 + \text{IR})^{-k-1}\sum_{i=k+1}^n P([\tau] = i) \\
&= P \cdot \sum_{k=0}^{n-1}(1 + \text{IR})^{-k-1}\bar{F}(k+1) \\
&= P \cdot \sum_{k=0}^{n-1}(1 + \text{IR})^{-k-1}\left[\prod_{j=0}^k(1 - \hat{\text{Q}}_{x+j,t+j})\right],
\end{aligned}$$

in agreement with (23).

In order to derive an expression for the annuity premium prediction variance above (25), we Taylor expand f. This gives

$$\begin{aligned}
\sigma^2_{\mathcal{P},xt} &= \text{Var}(\hat{\mathcal{P}}_{xt}(n)|\tilde{\boldsymbol{m}}_{xt}) \\
&= \text{Var}\left[f\left(\tilde{\boldsymbol{m}}_{xt} + \bar{\boldsymbol{\varepsilon}}_{xt}\right)|\tilde{\boldsymbol{m}}_{xt}\right] \\
&\approx \text{Var}\left(\sum_{j=0}^{n-1}\bar{\varepsilon}_{x+j,t+j}\frac{\partial f}{\partial\tilde{m}_{x+j,t+j}}\right) \\
&= \sum_{i,j=0}^{n-1}\text{Cov}(\bar{\varepsilon}_{x+i,t+i}, \bar{\varepsilon}_{x+j,t+j})\frac{\partial f}{\partial\tilde{m}_{x+i,t+i}}\frac{\partial f}{\partial\tilde{m}_{x+j,t+j}},
\end{aligned} \tag{41}$$

where $\bar{\boldsymbol{\varepsilon}}_{xt} = (\bar{\varepsilon}_{x+j,t+j})_{j=0}^{n-1}$, and $\bar{\varepsilon}_{xt} = \varepsilon_{xt} - (\tilde{m}_{xt} - m_{xt})$ is the prediction error at age x and time t. The next step is to find (estimates of) the terms in (41). First, for the prediction error covariances, we split the process error ε_{xt} into systematic and binomial parts (8), as in (22), and use the formula below (30). This gives a covariance

$$\begin{aligned}
\tilde{\sigma}_{xt,x't'} &= \widehat{\text{Cov}}(\bar{\varepsilon}_{xt}, \bar{\varepsilon}_{x't'}) \\
&= \dot{\boldsymbol{m}}_{xt}^T\left(\boldsymbol{X}^T\tilde{\boldsymbol{V}}^{-1}\boldsymbol{X}\right)^{-1}\dot{\boldsymbol{m}}_{x't'} \\
&\quad + 1_{\{(x,t)=(x',t')\}}\left[N_{xt}q_{xt}(\tilde{\boldsymbol{\theta}})(1 - q_{xt}(\tilde{\boldsymbol{\theta}}))\right]^{-1} \\
&\quad + 1_{\{(x,t)=(x',t')\}}\tilde{\sigma}_\eta^2 + 1_{\{t=t'\}}c_xc_{x'}\tilde{\sigma}_\zeta^2 \\
&\quad + d_xd_{x'}\min(t - t_T, t' - t_T)\tilde{\sigma}_\kappa^2
\end{aligned}$$

between cells (x, t) and (x', t'). Notice in particular that the prediction variance $\tilde{\sigma}_{xt,xt}$ equals $\tilde{\sigma}_{xt}^2$ in (22). Second, for the partial derivatives of f we get

$$\begin{aligned}
\frac{\partial f}{\partial\tilde{m}_{x+j,t+j}} &= -P \cdot \sum_{k=j}^{n-1}\left[\left(\prod_{i=0}^k\frac{1}{1 + e^{\tilde{m}_{x+i,t+i}}}\right)\cdot\frac{e^{\tilde{m}_{x+j,t+j}}}{1 + e^{\tilde{m}_{x+j,t+j}}}\right] \\
&\quad \times (1 + \text{IR})^{-k-1} \\
&= -P \cdot \tilde{\text{Q}}_{x+j,t+j}\sum_{k=j}^{n-1}\left(\prod_{i=0}^k\frac{1}{1 + e^{\tilde{m}_{x+i,t+i}}}\right)(1 + \text{IR})^{-k-1} \\
&= -P \cdot \tilde{\text{Q}}_{x+j,t+j}\prod_{i=0}^j(1 - \tilde{\text{Q}}_{x+i,t+i}) \\
&\quad \cdot \sum_{k=j}^{n-1}\left[\prod_{i=j+1}^k(1 - \tilde{\text{Q}}_{x+i,t+i})\right](1 + \text{IR})^{-k-1},
\end{aligned}$$

where the last product equals 1 when $k = j$. $\square$

Life expectancy formula. We will motivate the life expectancy formula described in the last paragraph of Section 4.2. Let $\tau = \tau_{xt}$ be the remaining life time of a randomly chosen individual of age x at the beginning of calendar year t. Since the integer valued maximal age is $x_{\max}$, we assume that τ has a continuous distribution on $(0, x_{\max} + 1)$, since an individual may live up to birthday $x_{\max} + 1$ is reached. Recalling that $n = x_{\max} - x$, we write

$$\tau = \tau_1 + \tau_2,$$

where $\tau_1 = \min(n, \tau)$ is the remaining life time within n years from the start of calendar year t, and $\tau_2 = \max(0, \tau - \tau_1)$ is the remaining life time between n and $n + 1$ years from this time point. Let $f(r)$ and $\bar{F}(r) = P(\tau \geq r)$ be the density and survival functions of τ, and let X represent the time of the year of this individual's birthday, uniformly distributed on $(0, 1)$. Assume that f is piecewise constant on all one year intervals $[0, 1], \ldots, [n - 1, n]$, so that $\bar{F}$ is piecewise linear on $[0, n]$. Assume further that τ_2 has a uniform distribution on $(0, X)$ for an individual with birthday X that survives n years. Then integrals can be replaced by Riemann sums, so that

$$\begin{aligned}
\hat{\mathring{e}}_{xt} &= E(\tau) \\
&= E(\tau_1) + E(\tau_2) \\
&= \left[\int_0^n r f(r) dr + n\bar{F}(n)\right] + \bar{F}(n) E(\tau - n | \tau > n) \\
&= \sum_{i=0}^{n-1} (i + 0.5) f(i + 0.5) + n\bar{F}(n) \\
&\quad + \bar{F}(n) E\left[E(\tau - n | \tau > n, X = x)\right] \\
&= \sum_{i=0}^{n-1} f(i + 0.5) \left(0.5 + \sum_{k=0}^{i-1} 1\right) + n\bar{F}(n) + \bar{F}(n) E(0.5X) \\
&= \sum_{k=0}^{n-1} \left[0.5 f(k + 0.5) + \sum_{i=k+1}^{n-1} f(i + 0.5)\right] \\
&\quad + n\bar{F}(n) + 0.25\bar{F}(n) \\
&= \sum_{k=0}^{n-1} \left[\bar{F}(k + 0.5) - \bar{F}(n)\right] + n\bar{F}(n) + 0.25\bar{F}(n) \\
&= \sum_{k=0}^{n-1} \bar{F}(k + 0.5) + 0.25\bar{F}(n) \\
&= \sum_{k=0}^{n-1} (1 - 0.5\hat{Q}_{x+k,t+k}) \prod_{j=0}^{k-1} (1 - \hat{Q}_{x+j,t+j}) \\
&\quad + 0.25 \prod_{j=0}^{n-1} (1 - \hat{Q}_{x+j,t+j}),
\end{aligned}$$

where any sum (product) for which the upper index is smaller than the lower one, is defined as 0 (1). □

References

Aitken, A.C., 1935. On least squares and linear combinations of observations. Proc. Roy. Soc. Edinburgh 55, 42–48.

Alho, J.M., 2000. Discussion of Lee (2000). N. Am. Actuar. J. 4, 91–93.

Barrieu, P., et al., 2012. Understanding, modelling and managing longevity risk: key issues and main challenges. Scand. Actuar. J. 2012 (3), 203–231.

Bell, W.R., 1997. Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates. J. Off. Stat. 13 (3), 279–303.

Bell, W.R., Monsell, B.C., 1991. Using principal components in time series modeling and forecasting of age specific mortality rates. In: Proceedings of the American Statistical Association, Social Statistics Session, pp. 154–159.

Booth, H., Tickle, L., 2008. Mortality modelling and forecasting: A review of methods. Ann. Actuar. Sci. 3 (I–II), 3–43.

Bozik, J.E., Bell, W.R., Forecasting age specific mortality using principal components. In: Proceedings of the American Statistical Association, Social Statistics Session, 1987, pp. 396–401.

Brass, W., 1971. On the scale of mortality. In: Brass, W. (Ed.), Biological Aspects of Demography. Taylor and Francis, London, pp. 69–110.

Breslow, N.E., Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. J. Amer. Statist. Assoc. 88 (421), 9–25.

Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods, second ed. Springer-Verlag, New York.

Brouhns, N., Denuit, M., van Keilegom, I., 2005. Bootstrapping the log-bilinear model for mortality forecasting. Scand. Actuar. J. 2005 (3), 212–224.

Brouhns, N., Denuit, M., Vermunt, J.K., 2002. A Poisson log-bilinear regression approach to the construction of projected lifetables. Insurance Math. Econom. 31, 373–393.

Cairns, A.J.G., 2013. Robust hedging of longevity risk. J. Risk Insurance 80 (3), 621–648.

Cairns, A.J.G., 2014. Modeling and management of longevity risk. In: Hammond, P.B., Maurer, R., Mitchell, O.S. (Eds.), Recreating Sustainable Retirement: Resilience, Solvency, and Tail Risk. Oxford University Press, Oxford, UK, pp. 71–88.

Cairns, A.J.G., Blake, D., Dowd, K., 2006. A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. J. Risk Insurance 73 (4), 687–718.

Cairns, A.J.G., Blake, D., Dowd, K., 2008. Modelling and management of mortality risk: a review. Scand. Actuar. J. 2008 (2–3), 79–113.

Cairns, A.J.G., Blake, D., Dowd, D., Couglan, G.D., Epstein, D., Ong, A., Balevich, I., 2009. A quantitative comparison of stochastic mortality models using data from England & Wales and the United States. N. Am. Actuar. J. 13, 1–35.

Cairns, A.J.G., Blake, D., Dowd, D., Couglan, G.D., Khalaf-Allah, M., 2011. Bayesian stochastic mortality modelling for two populations. ASTIN Bull. 41, 29–59.

Cairns, A.J.G., Dowd, D., Blake, D., Coughlan, G.D., 2014. Longevity hedge effectiveness: A decomposition. Quant. Finance 14 (2), 217–235.

Coelho, E., Nunes, L.C., 2011. Forecasting mortality in the event of a structural change. J. R. Stat. Soc. Ser. A 174 (3), 713–736.

Cossette, H., Delwarde, A., Denuit, M., Guillot, F., Marceau, E., 2007. Pension plan valuation and dynamic mortality tables. N. Am. Actuar. J. 11, 1–34.

Currie, I.D., 2013. Smoothing constrained generalized linear models with an application to the Lee–Carter model. Stat. Model. 13 (1), 69–93.

Currie, I.D., Durban, M., Eilers, P.H.C., 2004. Smoothing and forecasting mortality rates. Stat. Model. 4, 279–298.

Delwarde, A., Denuit, M., Partrat, D., 2007a. Negative binomial version of the Lee–Carter model for mortality forecasting. Appl. Stoch. Models Bus. Ind. 23, 385–401.

Delwarde, A., Denuit, M., Eilers, P., 2007b. Smoothing the Lee–Carter and Poisson log-bilinear models for mortality forecasting: a penalized log-likelihood approach. Stat. Model. 7 (1), 29–48.

Denton, F.T., Feaver, C.H., Spencer, B.G., 2004. Time series analysis and stochastic forecasting: An econometric study of mortality and life expectancy. J. Popul. Econ. 18, 223–227.

Djeundje, V.A.B., Currie, I.D., 2010. Smoothing dispersed counts with applications to mortality data. Ann. Actuar. Sci. 5 (1), 33–52.

Eilers, P.H.C., Marx, B.D., 1996. Flexible smoothing with B-splines and penalties. Statist. Sci. 11 (2), 89–121.

Eilers, P.H., Gampe, J., Marx, B.D., Rau, R., 2008. Modulation models for seasonal time series and incidence tables. Stat. Med. 27 (17), 3430–3441.

Ekheden, E., Hössjer, O., 2014. Analysis of the stochasticity of mortality using variance decomposition. In: Silvestrov, D., Martin-Löf, A. (Eds.), Modern Problems in Insurance Mathematics. In: EEA Series, Springer International Publishing, Cham, Switzerland, pp. 199–222. Chapter 13 in.

Goldstein, H., 1986. Multilevel mixed linear model analysis using iterative generalized least squares. Biometrika 73, 43–56.

Guerrero, V.M., Silva, E., 2010. Non-parametric and structured graduation of mortality rates. Popul. Rev. 49 (2), 13–26.

Hall, M., Friel, N., 2010. Mortality projections using generalized additive models with applications to annuity values for the Irish population. Ann. Actuar. Sci. 5 (1), 19–32.

Hansen, C.B., 2007. Generalized least squares inference in panel and multipanel models with serial correlation and fixed effects. J. Econometrics 140, 670–694.

Harville, D.A., 1977. Maximum likelihood approaches to variance component estimation and to related problems. J. Amer. Statist. Assoc. 72 (358), 320–337.

Henderson, C.R., 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31 (2), 423–447.

Hössjer, O., 2008. On the coefficient of determination for mixed regression models. J. Statist. Plann. Inference 138, 3022–3038. http://dx.doi.org/10.1016/j.jspi.2007.11.010.

Hössjer, O., Eriksson, B., Järnmalm, K., Ohlsson, E., 2009. Assessing individual unexplained variation in non-life insurance. ASTIN Bull. 39 (1), 249–273.

Hyndman, R.J., Booth, H., Tickle, L., Maindonald, J., 2014. R-package demography, version 1.18, September 15, 2014. http://robjhyndman.com/software/demography/.

Imoto, S., Konishi, S., 2003. Selection of smoothing parameters in B-spline nonparametric regression models using information criteria. Ann. Inst. Statist. Math. 55 (4), 671–687.

Koissi, M.-C., Shapiro, A.F., Högnäs, G., 2006. Evaluating and extending the Lee–Carter model for mortality forecasting: bootstrap confidence interval. Insurance Math. Econom. 39, 287–309.

Lee, R.D., Carter, L.R., 1992. Modelling and forecasting US mortality. J. Amer. Statist. Assoc. 87 (419), 659–671.
Lee, Y., Nelder, J.A., Pawitan, Y., 2006. Generalized Linear Models with Random Effects. Unified Analysis via *H*-Likelihood. Chapman and Hall/CRC, Boca Raton.
Li, J.S.H., Hardy, M.R., Tan, K.S., 2009. Uncertainty in mortality forecasting: an extension to the classic Lee–Carter approach. ASTIN Bull. 39, 137–164.
Li, J.S.H., Chan, W.S., Cheung, S.H., 2011. Structural changes in the Lee–Carter mortality indexes: detection and implications. N. Am. Actuar. J. 15 (1), 13–31.
Li, N., Lee, R., 2005. Coherent mortality forecasts for a group of populations: An extension of the Lee–Carter model. Demography 42 (3), 575–594.
Osmond, C., 1985. Using age, period and cohort models to estimate future mortality rates. Int. J. Epidemiol. 14, 124–129.
Parks, R., 1967. Efficient estimation of a system of regression functions when when disturbances are both serially correlated and contemporaneously correlated. J. Amer. Statist. Assoc. 62, 500–509.
Pedroza, C., 2006. A Bayesian forecasting model: predicting US male mortality. Biostatistics 7 (4), 530–550.
Renshaw, A.E., Haberman, S., 2003a. Lee–Carter mortality forecasting: a parallel generalized linear modelling approach for England and Wales mortality projections. Appl. Stat. 52 (1), 119–137.
Renshaw, A.E., Haberman, S., 2003b. On the forecasting of mortality reduction factors. Insurance Math. Econom. 32 (3), 379–401.
Renshaw, A.E., Haberman, S., 2003c. Lee–Carter mortality forecasting with age-specific enhancement. Insurance Math. Econom. 33, 255–272.
Renshaw, A.E., Haberman, S., 2006. A cohort-based extension to the Lee–Carter model for mortality reduction factors. Insurance Math. Econom. 38 (3), 556–570.
Robinson, G.K., 1991. That BLUP is a good thing: The estimation of random effects. Statist. Sci. 6 (1), 15–32.
Rue, H., Held, L., 2005. Gaussian Markov Random Fields: Theory and Applications. Chapman and Hall/CRC, Boca Raton, FL.
Sweeting, P.J., 2011. A trend-change extension of the Cairns–Blake–Dowd model. Ann. Actuar. Sci. 5 (2), 143–162.
van Berkum, F., Antonio, K., Vellekoop, M., 2013. Structural changes in mortality rates with an application to Dutch and Belgian data. Report AFI 1379. Faculty of Economics and Business, Katholieke Universiteit Leuven, Belgium.
White, K.M., 2002. Longevity advance in high-income countries, 1955–1996. Popul. Dev. Rev. 28 (1), 59–76.
Wilmoth, J.R., 1993. Computational methods for fitting and extrapolating the Lee–Carter model of mortality change. Technical Report. Dept. of Demography, University of California, Berkeley.