




# A Hermite spline approach for modelling population mortality

Sixian Tang<sup>1,2\*</sup> , Jackie Li<sup>1</sup>  and Leonie Tickle<sup>3</sup> 

<sup>1</sup>Department of Econometrics and Business Statistics, Monash University, Melbourne, VIC, Australia; <sup>2</sup>School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University, Perth, WA, Australia; and <sup>3</sup>Department of Actuarial Studies and Business Analytics, Macquarie University, Sydney, NSW, Australia

\*Corresponding author. E-mail: [sixian.tang@monash.edu](mailto:sixian.tang@monash.edu)

(Received 11 July 2021; revised 13 September 2022; accepted 16 September 2022; first published online 17 October 2022)

## Abstract

One complication in mortality modelling is capturing the impact of risk factors that contribute to mortality differentials between different populations. Evidence has suggested that mortality differentials tend to diminish over age. Classical methods such as the Gompertz law attempt to capture mortality patterns over age using intercept and slope parameters, possibly causing an unjustified mortality crossover at advanced ages when applied independently to different populations. In recent research, Richards (*Scandinavian Actuarial Journal* 2020(2), 110–127) proposed a Hermite spline (HS) model that describes the age pattern of mortality differentials using one parameter and circumvents an unreasonable crossover by default. The original HS model was applied to pension data at individual level in the age dimension only. This paper extends the method to model population mortality in both age and period dimensions. Our results indicate that in addition to possessing desirable fitting properties, the HS approach can produce accurate mortality forecasts, compared with the Gompertz and P-splines models.

**Keywords:** Mortality forecasting; Hermite splines; Mortality models; Longevity risk

## 1. Introduction

Over the past few decades, the reduction in human mortality has been accelerating far faster than expected. However, such improvement is not consistent across ages. Several studies found evidence that mortality in older populations (55+) has been declining faster than that in younger age groups (Christensen *et al.*, 2009). From 1970 onwards, the primary driving factor of rising human life expectancy in developed countries has shifted from decreasing mortality at younger ages to improving longevity at older ages (Cairns *et al.*, 2009).

Mortality modelling is crucial not only in demography but also in a wide range of areas such as public health, insurance and pension planning in both the private and public sectors. For instance, insurance companies that provide life annuities need to make accurate and sensible mortality predictions in pricing and reserving. Failing to do so might result in an unexpected loss on the annuity products that pay lifetime income to the annuitants on survival. These portfolios whose cash flows are linked to survival rates at higher ages are exposed to a risk of people living unexpectedly longer, i.e., longevity risk. In order to quantify risks associated with mortality rates, a number of stochastic mortality models have been proposed by researchers and practitioners. One category of stochastic mortality models called the extrapolative family makes assumptions about future age and period patterns of mortality rates based on past developments. According to Blake *et al.* (2018), extrapolative models may be broadly divided into three branches.

The first class originates from the Lee–Carter model (Lee & Carter, 1992) that analyses mortality rates in both age and time dimensions using multiplicative factors. Renshaw & Haberman (2003) extended the original Lee–Carter model to incorporate two factors and applied this extension to male mortality data in the United Kingdom and England and Wales. They proposed another extension of the original Lee–Carter model by adding a cohort factor to explain patterns specific to birth year (Renshaw & Haberman, 2006). A special case of the Lee–Carter model with a cohort effect is called the Age-Period-Cohort model (Currie, 2006). It replaces the age-specific sensitivity to changes in the period and cohort factors by the reciprocal of the number of ages in the data.

The original Lee–Carter model does not impose assumptions on the smoothness of fitted mortality rates across ages and periods. On the other hand, methods that are part of a second category of extrapolative approaches assume that mortality rates are smooth for adjacent ages but not years. An early method in this class was introduced by Cairns *et al.* (2006) who designed the so-called Cairns–Blake–Dowd (CBD) model for explaining mortality rates at higher ages. It is constructed based on the Gompertz law (Gompertz, 1825) that describes the logarithm of the mortality hazard in a given year as a linear function of age, while the CBD model works on the logit of death rates. The CBD model can also be regarded as an extension of the classic parametric mortality curves as described in McNown & Rogers (1989). Delwarde *et al.* (2007) applied a penalty structure to smooth out the irregular patterns in the age sensitivity factor of the Lee–Carter model while Hyndman and Ullah (2007) combined the non-parametric smoothing technique with the Lee–Carter model to deal with the randomness in the data. Another early model situated Lee–Carter within a state-space framework, enabling integrated estimation and forecasting, and imposed across-age smoothness using B-splines (De Jong & Tickle, 2006). Other variations in the second category have been developed by academics to cater for different needs. For example, Cairns *et al.* (2009) proposed a modification to the CBD model to accommodate cohort patterns. In the same paper, they further generalised the CBD model by adding a quadratic term to capture the potential curvature in the logit death rates.

The last group of extrapolative mortality models originates from splines-fitting techniques, imposing smoothness on both the age and time dimensions. Eilers & Marx (1996) imposed a difference penalty on the coefficients of B-splines formed by polynomials joined at different knots. This so-called penalised B-splines (P-splines (PS)) approach strikes a compromise between the adherence to data and smoothness of the fitted values. Then the one-dimensional regression splines were extended to fit temperature data in two dimensions (Eilers & Marx, 2003). Currie *et al.* (2004) further developed their work and considered the prediction method via PS matrix to forecast human mortality rates. A separate cohort factor is not included in the PS model, as patterns related to the year of birth are smoothed out. Dokumentov *et al.* (2018) applied a bivariate smoothing approach to produce mortality surfaces that allow for the period and cohort effects. Unlike the previous two classes of models that often predict future mortality scenarios by selecting appropriate time series processes for the period factors, under the PS method projection is a natural outcome of the smoothing process and does not rely on forecaster judgement. Under the PS method it is not possible to generate random sample paths of mortality rates without the use of methods such as bootstrap. One may use the variance–covariance matrix of the PS model to find the probability intervals of the fitted and predicted mortality rates. However, when it comes to pricing and hedging problems that require the simulated paths of mortality-/longevity-linked cash flows, the process uncertainty (variability in the time series) under the PS model cannot be examined.

Although there have been considerable developments in the methods of mortality modelling, no consensus has been achieved on the ideal choice so far. To draw a conclusion on the goodness of fit and forecast accuracy of mortality models, several studies reviewed and examined different extrapolative candidates both qualitatively and quantitatively (see Booth & Tickle, 2008; Cairns

*et al.*, 2009; Dowd *et al.*, 2010a, 2010b). Cairns *et al.* (2009) and (2011) developed a set of criteria to evaluate the performance of mortality models such as parameter parsimony, biological consistency and plausibility of prediction uncertainty. Thereinto, the biological reasonableness among multiple populations has been widely discussed in the literature. By way of illustration, it is commonly accepted that male life expectancy tends to be shorter than that of females at all ages (Alho & Spencer, 2006). To avoid the opposite situation, feasible mortality models should not include a crossover in age-related mortality rates by gender. The coherence between mortality rates of related populations can be achieved by employing multi-population models (e.g., Li & Lee, 2005; Hyndman *et al.*, 2013; Li, 2013; Haberman *et al.*, 2014; Villegas *et al.*, 2017).

When multiple populations are considered, it is vital to maintain features of risk factors contributing to mortality differentials for designing and pricing insurance products. A demographic theory called the compensation law of mortality (Gavrilov & Gavrilova, 1991) describes the convergence behaviour of mortality rates between populations. Specifically, the death rates of a population that are higher than those of related populations tend to increase with age at a lower speed (a lower aging rate). It follows that the mortality differentials between associated populations tend to diminish over age. The typical way of modelling mortality progression over age describes the mortality of a population using two parameters – level and slope. Under this setting, the fitted age-related mortality curves of related populations in a given year start from different intercepts and their gaps narrow when the populations get older. However, this method may cause a crossover problem. Richards (2020) illustrated this possible drawback by applying the widely used Gompertz law to data of two pension schemes with different pension sizes. The disadvantaged portfolio (small-size pension) exhibits heavier mortality at younger ages and a gradual increase over the age range. In contrast, the advantaged portfolio (large-size pension) displays lighter mortality at younger ages and steeper slope. The fitted mortality differential between them reduces with increasing age but eventually results in an unreasonable crossover at advanced ages.

Richards (2020) pointed out the imprudence of classical modelling approaches that potentially bring a crossover. He introduced a so-called Hermite spline (HS) model in which mortality differentials between two populations depend on one single parameter and take a strictly positive decreasing form, therefore resulting in convergence without crossover. Specifically, the full HS model of the log mortality hazard rate is constructed with four cubic polynomials in the Hermite form (Schoenberg, 1969; Kreyszig, 2009), which specifies the gradients approaching the start and end points of a given age interval. One may set one or both of the gradient parameters to zero to control the flexibility of the smoothed mortality path between two ages. The author considered four versions of the HS model, of varying flexibility. The risk factors (e.g., sample size, gender, smoking status, poverty, etc.) of a population can be incorporated by altering the coefficient (known as the intercept) of the first HS that is monotonically decreasing. Consequently, the fitted mortality differentials for each risk factor vanish as age increases, avoiding a crossover between populations by default.

Richards (2020) applied this method to smooth individual mortality data for pension schemes in the age dimension only. The task of projecting two-dimensional population mortality rates was left for future research. Richards (2021) further extended the original HS model by including a time trend into the gradient parameter and fitted the model using the individual data of a local pension scheme in England and Wales. Again, mortality forecasting was not addressed.

In addition to being an elegant smoothing approach for individual data, we show in this paper that the HS models also produce desirable forecast accuracy when country-level mortality experience is employed. In particular, the full HS model attains a good balance between goodness-and-fit and forecast accuracy. When multiple risk factors are considered, it tends to describe the patterns of mortality differentials well and to ensure convergence at advanced ages. The focus of this paper is on the longevity risk linked to higher-age survival probabilities; hence, we compare our model

to the widely applied Gompertz law and PS model. The former is well-known for its ability to model mortality data at higher ages, while the latter serves as a benchmark in spline fitting<sup>1</sup>.

The organisation of the paper is as follows. Section 2 briefly reviews the structure of HS s. It also introduces four variations of the two-dimensional HS models and the two benchmark candidates. We then demonstrate the estimation procedure and fitting results of the six models in section 3. The projection and backtesting performances are analysed in sections 4 and 5, respectively. A simulation study is presented in section 6 to examine further the predictive power of the six models. Finally, section 7 concludes the paper and provides some directions of further research.

## 2. Methodology

Before introducing the mortality models, we first make an assumption about the instantaneous mortality hazard rate. Since mortality data are mostly available for each single year of age, we assume that the force of mortality  $\mu_{x,t}$  at age  $x$  in year  $t$  remains constant over each integer age-period interval. It follows that the central death rate  $m_{x,t}$  is equivalent to the force of mortality  $\mu_{x,t}$ . For the rest of the paper, we employ  $m_{x,t}$  as the modelling measure.

### 2.1 HS models

Spline fitting is an interpolation method under which segments of low-degree polynomials connected at chosen control points are fitted to describe the data shape. One widely used category known as the cubic HS obtains the fitted curve using cubic polynomials defined in Hermite form (Hermite, 1864). In more detail, one specifies the values and first derivatives of the start and end points for each segment, and the shape of the fitted curve is controlled by the Hermite basis functions (Marschner & Shirley, 2018). In matrix form, the interpolation polynomial  $p(u)$  over  $u \in [0, 1]$  can be expressed as

$$p(u) = (u^3, u^2, u, 1) \begin{pmatrix} 2 & -2 & 1 & 1 \\ -3 & 3 & -2 & -1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ \delta_0 \\ \delta_1 \end{pmatrix} = \begin{pmatrix} 2u^3 - 3u^2 + 1 \\ -2u^3 + 3u^2 \\ u^3 - 2u^2 + u \\ u^3 - u^2 \end{pmatrix}' \begin{pmatrix} p_0 \\ p_1 \\ \delta_0 \\ \delta_1 \end{pmatrix} = \mathbf{f}(\mathbf{u})\mathbf{P}, \quad (1)$$

where  $\mathbf{P}$  is the  $(4 \times 1)$  control matrix,  $p_0$  and  $p_1$  represent the start and end points of the fitted function,  $\delta_0$  and  $\delta_1$  are the corresponding tangents at  $u = 0$  and  $u = 1$  and  $\mathbf{f}(\mathbf{u})$  contains basis functions of the cubic HS. The resulting HS  $p(u)$  is a linear combination of the four basis functions.

Richards (2020) proposed the HS model that smooths post-retirement mortality  $m_x$  in a given year using four third-order polynomials in a Hermite form,

$$\ln m_x = \alpha h_{00}(x_k) + \omega h_{01}(x_k) + s_0 h_{10}(x_k) + s_1 h_{11}(x_k), \quad (2)$$

<sup>1</sup>Cairns *et al.* (2011) investigated the prediction performance of Lee–Carter and CBD type models using mortality data of the elderly male population in England and Wales. They found that the Lee–Carter model produces implausibly small levels of uncertainty in mortality forecasts at advanced ages. For this reason, we do not consider the Lee–Carter model in our analysis which focuses on the modelling of higher age mortality. In this paper, the modelling variable is the log central death rates. For comparison purposes, we use the Gompertz model instead of the CBD model which works on the logit of death rates.

where

$$\begin{cases} h_{00}(x_k) = (1 + 2x_k)(1 - x_k)^2 \\ h_{01}(x_k) = x_k^2(3 - 2x_k) \\ h_{10}(x_k) = x_k(1 - x_k)^2 \\ h_{11}(x_k) = x_k^2(x_k - 1) \end{cases}, \tag{3}$$

$x_k = (x - x_0) / (x_1 - x_0)$  for a predetermined age interval  $[x_0, x_1]$  and  $x_k \in [0, 1]$ . There are four coefficients  $\alpha, \omega, s_0$  and  $s_1$  for the Hermite basis functions<sup>2</sup>  $h_{ij}(x_k), i = 0, 1, j = 0, 1$  to be estimated. Thereinto,  $\alpha$  and  $\omega$  correspond to  $p_0$  and  $p_1$  in (1) and reflect the level of  $\ln m_x$  at the two boundary ages  $x_0$  and  $x_1$ , respectively<sup>3</sup>.

We denote respectively  $D, E$  and  $M$  as the  $(n_a \times n_y)$  matrices of the death counts, central exposures and central death rates sorted by age and year, where the total number of ages (years) equals  $n_a(n_y)$ . It is assumed that the number of deaths is a random variable following a Poisson distribution. Suppose  $d, e$  and  $m$  are the vectors with a length of  $n_a$  in the above matrices corresponding to a given year. In a matrix form, one may write the HS model as

$$\ln m = X_0\beta, \tag{4}$$

where  $m$  is a vector of  $\ln m_x$  with a length of  $n_a = x_1 - x_0 + 1$ ,  $X_0$  refers to a  $(n_a \times 4)$  matrix containing the four Hermite basis functions at each age,  $\beta = (\alpha, \omega, s_0, s_1)'$  is the coefficient vector.

A smooth path between the youngest and oldest death rates in the interval is specified by  $s_0$  and  $s_1$  which give the gradients at the two boundary ages. As discussed in the original paper of the HS method, the fitted mortality curve may be too flexible and may deviate from the typical shape when both slope parameters are employed. Accordingly, four versions of the HS model can be constructed by setting one or both gradient coefficients to zero. The four models, namely HS1 ( $s_0 = s_1 = 0$ ), HS2 ( $s_1 = 0$ ), HS3 ( $s_0 = 0$ ) and HS4 can be defined by adjusting the coefficient vector  $\beta$ :

$$\beta' = \begin{cases} HS1: (\alpha, \omega, 0, 0) \\ HS2: (\alpha, \omega, s_0, 0) \\ HS3: (\alpha, \omega, 0, s_1) \\ HS4: (\alpha, \omega, s_0, s_1) \end{cases}.$$

As explained in section 1, the classical Gompertz law requires two parameters to model the decreasing mortality differentials over age and can potentially cause an unjustified crossover at advanced ages. On the other hand, the four HS models can ensure the convergence using only the parameter  $\alpha$ . In the work of Richards (2020), the initial mortality level  $\alpha^{(i)}$  for population  $i$  can be expressed as:

$$\alpha^{(i)} = \alpha^{(0)} + \sum_{j=1}^J \alpha^{(r_j)} z^{(ij)}, \tag{5}$$

where  $z^{(ij)}$  is an indicator variable which equals 1 if population  $i$  has risk factor  $j$  ( $j = 1, 2, \dots, J$ ) and 0 otherwise. When a population does not exhibit any of the  $J$  risk characteristics, the intercept of the fitted mortality curve is equivalent to the baseline value  $\alpha^{(0)}$ . The estimated mortality level at the starting age increases by  $\alpha^{(r_j)}$  for each additional risk factor  $j$ . By way of illustration, suppose that mortality data with the same gender of two states in the same country only differ in one risk factor – state. Under the simplest model HS1, their mortality differentials can be incorporated using the individual starting values  $\alpha^{(1)} = \alpha^{(0)}$  and  $\alpha^{(2)} = \alpha^{(0)} + \alpha^{(r_1)}$  and a common ending

<sup>2</sup>The basis functions in (3) are the factorised form of those presented in (1).

<sup>3</sup>When  $x = x_0, x_k = \frac{x_0 - x_0}{x_1 - x_0} = 0$  and  $\ln m_{x_0} = \alpha h_{00}(0) + \omega h_{01}(0) + s_0 h_{10}(0) + s_1 h_{11}(0) = \alpha(1 + 2 \times 0)(1 - 0) = \alpha$ .  
When  $x = x_1, x_k = \frac{x_1 - x_0}{x_1 - x_0} = 1$  and  $\ln m_{x_1} = \alpha h_{00}(1) + \omega h_{01}(1) + s_0 h_{10}(1) + s_1 h_{11}(1) = \omega \times 1^2 \times (3 - 2 \times 1) = \omega$ .

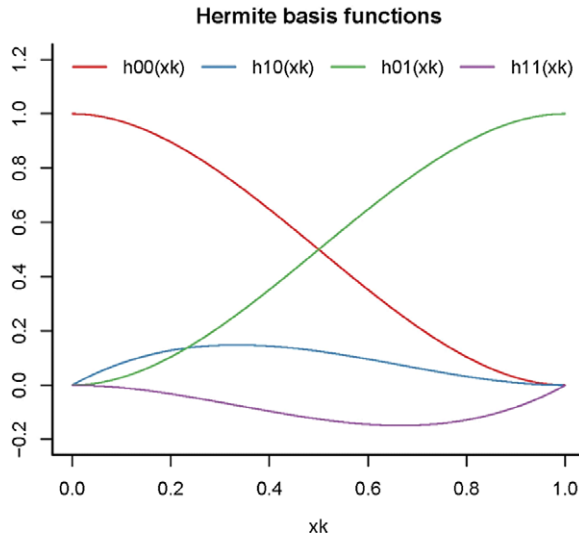


Figure 1. Hermite basis splines  $h_{ij}(x_k)$  for  $x_k$  ranging from 0 to 1,  $i = 0, 1$  and  $j = 0, 1$ .

value  $\omega$ . Under the compensation law of mortality, the two sets of state-level mortality rates (in a specific year) should converge gradually as age increases. As shown in Figure 1 that plots the basis functions of HSs,  $h_{00}(x_k)$  is monotonically decreasing over the defined age range. Consequently, the gap between the fitted mortality curves of the two populations vanishes over age by default, without the need of adopting a “slope” parameter.

In addition to the parsimonious modelling of mortality differentials, the potential crossover problem under the Gompertz model can be avoided due to the monotonicity of the Hermite function  $h_{00}(x_k)$ . Specifically, the difference between the log central death rates of the two populations at age  $x$  under the HS1 model is calculated as

$$\begin{aligned} \ln m_x^{(1)} - \ln m_x^{(2)} &= \left[ \alpha^{(1)} h_{00}(x_k) + \omega h_{01}(x_k) \right] - \left[ \alpha^{(2)} h_{00}(x_k) + \omega h_{01}(x_k) \right] \\ &= \left( \alpha^{(1)} - \alpha^{(2)} \right) h_{00}(x_k), \end{aligned} \tag{6}$$

where  $x_k = (x - x_0) / (x_1 - x_0)$ . Note that the two populations are assumed to share a common death rate  $\ln m_{x_1} = \omega$  at an advanced age  $x_1$ <sup>4</sup>. Since  $h_{00}(x_k)$  is a decreasing function ranging between 0 and 1, the difference in equation (6) narrows down to 0 as  $x$  reaches  $x_1$  but never switches its sign over the given range  $[x_0, x_1]$ . Nonetheless, one may argue that mortality differentials resulted from certain risk factors such as gender reduce at higher ages but do not vanish entirely (Tickle, 1997). That is, one population always has lower death rates than the other for the entire age range considered (assuming  $\alpha^{(1)} > \alpha^{(2)}$  and  $\omega^{(1)} > \omega^{(2)}$ ). To arrive at this pattern, a population-specific ending value  $\omega^{(i)}$  may be employed, producing mortality differentials as follows:

$$\begin{aligned} \ln m_x^{(1)} - \ln m_x^{(2)} &= \left[ \alpha^{(1)} h_{00}(x_k) + \omega^{(1)} h_{01}(x_k) \right] - \left[ \alpha^{(2)} h_{00}(x_k) + \omega^{(2)} h_{01}(x_k) \right] \\ &= \left( \alpha^{(1)} - \alpha^{(2)} \right) h_{00}(x_k) + \left( \omega^{(1)} - \omega^{(2)} \right) h_{01}(x_k). \end{aligned} \tag{7}$$

<sup>4</sup>No consensus has been reached on the existence of a limiting mortality level. Gampe (2010) found that the human force of mortality becomes a flat curve after age 110. Nevertheless, some research argued that such plateau is a consequence of demographic errors such as unreported deaths and the pooling of cohorts (Black *et al.*, 2017; Newman, 2018).

As illustrated in Figure 1,  $h_{00}(x_k)$  is monotonically decreasing, while  $h_{01}(x_k)$  is an increasing function. Besides, they are symmetric about the line  $h(x_k) = 0.5$  ( $h_{00}(x_k) + h_{01}(x_k) = 1$ ). It implies that when  $(\alpha^{(1)} - \alpha^{(2)}) = (\omega^{(1)} - \omega^{(2)})$ , the sum of the two weighted functions in (7) becomes a horizontal line. Since  $(\alpha^{(1)} - \alpha^{(2)})$  and  $(\omega^{(1)} - \omega^{(2)})$  refer to the differences in the log mortality rates between the two populations at the youngest and oldest ages, respectively, one would expect the former to be more pronounced than the latter due to the compensation law of mortality (Gavrilov & Gavrilova, 2001). Accordingly,  $h_{00}(x_k)$  contributes more to the weighted sum in equation (7), leading to declining but non-zero differentials. Again, the no-crossover property can be guaranteed, as the weighted sum of the two Hermite basis functions does not change in sign over the domain interval. The other three HS models also possess this property if one sets common gradient parameters  $s_0$  and  $s_1$  between related populations<sup>5</sup>.

The above HS models were initially proposed to model mortality over the age range rather than over time. Next, we show that the HS approach can be extended to model two-dimensional population mortality data. Similar to the widely adopted CBD model, the HS structure is fitted to the mortality data of each year. By doing so, the estimated values of each model parameter over different years form a time series, based on which the future death rates can be projected or simulated. The modified versions of the HS models are given below:

$$\ln M = XB = (I_{n_y} \otimes X_0) B, \tag{8}$$

where  $I_{n_y}$  is an identity matrix of size  $n_y$ ,  $X_0$  has a dimension of  $(n_a \times 4)$  and comprises the Hermite basis functions at each age, the design matrix  $X$  is the Kronecker product of  $I_{n_y}$  and  $X_0$ , the coefficient vector  $B = (\beta'_{t_0}, \beta'_{t_0+1}, \dots, \beta'_{t_1})'$  concatenates the year-specific HS coefficients  $\beta_t$ . Again, the four HS versions can be obtained by adjusting  $\beta_t$ . The modified versions of the HS models are given below:

$$\begin{cases} \text{HS1: } \ln m_{x,t} = \alpha_t h_{00}(x_k) + \omega_t h_{01}(x_k) \\ \text{HS2: } \ln m_{x,t} = \alpha_t h_{00}(x_k) + \omega_t h_{01}(x_k) + s_{0,t} h_{10}(x_k) \\ \text{HS3: } \ln m_{x,t} = \alpha_t h_{00}(x_k) + \omega_t h_{01}(x_k) + s_{1,t} h_{11}(x_k) \\ \text{HS4: } \ln m_{x,t} = \alpha_t h_{00}(x_k) + \omega_t h_{01}(x_k) + s_{0,t} h_{10}(x_k) + s_{1,t} h_{11}(x_k) \end{cases}$$

Similar to the one-dimensional case,  $\alpha_t$  and  $\omega_t$  can be interpreted as the fitted values of  $\ln m_{x_0,t}$  and  $\ln m_{x_1,t}$ . The other two parameters  $s_{0,t}$  and  $s_{1,t}$  specify the smoothed pathway between the two mortality rates at the two boundary ages in year  $t$ .

When the time dimension is included in modelling mortality data, avoiding any mortality crossover between related populations is not straightforward. Particularly, when the sample period covers more distant years, the observed mortality rates at the oldest ages are more volatile, and the data themselves may exhibit some crossover. Such features would not be helpful in determining the fitted mortality surface. We propose one possible solution to cope with this issue by imposing constraints on the second-level parameter  $\omega_t$ . Consider males (population 1) and females (population 2) in the same country, when the two populations are assumed to share (only) the common gradient parameters  $s_{0,t}$  and  $s_{1,t}$ , for year  $t$ , the initial mortality differential  $(\alpha_t^{(1)} - \alpha_t^{(2)})$  should decrease to  $(\omega_t^{(1)} - \omega_t^{(2)})$  without changing its sign over the age range considered. This property is true if  $(\alpha_t^{(1)} - \alpha_t^{(2)}) (\omega_t^{(1)} - \omega_t^{(2)}) > 0$ . Given  $\alpha_t^{(1)} > \alpha_t^{(2)}$  for all  $t$ ,  $\omega_t^{(1)} > \omega_t^{(2)}$  is required to avoid the potential crossover. In other words, the female population with lower mortality rates at the starting age is expected to maintain this ‘‘advantage’’ at the ending age. As we will discuss in

<sup>5</sup>Common gradient parameters imply that the fitted curves exhibit similar paths of leaving and approaching the start and end values of mortality rates, which is a sensible assumption for associated populations.

more detail later, the maximum likelihood estimation with the Newton–Raphson iterative scheme is used to calibrate the model parameters. Nevertheless, the inequality constraints are not guaranteed to be satisfied during the estimation procedure. Accordingly, we adopt the barrier method (Wright & Nocedal, 1999) under which the inequality constraints are incorporated using the so-called barrier functions that serve as a penalty term. It has been applied in mortality modelling by Li & Liu (2020, 2021) to constrain their heat wave model parameters. As the parameter estimates approach the inequality bounds, the magnitude of the barrier functions rises to infinity, and so the log-likelihood is penalised more by such estimates. Then the optimised parameters are “forced” to fall within the boundaries.

For ease of exposition, consider an optimisation problem searching for parameter values that maximise a function  $f(\theta)$ , subject to the constraint  $\theta > a$ . Under the barrier method, the optimisation problem is redesigned as maximising the penalised function  $f(\theta) - r\left(\frac{1}{\theta-a}\right)$ , where  $B(\theta) = r\left(\frac{1}{\theta-a}\right)$  refers to the barrier function with  $r > 0$ . Given a feasible initial value that over-satisfies the constraint, as  $\theta$  decreases to its lower bound  $a$ , the barrier function tends to infinity, which then contradicts the aim of maximisation. Accordingly, this formulation would avoid the violation of the parameter restriction. Similarly, the barrier functions for the constraints  $\theta < b$  and  $a < \theta < b$  can be constructed as  $r\left(\frac{1}{b-\theta}\right)$  and  $r\left(\frac{1}{b-\theta} + \frac{1}{\theta-a}\right)$ , respectively.

Setting the objective function to the Poisson log-likelihood function, multiple barrier functions can be embedded to manage the inequality constraints imposed on the population-specific  $\omega_t^{(i)}$ . For example, when the male and female populations in the same country are considered, the following constraints are imposed:

$$\begin{cases} \omega_t^{(1)} > a_t^{(1,2)} \\ \omega_t^{(2)} < a_t^{(1,2)} \end{cases},$$

where  $\omega_t^{(1)}$  and  $\omega_t^{(2)}$  are the male and female level parameters at the ending age, respectively,  $a_t^{(1,2)}$  is calculated as the average of the initial values of  $\omega_t^{(1)}$  and  $\omega_t^{(2)}$ . Under the careful design of the barrier method, the no-crossover condition  $\omega_t^{(1)} > \omega_t^{(2)}$  can be met. The specifications of the barrier functions, initial values, penalised log-likelihood function and the updating equations for the HS4 model are provided in Appendix A. As we will show in section 3.2, the above barrier constraints can be applied similarly to risk factors other than gender.

### 2.2 Gompertz model

Gompertz (1825) discovered that the age-specific hazard rate of adult mortality increases exponentially with age and proposed a two-parameter model

$$\mu_x = Bc^x$$

or equivalently

$$\ln \mu_x = k_1 + k_2x. \tag{9}$$

Since this discovery, actuaries and researchers have developed various modifications to the original design<sup>6</sup>. For instance, Makeham (1867) added an age-independent term to the above model for the age-specific hazard to account for deaths that do not result from senescence. Besides including more parameters, adjustments were made to the format of the original function. The

<sup>6</sup>See Forfar (2014) for a summary of the evolution of mortality laws.



so-called Perks model (Perks, 1932) with four parameters adopted a logistic function, which is described as follows:

$$\mu_x = \frac{A + Bc^x}{1 + Dc^x}.$$

The well-known CBD model (Cairns *et al.*, 2006) can be obtained by replacing the hazard rate with the death rate and setting  $A = 0$  and  $B = D$  under the Perks law and adding a time dimension. Specifically, the probability of death  $q_{x,t}$  is modelled as:

$$q_{x,t} = \frac{e^{k_{1,t} + k_{2,t}x}}{1 + e^{k_{1,t} + k_{2,t}x}}$$

or equivalently

$$\text{logit } q_{x,t} = \ln \left( \frac{q_{x,t}}{1 - q_{x,t}} \right) = k_{1,t} + k_{2,t}x.$$

It is evident that the CBD model can be regarded as an application of the Gompertz law (equation (9)) applied to a transformed measure, in modelling population mortality data with both the age and time dimensions. This model has been widely applied and extended in the literature and has been found particularly suitable for explaining mortality rates at higher ages (Cairns *et al.*, 2009).

For comparison purposes, we utilise the logarithm of the central death rate rather than the logit probability of death for the two-dimensional Gompertz law in this paper. The log central death rate  $\ln m_{x,t}$  of a life aged  $x$  in year  $t$  is expressed as

$$\ln m_{x,t} = k_{1,t} + k_{2,t}x, \quad (10)$$

where  $k_{1,t}$  refers to the level of the mortality curve in year  $t$  and  $k_{2,t}$  describes the speed at which mortality increases for each extra year of age in year  $t$ . The fitted log central death rates in a given year are assumed to grow linearly with age.

### 2.3 PS model

The two mortality models introduced above make assumptions about the functional form of mortality rates and only impose smoothness across adjacent ages. However, some academics are sceptical about specifying the mortality surface using defined functions. To address this concern, one branch of mortality models originated from the spline fitting technique that smooths mortality rates only based on features of data and avoids the identification of the model structure.

One of the most commonly applied methods in the actuarial literature is the PS model (Currie *et al.*, 2004). The concept of PS was initially proposed by Eilers & Marx (1996). They applied difference penalties on the coefficients of B-splines to strike a balance between the goodness of fit and smoothness of the fitted curve. This one-dimensional spline fitting approach was extended to a bivariate regression model in Eilers & Marx (2003) and Eilers *et al.* (2006). Using the same notation as in section 2.1,  $\mathbf{D}$ ,  $\mathbf{E}$  and  $\mathbf{M}$  represent the matrices of the death counts, central exposures and central death rates, and  $\mathbf{d}$ ,  $\mathbf{e}$  and  $\mathbf{m}$  are the vector components corresponding to a given year  $t$ .

The log death counts in year  $t$  can be modelled as:

$$\ln \mathbf{d} = \ln \mathbf{e} + \ln \mathbf{m} = \ln \mathbf{e} + \mathbf{B}\mathbf{a}, \quad (11)$$

where  $\mathbf{B}$  is the  $(n_a \times k)$  B-spline regression matrix containing polynomials connected at equally spaced knots, and  $\mathbf{a}$  is the  $(k \times 1)$  coefficient vector. The column dimension  $k$  of  $\mathbf{B}$  is equivalent to the sum of the polynomial degree  $q$  and the number of (internal) knots  $n_{d_x}$  (i.e.,  $k = q + n_{d_x}$ ). The more knots the regression employs, the more flexible the fitted curve is. One may adopt a

sufficient number of knots for B-splines to provide more flexibility than needed and avoid over-fitting by constructing a difference penalty term that imposes smoothness (Eilers & Marx, 2002). Accordingly, the coefficient vector  $a$  can be estimated by maximising the penalised log-likelihood

$$l_p = l(a; d) - \frac{1}{2} a' P a,$$

where  $l(a; d)$  refers to the log-likelihood of Poisson distribution and the penalty matrix  $P = \lambda \Delta' \Delta$  is composed of the smoothing parameter  $\lambda$  and difference matrix  $\Delta$  with order 2.

The PS approach can also be applied to fitting the mortality surface across ages and years. We denote the  $(n_a \times k_a)$  and  $(n_y \times k_y)$  B-spline basis matrices associated with age  $x$  and year  $t$  by  $B_a$  and  $B_y$ , respectively. The number of equally spaced knots can be determined separately for the age and year horizons based on the data range. Then the B-spline matrix in the two-dimensional setting is constructed as the Kronecker product

$$B^{(2)} = B_y \otimes B_a.$$

Given mortality data across both ages and years, the PS model is expressed as:

$$\ln D = \ln E + \ln M = \ln E + B^{(2)} a = \ln E + B_a A B_y', \tag{12}$$

where  $A$  is the  $(k_a \times k_y)$  matrix obtained by rearranging the coefficient vector  $a$  with a length of  $k_a k_y$ . The parameters in the PS model are estimated by maximising the same Poisson log-likelihood function using a smoothing matrix that penalises the two dimensions of  $A$  separately. The penalty matrix is given as

$$P = \lambda_a (I_{k_y} \otimes \Delta'_a \Delta_a) + \lambda_y (\Delta'_y \Delta_y \otimes I_{k_a}),$$

where  $I_{k_y}$  ( $I_{k_a}$ ),  $\lambda_a$  ( $\lambda_y$ ) and  $\Delta_a$  ( $\Delta_y$ ) are the identity matrix, smoothing parameter and difference matrix contributing to the columns (rows) of  $A$ .

We have briefly reviewed the theoretical background of the PS method, more technical details and improvements on the fitting algorithm can be found in Currie *et al.* (2004, 2006) and Eilers *et al.* (2006).

### 3. Fitting

#### 3.1 Single-population fitting

To fit the above three models, we collect population mortality data of Australia, England and Wales, France, Japan and the United States from the Human Mortality Database (HMD, 2021). Specifically, the death counts and exposure to risk of both genders from 1950 to 2018 over ages 56–95 are employed for this analysis. Following Brouhns *et al.* (2002), we assume that the number of deaths is a random variable following a Poisson distribution with the mean equal to the expected number of deaths. This Poisson assumption enables a rigorous framework to estimate the model parameters by maximising the log-likelihood function

$$l = \ln L = \sum_{x,t} \left( d_{x,t} \ln \hat{d}_{x,t} - \hat{d}_{x,t} - \ln(d_{x,t}!) \right), \tag{13}$$

where  $d_{x,t}$  represents the observed death counts. The fitted number of deaths  $\hat{d}_{x,t}$  is calculated as the product of the observed exposures and fitted mortality rates  $(\hat{d}_{x,t} = E_{x,t} \hat{m}_{x,t})$ . As an illustration, the fitted death count for lives aged  $x$  in year  $t$  under the HS4 model is  $\hat{d}_{x,t} = E_{x,t} \hat{m}_{x,t} = E_{x,t} \exp(\alpha_t h_{00}(x_k) + \omega_t h_{01}(x_k) + s_{0,t} h_{10}(x_k) + s_{1,t} h_{11}(x_k))$ . For a parameter  $\theta$ ,

the Newton–Raphson method is adopted to update its estimate iteratively as  $\theta^* = \theta - \frac{\partial l}{\partial \theta} / \frac{\partial^2 l}{\partial \theta^2}$ .

We provide below a summary of estimation steps under the HS4 model.

- (1) Assign initial values to the parameters under the HS4 model and compute using the initial estimates. Specifically, we set  $\hat{s}_{0,t} = \hat{s}_{1,t} = 0$ . The initial values of the other two parameters  $\alpha_t$  and  $\omega_t$  are set to the observed mortality levels at the two boundary ages  $x_0$  and  $x_1$  of the sample range in year  $t$  (i.e.,  $\hat{\alpha}_t = \ln m_{x_0,t}$  and  $\hat{\omega}_t = \ln m_{x_1,t}$ ). For our sample,  $x_k = \frac{x - x_0}{x_1 - x_0} = \frac{x - 56}{95 - 56}$ .
- (2) Update  $\alpha_t$  as  $\hat{\alpha}_t^* = \hat{\alpha}_t + \frac{\sum_x (d_{x,t} - \hat{d}_{x,t}) h_{00}(x_k)}{\sum_x \hat{d}_{x,t} [h_{00}(x_k)]^2}$ .
- (3) Update  $\omega_t$  as  $\hat{\omega}_t^* = \hat{\omega}_t + \frac{\sum_x (d_{x,t} - \hat{d}_{x,t}) h_{01}(x_k)}{\sum_x \hat{d}_{x,t} [h_{01}(x_k)]^2}$ .
- (4) Update  $s_{0,t}$  as  $\hat{s}_{0,t}^* = \hat{s}_{0,t} + \frac{\sum_x (d_{x,t} - \hat{d}_{x,t}) h_{10}(x_k)}{\sum_x \hat{d}_{x,t} [h_{10}(x_k)]^2}$ .
- (5) Update  $s_{1,t}$  as  $\hat{s}_{1,t}^* = \hat{s}_{1,t} + \frac{\sum_x (d_{x,t} - \hat{d}_{x,t}) h_{11}(x_k)}{\sum_x \hat{d}_{x,t} [h_{11}(x_k)]^2}$ .
- (6) Calculate the log-likelihood function  $l = \sum_{x,t} (x_t \ln \hat{d}_{x,t} - \hat{d}_{x,t} - \ln (d_{x,t}!))$  using the parameters estimated in (2)–(5).
- (7) Repeat steps (2)–(6) until the improvement in  $l$  is small enough, e.g., the change in  $l$  between two iterations is less than  $10^{-7}$ .

This study calibrates the HS and Gompertz models using the Poisson updating approach described above, while the PS model is fitted and forecasted using an R package “MortalitySmooth<sup>7</sup>”. Both fitting methods are based on the Poisson maximum likelihood estimation, while the latter uses a so-called generalised linear array modelling that acts on matrices (Camarda, 2012). To examine the goodness of fit of the candidates, we calculate two measures – Akaike information criterion (AIC) and Bayesian information criterion (BIC) – as below:

$$AIC = -2l + n_p \times 2$$

$$BIC = -2l + n_p \ln n_d,$$

where  $l$  refers to the Poisson log-likelihood function, and  $n_p$  is the number of effective parameters. These two criteria penalise the likelihood by the number of parameters involved in the model with a penalty factor of 2 and the log of data counts ( $n_d$ ), respectively. The information criteria and residuals plots are provided in Table 1 and Figure 2, respectively.

Some interesting remarks can be made from Table 1 that lists the model comparison criteria values for both genders of the ten datasets. Firstly, HS1-3 models always produce greater AICs and BICs than the full HS4 model. Note that  $\alpha_t$  and  $\omega_t$  are interpreted as the mortality rates at the two boundaries of the age interval. Essentially, the fitted mortality curves in each year under the HS1 model are assumed to have a shape determined by the two Hermite functions linking the starting and ending points. Such a specification may be too rigid for modelling population mortality data,

<sup>7</sup>The R package allows the user to determine the smoothing parameters based on different criteria. In this study, we use the “optimal” smoothing parameters that lead to the lowest Bayesian information criterion (BIC) score. Two knot choices are considered for the PS method – one internal knot per 10 data points and one internal knot per 3 data points in each dimension. Furthermore, the default order of differences  $d = 2$  is adopted to form a linear extrapolation of the coefficients of the cubic ( $q = 3$ ) B-splines. Note that the number of effective parameters under the PS method depends on the estimated penalty parameter, which can vary with datasets.

**Table 1.** Effective number of parameters ( $n_p$ ), log-likelihood values ( $l$ ), AIC and BIC values under the HS1-4, Gompertz, PS10 and PS3 models. The mortality models are ranked from 1 (the lowest AIC/BIC) to 6 (the highest AIC/BIC) and the ranks are given in brackets.

Model	HS1	HS2	HS3	HS4	Gompertz	PS10	PS3
<b>Australia</b>							
$n_p$	138	207	207	276	138	36	86
$l$	-36,565	-26,575	-30,465	-14,597	-17,285	-15,378	-14,974
AIC	73,406 (7)	53,564 (5)	61,344 (6)	<b>29,746</b> (1)	34,847 (4)	30,829 (3)	30,121 (2)
BIC	74,224 (7)	54,790 (5)	62,570 (6)	31,380 (3)	35,664 (4)	31,044 (2)	<b>30,632</b> (1)
<b>England and Wales</b>							
$n_p$	138	207	207	276	138	59	170
$l$	-123,915	-61,640	-105,073	-18,807	-30,206	-21,662	-19,713
AIC	248,106 (7)	123,694 (5)	210,560 (6)	<b>38,166</b> (1)	60,687 (4)	43,444 (3)	39,766 (2)
BIC	248,923 (7)	124,920 (5)	211,786 (6)	<b>39,801</b> (1)	61,504 (4)	43,796 (3)	40,774 (2)
<b>France</b>							
$n_p$	138	207	207	276	138	59	139
$l$	-100,986	-80,258	-50,664	-18,648	-54,468	-20,160	-18,574
AIC	202,247 (7)	160,930 (6)	101,742 (4)	37,848 (2)	109,213 (5)	40,438 (3)	<b>37,425</b> (1)
BIC	203,064 (7)	162,156 (6)	102,968 (4)	39,483 (2)	110,030 (5)	40,787 (3)	<b>38,248</b> (1)
<b>Japan</b>							
$n_p$	138	207	207	276	138	62	176
$l$	-154,116	-100,804	-98,878	-21,985	-46,949	-23,790	-21,090
AIC	308,508 (7)	202,023 (6)	198,170 (5)	44,521 (2)	94,174 (4)	47,703 (3)	<b>42,531</b> (1)
BIC	309,325 (7)	203,249 (6)	199,396 (5)	46,156 (2)	94,991 (4)	48,070 (3)	<b>43,573</b> (1)
<b>United States</b>							
$n_p$	138	207	207	276	138	62	264
$l$	-329,226	-253,842	-233,229	-31,095	-85,917	-31,663	-24,724
AIC	658,729 (7)	508,099 (6)	466,871 (5)	62,742 (2)	172,111 (4)	63,450 (3)	<b>49,975</b> (1)
BIC	659,546 (7)	509,325 (6)	468,097 (5)	64,377 (3)	172,928 (4)	63,816 (2)	<b>51,538</b> (1)
<b>(a) Males</b>							
Model	HS1	HS2	HS3	HS4	Gompertz	PS10	PS3
<b>Australia</b>							
$n_p$	138	207	207	276	138	38	87
$l$	-40,171	-35,134	-21,569	-14,534	-19,633	-15,319	-14,844
AIC	80,618 (7)	70,683 (6)	43,552 (5)	<b>29,620</b> (1)	39,542 (4)	30,714 (3)	29,861 (2)
BIC	81,436 (7)	71,909 (6)	44,778 (5)	31,255 (3)	40,360 (4)	30,940 (2)	<b>30,375</b> (1)
<b>England and Wales</b>							
$n_p$	138	207	207	276	138	54	172
$l$	-116,748	-99,374	-63,802	-18,822	-31,094	-22,867	-20,371
AIC	233,772 (7)	199,163 (6)	128,019 (5)	<b>38,196</b> (1)	62,464 (4)	45,841 (3)	41,087 (2)
BIC	234,589 (7)	200,389 (6)	129,245 (5)	<b>39,831</b> (1)	63,281 (4)	46,162 (3)	42,108 (2)
<b>France</b>							
$n_p$	138	207	207	276	138	46	138
$l$	-158,050	-107,341	-32,427	-17,801	-73,551	-21,559	-19,728

Table 1. Continued.

(a) Males														
Model	HS1		HS2		HS3		HS4		Gompertz		PS10		PS3	
AIC	316,376	(7)	215,095	(6)	65,268	(4)	<b>36,155</b>	(1)	147,377	(5)	43,209	(3)	39,733	(2)
BIC	317,193	(7)	216,321	(6)	66,494	(4)	<b>37,790</b>	(1)	148,195	(5)	43,478	(3)	40,553	(2)
Japan														
$n_p$	138		207		207		276		138		62		160	
$l$	-249,457		-170,454		-55,538		-19,591		-86,174		-23,169		-21,428	
AIC	499,190	(7)	341,323	(6)	111,491	(4)	<b>39,733</b>	(1)	172,623	(5)	46,462	(3)	43,175	(2)
BIC	500,008	(7)	342,549	(6)	112,717	(4)	<b>41,368</b>	(1)	173,441	(5)	46,831	(3)	44,120	(2)
United States														
$n_p$	138		207		207		276		138		63		264	
$l$	-451,804		-414,642		-161,650		-31,941		-115,397		-35,060		-25,586	
AIC	903,883	(7)	829,698	(6)	323,713	(5)	64,434	(2)	231,070	(4)	70,245	(3)	<b>51,699</b>	(1)
BIC	904,701	(7)	830,924	(6)	324,939	(5)	66,069	(2)	231,887	(4)	70,617	(3)	<b>53,260</b>	(1)
(b) Females														

which is vividly demonstrated by the worst performance of the HS1 model. Employing additional gradient factors can introduce more flexibility to the model and so can improve the goodness-of-fit, which can be seen from the better ranks of the HS2-4 models.

Comparing between the HS2 and HS3 models with one gradient factor  $s_{0,t}$  and  $s_{1,t}$ , respectively, the former only outperforms the latter for males in Australia and England and Wales. Since  $s_{0,t}$  ( $s_{1,t}$ ) controls the pattern of the fitted mortality curves leaving from (approaching) the starting (ending) age, one may suspect that there have been more variations in the mortality development of older age groups. Under the full HS model, employing two gradient factors yields significantly better fitting performance than the other HS candidates. As a benchmark, the Gompertz model is specified by two time-varying components and gives smaller AIC and BIC values than the HS1-2 models in all cases. However, it consistently underperforms the HS4 model (and HS3 model in some cases).

The above five models all propose a functional form for mortality rates in the age dimension, while no smoothness assumption across periods is imposed. On the other hand, the PS approach produces a smoothed mortality surface using splines connected by knots. We consider the PS10 and PS3 models with a different number of knots. Specifically, the PS10 (PS3) refers to the PS model with one internal knot per 10 (3) data points in each dimension. The PS3 model tends to be the optimal choice among the seven candidates for males in four out of the five countries, while the HS4 model appears to be preferred for female populations. When fewer internal knots are adopted, the PS10 model tends to underperform the HS4 model but still performs better than the Gompertz model in all cases.

Figure 2 depicts the deviance residuals against age, period and cohort year for Australia. Only the HS4, Gompertz, PS10 and PS3 models, which have similar AIC/BIC values, are presented. It is apparent that the residuals are generally quite random under the HS4 and two PS models. However, those produced by the Gompertz method do exhibit obvious patterns, which may suggest the need for more parameters. Moreover, the residuals plotted against age under the four models (and the residuals against year under the PS models) show some “bars” at certain ages (in certain years). Since the HS4 and Gompertz models do not assume smoothed mortality rates in the period dimension, the time-varying parameters would capture the pattern in each year separately and lead to more scattered residuals against year. This difference is a direct consequence of different smoothness assumptions between the mortality models.

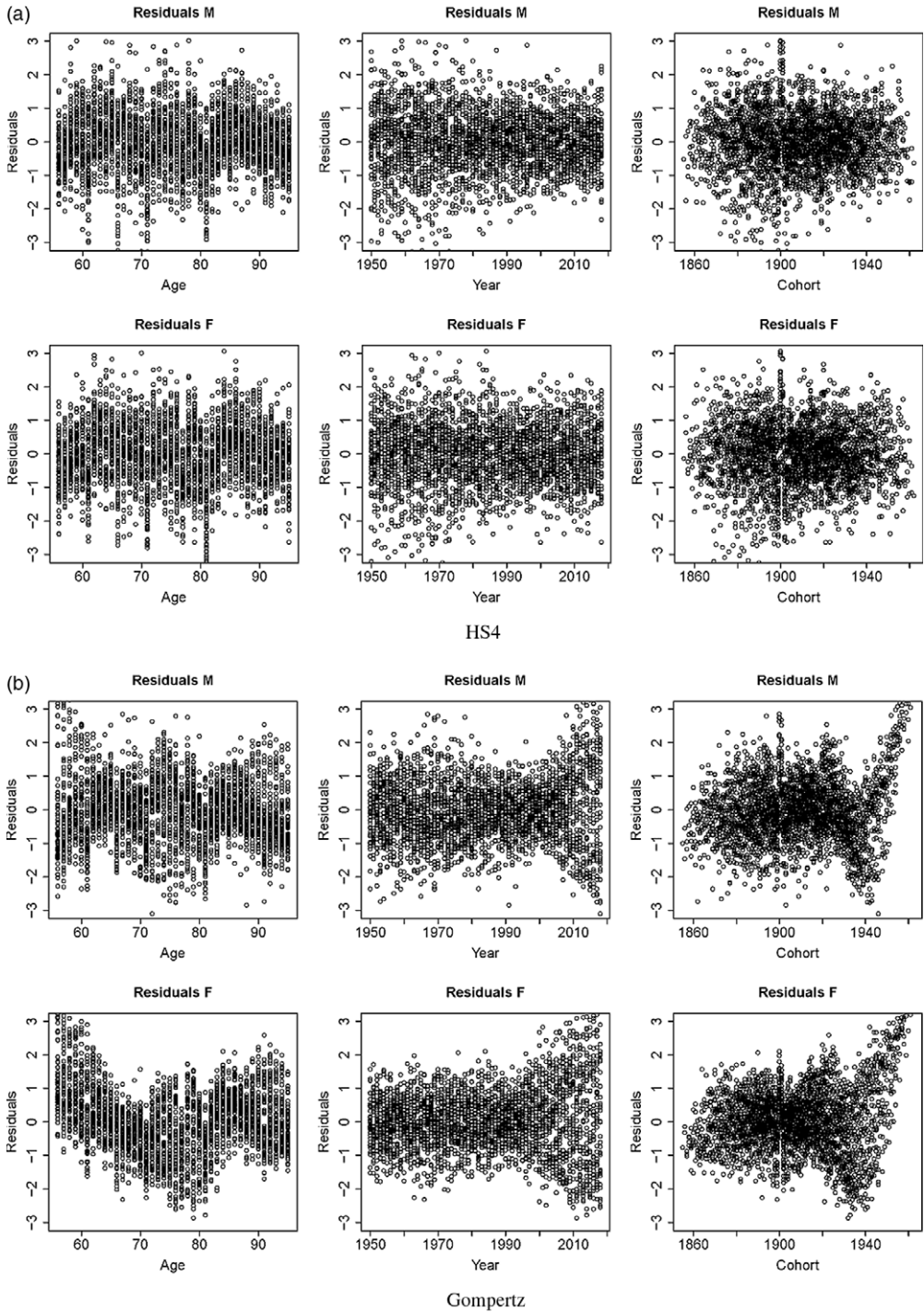


Figure 2. Standardised deviance residuals plots under the HS4, Gompertz, PS10 and PS3 models for Australian males (top panel) and females (bottom panel).

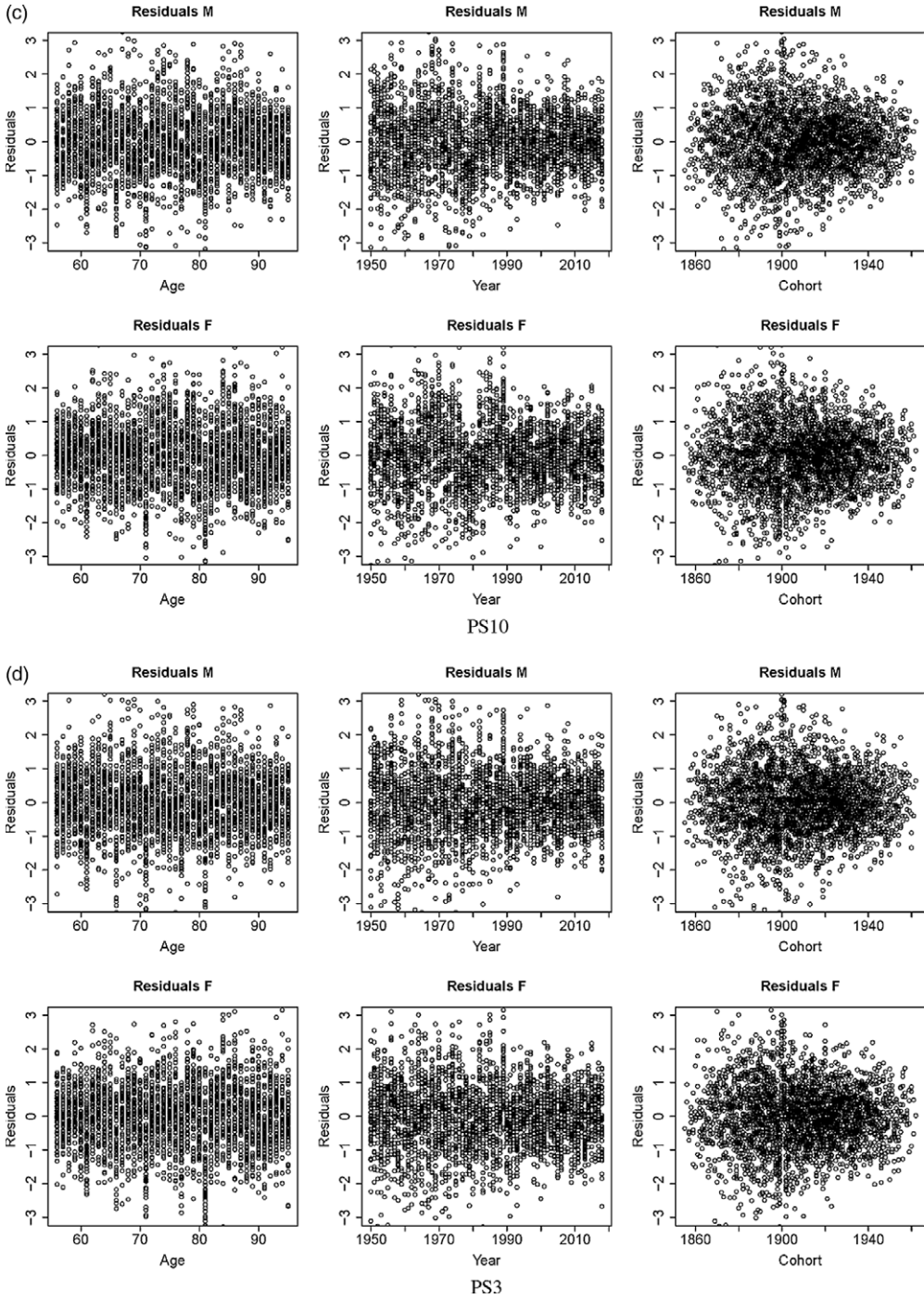


Figure 2. Continued.

Based on these observations, one can see that imposing smoothness on both the age and time dimensions can lead to missing out some important patterns, though having the advantage of a smaller number of effective parameters. For those mortality models that only impose the smoothness assumption on the age dimension (HS and Gompertz models), more parameters are used to describe the temporal developments more adequately. It is noteworthy that an outstanding fitting performance does not guarantee accurate forecasting. As we will see later, the forecasting performance under the PS approach is subject to great uncertainty, though the fitting results are superior. Backtesting will be conducted and presented in section 5 to investigate the forecast accuracy of the mortality models.

### 3.2 No-crossover property of the HS models

As discussed in section 2.1, the key strength of the HS approach is that mortality crossover is avoided via the properties of Hermite basis functions. Mortality differentials due to risk factors other than gender decrease as age increases, portrayed by the level parameter  $\alpha_t$  specifically for each population. When the gender effect is considered, mortality differentials diminish with increasing age but never vanish entirely. By imposing constraints on the level parameter specifically for each population  $\left\{ \begin{array}{l} \omega_t^{(1)} > a_t^{(1,2)} \\ \omega_t^{(2)} < a_t^{(1,2)} \end{array} \right.$ , the HS models can capture such a feature even if there is occasional mortality crossover in the data. In contrast, the Gompertz and PS models are incapable of modelling this pattern.

As an illustration of this crossover issue, we choose the mortality data for ages between 56 and 108<sup>8</sup> and years from 1950 to 2018. Figure 3 presents the observed and fitted log central death rates for the United States in 2018 under the six models. The PS approach here adopts one internal knot per 10 data points in each dimension to alleviate the potential crossover issue caused by overfitting. It can be seen that the HS1-3 models tend to underestimate the mortality rates at younger ages. The constructed shape of the HS1 and HS2 curves fits poorly at the oldest ages. On the other hand, the Gompertz model proposes a linear mortality curve of age and fails to capture the curvature observed at the highest ages. By comparison, the fitted curves under the HS4 and PS models follow the historical patterns more closely. However, the curves fitted under the PS model adhere to the data very closely, potentially indicating overfitting and lack of smoothness, even after applying the penalty structure. Another concern from Figure 3 is the intersection of the fitted female and male central death rates under the Gompertz and PS models at advanced ages. Under the compensation law of mortality, one would expect the gap between two genders to decline gradually over increasing age. But there is no biological basis to justify any crossover. Overall, the four HS models can generate a smoothed mortality curve that captures the relationships between males and females in the same country. Given appropriate constraints imposed on  $\omega_t^{(i)}$ , the no-crossover property can be ensured.

In addition to gender, the no-crossover property of the HS method holds when more risk factors are considered. As explained in equation (7) and footnote 6, when common gradient parameters are assumed for associated populations, the fitted mortality curves from the HS models would demonstrate narrowed gaps without an unreasonable crossover. This assumption should only be imposed if one believes that the populations under study exhibit similar paths of reaching the boundary mortality rates over the age interval. We have illustrated that the HS4 model tends to produce a good fit for mortality rates of males and females in the United States. Next, the fitting performances of the HS4, Gompertz and PS models are investigated when two risk

<sup>8</sup>This age interval is applied to demonstrate the possibility of mortality crossover under the Gompertz and PS models. But since mortality rates at advanced ages tend to be more volatile, we exclude ages above 95 when examining model performance with AIC and BIC.



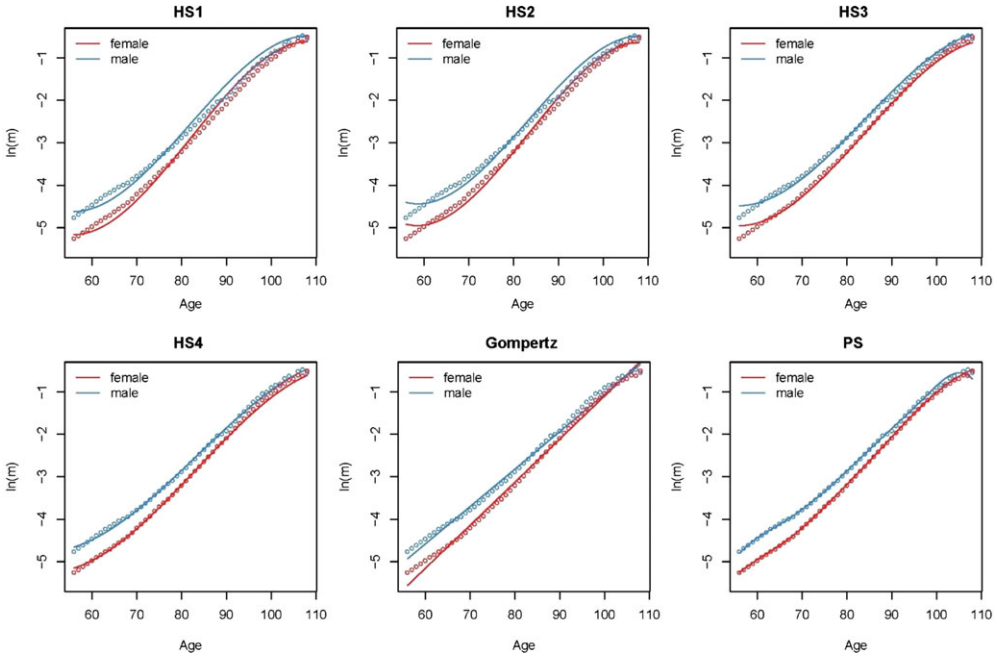


Figure 3. Historical (dots) and fitted (solid lines) log central death rates in 2018 for the United States.

factors – gender and country – are employed. Since the five countries considered in this study may not be highly related, the common-gradient assumption can be inappropriate. Accordingly, mortality data between 2008 and 2018 of females and males in Scotland and England and Wales are used for this particular illustration<sup>9</sup>. Based on the observations on  $\alpha_t^{(i)}$  of the four populations, the following restriction is required for the HS4 model to avoid the potential crossover:

$$\omega_t^{(m2)} > \omega_t^{(m1)} > \omega_t^{(f2)} > \omega_t^{(f1)},$$

where  $\omega_t^{(m1)}$  and  $\omega_t^{(f1)}$  ( $\omega_t^{(m2)}$  and  $\omega_t^{(f2)}$ ) represent the level parameter of the female and male populations in England and Wales (Scotland). To conform with the constraint format of the barrier method, the above inequality can be converted into:

$$\begin{cases} \omega_t^{(m2)} > a_t^{(1,2)} \\ a_t^{(2,3)} < \omega_t^{(m1)} < a_t^{(1,2)} \\ a_t^{(3,4)} < \omega_t^{(f2)} < a_t^{(2,3)} \\ \omega_t^{(f1)} < a_t^{(3,4)} \end{cases},$$

where  $a_t^{(jj+1)}$  is computed as the average of the  $j^{th}$  and  $(j + 1)^{th}$  ( $j = 1, 2, 3$ ) highest initial values of the level parameter at the ending age. Note that other relationships among  $\omega_t^{(i)}$ 's can be assumed. For example, if one believes that populations with the same gender will converge to

<sup>9</sup>For the PS model, we use one internal knot per 10 ages and per 3 years (instead of 10), as only 10 years of mortality data are considered.

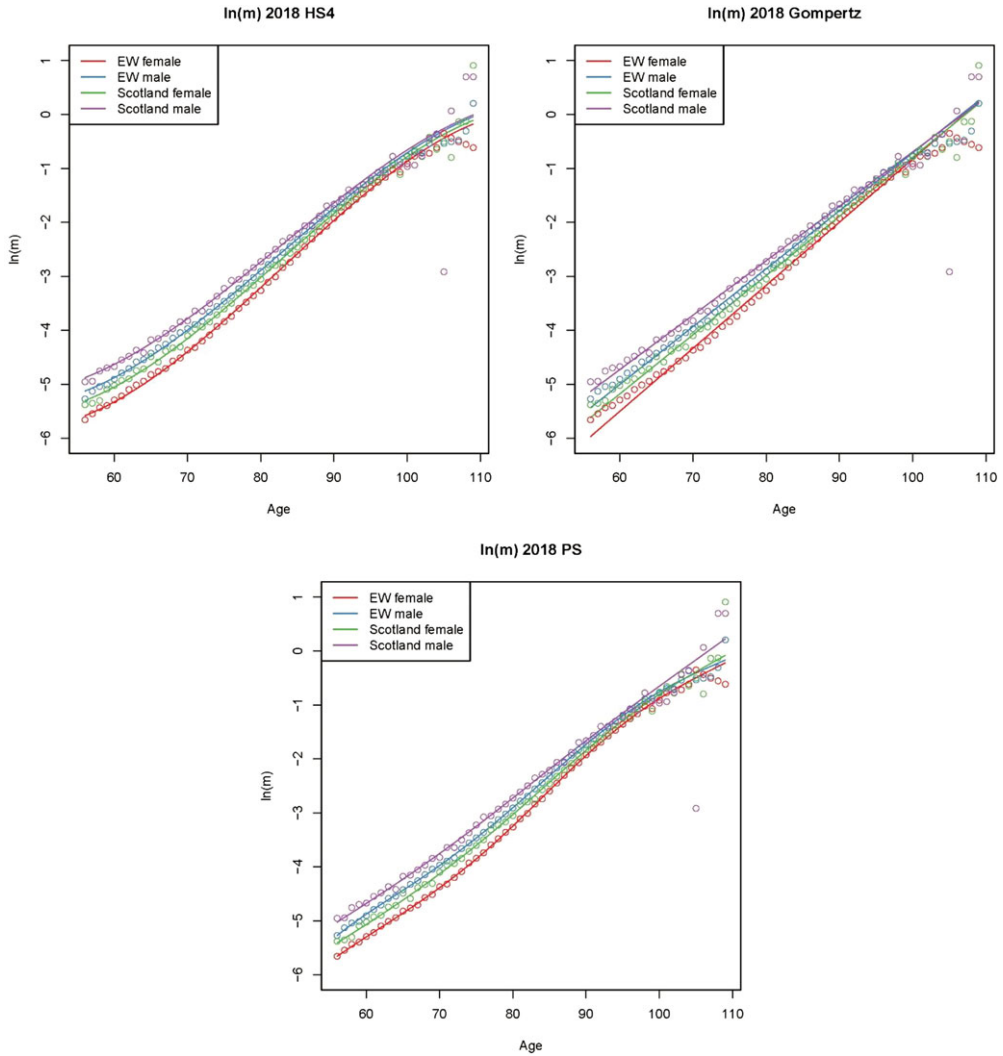


Figure 4. Fitted log central death rates in 2018 of females and males in Scotland and England and Wales.

the same level at the ending age (regardless of the country), a common  $\omega_t^{(m)} \left( \omega_t^{(f)} \right)$  could be employed.

Figure 4 displays the observed and fitted mortality rates in 2018 under the three mortality models. The PS curves show unwarranted bends due to the lack of smoothness. Although the Gompertz model captures the decreasing mortality differentials over age, evident crossovers are observed at around age 100. Under the HS4 model, the fitted mortality differentials for populations with the same gender in the two countries (red and green curves, blue and purple curves) tend to reduce towards a small gap at advanced ages. For males and females in the same country, the differences in fitted mortality rates diminish with increasing age but never vanish entirely. To conclude, the HS4 model seems to strike a good balance between the goodness-of-fit and smoothness. More importantly, the HS4 model can describe the law of mortality properly and avoid the unjustified mortality crossover that presents under the Gompertz and PS models.

## 4. Long-Term Prediction Performance

### 4.1 Time series models

This study now proceeds to examine the long-term performance of the six mortality models. The PS model extrapolates mortality rates not covered by the sample range during the fitting process by treating them as missing values. The R package “MortalitySmooth” offers the facility of both model estimation and projection, given a predetermined penalty order. In this study, we follow the suggestion by Currie *et al.* (2004) and use the default penalty order of 2 for both the age and period dimensions, which implies that (log) mortality rates are predicted via (approximately) linear extrapolation. Under the HS and Gompertz models, the time-varying parameters are modelled by time series processes and projected to a future period to predict mortality levels. We first conduct a unit root test for each of the period factors  $\alpha_t$ ,  $\omega_t$ ,  $s_{0,t}$ ,  $s_{1,t}$ ,  $k_{1,t}$  and  $k_{2,t}$ .

The results suggest that all the period factors are not stationary at 5% significance level under the five models except for the female  $\omega_t$  of Australia under the HS3 model, female  $\omega_t$  of France under the HS4 model, female  $s_{1,t}$  of Japan under the HS3 and HS4 models and male  $s_{0,t}$  of Japan under the HS4 model. When using a 1% significance level, the unit root is detected for all the time series processes. We decide to fit the time-varying factors in the same mortality model by a multivariate (bivariate) random walk with drift (RWD):

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\mu} + \mathbf{Z}_t, \quad (14)$$

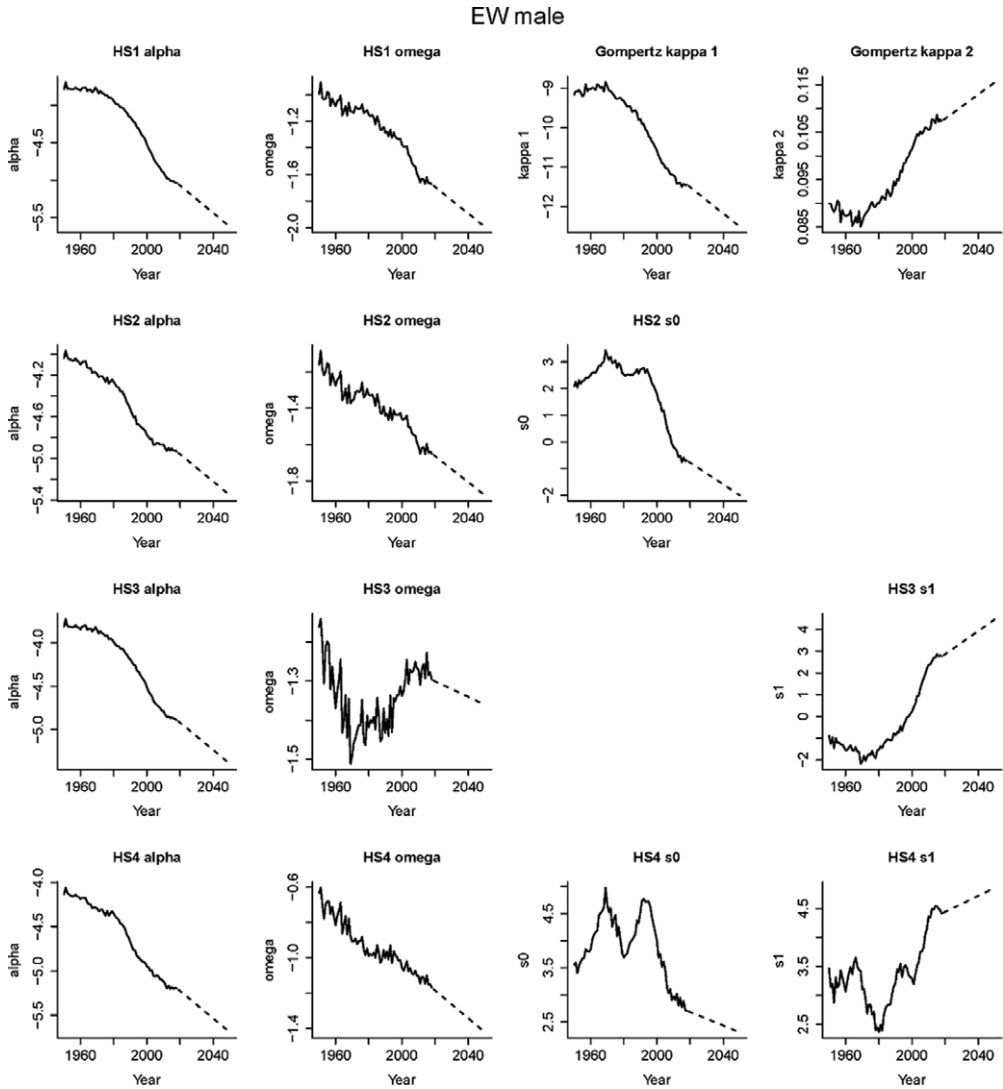
where  $\boldsymbol{\theta}_t$  is a  $(n \times 1)$  vector of the  $n$  time-varying factors in a mortality model in year  $t$ ,  $\boldsymbol{\mu}$  refers to the corresponding drift vector,  $\mathbf{Z}_t$  contains  $n$  multivariate Gaussian error terms with mean zero. Although other time series choices can be considered, the RWD model has been widely adopted in the literature to extrapolate the decreasing trend of mortality (e.g., Cairns *et al.*, 2006; Dowd *et al.*, 2010c). In addition, the assumed linear extrapolation under the multi-dimensional RWD process allows for a fair comparison between the PS approach with the second-order penalty and the other five models.

Figure 5 displays the estimated and projected parameters in the HS1-4 and Gompertz<sup>10</sup> models for males in England and Wales. As illustrated, the two level parameters  $\alpha_t$  and  $\omega_t$  show a clear declining trend under the HS models, though the downward tendency in  $\omega_t$  becomes less prominent under the HS3 model. This observation aligns with our expectations because these parameters are interpreted as the fitted mortality rates at the youngest and oldest ages of the sample which have been improving over time. Furthermore, the behaviour of  $s_{0,t}$  and  $s_{1,t}$  varies to some extent under different models. For example, the estimated  $s_{0,t}$  values under the HS2 model (second row) demonstrate a generally decreasing trend over time. However, under the full model (fourth row), the declining tendency becomes less clear and more fluctuations can be found in the estimated  $s_{0,t}$  values. Similarly, although the estimated  $s_{1,t}$  values have been increasing since around 1980 under both the HS3 (third row) and HS4 (fourth row) models, the latter tends to show more fluctuating patterns. Besides, one may notice that the intercept and slope parameters  $k_{1,t}$  and  $k_{2,t}$  in the Gompertz model (first row) show an opposite trend to each other. It indicates that the average mortality level has been improving, and this reduction has been larger at the younger ages. Though not presented here, the parameter estimates exhibit similar trends for females and males in the five countries. Then we model the time-varying parameters as the multi-dimensional RWD. The estimated drift terms, standard deviations and correlation matrices are presented in Table B.1 in Appendix B.

### 4.2 Long-term projections

After obtaining the time series forecasts, future mortality rates can be calculated by updating the predicted time-varying components. It is of interest to investigate the impact of employing

<sup>10</sup>Since the B-spline coefficients in the PS model do not have meaningful interpretation, they are not plotted here.

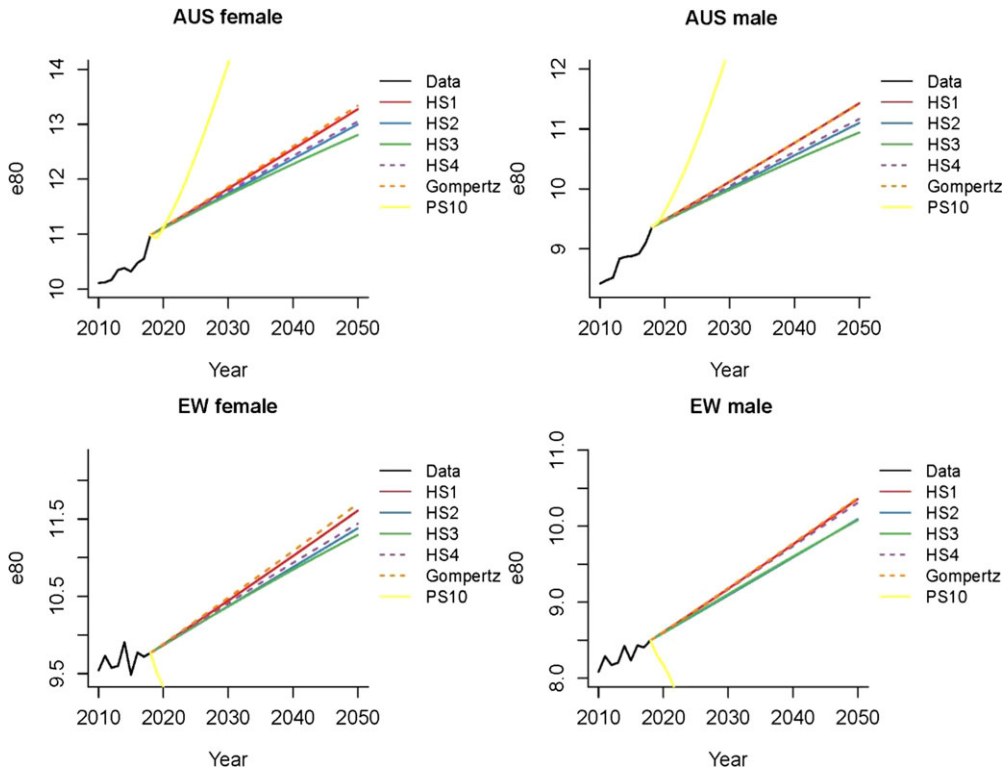


**Figure 5.** Estimated (solid lines) and projected (dashed lines) period factors in the HS1-4 and Gompertz models for males in England and Wales.

different model structures on mortality forecasts. We calculate life expectancies at key ages over the projection period. Since our data only cover ages up to 95, the Coale–Kisker approach (Coale & Kisker, 1990) is applied to extend the mortality rates to an “ultimate” age 110 with a presumed log central death rate of 0.7 (Gampe, 2010)<sup>11</sup>. Figure 6 shows the Australia and England and Wales observed life expectancy at age 80 ( $e_{80}$ ), with projections up to the year 2050.

It is apparent that the PS10 model (yellow curves) predicts significantly longer life expectancy than the other model candidates for Australia. We remark that the projections produced under the PS10 model become particularly unreasonable when mortality data of England and Wales are employed – declining life expectancies over time. As demonstrated in Figure B.1 in Appendix B that plots the fitted and extrapolated mortality rates at ages 60, 70 and 80, the projected mortality

<sup>11</sup> Although the PS approach can extend the mortality forecasts at advanced ages using its own matrix approach, we apply the same Coale-Kisker method of closing the life table for comparison purpose.



**Figure 6.** Observed and projected life expectancy at age 80 for Australia and England and Wales from 2010 to 2050. For demonstration purposes, the projections under the HS4 and Gompertz models are displayed in dashed lines.

rates for England and Wales exhibit an increasing tendency, which is not the case for Australia. The potential reason could be the slowing down of mortality improvement over the most recent few years in England and Wales. Since mortality rates are not predicted by selecting appropriate time series processes to ensure biological reasonableness, it is possible to observe unrealistic patterns in the forecasts under the PS approach.

The potentially unstable projections of the PS model were highlighted in previous studies. For example, based on Italian populations, Cocevar (2007) obtained insensible mortality forecasts for females, while the performance is reasonable for males. Richards *et al.* (2007) examined the prediction performance of the two-dimensional PS method using mortality data of seven countries and found erratic forecasts in some cases. They argued that the PS method tends to balance irregular effects in data with smoothing and those features are carried through to projections.

We also investigate the long-term prediction performance under the PS model using different number of knots, while there is no optimal knot choice that can produce sensible forecasts for all datasets. In addition, though excessive knots are penalised to generate a reasonable effective dimension, they tend to yield even more volatile forecasts. Therefore, only the prediction performance of the PS10 model is presented in the remainder of the paper.

Besides this noticeable (and unreasonable) deviation under the PS method, there exist some differences among the other models. Firstly, the HS1 model that obtains future mortality rates by projecting the mortality levels at the initial (56) and ending (95) ages results in the longest life expectancy of the four HS models. As depicted in Figure 6, adding one or two gradient parameters tends to reduce the predicted values. For example, the green curves under the HS3 model almost always lie below the other four sets of predictions (excluding the PS method) for the two ages. This can be attributed primarily to the less linear trend of the second level parameter  $\omega_t$ . For

example, it can be seen from Figure 5 that the projected values of  $\hat{\omega}_t$  (the projected  $\ln \hat{m}_{x_1,t}$ ) from the HS3 model are generally higher than those from the other HS candidates, while the HS3  $\hat{\omega}_t$  (the projected  $\ln \hat{m}_{x_0,t}$ ) values are similar to those from the HS2 model. One would expect the former to predict longer life expectancy than the latter. Although not presented here, for all the datasets the projections under the HS2 (blue) and HS4 (purple) models tend to lie within the two boundary curves formed by the other two HS candidates for all the datasets. Gompertz projections tend to be similar to those for HS1, with differences of less than 0.1–0.2 years in 2050.

Overall, the HS1-4 and Gompertz models give sensible forecasts of life expectancies. The PS model predicts unreasonably high or decreasing life expectancies, possibly caused by the automatic extrapolation process that may have overfitted the data and misused the random noises therein.

## 5. Backtesting

In the above analysis, we notice some similarities and discrepancies among the mortality forecasts under the six models. This section investigates the accuracy of their predictions via backtesting (also known as out-of-sample testing). In more detail, the sample period is split into two parts for calibrating the model parameters and testing the forecasting performance, respectively. Since the actual mortality rates are given over the testing period, the forecast accuracy of each model can be assessed by comparing the projected values against the observations. We adopt the root mean squared error (RMSE) to examine the forecasting performance quantitatively. To attain a better insight into the capability of generating sensible forecasts, the comparison is made across age, time and both dimensions using the following three measures:

$$\begin{aligned}
 RMSE_x &= \sqrt{\frac{1}{N} \sum_{h=1}^N (\ln m_{x,T+h} - \ln \hat{m}_{x,T+h})^2} \\
 RMSE_h &= \sqrt{\frac{1}{(x_1 - x_0 + 1)} \sum_{x=x_0}^{x_1} (\ln m_{x,T+h} - \ln \hat{m}_{x,T+h})^2} \\
 RMSE_{all,h} &= \sqrt{\frac{1}{(x_1 - x_0 + 1) \times h} \sum_{j=1}^h \sum_{x=x_0}^{x_1} (\ln m_{x,T+j} - \ln \hat{m}_{x,T+j})^2},
 \end{aligned} \tag{15}$$

where  $T$  is the ending year of the calibration period,  $N$  is the total length of the testing period,  $x_0(x_1)$  corresponds to the minimum (maximum) value in the age interval and  $m_{x,T+h}(\hat{m}_{x,T+h})$  represents the observed (predicted) central death rate.  $RMSE_x(RMSE_h)$  refers to the root mean squared error at age  $x$  (in the  $h^{\text{th}}$  year of forecasting) averaged across all projection years (ages), and  $RMSE_{all,h}$  gives a measure considering all ages and predictions up to year  $T + h$ . The results of a 10-step forecasting horizon are provided in Figures 7–8 and Table 2.

Figure 7 plots the 10-step  $RMSE_x$  under the six mortality models against age. Specifically, the prediction errors under the HS2 and HS4 models are somewhat similar for all the datasets. At the youngest and oldest ages, they deviate slightly, in which the HS2 errors are smaller than the HS4 errors. Over older ages, the HS3 model tends to produce the highest RMSE values among the HS candidates for four out of the five countries. Interestingly, the simplest model does not always lead to the lowest forecast accuracy. For example, the HS1 errors tend to be larger than the other HS errors at younger ages but performs better as age increases. One exception is Japan, for which the HS1 curves lie above the others at the oldest ages.

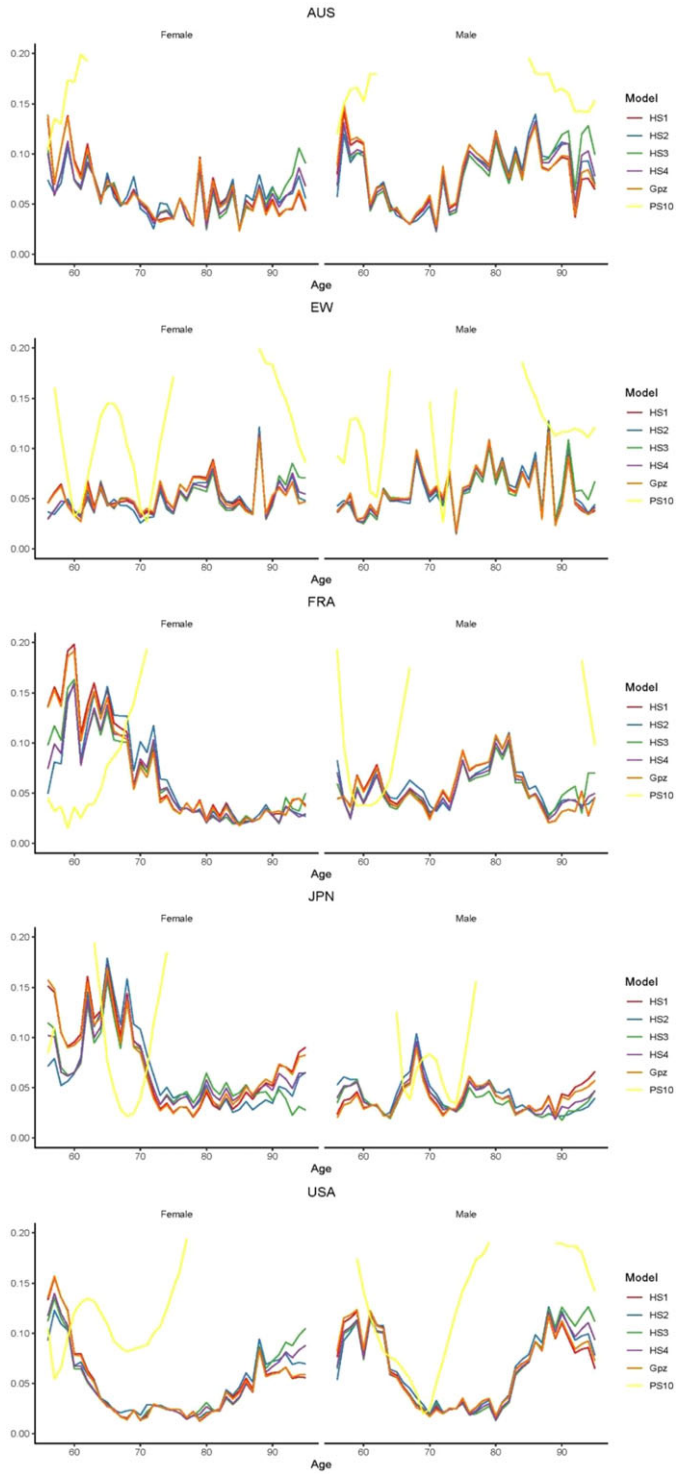


Figure 7.  $RMSE_x$  values averaged over 10 forecasting steps.

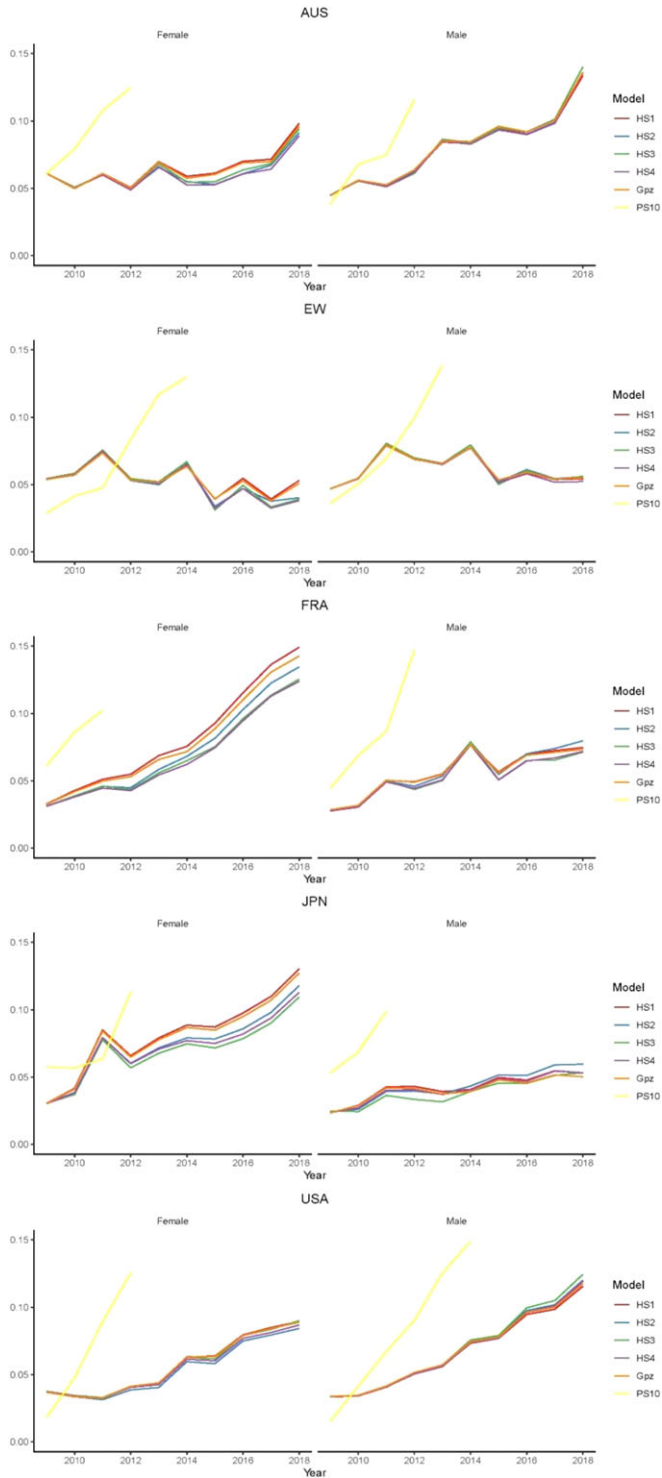


Figure 8.  $RMSE_t$  values averaged across ages 56–95.



**Table 2.**  $RMSE_{all,10}$  (%) under the HS1-4, Gompertz and PS10 models. The mortality models are ranked from 1 (the lowest prediction error) to 6 (the highest prediction error) and the ranks are given in brackets.

	HS1	HS2	HS3	HS4	Gompertz	PS10
<b>Females</b>						
Australia	6.65% (5)	6.24% (2)	6.37% (3)	<b>6.15%</b> (1)	6.56% (4)	24.63% (6)
England and Wales	5.49% (5)	5.26% (2)	5.32% (3)	<b>5.23%</b> (1)	5.43% (4)	16.39% (6)
France	9.04% (5)	8.06% (3)	7.58% (2)	<b>7.46%</b> (1)	8.66% (4)	28.73% (6)
Japan	8.63% (5)	7.79% (3)	<b>7.29%</b> (1)	7.55% (2)	8.44% (4)	25.87% (6)
United States	6.05% (5)	<b>5.70%</b> (1)	6.03% (3)	5.87% (2)	6.03% (4)	25.46% (6)
<b>Males</b>						
Australia	8.41% (2)	8.42% (3)	8.56% (5)	<b>8.40%</b> (1)	8.52% (4)	20.58% (6)
England and Wales	6.21% (2)	6.30% (5)	6.27% (4)	<b>6.15%</b> (1)	6.23% (3)	17.80% (6)
France	5.88% (4)	5.92% (5)	5.57% (2)	<b>5.56%</b> (1)	5.83% (3)	24.85% (6)
Japan	4.34% (4)	4.47% (5)	<b>3.97%</b> (1)	4.24% (3)	4.17% (2)	33.77% (6)
United States	<b>7.28%</b> (1)	7.47% (4)	7.63% (5)	7.39% (2)	7.39% (3)	16.86% (6)

As a benchmark, the Gompertz model exhibits comparable results to the HS1 model. Comparing with the HS family, this baseline model tends to produce moderately more accurate mortality forecasts at advanced ages (greater than age 90), whereas its performance is less desirable at younger ages (before age 65). Unlike the above five models that generate similar  $RMSE_x$  patterns across genders and countries, the performance of the PS10 model appears to be highly dependent on the dataset involved. When French and Japanese male data are employed, the PS approach seems to produce appealing forecast accuracy for certain age groups. Nevertheless, the RMSE values under the PS method are unusually high for the other datasets, casting doubt on the robustness of its forecasting performance.

Averaging over all ages, one can compare the prediction errors under the six mortality models over time horizons. As depicted in Figure 8, the curves generally show an upward trend, indicating the increasing uncertainty in mortality forecasts in the more distant future. Based on the mortality data in Australia, England and Wales and France, the HS4 model appears to be the best performing candidate amongst the six, while HS3 and HS2 are preferred for Japan and the United States, respectively. The benchmark Gompertz method has a prediction performance similar to the HS1 model and is almost always worse than the HS2 and HS4 models. Again, the performance of the PS approach depends heavily on the dataset used to calibrate the parameters. Overall, it fails to reliably predict mortality rates for the five countries.

After examining the model performance in the age and time dimensions separately, we now move on to a two-dimensional measure  $RMSE_{all,h}$  to make a more comprehensive assessment. The  $RMSE_{all,10}$  values averaged over all ages and years for the five countries are tabulated in Table 2. For both genders, the HS4 model tends to be the best performing candidate for three out of the five datasets, and it also produces decent forecast accuracy for the other two countries. Interestingly, the parsimonious HS1 model that is not often preferred in the previous analysis produces the best forecast accuracy for males in the United States. The Gompertz approach ranks in the middle among the six mortality models. In line with the above findings, the PS model always has notably high prediction errors, which highlights the robustness issue of the forecasting performance of the PS model.

One may argue that the forecasting performances of the mortality models might vary significantly with the split between the calibration and testing periods. In order to examine the robustness of the backtesting results, we repeat the analysis using a 20-step forecasting horizon. For the sake of brevity, only the overall performance measure  $RMSE_{all,20}$  is presented in Table 3. With a longer forecasting horizon, the HS2 model is preferred for females (males) in three (two)

**Table 3.**  $RMSE_{all,20}$  (%) under the HS1-4, Gompertz and PS10 models. The mortality models are ranked from 1 (the lowest prediction error) to 6 (the highest prediction error) and the ranks are given in brackets.

	HS1	HS2	HS3	HS4	Gompertz	PS10
<b>Females</b>						
Australia	8.11% (4)	<b>7.33%</b> (1)	7.52% (3)	7.47% (2)	8.12% (5)	52.15% (6)
England and Wales	12.62% (2)	<b>12.37%</b> (1)	12.83% (5)	12.70% (3)	12.71% (4)	25.36% (6)
France	12.39% (5)	9.80% (2)	9.83% (3)	<b>9.34%</b> (1)	11.92% (4)	39.61% (6)
Japan	12.46% (5)	11.02% (3)	<b>10.50%</b> (1)	10.58% (2)	12.34% (4)	50.26% (6)
United States	5.31% (4)	<b>4.70%</b> (1)	5.11% (3)	4.90% (2)	5.32% (5)	59.58% (6)
<b>Males</b>						
Australia	16.78% (2)	<b>16.72%</b> (1)	16.93% (5)	16.83% (4)	16.80% (3)	36.54% (6)
England and Wales	20.63% (3)	20.76% (5)	20.75% (4)	20.61% (2)	<b>20.60%</b> (1)	31.31% (6)
France	13.18% (5)	13.08% (3)	13.12% (4)	<b>12.87%</b> (1)	13.07% (2)	47.69% (6)
Japan	5.15% (5)	<b>3.94%</b> (1)	4.68% (3)	4.08% (2)	4.98% (4)	81.13% (6)
United States	<b>12.39%</b> (1)	12.56% (4)	12.43% (2)	12.52% (3)	12.62% (5)	24.30% (6)

out of the five countries. The HS4 model gives the lowest prediction error for both genders in France and is mostly the second or third best method for the other datasets. In comparison, the other two HS members demonstrate more variations in the rank of the forecasting performances.

A sensible mortality model is expected to generate stable predictions. To further examine the forecast uncertainty of the six mortality models, a contracting horizon backtesting (Dowd *et al.*, 2010c) is conducted. In more detail, we calibrate the mortality models on 20-year sequential subsamples and obtain mortality predictions for a pre-specified “future” year (year 2018). The first subsample covers data from 1950 to 1969, then from 1951 to 1970, and so on until the ending year (stepping-off year) of the fitting period reaches 2017. One would expect the predicted mortality rates on the forecast date to converge steadily to the observed values as the stepping-off year increases. Figure 9 visualises the forecasts in 2018 under the six models, using the mortality rates at age 65 as an example. All the mortality models except the PS approach illustrate similar paths of the projected central death rates. For most populations, the forecasts show a smooth curve that gradually converges towards the observed values, suggesting a stable progression of the model parameters between the sequential subsamples. As age increases (not presented here), the mortality data are more volatile, and the paths of forecasts tend to be less smooth. Nevertheless, the HS1-4 and Gompertz models still tend to share similar patterns. For the PS model, the paths fluctuate extensively and are not consistent with other methods, suggesting that the robustness of the prediction performance is a concern.

In summary, our out-of-sample testing indicates that the HS family, especially the HS4 model, tends to generate a comparable or higher degree of forecast accuracy than the benchmark Gompertz model. Furthermore, the unstable performance of the PS model suggests that extrapolating patterns without considering the temporal structure of mortality developments explicitly can give rise to unreasonable predictions<sup>12</sup>. The lack of robustness of the forecasting performance of the PS model is also addressed in Bohk-Ewald and Rau (2017) and Camarda (2019).

## 6. Simulation study

So far, we have investigated the predictive power of the six mortality models using backtesting and RMSE. Another crucial aspect of model examination lies in the level of uncertainty of the

<sup>12</sup>We also tested the performance of the PS approach using different number of knots, while no knot choice can ensure stable forecast accuracy for the datasets considered in this paper.

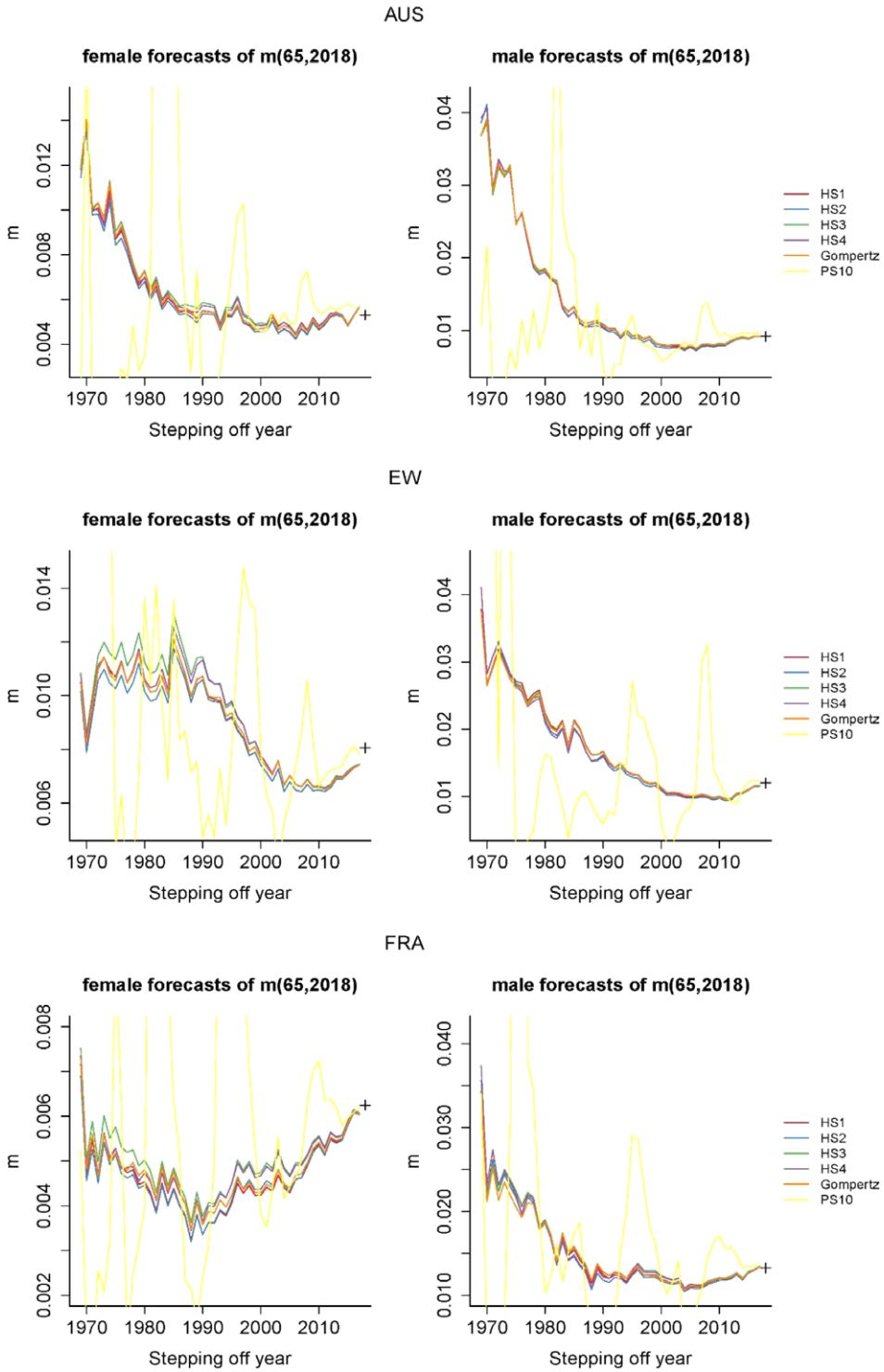


Figure 9. Forecasts of the central death rates in 2018. Mortality models are calibrated from 20-year sequential subsamples ending in the stepping-off year. The plus sign refers to the observed  $m_{65,2018}$ .

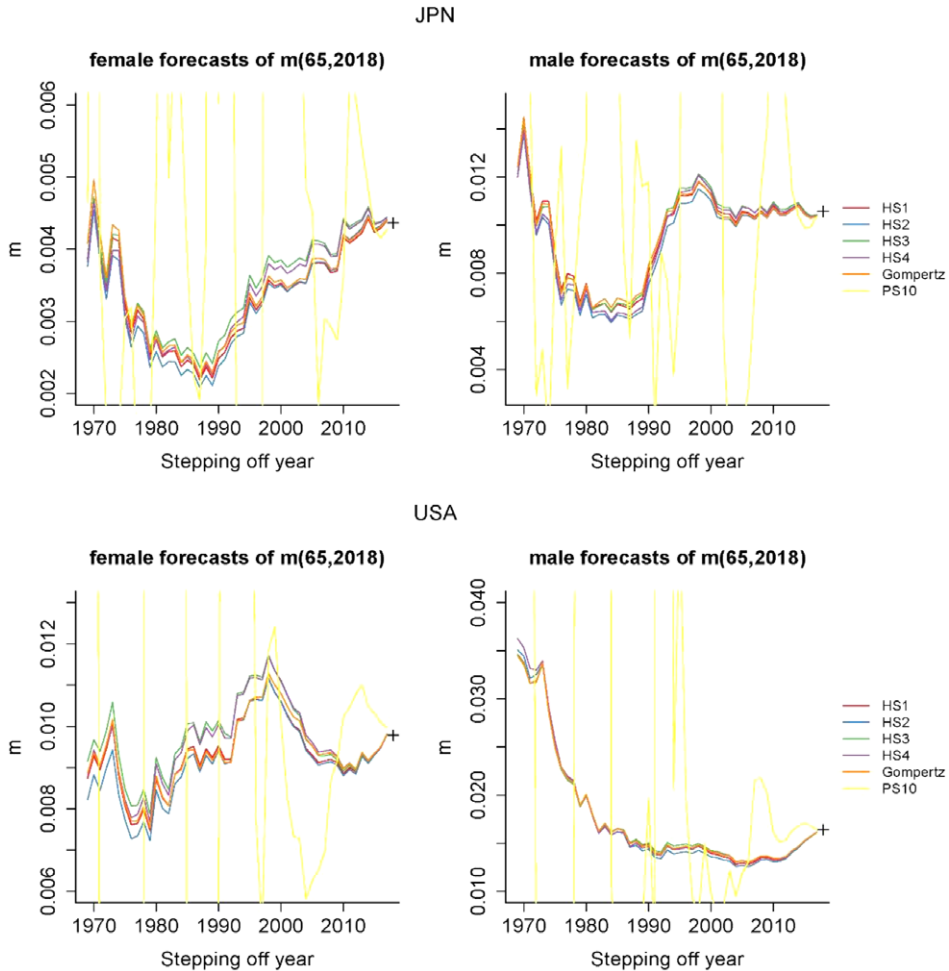
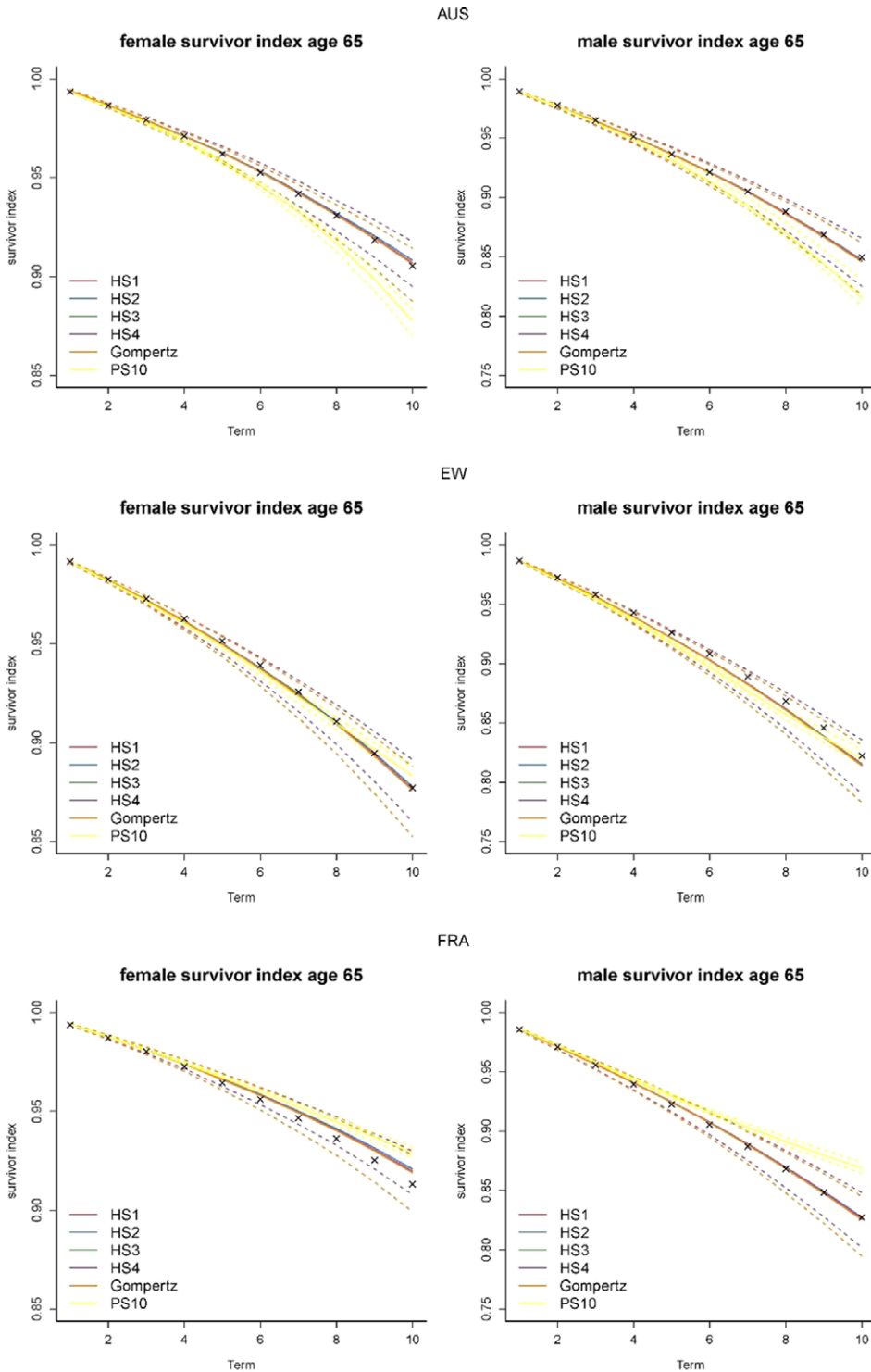


Figure 9. Continued.

generated mortality scenarios. In practice, life insurance companies are required to prepare capital for mortality and longevity risk based on risk measures such as the 99.5% Value-at-Risk (VaR) and expected shortfall (conditional VaR). Therefore, the degree of variation in mortality simulations under a mortality model plays a vital role in pricing longevity-/mortality-linked products and preparing reserves. This section studies the performances of the candidate models using the semi-parametric bootstrap (Brouhns *et al.*, 2005) that incorporates both the process error (uncertainty in time series processes for the HS and Gompertz models) and parameter error (uncertainty in parameter estimation for all the six models) into the analysis. Firstly, a pseudo sample of death counts is simulated from a Poisson distribution with the mean being the observed number of deaths. Then we fit the six mortality models to the pseudo sample and generate one set of future mortality scenarios. This process is repeated to generate 5,000 scenarios. Note that the PS model does not involve simulating future paths of time series processes, so the only source of uncertainty for the PS model is the parameter error. Following Cairns *et al.* (2009), the survivor index  $S(x, t) = \prod_{i=0}^t (1 - q_{x+i, T+i})$  representing the probability that the cohort aged  $x$  in year  $T$  survives to year  $T + t$  is employed to compare the model performance, where  $q_{x, T}$  is the probability that a life aged  $x$  in year  $T$  will die in one year. Using a 10-year testing period, we plot in Figure 10 the projected



**Figure 10.** Projected values (solid lines) and 95% prediction intervals (dashed lines) of survivor index  $S(65, t)$  of the cohort aged 65 in 2009. The cross sign in the figure denotes the observed survivor index. The mortality models are calibrated on years 1950–2008 and projected/simulated over a 10-year forecasting horizon.

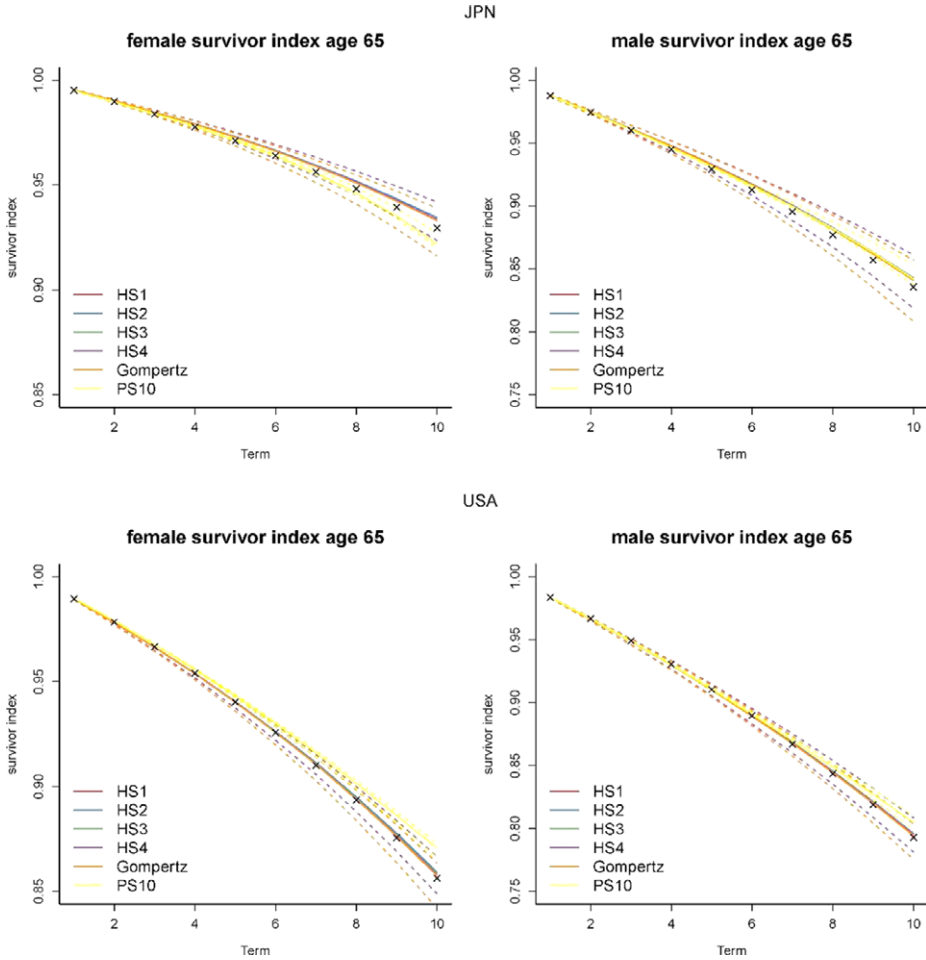
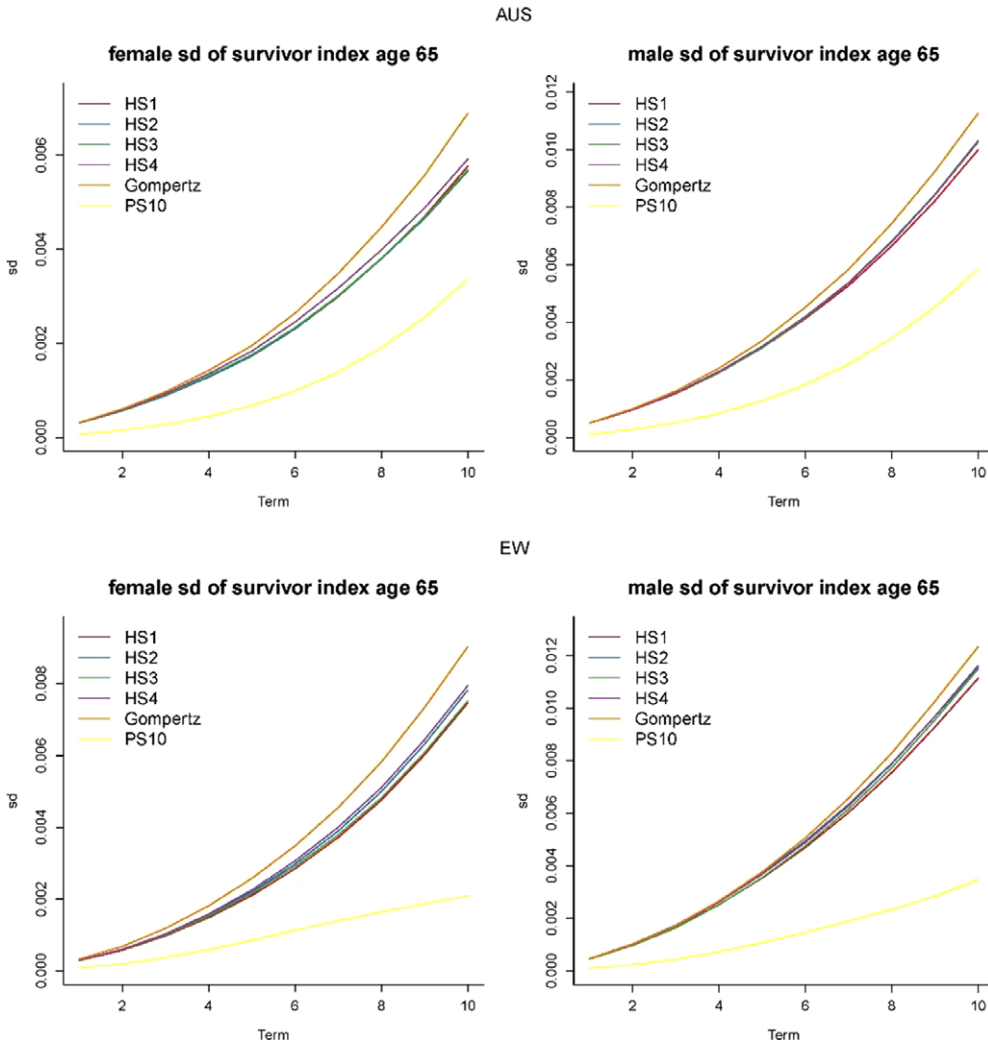


Figure 10. Continued.

$S(65, t)$  for the cohort aged 65 in the first prediction year. For illustration purposes, only the 95% prediction intervals of the HS4, Gompertz and PS models are presented.

For the ten datasets, the predicted survivor indices under the HS1-4 and Gompertz models follow roughly the same path as the observed values (cross signs). But when the PS model (yellow) is applied, there exists an obvious overestimation (underestimation) of the death rates for France and females in the United States (Australia). Besides, the simulated mortality distribution from the PS model only incorporates the parameter uncertainty, so its prediction intervals tend to be narrower than the others. Under the Gompertz model, the 95% prediction intervals are slightly broader than those under the other four HS candidates. Within the HS family, the prediction bounds are rather close to one another, though the full model appears to generate the widest prediction bounds because it involves more time series processes and potentially greater process error. All the mortality models except the PS model can produce prediction intervals that capture the observed values (the prediction intervals under the HS1-3 models are not displayed).

A more pronounced reflection is illustrated in Figure 11 that displays the standard deviations of  $S(65, t)$ . At the end of the testing period ( $t = 10$ ), the Gompertz model demonstrates the greatest variations, followed by the HS candidates and the PS model. As discussed above, the HS4 model tends to create more variations in the simulated scenarios of  $S(65, t)$ . However, the differences



**Figure 11.** Standard deviation of survivor index  $S(65, t)$  of the cohort aged 65 in 2009,  $t = 1, 2, \dots, 10$ . The mortality models are calibrated on mortality data in Australia and England and Wales over 1950–2008 and projected/simulated over a 10-year forecasting horizon. Multivariate RWD is fitted to the time-varying parameters.

in the standard deviations of  $S(65, t)$  within the HS family are not significant, and no general conclusion can be made on their ordinal relationship.

The above results for the HS1–4 and Gompertz models are based on the multivariate RWD model. If the time-varying parameters are modelled separately by univariate RWD processes regardless of the strong correlation, the level of process uncertainty may differ significantly. We then investigate the impact of the correlation structure of the time series models on the prediction intervals. Since the projection results are not affected under this independence setting, we only present the standard deviations of  $S(65, t)$  for the same populations in Figure B.2 in Appendix B. At the end of the testing period ( $t = 10$ ), the Gompertz model demonstrates the greatest variations, followed roughly by the HS4, HS2, HS3 and HS1 models. Although the estimated standard errors of the time series models for the Gompertz parameters have a similar magnitude to those for the HS parameters, the time-varying parameters in these two types of models are scaled differently. Under the Gompertz model,  $k_{1,t}$  and  $k_{2,t}$  are multiplied by 1 and  $x$  (from 56–95), respectively.

On the other hand, the HS parameters are scaled by the Hermite basis functions that are less than 1 in magnitude. Without considering the correlation structure, the 95% prediction intervals under the Gompertz model would be much broader than those under the other four HS candidates. Within the HS family, the full (simplest) model generates the widest (narrowest) prediction bounds because it involves more (fewer) time series processes and greater (smaller) process error. Although both the HS2 and HS3 models have three time-varying components, the width of the 95% prediction intervals under the former tends to be larger than that under the latter. One potential explanation could be that the gradient parameter in the HS2 model acts on the younger age range and creates more variations in mortality simulations at younger ages (e.g., younger than age 75). Since Figure B.2 considers  $S(65, t)$  for  $1 \leq t \leq 10$  only, simulations under the HS2 model would demonstrate a higher level of uncertainty than those under the HS3 model. When older age groups are considered, the situation would be reversed.

It is interesting to observe that incorporating the correlation structure between the time-varying components leads to notably different standard deviations of the simulated mortality scenarios between the HS1-4 and Gompertz models. As presented in Table B.1 in Appendix B,  $k_{1,t}$  and  $k_{2,t}$  in the Gompertz model exhibit a highly negative correlation. The standard deviations of simulated mortality rates under the multivariate RWD model are then smaller than that in the independent case (univariate RWD). By contrast, the two level parameters in the HS1 model are positively correlated, and the standard deviations in Figure 11 (correlated) are slightly greater than those in Figure B.2 (independent). In other words, when a positive (negative) relationship exists between two time-varying parameters, the prediction intervals generated from multivariate RWD tend to be wider (narrower) than those from the independent setting. Note that there is one exception in the HS3 and HS4 models. Since the gradient parameter  $s_{1,t}$  is multiplied by the function  $h_{11}(x_k)$  that remains negative over the age range, a positive correlation between  $s_{1,t}$  and other time-varying factors would lead to narrower prediction intervals. Overall, more variations can be observed in the simulated scenarios under the univariate RWD than those from the multivariate RWD model.

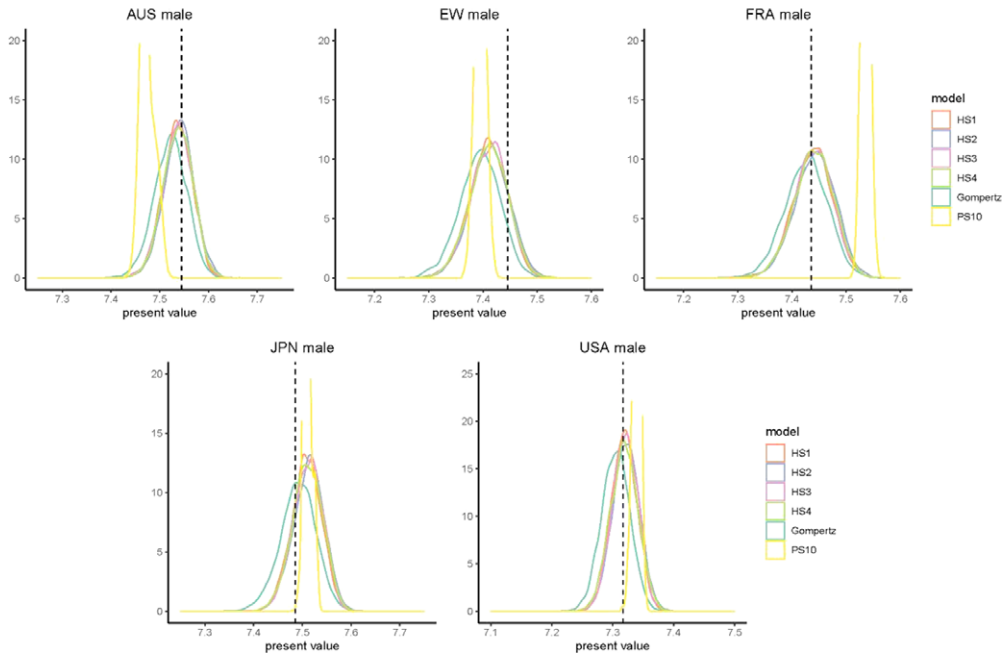
Given the same central projections, wider prediction intervals suggest a larger range in which the future values are likely to fall. Although there is no “correct” level of uncertainty in mortality forecasting, an exceedingly broad range means that the predictions are imprecise and the final outcomes are highly uncertain. For instance, the backtesting results in Figure 10 show that the historical survivor index is well captured even by the narrowest 95% prediction intervals under the HS1 model. Such a high level of uncertainty of mortality forecasts in Figure B.2, especially under the Gompertz model, seems unnecessary.

As an additional comparison, we calculate the present values of the simulated 10-year annuity for the male cohort aged 65 in 2009. Given the simulated scenarios and a predetermined discount rate, the term annuity that pays \$1 at the end of each future year on survival can be valued as follows:

$$a_{65:10|} = \sum_{t=1}^{10} \frac{S(65, t)}{(1+r)^t}, \quad (16)$$

where  $a_{65:10|}$  represents the present value of the 10-year term annuity payable to the 1944 (2009 – 65) birth cohort, and  $S(65, t)$  is the corresponding survivor index. We use a 4% discount rate ( $r$ ) in this analysis. The simulated density distributions of  $a_{65:10|}$  for the six mortality models are plotted in Figure 12. It can be observed that both the central tendencies and dispersions somewhat depend on the underlying model. Within the HS category, mortality distributions under the four models share a rather similar location of the peak. As to the spread, the full model with more time-varying parameters tends to generate more dispersed distributions, though the difference is not pronounced. Between the three types of mortality models, the Gompertz model tends to simulate the annuity distributions with the heaviest tails, which is in line with the findings on the survivor indices. The density distributions under the Gompertz model seem to lie on the left





**Figure 12.** Simulated distributions of 10-year annuity values for the male cohort aged 65 in 2009. The mortality models are calibrated on mortality data over 1950–2008. The dashed lines correspond to the annuity value calculated from historical data over 2009–2018.

of the HS distributions for all the five countries, which implies that the HS annuity values are marginally higher than the benchmark. On the other hand, the situation is more erratic under the PS model. It produces annuity distributions that often fail to cover the present value calculated from historical data. Both the evident concentration on the peak and the “wrong” location of the central tendencies are responsible for this poor performance.

## 7. Conclusions

In this paper, we have extended the one-dimensional HS model proposed by Richards (2020) to cater for the modelling of country-level mortality data over time. Four versions of the HS model are considered and compared with the Gompertz and PS models. Unlike the PS model that makes no assumptions about the functional form of mortality rates across both ages and years, the Gompertz and HS methods impose smoothness across ages only. The Gompertz model assumes that the log mortality rates over an age range follow a straight line, while the HS models smooth the mortality rates between two boundary ages using Hermite basis functions. Based on the mortality data of Australia, England and Wales, France, Japan and the United States, the full model of the HS family produces smaller AIC and BIC values than the Gompertz model in all cases. By imposing smoothness on both the age and period dimensions, the PS model tends to achieve a satisfactory fit with smaller effective dimensions. Nonetheless, when advanced ages are included in the sample range, the potential crossover between the fitted mortality rates of populations with different risk factors can be circumvented only under the HS models.

Although the PS model exhibits a decent fit, it often fails to generate realistic projections by expanding the fitted mortality matrix. On the other hand, the HS and Gompertz models utilise time series models and result in reasonable forecasts in the long run. Our backtesting results reveal

that the forecasting performance of the PS model appears to be highly unstable. Fairly high prediction errors are observed under the PS approach. Comparatively, the other five mortality models demonstrate more robust backtesting performance. Thereinto, the HS1 and Gompertz models are quite similar in prediction accuracy. The HS4 and HS2 models tend to perform well for 10-year and 20-year forecasting horizons, respectively. Next, we simulate the survivor index of lives aged 65 over the last 10 years of the sample period and compare the results with the observed values. It is found that the prediction intervals well capture the historical survivor index of the cohort under all the models except the PS method. Furthermore, there are differences in the degree of variations in the simulated scenarios between the mortality models, which would affect the pricing and reserving of longevity-/mortality-linked products in practice.

Overall, amongst the six models, the HS4 model achieves a good balance between different aspects of the mortality modelling exercise. It fits the historical data reasonably well for several countries. It gives rise to decent prediction accuracy. It induces a suitable level of uncertainty in the simulations. And it ensures the gradual convergence of mortality rates between populations at older ages.

There are some potential areas for future research. This paper focuses on longevity risk and only considers populations aged 56 and above. When a wider age range is under interest, the HS approach with a maximum of 4 parameters for each year may not capture all the features of the mortality curves. To overcome this limitation, one may employ multiple HSs that are continuously joined at different knots. It would be interesting to examine the performance of this extension and compare it with other mortality models. More investigation is needed to determine the optimal number and location of knots for different datasets. By default, the HS approach has the advantage of avoiding mortality crossover in the age dimension. Another modification is constructing a multi-population HS model that ensures convergence in both age and time dimensions. This property would require a careful choice of time series models, but more investigation is needed. Note that the fitted mortality rates under the HS models start and end at the corresponding observations, and appropriate endpoints need to be determined when multiple populations are involved. Furthermore, we have noticed that the HS models are capable of capturing the relationship between different populations. One may apply this method to evaluate the effectiveness of longevity risk hedge affected by different types of risk factors.

**Acknowledgement.** The authors thank the editor and reviewers for their valuable comments.

## References

- Alho, J. & Spencer, B. (2006). *Statistical Demography and Forecasting*. Springer, New York.
- Black, D.A., Hsu, Y.-C., Sanders, S.G., Schofield, L.S. & Taylor, L.J. (2017). The Methuselah effect: the pernicious impact of unreported deaths on old-age mortality estimates. *Demography* 54(6), 2001–2024.
- Blake, D., Cairns, A.J.G., Dowd, K. & Kessler, A.R. (2018). Still living with mortality: the longevity risk transfer market after one decade. *British Actuarial Journal* 24(1), 1–80. doi: [10.1017/S1357321718000314](https://doi.org/10.1017/S1357321718000314)
- Bohk-Ewald, C. & Rau, R. (2017). Probabilistic mortality forecasting with varying age-specific survival improvements. *Genus* 73(1), 1–37.
- Booth, H. & Tickle, L. (2008). Mortality modelling and forecasting: a review of methods. *Annals of Actuarial Science* 3(1–2), 3–43.
- Brouhns, N., Denuit, M. & Van Keilegom, I. (2005). Bootstrapping the Poisson log-bilinear model for mortality forecasting. *Scandinavian Actuarial Journal* 2005(3), 212–224. doi: [10.1080/03461230510009754](https://doi.org/10.1080/03461230510009754)
- Brouhns, N., Denuit, M. & Vermunt, J.K. (2002). A Poisson log-bilinear regression approach to the construction of projected life tables. *Insurance: Mathematics and Economics* 31(3), 373–393. Available online at the address <https://EconPapers.repec.org/RePEc:eee:insuma:v:31:y:2002:i:3:p: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(4), 687–718. doi: [10.1111/j.1539-6975.2006.00195.x](https://doi.org/10.1111/j.1539-6975.2006.00195.x)

- Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D. & Khalaf-Allah, M. (2011). Mortality density forecasts: an analysis of six stochastic mortality models. *Insurance: Mathematics and Economics* 48(3), 355–367. Available online at the address <https://doi.org/10.1016/j.insmatheco.2010.12.005%3cb%3e>
- 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 and Wales and the United States. *North American Actuarial Journal* 13(1), 1–35. doi: [10.1080/10920277.2009.10597538](https://doi.org/10.1080/10920277.2009.10597538)
- Camarda, C.G. (2012). MortalitySmooth: an R package for smoothing Poisson counts with P-splines. *Journal of Statistical Software* 50(1), 1–24.
- Camarda, C.G. (2019). Smooth constrained mortality forecasting. *Demographic Research* 41, 1091–1130.
- Christensen, K., Doblhammer, G., Rau, R. & Vaupel, J.W. (2009). Ageing populations: the challenges ahead. *The Lancet* 374(9696), 1196–1208.
- Coale, A.J. & Kisker, E. (1990). *Defects in data on old-age mortality in the United States: New procedures for calculating mortality schedules and life tables at the highest ages* (Vol. 4).
- Cocovar, P. (2007). An analysis of recent mortality trends in the Italian population using penalised B-spline regression. *Giornale dell'Istituto Italiano degli Attuari* 70, 21–43.
- Currie, I.D. (2006). Smoothing and forecasting mortality rates with P-splines. *Talk given at the Institute of Actuaries*.
- Currie, I.D., Durban, M. & Eilers, P.H. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling An International Journal* 4(4), 279–298.
- Currie, I.D., Durban, M. & Eilers, P.H. (2006). Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68(2), 259–280.
- De Jong, P. & Tickle, L. (2006). Extending Lee–Carter mortality forecasting. *Mathematical Population Studies* 13(1), 1–18. doi: [10.1080/08898480500452109](https://doi.org/10.1080/08898480500452109)
- Delwarde, A., Denuit, M. & Eilers, P. (2007). Smoothing the Lee–Carter and Poisson log-bilinear models for mortality forecasting: a penalized log-likelihood approach. *Statistical Modelling* 7(1), 29–48.
- Dokumentov, A., Hyndman, R.J. & Tickle, L. (2018). Bivariate smoothing of mortality surfaces with cohort and period ridges. *Stat* 7(1), e199.
- Dowd, K., Cairns, A.J., Blake, D., Coughlan, G.D., Epstein, D. & Khalaf-Allah, M. (2010a). Backtesting stochastic mortality models: an ex post evaluation of multiperiod-ahead density forecasts. *North American Actuarial Journal* 14(3), 281–298.
- Dowd, K., Cairns, A.J., Blake, D., Coughlan, G.D., Epstein, D. & Khalaf-Allah, M. (2010b). Evaluating the goodness of fit of stochastic mortality models. *Insurance: Mathematics and Economics* 47(3), 255–265.
- Dowd, K., Cairns, A.J.G., Blake, D., Coughlan, G.D., Epstein, D. & Khalaf-Allah, M. (2010c). Backtesting stochastic mortality models. *North American Actuarial Journal* 14(3), 281–298. doi: [10.1080/10920277.2010.10597592](https://doi.org/10.1080/10920277.2010.10597592)
- Eilers, P.H., Currie, I.D. & Durban, M. (2006). Fast and compact smoothing on large multidimensional grids. *Computational Statistics & Data Analysis* 50(1), 61–76.
- Eilers, P.H. & Marx, B.D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* 11(2), 89–121.
- Eilers, P.H. & Marx, B.D. (2003). Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometrics and Intelligent Laboratory Systems* 66(2), 159–174.
- Eilers, P.H.C. & Marx, B.D. (2002). Generalized linear additive smooth structures. *Journal of Computational and Graphical Statistics* 11(4), 758–783.
- Forfar, D.O. (2014). Mortality laws. In *Wiley StatsRef: Statistics Reference Online* (pp. 1–11).
- Gampe, J. (2010). Human mortality beyond age 110. In H. Maier, J. Gampe, B. Jeune, J.-M. Robine & J.W. Vaupel (Eds.), *Supercentenarians* (pp. 219–230). Springer Berlin Heidelberg, Berlin, Heidelberg.
- Gavrilov, L.A. & Gavrilova, N.i.a.S. (1991). *The Biology of Life Span: A Quantitative Approach*. Harwood Academic Publishers, New York.
- Gavrilov, L.A. & Gavrilova, N.S. (2001). The reliability theory of aging and longevity. *Journal of theoretical Biology* 213(4), 527–545.
- Gompertz, B. (1825). XXIV. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. FRS &c. *Philosophical Transactions of the Royal Society of London* 115, 513–583.
- Haberman, S., Kaishev, V., Millosovich, P., Villegas, A., Baxter, S., Gaches, A., Gunlaugsson, S., & Sison, M. (2014). *Longevity Basis Risk – A Methodology for Assessing Basis Risk*. Available online at the address <http://openaccess.city.ac.uk/id/eprint/16835/>
- Hermite, M. (1864). *Sur un nouveau développement en série des fonctions*. Imprimerie de Gauthier-Villars.
- Human Mortality Database. (2021). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available online at the address [www.mortality.org](http://www.mortality.org) [accessed 4-Sep-2020].
- Hyndman, R.J., Booth, H. & Yasmeen, F. (2013). Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography* 50(1), 261–283.
- Hyndman, R.J. & Ullah, M.S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis* 51(10), 4942–4956.

- Kreyszig, E. (2009). *Advanced Engineering Mathematics*, 10th edition. Wiley, New York.
- Lee, R.D. & Carter, L.R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association* 87(419), 659–671. doi: [10.1080/01621459.1992.10475265](https://doi.org/10.1080/01621459.1992.10475265)
- Li, J. (2013). A Poisson common factor model for projecting mortality and life expectancy jointly for females and males. *Population Studies* 67(1), 111–126. doi: [10.1080/00324728.2012.689316](https://doi.org/10.1080/00324728.2012.689316)
- Li, J.S.-H. & Liu, Y. (2020). The heat wave model for constructing two-dimensional mortality improvement scales with measures of uncertainty. *Insurance: Mathematics and Economics* 93, 1–26. Available online at the address <https://doi.org/10.1016/j.insmatheco.2020.04.001>
- Li, J.S.-H. & Liu, Y. (2021). Recent declines in life expectancy: implication on longevity risk hedging. *Insurance: Mathematics and Economics* 99, 376–394.
- Li, N. & Lee, R. (2005). Coherent mortality forecasts for a group of populations: an extension of the lee-carter method. *Demography* 42(3), 575–594. doi: [10.1353/dem.2005.0021](https://doi.org/10.1353/dem.2005.0021)
- Makeham, W.M. (1867). On the law of mortality. *Journal of the Institute of Actuaries (1866-1867)* 13(6), 325–358.
- Marschner, S. & Shirley, P. (2018). *Fundamentals of Computer Graphics*. CRC Press, Natick, Massachusetts.
- McNown, R. & Rogers, A. (1989). Forecasting mortality: a parameterized time series approach. *Demography* 26(4), 645–660.
- Newman, S.J. (2018). Errors as a primary cause of late-life mortality deceleration and plateaus. *PLoS Biology* 16(12), e2006776.
- Perks, W. (1932). On some experiments in the graduation of mortality statistics. *Journal of the Institute of Actuaries (1886-1994)* 63(1), 12–57.
- Renshaw, A.E. & Haberman, S. (2003). Lee–Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics* 33(2), 255–272.
- Renshaw, A.E. & Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics* 38(3), 556–570. Available online at the address <https://EconPapers.repec.org/RePEc:eee:insuma:v:38:y:2006:i:3:p:556-570>
- Richards, S.J. (2020). A Hermite-spline model of post-retirement mortality. *Scandinavian Actuarial Journal* 2020(2), 110–127.
- Richards, S.J. (2021). A value-at-risk approach to mis-estimation risk. *Institute and Faculty of Actuaries*. Available online at the address <https://ifoadd7.ifoaprod1.com/system/files/field/document/Sessional%20Final%20Paper.pdf>.
- Richards, S.J., Ellam, J.R., Hubbard, J., Lu, J.L.C., Makin, S.J., & Miller, K.A. (2007). Two-dimensional mortality data: patterns and projections. *British Actuarial Journal* 13(3), 479–536. doi: [10.1017/S1357321700001537](https://doi.org/10.1017/S1357321700001537)
- Schoenberg, I. (1969). Cardinal interpolation and spline functions. *Journal of Approximation Theory* 2(2), 167–206.
- Tickle, L. (1997). *The Sex Differential in Mortality Rates and Underlying Factors*. School of Economic and Financial Studies, Macquarie University.
- Villegas, A., Haberman, S., Kaishev, V.K. & Millossovich, P. (2017). A comparative study of two population models for the assessment of basis risk in longevity hedges. *ASTIN Bulletin* 47(3), 631–679.
- Wright, S. & Nocedal, J. (1999). Numerical optimization. *Springer Science* 35(67–68), 7.

## Appendix

### Appendix A. Maximum likelihood estimation with the barrier method

Suppose that the mortality rates of male ( $i = 1$ ) and female ( $i = 2$ ) populations in the same country are fitted by the HS4 model with common  $s_{0,t}$  and  $s_{1,t}$ . The log central death rates for the  $i^{\text{th}}$  population can be expressed as:

$$\ln m_{x,t}^{(i)} = \alpha_t^{(i)} h_{00}(x_k) + \omega_t^{(i)} h_{01}(x_k) + s_{0,t} h_{10}(x_k) + s_{1,t} h_{11}(x_k),$$

where  $\alpha_t^{(i)}$  and  $\omega_t^{(i)}$  are the population-specific level parameters,  $s_{0,t}$  and  $s_{1,t}$  refer to the common gradient parameters. Thereinto,  $\omega_t^{(1)}$  and  $\omega_t^{(2)}$  are inhibited by the constraint  $\omega_t^{(1)} > \omega_t^{(2)}$  to ensure the no-crossover property. Following the assumptions discussed in section 3.1, the model parameters can be calibrated by maximising the Poisson log-likelihood function  $l$ , using the Newton–Raphson iteration scheme.

However, it is not straightforward to implement optimisation with inequality constraints. To address this issue, we follow Li & Liu (2020, 2021) and employ the barrier method under which the above optimisation is converted to maximising the penalised log-likelihood function  $l^{(B)}$ . The

barrier method allows us to set the upper and/or lower bounds of the parameters. Instead of restricting the direct relationship between  $\omega_t^{(1)}$  and  $\omega_t^{(2)}$ , two separate constraints are introduced,

$$\begin{cases} \omega_t^{(1)} > a_t^{(1,2)} \\ \omega_t^{(2)} < a_t^{(1,2)} \end{cases},$$

where  $\omega_t^{(1)}$  and  $\omega_t^{(2)}$  are the male and female level parameters at the ending age, respectively,  $a_t^{(1,2)}$  is calculated as the average of the initial values of  $\omega_t^{(1)}$  and  $\omega_t^{(2)}$ . Accordingly, the penalised log-likelihood function can be written as:

$$l^{(B)} = l - B^{(1)}(\vec{\omega}^{(1)}; r_\omega^{(1)}) - B^{(2)}(\vec{\omega}^{(2)}; r_\omega^{(2)}),$$

where  $l$  is defined in equation (13),  $\vec{\omega}^{(1)}(\vec{\omega}^{(2)})$  is a vector containing  $\omega_t^{(1)}(\omega_t^{(2)})$  values for  $t \in [t_0, t_1]$ . The barrier functions  $B^{(1)}(\vec{\omega}^{(1)}; r_\omega^{(1)})$ ,  $B^{(2)}(\vec{\omega}^{(2)}; r_\omega^{(2)})$  and the so-called resistance levels  $r_\omega^{(1)}, r_\omega^{(2)}$  are defined as follows:

$$\begin{cases} B^{(1)}(\vec{\omega}^{(1)}; r_\omega^{(1)}) = \sum_{t=t_0}^{t_1} r_\omega^{(1)} \left( \frac{1}{\omega_t^{(1)} - a_t^{(1,2)}} \right) \\ B^{(2)}(\vec{\omega}^{(2)}; r_\omega^{(2)}) = \sum_{t=t_0}^{t_1} r_\omega^{(2)} \left( \frac{1}{a_t^{(1,2)} - \omega_t^{(2)}} \right) \end{cases}, \text{ with } \begin{cases} r_\omega^{(1)} = 0.0001 |\tilde{l}| / \left( \sum_{t=t_0}^{t_1} \frac{1}{\omega_t^{(1)} - a_t^{(1,2)}} \right) \\ r_\omega^{(2)} = 0.0001 |\tilde{l}| / \left( \sum_{t=t_0}^{t_1} \frac{1}{a_t^{(1,2)} - \omega_t^{(2)}} \right) \end{cases},$$

where  $\tilde{\omega}_t^{(i)}$  is the initial value of the second level parameter in year  $t$  for the  $i^{th}$  population,  $\tilde{l}$  is calculated using the initial values of all the model parameters.

As suggested in Li & Liu (2020, 2021), the resistance levels are determined using the initial value of the log-likelihood function. The above specification implies that each barrier function accounts for 0.1% of the magnitude of  $l$  at the start of the estimation process<sup>13</sup>. At the end of each iteration, the resistance levels  $r_\omega^{(1)}$  and  $r_\omega^{(2)}$  are reduced by 15% so that the penalty term contributes less and less to the log-likelihood function as it converges.

### A.1 Initial values of $\omega_t^{(i)}$

Since  $\omega_t^{(i)}$  represents the mortality levels at the ending age  $x_1$ , one may simply set  $\ln m_{x_1,t}^{(i)}$  as its initial values. Nonetheless, in the presence of crossovers in the mortality data ( $\ln m_{x_1,t}^{(1)} < \ln m_{x_1,t}^{(2)}$ ) in some years, such conditions can violate the requirements of the barrier method (starting from the interior points of the bounds). To avoid such a violation, we employ the following initial values:

$$\begin{cases} \omega_t^{(1)} = \max(\ln m_{x_1,t}^{(1)}, \ln m_{x_1,t}^{(2)}) \\ \omega_t^{(2)} = \min(\ln m_{x_1,t}^{(1)}, \ln m_{x_1,t}^{(2)}) \end{cases}.$$

When the sample covers early periods (e.g., before 1970), mortality data at advanced ages may demonstrate some fluctuations. Using the above initial values can result in unreasonable

<sup>13</sup>Note that the number 0.1% is chosen based on empirical analysis. If the proportion is too high, the parameters are estimated in a way that is more prone to satisfying the constraints rather than producing the “best” fit. If an unreasonably low proportion is used, the penalty functions would have a minor effect on the optimisation, which may lead to parameters falling outside the bounds. After testing a set of values, we find that 0.1% strikes a good balance between the goodness-of-fit and satisfaction of the constraints for the mortality data of United States. Though the proportion seems low, no constrained parameters are found to escape from their bounds.

bounds  $a_t^{(1,2)}$  at the ending age  $x_1$ , which affects the goodness-of-fit. To develop proper constraints, one may also consider using the maximum likelihood estimates (without the no-crossover restrictions) as the initial values.

### A.2 Updating equations of $\omega_t^{(i)}$

As before, the parameters under the HS4 model are estimated iteratively using the Newton-Raphson method. Combining with the barrier approach, the general updating equation becomes

$\theta^* = \theta - \frac{\partial l^{(B)}}{\partial \theta} / \frac{\partial^2 l^{(B)}}{\partial \theta^2}$ . Because the barrier functions in  $l^{(B)}$  are only associated with  $\omega_t^{(i)}$ , the updating equations remain the same for the other parameters. The first and second partial derivatives for  $\omega_t^{(i)}$  are summarised below:

$$\left\{ \begin{array}{l} \frac{\partial l^{(B)}}{\partial (\omega_t^{(1)})} = \frac{\partial l}{\partial (\omega_t^{(1)})} + r_\omega^{(1)} \times \frac{1}{(\omega_t^{(1)} - a_t^{(1,2)})^2} \\ \frac{\partial l^{(B)}}{\partial (\omega_t^{(2)})} = \frac{\partial l}{\partial (\omega_t^{(2)})} - r_\omega^{(2)} \times \frac{1}{(a_t^{(1,2)} - \omega_t^{(2)})^2} \\ \frac{\partial^2 l^{(B)}}{\partial (\omega_t^{(1)})^2} = \frac{\partial^2 l}{\partial (\omega_t^{(1)})^2} - r_\omega^{(1)} \times \frac{2}{(\omega_t^{(1)} - a_t^{(1,2)})^3} \\ \frac{\partial^2 l^{(B)}}{\partial (\omega_t^{(2)})^2} = \frac{\partial^2 l}{\partial (\omega_t^{(2)})^2} - r_\omega^{(2)} \times \frac{2}{(a_t^{(1,2)} - \omega_t^{(2)})^3} \end{array} \right.$$

## Appendix B. Tables and Figures

**Table B.1.** Estimates of multivariate (bivariate) random walk with drift.  $\hat{\mu}$  and  $\hat{\sigma}$  refer to the drift and standard deviation of the fitted processes. Values of the correlation matrix are given in italic.

Australia												
Male						Female						
	$\hat{\mu}$	$\hat{\sigma}$	$\alpha_t$	$\omega_t$	$s_{0,t}$	$s_{1,t}$	$\hat{\mu}$	$\hat{\sigma}$	$\alpha_t$	$\omega_t$	$s_{0,t}$	$s_{1,t}$
HS1							HS1					
$\alpha_t$	-0.022	0.030	1.000	0.406			-0.023	0.026	1.000	0.276		
$\omega_t$	-0.010	0.049	0.406	1.000			-0.011	0.049	0.276	1.000		
HS2							HS2					
$\alpha_t$	-0.015	0.037	1.000	0.482	-0.610		-0.012	0.041	1.000	0.594	-0.781	
$\omega_t$	-0.007	0.055	0.482	1.000	-0.364		-0.008	0.054	0.594	1.000	-0.538	
$s_{0,t}$	-0.055	0.189	-0.610	-0.364	1.000		-0.074	0.237	-0.781	-0.538	1.000	
HS3							HS3					
$\alpha_t$	-0.019	0.029	1.000	0.384		0.138	-0.018	0.027	1.000	0.503		0.361
$\omega_t$	0.000	0.078	0.384	1.000		0.779	-0.003	0.069	0.503	1.000		0.763
$s_{1,t}$	0.071	0.295	0.138	0.779		1.000	0.071	0.236	0.361	0.763		1.000
HS4							HS4					
$\alpha_t$	-0.018	0.043	1.000	0.040	-0.732	-0.401	-0.016	0.045	1.000	0.180	-0.801	-0.347
$\omega_t$	-0.005	0.071	0.040	1.000	0.280	0.668	-0.007	0.066	0.180	1.000	0.128	0.619

Table B.1. Continued.

Australia												
Male						Female						
	$\hat{\mu}$	$\hat{\sigma}$	$\alpha_t$	$\omega_t$	$s_{0,t}$	$s_{1,t}$	$\hat{\mu}$	$\hat{\sigma}$	$\alpha_t$	$\omega_t$	$s_{0,t}$	$s_{1,t}$
$s_{0,t}$	-0.019	0.258	-0.732	0.280	1.000	0.653	-0.022	0.309	-0.801	0.128	1.000	0.637
$s_{1,t}$	0.032	0.344	-0.401	0.668	0.653	1.000	0.038	0.296	-0.347	0.619	0.637	1.000
Gompertz						Gompertz						
	$\hat{\mu}$	$\hat{\sigma}$	$k_{1,t}$	$k_{2,t}$			$\hat{\mu}$	$\hat{\sigma}$	$k_{1,t}$	$k_{2,t}$		
$k_{1,t}$	-0.043	0.100	1.000	-0.957			-0.042	0.100	1.000	-0.969		
$k_{2,t}$	0.000	0.001	-0.957	1.000			0.000	0.002	-0.969	1.000		
England and Wales												
Male						Female						
	$\hat{\mu}$	$\hat{\sigma}$	$\alpha_t$	$\omega_t$	$s_{0,t}$	$s_{1,t}$	$\hat{\mu}$	$\hat{\sigma}$	$\alpha_t$	$\omega_t$	$s_{0,t}$	$s_{1,t}$
HS1							HS1					
$\alpha_t$	-0.018	0.028	1.000	0.507			-0.018	0.025	1.000	0.428		
$\omega_t$	-0.010	0.045	0.507	1.000			-0.010	0.049	0.428	1.000		
HS2							HS2					
$\alpha_t$	-0.013	0.026	1.000	0.592	-0.294		-0.008	0.026	1.000	0.448	-0.434	
$\omega_t$	-0.007	0.048	0.592	1.000	-0.296		-0.007	0.049	0.448	1.000	-0.089	
$s_{0,t}$	-0.041	0.156	-0.294	-0.296	1.000		-0.068	0.130	-0.434	-0.089	1.000	
HS3							HS3					
$\alpha_t$	-0.016	0.026	1.000	0.277		-0.263	-0.014	0.024	1.000	0.498		-0.011
$\omega_t$	-0.002	0.061	0.277	1.000		0.680	-0.001	0.051	0.498	1.000		0.350
$s_{1,t}$	0.054	0.196	-0.263	0.680		1.000	0.071	0.113	-0.011	0.350		1.000
HS4							HS4					
$\alpha_t$	-0.016	0.027	1.000	0.377	-0.375	-0.248	-0.012	0.028	1.000	0.198	-0.545	-0.270
$\omega_t$	-0.008	0.051	0.377	1.000	-0.003	0.427	-0.004	0.052	0.198	1.000	0.326	0.422
$s_{0,t}$	-0.012	0.161	-0.375	-0.003	1.000	0.389	-0.017	0.168	-0.545	0.326	1.000	0.605
$s_{1,t}$	0.014	0.157	-0.248	0.427	0.389	1.000	0.042	0.130	-0.270	0.422	0.605	1.000
Gompertz						Gompertz						
	$\hat{\mu}$	$\hat{\sigma}$	$k_{1,t}$	$k_{2,t}$			$\hat{\mu}$	$\hat{\sigma}$	$k_{1,t}$	$k_{2,t}$		
$k_{1,t}$	-0.034	0.082	1.000	-0.944			-0.030	0.088	1.000	-0.961		
$k_{2,t}$	0.000	0.001	-0.944	1.000			0.000	0.001	-0.961	1.000		
France												
Male						Female						
	$\hat{\mu}$	$\hat{\sigma}$	$\alpha_t$	$\omega_t$	$s_{0,t}$	$s_{1,t}$	$\hat{\mu}$	$\hat{\sigma}$	$\alpha_t$	$\omega_t$	$s_{0,t}$	$s_{1,t}$
HS1							HS1					
$\alpha_t$	-0.016	0.030	1.000	0.623			-0.023	0.029	1.000	0.463		
$\omega_t$	-0.012	0.048	0.623	1.000			-0.013	0.056	0.463	1.000		
HS2							HS2					
$\alpha_t$	-0.009	0.030	1.000	0.707	-0.291		-0.007	0.031	1.000	0.689	-0.468	
$\omega_t$	-0.009	0.050	0.707	1.000	-0.298		-0.010	0.058	0.689	1.000	-0.397	
$s_{0,t}$	-0.055	0.140	-0.291	-0.298	1.000		-0.104	0.143	-0.468	-0.397	1.000	

Table B.1. Continued.

France												
Male						Female						
	$\hat{\mu}$	$\hat{\sigma}$	$\alpha_t$	$\omega_t$	$s_{0,t}$	$s_{1,t}$	$\hat{\mu}$	$\hat{\sigma}$	$\alpha_t$	$\omega_t$	$s_{0,t}$	$s_{1,t}$
HS3						HS3						
$\alpha_t$	-0.013	0.030	1.000	0.516		-0.068	-0.016	0.027	1.000	0.530		-0.075
$\omega_t$	-0.001	0.056	0.516	1.000		0.536	-0.003	0.057	0.530	1.000		0.270
$s_{1,t}$	0.076	0.168	-0.068	0.536		1.000	0.089	0.109	-0.075	0.270		1.000
HS4						HS4						
$\alpha_t$	-0.012	0.031	1.000	0.532	-0.383	-0.236	-0.012	0.033	1.000	0.595	-0.582	-0.435
$\omega_t$	-0.006	0.052	0.532	1.000	0.107	0.390	-0.007	0.052	0.595	1.000	-0.212	-0.090
$s_{0,t}$	-0.007	0.183	-0.383	0.107	1.000	0.638	-0.036	0.182	-0.582	-0.212	1.000	0.682
$s_{1,t}$	0.049	0.206	-0.236	0.390	0.638	1.000	0.052	0.152	-0.435	-0.090	0.682	1.000
Gompertz						Gompertz						
	$\hat{\mu}$	$\hat{\sigma}$	$k_{1,t}$	$k_{2,t}$			$\hat{\mu}$	$\hat{\sigma}$	$k_{1,t}$	$k_{2,t}$		
$k_{1,t}$	-0.024	0.077	1.000	-0.924			-0.037	0.091	1.000	-0.954		
$k_{2,t}$	0.000	0.001	-0.924	1.000			0.000	0.001	-0.954	1.000		
Japan												
Male						Female						
	$\hat{\mu}$	$\hat{\sigma}$	$\alpha_t$	$\omega_t$	$s_{0,t}$	$s_{1,t}$	$\hat{\mu}$	$\hat{\sigma}$	$\alpha_t$	$\omega_t$	$s_{0,t}$	$s_{1,t}$
HS1						HS1						
$\alpha_t$	-0.023	0.023	1.000	0.711			-0.034	0.026	1.000	0.684		
$\omega_t$	-0.014	0.055	0.711	1.000			-0.016	0.051	0.684	1.000		
HS2						HS2						
$\alpha_t$	-0.016	0.027	1.000	0.643	-0.550		-0.017	0.031	1.000	0.502	-0.537	
$\omega_t$	-0.009	0.058	0.643	1.000	-0.169		-0.012	0.051	0.502	1.000	0.085	
$s_{0,t}$	-0.055	0.147	-0.550	-0.169	1.000		-0.107	0.124	-0.537	0.085	1.000	
HS3						HS3						
$\alpha_t$	-0.019	0.023	1.000	0.627		0.098	-0.027	0.026	1.000	0.661		0.136
$\omega_t$	0.002	0.065	0.627	1.000		0.518	-0.005	0.053	0.661	1.000		0.279
$s_{1,t}$	0.092	0.192	0.098	0.518		1.000	0.090	0.133	0.136	0.279		1.000
HS4						HS4						
$\alpha_t$	-0.020	0.028	1.000	0.529	-0.596	-0.322	-0.025	0.028	1.000	0.482	-0.379	0.024
$\omega_t$	-0.009	0.053	0.529	1.000	0.055	0.124	-0.012	0.052	0.482	1.000	0.291	0.299
$s_{0,t}$	-0.008	0.186	-0.596	0.055	1.000	0.596	-0.030	0.124	-0.379	0.291	1.000	0.404
$s_{1,t}$	0.028	0.182	-0.322	0.124	0.596	1.000	0.039	0.147	0.024	0.299	0.404	1.000
Gompertz						Gompertz						
	$\hat{\mu}$	$\hat{\sigma}$	$k_{1,t}$	$k_{2,t}$			$\hat{\mu}$	$\hat{\sigma}$	$k_{1,t}$	$k_{2,t}$		
$k_{1,t}$	-0.042	0.078	1.000	-0.962			-0.065	0.074	1.000	-0.945		
$k_{2,t}$	0.000	0.001	-0.962	1.000			0.001	0.001	-0.945	1.000		



Table B.1. Continued.

United States												
Male						Female						
	$\hat{\mu}$	$\hat{\sigma}$	$\alpha_t$	$\omega_t$	$s_{0,t}$	$s_{1,t}$	$\hat{\mu}$	$\hat{\sigma}$	$\alpha_t$	$\omega_t$	$s_{0,t}$	$s_{1,t}$
HS1						HS1						
$\alpha_t$	-0.013	0.018	1.000	0.429			-0.014	0.014	1.000	0.418		
$\omega_t$	-0.008	0.022	0.429	1.000			-0.009	0.026	0.418	1.000		
HS2						HS2						
$\alpha_t$	-0.009	0.018	1.000	0.533	-0.346		-0.007	0.020	1.000	0.679	-0.662	
$\omega_t$	-0.006	0.024	0.533	1.000	-0.357		-0.006	0.028	0.679	1.000	-0.607	
$s_{0,t}$	-0.030	0.088	-0.346	-0.357	1.000		-0.048	0.091	-0.662	-0.607	1.000	
HS3						HS3						
$\alpha_t$	-0.011	0.017	1.000	0.348		0.024	-0.011	0.015	1.000	0.554		0.300
$\omega_t$	-0.002	0.034	0.348	1.000		0.753	-0.002	0.034	0.554	1.000		0.774
$s_{1,t}$	0.043	0.123	0.024	0.753		1.000	0.053	0.098	0.300	0.774		1.000
HS4						HS4						
$\alpha_t$	-0.011	0.018	1.000	0.241	-0.388	-0.155	-0.010	0.019	1.000	0.407	-0.628	-0.120
$\omega_t$	-0.005	0.033	0.241	1.000	0.263	0.704	-0.005	0.033	0.407	1.000	0.057	0.637
$s_{0,t}$	-0.007	0.099	-0.388	0.263	1.000	0.567	-0.013	0.102	-0.628	0.057	1.000	0.535
$s_{1,t}$	0.021	0.140	-0.155	0.704	0.567	1.000	0.029	0.106	-0.120	0.637	0.535	1.000
Gompertz						Gompertz						
	$\hat{\mu}$	$\hat{\sigma}$	$k_{1,t}$	$k_{2,t}$			$\hat{\mu}$	$\hat{\sigma}$	$k_{1,t}$	$k_{2,t}$		
$k_{1,t}$	-0.021	0.050	1.000	-0.947			-0.021	0.048	1.000	-0.957		
$k_{2,t}$	0.000	0.001	-0.947	1.000			0.000	0.001	-0.957	1.000		

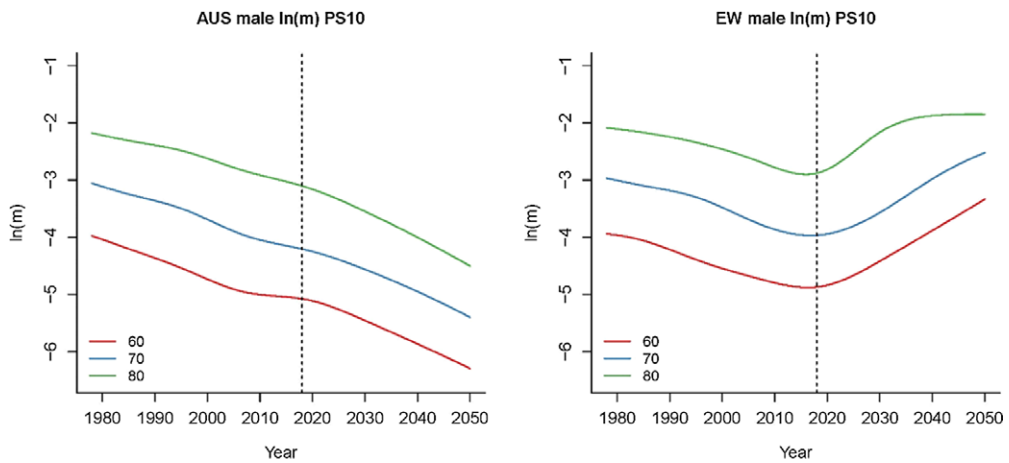
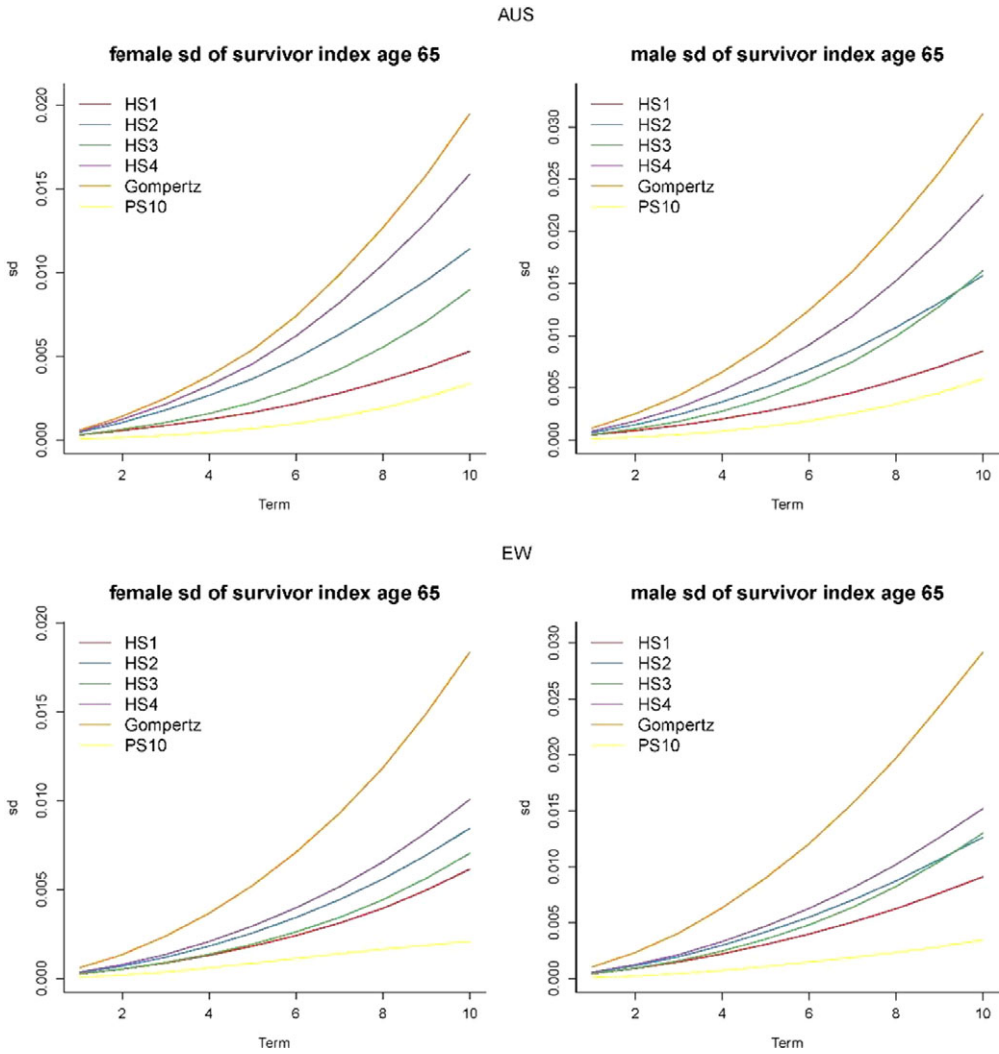


Figure B.1. Fitted (before the dashed line) and extrapolated (after the dashed line) log central death rates under the PS10 model for males in Australia and England and Wales. The dashed line refers to the ending year (2018) of the sample period.



**Figure B.2.** Standard deviation of survivor index  $S(65, t)$  of the cohort aged 65 in 2009,  $t = 1, 2, \dots, 10$ . The mortality models are calibrated on mortality data in Australia and England and Wales over 1950–2008 and projected/simulated over a 10-year forecasting horizon. Univariate RWD is fitted to the time-varying parameters.