Institute
and Faculty
of Actuaries

ORIGINAL RESEARCH PAPER

On the integration of deterministic opinions into mortality smoothing and forecasting

Viani Biatat Djeundje [ID]

University of Edinburgh, United Kingdom
E-mail: viani.djeundje@ed.ac.uk

Abstract

Modelling and forecasting mortality is a topic of crucial importance to actuaries and demographers. However, forecasts from the majority of mortality projection models are continuations of past trends seen in the data. As such, these models are unable to account for external opinions or expert judgement. In this work, we present a method for the incorporation of deterministic opinions into the smoothing and forecasting of mortality rates using constraints. Not only does our approach yield a smooth transition from the past into the future, but also, the shapes of the resulting forecasts are governed by a combination of the opinion inputs and the speed of improvements observed in the data. In addition, our approach offers the possibility to compute the amount of uncertainty around the projected mortality trends conditional on the opinion inputs, and this allows us to highlight some of the pitfalls of deterministic projection methods.

Keywords: Mortality modelling; Forecasting; Expert opinions; Smoothing; Conditional uncertainty

1. Introduction

Most actuarial calculations and demographic projections involve assumptions about the future. In life insurance in particular, the future trajectories of mortality trends are of crucial importance because unanticipated mortality improvements (Willets, 1999) can have detrimental effects not only on the planning of national social programs but also on the financial stability of the insurance industry, including pension plans and annuities.

Various mortality projection models have been proposed and developed successfully, including the Lee–Carter model and its extensions (Lee & Carter, 1992; Renshaw & Haberman, 2006; Brouhns *et al.*, 2002; Delwarde *et al.*, 2007; Debón *et al.*, 2010), the Cairns-Blake-Dowd model and its extensions (Cairns *et al.*, 2006; Cairns *et al.*, 2009; Plat, 2009), and P-splines smooth models (Currie *et al.*, 2004; Currie *et al.*, 2006; CMI, 2006; Djeundje & Currie, 2010; Biatat & Currie, 2010) among others. These models are built on different assumptions and they often yield different projected trends. Thus, in practice, using several projection models generally gives more insight into the potential directions of mortality improvements than using a single model; see Debón *et al.* (2010), Woods (2016), Richards *et al.* (2014), and Richards *et al.* (2017) for discussion and illustration.

There is a role for flexible models that can integrate emerging information or expert opinions in such a way that the impact of these opinions can be switched on or off. Indeed, this would allow better understanding and greater scrutiny of the financial implications of potential mortality trajectories. However, a common feature of current mortality projection models is that they are extrapolative. That is, they are built on historical data and the resulting projected trends are continuations of past behaviour in mortality rates. As such, these models are not flexible enough to

allow a direct incorporation of external opinions (or expert judgement) in the calibration process or in the exploration of 1-in-200 years type events.

The desire to incorporate opinions into mortality forecasting is not new (Janssen & Kunst, 2007; Stoeldraijer *et al.*, 2013). In this context, short-term projections are usually treated separately from long-term forecasts, and the former is based on the general principle that the near future resembles the recent past (Andreev & Vaupel, 2006). However, long-term projections are tricky and the outcome crucially depends on the choices made as part of the forecasting procedure. For example, it is well known that the use of general extrapolative mortality forecasting methods can yield unlikely future patterns in the long run (Currie, 2015; Janssen *et al.*, 2013).

One way to overcome these issues is to include information about the determinants of past trends in the mortality forecast. For example, several studies have attempted to incorporate epidemiological information into mortality forecasting; see French & O'Hare (2014), Janssen *et al.* (2013) or Preston *et al.* (2014) among others.

In the context of longevity insurance, actuaries find alternative ways to account for expert opinions in their pricing and liability calculations. For longevity risk in the UK for instance, mortality opinions are often expressed in terms of long-term mortality improvements and are incorporated into the calculation process via a two-stage procedure. The first stage consists of fitting a mortality model without forecasting, and the second stage is to draw *sensible lines* between the fitted rates and the assumed long-term rate (CMI, 2006, 2016). However, there are caveats with this type of procedure. For example, it can lead to discontinuities as one moves from the past into the future. In addition, the speed or direction of travel are subjective. Furthermore, it is difficult to quantify the uncertainty around the projected mortality trends from such a procedure, unless one relies on some post adjustments of the output from a separate stochastic mortality model; see for example Cairns (2017).

In this work, we present a direct method for the incorporation of deterministic opinions into the smoothing and forecasting of mortality rates. Our method is built in the framework of P-splines (Eilers & Marx, 1996; Currie *et al.*, 2004) and the estimation methodology is adapted from the contained equations in Currie (2013).

The reason for choosing P-splines framework is that, not only does it allow the smoothing of the data and forecasting to be performed in one stage, but also, the framework is flexible enough to handle even extreme opinion specifications. Our contribution is (*a*) to show how to appropriately reparameterise deterministic opinions stated in terms of standard mortality metrics into the model, (*b*) to describe how to calibrate a P-splines mortality model in conjunction with opinion inputs in such a way that the resulting mortality trends are a combination of the signals from past trends and the expert opinions, (*c*) to quantify the amount of uncertainty around the central projected trends conditional on the opinion inputs, and (*d*) to apply our method to real-world mortality data under various opinion scenarios.

There are many alternative smoothing frameworks in the literature including kernel smoothing, splines smoothing, locally weighted regression, direct smoothing, etc. A comparison of some of these methods in the mortality context was carried out by Debón *et al.* (2006). More recently, Ludkovski *et al.* (2018) described how mortality smoothing can be carried out through a kernel method within a Gaussian Process framework, and illustrated how this can be implemented in a Bayesian way using Markov Chain Monte Carlo method. An attractive feature of the Bayesian method resides in the possibility to incorporate stochastic prior information by specifying prior distributions for the model parameters.

In practice in the longevity industry however, the use of deterministic opinions remains very popular (CMI, 2006, 2016). In addition, prior opinions are often formulated on the scale of standard mortality metrics; the conversion of such prior specifications into prior distributions on model parameters is not always straightforward. Our work in this study focuses on the incorporation of deterministic opinions via penalty and constraints. In particular, we show how opinions stated directly in terms of popular mortality metrics can be built into a smooth mortality model

using well-known statistical techniques. The method of P-spline is our underlying smoothing method. A detailed exploration of the benefits of P-splines compared to other smoothing methods can be found in Eilers & Marx (2010). For example, the resulting smooth surface can be decomposed into an age component, a time component, and an interaction term through the underlying difference penalty. A thorough description and illustration of this within the spatio-temporal framework can be found in Lee & Durban (2011).

This paper is organised as follows. Section 2 provides a brief overview of mortality smoothing using the method of P-splines and discusses some practical challenges often encountered in the actuarial mortality context. Section 3 describes how to incorporate deterministic opinions into mortality smoothing and forecasting. Section 4 presents some applications of our method, and we close with some concluding remarks in Section 5.

2. Smoothing Mortality Data

Our approach to the incorporation of opinion inputs requires a flexible smooth modelling framework to start with.

Thus, in this section, we begin with a brief description of the flexible method of P-splines smoothing as it applies to mortality data, and set up some notation for the rest of the paper. For ease of presentation, we shall start in one dimension and then move to two dimensions.

2.1 Smoothing mortality data in one dimension

Let us consider mortality at a given age, x, for calendar years from t_1 to t_n in ascending order, and let us denote by $\boldsymbol{D}_x = (D_1, ..., D_n)$ and $\boldsymbol{E}_x = (E_1, ..., E_n)$ the vectors of death counts and central exposed-to-risk, respectively.

A standard way to estimate trends from aggregated mortality data is based on the assumption that the death experience follows a Poisson[1] distribution with mean proportional to the exposed-to-risk:

$$D_j \sim \mathcal{Poisson}\left(E_j \times \mu_j\right), \; g(\mu_j) = \mathcal{S}(t_j) \tag{1}$$

where μ_j represents the force of mortality in calendar year t_j, g is the link function, and $\mathcal{S}$ is a smooth function that we want to estimate and extrapolate.

There are several ways to estimate a smooth function, one of the most appealing approaches being the method of P-splines (Eilers & Marx, 1996; Currie et al., 2004).

The method shares several features with standard regression. In particular, it involves expressing the smooth function $\mathcal{S}$ as a linear combination of a basis of B-splines:

$$\mathcal{S}(t) = \sum_{k=1}^{c} B^{\{k\}}(t) \times \theta_k \tag{2}$$

where the $B^{\{k\}}(t)$, $1 \le k \le c$, represent the values of the B-spline functions at time t, and the θ_k denote the regression coefficients associated with the B-spline basis.

In practice, the smooth aspect of the model is achieved by penalising differences in adjacent coefficients via the optimisation of the penalised log-likelihood, ℓ_P, given by

$$\ell_P(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}) - \boldsymbol{\theta}' \boldsymbol{P}_\lambda \boldsymbol{\theta}, \quad \text{with } \boldsymbol{P}_\lambda = \lambda \boldsymbol{\Delta}' \boldsymbol{\Delta} \tag{3}$$

In this expression, $\boldsymbol{\theta} = (\theta_1, ..., \theta_c)'$ represents the joint vector of coefficients, $\ell(\boldsymbol{\theta})$ is the ordinary Poisson log-likelihood arising from (1), $\boldsymbol{P}_\lambda$ is the roughness penalty matrix, λ is the scalar

---

[1]In practice, some aggregated mortality data are prone to over-dispersion. This invalidates the Poisson assumption and can lead to under-smoothing and unstable forecasts. A simple way to circumvent this violation is to replace the Poisson distribution by the quasi-Poisson (Djeundje & Currie, 2010; CMI, 2014).

smoothing parameter, and Δ is the difference matrix operator. In practice, the second-order difference is often preferred because it produces sufficient flexibility over the data range, and when used for forecasting, the shape of the extrapolated spline coefficients ties reasonably well with that of the fitted coefficients, provided there is a sufficiently strong signal in the forecasting direction.

If we denote by t the joint vector of time points, i.e. $t = (t_1, ..., t_n)$, by B_t the matrix of B-splines along the time points t, i.e. $B_t = [B^{\{1\}}(t) : \cdots : B^{\{c\}}(t)]$, then, the value of the coefficient vector that maximises the penalised log-likelihood in (3) is found by solving the following penalised version of the scoring algorithm:

$$\left(B_t' \, \tilde{W} B_t \, + \, P_\lambda \right) \hat{\theta} = B_t' \, \tilde{W} \, \tilde{z}, \tag{4}$$

where W is the diagonal weight matrix and z is the so-called working variable; tilde ($\tilde{\ }$) refers to an approximate solution, and hat ($\hat{\ }$) refers to an improved estimate (Currie *et al.*, 2004).

Actuaries often need to forecast mortality into the far future when calculating present values of pension and annuity liabilities. With the method of P-splines, mortality projection is treated as a missing data problem, and the penalty is used to fill-in the missing data. For example, let us consider mortality data for calendar years $t = (t_1, ..., t_n)$. To forecast for r years into the future, these r future years are first appended to t and the B-splines are computed along the augmented time vector $t_+ = (t_1, ..., t_n, t_{n+1}, ..., t_{n+r})$. Thus, the original $n \times c$ B-spline matrix B_t becomes an augmented $(n + n_+) \times c_+$ matrix of B-splines in time. Accordingly, an augmented iterative system similar to (4) is solved, yielding estimates of the spline regression coefficients for both the data and forecasting regions.

So far, we have overlooked a very important issue: the choice of the smoothing parameter, λ. To provide a reasonable balance between the conflicting characteristics of goodness-of-fit and parsimony, λ is often selected as the minimiser of the Bayesian Information Criteria (BIC) given by

$$\text{BIC} = \text{Deviance} + \log(n) \times \text{ED} \tag{5}$$

where ED represents the effective dimension of the model.

2.2 Smoothing mortality data in two dimensions

Population mortality data are generally available in a two-dimensional form.

The P-spline machinery in this context works by analogy to the one-dimensional case.

Let us denote by $x = (x_1, ..., x_{n_1})$ and $t = (t_1, ..., t_{n_2})$ the vectors of age and year indices. Also, let D represents the $n_1 \times n_2$ table of death counts and D_{ij} the entry of D corresponding to age x_i in calendar year t_j. Similarly, let E and E_{ij} represent corresponding quantities in the exposure data; i.e. E is the $n_1 \times n_2$ matrix of exposed-to-risk of death, and E_{ij} is its entry corresponding to age x_i and calendar year t_j. The basic model assumption (1) is extended to

$$D_{ij} \sim \mathcal{Poisson}\left(E_{ij} \times \mu_{ij} \right), \; g(\mu_{ij}) = \mathcal{S}(x_i, t_j) \tag{6}$$

where $\mathcal{S}$ is now a bivariate smooth function.

To estimate $\mathcal{S}$, we express it in terms of marginal bases of B-splines in age and time; i.e.

$$\mathcal{S}(x_i, t_j) = \sum_{k=1}^{c_1} \sum_{l=1}^{c_2} B_1^{\{k\}}(t_j) \times B_2^{\{l\}}(x_i) \times \Theta_{kl} \tag{7}$$

where the $B_1^{\{k\}}$ and $B_2^{\{l\}}$ are marginal B-splines in age and time, respectively, and the Θ_{kl} are coefficients to be estimated.

Let us denote by Θ the $c_1 \times c_2$ matrix whose entries are the Θ_{kl}. By analogy to the one-dimensional case, the smoothness of the model is achieved by penalising the rows and columns of Θ. If we denote by θ the $c_1 c_2$-length vector obtained by stacking the columns of Θ, i.e. $\theta = vec(\Theta)$,

then, this two-dimensional penalisation of the rows and the columns of $\boldsymbol{\Theta}$ is equivalent to applying the penalty matrix $\boldsymbol{P}_{\lambda_1,\lambda_2}$ on the coefficient vector $\boldsymbol{\theta}$, where

$$\boldsymbol{P}_{\lambda_1,\lambda_2} = \lambda_1 \left(\boldsymbol{I}_{c_1} \otimes \boldsymbol{\Delta}_1' \boldsymbol{\Delta}_1 \right) + \lambda_2 \left(\boldsymbol{\Delta}_2' \boldsymbol{\Delta}_2 \otimes \boldsymbol{I}_{c_1} \right). \tag{8}$$

In this expression, λ_1 and λ_2 are smoothing parameters in age and time; $\boldsymbol{\Delta}_1$ and $\boldsymbol{\Delta}_2$ are second order difference operators in age and time; $\boldsymbol{I}_c$ is the $c \times c$ identity matrix, and $\otimes$ is the Kronecker product.

With this in place, model fitting is carried out as in the one-dimensional case, but, with death data given by $vec(\boldsymbol{D})$, exposure data $vec(\boldsymbol{E})$, regression matrix $\boldsymbol{B_t} \otimes \boldsymbol{B_x}$, link function g, and penalty matrix $\boldsymbol{P}_{\lambda_1,\lambda_2}$. Projection can also take place not only in time, but also in age, by augmentation of the relevant marginal bases as described in Section 2.1. This shall be illustrated in Section 4.

One challenge with P-splines reported in the actuarial context is what is known as the *edge effects*. That is, the fitted mortality rates toward the edge of the data can be unduly influenced by the relative level of experience in that final year (CMI, 2006). One may argue that this can potentially lead to unrealistic mortality forecasts.

However, it is important to bear in mind that in any forecasting method, beside the data, forecasts are also controlled by a number of key parameters. With P-splines in particular, forecasts can be controlled by the splines knots spacing, the smoothing parameters and the difference order of the penalty. Stable mortality forecasts can be achieved through appropriate selection of these parameters. See for example Richards (2009) on the illustration of the importance of the knots spacing, Djeundje (2011, section 5.3) on the importance of appropriate smoothing parameters, or Carballoa *et al.* (2017) on the quantification of the impact of individual data points on the forecast. In addition, the incorporation of opinion inputs into the estimation process (as we shall see in the next sections) does reduce the impact of the data at the *edges* on the direction of the forecast, especially as one approaches the opinion points.

3. Integrating opinions into the Model

Actuaries like to investigate mortality scenarios subject to some prior or expert opinions, in conjunction with scenarios produced by full stochastic mortality projection models. These investigations are usually carried out through various metrics, including the force of mortality, mortality rates, mortality improvements, or mortality reduction factors. In the projection method used by CMI (2006, 2016) for instance, an important component of the user's opinions is specified in terms of long-term rates of mortality improvements.

In this section, we look at how to build deterministic opinions into mortality smoothing and forecasting. For ease of presentation, we shall assume that any opinions about mortality can be expressed clearly on the scale of the force of mortality (equivalently, mortality rates) or in terms of mortality improvements (equivalently, mortality reduction factors). We start in one dimension as in Section 2. Thus, let us consider mortality data $\boldsymbol{D}_x$ and $\boldsymbol{E}_x$ at a given age, x, for calendar years $\boldsymbol{t} = (t_1, ..., t_n)$.

3.1 Incorporating deterministic opinions about the force of mortality

On the scale of the force of mortality, we are interested in projection scenarios in which the force of mortality at age x in calendar years $\boldsymbol{t}_o = (t_{o1}, ..., t_{ou})$ is known a priori, where u is an integer. We shall denote by, $\mathring{\mu}_{x,t_{ok}}$, the a priori value of the force of mortality at age x and calendar year t_{ok}, $k = 1, ..., u$.

For example, let us consider mortality data at age, $x = 60$, for calendar years, $\boldsymbol{t} = (1970, ..., 2010)$. We might want to fit a smooth line to these data and forecast the trend into future in such a way that the forecast values of the force of mortality at calendar years $\boldsymbol{t}_o = (2025, 2030)$ are set to $\mathring{\mu}_{60,2025} = 0.006$ and $\mathring{\mu}_{60,2030} = 0.005$.

In general, let us assume that we want to estimate and forecast a smooth mortality trend to the data as described in Section 2, but, such that the mortality trend fulfils the following conditions:

$$\mu_{x,t_{ok}} = \mathring{\mu}_{x,t_{ok}}, \ 1 \le k \le u \tag{9}$$

where $\mathring{\mu}_{x,t_{ok}}$ are known values of the force of mortality at age x in calendar years t_{ok}, and $\mu_{x,t_{ok}}$ denotes the fitted force of mortality. The $\mathring{\mu}_{x,t_{ok}}$ represent prior opinion inputs on the scale of the force of mortality.

This prior opinion (9) imposes some restriction on the shape and trajectory of the fitted mortality trend. We will refer to this type of equations as the *opinion constraints*. There is no restriction on the time locations t_{ok}'s of these opinion constraints.

The opinion constraints (9) can be expressed compactly as

$$\boldsymbol{B}_{t_o}\boldsymbol{\theta} = g\left(\mathring{\boldsymbol{\mu}}_{x,t_o}\right) \tag{10}$$

where $\boldsymbol{B}_{t_o}$ denotes the $u \times c$ sub-matrix of the B-spline matrix in (2) whose rows correspond to the opinion times $\boldsymbol{t}_o$; g is the link function as in (1), and $\mathring{\boldsymbol{\mu}}_{x,t_o} = (\mathring{\mu}_{x,t_{o1}}, ..., \mathring{\mu}_{x,t_{ou}})$ is the joint vector of known forces of mortality.

Thus, we have defined a Penalised Generalised Linear Model with Poisson error (1)–(2), penalty matrix $\boldsymbol{P}_\lambda$, link function g, and deterministic opinions given by (10). The estimation process is described in Section 3.3 below.

3.2 Incorporating deterministic opinions about mortality improvements

Alternatively, one may want to investigate a mortality scenario in which the mortality improvement rates at a given age x in calendar years $\boldsymbol{t}_o = (t_{o1}, ..., t_{ou})$ are known a priori. We shall denote by, $\mathring{\imath}_{x,t_{ok}}$, the a priori value of the mortality improvement rate at age x and calendar year t_{ok}, $k = 1, ..., u$.

For example, let us consider mortality data at age, $x = 60$, for calendar years, $\boldsymbol{t} = (1970, \ldots, 2010)$. We might want to fit a smooth line to these data and forecast the trend into future in such a way that the resulting forecast of the mortality improvement rates in calendar years $\boldsymbol{t}_o = (2025, 2030)$ is set to $\mathring{\imath}_{60,2025} = 1.5\%$ and $\mathring{\imath}_{60,2030} = 1.1\%$.

In general, let us assume that we want to fit and forecast a smooth mortality trend to the data as described in Section 2.1, but such that

$$\frac{q_{x,t_{ok}-1} - q_{x,t_{ok}}}{q_{x,t_{ok}-1}} = \mathring{\imath}_{x,t_{ok}}, \ 1 \le k \le u \tag{11}$$

where $\mathring{\imath}_{x,t_{ok}}$ are known values of mortality improvements at age x in calendar years t_{ok}.

We propose two methods to express the opinion constraints (11) in a similar form as (10). The first method is an approximation whereas the second method is exact.

3.2.1 Approximate incorporation of mortality improvements opinions

The left-hand side of equation (11) can be expressed in terms of the force of mortality as

$$\frac{q_{x,t_{ok}-1} - q_{x,t_{ok}}}{q_{x,t_{ok}-1}} = 1 - \frac{q_{x,t_{ok}}}{q_{x,t_{ok}-1}}$$

$$= 1 - \frac{1 - \exp\left(-\mu_{x,t_{ok}}\right)}{1 - \exp\left(-\mu_{x,t_{ok}-1}\right)}$$

Applying the first-order Taylor expansion for "exp (y)" yields

$$\frac{q_{x,t_{ok}-1} - q_{x,t_{ok}}}{\hat{q}_{x,t_{ok}-1}} \approx 1 - \frac{\mu_{x,t_{ok}}}{\mu_{x,t_{ok}-1}}$$

Equation (11) becomes

$$1 - \mathring{i}_{x,t_{ok}} \approx \frac{\mu_{x,t_{ok}}}{\mu_{x,t_{ok}-1}} \tag{12}$$

Thus, assuming the Poisson canonical link, i.e. $g(\mu) = \ln(\mu)$, the opinion constraints (11) can be rearranged compactly as

$$\left(\boldsymbol{B}_{t_o} - \boldsymbol{B}_{t_o-1}\right)\boldsymbol{\theta} \approx \ln\left(1 - \mathring{i}_{x,to}\right) \tag{13}$$

where $\boldsymbol{B}_{t_o}$ and $\boldsymbol{B}_{t_o-1}$ denote the $u \times c$ sub-matrices of the B-spline matrix in (2) corresponding to the opinion time vectors $\boldsymbol{t}_o$ and $(\boldsymbol{t}_o - \boldsymbol{1})$, and $\boldsymbol{1}$ is a vector of 1's.

Hence, we have defined a Penalised Generalised Linear Model with Poisson error (1)–(2), penalty matrix $\boldsymbol{P}_\lambda$, log link, and deterministic opinions expressed in (13). The estimation process is described in Section 3.3 below.

3.2.2 Exact incorporation of mortality improvements opinions

Approximation (12) is valid only when the $\mu_{x,t_{o1}}$ and $\mu_{x,t_{o1}-1}$ are small. Therefore, the method in Section 3.2.1 might not be suitable for high mortality (e.g. the mortality of the very old). Alternatively, let us set the link function to $g(\mu) = \ln(1 - \exp(-\mu))$; that is

$$\ln\left(1 - \exp\left(-\boldsymbol{\mu}_{x,t}\right)\right) = \boldsymbol{B}_t\,\boldsymbol{\theta} \tag{14}$$

The improvement opinions in (11) can then be rearranged and expressed compactly as

$$\left(\boldsymbol{B}_{t_o} - \boldsymbol{B}_{t_o-1}\right)\boldsymbol{\theta} = \ln\left(1 - \mathring{i}_{x,t_o}\right) \tag{15}$$

Thus, we have a Penalised Generalised Linear Model with Poisson error (1)–(2), penalty matrix $\boldsymbol{P}_\lambda$, link function $g(\mu) = \ln(1 - \exp(-\mu))$, and deterministic opinions given by (15). In this case, the link function corresponds to the logarithmic transform of the mortality rates.

It is worth clarifying that although the approximate constraint equation (13) and its exact counterpart (15) have the same form, the underlying models use different link functions. An illustration of the difference arising from the two approaches will be shown in Section 4.

3.3 Fitting the model, standard errors, and extension to two dimensions

We have established that deterministic opinions can be expressed as a set of linear constraints on the regression or spline coefficients as in (12) or (15); that is,

$$\boldsymbol{C}_o\boldsymbol{\theta} = \boldsymbol{z}_o \tag{16}$$

where $\boldsymbol{C}_o$ is a matrix made up of a simple combination of the rows of the B-spline regression matrix and $\boldsymbol{z}_o$ is the joint vector of opinion inputs.

Thus, fitting the model under these opinions reduces to the optimisation of the penalised log-likelihood ℓ_P subject to the constraints (16). Many strategies have been developed for constrained optimisation problems; see for example Strang (1986), Bjørck (1996) and Currie (2013) among others. Here, we use the Lagrange multiplier method. Hence, we consider the Lagrange objective function

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\delta}) = \ell_P(\boldsymbol{\theta}) + (\boldsymbol{C}_o\boldsymbol{\theta} - \boldsymbol{z}_o)'\boldsymbol{\delta} \tag{17}$$

where $\boldsymbol{\delta}$ is the vector of Lagrange multipliers.

The value of $(\boldsymbol{\theta}, \boldsymbol{\delta})$ that maximises $\mathcal{L}$ can be obtained using the Newton–Raphson method (Strang, 1986; Currie, 2013). In particular, by adapting the result in Currie (2013), it can be shown that the value of $(\boldsymbol{\theta}, \boldsymbol{\delta})$ that maximises $\mathcal{L}$ corresponds to the solution of the following extended penalised scoring equations

$$\begin{pmatrix} \boldsymbol{B}_t' \, \widetilde{\boldsymbol{W}} \boldsymbol{B}_t + \boldsymbol{P}_\lambda & \boldsymbol{C}_o' \\ \boldsymbol{C}_o & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\theta}} \\ \widehat{\boldsymbol{\delta}} \end{pmatrix} \approx \begin{pmatrix} \boldsymbol{B}_t' \, \widetilde{\boldsymbol{W}} \widetilde{\boldsymbol{z}} \\ \boldsymbol{z}_o \end{pmatrix} \tag{18}$$

At convergence, the conditional effective dimension and conditional covariance matrix of the fitted mortality trend can be computed based on the inverse of the 2 by 2 block matrix on the left-hand side of (18). If we set $\boldsymbol{A} = \boldsymbol{B}_t' \, \widehat{\boldsymbol{W}} \boldsymbol{B}_t + \boldsymbol{P}_\lambda$, and apply the formula for the inverse of block matrices (Searle *et al.*, 2006), we obtain

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{C}_o' \\ \boldsymbol{C}_o & \boldsymbol{0} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1} \boldsymbol{C}_o' (\boldsymbol{C}_o \boldsymbol{A}^{-1} \boldsymbol{C}_o')^{-1} \boldsymbol{C}_o \boldsymbol{A}^{-1} & \boldsymbol{A}^{-1} \boldsymbol{C}_o' (\boldsymbol{C}_o \boldsymbol{A}^{-1} \boldsymbol{C}_o')^{-1} \\ (\boldsymbol{C}_o \boldsymbol{A}^{-1} \boldsymbol{C}_o')^{-1} \boldsymbol{C}_o \boldsymbol{A}^{-1} & -(\boldsymbol{C}_o \boldsymbol{A}^{-1} \boldsymbol{C}_o')^{-1} \end{pmatrix}$$

Hence, the covariance of $\hat{\boldsymbol{\theta}}$ is

$$Cov\left(\hat{\boldsymbol{\theta}} \mid \text{opinions}\right) = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1} \boldsymbol{C}_o' \left(\boldsymbol{C}_o \boldsymbol{A}^{-1} \boldsymbol{C}_o'\right)^{-1} \boldsymbol{C}_o \boldsymbol{A}^{-1} \tag{19}$$

and the effective dimension (ED) of the model is

$$\text{ED} \mid \text{opinions} = trace \left\{ \left(\boldsymbol{A}^{-1} - \boldsymbol{A}^{-1} \boldsymbol{C}_o' \left(\boldsymbol{C}_o \boldsymbol{A}^{-1} \boldsymbol{C}_o'\right)^{-1} \boldsymbol{C}_o \boldsymbol{A}^{-1}\right) \boldsymbol{B}_t' \widehat{\boldsymbol{W}} \boldsymbol{B}_t \right\} \tag{20}$$

From expression (19), the covariance matrix of the estimator of the fitted mortality curve $\boldsymbol{B}_t \hat{\boldsymbol{\theta}}$ can be computed as

$$Cov\left(\boldsymbol{B}_t \hat{\boldsymbol{\theta}} \mid \text{opinions}\right) = \boldsymbol{B}_t \, Cov\left(\hat{\boldsymbol{\theta}} \mid \text{opinions}\right) \boldsymbol{B}_t'$$
$$= \boldsymbol{B}_t \boldsymbol{A}^{-1} \boldsymbol{B}_t' - \boldsymbol{B}_t \boldsymbol{A}^{-1} \boldsymbol{C}_o' \left(\boldsymbol{C}_o \boldsymbol{A}^{-1} \boldsymbol{C}_o'\right)^{-1} \boldsymbol{C}_o \boldsymbol{A}^{-1} \boldsymbol{B}_t' \tag{21}$$

The square roots of the diagonal elements of expression (21) are the estimated standard errors of the fitted mortality trends. We shall illustrate these standard errors in Section 4.2.

In general, the deterministic opinions tend to reduce the spread of the confidence intervals around the estimated mortality curves. In particular, when the opinions are formulated on the scale of the force of mortality or mortality rates as in Section 3.1, the standard errors of the fitted mortality curves vanish at the opinion time points t_o. Indeed,

$$Cov\left(\boldsymbol{B}_{t_o} \hat{\boldsymbol{\theta}} \mid \text{opinions}\right) = \boldsymbol{B}_{t_o} \, Cov\left(\hat{\boldsymbol{\theta}} \mid \text{opinions}\right) \boldsymbol{B}_{t_o'}$$
$$= \boldsymbol{B}_{t_o} \boldsymbol{A}^{-1} \boldsymbol{B}_{t_o'} - \boldsymbol{B}_{t_o} \boldsymbol{A}^{-1} \boldsymbol{C}_o' \left(\boldsymbol{C}_o \boldsymbol{A}^{-1} \boldsymbol{C}_o'\right)^{-1} \boldsymbol{C}_o \boldsymbol{A}^{-1} \boldsymbol{B}_{t_o'} \tag{22}$$
$$= \boldsymbol{0} \quad \text{if} \quad \boldsymbol{C}_o = \boldsymbol{B}_{t_o}$$

We emphasise that the standard errors arising from (19) must be interpreted with caution because they are conditional on the opinion inputs. This is illustrated and discussed in Section 4.2.

So far we have described how to build deterministic opinions into the model in one-dimensional settings. This can be extended to two dimensions. As in Section 2.2, let us consider mortality data $\boldsymbol{E}$ and $\boldsymbol{D}$, each stored in a matrix form for ages $\boldsymbol{x} = (x_1, ..., x_{n_1})$ and calendar years $\boldsymbol{t} = (t_1, ..., t_{n_2})$. Without loss of generality, let us suppose that one is interested in modelling and forecasting mortality rates under the assumption that mortality improvements in given cells $(x_{o1}, t_{o1}), ..., (x_{ou}, t_{ou})$ are known a priori. That is, one would like to fit a mortality surface to the data such that

$$1 - \frac{q_{x_{ok}, t_{ok}}}{q_{x_{ok}, t_{ok}-1}} = \mathring{\imath}_{x_{ok}, t_{ok}} \tag{23}$$

where the $\hat{i}_{x_{ok},t_{ok}}$ are known values of the mortality improvements, and the $q_{x,t}$ are fitted mortality rates.

Note that the deterministic opinions can be formulated on different scales as in Section 3.1. Also, there is no restriction on the locations (x_{ok}, t_{ok}) of the opinion constraints in the sense that some opinion cells (x_{ok}, t_{ok}) can be within the data regions as well as outside of the data region.

If we denote by $\boldsymbol{\Theta}$ the matrix of spline coefficients as in Section 2.2, by $\boldsymbol{C}_o$ the matrix obtained by stacking the rows of the one-row matrices $\boldsymbol{B}_{t_{ok}} \otimes \boldsymbol{B}_{x_{ok}}$, $1 \leq k \leq u$, on top of each other, and we set $\boldsymbol{\theta} = vec(\boldsymbol{\Theta})$ and $\boldsymbol{z}_o = (\hat{i}_{x_{o1},t_{o1}}, \hat{i}_{x_{o2},t_{o2}}, ..., \hat{i}_{x_{ou},t_{ou}})$, then, the two-dimensional opinion (23) takes the matrix form of (16).

Hence, the fitted mortality surface encapsulating the data and the opinion inputs is obtained by solving an augmented system of iterative equations as in (18), but with $\boldsymbol{B}_t$ replaced by the two-dimensional B-splines basis $\boldsymbol{B}_t \otimes \boldsymbol{B}_x$, and $\boldsymbol{P}_\lambda$ replaced by the penalty matrix $\boldsymbol{P}_{\lambda_1,\lambda_2}$ defined in (8).

Furthermore, the conditional covariance matrix and effective dimension of the fitted mortality surface can be computed as in (11), but, with $\boldsymbol{B}_t$ and $\boldsymbol{P}_\lambda$ substituted by $\boldsymbol{B}_t \otimes \boldsymbol{B}_x$ and $\boldsymbol{P}_{\lambda_1,\lambda_2}$.

We close this section by noting that setting very dissimilar opinion inputs in adjacent ages/years can yield an unsmooth mortality surface, especially in the vicinity of the opinion locations. Also, setting a large number of opinion inputs (e.g. opinions at every single age/year) can cause singularities in the extended penalised scoring equation (18). These problems can be tackled by adding a small ridge penalty to the penalty matrix (Hoerl *et al.*, 1970), by selecting a subset of the opinion inputs to work with, or by increasing the number of B-splines.

3.4 Similarities and differences with CMI projection approach

The Continuous Mortality Investigation (CMI) carries out research into mortality and morbidity experience on behalf of the Institute and Faculty of Actuaries, and produces practical tools for mortality projection that are widely used in the insurance industry to support the pricing and valuation of pension and annuity business. The projection methodology used by CMI has evolved over time, taking into account relevant literature on the mechanisms that determine ageing and longevity, including empirical features such as cohort effects (Willets, 2004), as well as latest work on mortality forecasting methods.

In the most recent CMI projection models (CMI, 2014; CMI, 2016; CMI, 2020), the basic approach is to project mortality improvement rates by interpolating between *current improvement rates* and some assumed *long-term improvement rates*. The current improvement rates are estimated from historical data whereas the long-term improvement rates are set by users of the CMI model. This interpolation process is carried out separately for the age period and cohort components, and these components are then summed to give the overall mortality improvements. In this process, the shape of the projected rates is driven by a large number of parameters, including the initial direction of travel of mortality improvements, the convergence period, and the proportion of the remaining improvements at the midpoint. The users of the tools can then control projected mortality improvement patterns by playing with these parameters.

Some of features of the CMI approach can be found in the method presented in this paper. In particular, the initial and long-term improvement rates in the CMI approach can be fed into our framework as opinion inputs; the age-dependent convergence period can also be specified. By default, the initial direction of travel and the proportion of the remaining improvements at the midpoint are controlled by the smoothing process. However, the users of our approach can alter these defaults. For example, let us denote by t_0 the initial calendar year, by T_x the age-dependent convergence period, by i_{x,t_0} the initial improvement rates, and by i_{x,t_0+T_x} the long-term

improvement rates. A proportion of remaining improvements of a% at midpoint is achieved within our framework by setting the deterministic input in equation (12) to

$$\mathring{i}_{x,t_o+\frac{T_x}{2}} = i_{x,t_0+T_x} + \frac{a\left(i_{x,t_0} - i_{x,t_0+T_x}\right)}{100} \tag{24}$$

A major criticism of the CMI projection methodology is that the shape of the forecast is subjective and comes without any uncertainty measure. The methodology presented in this paper (i) combines smooth patterns from the data together with the opinion inputs to derive forecasts and (ii) also allow us to compute the amount of uncertainty around the forecasts conditional on the input opinions. Moreover, unlike the CMI model, the approach presented in this study allows us to specify opinions not only in terms of mortality improvements, but also in terms of other mortality metrics by ages and calendar years. In the application Section 4.1 for instance, we shall specify opinion directly on mortality rates as well as on mortality improvements.

4. Applications

For illustrations, we use mortality data for UK males, ages 50–95 and calendar years 1970–2010 from the Human Mortality Database. We start by fitting and forecasting mortality in age and time without opinion constraints. We use cubic B-splines with 5-year equi-spaced knots in age and time, and apply second-order difference penalties to achieve smoothing. In the time direction, we project 35 years into the future, and in the age direction, we project from ages 95 to 105. The output of the fitted model is shown in Figure 1. This is broadly as expected: increasing mortality rates with age, and mortality reduction over time.

4.1 Some scenarios involving opinion inputs

We now turn to the model involving opinion inputs, and we consider two scenarios. The first scenario states opinion inputs in terms of mortality improvement rates whereas the second expresses opinions in terms of mortality rates.

4.1.1 Scenario 1

In this first scenario, we set opinions about mortality improvements rates according to equation (25). In order words, we assume that the mortality improvement rate from age 50 to 85 in 2029 is set to 1.5%, and then decreases linearly from that value of 1.5% at age 85 down to 0.6% by age 95. A similar pattern is used to set long-term improvements throughout the CMI models.

$$\mathring{i}_{x,2029} = \begin{cases} 1.5\% & \text{if } 50 \leq x \leq 85 \\ \dfrac{(95-x) \times 1.5\% + (x-85) \times 0.6\%}{10} & \text{if } 85 < x \leq 95 \end{cases} \tag{25}$$

We then incorporate opinions (25) into the model using the exact method presented in Section 3.2.2.

A comparison of the outputs from the model encapsulating these options is shown in Figure 2, against the output of the model without deterministic opinions shown in Figure 1. In particular, the lower panel on the right-hand side shows the convergence of the fitted mortality improvement rates towards the deterministic opinion pattern specified in equation (25). Comparing this panel to the corresponding panel in Figure 1 illustrates the impact of the opinion inputs (25) on the fitted mortality improvement rates. Similarly, the resulting impact in terms of mortality rates is visualised by comparing the upper panels of Figures 1 and 2.
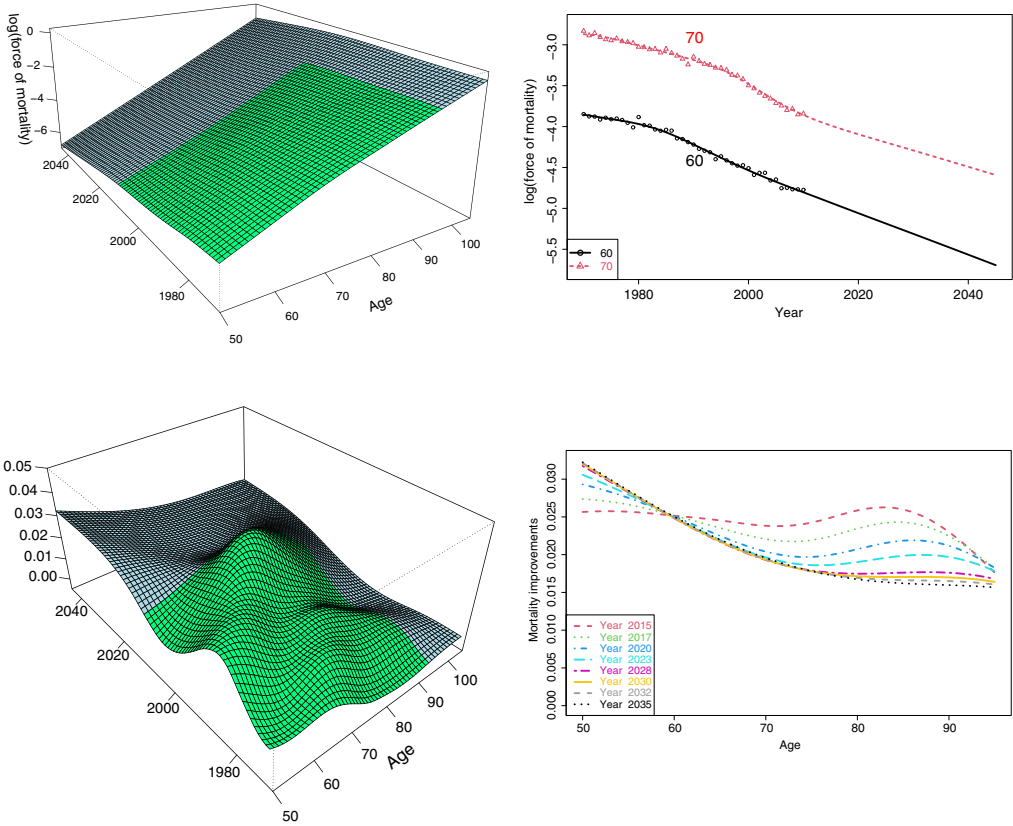
Figure 1. Output from the model without opinion constraints. Upper panels: fitted and projected mortality rates. Lower panels: mortality improvements.

4.1.2 Scenario 2

In a second scenario, we set opinion inputs in terms of mortality rates according to Table 1. These rates were obtained by projecting the "NLT16-18 (E&W)" mortality base table through the CMI model with core values specifications and a long-term mortality improvement rate of 1.5%.

The output from the model encapsulating the deterministic opinions in Table 1 is shown in Figure 3. By contrasting the panels of this figure against corresponding panels in Figure 1 and Figure 2, we can see how various opinion inputs affect projections. In particular, the lower panel on the right in Figure 3 shows an overall upward pattern of improvement rates resulting from the fact that the improvement opinion rates at age 60 (in Table 1) are lower compared to those around the same age from the model without opinion constraint (shown in Figure 1).

4.2 Conditional uncertainty

An essential aid to the interpretation of estimated mortality trends is the uncertainty. The deterministic opinion constraints affect not only the fitted and projected mortality surface, but also its standard error as well.

In the one-dimensional case as in Section 3.3, the variances of the fitted mortality curve, $g(\hat{\mu}_t)$, are the diagonal elements of the covariance matrix (21), where g represents the link function used throughout Section 3. The computation is identical in the two-dimensional case, except that the B-spline matrix $\boldsymbol{B}_t$ is substituted by its two-dimensional counterpart as described in Section 3.3.

Table 1. Opinion inputs about mortality rates for scenario 2.

| Age	Calendar year	Mortality rate
60	2029	0.007
60	2040	0.006
70	2029	0.015
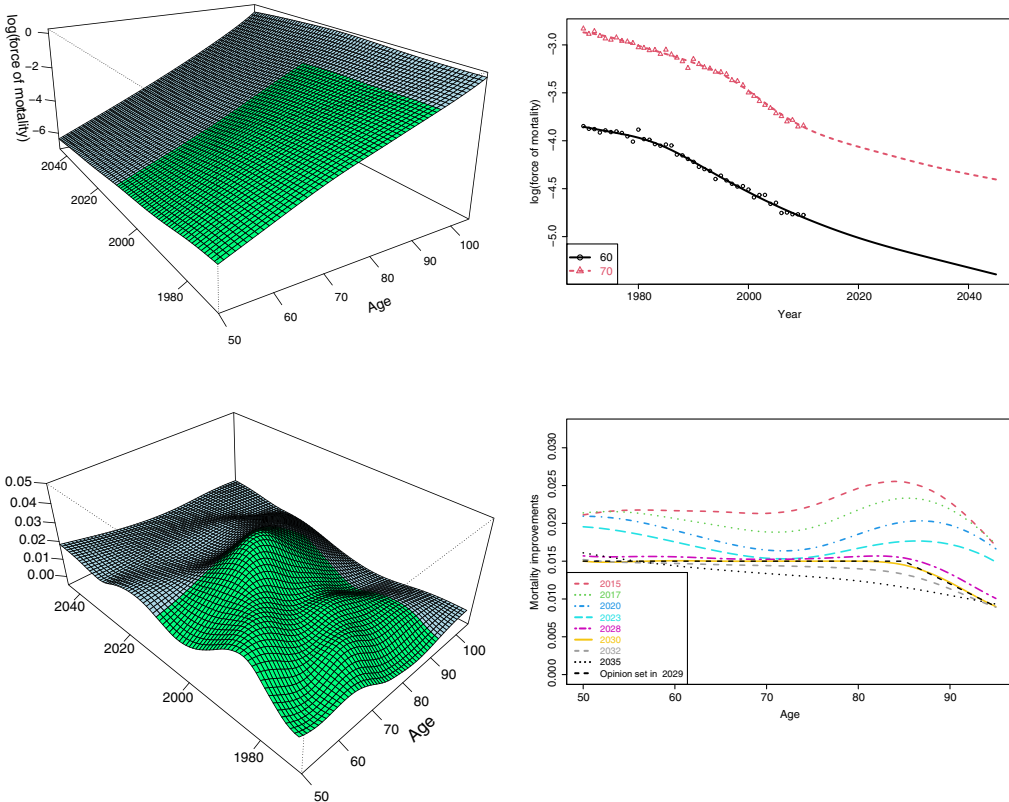
Figure 2. Outputs from the model with opinions on improvements according to equation (25). Upper panels: fitted and projected force of mortality. Lower panels: mortality improvements.

In general, incorporation of opinion constraints tend to reduce the spread of the conditional standard error around the projected mortality curves; this is illustrated in Figure 4. Many comments can be made about this figure.

First, the panel on the top-left shows that the standard error about the projected trends from the model without opinion constraints increases over time, as expected: as we travel further into the future, the projected trends become more uncertain.

Second, the top-right panel on the shows increasing standard errors during the early years of projection and then a slowdown and flattening of these standard errors due to the incorporation of opinions constraints on mortality improvements in calendar year 2029.

Third, this slowdown of standard errors is also seen in the bottom panel. In particular, this panel shows that imposing deterministic opinion constraints on mortality rates implies that one is claiming to know the exact values of mortality rates at the opinion locations, and as a result,
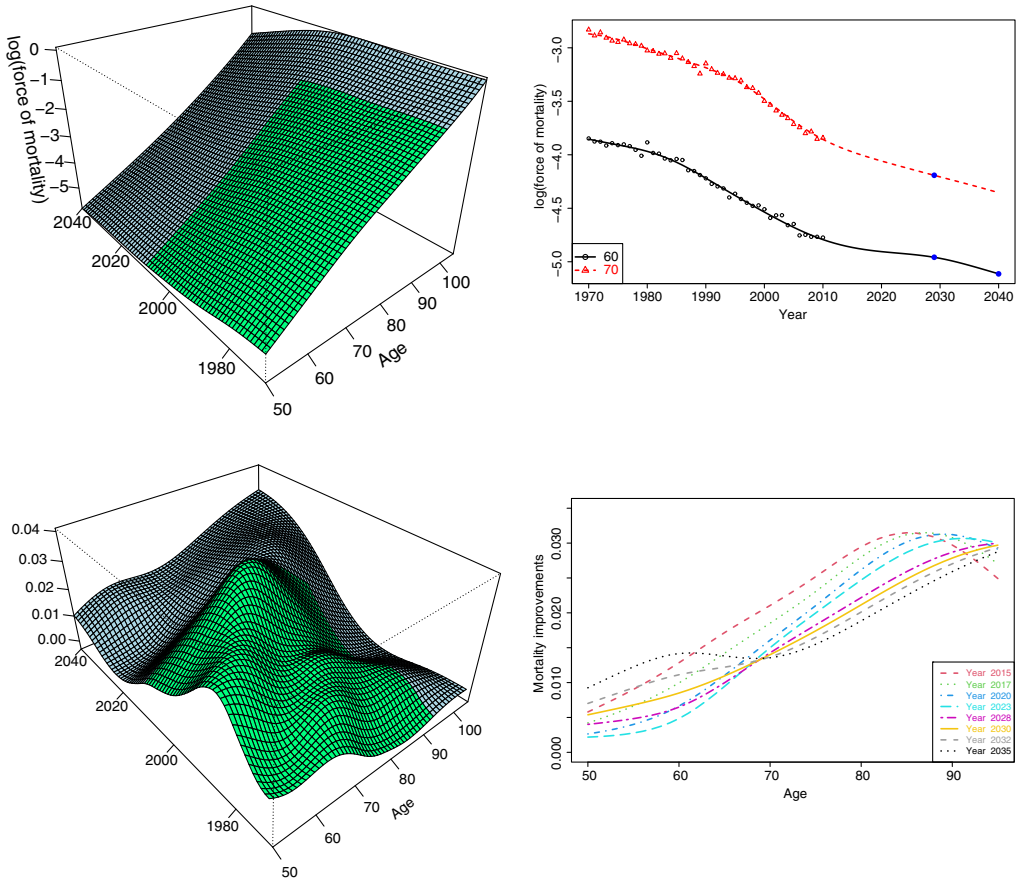
Figure 3. Outputs from the model with opinions on mortality rates according to scenario 2. Upper panels: fitted and projected force of mortality; the blue dots correspond to the opinion inputs. Lower panels: mortality improvements.

the standard error around the projected mortality rates decrease and vanishes as we approach the opinion locations.

In general, imposing deterministic opinion constraints on a given mortality metric causes standard error about the fitted/projected values of that metric to reduce and to vanish at the opinion locations. However, the standard errors in respect of other resulting mortality metrics do not necessarily vanish. For example, the top-right panel in Figure 4 reveals that, although deterministic constraints are placed on mortality improvements, there are still some amounts of uncertainty around the fitted mortality rates at the opinion locations. This can be ascribed to the fact that the value of the mortality improvement rate in a given year is driven by a combination of mortality rates in two successive years.
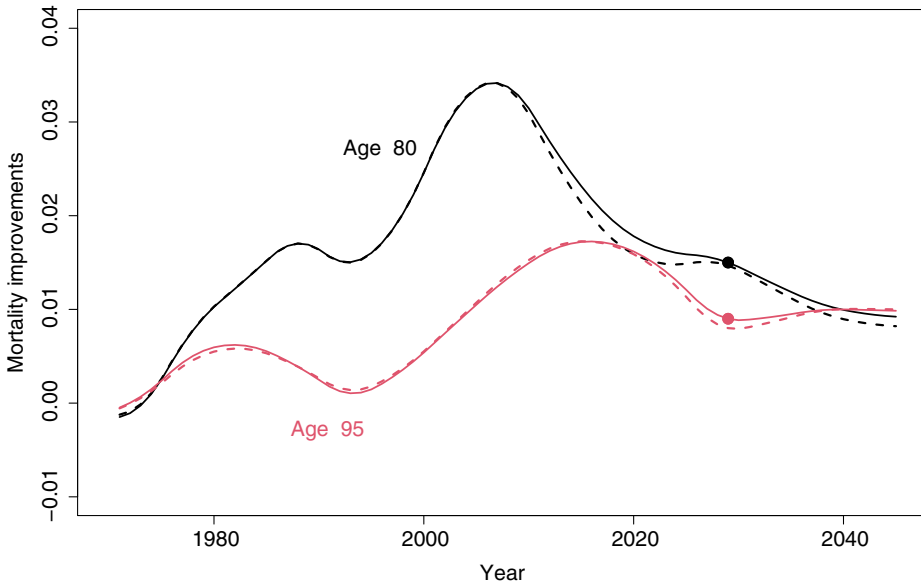
4.3 Exact method versus approximate method

In Section 3.2, we presented two ways to integrate opinions on the scale of mortality improvements. In practice, the difference between these two approaches is relatively small. For illustration let us reconsider the opinion inputs in Scenario 1 – See Equation (25).

An illustration of the difference between the exact method and the approximate method based on the opinion specification of scenario 1 is shown in Figure 5. On this graphic, the continuous lines correspond to the exact method and the dashed lines represent the approximate
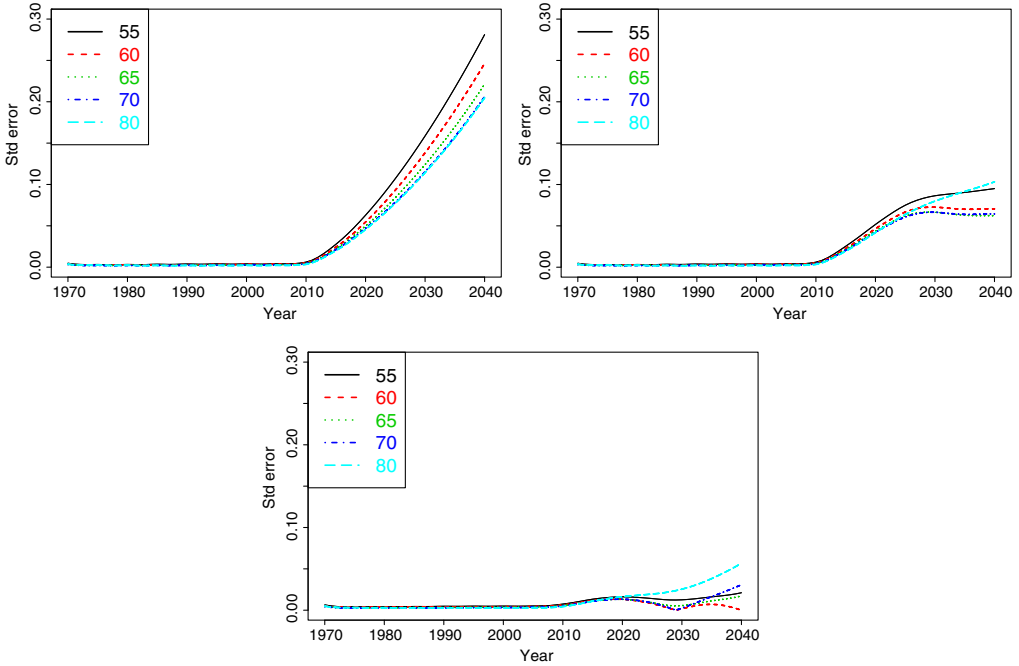
Figure 4. Illustration of the uncertainty around log $(\hat{\mu}_{x,t})$ arising from fitting models to UK mortality data for males from calendar year 1970–2010. Top left: model without opinion constraints. Top right: model with opinion inputs on mortality improvements – See equation (25). Bottom: model subject to opinion constraints on mortality rates – See Scenario 2 in Table 1.

Figure 5. Illustration of the difference between the exact method (continuous lines) and its approximate counterpart (dashed lines), from model fitted under Scenario 1. The dots represent the opinion inputs in 2029.
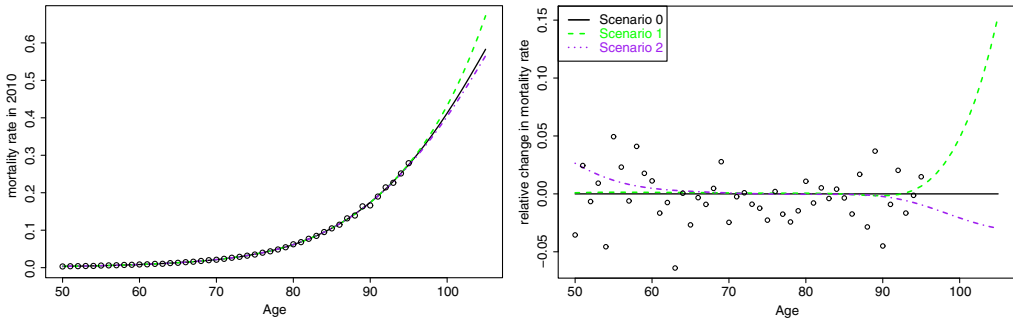
Figure 6. Illustration of the impact of opinion inputs at the edge of the data region (i.e. calendar year 2010) under the deterministic opinions in scenarios 1 and 2. Scenario 0 refers to the model without opinion inputs. The dots represent the observed data in 2010. Left-hand panel: fitted mortality rates. Right-hand panel: change in the fitted mortality rates relative to Scenario 0.

method. Although the exact method appears to be more accurate at hitting the targets, this graphic indicates that the difference between the two methods is relatively small. Nonetheless, the approximate method is slightly easier to implement because it uses the Poisson canonical link function.

4.4 Impact of deterministic opinion on the edges

Opinion input about future mortality rates can potentially impact the in-sampling smoothing of historical data, especially towards the edge of the data region. In this section, we illustrate the magnitude of such influence using the two opinion scenarios shown in Section 4. Under each scenario, we compared the fitted mortality rates at the edge of the data (that is ages 50–95 in calendar year 2010) to the fitted mortality rates from the model without opinions. This comparison is shown on graphics in Figure 6.

These graphs highlight that the impact of the opinion inputs on the in-sampling smoothed surface is relatively minor under the two scenarios considered (except perhaps at ages 50–60 for scenario 2). A further investigation using more extreme opinion specifications yielded a similar conclusion. The choice of the smoothing parameters plays a role here. On the one hand, large values of the smoothing parameters would tend to yield a smoother mortality surface and this could increase the remote impact of opinion inputs on the in-sampling smoothing. On the other hand, lower smoothing parameters increase the roughness/flexibility of the fitted mortality surface and, therefore, tend to reduce the remote impact of the opinion inputs. In our experience, selecting smoothing parameters through BIC adjusted for overdispersion allows us to reduce the remote impact of the opinion inputs. Nonetheless, it is good to bear in mind that a very extreme opinion specification near the edge of the data could yield a larger impact especially on the edges.

5. Concluding remarks

The main objective of this work was to present a simple way of integrating deterministic opinions into flexible mortality projection models. This has been achieved by expressing the opinion inputs as a system of constraints, and building them into the model using the standard machinery of iterative weighted least squares. This integrated approach addresses many limitations of current deterministic projection methods. In particular, the fitted and projected mortality trends arising from our method are driven by a combination of the speed of improvements from the data and the opinion inputs. Additionally, our approach provides a statement of conditional uncertainty around the mortality trends.

In this paper, we have focussed on opinion inputs expressed in terms of the actual value of widely used mortality metrics. Other types of constraint can be considered. For example, incorporating opinion constraints on the gradient of mortality rates has the potential to address the problem of crossing over of mortality forecasts at adjacent ages often found in some mortality models. Also, extending the work presented in this paper to account for non-deterministic opinion will be a valuable addition to the topic of mortality forecasting.

References

Andreev, **K.F.** & **Vaupel**, **J.W.** (2006). Forecasts of Cohort Mortality after Age 50. Max Planck Institute for Demographic Research.

Biatat, **V.** & **Currie**, **I.D.** (2010). Joint models for classification and comparison of mortality in different countries. In *Proceedings of 25rd International Workshop on Statistical Modelling* (pp. 89–94).

Bjõrck, **A.** (1996). Numerical Methods for Least Squares Problems. Society for Industrial and Applied Mathematics.

Brouhns, **N.**, **Denuit**, **M.** & **Vermunt**, **J.** (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, **31**, 373–393.

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

Cairns, **A.J.G.**, **Blake**, **D.**, **Dowd**, **K.**, **Coughlan**, **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. *North American Actuarial Journal*, **13**, 1–35.

Cairns, **A.J.G.** (2017). A Flexible and Robust Approach to Modelling Single Population Mortality. *Longevity* **13**, Taipei, September 2017.

Carballoa, **A.**, **Durbán**, **M.** & **Lee**, **D.J.** (2017). A general framework for prediction in penalized regression. Universidad Carlos III de Madrid Working Papers.

Continuous Mortality Investigation (2006). Working Paper 38. A Prototype Mortality Projections Model: Part One - An Outline of the Proposed Approach.

Continuous Mortality Investigation Bureau (2009). Working Papers 39. A Prototype Mortality Projections Model, Part Two - Detailed Analysis.

Continuous Mortality Investigation (2014). Working Paper 74. The CMI Mortality Projections Model, CMI2014.

Continuous Mortality Investigation (2016). Working Paper 91. CMI Mortality Projections Model consultation - technical paper.

Continuous Mortality Investigation (2020). Working Paper 129: "CMI Mortality Projections Model: CMI 2019" (2020)

Currie, **I.D.** (2015). Smoothing constrained GLMs and improving Lee-Carter Forecasts. Universidad Carlos III de Madrid, May 2015.

Currie, **I.D.**, **Durban**, **M.** & **Eilers**, **P.H.C.** (2004). Smoothing and forecasting mortality rates. *Statistical Modelling*, **4**, 279–298.

Currie, **I.D.**, **Durban**, **M.** & **Eilers**, **P.H.C.** (2006). Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society (Series B)*, **68**, 259–280.

Currie, **I.** (2013). Smoothing constrained generalized linear models with an application to the Lee-Carter model. *Statistical Modelling*, **13**, 69–93.

Debón, **A.**, **Martnez-Ruiz**, **F.** & **Montes**, **F.** (2010). A geostatistical approach for dynamic life tables: the effect of mortality on remaining lifetime and annuities. *Insurance: Mathematics and Economics*, **47**, 327–336.

Debón, **A.**, **Montes**, **F.** & **Sala**, **R.** (2006). A comparison of nonparametric methods in the graduation of mortality: application to data from the Valencia region (Spain). *International Statistical Review*. Available online at the address https://doi.org/10.1111/j.1751-5823.2006.tb00171.x.

Delwarde, **A.**, **Denuit**, **M.** & **Eilers**, **P.H.C.** (2006). Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting: a penalised likelihood approach. *Statistical Modelling*, **7**, 29–48.

Djeundje, **V.A.B.** & **Currie**, **I.D.** (2010). Smoothing dispersed counts with applications to mortality data. *Annals of Actuarial Science*, **5**, 33–52.

Djeundje, **V.A.B.** (2011). *Hierarchical and Multidimensional Smoothing with Applications to Longitudinal and Mortality Data*. Heriot-Watt University, UK. Willey.

Eilers, **P.H.C.** & **Marx**, **B.D.** (1996). Flexible smoothing with B-splines and penalties. *Statistical Sciences*, **11**, 89–121.

Eilers, P.H.C. & **Marx**, B.D. (2010). Splines, knots, and penalties. *Computational Statistics*, **2**, 637–653.

French, D. and **O'Hare**, C. (2014). Forecasting death rates using exogenous determinants. *Journal of Forecasting*, **33**, 640–650.

Hoerl, A.E. & **Kennard**, R.W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, **12**, 55–67.

Human Mortality Database. University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany). Available online at the address www.mortality.org.

Janssen, F. & **Kunst**, A. (2007). The choice among past trends as a basis for the prediction of future trends in old-age mortality. *Popul Stud (Camb)*, **61**, 315–326.

Janssen, F., **Van Wissen**, L. & **Kunst**, A. (2013). Including the smoking epidemic in internationally coherent mortality projections. *Demography*, **50**, 1341–1362.

Lee, D.J. & **Durban**, M. (2011). P-spline ANOVA-type interaction models for spatio-temporal smoothing. *Statistical Modelling*, **11**, 49–69.

Lee, R.D. & **Carter**, L.R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, **87**, 659–671.

Ludkovski, M., **Risk**, J. & **Zail**, H. (2018). Gaussian process models for mortality rates and improvement factors. *Astin Bulletin*, **48**, 1307–1347.

McCullagh, P. & **Nelder**, J.A. (1989). *Generalized Linear Models*. Chapman and Hall.

O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems (with discussion). *Statistical Sciences*, **1**, 505–527.

Peristera, P. & **Kostaki**, A. (2005). An evaluation of the performance of kernel estimators for graduating mortality data. *Journal of Population Research*, **22**, 185–197.

Plat, R. (1986). On stochastic mortality modeling. *Insurance: Mathematics and Economics*, **45**, 393–404.

Preston, S.H., **Stokes**, A., **Mehta**, N.K. & **Cao**, B. (2014). Projecting the effect of changes in smoking and obesity on future life expectancy in the United States. *Demography*, **51**, 27–49.

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

Richards, S.J. (2009). Stabilising projections. Available online at the address https://www.longevitas.co.uk/site/information matrix/stabilisingprojections.html4

Richards, S.J., **Currie**, I.D., **Kleinow**, T. & **Ritchie**, G.P. (2017). A stochastic implementation of the APCI model for mortality projections. Actuarial Research Centre (ARC).

Richards, S.J., **Currie**, I.D. & **Ritchie**, G.P. (2014). A Value-at-Risk framework for longevity trend risk. *British Actuarial Journal*, **19**, 116–139.

Searle, S.R., **Casella**, G. & **McCulloch**, C.E. (2006). *Variance Components*. Willey.

Stoeldraijer, L., **van Duin**, C., **van Wissen**, L. & **Janssen**, F. (2013). Impact of different mortality forecasting methods and explicit assumptions on projected future life expectancy: the case of the Netherlands. *Demographic Research*, **29**, 323–354.

Strang, G. (1986). *Introduction to Applied Mathematics* (pp. 96–107). Wellesley-Cambridge Press.

Willets, R.C. (1999). Mortality in the next millennium. Staple Inn Actuarial Society.

Willets, R. (2004). The cohort effect: insights and explanations. *British Actuarial Journal*, **10**, 833–877.

Woods, S. (2016). Reflections on the 2015 Solvency II internal model approval process. Bank of England Prudential Regulation Authority.