Smooth projection of mortality improvement rates: a Bayesian two-dimensional spline approach

Xiaobai Zhu¹ · Kenneth Q. Zhou²

Received: 19 January 2022 / Revised: 25 April 2022 / Accepted: 14 June 2022 / Published online: 27 June 2022
© EAJ Association 2022

Abstract
This paper proposes a spline mortality model for generating smooth projections of mortality improvement rates. In particular, we follow the two-dimensional cubic B-spline approach developed by Currie et al. (Stat Model 4(4):279–298, 2004), and adopt the Bayesian estimation and LASSO penalty to overcome the limitations of spline models in forecasting mortality rates. The resulting Bayesian spline model not only provides measures of stochastic and parameter uncertainties, but also allows external opinions on future mortality to be consistently incorporated. The mortality improvement rates projected by the proposed model are smoothly transitioned from the historical values with short-term trends shown in recent observations to the long-term terminal rates suggested by external opinions. Our technical work is complemented by numerical illustrations that use real mortality data and external rates to showcase the features of the proposed model.

Keywords Spline mortality modeling · Bayesian estimation · Stochastic mortality forecasting · LASSO penalty · External information · The MIM-2021 model

1 Introduction

1.1 Background

Due to the steady increases in life expectancy in recent decades, incorporating future mortality improvements in actuarial pricing and valuation has become an increasingly important task for actuaries. For instance, life and pension actuaries often need to use improvement scales to adjust their baseline life table to account for the foreseen deductions in future mortality rates. To facilitate this practical
need, the actuarial professions in different countries have been actively exploring new methodologies for generating mortality improvement (MI) scales.

Historically, MI scales are made using simple one-dimensional formulas, for example, the AA scale of the Society of Actuaries (SOA), in which only the effect of age is considered. In 2014, following the conceptual framework developed by the Continuous Mortality Investigation (CMI) of the Institute and Faculty of Actuaries (IFoA) in the UK, the Retirement Plans Experience Committee (RPEC) of the SOA released the RPEC model to address the issues of Scale AA. In particular, the RPEC model exploits expert opinions on long-term mortality improvement rates to produce arrays of two-dimensional (age and time) MI scales. The resulting MI rates converge gradually from recent observations to the long-term externally given values. The RPEC model was updated annually with Scale MP-2020 being the most recent one of the ‘MP’ series. In addition, other actuarial professions have also developed MI scales for their own populations, for example, the MI-2017 scale and the CPM-B scale from the Canadian Institute of Actuaries (CIA).

More recently, the SOA promulgated a new methodology for generating two-dimensional MI scales, called the MIM-2021 model. Compared to its predecessor (the RPEC model), several new features were added. First, the MIM-2021 model uses historical mortality data from not only the Social Security Administration (SSA) of the US, but also the National Center for Health Statistics (NCHS), where the dataset is stratified into socioeconomic categories. Second, the MIM-2021 model allows the user to externally specify an optional set of intermediate-term MI rates, in addition to the long-term ultimate ones, so that a more flexible setting of MI scales can be achieved. Third, using polynomial functions, the resulting MI scales are smoothly transitioned from recently observed values to the intermediate ones and then ultimately to the long-term rates.

Despite being an improved model for generating MI scales, the MIM-2021 model is subject to several limitations. First, the model’s graduation and projection methods are not the same. To have smooth mortality data, the Whittaker–Henderson graduation method is used on the historical mortality rates, whereas the method for projecting future MI rates is based on two polynomial functions that are separately applied to the age and cohort dimensions. Second, the polynomial-based projection method is not flexible. Although the model user can specify intermediate- and long-term values, the rate of transition between different periods cannot be controlled. Third, due to the deterministic projection method, the MIM-2021 model can only produce point estimates for future MI rates, but not provide measures of uncertainty surrounding the projected values. Moreover, the externally given rates are also treated as deterministic inputs with no measure of uncertainty.

The research objective of this paper is to develop a mortality improvement model that carries the features of the MIM-2021 model and also overcomes its limitations. In particular, we aim to build a mortality model that can (1) statistically incorporate external opinions on intermediate- and long-term MI rates, (2) smoothly project MI rates from recent observations to external rates on a two-dimensional surface, and (3) provide measures of uncertainty for the projected MI rates and the externally given rates.
1.2 Literature review

Broadly speaking, there are two types of approaches that can be adopted to achieve our research objective. The first one models MI rates using factor-based mortality models, such as the Lee–Carter model [27] or the Cairns–Blake–Dowd model [3]. For example, Li et al. [30] used the Lee–Carter model to generate MI rates for Canadian insured lives. Haberman and Renshaw [19] considered using the Lee–Carter model to forecast two-dimensional MI rates. Following the work of Hyndman and Ullah [25], Shang [41] proposed to use dynamic principal component regression to forecast MI rates. More recently, to generate two-dimensional MI scales, Li and Liu [31] proposed the heat-wave model, which uses normal distributions to describe the short-term pattern in MI rates. Other examples include those considered by Haberman and Renshaw [20], Hunt and Villegas [24], Renshaw and Haberman [38] and Li and Liu [32].

This first type of approaches is subject to two major issues that are hindering the research objective. First, the MI rates projected by these factor-based models are likely to be jagged across ages and/or cohorts, due to the non-smooth factor estimates. To overcome this problem, studies, such as Renshaw and Haberman [39], Delwarde et al. [10], Currie [7], Pitt et al. [36] and Li et al. [33], have considered using penalization methods or splines to smooth the age-specific factors. Nevertheless, such smoothing methods can hardly be applied simultaneously to both age- and period-specific factors to achieve a smooth two-dimensional MI surface.

Second, the factor-based models often suggest that the high level of mortality improvements observed in recent decades will continue indefinitely into the future. This conjecture is not aligned with the industry’s view that MI rates should gradually converge from recent observations to certain long-term terminal values. More importantly, it is not clear how external opinions on long-term rates can be statistically incorporated in factor-based models. Dowd et al. [14] proposed to deliberately alter the drift values of the period-specific factors. Dodd et al. [13] considered using a weighted average method to blend the external rates and internal results. Moreover, although factor-based models can provide stochastic uncertainty to the projected rates, uncertainty surrounding the external rates cannot be easily measured.

The second type of approaches focuses on using non-parametric methods, such as applying spline functions to directly model and project mortality rates. Under the roof of generalized additive models, Dodd et al. [12] and Hilton et al. [21] considered using one-dimensional P-splines to smooth mortality rates. On two-dimensional modeling, Currie et al. [8] developed a two-dimensional B-spline mortality model. Extending the work of Currie et al. [8], Luoma et al. [34] considered natural Bayesian splines, Camarda et al. [5] used P-splines, while Camarda [4] proposed to use constrained P-splines instead. Other studies that also considered non-parametric methods include Currie et al. [9], Huang and Browne [23], Richards [40] and Tang et al. [43]. Lastly, we remark that the Whittaker–Henderson graduation method used in the MIM-2021 model belongs to this type of approaches.
1.3 Our contributions

In this paper, we focus on the two-dimensional setting of spline mortality modeling. In particular, we follow the seminal work of Currie et al. [8] to construct a Bayesian two-dimensional B-spline model for smoothing and projecting MI rates. To achieve our research objective and also overcome some limitations of spline mortality modeling, we develop our spline MI model with several adaptations, which are outlined as follows.

Firstly, a two-dimensional B-spline can naturally provide a smooth surface of MI rates. However, to reach an appropriate balance between smoothness and goodness-of-fit, a penalty term needs to be added to the likelihood function, which in turn leads to a P-spline. When applied to mortality modeling, a P-spline model is often criticized for its subjective knots, smoothing parameter and limited forecast ability. As discussed in Currie et al. [8], the order of penalty needs to be subjectively determined and has dominant effects on the forecast results. Setting the order of penalty to 1, 2 and 3 is equivalent to assuming no future MI, linear MI and quadratic MI, respectively. To address this problem, we implement LASSO penalty to spline mortality modeling by adopting the method of Park and Casella [35]. Compared to the order-based penalties, the LASSO penalty is able to provide a more robust and flexible setting for projecting MI rates, such as capturing short-term mortality shocks and allowing terminal long-term rates. For autoregressive mortality models, the use of LASSO penalty in parameter estimation has been considered by studies, such as Li and Lu [28], Chang and Shi [6] and Li and Shi [29].

Secondly, to have a joint estimation and projection framework, we adopt the Bayesian method to estimate our model and project MI rates. The idea of applying spline models under the Bayesian method is well explored in the statistical literature (see, e.g.,[11, 26, 42]). Adopting the existing statistical work, the Bayesian method allows us to obtain a joint posterior distribution of the entire MI surface, including both the estimated historical rates and the projected future rates. The resulting predictive distribution will appropriately reflect the extent of stochastic uncertainties in future MI rates. Furthermore, the Bayesian results provide an avenue for us to analyze the different levels of uncertainties, originated from either parameter estimation, external opinions or random noises.

Thirdly, we modify the Bayesian B-spline projection method to develop a straightforward approach of incorporating external opinions on future MI rates. This approach treats any externally given intermediate-term and long-term MI rates as a reliable source of “additional data”. Multiple years of external rates and multiple opinions on the same year-age cell are allowed. This feature is compatible with the industry’s practice, where distinct expert opinions on future MI rates might be collected in a survey. Under the Bayesian method, the uncertainty resulting from different external opinions can be properly reflected by the posterior distribution of the external rates.

Lastly, the technical contributions of this paper are complemented by a numerical illustration that showcases how the proposed model can be used to generate a smooth surface of MI rates using real mortality data and external opinions. The projected MI rates are smoothly transitioned from the short-term patterns shown in recent mortality
observations to the externally given intermediate- and long-term rates. The mean forecasts are accompanied by measures of uncertainties generated from different sources. To demonstrate the flexibility of the proposed model on incorporating external opinions, we further consider other scenarios where multiple years or multiple input of external rates are provided.

The rest of this paper is organized as follows. Section 2 provides the setup of our modeling approach, including a brief review of the MIM-2021 model. Section 3 outlines the proposed spline MI model and the technical details. Section 4 discusses the Bayesian estimation and projection methods developed for our model. Section 5 provides a numerical illustration using real data and external rates. Lastly, Sect. 6 concludes the paper with limitations and possible future work.

2 The setup

In this section, we first set up the notation to be used in this paper and provide a brief review of the MIM-2021 model. We then discuss the features and limitations of the MIM-2021 model in detail, and our motivations and objectives for this study.

2.1 Notation

Let $m_{x,t}$ denote the central death rate for age $x \in \{x_1, \ldots, x_{n_a}\}$ in year $t \in \{t_1, \ldots, t_{n_y}\}$, where $n_a$ and $n_y$ are the total number of ages and years in the mortality dataset, respectively. The mortality improvement (MI) rate for age $x$ in year $t$ is defined as

$$y_{x,t} = 1 - \frac{m_{x,t}}{m_{x,t-1}}.$$  

We can interpret $y_{x,t}$ as the relative change in the central death rates between year $t$ and year $t - 1$ for age $x$. A positive value of $y_{x,t}$ indicates that mortality improved from year $t - 1$ and year $t$ (i.e., mortality rate decreased), whereas a negative value of $y_{x,t}$ indicates a year of mortality deterioration.

2.2 The MIM-2021 model

Before applying cubic polynomial functions to project future MI rates, the MIM-2021 model first uses the Whittaker–Henderson method to graduate the historical values of $m_{x,t}$ to obtain the smoothed rates, denoted as $m_{x,t}^s$. In particular, the following objective function is minimized in order to smooth the central death rates:

$$\sum_{t=t_1}^{t_{n_y}} \sum_{x=x_1}^{x_{n_a}} (\ln m_{x,t}^s - \ln m_{x,t})^2 + v_a \sum_{t=t_1}^{t_{n_y}} \sum_{x=x_1+3}^{x_{n_a}} (\Delta^3_x (\ln m_{x,t}^s))^2 + v_y \sum_{t=t_1+3}^{t_{n_y}} \sum_{x=x_1}^{x_{n_a}} (\Delta^3_t (\ln m_{x,t}^s))^2.$$  

(1)

The parameters $v_a$ and $v_y$ are penalty weights that balance the closeness of $m_{x,t}^s$ to $m_{x,t}$ and the smoothness of $m_{x,t}^s$. The functions $\Delta^3_x(\cdot)$ and $\Delta^3_t(\cdot)$ are the third-order
difference along the age and time dimensions, respectively. The MIM-2021 model uses a value of $\nu_y = 100$ over the time dimension, and a value of $\nu_a = 400$ over the age dimension.

Based on the smoothed values $m_{x,t}$, the smooth MI rates are obtained from $y_{x,t} = 1 - \frac{m_{x,t}}{m_{x,t-1}}$. To project rates beyond year $t_{n_y}$ (the last year of the dataset), the MIM-2021 model supposes that the MI rates $n_p$ years after year $t_{n_y}$ are externally given; that is, the MI rates for all ages in year $t_{n_y} + n_p$ are externally determined by the model user. The MIM-2021 model then uses two cubic polynomial functions, one along the age dimension and the other one along the cohort dimension, to interpolate MI rates for years between $t_{n_y}$ and $t_{n_y} + n_p$. For each age $x$, the cubic polynomial over the time dimension is specified as

$$
\begin{align*}
\gamma_{x,t}^{\text{age}} &= y_{x,t_{n_y}} + l_x (t - t_{n_y}) - \left( \frac{2n_p l_x - 3 (y_{x,t_{n_y}+n_p} - y_{x,t_{n_y}})}{n_p^2} \right) (t - t_{n_y})^2 \\
&\quad + \left( \frac{n_p l_x - 2 (y_{x,t_{n_y}+n_p} - y_{x,t_{n_y}})}{n_p^3} \right) (t - t_{n_y})^3,
\end{align*}
$$

where $l_x$ is an initial slope externally specified by the model user, or set to $y_{x,t_{n_y}} - y_{x,t_{n_y}-1}$ as the default value. It is further assumed that the ending slope at year $t_{n_y} + n_p$ is zero for all ages. Along the cohort dimension, a similar but more complicated cubic polynomial is used to obtain $y_{x,t}^{\text{cohort}}$. We omit the function here.

The MIM-2021 model is a two-dimensional model in the sense that it separately interpolates MI rates on the age and cohort dimensions. To obtain $\gamma_{x,t}^{\text{age}}$, an age-specific cubic polynomial function over time is used, while a cohort-specific cubic polynomial function over time is used to obtain $y_{x,t}^{\text{cohort}}$. To produce the final MI scales, the interpolated values along the two dimensions are combined using

$$
\gamma_{x,t}^{\text{MIM}} = \omega_{x,t} \gamma_{x,t}^{\text{age}} + (1 - \omega_{x,t}) \gamma_{x,t}^{\text{cohort}},
$$

where $\omega_{x,t}$ is a weight given again by the model user. The default value in the MIM-2021 model for $\omega_{x,t}$ is 50% for all $x$ and $t$. Lastly, if the intermediate-term rates are also externally given, then the two cubic polynomials will be applied from year $t_{n_y}$ to the year of the intermediate-rates, and afterwards linear interpolations will be used to connect the intermediate-term and long-term rates.

### 2.3 Our motivations

To motivate our study, we now discuss the features and limitations of the MIM-2021 model. An illustration of the interpolation methods used in the MIM-2021 model is provided in Fig. 1. It is assumed in the MIM-2021 model that $t_1 = 1982$.
and \( t_n = 2018 \). The end of the projection period \( t_n + n_p \) is set by the model user as the long-term terminal point. To project MI rates from \( t_n \) to \( t_n + n_p \), the MIM-2021 model relies on either linear or cubic polynomial functions.

As shown in the left panel of Fig. 1, with only long-term rates externally given, the cubic polynomial function is used to first extend the recently observed and smoothed MI rates to a set of projected short-term rates. These rates highly depend on the recent MI trends formed by the smoothed data. The same function then continues to project MI rates to the externally given long-term rates, which in turn produces a smooth transition between data and external rates. If intermediate-term rates are also given, then the short-term rates will first be connected with the intermediate-term rates, and then further extended to the long-term ones using a linear interpolation, as shown in the right panel of Fig. 1.

Although the MIM-2021 model can produce a smooth projection of future MI rates, it is subject to several limitations. First, as indicated by Eqs. (1) and (2), the graduation method and the projection method used in the MIM-2021 model are not the same method. The data are smoothed by the Whittaker–Henderson method, while the projection is produced by polynomial interpolations. This inconsistent use of two methods may lead to a jagged transition between the last year of data and the first year of projection, known as the “edge effects”.

Second, the MIM-2021 model’s projection of future MI rates is not flexible. The two-dimensional projection is obtained from a weighted average of two independent one-dimensional polynomial functions, as suggested by Eq. (3). The interactions of MI rates across ages and cohorts cannot be captured under this approach. In addition, the projected MI trends will be stagnant, since the projecting function used is either a cubic or linear polynomial.

Third, the MIM-2021 model can only provide point estimates of future MI rates, but not a measure of uncertainty surrounding the point estimates. This shortcoming greatly limits the model’s use in risk management tasks, such as calculating solvency capital requirements. Further, the MIM-2021 model treats
external opinions on future MI rates as deterministic ‘true’ values with no uncertainty assumed to these externally given rates.

In the rest of this paper, we aim to develop a mortality improvement model that can produce a smooth projection of two-dimensional future MI rates with external opinions on intermediate- and long-term rates incorporated and appropriate measures of uncertainty for the observed, projected and externally given rates.

3 The proposed model

This section presents the proposed two-dimensional cubic B-spline MI projection model. A brief review of the spline mortality modeling technique developed in Currie et al. [8] is first provided. We then discuss the implementation of LASSO penalty and the incorporation of external opinions under a two-dimensional cubic B-spline. Lastly, the smooth projection method of the final model is provided.

3.1 Spline mortality modeling

3.1.1 One-dimensional smoothing

We first illustrate the idea of spline smoothing on the one-dimensional case. For a particular year $t$, suppose we want to smooth mortality data along the age dimension from age $x_1$ to $x_{n_a}$. Using a one-dimensional spline function over age, we can obtain smoothed values of $(y_{x_1,t}, \ldots, y_{x_{n_a},t})$ from

$$y_{x,t} = \sum_{i=1}^{c_a} \beta_i b_i(x) + z_{x,t}$$

(4)

where $c_a$ is the number of basis functions used in the spline, $b_i(x)$ is the $i$-th basis function evaluated at age $x$, $\beta_i$ is the coefficient associated with the $i$-th basis function, and $\{z_{x,t}\}$ are independent and identically distributed (i.i.d.) normal errors with mean 0 and variance $\sigma^2$.

Putting the MI rates in a $(n_a \times 1)$ vector form, denoted by $y_t = (y_{x_1,t}, \ldots, y_{x_{n_a},t})^T$, Eq. (4) can be expressed in a matrix form as

$$y_t = B_t \times \beta_t + z_t,$$

(5)

where $B_t$ is a $(n_a \times c_a)$ matrix known as the basis matrix, $\beta_t$ is a $(c_a \times 1)$ vector consists of the coefficients, and $z_t$ is a $(n_a \times 1)$ vector of i.i.d. normal errors. The choice of the basis matrix $B_t$ will be discussed later in this subsection.

3.1.2 Two-dimensional smoothing

For the two-dimensional case, we follow the idea of Currie et al. [8], which uses two one-dimensional marginal splines on each of the age and time dimensions to build a two-dimensional spline for smoothing mortality rates.
Consider the MI rates in a two-dimensional \((n_a \times n_y)\) matrix form, denoted by
\[
Y = (y_{t_1}, \ldots, y_{t_n}).
\]
Let \(B_a\) be the \((n_a \times c_a)\) basis matrix of a one-dimensional marginal spline along the age dimension, and \(B_y\) be the corresponding \((n_y \times c_y)\) basis matrix along the time dimension, where \(c_y\) is the number of basis functions in \(B_y\). The two-dimensional spline model for \(Y\) is defined as
\[
y = B \times \beta + z,
\]
where
- \(y = \text{vec}(Y)\) is the vectorization of \(Y\),
- \(B = B_a \otimes B_y\) is the basis matrix derived from the Kronecker product of \(B_a\) and \(B_y\),
- \(\beta\) is a \((c_ac_y \times 1)\) vector consists of all the coefficients for the basis functions in \(B\), and
- \(z\) is a \((n_an_y \times 1)\) vector of i.i.d. normal errors.

### 3.1.3 The basis matrix

The basis functions can take various different forms. For our modeling and projecting purposes, we choose to use B-spline. Among numerous choices, B-spline has particular popularity due to its sparse structure. This feature is particularly useful to us, since it greatly enhances the numerical stability of the estimation process when the number of coefficients used is large.

The definition of B-spline is expressed in a recursive form, starting from a B-spline with an order of zero and its \(i\)-th basis function being defined as
\[
b_{i,0}(x) := \begin{cases} 
1 & \text{if } k_i \leq x < k_{i+1}, \\
0 & \text{otherwise}
\end{cases}
\]
where \(k_i\) is called a chosen knot. For the higher orders, \(h \geq 1\), the basis function follows
\[
b_{i,h}(x) := \frac{x - k_i}{k_{i+h} - k_i} b_{i,h-1}(x) + \frac{k_{i+h+1} - x}{t_{i+h+1} - t_{i+1}} b_{i+1,h-1}(x).
\]
Notice that B-spline has only local support, which means the impact of the \(i\)-th basis function becomes zero if \(x \geq k_{i+h+1}\) or \(x \leq k_i\). This feature makes B-spline particularly suitable for modeling temporary short-term shocks because the effect of shocks will diminish to zero over time.

The common choice of the highest order is \(h = 3\), which in turn leads to a cubic B-spline. We choose to use \(h = 3\) for our MI B-spline model as well. If we select \(c_a + 3\) knots along the age dimension and \(c_y + 3\) knots along the time dimension, then the total number of basis functions in \(B\) and coefficients in \(\beta\) is \(c_ac_y\). We thus need to estimate \(c_ac_y\) numbers of coefficients for our cubic B-spline MI model.
3.2 LASSO penalty

The coefficients in a spline model can be estimated via minimizing the following objective function over $\beta$:

$$(y - B \times \beta)^T (y - B \times \beta).$$

However, to achieve an appropriate balance between the smoothness and goodness-of-fit, a penalty term is often added, which results in a P-spline model. In particular, the objective function becomes

$$(y - B \times \beta)^T (y - B \times \beta) + P(\beta),$$

where $P(\beta)$ is a penalty function with an order of 1, 2 or 3.

We remark that P-spline is the model originally used in Currie et al. [8]. However, for our objectives of modeling MI rates, P-spline is subject to a number of caveats. First, the number and position of knots are restrictive. On this matter, Currie et al. [8] suggested that “the following simple rule-of-thumb is often sufficient: with equally spaced data, use one knot for every four or five observations up to a maximum of 40 knots ...” We cannot follow this rule-of-thumb, because, when intermediate- and/or long-term external rates are given, the resulting data structure will no longer be equally spaced.

Second, the order of the penalty function in P-spline needs to be subjectively determined, and will greatly affect the projection results. As discussed in Currie et al. [8], setting the penalty order to 1, 2 and 3 is equivalent to assuming no MI, linear MI and quadratic MI in the future, respectively. Thus, a P-spline model may perform well in terms of smoothing historical data, but it cannot achieve our desired feature of smoothly transitioning MI rates from recent observations to external rates.

To address the issues, we choose to adopt the LASSO penalty for our cubic B-spline MI model. The objective function is specified as

$$(y - B \times \beta)^T (y - B \times \beta) + \lambda \| \beta \|_1$$

where $\lambda$ is the smoothing parameter and $\| \beta \|_p$ is the $p$-norm of $\beta$. Under LASSO penalty, the number of basis functions used can dramatically increase, the position of knots no longer needs to be equidistant, and there is no penalty order to determine.

The LASSO penalty will shrink unnecessary coefficients towards zero, especially for those that have no data nearby. This feature is particularly useful to us, because a large number of knots (likely more than the number of data points we have) will need to be placed between the data and external rates achieve a smooth transition. Moreover, our preliminary analysis indicates that a P-spline model may encounter computational errors when the number of knots used is too large, but switching to LASSO penalty will solve the problem. In short, the flexibility and stability of LASSO penalty make it an excellent penalty option for our modeling and projecting objectives.
3.3 External rates

Aside from the observed mortality rates, expert opinions on future mortality rates collected from external sources can be valuable information in modeling and projecting mortality rates. In the literature, many studies have considered incorporating external information in mortality modeling, for example, Billari et al. [1], Billari et al. [2], Dowd et al. [14] and Dodd et al. [13]. In this paper, we propose an alternative approach that is compatible with the cubic B-spline MI model, and also in line with the needs of the insurance industry’s practice.

Suppose we have collected $K$ sets of external opinions on MI rates for $p$ future years, denoted as $\tilde{y}_k^{(\tilde{t})} = (y_{x_1, \tilde{t}}, \ldots, y_{x_a, \tilde{t}})^T$, $\tilde{t} \in \{\tilde{t}_1, \ldots, \tilde{t}_p\}$ and $k = 1, \ldots, K$, where $\tilde{t}_p$ is assumed to be the last year of the projection period. By considering these external rates as an additional source of ‘data’, we combine the observed rates with the external rates to create a synthetic dataset, denoted by

$$\tilde{Y} = \left( Y, y_1^{(1)}, \ldots, y_1^{(K)}, \ldots, y_p^{(1)}, \ldots, y_p^{(K)} \right).$$

The corresponding basis matrix becomes

$$\tilde{B} = \begin{pmatrix} B \\ J_{K,1} \otimes B_{\tilde{t}_1} \\ \vdots \\ J_{K,1} \otimes B_{\tilde{t}_p} \end{pmatrix},$$

where $J_{K,1}$ is a $(K \times 1)$ vector of ones, and $B_{\tilde{t}}$ is the corresponding $(n_a \times c_a c_y)$ basis matrix for the external rates in year $\tilde{t} \in \{\tilde{t}_1, \ldots, \tilde{t}_p\}$. Since the number of columns in basis matrices $\tilde{B}$ and $B$ is the same, the number of coefficients $\beta$ used in the B-spline model is unchanged, and the LASSO penalty of Eq. (8) can be simply adjusted to

$$(\tilde{y} - \tilde{B} \times \beta)^T(\tilde{y} - \tilde{B} \times \beta) + \lambda \| \beta \|_1$$  \hspace{-1cm} (9)

where $\tilde{y} = \text{vec}(\tilde{Y})$.

From a practical point of view, our approach is very sensible because it can simultaneously consider multiple external opinions collected from a single survey. Whether these opinions provides the same or different future MI rates, the proposed approach will consistently incorporate each of them to form an aggregated opinion. More specifically, our Bayesian estimation approach (discussed in Sect. 4) will integrate the similarity and/or difference among the external opinions to form a joint prior belief of future MI rates.

Furthermore, the external opinions collected do not need to include range or volatility estimates, because the point estimates of multiple opinions can naturally reflect a certain level of uncertainty for future rates. In comparison, the MIM-2021 model can be regarded as using only one external opinion (specified by the model user) or using one “consensus” of all external opinions collected. Since the difference among external opinions is not reflected, the uncertainty underlying the external rates will be highly underestimated.
3.4 Smooth projection

We are now ready to present the final smooth projection method of the proposed cubic B-spline model with LASSO penalty and external rates incorporated. Let

\[ \hat{Y} = (\hat{y}_{t_1}, \ldots, \hat{y}_{t_{n_y}}, \hat{y}_{t_{n_y}+1}, \ldots, \hat{y}_{\bar{t}_p}), \]

and \( \hat{y} = \text{vec}(\hat{Y}) \) denote the smoothed/estimated values of past MI rates all ages from year \( t_1 \) to \( t_{n_y} \), combined with the projected values of future MI rates for all ages from year \( t_{n_y} + 1 \) and \( \bar{t}_p \) (the last year of the projection period). The proposed two-dimensional cubic B-spline model obtains \( \hat{y} \) as

\[ \hat{y} = \hat{B} \times \hat{\beta}, \]

where

- \( \hat{\beta} \) is a vector consists of all the coefficients estimated from minimizing Eq. (9), and
- \( \hat{B} = (B^T, B_{t_{n_y}+1}^T, \ldots, B_{\bar{t}_p}^T)^T \), with \( B \) being the basis matrix defined in Eq. (6) and \( B_t \) being the corresponding basis matrix for future year \( t \in \{t_{n_y} + 1, \ldots, \bar{t}_p\} \).

4 The Bayesian estimation

As a flexible modeling method, B-spline models can provide accurate in-sample fitting and smooth out-of-sample projections. However, because a large number of parameters/coefficients is used, the parameter uncertainty of a B-spline model should not be ignored in its model estimation and projection process. For instance, the smoothing parameter \( \lambda \) of LASSO penalty controls the smoothness and goodness-of-fit of a B-spline model. The frequentist approach often determines the value of \( \lambda \) using cross-validation, which provides no straightforward measure of parameter uncertainty for \( \lambda \).

Another issue that we encountered in our preliminary analysis is that, when minimizing a penalized objective function (e.g., Eq. (7)), the inversion of a large matrix may become unstable. Since the mortality data used has \( n_a \) ages and \( n_y \) years, our basis matrix for the sample period already has a large size of \( (n_a n_y \times c_a c_y) \). This size will further expand when the projection period and external rates are considered.

The Bayesian estimation approach allows us to overcome the aforementioned issues in a statistically rigorous manner. We thus adopt the Bayesian approach to estimate our proposed model, and present the technical details of our Bayesian estimation procedure in this section.

4.1 Priors

Under the Bayesian approach, all model parameters are considered to be random variables, of which prior distributions must be provided. Our model parameters
include the coefficient vector \( \mathbf{\beta} = (\beta_1, \ldots, \beta_{c acy})^T \), the smoothing parameter \( \lambda \) for LASSO penalty, and the variance parameter \( \sigma^2 \) of i.i.d. normal errors.

We start with the \( \mathbf{\beta} \) coefficients. As noted by Tibshirani [44], the LASSO penalty term in Eq. (8) can be interpreted as assuming a Laplace prior to the coefficients. We adopt this result for our model’s LASSO penalty term in Eq. (9). Following the work of Park and Casella [35], the prior of each coefficient in \( \mathbf{\beta} \) is defined as

\[
\pi(\beta_i | \sigma^2, \lambda) = \frac{\lambda}{2\sqrt{\sigma^2}} \exp\left(-\frac{\lambda \cdot |\beta_i|}{\sqrt{\sigma^2}}\right),
\]

for \( i = 1, \ldots, c acy \). The reason for including \( \sigma^2 \) in the prior of \( \beta_i \) is to guarantee a unimodal posterior for \( \beta_i \).

To simplify the implementation of the estimation procedure, an equivalent but hierarchical representation for the prior of \( \beta_i \) is adopted:

\[
\pi(\beta_i | \sigma^2, \lambda) = \int_0^\infty \left\{ \frac{1}{\sqrt{2\pi\sigma^2 \tau_i}} \exp\left(-\frac{\beta_i^2}{2\sigma^2 \tau_i}\right) \right\} \times \left\{ \frac{\lambda^2}{2} \exp\left(-\frac{\lambda^2 \tau_i}{2}\right) \right\} d\tau_i,
\]

where the first term inside the integral indicates a normal distribution, while the second term indicates an exponential distribution. Here a new auxiliary variable \( \tau_i \) is introduced for each \( \beta_i \). The prior for \( \beta_i \) is indeed a normal distribution with mean zero and variance \( \sigma^2 \tau_i \), and the hyper-prior for \( \tau_i \) follows an exponential distribution with scale parameter \( \frac{2}{\lambda^2} \).

Denote \( \mathbf{\tau} = (\tau_1, \ldots, \tau_{c acy})^T \) and \( \mathbf{D}_\tau = \text{diag}(\tau) \). We summarize the prior for \( \mathbf{\beta} \) in the following vector form:

\[
\begin{align*}
\beta | \sigma^2, \mathbf{\tau} &\sim \text{MVN}_{c acy}(0, \sigma^2 \mathbf{D}_\tau), \\
\mathbf{\tau} | \lambda &\sim \text{i.i.d. Exp} \left( \frac{2}{\lambda^2} \right).
\end{align*}
\]

For \( \lambda^2 \) and \( \sigma^2 \), we use a non-informative prior, such that \( \sigma^2 \propto \frac{1}{\sigma^2} \) and \( \lambda^2 \propto \frac{1}{\lambda^2} \).

### 4.2 Full conditionals

Following the priors assumed in the previous subsection, the full conditionals of \( \mathbf{\beta} \), \( \mathbf{\tau} \), \( \lambda \) and \( \sigma^2 \) can be derived as follows.

- The full conditional of \( \mathbf{\beta} \) is a multivariate normal distribution with a dimension of \( c acy \), a mean vector

\[
\mathbf{\mu}_\beta = \left( \mathbf{\bar{B}}^T \mathbf{\bar{B}} + \mathbf{D}_\tau^{-1} \right)^{-1} \mathbf{\bar{B}}^T \mathbf{\bar{y}},
\]

and a covariance matrix of
\[ \sum_\beta = \sigma^2 \left( \tilde{\beta}^T \tilde{\beta} + D_{\tau}^{-1} \right)^{-1}. \]

- The full conditional of \( \sigma^2 \) is an inverse-gamma distribution with shape parameter
  \[ a_{\sigma^2} = \frac{n_an_y + c_a c_y + n_a \times K \times p}{2}, \]
  and scale parameter
  \[ b_{\sigma^2} = \frac{(\bar{y} - \tilde{\beta} \beta)^T(\bar{y} - \tilde{\beta} \beta) + \beta^T D_{\tau}^{-1} \beta}{2}. \]

- The full conditional of \( \tau_i^{-1}, i = 1, \ldots, c_a c_y \), is an inverse-Gaussian distribution with mean parameter
  \[ a_{\tau_i} = \sqrt{\frac{\lambda^2 \sigma^2}{\beta^2}}, \]
  and shape parameter \( b_{\tau_i} = \lambda^2 \).

- The full conditional of \( \lambda^2 \) is a gamma distribution with shape parameter \( a_{\lambda^2} = c_a c_y \)
  and rate parameter
  \[ b_{\lambda^2} = \frac{\text{tr}(D_{\tau})}{2}. \]

### 4.3 Gibbs sampling

Since the full conditionals of all parameters are explicitly derived, Gibbs sampling can be directly applied. The total number of parameters (including the auxiliary ones) is \( 2c_a c_y + 2 \). We denote all of the parameters in a single set as

\[ \Theta = \left\{ \beta_1, \ldots, \beta_{c_a c_y}, \tau_1, \ldots, \tau_{c_a c_y}, \sigma^2, \lambda \right\}. \]

In each iteration of Gibbs sampling, a sample of parameter \( \theta_j \), for \( j = 1, \ldots, 2c_a c_y + 2 \), is drawn from its full conditional, specified in the previous subsection,

\[ \pi\left( \theta_j \mid \bar{y}, \Theta_{-\theta} \right), \]

where \( \Theta_{-\theta} \) is the set \( \Theta \) excluding the \( j \)-th element \( \theta_j \). After a large number of iterations, the drawn samples will converge to the joint posterior distribution of \( \pi(\Theta | \bar{y}) \).

Finally, denoting \( \hat{y}_{x,t}, x = x_1, \ldots, x_n \) and \( t = t_1, \ldots, t_n \), as the smoothed MI rate in the sample period, the posterior distribution of \( \hat{y}_{x,t} \) given \( \bar{y} \) is obtained from
It is well known that samples drawn from Gibbs sampling exhibit auto-correlation, and hence thinning must be applied to avoid bias in assessing the posterior distribution. We use the method of Effective Sample Size (ESS) to decide the number of effective samples for our final Bayesian analysis. In addition, both the single-chain convergence test Geweke [17] and the multi-chain convergence test Gelman and Rubin [16] are applied to verify the convergence of the effective samples.

### 4.4 Predictive distribution

The projections of future MI rates and also central death rates are obtained from the posterior predictive distribution. Let \( \tilde{y}_{x,t} \) and \( \tilde{m}_{x,t} \), \( x = x_1, \ldots, x_{n_a} \) and \( t = t_n + 1, \ldots, \tilde{t}_p \), denote the projected future MI rates and central death rates, respectively. Assuming that the past and future rates are independently distributed given \( \Theta \), the posterior predictive distribution of \( \tilde{y}_{x,t} \) and \( \tilde{m}_{x,t} \) are obtained from

\[
f(\tilde{y}_{x,t} | \tilde{y}) = \int f(\tilde{y}_{x,t} | \Theta) \cdot \pi(\Theta | \tilde{y}) \cdot d\Theta
\]

and

\[
f(\tilde{m}_{x,t} | \tilde{y}) = \int f(\tilde{m}_{x,t} | \tilde{y}^{-}) \cdot f(\tilde{y}^{-} | \Theta) \cdot \pi(\Theta | \tilde{y}) \cdot d\Theta d\tilde{y}^{-},
\]

respectively, where \( \tilde{y}^{-} = (\tilde{y}_{x,t_{n_a}+1}, \ldots, \tilde{y}_{x,t})^T \).

Empirically, we first draw a realization of \( \beta \) and \( \sigma^2 \), denoted respectively as \( \hat{\beta} \) and \( \hat{\sigma}^2 \), from their joint posterior distribution \( \pi(\beta, \sigma^2 | \tilde{y}) \). Then, we simulate \( \tilde{y}_{x,t} \) from a normal distribution with mean \( \hat{b}(x, t)\hat{\beta} \) and variance \( \hat{\sigma}^2 \), where \( \hat{b}(x, t) \) is the corresponding row from the basis matrix \( \hat{B} \) for age \( x \) and year \( t \). Lastly, we set

\[
\hat{m}_{x,t} = m_{x,t_n} \left( \prod_{s=t_n+1}^{t} 1 - \tilde{y}_{x,s} \right)
\]

In our Gibbs sampling procedure, the values of \( \tilde{y}_{x,t} \) and \( \hat{m}_{x,t} \), \( x = x_1, \ldots, x_{n_a} \) and \( t = t_n + 1, \ldots, \tilde{t}_p \), are obtained at the end of each iteration. Note that the parameters \( \tau \) and \( \lambda \) do not directly concern the projections.

### 5 A numerical illustration

In this section, we use real mortality data and external rates to illustrate how the proposed cubic B-spline model can be used to generate smoothed two-dimensional MI rates.
5.1 Data

The mortality data used are obtained from the United States Social Security Administration (SSA). This dataset contains US male and female mortality experience for years 1982 to 2018 and ages 0 to 110. In this illustration, we consider the female population and the entire sample period 1982–2018, but reduce the age range to 35–99 to cover the typical ages of insureds. We further assume that expert opinions on future intermediate-term and long-term MI rates are collected from an external source. For illustrative purposes, we use the intermediate-term and long-term rates adopted from the MIM-2021 model for years 2030 and 2042, respectively.

Figure 2 depicts the MI rates of the assumed data in a surface plot (left panel) and a heat map (right panel). We observe that the historical MI rates are not smooth over the age and time dimensions. Both negative and positive values were observed over the sample period. Over the age range, stronger MI rates were observed for old ages in the recent years. Moreover, cohort effect can be detected from the heat map, where regions of ‘hot’ and ‘cold’ colors are formed along the diagonal line. Compared to the historical rates, the external rates are relatively smooth and positive for all ages.

We restate here that the objective of our spline modeling approach is to graduate the historical rates on a two-dimensional scale, and smoothly project the rates from recent observations to the externally given ones with measures of uncertainty.

5.2 Estimation results

The Bayesian estimation results of the model parameters \((\beta_1, \ldots, \beta_{c_1c_2}, \sigma^2, \lambda)\) are presented in this subsection. From a total of 26,000 iterations of Gibbs sampling, we collected 5000 drawn samples with a burn-in period of 1000 iterations and a thinning frequency of 5 iterations per sample. These 5000 drawn samples form an empirical posterior distribution for all the model parameters.

Figure 3 shows two histograms, depicting the posterior distribution of \(\sigma^2\) (left panel) and \(\lambda\) (right panel). For the LASSO penalty parameter \(\lambda\), the empirical
distribution has a mean of 0.8456, a variance of 0.0405, and is slightly skewed to the right. A higher value of $\lambda$ corresponds to a stronger weight on the penalty term, and consequently a smoother two-dimensional MI surface. For the variance parameter $\sigma^2$, the empirical distribution is mostly symmetric and stable, with a mean of $2.6424 \times 10^{-4}$ and a variance of $6.1191 \times 10^{-11}$.

Figure 4 depicts the posterior distribution of three selected $\beta$ parameters. Recall from Eq. (6) that $\beta$ parameters are the coefficients of basis functions. We have used $c_a = 16$ basis functions with knots evenly spaced over the age range 35–99, and $c_y = 12$ basis functions with knots evenly spaced over the sample period 1982–2018 and also on the external years 2030 and 2042, which in total produces $c_a c_y = 192$ numbers of $\beta$ parameters. Since LASSO penalty will shrink certain coefficients to zero, not all of the estimated $\beta$ parameters have non-zero values. Further, because Bayesian estimation is used, the $\beta$ parameters that should have a value of zero will instead have a posterior distribution concentrated around a mean of zero. Other $\beta$ parameters may have a posterior distribution with either a positive or negative mean. The three selected $\beta$ parameters shown in Fig. 4 correspond to these three cases.

5.3 Projection results

We now examine the projected MI rates, the projected central death rates, and also the smooth transition between the historical values in the dataset and the intermediate-term and long-term rates given by external opinions.

Figure 5 illustrates the estimated and projected two-dimensional MI rates produced by the proposed cubic B-spline model. Compared to Fig. 2 which depicts the raw data, the surface plot (left panel) shows that our model is able to generate a smooth two-dimensional MI surface. From the heat map (right panel), we observe that the smoothed and projected MI rates are not only age- and time-dependent, but also exhibit cohort effects as indicated by the ‘hot’ and ‘cold’ regions formed diagonally. Moreover, one can observe that the short-term MI trends shown in recent

![Fig. 3](image)

The empirical posterior distribution of $\sigma^2$ (left) and $\lambda$ (right), generated by 5000 drawn samples
Fig. 4. The empirical posterior distribution of three selected $\beta$ parameters with a mean of negative (left), zero (middle) and positive (right) value, generated by 5000 drawn samples.
observations fade away over time, and the projected MI rates smoothly converge to the intermediate-term and long-term rates.

The smooth transition between recent observations and external rates is further examined in Fig. 6. We use fan charts to show the posterior distribution of the estimated and projected MI rates for ages 45, 55, 65 and 75. The gray fan charts represent the estimated historical rates and the externally given rates, while the green fan charts represent the projected rates. The mean of the posterior predictive distribution is displayed by a solid red line. As shown in Fig. 6, the projected MI rates are smoothly transitioned from recent observations (ended at year 2018) to the intermediate-term rates (in year 2030), and then ultimately converged to the long-term rates (in year 2042). More specifically, the green fan chart shows that the projected MI rates for years 2019–2029 are heavily driven by the MI trends formed before year 2018, but will eventually connect to the intermediate-term rate in year 2030. After year 2030, the projected MI rates are further extended to the long-term rate in year 2042, following the smooth pattern of previous projections.

Figure 7 displays the posterior distribution of the projected log central death rates for ages 45, 55, 65 and 75. The mean forecasts are indicated by the darkest region in the center of the fan charts, while the level of uncertainty surrounding the mean forecasts is represented by the width of the fan charts. From Fig. 7, we observe that there is a smooth transition of mortality changes in the projection period. For the first few years after 2018, the projected rates follow the trend indicated by the observed values before year 2018. This period corresponds to the first green fan chart region in Fig. 6 for years 2019–2029. As the projection passes through year 2030, the rate of mortality changes starts to follow the intermediate-term rate, and then ultimately approach to the long-term rate.

To further examine the composition of the uncertainty surrounding the mean forecasts, we show two sets of predictive intervals at the 90% confidence level for ages 45, 55, 65 and 75 in Fig. 8. The one with solid lines represents the predictive intervals with uncertainty generated from $\beta$ parameters only, while the one with dash lines includes the uncertainty added from random noises. Recall...
from Eq. (6) that the spline mortality model is subject to random normal noises $\zeta_{x,t}$. From Fig. 8, we observe that the impact of $\zeta_{x,t}$ on predictive intervals is obvious, but not as significant as the uncertainty generated by $\beta$ parameters.

We now compare the projected rates between our Bayesian spline model and the MIM-2021 model. Figure 9 shows the log central death rates projected by the MIM-2021 model, and also those projected by our Bayesian spline model with 90% predictive intervals and random noises. First, we observe that, while the projected rates are different between the two models, the ones projected by the MIM-2021 model lie within the 90% predictive intervals of our model. Second, unreasonable projection trends could be produced by the MIM-2021 model. More specially, for age 45, we see that the MIM-2021 model’s projected death rates are increasing for the first few years. Such trend clearly does not align with the historical rates from the SSA dataset. A similar situation can also be observed for age 55, where the decreasing trend is much slower than the one formed by recent historical rates. Third, we remark that the MIM-2021 model is only able to generate the mean forecasts but not the predictive intervals.
5.4 The impact of external opinions

In this subsection, we use two different modeling scenarios to showcase the flexibility of the proposed model in incorporating external opinions. In particular, we consider (a) the baseline scenario, which uses the same model settings as in the previous subsections, (b) Scenario I, in which multiple years of external rates are given, and (c) Scenario II, in which multiple external opinions for the same years are given. Figure 10 shows the posterior distribution of the estimated and projected MI rates for age 55 under the baseline scenario and the two alternative scenarios.

Scenario I represents the case that the modeler wants to specify multiple years of intermediate-term and long-term MI rates from external sources. Our model can accommodate this situation by simply adding multiple years of external rates to the synthetic dataset, as discussed in Sect. 3.3. Panel (b) of Fig. 10 shows the Bayesian estimation and projection results for age 55, assuming that the intermediate-term rates for years 2029–2031 and the long-term rates for years 2041–2043 are externally given. Compared to the baseline scenario shown in panel (a), a lower level of uncertainty is observed for years that have external information available in its own or adjacent years. The mean forecast still follows the trend of recent observations, and smoothly connects to the intermediate-term and long-term rates over time.

Scenario II represents the case that the modeler wants to specify multiple external opinions for the same intermediate-term and long-term years. In practice, the
The modeler may collect multiple opinions about future MI rates. As discussed in Sect. 3.3, by treating these opinions as multiple inputs for the same year in the synthetic dataset, the proposed model is able to simultaneously incorporate all of them. If similar values are provided by multiple opinions, then the model should reflect the fact that we are highly confident about the MI rates of those years. Panel (c) of Fig. 10 shows the results for age 55, assuming that the modeler has collected 10 sets of the same opinions for the intermediate-term year 2030 and the long-term year 2042. It is clear that, when multiple external opinions are used, the level of uncertainty surrounding the mean forecast greatly decreases comparing to the baseline scenario.

5.5 The impact of basis functions

We now examine the impact of basis functions on the projected MI rates. Depending the modeler’s prior beliefs, a different transition rate might be desired between historical values and external rates. For example, assuming that a recent mortality shock was observed in historical values, the modeler may believe that the shock is transitional and will soon fade away while converging to the external rates. Conversely, the modeler may also believe that the shock will last for a while or even further extend to a more extreme shock before connecting back to the external rates.
Moreover, as a practical need, the MIM-2021 model also allows the user to specify an external transition rate at the beginning of the projection period. Under the proposed B-spline model, we can change the location of basis functions and knots to achieve this modeling requirement.

Figure 11 shows the posterior distribution of the estimated and projected MI rates for age 55, with basis functions and knots placed in (a) year 2018 (Baseline), (b) year 2015 (Case I), and (c) year 2021 (Case II). The other basis functions used in the sample period and the external years are unchanged. Note that the basis functions and knots are no longer evenly spaced over the sample period under Case I and Case II.

For Case I, panel (b) of Fig. 11 shows that the rate of convergence from the historical values before year 2018 to the intermediate-term rate in year 2030 is faster than that of the baseline case. The transition during years 2019–2029 is smoother and less volatile, and the erratic MI trend observed during years 2003–2018 has little impact on the projected MI rates. In contrast, the results of Case II shown in panel (c) of Fig. 11 indicate a more pronounced transition from the historical years to the intermediate-term year. The recent erratic MI trend is projected to continue for about 5 more years, and will then converge to the external intermediate-term rate. Lastly, note that the estimated MI rates before year 2013 are largely unchanged among the three cases, because the basis functions used before year 2013 are the same.
The impact of penalty choice

We now compare the impact of penalty choice for a spline model on the estimation and projection results. Figure 12 shows the posterior distribution of the estimated and projected MI rates for age 55 when the penalty used is (a) LASSO (left penal), (b) P-spline with an order of 2 (middle penal), and (c) no penalty (right penal).

From panel (b), we see that the MI rates projected by a P-spline model with a penalty order of 2 have a linear upward trend that is extrapolated from the estimated rates before year 2018 (within the sample period). This outcome is expected because, as discussed previously in Sect. 3.2 and also in Currie et al. [8], a P-spline model’s penalty order greatly affects the projection results and setting the penalty order to 2 will produce a linear projection trend. Compared to panel (a) where the results from LASSO penalty is shown, we can conclude that our B-spline model with LASSO penalty is able to provide a higher level of flexibility in terms of the smoothness of both estimated and projected MI rates. This conclusion also confirms

![Fig. 10](image-url) Fan charts showing the posterior distribution of the estimated and projected MI rates for age 55, using single year and single opinion of external rates (Baseline), multiple years of external rates (Scenario I), and multiple opinions of external rates (Scenario II)

5.6 The impact of penalty choice

We now compare the impact of penalty choice for a spline model on the estimation and projection results. Figure 12 shows the posterior distribution of the estimated and projected MI rates for age 55 when the penalty used is (a) LASSO (left penal), (b) P-spline with an order of 2 (middle penal), and (c) no penalty (right penal).

From panel (b), we see that the MI rates projected by a P-spline model with a penalty order of 2 have a linear upward trend that is extrapolated from the estimated rates before year 2018 (within the sample period). This outcome is expected because, as discussed previously in Sect. 3.2 and also in Currie et al. [8], a P-spline model’s penalty order greatly affects the projection results and setting the penalty order to 2 will produce a linear projection trend. Compared to panel (a) where the results from LASSO penalty is shown, we can conclude that our B-spline model with LASSO penalty is able to provide a higher level of flexibility in terms of the smoothness of both estimated and projected MI rates. This conclusion also confirms
our objective for developing a LASSO penalty for the Bayesian two-dimensional B-spline model.

Panel (c) of Fig. 12 shows the estimated and projected MI rates when no penalty is implemented. It is no surprise that the projection results are unreasonable under this case. The magnitude of the projected MI rates is outrageously large as compared to that of the LASSO and P-spline penalties. Thus, we conclude that a penalty function is certainly necessary when using a spline model to project future MI rates.

6 Conclusion

This paper proposed a two-dimensional cubic B-spline model for generating smooth projections of mortality improvement rates. Our study was motivated by the recently proposed MIM-2021 model, which generates mortality improvement scales for actuarial professions. To build a spline mortality model that addresses the shortcomings of the MIM-2021 model and also maintains its practical features, we adopted the
two-dimensional B-spline approach proposed by Currie et al. [8]. To overcome the limitations of spline models in mortality forecasting, we implemented the LASSO penalty and Bayesian estimation for our model. The projected MI rates are smoothly transitioned from historical values with short-term trends to any externally given intermediate- and long-term rates. We numerically demonstrated the features of the proposed model using real mortality data and external opinions.

We acknowledge that the proposed model is subject to several caveats. First, a number of model inputs needs to be subjectively determined. Although we used Bayesian estimation and LASSO penalty to reduce the selection of smoothing parameter and penalty order, as a spline model, one still needs to subjectively specify the position of basis functions (as known as knot selection). We demonstrated that the position of basis functions can bring flexibility to the projected MI rates, but the selection process will require expert knowledge from the model user.

Second, the forecasting performance of our model is contingent on the quality of the external rates incorporated. To fulfill the practical needs of having

![Fan charts showing the posterior distribution of the estimated and projected MI rates for age 55 when the penalty used is a LASSO, b P-spline with an order of 2, and c no penalty](image)

**Fig. 12** Fan charts showing the posterior distribution of the estimated and projected MI rates for age 55 when the penalty used is a LASSO, b P-spline with an order of 2, and c no penalty
intermediate- and long-term MI rates, our model projects future MI rates from historical values to any externally given ones. The accuracy of these long-term forecasts is thus highly driven by the quality of the external opinions used. We acknowledge that there are other mortality models that may statistically outperform our model, and do not claim that our model is better than others in term of forecasting performance. Instead, our objective is to provide a modeling approach that satisfies the requirements of actuarial professions for generating future MI rates.

Third, our numerical illustration does not include mortality data from the Covid-19 period. Given that the pandemic data is not fully collected yet, we choose to postpone the study of Covid-19 mortality data for future research. At the time of writing, the MIM-2021 model was updated to the MIM-2021-v2 model by the Society of Actuaries to include data from year 2019 and also adjustments for the pandemic effects. A future study could further analyze the proposed model using a complete Covid-19 mortality dataset along with the MIM-2021-v2 model.

Other future work may focus on addressing some of the other aforementioned limitations. For example, the model’s performance may be improved by considering other penalty options and basis functions, such as ridge regression [22], roughness penalty [15], or elastic net [45]. Further improvement options include applying LASSO regularization to pre-process mortality data [37], or applying elastic net to reduce dimensions in MI modeling [18]. Lastly, to strengthen the cohort effect in mortality dynamics, one may consider applying the proposed B-spline model on the age-cohort dimension instead of the age-period dimension.

Acknowledgements The authors would like to thank the two anonymous reviewers for their constructive comments and suggestions that improved the quality of this paper. The authors also greatly appreciate the comments received from Hua Chen and other participants at the 24th International Congress on Insurance: Mathematics and Economics.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

References

1. Billari FC, Graziani R, Melilli E (2012) Stochastic population forecasts based on conditional expert opinions. J R Stat Soc Ser A (Stat Soc) 175(2):491–511
2. Billari FC, Graziani R, Melilli E (2014) Stochastic population forecasting based on combinations of expert evaluations within the Bayesian paradigm. Demography 51(5):1933–1954
3. Cairns AJ, Blake D, Dowd K (2006) A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. J Risk Insur 73(4):687–718
4. Camarda CG (2019) Smooth constrained mortality forecasting. Demogr Res 41:1091–1130
5. Camarda CG et al (2012) Mortalitysmooth: an R package for smoothing Poisson counts with p-splines. J Stat Softw 50(1):1–24
6. Chang L, Shi Y (2020) Dynamic modelling and coherent forecasting of mortality rates: a time-varying coefficient spatial-temporal autoregressive approach. Scand Actuar J 2020(9):843–863
7. Currie ID (2013) Smoothing constrained generalized linear models with an application to the Lee–Carter model. Stat Model 13(1):69–93
8. Currie ID, Durban M, Eilers PH (2004) Smoothing and forecasting mortality rates. Stat Model 4(4):279–298
9. Currie ID, Durban M, Eilers PH (2006) Generalized linear array models with applications to multidimensional smoothing. J R Stat Soc Ser B (Stat Methodol) 68(2):259–280
10. Delwarde A, Denuit M, Eilers P (2007) Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting: a penalized log-likelihood approach. Stat Model 7(1):29–48
11. DiMatteo I, Genovese CR, Kass RE (2001) Bayesian curve-fitting with free-knot splines. Biometrika 88(4):1055–1071
12. Dodd E, Forster JJ, Bijak J, Smith PW (2018) Smoothing mortality data: the English life tables, 2010–2012. J R Stat Soc Ser A (Stat Soc) 181(3):717–735
13. Dodd E, Forster JJ, Bijak J, Smith PW (2021) Stochastic modelling and projection of mortality improvements using a hybrid parametric/semi-parametric age-period-cohort model. Scand Actuar J 2:134–155
14. Dowd K, Cairns A, Blake D (2020) CBDX: a workhorse mortality model from the Cairns–Blake–Dowd family. Ann Actuar Sci 14(2):445–460
15. Eilers PH, Marx BD (1996) Flexible smoothing with B-splines and penalties. Stat Sci 11(2):89–121
16. Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. Stat Sci 7(4):457–472
17. Geweke JF (1991) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, vol 148. Federal Reserve Bank of Minneapolis, Research Department, Minneapolis
18. Guibert Q, Lopez O, Piette P (2019) Forecasting mortality rate improvements with a high-dimensional VAR. Insur Math Econ 88:255–272
19. Haberman S, Renshaw A (2012) Parametric mortality improvement rate modelling and projecting. Insur Math Econ 50(3):309–333
20. Haberman S, Renshaw A (2013) Modelling and projecting mortality improvement rates using a cohort perspective. Insur Math Econ 53(1):150–168
21. Hilton J, Dodd E, Forster JJ, Smith PWF (2019) Projecting UK mortality by using Bayesian generalized additive models. J R Stat Soc Ser C (Appl Stat) 68(1):29–49
22. Hoerl AE, Kennard RW (1970) Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12(1):55–67
23. Huang F, Browne B (2017) Mortality forecasting using a modified continuous mortality investigation mortality projections model for China I: methodology and country-level results. Ann Actuar Sci 11(1):20–45
24. Hunt A, Villegas AM (2017) Mortality improvement rates: modeling, parameter uncertainty and robustness. Presented at the living to 100 symposium
25. Hyndman RJ, Ullah MS (2007) Robust forecasting of mortality and fertility rates: a functional data approach. Comput Stat Data Anal 51(10):4942–4956
26. Lang S, Brezger A (2004) Bayesian p-splines. J Comput Graph Stat 13(1):183–212
27. Lee RD, Carter LR (1992) Modeling and forecasting us mortality. J Am Stat Assoc 87(419):659–671
28. Li H, Lu Y (2017) Coherent forecasting of mortality rates: a sparse vector-autoregression approach. ASTIN Bull J IAA 47(2):563–600
29. Li H, Shi Y (2021) Mortality forecasting with an age-coherent sparse var model. Risks 9(2):35
30. Li JS-H, Hardy M, Tan KS (2010) Developing mortality improvement formulas: the Canadian insured lives case study. North Am Actuar J 14(4):381–399
31. Li JS-H, Liu Y (2020) The heat wave model for constructing two-dimensional mortality improvement scales with measures of uncertainty. Insur Math Econ 93:1–26
32. Li JS-H, Liu Y (2021) Recent declines in life expectancy: implication on longevity risk hedging. Insur Math Econ 99:376–394
33. Li JS-H, Zhou QK, Zhu X, Chan W-S, Chan FW-H (2019) A Bayesian approach to developing a stochastic mortality model for China. J R Stat Soc Ser A (Stat Soc) 182(4):1523–1560
34. Luoma A, Puustelli A, Koskinen L (2012) A Bayesian smoothing spline method for mortality modelling. Ann Actuar Sci 6(2):284–306
35. Park T, Casella G (2008) The Bayesian Lasso. J Am Stat Assoc 103(482):681–686
36. Pitt D, Li J, Lim TK (2018) Smoothing Poisson common factor model for projecting mortality jointly for both sexes. ASTIN Bull J IAA 48(2):509–541
37. Rabbi AMF, Mazzuco S (2021) Mortality forecasting with the Lee-Carter method: adjusting for smoothing and lifespan disparity. Eur J Popul 37(1):97–120
38. Renshaw A, Haberman S (2021) Modelling and forecasting mortality improvement rates with random effects. Eur Actuar J 11:381–412
39. Renshaw AE, Haberman S (2003) On the forecasting of mortality reduction factors. Insur Math Econ 32(3):379–401
40. Richards SJ (2020) A Hermite-spline model of post-retirement mortality. Scand Actuar J 2:110–127
41. Shang HL (2019) Dynamic principal component regression: application to age-specific mortality forecasting. ASTIN Bull J IAA 49(3):619–645
42. Speckman PL, Sun D (2003) Fully Bayesian spline smoothing and intrinsic autoregressive priors. Biometrika 90(2):289–302
43. Tang KH, Dodd E, Forster JJ (2021) Joint modelling of male and female mortality rates using adaptive p-splines. Ann Actuar Sci 16:119–135
44. Tibshirani R (1996) Regression shrinkage and selection via the lasso. J R Stat Soc Ser B (Methodol) 58(1):267–288
45. Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. J R Stat Soc Ser B (Stat Methodol) 67(2):301–320

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