International Journal of Forecasting 39 (2023) 450–469

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Multi-population mortality projection: The augmented common factor model with structural breaks

Pengjie Wang, Athanasios A. Pantelous *, Farshid Vahid

Department of Econometrics and Business Statistics, Monash University, Victoria, Australia

ARTICLE INFO

Keywords:
Multi-population mortality projection
Augmented common factor (ACF) model
Structural/trend change
Bayesian statistics
Bayesian forecasting

ABSTRACT

Multi-population mortality forecasting has become an increasingly important area in actuarial science and demography, as a means to avoid long-run divergence in mortality projections. This paper aims to establish a unified state-space Bayesian framework to model, estimate, and forecast mortality rates in a multi-population context. In this regard, we reformulate the augmented common factor model to account for structural/trend changes in the mortality indexes. We conduct a Bayesian analysis to make inferences and generate forecasts so that process and parameter uncertainties can be considered simultaneously and appropriately. We illustrate the efficiency of our methodology through two case studies. Both point and probabilistic forecast evaluations are considered in the empirical analysis. The derived results support the fact that the incorporation of stochastic drifts mitigates the impact of the structural changes in the time indexes on mortality projections.

© 2021 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

Worldwide mortality rates are consistently in decline across all age groups and have in particular been so during the last two decades. Although this decrease in mortality rates is undoubtedly seen as a great achievement of contemporary society, it may impose an inherently disproportionate risk on the liability of public pension schemes and the insurance industry if the characteristics around the mortality improvement are not precisely captured in modelling and forecasting.

A considerable amount of literature has been published on mortality forecasting since the seminal work of Lee and Carter (1992). Their celebrated Lee–Carter (LC) model decomposes the age–time matrix of mortality rates into the age factors and a time index based on principal component analysis, and then extrapolates the motion of the time index to generate age-specific mortality forecasts. Compared with the deterministic approach to forecasting mortality, the LC model, which comprises so-called stochastic mortality modelling, produces not only a single point estimate but also probabilistic credible intervals to quantify the degree of uncertainty underlying the mortality evolution, which is useful in risk management, such as in value-at-risk and hedge-ratio calculations; see for example (Leung et al., 2018, 2021). The LC model has inspired numerous modifications, variations, and extensions. Examples encompass adding additional age–time components (Renshaw & Haberman, 2003), modelling cohort effects (Renshaw & Haberman, 2006), proposing alternative estimation approaches (Brouhns et al., 2002; Czado et al., 2005; Pedroza, 2006), and incorporating climate and economic changes (Boonen & Li, 2017; Dutton et al., 2020; Hanewald, 2011; Hanewald et al., 2011; Niu & Melenberg, 2014; Seklecka et al., 2019, 2017). All of the above models concentrate on mortality forecasting for a single population.

However, mortality improvements across different populations—e.g., both genders in a country, or national populations possessing similar characteristics—are intrinsically correlated. For instance, the spread of COVID-19

* Corresponding author.
E-mail addresses: Pengjie.Wang@monash.edu (P. Wang), Athanasios.Pantelous@monash.edu (A.A. Pantelous), Farshid.Vahid@monash.edu (F. Vahid).

https://doi.org/10.1016/j.ijforecast.2021.12.008
0169-2070/© 2021 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

poses a threat to health worldwide, while the breakthrough in a vaccine against the virus will reduce the infection and hence the fatality rates across countries. The goal of capturing the cross-population dependence in mortality forecasting has led to a proliferation of studies. Cairns et al. (2011) extend the Cairns–Blake–Dowd (CBD) model (Cairns et al., 2006) to a two-population setting to avoid the crossing-over or divergence of joint mortality trends in an unrealistic way. Hyndman et al. (2013) utilise the product-ratio approach to generate coherent forecasts for multiple populations. In a recent paper, to enhance the forecast credibility with the hope of improving the prediction accuracy, Huynh and Ludkovski (2020) propose multi-output Gaussian process models to joint model mortality data, as the mortality experience of a certain group of populations may be a better guide to their future than their own histories would be. Finally, multi-population models can help researchers and end-users gauge the population basis risk (Coughlan et al., 2011), which is defined as the discrepancy between the mortality experience of the exposed population (e.g., the members or small population groups of a specific pension plan) and that of the hedging population (e.g., the national population that determines the time indexes). This enables us to securitise and hedge longevity risks correlated with more than one population or smaller groups of the population where their mortality forecasts should be consistent with those of the national or larger population; see, for example, Li and Hardy (2011), Liu and Li (2016), Pantelous and Zimbidis (2008) and Yang and Wang (2013).

As one of the most popular multi-population mortality models, the augmented common factor (ACF) model proposed by Li and Lee (2005) has been extensively considered in the literature (Antonio et al., 2015; Enchev et al., 2017; Kang et al., 2018; Kleinow, 2015; Li, 2013) and used in practice. One important reason is that the ACF model can be recognised as a natural generalisation of the LC model for multiple populations. Within the ACF model setting, there is one common time index and N population-specific time indexes that govern the change in the mortality experience. Li and Lee (2005) employ the random walk with a constant drift assuming a linear correlation structure in the common time effects. Further, they meticulously model the rest of the time indexes using the mean-reverting process to avoid indefinitely diverging projection results in the long run. Despite its popularity in academia and industry, however, the structural changes exhibited in the common time effects violate the linear correlation assumption, which is so far understudied.

To identify the structural changes in mortality projections for a single population, Van Berkum et al. (2016) use the Bayesian information criterion to determine the number of changes in time indexes. Although they incorporate the multiple trend changes into the mortality models in the fitting period, their work makes "no reference to possible future structural breaks" (see, O'Hare & Li, 2015) because the future trend is expected to remain the same as the historical one after the last detected breakpoint. The other main class of methods for dealing with structural change involves assigning more weight (see, Hyndman et al., 2013; Hyndman & Ullah, 2007) to the most recent mortality data so that the forecasting results will not be distorted by the historical data sets. However, the methods in the above two categories neither acknowledge the stochastic nature of the structural change nor allow the mortality trend to change randomly in the future. Therefore, the first contribution of this study is to reformulate the ACF model to explicitly model the dynamic process of the drifts in the common time indexes, and thus to examine the impact of the structural changes on mortality projections.

This paper also contributes to the Bayesian mortality forecasting literature. Bayesian statistics provides an excellent framework for accounting for all sources of uncertainty, including uncertainty from the stochastic process itself, parameter uncertainty, and model risks, when generating probabilistic forecasts (see e.g., Alexopoulos et al., 2019; Antonio et al., 2015; Czado et al., 2005; Dellaportas et al., 2001; Hunt & Blake, 2020a; Pedroza, 2006; Wiśniowski et al., 2015; Wong et al., 2018). Hence, the computing approach we employ here yields a full uncertainty quantification. Bayesian inference is conducted to obtain parameter estimates and generate mortality forecasts. Particularly, we utilise the square-root-form Kalman Filter to boost robustness in sampling the latent states, consisting of the joint distribution of time indexes with stochastic drifts. To illuminate the existence of structural changes in the mortality indexes, we conduct two case studies: one for Australian two-gender mortality projection, and the other to project age-specific mortality for males in selected eurozone countries. In the second case study, we perform exploratory data analysis to categorise the eurozone countries so that the mortality forecast will not be distorted by other populations with marked dissimilarities. Although the model proposed by Liu and Li (2017) permits time-varying drifts to model structural changes in time indexes in the CBD model, the difference between our model and theirs is twofold. Firstly, we work under the ACF model setting for multiple populations. Secondly, in addition to using point forecasts for model comparisons, with the advantage of Bayesian statistics, we demonstrate probabilistic forecast comparisons. The empirical analysis conducted with real data clearly supports the fact that the incorporation of stochastic drifts mitigates the impact of the structural change in the time indexes on the mortality projection.

The outline of this paper is as follows. We firstly formulate the ACF model and its variant in the state-space modelling framework. Next, we explain the methodology for inference and forecasting in detail. We present two case studies to elaborate the application of the models. Finally, we compare the proposed ACF model with the conventional model. We conclude the paper in the last section.

2. Model formulation

2.1. Definitions and notation

The observed central mortality rates for an x-year-old from population i in year t, denoted by $\hat{m}_{x,t,i}$, is calculated

as

$$\hat{m}_{x,t,i} = \frac{d_{x,t,i}}{e_{x,t,i}},$$

where $d_{x,t,i}$ is the observed number of deaths in population i occurring at age x last birthday in year t, and $e_{x,t,i}$ is the corresponding exposure to risk, i.e., the average population i aged x during calendar year t.

Vectors and matrices are written in bold throughout the rest of the paper. log denotes the natural logarithm unless otherwise stated. All other variables, parameters, distributions, and notations presented below are defined accordingly.

2.2. Augmented common factor model

Let $y_{x,t,i} = \log \hat{m}_{x,t,i}$. The augmented common factor (ACF) model proposed by Li and Lee (2005) assumes that

$$y_{x,t,i} = \alpha_{x,i} + \beta_{x,c}\kappa_{t,c} + \beta_{x,i}\kappa_{t,i} + \epsilon_{x,t,i}, \quad \epsilon_{x,t,i} \sim \mathcal{N}(0, \sigma^2_{\epsilon,i}), \tag{1}$$

for age groups $x = x_1, x_2, \ldots, x_p$, time periods $t = t_0 + 1, t_0 + 2, \ldots, t_0 + n$, and populations $i = 1, \ldots, N$, where $\mathcal{N}(\cdot, \cdot)$ denotes the Gaussian distribution. All the sets of parameters in the ACF model have interpretable meanings: $\alpha_{x,i}$ denotes the baseline shape of $y_{x,t,i}$ over time t for population i; $\beta_{x,c}$ represents the age-specific sensitivity with respect to changes in the common mortality index $\kappa_{t,c}$ for all sub-populations; $\kappa_{t,i}$ measures the population-specific change in mortality deviating from $\kappa_{t,c}$; and $\beta_{x,i}$ denotes the sensitivity of $y_{x,t,i}$ to a change in $\kappa_{t,i}$. The error term, $\epsilon_{x,t,i}$, is assumed to be normally distributed with zero mean and variance $\sigma^2_{\epsilon,i}$ for all ages in population i. Furthermore, we assume that $\epsilon_{x,t,i}$ is independent across different populations, ages, and years.

In Li and Lee (2005), the first-order singular value decomposition (SVD) approach is applied to estimate the parameters. The evolution of $\kappa_{t,c}$ is then extrapolated using a random walk with a constant drift, namely

$$\kappa_{t,c} = \mu_c + \kappa_{t-1,c} + \omega_{t,c}, \quad \omega_{t,c} \sim \mathcal{N}(0, \sigma^2_{\omega,c}), \tag{2}$$

where μ_c is the drift term for the common time-specific factor.

To achieve coherence in the projection, the first-order autoregressive (AR(1)) model is employed to forecast $\kappa_{t,i}$:

$$\kappa_{t,i} = \mu_i + \xi_i\kappa_{t-1,i} + \omega_{t,i}, \quad \omega_{t,i} \sim \mathcal{N}(0, \sigma^2_{\omega,i}), \tag{3}$$

where μ_i is the static drift term for population i, and ξ_i is a constant, with $|\xi_i| < 1$.

It is assumed that the error terms $\omega_{t,c}$ and $\omega_{t,i}$ are independent, since the former is the random effect for the group, and the latter that for the specific population i. Moreover, $\omega_{t,i}$ for the N populations in the group is mutually independent because each error term represents the stochastic innovation in the respective populations. The conventional ACF model specified in (1) is not identifiable until we impose the following parameter constraints: $\sum_{x=x_1}^{x_p} \beta_{x,c} = 1$, $\sum_{t=t_0+1}^{t_0+n} \kappa_{t,c} = 0$, $\sum_{x=x_1}^{x_p} \beta_{x,i} = 1$, and $\sum_{t=t_0+1}^{t_0+n} \kappa_{t,i} = 0$.

It is straightforward to reformulate the ACF model specified by Eqs. (1)–(3) in a state-space representation, which generally includes two components: a state equation that describes the dynamics of the latent states through a Markov process, and a measurement equation that shows the connection between the observations and latent states. Let $\boldsymbol{y}_t = [\boldsymbol{y}_{t,1}^T, \ldots, \boldsymbol{y}_{t,N}^T]^T$, where T denotes the transpose of the matrix/vector, with $\boldsymbol{y}_{t,i} = [y_{x_1,t,i}, \ldots, y_{x_p,t,i}]^T$ for population i, and $\boldsymbol{\kappa}_t = [\kappa_{t,c}, \kappa_{t,1}, \ldots, \kappa_{t,N}]^T$. Then,

$$\begin{cases} \boldsymbol{\kappa}_t = \boldsymbol{D} + \boldsymbol{C}\boldsymbol{\kappa}_{t-1} + \boldsymbol{\omega}_t, & \boldsymbol{\omega}_t \sim \mathcal{N}_{N+1}(\boldsymbol{0}, \boldsymbol{W}), \\ \boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}\boldsymbol{\kappa}_t + \boldsymbol{\epsilon}_t, & \boldsymbol{\epsilon}_t \sim \mathcal{N}_{p\times N}(\boldsymbol{0}, \boldsymbol{V}), \end{cases} \tag{4}$$

where the static parameter vector $\boldsymbol{\Theta} = \{\boldsymbol{D}, \boldsymbol{C}, \boldsymbol{W}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{V}\}$ includes $\boldsymbol{D} = [\mu_c, \mu_1, \ldots, \mu_N]^T$, $\boldsymbol{C} = \text{diag}(1, \xi_1, \xi_2, \ldots, \xi_N)$, and $\boldsymbol{W} = \text{diag}(\sigma^2_{\omega,c}, \sigma^2_{\omega,1}, \ldots, \sigma^2_{\omega,N})$, $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1^T, \ldots, \boldsymbol{\alpha}_N^T]^T$ with $\boldsymbol{\alpha}_i = [\alpha_{x_1,i}, \ldots, \alpha_{x_p,i}]^T$, $\boldsymbol{\beta}_c = [\beta_{x_1,c}, \ldots, \beta_{x_p,c}]^T$, $\boldsymbol{\beta}_i = [\beta_{x_1,i}, \ldots, \beta_{x_p,i}]^T$ and

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_c & \boldsymbol{\beta}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{\beta}_c & \boldsymbol{0} & \boldsymbol{\beta}_2 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\beta}_c & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{\beta}_N \end{bmatrix},$$

here $\boldsymbol{0}$ indicates a column vector with p rows, and $\boldsymbol{V} = \text{diag}(\boldsymbol{V}_1, \boldsymbol{V}_2, \ldots, \boldsymbol{V}_N)$ is a block diagonal matrix with $\boldsymbol{V}_i = \sigma^2_{\epsilon,i}\boldsymbol{1}_p$, and $\boldsymbol{1}_p$ is a p-dimensional identity matrix.

2.3. Augmented common factor model with stochastic drifts

Several researchers (Li et al., 2011; Li & Wong, 2020; O'Hare & Li, 2015; Van Berkum et al., 2016) suggest that there are permanent structural/trend changes presenting in common time effects, which suggests that the random walk with a constant drift may not capture the non-linearity in $\kappa_{t,c}$. It may result in highly significant bias in mortality projections in the long run if we linearly extrapolate the mortality index from the fitting period to the future. Models allowing for the structural changes in the mortality index can be divided into two categories. The first involves identifying the presence of structural breaks in the latent time indexes, and presuming the future morality trend to be fixed after the last break occurs. The second category consists of assigning more weight to the more recent observations of mortality rates. By doing this, researchers can reduce the sensitivity of the parameter estimates to the choice of initial year of the calibration window so that the projection results will be more in line with the most recently observed trends. The question underlying the second way is that it might not be straightforward to optimise the weights based on a rigorous statistical criterion. Besides, neither method acknowledges the stochastic nature of the structural changes nor allows the mortality trend to change randomly in the future.

As it is observed, advances in medicine and a continually improving healthcare system have a significant impact on the speed at which mortality improves. Therefore, it is essential in what follows to model structural changes in common time effects by waiving the random walk assumption with a constant drift. Motivated by the

local-level trend model (Durbin & Koopman, 2012; Petris et al., 2009), we incorporate a time-varying slope in $\kappa_{t,c}$ with all the other model setup unchanged in the original ACF model:

$$\begin{aligned} \kappa_{t,c} &= \mu_{t-1,c} + \kappa_{t-1,c} + \omega_{t,c}, \quad & \omega_{t,c} &\sim \mathcal{N}(0, \sigma^2_{\omega,c}), \\ \mu_{t,c} &= \mu_{t-1,c} + \nu_{t,c}, & \nu_{t,c} &\sim \mathcal{N}(0, \sigma^2_{\nu,c}). \end{aligned} \tag{5}$$

This model specification has several attractive features. The latent dynamic process $\kappa_{t,c}$ can be interpreted as an integrated random walk process. In this regard, we consider that $\omega_{t,c}$ and $\nu_{t,c}$ in the integrated random walk process, $\kappa_{t,c}$, will not be perfectly identified, as the main interest is to gauge the compound effects of the random variations, $\omega_{t,c}$ and $\nu_{t,c}$, on the dynamics of common time effects.[1] Thus, the structural changes in the mortality indexes are modelled in a stochastic framework, which provides more realistic forecasts than the original ACF model does, in which the assumption of the constant drift is not justifiable. The new model offers another advantage, as it avoids discarding the mortality data in the more distant past, which can help forecasters minimise sampling errors and information loss due to a small data size. Last but not least, the proposed model is able to generate coherent predictions, as the ACF model does, when $\kappa_{t,i}$ is a mean-reverting process, for $i \in \{1, 2, \ldots, N\}$. The proposed model can be represented as a state-space form with

$$\begin{cases} \tilde{\boldsymbol{\kappa}}_t = \tilde{\boldsymbol{D}} + \tilde{\boldsymbol{C}}\tilde{\boldsymbol{\kappa}}_{t-1} + \tilde{\boldsymbol{\omega}}_t, & \tilde{\boldsymbol{\omega}}_t \sim \mathcal{N}_{N+2}(\boldsymbol{0}, \tilde{\boldsymbol{W}}), \\ \boldsymbol{y}_t = \boldsymbol{\alpha} + \tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\kappa}}_t + \boldsymbol{\epsilon}_t, & \boldsymbol{\epsilon}_t \sim \mathcal{N}_{p\times N}(\boldsymbol{0}, \boldsymbol{V}), \end{cases} \tag{6}$$

where $\tilde{\boldsymbol{\kappa}}_t = [\kappa_{t,c}, \kappa_{t,1}, \ldots, \kappa_{t,N}, \mu_{t,c}]^T$, the parameter vector in the proposed model $\boldsymbol{\Theta} = \{\tilde{\boldsymbol{D}}, \tilde{\boldsymbol{C}}, \tilde{\boldsymbol{W}}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{V}\}$ with $\tilde{\boldsymbol{D}}^T = [0, \mu_1, \ldots, \mu_N, 0]$, $\tilde{\boldsymbol{W}} = \text{diag}(\sigma^2_{\omega,c}, \sigma^2_{\omega,1}, \ldots, \sigma^2_{\omega,N}, \sigma^2_{\nu,c})$,

$$\tilde{\boldsymbol{C}} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 1 \\ 0 & \xi_1 & \ddots & \vdots & 0 \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & \xi_N & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix},$$

and $\tilde{\boldsymbol{\beta}} = [\boldsymbol{\beta}, \boldsymbol{0}]$, where $\boldsymbol{0}$ is a column vector with dimension $Np \times 1$.

To be consistent with the conventions of the ACF model, we adopt the parameter constraints scheme under the state-space formulation. However, our preliminary simulation results show that the constraints proposed by Li and Lee (2005) will lead to numerical instability when sampling $\beta_{x,i}$. A possible explanation for this might be that $\beta_{x,i}$ can be less than zero for some age groups, and when we impose $\sum_{x=x_1}^{x_p} \beta_{x,i} = 1$ on the series of parameters, the simulation results are no longer limited to a particular range (see, also Hunt & Blake, 2020b). To circumvent this numerical issue, we instead propose that the constraint for $\beta_{x,i}$ is given by

$$\sum_{x=x_1}^{x_p} |\beta_{x,i}| = 1,$$

which has been used in practice, for example, in Hunt and Blake (2015). The above parameter constraints imposed in both the classical model and the updated one do not affect the models' fit and mortality prediction.

At this point, we should report that the stochastic drifts can also be modelled by combining a Bernoulli random variable, which indicates whether the trend change occurs, with a Gaussian random variable, which denotes the magnitude of the trend change. As far as we are concerned, it is a suitable choice for capturing the transitory change in time effects arising from short-term catastrophic mortality events.[2]

3. Bayesian inference and forecasting

Bayesian approaches are adopted in the relevant literature to perform parameter inference and mortality forecasting. A key benefit is that when generating forecasting results they take into consideration all sources of risk, encompassing uncertainties arising from the underlying stochastic processes, those due to model parameters, and, of course, those due to the model itself. In this section, we discuss a block-based Gibbs sampler for model estimation, together with details regarding Bayesian forecasting.

Within the Bayesian regime, we aim at determining the joint posterior density $p(\boldsymbol{\kappa}_{t_0:t_0+n}, \boldsymbol{\Theta}|\boldsymbol{y}_{t_0+1:t_0+n})$ in the ACF model with the initial distribution $\boldsymbol{\kappa}_{t_0} \sim \mathcal{N}_{N+1}(\boldsymbol{m}_{t_0}, \boldsymbol{C}_{t_0})$, where $\boldsymbol{\kappa}_{t_0:t_0+n} := \{\boldsymbol{\kappa}_{t_0}, \boldsymbol{\kappa}_{t_0+1}, \ldots, \boldsymbol{\kappa}_{t_0+n}\}$, $\boldsymbol{y}_{t_0+1:t_0+n} := \{\boldsymbol{y}_{t_0+1}, \ldots, \boldsymbol{y}_{t_0+n}\}$. Correspondingly, in the proposed ACF model, we need to determine the distribution of interest $p(\tilde{\boldsymbol{\kappa}}_{t_0:t_0+n}, \boldsymbol{\Theta}|\boldsymbol{y}_{t_0+1:t_0+n})$ with $\tilde{\boldsymbol{\kappa}}_{t_0} \sim \mathcal{N}_{N+2}(\tilde{\boldsymbol{m}}_{t_0}, \tilde{\boldsymbol{C}}_{t_0})$, where $\tilde{\boldsymbol{\kappa}}_{t_0:t_n}$ denotes the collective latent states involving the stochastic drifts $\boldsymbol{\mu}_{t_0:t_0+n} := \{\mu_{t_0,c}, \mu_{t_0+1,c}, \ldots, \mu_{t_0+n,c}\}$.

3.1. Sampling for latent states

In this section, we present the sampling algorithm for the latent states in the ACF model without stochastic drift, which is extended in a straightforward manner to its variant that considers structural changes in time effects. The sampling approaches based on the Metropolis–Hastings algorithm for the latent states may lead to computing inefficiency; see Antonio et al. (2015) and Czado et al. (2005). Thus, following Carter and Kohn (1994) and Frühwirth-Schnatter (1994), we obtain samples of the latent states as a block using the forward filtering backward smoothing (FFBS) algorithm.

In the filtering phase of FFBS, we derive $p(\boldsymbol{\kappa}_t|\boldsymbol{y}_{t_0+1:t})$ from $p(\boldsymbol{\kappa}_{t-1}|\boldsymbol{y}_{t_0+1:t-1})$ for $t = t_0 + 1, t_0 + 2, \ldots, t_0 + n$. Let $\boldsymbol{\kappa}_{t-1}|\boldsymbol{y}_{t_0+1:t-1} \sim \mathcal{N}(\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1})$. The predicted distributions of $\boldsymbol{\kappa}_t$ and $\boldsymbol{y}_t$, given all the observations up to $t-1$, are provided by

$$\boldsymbol{\kappa}_t|\boldsymbol{y}_{t_0+1:t-1} \sim \mathcal{N}_{N+1}(\boldsymbol{a}_t, \boldsymbol{R}_t), \tag{7}$$

with

$$\boldsymbol{a}_t = \boldsymbol{D} + \boldsymbol{C}\boldsymbol{m}_{t-1} \quad \text{and} \quad \boldsymbol{R}_t = \boldsymbol{C}\boldsymbol{C}_{t-1}\boldsymbol{C}^T + \boldsymbol{W},$$

[1] See the estimation algorithm procedure discussed in Section 3.

[2] However, this direction is beyond of the scope of this paper, as the main target of it is to examine the impact of *permanent* trend changes on the future mortality experience. Modelling transitory trend changes in time effects under the Bayesian paradigm will be the focus of a follow-up paper.

and

$$\boldsymbol{y}_t|\boldsymbol{y}_{t_0+1:t-1} \sim \mathcal{N}_{p\times N}(\boldsymbol{f}_t, \boldsymbol{Q}_t), \tag{8}$$

with

$$\boldsymbol{f}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}\boldsymbol{a}_t \quad \text{and} \quad \boldsymbol{Q}_t = \boldsymbol{\beta}\boldsymbol{R}_t\boldsymbol{\beta}^T + \boldsymbol{V}.$$

We recognise that the filtering density can be factorised as

$$p(\boldsymbol{\kappa}_t|\boldsymbol{y}_{t_0+1:t}) \propto p(\boldsymbol{y}_t|\boldsymbol{\kappa}_t)p(\boldsymbol{\kappa}_t|\boldsymbol{y}_{t_0+1:t-1}),$$

where $p(\boldsymbol{y}_t|\boldsymbol{\kappa}_t)$ is regarded as the likelihood function and $p(\boldsymbol{\kappa}_t|\boldsymbol{y}_{t_0+1:t-1})$ as the prior. Hence, the posterior $p(\boldsymbol{\kappa}_t|\boldsymbol{y}_{t_0+1:t})$ is a multivariate normal with mean $\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{R}_t\boldsymbol{\beta}^T\boldsymbol{Q}_t^{-1}(\boldsymbol{y}_t - \boldsymbol{f}_t)$ and covariance $\boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{R}_t\boldsymbol{\beta}^T\boldsymbol{Q}_t^{-1}\boldsymbol{\beta}\boldsymbol{R}_t$.

Moving on to the smoothing phase, we note that the full conditional density of $\boldsymbol{\kappa}_{t_0:t_0+n}$ given all other static parameters and observations can be decomposed as

$$\begin{aligned} &p(\boldsymbol{\kappa}_{t_0:t_0+n}|\boldsymbol{\Theta}, \boldsymbol{y}_{t_0+1:t_0+n}) \\ &= \left[\prod_{t=t_0}^{t_0+n-1} p(\boldsymbol{\kappa}_t|\boldsymbol{y}_{t_0+1:t_0+n}, \boldsymbol{\kappa}_{t_0+1:t_0+n}, \boldsymbol{\Theta})\right] \\ &\quad \times p(\boldsymbol{\kappa}_{t_0+n}|\boldsymbol{y}_{t_0+1:t_0+n}, \boldsymbol{\Theta}) \\ &= \left[\prod_{t=t_0}^{t_0+n-1} p(\boldsymbol{\kappa}_t|\boldsymbol{y}_{t_0+1:t}, \boldsymbol{\kappa}_{t+1}, \boldsymbol{\Theta})\right] \\ &\quad \times p(\boldsymbol{\kappa}_{t_0+n}|\boldsymbol{y}_{t_0+1:t_0+n}, \boldsymbol{\Theta}). \end{aligned} \tag{9}$$

Eq. (9) provides an alternative way to calculate the full conditional density of the time effects directly. Once we have determined the filtering density of $\boldsymbol{\kappa}_{t_0+n}$, we can draw samples for $\boldsymbol{\kappa}_{t_0:t_0+n-1}$, moving backwards via the smoothing density function, $p(\boldsymbol{\kappa}_t|\boldsymbol{y}_{t_0+1:t}, \boldsymbol{\kappa}_{t+1})$. For $t = t_0+n-1, t_0+n-2, \ldots, t_0$,

$$p(\boldsymbol{\kappa}_t|\boldsymbol{y}_{t_0+1:t}, \boldsymbol{\kappa}_{t+1}) \propto p(\boldsymbol{\kappa}_{t+1}|\boldsymbol{\kappa}_t)p(\boldsymbol{\kappa}_t|\boldsymbol{y}_{t_0+1:t}),$$

where $\boldsymbol{\kappa}_{t+1}|\boldsymbol{\kappa}_t \sim \mathcal{N}_{N+1}(\boldsymbol{D} + \boldsymbol{C}\boldsymbol{\kappa}_t, \boldsymbol{W})$ and $\boldsymbol{\kappa}_t|\boldsymbol{y}_{t_0+1:t} \sim \mathcal{N}_{N+1}(\boldsymbol{m}_t, \boldsymbol{C}_t)$. Therefore, it is straightforward to show that

$$\boldsymbol{\kappa}_t|\boldsymbol{y}_{t_0+1:t}, \boldsymbol{\kappa}_{t+1} \sim \mathcal{N}_{N+1}(\boldsymbol{h}_t, \boldsymbol{H}_t) \tag{10}$$

with $\boldsymbol{h}_t = \boldsymbol{m}_t + \boldsymbol{C}_t\boldsymbol{C}^T\boldsymbol{R}_{t+1}^{-1}(\boldsymbol{\kappa}_{t+1} - \boldsymbol{a}_{t+1})$ and $\boldsymbol{H}_t = \boldsymbol{C}_t - \boldsymbol{C}_t\boldsymbol{C}^T\boldsymbol{R}_{t+1}^{-1}\boldsymbol{C}\boldsymbol{C}_t$.

Direct programming of the above algorithm may lead to numerical instability, since the floating error may yield non-symmetric and non-positive-definite covariance matrices. Thus, we exploit the square-root form of the Kalman filter and smoother based on SVD (Wang et al., 1992; Zhang & Li, 1996) to compute the filtering and smoothing covariance matrices. The square-root form of the Kalman filter and smoother can be found in Appendix A in the online supplementary materials.

3.2. Sampling for static parameters

3.2.1. Prior assumptions

We exploit the Metropolis-within-Gibbs method to sample from the conditional density of the static parameters given the latent states and observations. We assume the same conjugate priors for the common static parameters in the ACF model and its variant, which are $\alpha_{x,i} \sim \mathcal{N}(\tilde{\mu}_{\alpha,i}, \tilde{\sigma}^2_{\alpha,i})$, $\beta_{x,c} \sim \mathcal{N}(\tilde{\mu}_{\beta,c}, \tilde{\sigma}^2_{\beta,c})$, $\beta_{x,i} \sim \mathcal{N}(\tilde{\mu}_{\beta,i}, \tilde{\sigma}^2_{\beta,i})$, $\sigma^2_{\epsilon,i} \sim$ inv-$\Gamma(\tilde{a}_{\epsilon,i}, \tilde{b}_{\epsilon,i})$, $\xi_i \sim \mathcal{N}(\tilde{\mu}_{\xi,i}, \tilde{\sigma}^2_{\xi,i})$ truncated to the interval $(-1, 1)$, $\sigma^2_{\omega,c} \sim$ inv-$\Gamma(\tilde{a}_{\omega,c}, \tilde{b}_{\omega,c})$, and $\sigma^2_{\omega,i} \sim$ inv-$\Gamma(\tilde{a}_{\omega,i}, \tilde{b}_{\omega,i})$, where inv-$\Gamma(.,.)$ denotes the inverse-Gamma distribution.[3]. The conjugate normal priors are employed for the fixed drifts in the ACF model, $\mu_c \sim \mathcal{N}(\tilde{\mu}_c, \tilde{\sigma}^2_c)$ and $\mu_i \sim \mathcal{N}(\tilde{\mu}_i, \tilde{\sigma}^2_i)$. Meanwhile, for the additional parameter in the ACF model with stochastic drifts, we propose using $\sigma^2_{v,c} \sim$ inv-$\Gamma(\tilde{a}_{v,c}, \tilde{b}_{v,c})$. All priors are assumed to be independent with each other.

Table 1
Prior assumptions for the static parameters of the ACF model and its variant.

Parameter	Prior distributions
$\alpha_{x,i}$	$\mathcal{N}(0, 10)$
$\beta_{x,c}$	$\mathcal{N}(0, 10)$
$\beta_{x,i}$	$\mathcal{N}(0, 10)$
$\sigma^2_{\epsilon,i}$	inv-$\Gamma(2.1, 0.1)$
μ_c	$\mathcal{N}(0, 10)$
μ_i	$\mathcal{N}(0, 10)$
ξ_i	$\mathcal{N}(0, 10)$ truncated to $[-1, 1]$
$\sigma^2_{\omega,c}$	inv-$\Gamma(2.1, 0.1)$
$\sigma^2_{\omega,i}$	inv-$\Gamma(2.1, 0.1)$
$\sigma^2_{v,c}$	inv-$\Gamma(2.1, 0.1)$

There are two more remarks to be specified. Firstly, we would like the mortality data to determine the projection results; therefore, weakly informative priors are used for the parameters (see Table 1 concerning hyperparameter settings). Secondly, the conjugate and independent priors are exploited mostly for illustrative purposes.

3.2.2. Posterior distributions and sampling

There are analytical forms of posterior distributions for all of the time-fixed parameters except $\beta_{x,c}$ for $x \in \{x_1, \ldots, x_p\}$. Therefore, we need to insert a Metropolis sub-step into the Gibbs sampling method to draw samples for $\beta_{x,c}$. The sampling algorithm for $\beta_{x,c}$ can be found in Appendix B in the online supplementary materials.

3.3. Bayesian forecasting

The h-step-ahead forecast of the mortality rate in the original ACF model, given the available data, can be derived from

$$\begin{aligned} p(\boldsymbol{y}_{t_0+n+h}|\boldsymbol{y}_{t_0+1:t_0+n}) = \int\ & p(\boldsymbol{y}_{t_0+n+h}|\boldsymbol{\kappa}_{t_0+n+h}, \boldsymbol{\Theta}) \\ & \times p(\boldsymbol{\kappa}_{t_0+n+h}|\boldsymbol{\kappa}_{t_0+n+h-1}, \boldsymbol{\Theta}) \\ & \times \cdots \times p(\boldsymbol{\kappa}_{t_0+n}, \boldsymbol{\Theta}|\boldsymbol{y}_{t_0+1:t_0+n}) \\ & \times d\boldsymbol{\kappa}_{t_0+n:t_0+n+h} d\boldsymbol{\Theta}. \end{aligned} \tag{11}$$

Hence, for $h \geq 1$, we can recursively obtain the posterior predictive distributions of y_{t_0+n+h} given $y_{t_0+1:t_0+n}$ using the original ACF model, as shown in Algorithm 1. The

[3] The probability density function of the inverse Gamma random variable with shape $a > 0$ and scale $b > 0$ is $f(x|a, b) \propto x^{-a-1}\exp(-b/x)$.

forecasting distribution in the proposed ACF model with stochastic drifts can be derived accordingly.

Algorithm 1 Bayesian forecast of mortality rates (in logarithm) using the original ACF model.

1: **for** t in $1, \ldots, h$ **do**
2: **for** m in $1, \ldots, M$, where M is the number of simulations after burn-in and thinning **do**
3: Draw $\boldsymbol{\kappa}^{(m)}_{t_0+n+t} \sim \mathcal{N}(\boldsymbol{D}^{(m)} + \boldsymbol{C}^{(m)}\boldsymbol{\kappa}^{(m)}_{t_0+n+t-1}, \boldsymbol{W}^{(m)})$.
4: Draw $\boldsymbol{y}^{(m)}_{t_0+n+t} \sim \mathcal{N}(\boldsymbol{\alpha}^{(m)} + \boldsymbol{\beta}^{(m)}\boldsymbol{\kappa}^{(m)}_{t_0+n+t}, \boldsymbol{V}^{(m)})$.
5: **end for**
6: **end for**

4. Case studies

In this section, we provide a comprehensive illustration using two mortality data sets. The first one contains female and male mortality data from Australia, while the second consists of data from different national populations.

4.1. Australian two-gender mortality projection

4.1.1. Data

In our first case study, we use crude mortality rates for the Australian population ($i = 1$ for males and $i = 2$ for females) from the (Human Mortality Database, 2019) (HMD). The data set covers a period of $n = 60$ years from $t_0 + 1 = 1951$ to $t_0 + n = 2010$ and $p = 90$ age groups ranging from $x_1 = 0, x_2 = 1, \ldots, x_p = 89$. We excluded age groups above 89 in the following empirical analysis due to large sampling errors in senior age groups. The data were accessed in December 2019.

4.1.2. Simulation results

We collect 200,000 sample sets from the joint posterior distribution of the latent states and static parameters. We discard the first half of the simulations, which is treated as the "burn-in", and keep every 100 values after this, to reduce auto-correlation in the simulated samples. We further perform a convergence diagnostic check on the simulated parameters, consisting of trace plots, autocorrelation plots, and Geweke's convergence diagnostic. There is no significant evidence that the designed Markov sampler does not reach convergence.

Figs. 1–4 present the posterior distributions of time effects, age effects, and stochastic drifts from our proposed ACF model, in which structural changes in the time effects are considered, using the Australian two-gender mortality data. Each plot depicts the posterior median (solid line) as well as the equal-tail 95% credible interval (black dashed lines). The width of each credible interval indicates the level of parameter uncertainty entailed. As there are no distinguishable discrepancies between the posterior distributions of time and age effects generated from the original ACF model and from its variant, we display the figures from the latter model for brevity.

As we can see from the top-left corner in Fig. 1, the common hidden states exhibit a downward trend, which means that there is an overall decrease in mortality for Australians within the fitted period. However, the dynamics of $\kappa_{t,c}$ are not perfectly linear, as there is an observed decrease in the slope around the 1970s, demonstrating an accelerating mortality improvement. In other words, the rate of mortality improvement is not uniform across the calibration window. This finding is further confirmed by Fig. 2, which demonstrates the considerable decline in the drift of the prior model for $\kappa_{t,c}$. Compared with the proposed ACF model, the posterior median of μ_c is constant over time, and biased higher than the posterior median of $\mu_{t,c}$.

The pattern of the posterior distributions of $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ (Fig. 3) for both genders coincides with the overall mortality experience concerning age. Moreover, the average male mortality across ages is slightly higher than that for females, and the mortality hump for young 20-year-old males is more pronounced than that of young females. From the top-left plot in Fig. 4, we can see that there is a more rapid decrease in mortality for infants and people around 60 years of age in contrast to the other age groups, while the sensitivity of mortality decline reaches a low point at approximately 26 years old. However, Fig. 4 shows apparent jaggedness in the Bayesian estimation of $\boldsymbol{\beta}_c, \boldsymbol{\beta}_1$, and $\boldsymbol{\beta}_2$ which is highly likely to lead to an unsmoothed mortality projection. The jagged pattern in the posterior estimate is due to our exploitation of the observed mortality rate as the true, unknown value. We argue that the jagged pattern is not reasonable, since we expect that the mortality rate will evolve at a similar speed in adjacent age groups. In the current study, we assume the priors of the age parameters to be independent across age groups. However, this does not allow for smoothness across ages, which results in a jagged pattern in the mortality forecasts and prevents the use of our proposed methods in practice. Further research could be carried out to embody our prior beliefs in mortality data and thereby determine the impact of prior information on the forecasts. Another remark here is that we can incorporate a common age-sensitivity effect (Kleinow, 2015) into our model in future work, as the two genders are seen to share similar age-specific sensitivity in the remaining two figures in Fig. 4, and this may help improve parameter parsimony.

4.1.3. Forecasts

Fig. 5 depicts the posterior median of time effects (represented by the solid line) in the fitted period. We present the posterior predictive distribution with the predictive median of 60-year-ahead forecasts and 95% forecast intervals respectively using solid and dashed lines.

The differences between the dynamics of time effects are highlighted in Fig. 5, which demonstrates the following aspects:

1. The two models yield similar inference results, but highly different predictive distributions, especially for the common time effects. The projection of $\kappa_{t,c}$ in the original ACF model is slightly biased such that it is higher than that in the updated model, which is due to the constant drift being larger than

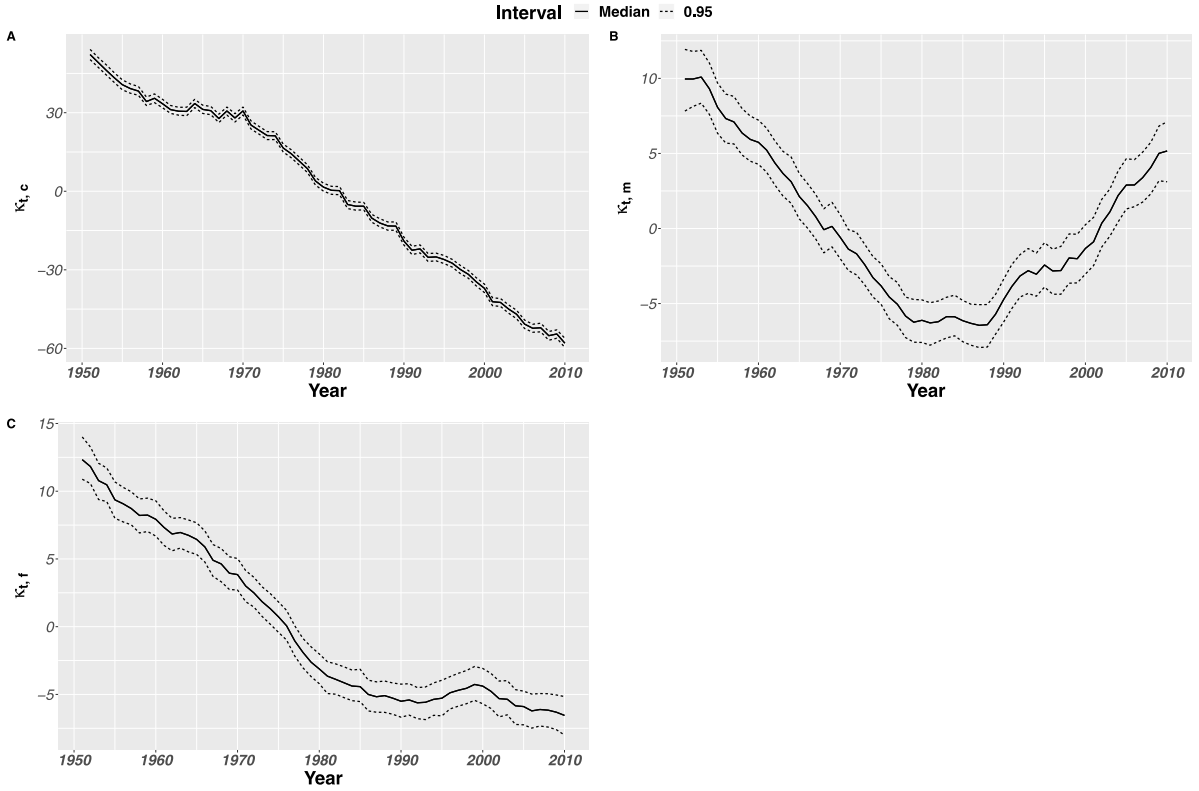


Fig. 1. Posterior distributions of time effects in the ACF model with stochastic drift given Australian two-gender mortality data. The solid lines represent posterior medians, and the dashed lines are 95% credible intervals.

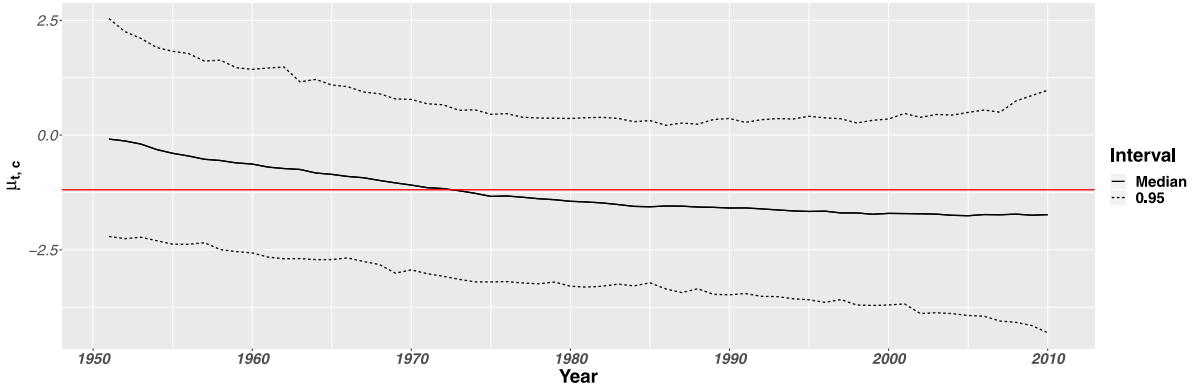


Fig. 2. Posterior distributions of the stochastic drift in the proposed ACF model given Australian two-gender mortality data. The solid lines represent posterior medians, and the dashed lines are 95% credible intervals. The red horizontal solid lines show the corresponding posterior medians of fixed drift in the ACF model.

the forecasted values of stochastic drift. In other words, we expect to see a significant acceleration in the rate of mortality improvement based on the projection results when using the updated model.

2. The choice of the time series model for population-specific time effects means that $\kappa_{t,i}$ will fluctuate around $\mu_i/(1-\xi_i)$ in the future, for $t \to \infty$ and $i \in \{1, 2, \ldots, N\}$. Therefore, the evolution of the predictive mortality rates will be dominated by the common time effects in the long run. It allows us to produce coherent mortality projections for all populations.
3. The third one revealed by Fig. 5 is the level of uncertainty. It is apparent that the updated model results in much wider prediction intervals for $\kappa_{t,c}$, since stochastic drifts are permitted. The randomness from $\mu_{t,c}$ contributes to the width of the forecasting intervals.

Figs. 6 and 7 demonstrate the 60-year-ahead forecasts of central mortality rates (in logarithm) at ages

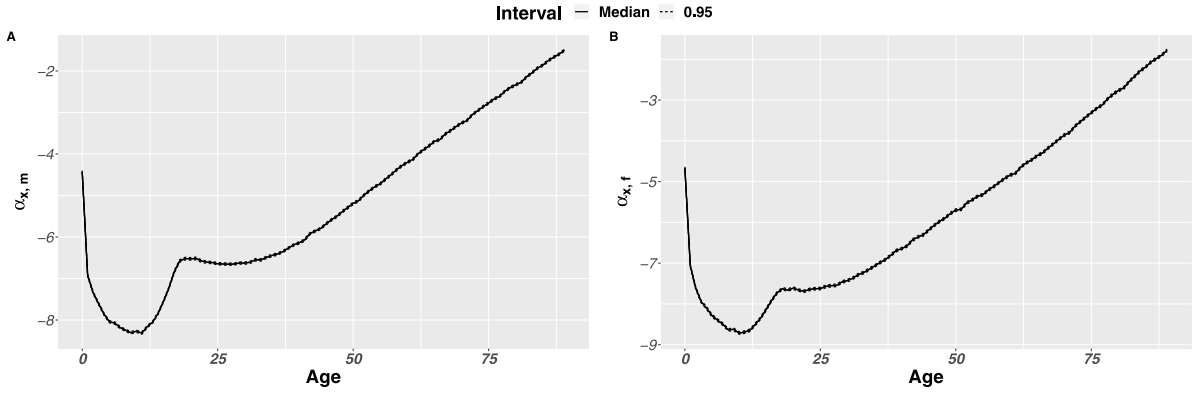


Fig. 3. Posterior distributions of the age-specific mortality profile in the ACF model with stochastic drift given Australian two-gender mortality data. The solid lines represent posterior medians, and the dashed lines are 95% credible intervals.

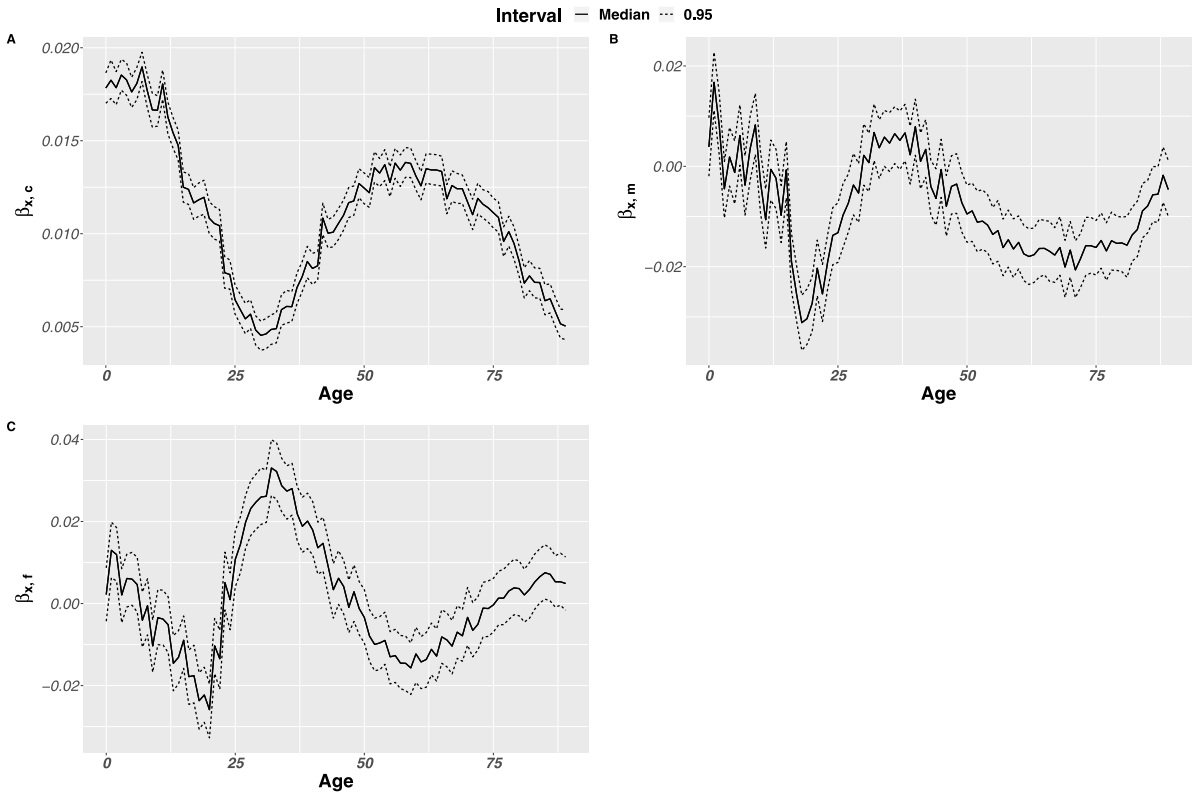


Fig. 4. Posterior distributions of age-specific sensitivity to the change of mortality in the ACF model with stochastic drift given Australian two-gender mortality data. The solid lines represent posterior medians, and the dashed lines are 95% credible intervals.

0, 10, 20, . . . , 80 years, which are respectively generated by the ACF model (red) and the ACF model with stochastic drift (green). The 95% prediction intervals (dashed lines) are displayed in each panel with the posterior median forecasts (solid lines) in the middle of the prediction intervals. The observed historical mortality data (solid points) are also depicted in Figs. 6 and 7. We can observe the following from the two plots:

1. The median forecast of the age-specific mortality rate has improved, along with the dynamics of the common time effects, at all ages, regardless of gender.

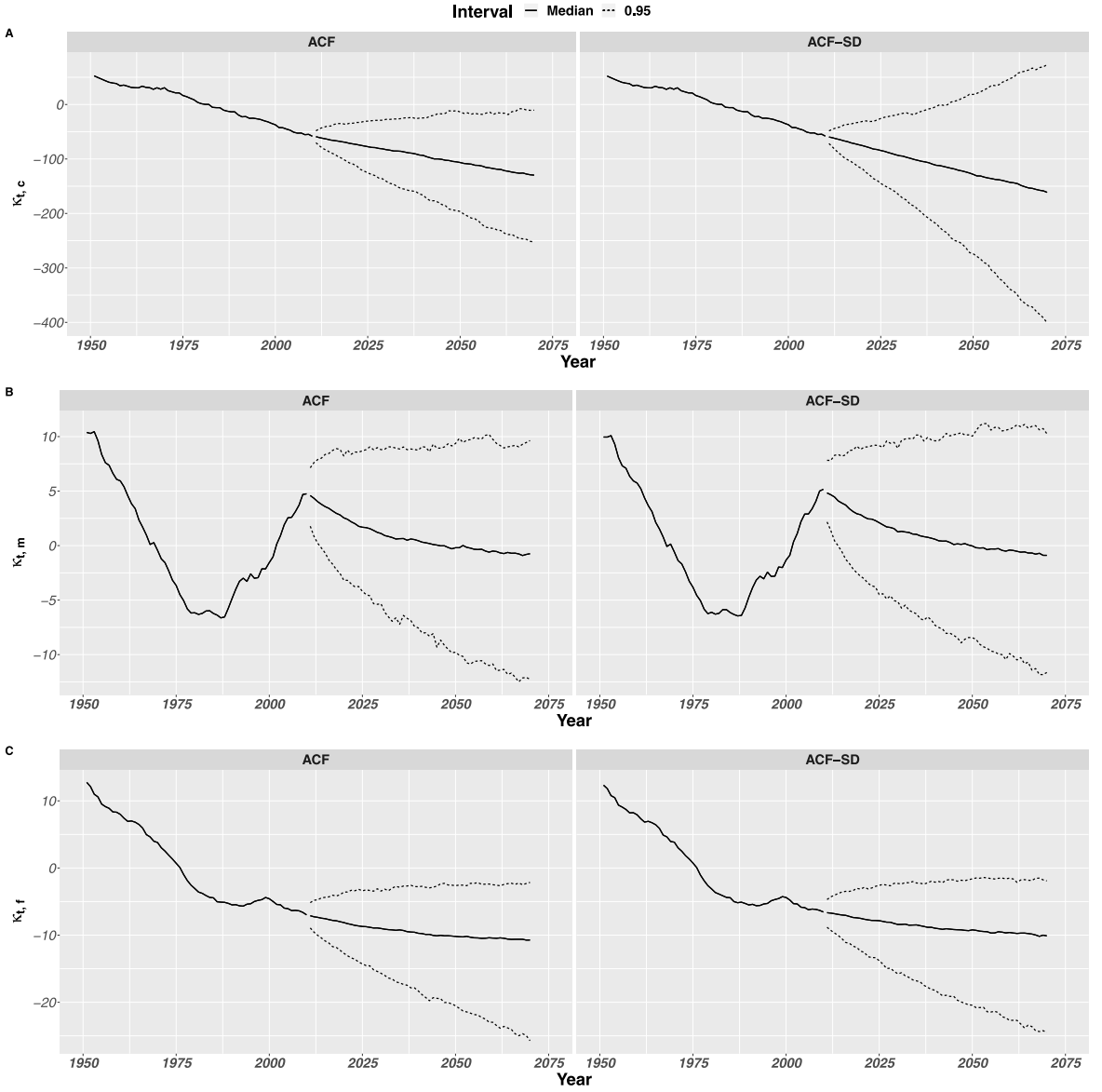


Fig. 5. Predictive distributions of time effects given Australian two-gender mortality data. Panel A shows common time effects $\kappa_{t,c}$, Panel B shows $\kappa_{t,m}$, and Panel C shows $\kappa_{t,f}$. The solid lines in the fitting period (1951–2010) are posterior medians of hidden states. The 95% predictive intervals (dashed lines) with the posterior predictive medians (solid lines) are depicted in the forecasting period (2011–2070).

2. Although the gradient of the central prediction produced from the two models is close to that of the line joining the solid points, particularly for the advanced age groups, our proposed model is able to generate a projection that is more consistent with historical mortality experience than the classical ACF model does.
3. The widths of the prediction intervals are in line with the existence of the stochastic drifts. The prediction intervals generated by the proposed model seem to be adequate with regard to the variation in the historical observed values, but this is not true for the classical ACF model.

4.2. Mortality forecasts for eurozone countries

4.2.1. Data

We use a male mortality data set with a total length of $n = 50$ years, from $t_0+1 = 1959$ to $t_0+n = 2008$ and $p = 90$ age groups from $x_1 = 0$ to $x_p = 89$ from five eurozone countries, namely, France (FRA), (West) Germany (DE), Belgium (BEL), Italy (ITA), and Spain (ESP), as a robustness check to show the proposed model's efficacy.

4.2.2. Exploratory data analysis: Population clustering

One important objective in multi-population models is—by taking commonalities of the historical mortality

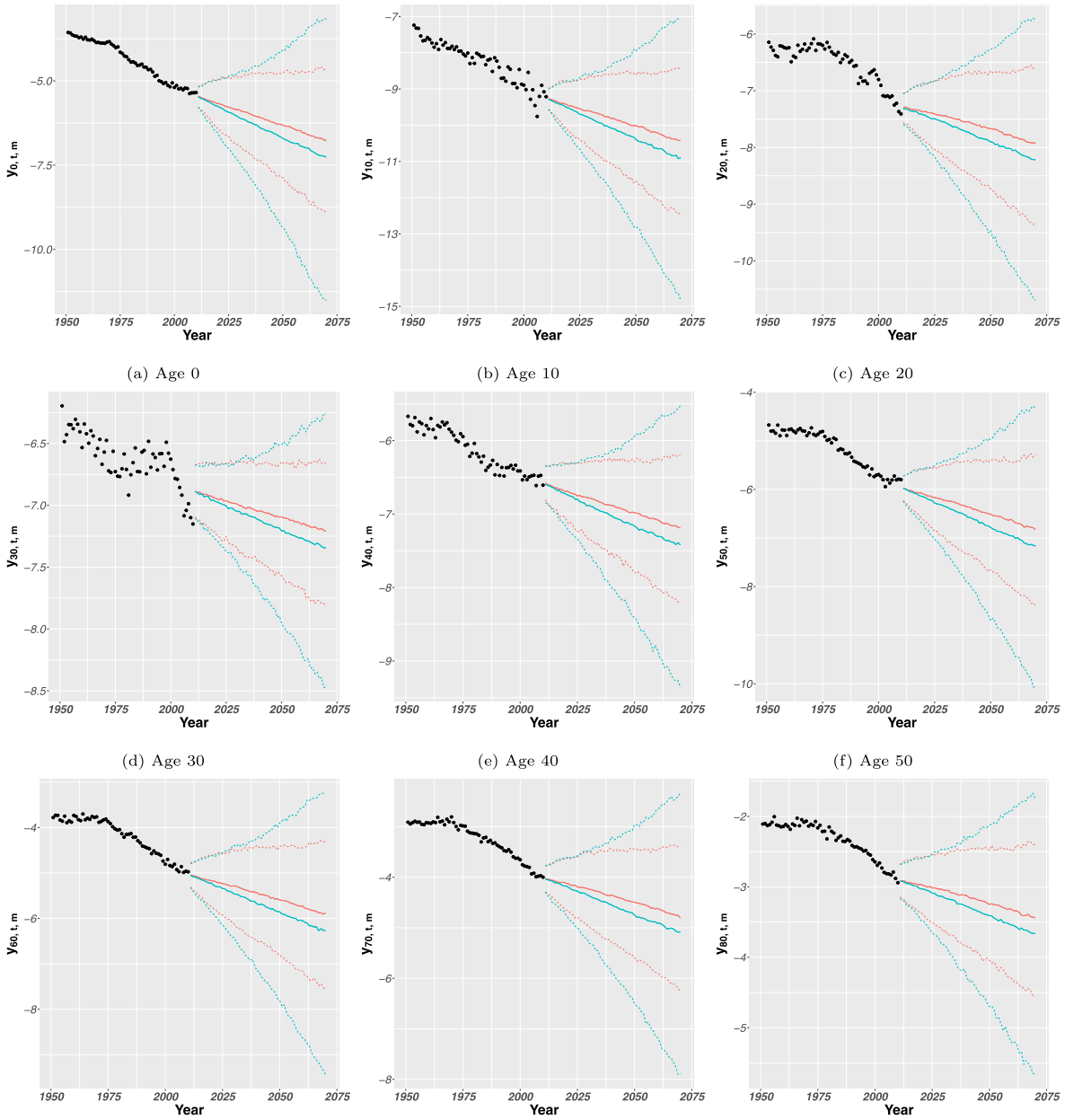
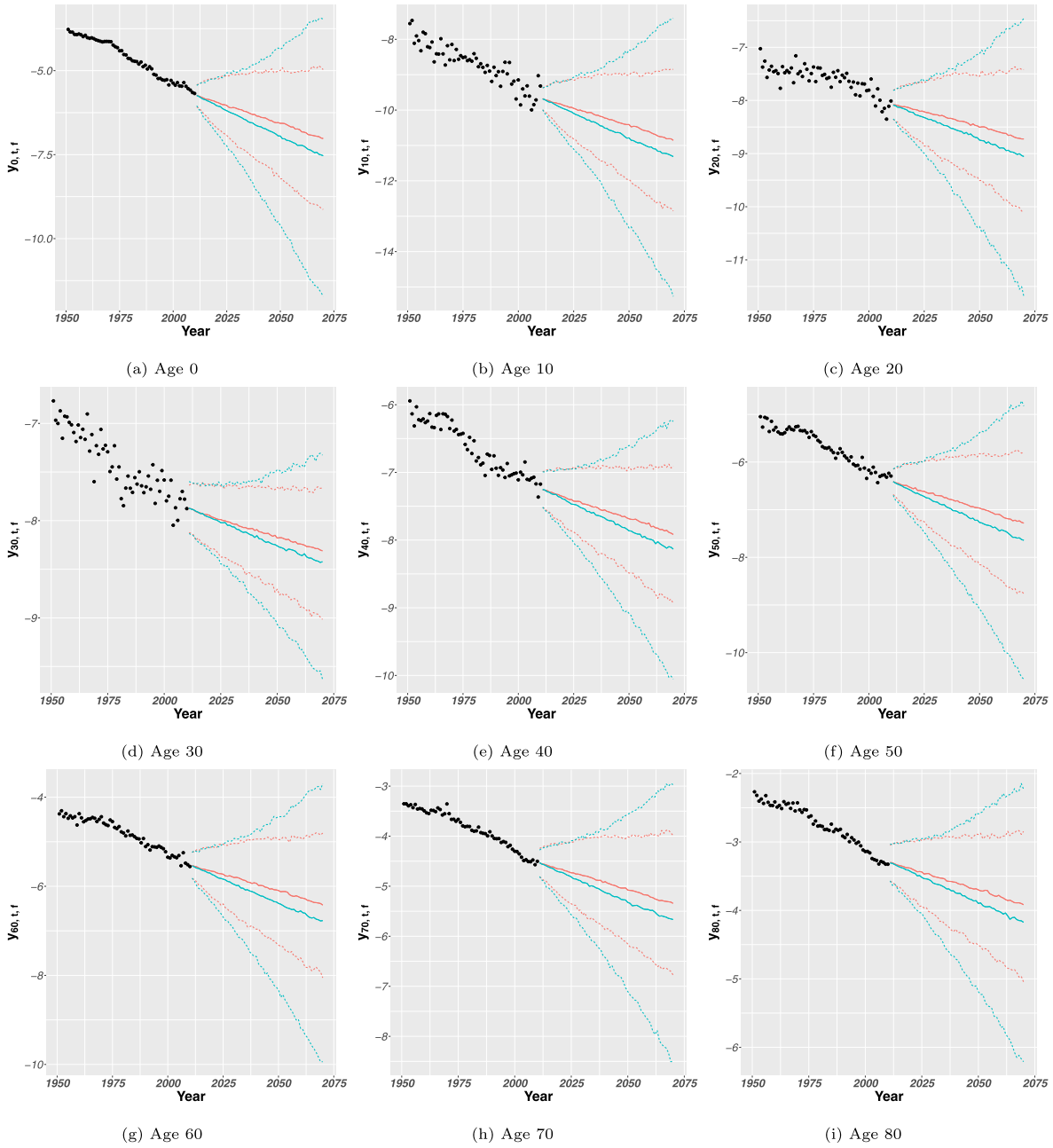


Fig. 6. Mortality forecasts for the central death rates (in logarithm) of Australian males at ages (a) 0 years, (b) 10 years, (c) 20 years, (d) 30 years, (e) 40 years, (f) 50 years, (g) 60 years, (h) 70 years, and (i) 80 years for the 2011–2070 period using the ACF model (red) and the updated ACF model (green). 95% prediction intervals (dashed lines) with the posterior median forecast (solid lines) are depicted in each sub-plot. Solid circles represent historical mortality experience. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

dynamics, namely mortality data integration—to obtain an improvement in predictive credibilities and accuracy. Mortality data fusion in the multi-population models is indeed related to the data selection process. However, it is less clear how to define a group of populations with similar characteristics in their mortality experience so that the modelling and forecasting for a sub-population of the group will not be distorted.

Cluster analysis is a reasonable approach to use to separate populations into homogeneous groups with similar mortality profiles. The mortality profile is depicted by the mortality surface, which is a function of time and age. Borrowing the idea from image recognition,[4] we propose

[4] We are aware that pixels in image recognition are independent of each other. The inter-independence does not apply to time-series



(a) Age 0 (b) Age 10 (c) Age 20

(d) Age 30 (e) Age 40 (f) Age 50

(g) Age 60 (h) Age 70 (i) Age 80

Fig. 7. Mortality forecast for the central death rates (in logarithm) of Australian females at ages (a) 0 years, (b) 10 years, (c) 20 years, (d) 30 years, (e) 40 years, (f) 50 years, (g) 60 years, (h) 70 years, and (i) 80 years for the 2011–2070 period using the ACF model (red) and the updated ACF model (green). 95% prediction intervals (dashed lines) with the posterior median forecast (solid lines) are depicted in each sub-plot. Solid circles represent historical mortality experience. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

a heuristic method for identifying the differences in mortality experience patterns based on hierarchical clustering using principal components (PCs). Population clustering can be conducted as follows:

1. We vectorise the mortality surface of population $i \in \{1, 2, \ldots, N\}$, which is represented by a matrix with dimensions given by period length n and the number of age groups p, into a row vector, denoted as 'Mortality Vec$_i$', with length np. We can then put all the mortality surfaces of N populations together

data such as mortality, as time-series data have high inner dependence. However, establishing a novel clustering method for classifying homogeneous populations with dependence of mortality data is beyond the scope of our paper, and can be left for future research.

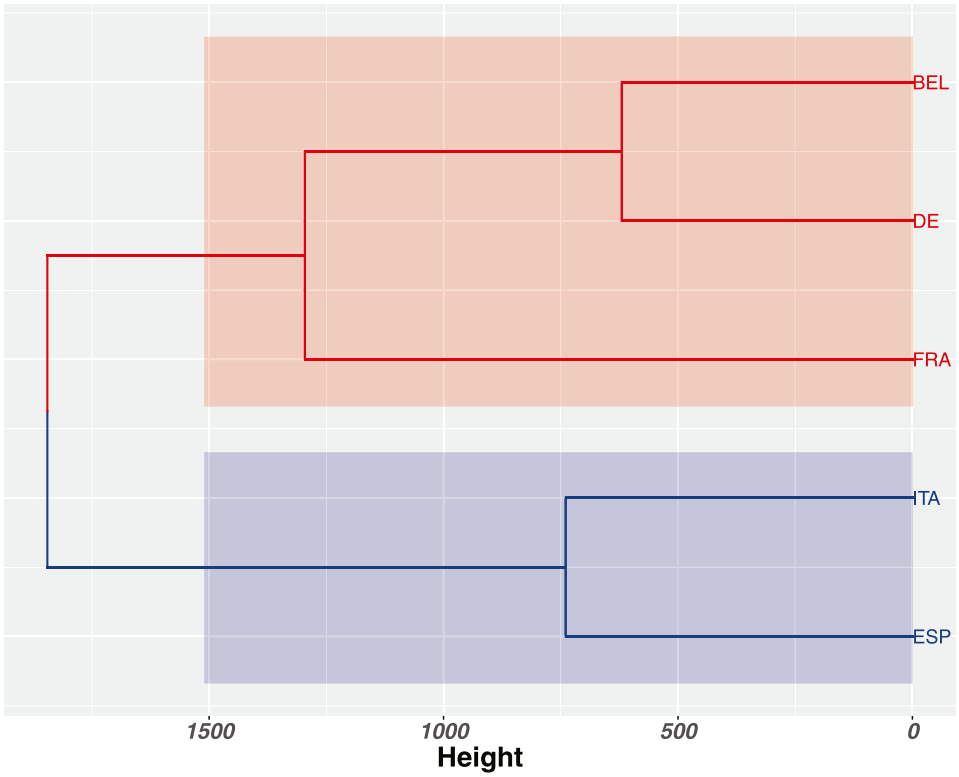


Fig. 8. Dendrogram of the hierarchical clustering of the selected eurozone male mortality profiles. The clustering result is based on mortality data sets for people aged 0–89 years in the period from 1959–2008. Belgium, (West) Germany, and France are classified within Cluster I, while Italy and Spain are within Cluster II.

in one large mortality matrix,

$$\text{Mortality Matrix} = \begin{bmatrix} \text{Mortality Vec}_1 \\ \text{Mortality Vec}_2 \\ \vdots \\ \text{Mortality Vec}_N \end{bmatrix}.$$

2. We perform principal component analysis (PCA) on the above mortality matrix. The number of PCs included in the clustering analysis is dependent on the cumulative percentage of explained variance. The criterion used here is to consider all those PCs up to a predetermined total variance, say 95%.
3. Using the squared Euclidean distance as the metric, and Ward's criterion as the linkage criterion, we perform hierarchical agglomerative clustering on the PCs of each country.

We can utilise the function `HCPC` in the package `FactoMineR` (see, Husson et al., 2010, for details) to implement the hierarchical clustering on the PCs. The result of the clustering is presented in Fig. 8. We can see that the five countries can be separated into two groups according to their distinguishable mortality evolution, albeit all of the countries are within the same monetary union. Cluster I includes Belgium, (West) Germany, and France; the other two countries, Italy and Spain, are classified into the second group. One possible way to explain the dendrogram is through the impact of socioeconomic status on mortality experience. The states in Cluster I have experienced long-term economic development; particularly, significant growth in gross domestic product (GDP) can be seen after the adoption of the euro in 1999. Although there is an increasing trend in GDP initially, due to the growing debt and economic vulnerability in the two southern European countries, Italy and Spain struggled from low GDP growth. Our clustering result also coincides with (Seklecka et al., 2019), who illuminate the discrepancies between and categorise the above five countries based on GDP.[5]

Parameter estimation and projection results can be fully retrieved from the online supplementary materials (see Appendix C).

5. Model comparison

In this section, we compare the original ACF model with its variant using the Bayesian state-space approach in terms of two aspects: in-sample goodness of fit and out-of-sample forecasting performance. Particularly, we treat the forecasting results from the (Lee & Carter, 1992) model for a single population as a benchmark in out-of-sampling validation. The LC model without structural

[5] The clustering method presented here can be regarded as an exploratory data analysis prior to mortality modelling and projection. One promising research direction is to encapsulate population clustering within modelling and forecasting.

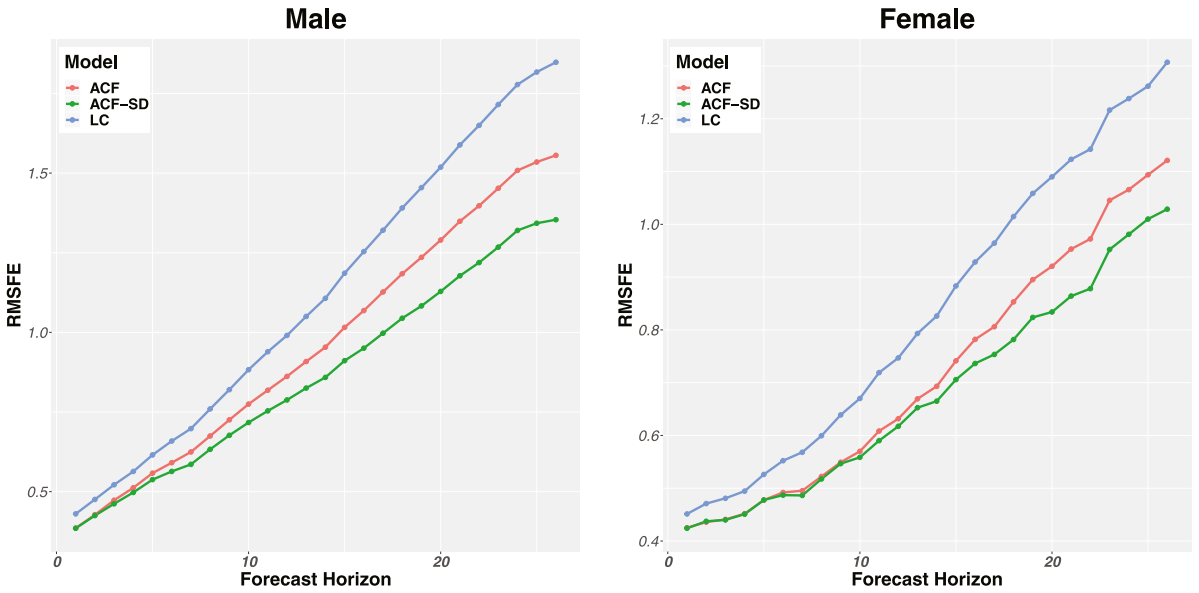


Fig. 9. Comparisons of out-of-sample point forecasts of central mortality rates for Australian males and females, using the classical ACF model (red), the ACF model with stochastic drifts (green), and the LC model (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

change involved is fitted using Pedroza's methodology (Pedroza, 2006), except we adopt the same prior specification as in the proposed model to facilitate comparison later on. Rather than choosing the optimal model, we aim to demonstrate the reasonable forecasts generated by the proposed model. Simultaneously, we attempt to justify whether the incorporation of the stochastic drift can improve model fitting and projection in our empirical analysis.

We evaluate the forecasting performance of the two models based on the Australian two-gender mortality data as follows. We first perform Bayesian inference using the 11 training sets separately, which cover the first s years in the data set with starting year $t_0 + 1 = 1951$, $s = \{30, 31, \ldots, 40\}$. Then, the two models generate the h-step-ahead 'forecast' of the observed crude mortality rate in the testing sets, with $h = 26$ in the first empirical study as it covers the most recently available Australian mortality data from the HMD. A similar validation procedure can be conducted for the eurozone male mortality data sets, with $t_0 + 1 = 1959$, $s = \{20, 21, \ldots, 30\}$, and $h = 28$.

5.1. In-sample fitting

To assess the model fit to the historical data with a penalty for model complexity, we employ the deviance information criterion (DIC) as the Bayesian measure of model complexity and fit. The DIC is defined as

$$\text{DIC} := \mathbb{E}[D(\boldsymbol{\vartheta})|\boldsymbol{y}_{t_0+1:t_0+n}] + p_D,$$

where

- $\boldsymbol{\vartheta}$ is a collection of latent states and static parameters;
- the deviance is defined as

$$D(\boldsymbol{\vartheta}) := -2\log p(\boldsymbol{y}_{t_0+1:t_0+n}|\boldsymbol{\vartheta}) + 2\log h(\boldsymbol{y}_{t_0+1:t_0+n});$$

- $\mathbb{E}[D(\boldsymbol{\vartheta})|\boldsymbol{y}_{t_0+1:t_0+n}]$ evaluates the goodness of fit;
- and $p_D = \mathbb{E}[D(\boldsymbol{\vartheta})|\boldsymbol{y}_{t_0+1:t_0+n}] - D(\mathbb{E}[\boldsymbol{\vartheta}|\boldsymbol{y}_{t_0+1:t_0+n}])$ is the number of effective parameters, and penalises over-fitting issues.

There is no need to consider $h(\boldsymbol{y}_{t_0+1:t_0+n})$ since it is nothing but a constant, which will be cancelled out in the model comparison. The lower the DIC, the better the model fit. The resulting DIC values are shown in Table 2. It is interesting to find that, in most of the training periods, the conventional ACF model performs better than the model with stochastic drifts in terms of in-sample model fit, which is because of model parsimony. However, in-sample fitting does not provide us a piece of evidence regarding which model will outperform in out-of-sample prediction.

5.2. Out-of-sample forecasting

5.2.1. Evaluation of point forecast

We evaluate the predictive accuracy for the age-specific mortality rates in logarithm form using the h-step root mean squared forecast error (RMSFE), which is defined as

$$\text{RMSFE}(h) = \sqrt{\frac{1}{n_p n_s}\sum_x\sum_s (y_{x,t_0+s+h,i} - \hat{y}_{x,t_0+s+h,i})^2}$$

for the forecast of the mortality rate for population i, where n_s is the length of the period in each training set, and $\hat{y}_{x,t_0+s+h,i}$ is the posterior median of the h-step-ahead posterior predictive distribution of the observed $y_{x,t_0+s+h,i}$ given the absolute error loss function employed. The model with smaller RMSFE(h) is preferred. Figs. 9, 10,

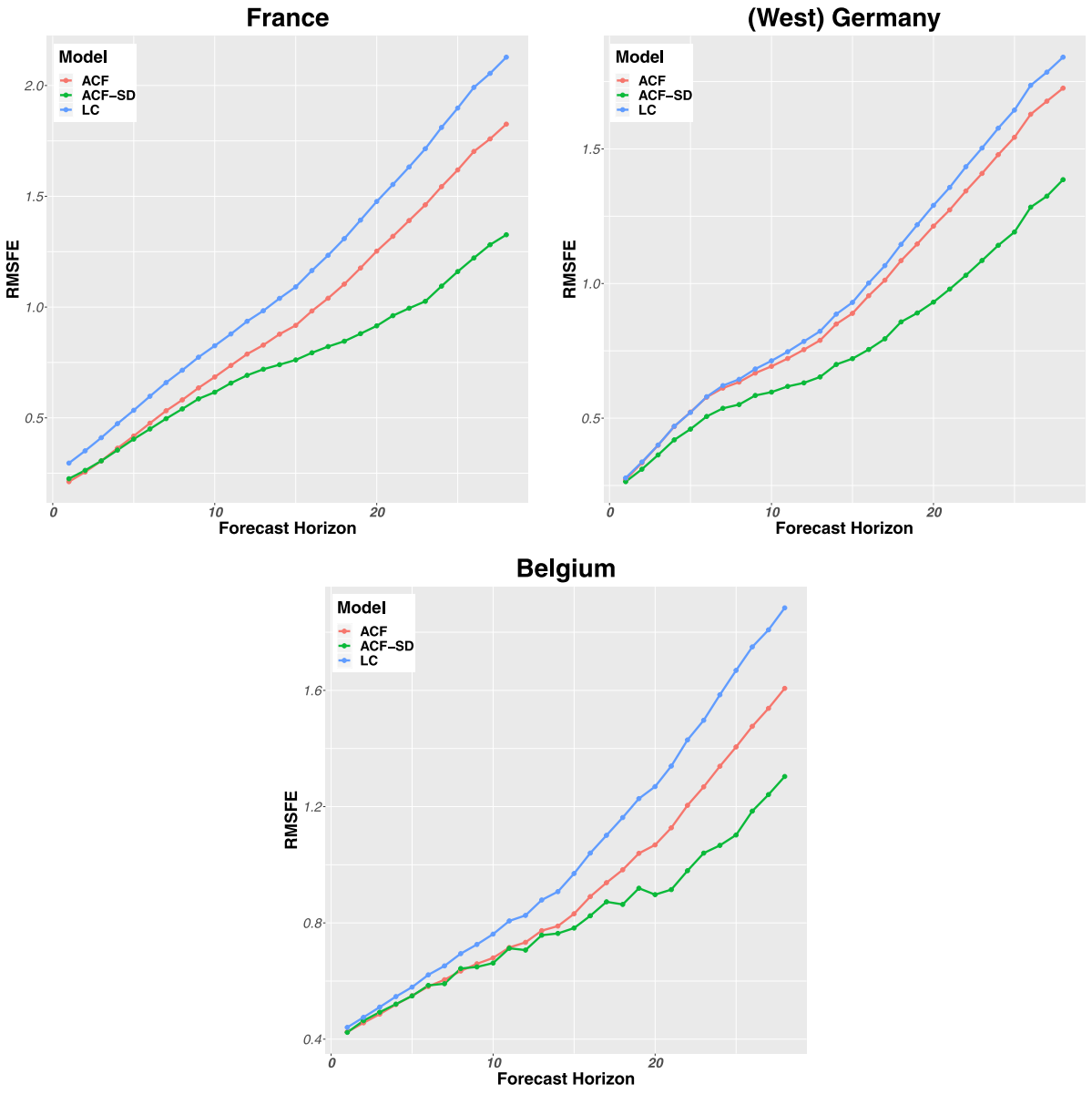


Fig. 10. Comparisons of out-of-sample point forecasts of central mortality rates for eurozone males in Cluster I, using the classical ACF model (red), the ACF model with stochastic drifts (green), and the LC model (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and 11 display the root mean squared prediction error generated by the two models as well as the LC model.

5.2.2. Evaluation of probabilistic forecast

It is acknowledged that the point forecast does not take the full uncertainty over $y_{x,t_0+s+h,i}$ into account, although it is straightforward to compute and, more importantly, to understand. In contrast to most of the existing literature on mortality projection, we exploit the two measures of the predictive accuracy of probabilistic forecasting so as to assess the model performance with more detailed uncertainty quantification. The measures employed here are (proper) scoring rules; see Gneiting and Raftery (2007).

Definition 1 (*Proper Scoring Rule*)**.** Let $y \in \mathbb{R}$ be the observations of the quantity of interest, Y, and $F \in \mathcal{F}$ be the corresponding probabilistic forecast. A scoring rule is a function that assigns a numerical score $S(F, y)$ to each pair (F, y). Furthermore, we say that a scoring rule is proper if the expectation of the scoring rule is optimised when quoting the true distribution as the forecast distribution, namely

$$\mathbb{E}_{Y\sim G}[S(G, Y)] \leq \mathbb{E}_{Y\sim G}[S(F, Y)].$$

The scoring rule is strictly proper only if the above equality holds when $F = G$.

The logarithmic score (LOGS) and continuous ranked probability score (CRPS) are widely used proper scoring

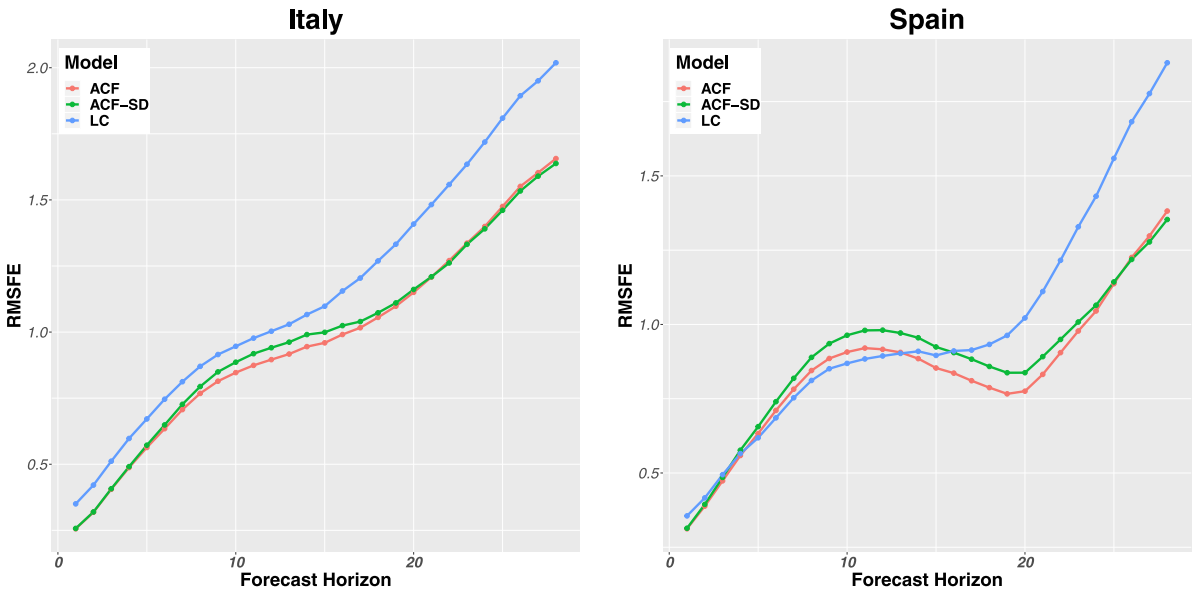


Fig. 11. Comparisons of out-of-sample point forecasts of central mortality rates for eurozone males in Cluster II, using the classical ACF model (red), the ACF model with stochastic drifts (green), and the LC model (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Deviance information criterion for the assessment of the original ACF model and the ACF model with stochastic drift, using data over different calibration windows.

Training period	Australian two-gender case		Training period	Eurozone male case			
				Cluster I		Cluster II	
	ACF	ACF-SD		ACF	ACF-SD	ACF	ACF-SD
1951–1980	−9828.09	−9810.56	1959–1978	−14075.16	−14042.72	−10418.87	−10413.34
1951–1981	−10125.09	−10106.59	1959–1979	−14814.55	−14779.37	−10943.22	−10941.09
1951–1982	−10475.81	−10520.13	1959–1980	−15454.66	−15425.03	−11391.65	−11370.56
1951–1983	−10810.68	−10796.05	1959–1981	−16198.36	−16138.6	−11870.12	−11865.46
1951–1984	−11131.04	−11113.63	1959–1982	−16906.53	−16838.62	−12333.77	−12311.32
1951–1985	−11454.75	−11443.55	1959–1983	−17575.04	−17528.15	−12782.56	−12778.03
1951–1986	−11670.39	−11646.12	1959–1984	−18231.95	−18184.74	−13239.92	−13223.25
1951–1987	−11985.73	−11969.55	1959–1985	−18838.82	−18817.81	−13742.54	−13724.54
1951–1988	−12266.7	−12262.83	1959–1986	−18788.65	−18701.31	−14182.44	−14174.13
1951–1989	−12476.55	−12468.74	1959–1987	−19962.74	−19373.09	−14512.86	−14503.84
1951–1990	−12670.17	−12649.89	1959–1988	−20487.79	−20452.08	−14846.01	−14844.62

Note: Smaller DIC values indicate a better model fit.

rules for evaluating probabilistic forecast performance. The formal definitions are as follows.

Definition 2 (*Logarithmic Score*)**.**The logarithmic score is defined in terms of f, the associated probability density function of F, as

$$\mathrm{LogS}(F, y) := -\log f(y).$$

Definition 3 (*Continuous Ranked Probability Score*)**.** The continuous ranked probability score can be defined as

$$\mathrm{CRPS}(F, y) := \int_{\mathbb{R}} (F(z) - \mathbb{1}_{\{y \le z\}})^2 dz,$$

where $\mathbb{1}$ is the indicator function.

In practice, we cannot know the predictive density f, since the parameters in the density are not known. Moreover, the forecast distribution F may not be analytical either. For these reasons, we need to compute the values of the scoring rules based on the Markov chain Monte Carlo (MCMC) output. The first probabilistic measure of the out-of-sample performance of the mortality rate using the logarithmic score is the h-step-ahead mean logarithmic score,

$$\begin{aligned}\mathrm{LogS}_1(h) &= -\frac{1}{n_p n_s} \log \prod_x \prod_s p_{\mathrm{post}}(y_{x,t_0+s+h,i}) \\ &= -\frac{1}{n_p n_s} \sum_x \sum_s \log \int p(y_{x,t_0+s+h,i}|\boldsymbol{\vartheta}) \\ &\quad \times p(\boldsymbol{\vartheta}|\boldsymbol{y}_{t_0+1:t_0+n}) d\boldsymbol{\vartheta} \\ &= -\frac{1}{n_p n_s} \sum_x \sum_s \log\left(\frac{1}{M} \sum_{m=1}^{M} p(y_{x,t_0+s+h,i}|\boldsymbol{\vartheta}^{(m)})\right),\end{aligned}$$

where $p_{\mathrm{post}}(y_{x,t_0+s+h,i})$ is the predictive density for the observed mortality rate in the out-of-sample data set

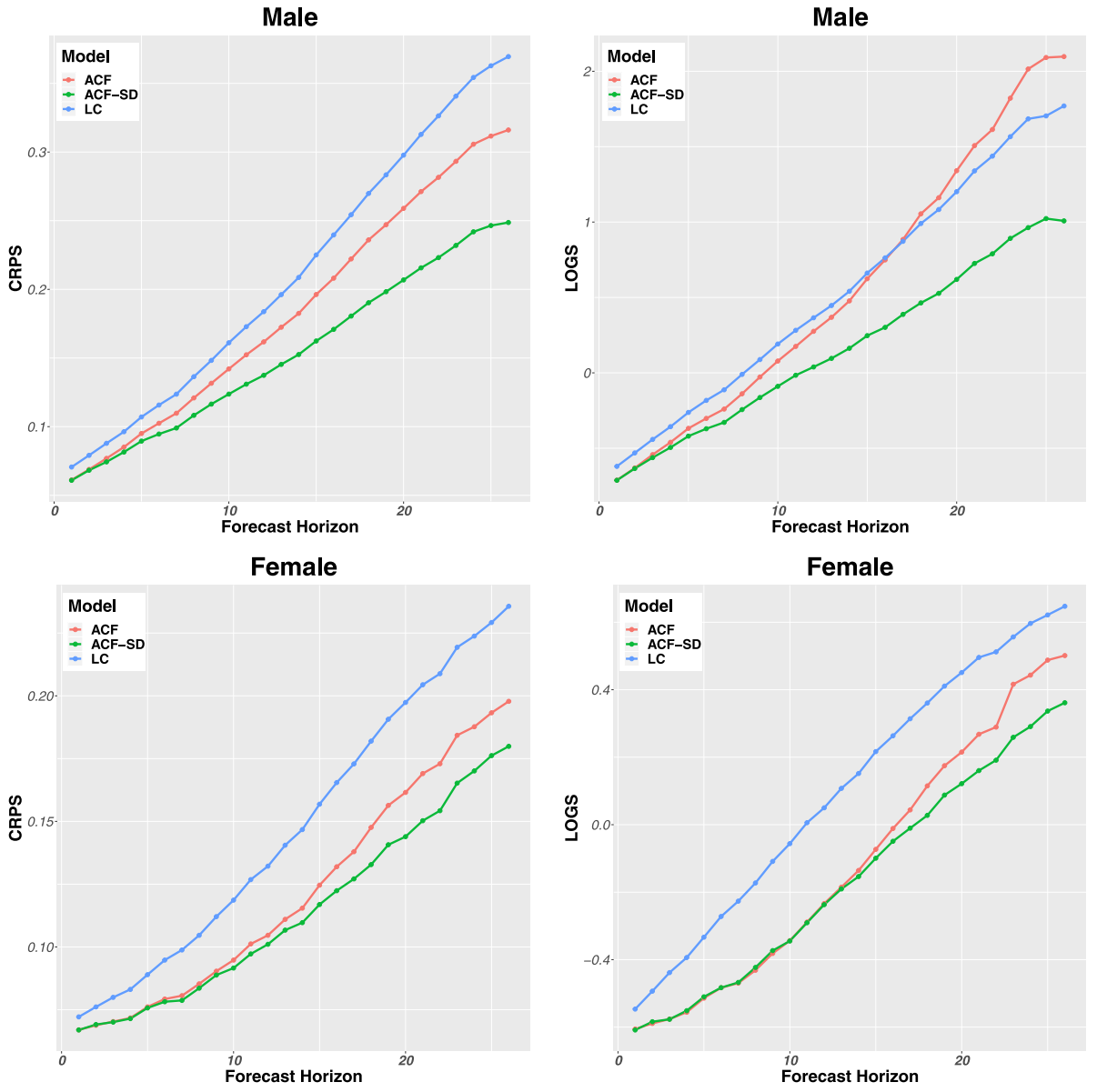


Fig. 12. Comparisons of out-of-sample probabilistic forecasts of central mortality rates for Australian males and females, using the classical ACF model (red), the ACF model with stochastic drifts (green), and the LC model (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

induced by the posterior distribution $p(\boldsymbol{\vartheta}|\boldsymbol{y}_{t_0+1:t_0+n})$, given the sequence of simulations of the parameters $\{\boldsymbol{\vartheta}^{(m)}\}_{m=1}^{M}$. To derive the second measure using CRPS, we need the empirical cumulative density function (CDF) to approximate the predictive CDF, which is

$$\hat{F}_M^{\text{ECDF}}(y) = \frac{1}{M}\sum_{m=1}^{M} \mathbb{1}_{\left\{y_{x,t_0+s+h,i}^{(m)} \le y\right\}},$$

where $\left\{y_{x,t_1+s-1+h,i}^{(m)}\right\}_{m=1}^{M}$ is a group of posterior forecasting samples. The h-step-ahead mean CRPS of the mortality rate can be written as

$$\begin{aligned}
\text{CRPS}_1(h) &= \frac{1}{n_p n_s}\sum_x\sum_s\left(\frac{1}{M}\sum_{m=1}^{M}\left|y_{x,t_0+s+h,i}^{(m)} - y_{x,t_0+s+h,i}\right|\right.\\
&\quad\left. - \frac{1}{2M^2}\sum_{m_1=1}^{M}\sum_{m_2=1}^{M}\left|y_{x,t_0+s+h,i}^{(m_1)} - y_{x,t_0+s+h,i}^{(m_2)}\right|\right)\\
&= \frac{1}{n_p n_s}\sum_x\sum_s\frac{2}{M^2}\\
&\quad\times\sum_{m=1}^{M}\left[\left(y_{x,t_0+s+h,i}^{(m)} - y_{x,t_0+s+h,i}\right)\right.\\
&\quad\left.\times\left(M\mathbb{1}_{\left\{y_{x,t_0+s+h,i}<y_{x,t_0+s+h,i}^{(m)}\right\}} - m + 0.5\right)\right].
\end{aligned}$$

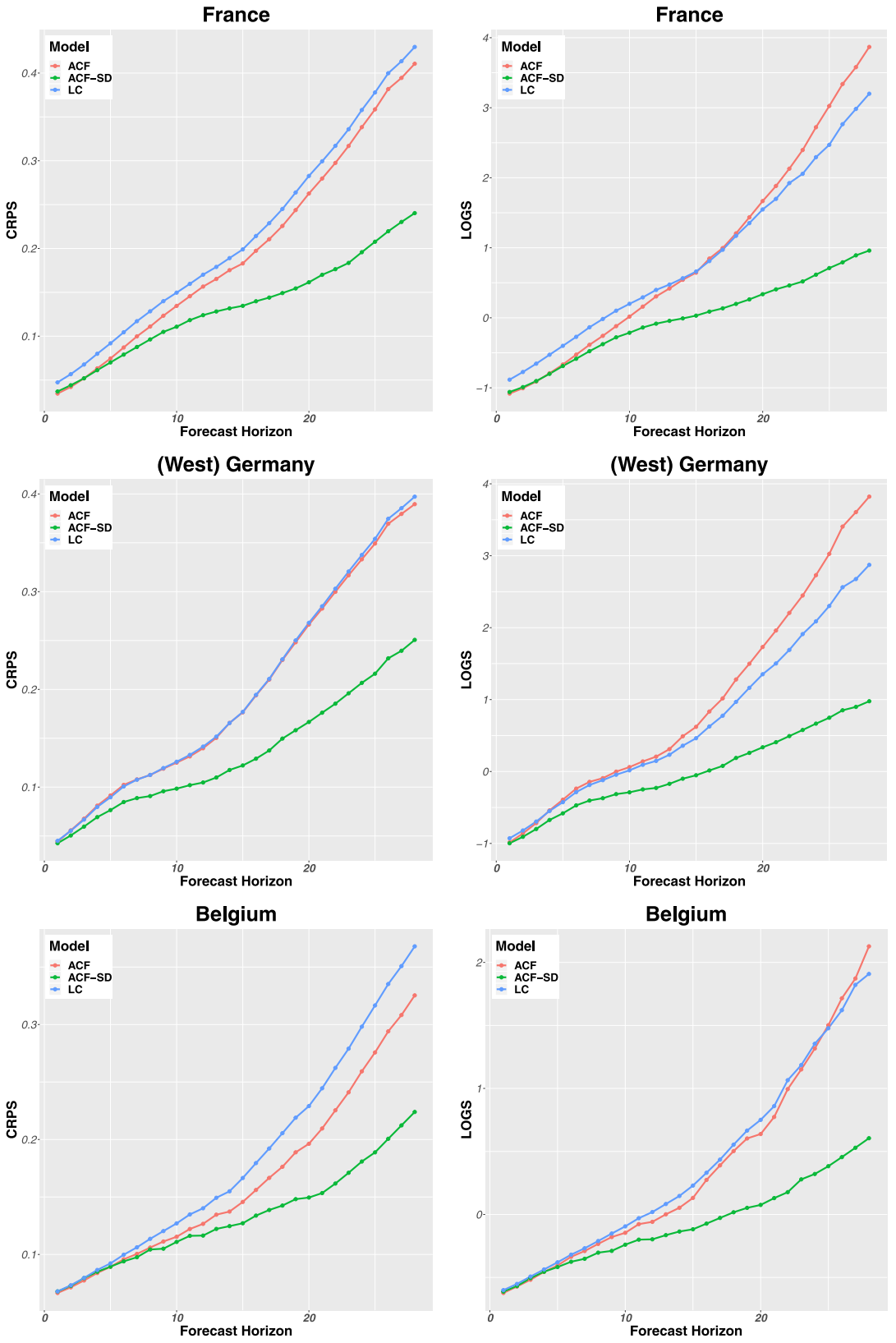


Fig. 13. Comparisons of out-of-sample probabilistic forecasts of central mortality rates for eurozone males in Cluster I, using the classical ACF model (red), the ACF model with stochastic drifts (green), and the LC model (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

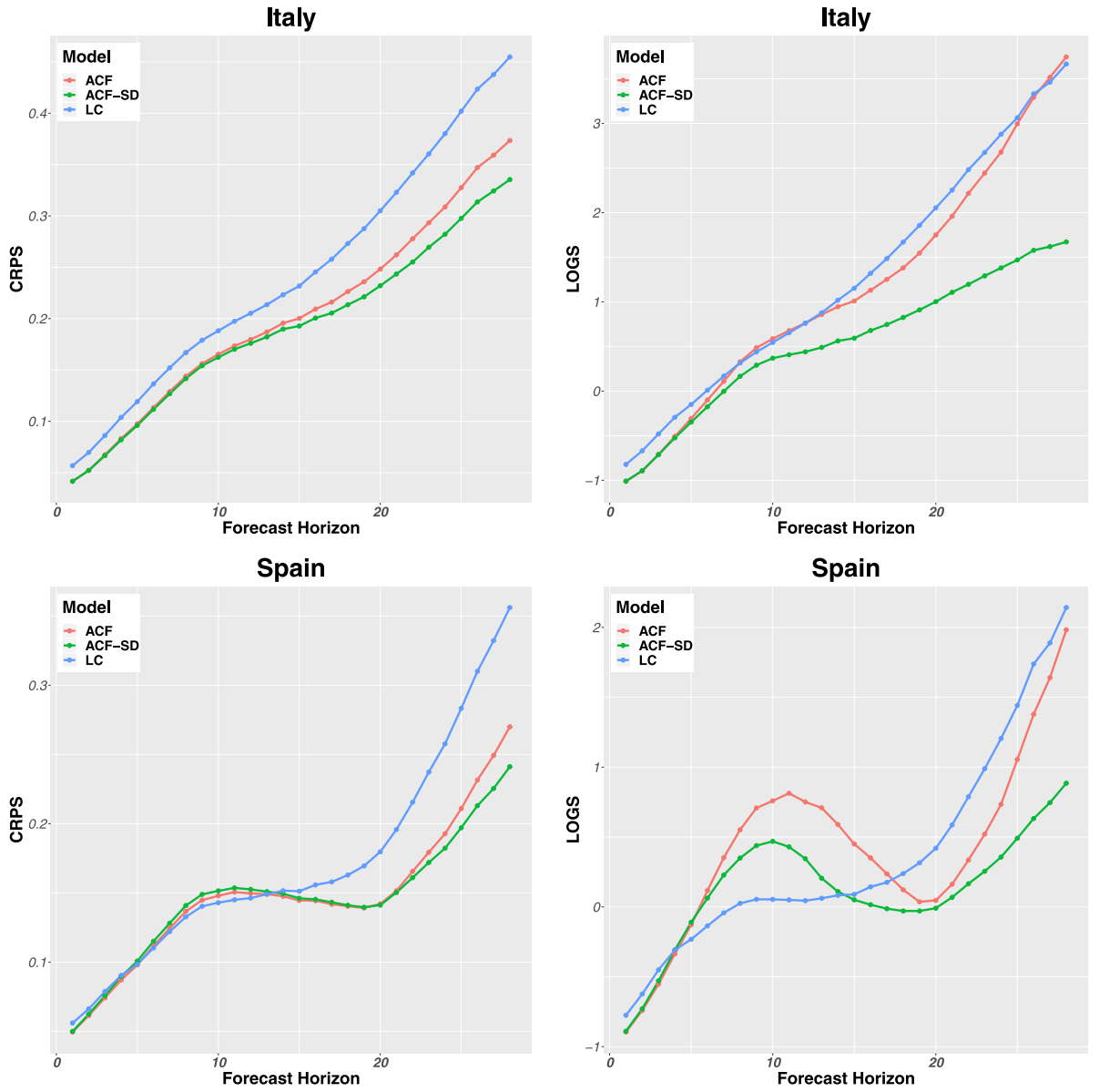


Fig. 14. Comparisons of out-of-sample probabilistic forecasts of central mortality rates for eurozone males in Cluster II, using the classical ACF model (red), the ACF model with stochastic drifts (green), and the LC model (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The final remark regarding the scoring rule is that the rules we consider in our work are negatively oriented, so a model with a lower mean LOGS and CRPS will be judged to be better. Fig. 12 shows the scores for the Australian population, while the following two plots (Figs. 13 and 14) show the scores for the selected eurozone countries.

5.3. Discussion

A comparison of the two models and the (Lee & Carter, 1992) model for a single population in terms of in-sample fitting and out-of-sample forecasting performance reveals the following aspects:

1. There is no doubt that the incorporation of stochastic effects, which may mitigate the impact of the structural change in latent states on mortality projection, will add complexity to statistical inference. However, it is somewhat surprising that the model assessment results using the data in the training period are comparable, illustrated by the indistinguishable DIC values generated by the two models (see Table 2 for reference), although the original ACF model dominates the updated one when using the vast training periods in our two empirical studies. This indicates that the increasing model

complexity due to the introduction of stochastic drifts does not cause a substantial over-fitting issue.

2. The point predictive accuracy comparison is given by Fig. 9 for the Australian mortality data and Figs. 10 and 11 for eurozone males. Firstly, the original ACF and the updated model generate comparable predictive accuracy, as measured by RMSFE(h), although the latter performs better in most cases. The key underlying point here is that the stochastic drifts in the time effects account for the risk in the change of the mortality improvement rate. Another interesting finding emerging from the analysis is that, as time goes by, the domination of the proposed model becomes more and more noticeable, which means that our model can be expected to generate more precise point forecasts than the original ACF model in the long run.
3. The probabilistic forecast evaluation provides us with a more consistent result, showing that our proposed model with stochastic drifts outperforms the original model based on the LOGS and CRPS scores, in both the Australian two-gender case (Fig. 12) and the grouped eurozone male case (Figs. 13 and 14). This empirical result again confirms the value of considering stochastic drifts within the ACF model to enhance the forecasting results.

6. Concluding remarks

The main goal of the present research was to incorporate structural changes into multi-population mortality modelling and forecasting, based on the conventional ACF model. To achieve our goal, we re-formulated the models as the state-space form and implemented Bayesian methodology throughout the paper. The Bayesian state-space method not only combines modelling, estimation, and forecasting but also integrates various sources of uncertainties in natural manners through the computation of joint posterior distributions. This is our second contribution to Bayesian mortality modelling and forecasting. Two case studies identified that our proposed model generates a significant and non-linear mortality improvement in contrast to the original ACF model, especially in the long run, with the incorporation of the concept of structural changes in mortality forecasting. The empirical findings from the comparison of the ACF model to the proposed model provide a new understanding of the risk associated with the trend in mortality improvement, which influences the mortality dynamics dramatically in the long term. This study laid the groundwork for future research into the impact of structural changes in mortality indexes on mortality projection.

A natural progression of this study would be to investigate how to impose the stochastic drifts in other mortality models (Cairns et al., 2006; Kleinow, 2015) for multiple populations. The present paper should also stimulate considerably more work in the field of pricing, reserving, and mortality/longevity risk management.

Appendix A. Supplementary materials

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2021.12.008.

References

Alexopoulos, A., Dellaportas, P., & Forster, J. J. (2019). Bayesian forecasting of mortality rates by using latent Gaussian models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *182*(2), 689–711.

Antonio, K., Bardoutsos, A., & Ouburg, W. (2015). Bayesian Poisson log-bilinear models for mortality projections with multiple populations. *European Actuarial Journal*, *5*(2), 245–281.

Boonen, T. J., & Li, H. (2017). Modeling and forecasting mortality with economic growth: A multi-population approach. *Demography*, *54*(5), 1921–1946.

Brouhns, N., Denuit, M., & Vermunt, J. K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics & Economics*, *31*(3), 373–393.

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

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., & Khalaf-Allah, M. (2011). Bayesian stochastic mortality modelling for two populations. *Astin Bulletin*, *41*(1), 29–59.

Carter, C. K., & Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika*, *81*(3), 541–553.

Coughlan, G. D., Khalaf-Allah, M., Ye, Y., Kumar, S., Cairns, A. J. G., Blake, D., & Dowd, K. (2011). Longevity hedging 101: A framework for longevity basis risk analysis and hedge effectiveness. *North American Actuarial Journal*, *15*(2), 150–176.

Czado, C., Delwarde, A., & Denuit, M. (2005). Bayesian Poisson log-bilinear mortality projections. *Insurance: Mathematics & Economics*, *36*(3), 260–284.

Dellaportas, P., Smith, A. F. M., & Stavropoulos, P. (2001). Bayesian analysis of mortality data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *164*(2), 275–291.

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford University Press.

Dutton, L., Pantelous, A. A., & Seklecka, M. (2020). The impact of economic growth in mortality modelling for selected OECD countries. *Journal of Forecasting*, *39*(3), 533–550.

Enchev, V., Kleinow, T., & Cairns, A. J. G. (2017). Multi-population mortality models: Fitting, forecasting and comparisons. *Scandinavian Actuarial Journal*, *2017*(4), 319–342.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, *15*(2), 183–202.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378.

Hanewald, K. (2011). Explaining mortality dynamics: The role of macroeconomic fluctuations and cause of death trends. *North American Actuarial Journal*, *15*(2), 290–314.

Hanewald, K., Post, T., & Gründl, H. (2011). Stochastic mortality, macroeconomic risks and life insurer solvency. *The Geneva Papers on Risk and Insurance-Issues and Practice*, *36*(3), 458–475.

Human Mortality Database (2019). University of california, berkeley, and max Planck institute for demographic research. Data downloaded from https://www.mortality.org/ in December, 2019.

Hunt, A., & Blake, D. (2015). Modelling longevity bonds: Analysing the Swiss re Kortis bond. *Insurance: Mathematics & Economics*, *63*, 12–29.

Hunt, A., & Blake, D. (2020). A Bayesian approach to modeling and projecting cohort effects. *North American Actuarial Journal*, 1–20.

Hunt, A., & Blake, D. (2020). Identifiability in age/period mortality models. *Annals of Actuarial Science*, *14*(2), 461–499.

Husson, F., Josse, J., & Pages, J. (2010). Principal component methods-hierarchical clustering-partitional clustering: Why would we need to choose for visualizing data. *Applied Mathematics Department*, 1–17.

Huynh, N., & Ludkovski, M. (2020). Multi-output Gaussian processes for multi-population longevity modeling. arXiv preprint arXiv:2003.02443.

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.

Kang, M., Liu, Y., Li, J. S.-H., & Chan, W.-S. (2018). Mortality forecasting for multiple populations: An augmented common factor model with a penalized log-likelihood. *Communications in Statistics: Case Studies, Data Analysis and Applications*, *4*(3–4), 118–141.

Kleinow, T. (2015). A common age effect model for the mortality of multiple populations. *Insurance: Mathematics & Economics*, *63*, 147–152.

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, *87*(419), 659–671.

Leung, M., Fung, M. C., & O'Hare, C. (2018). A comparative study of pricing approaches for longevity instruments. *Insurance: Mathematics & Economics*, *82*, 95–116.

Leung, M., Li, Y., Pantelous, A. A., & Vigne, S. A. (2021). Bayesian value-at-risk backtesting: the case of annuity pricing. *European Journal of Operational Research*, *293*(2), 786–801.

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.

Li, J. S.-H., Chan, W.-S., & Cheung, S.-H. (2011). Structural changes in the Lee-Carter mortality indexes: Detection and implications. *North American Actuarial Journal*, *15*(1), 13–31.

Li, J. S.-H., & Hardy, M. R. (2011). Measuring basis risk in longevity hedges. *North American Actuarial Journal*, *15*(2), 177–200.

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.

Li, J., & Wong, K. (2020). Incorporating structural changes in mortality improvements for mortality forecasting. *Scandinavian Actuarial Journal*, 1–16.

Liu, Y., & Li, J. S.-H. (2016). It's all in the hidden states: A longevity hedging strategy with an explicit measure of population basis risk. *Insurance: Mathematics & Economics*, *70*, 301–319.

Liu, Y., & Li, J. S.-H. (2017). The locally linear Cairns–Blake–Dowd model: A note on delta–nuga hedging of longevity risk. *Astin Bulletin*, *47*(1), 79–151.

Niu, G., & Melenberg, B. (2014). Trends in mortality decrease and economic growth. *Demography*, *51*(5), 1755–1773.

O'Hare, C., & Li, Y. (2015). Identifying structural breaks in stochastic mortality models. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, *1*(2).

Pantelous, A. A., & Zimbidis, A. A. (2008). Dynamic reforming of a quasi pay-as-you-go social security system within a discrete stochastic multidimensional framework using optimal control methods. *Applicationes Mathematicae*, *35*, 121–144.

Pedroza, C. (2006). A Bayesian forecasting model: predicting US male mortality. *Biostatistics*, *7*(4), 530–550.

Petris, G., Petrone, S., & Campagnoli, P. (2009). Dynamic linear models. In *Dynamic linear models with R* (pp. 31–84). Springer.

Renshaw, A. E., & Haberman, S. (2003). Lee–Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics & 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 & Economics*, *38*(3), 556–570.

Seklecka, M., Lazam, M. N., Pantelous, A. A., & O'Hare, C. (2019). Mortality effects of economic fluctuations in selected eurozone countries. *Journal of Forecasting*, *38*(1), 39–62.

Seklecka, M., Pantelous, A. A., & O'Hare, C. (2017). Mortality effects of temperature changes in the United Kingdom. *Journal of Forecasting*, *36*(7), 824–841.

Van Berkum, F., Antonio, K., & Vellekoop, M. (2016). The impact of multiple structural changes on mortality predictions. *Scandinavian Actuarial Journal*, *2016*(7), 581–603.

Wang, L., Libert, G., & Manneback, P. (1992). Kalman filter algorithm based on singular value decomposition. In *Proceedings of the 31st ieee conference on decision and control* (pp. 1224–1229). IEEE.

Wiśniowski, A., Smith, P. W. F., Bijak, J., Raymer, J., & Forster, J. J. (2015). Bayesian population forecasting: Extending the Lee-Carter method. *Demography*, *52*(3), 1035–1059.

Wong, J. S. T., Forster, J. J., & Smith, P. W. F. (2018). Bayesian mortality forecasting with overdispersion. *Insurance: Mathematics & Economics*, *83*, 206–221.

Yang, S. S., & Wang, C.-W. (2013). Pricing and securitization of multi-country longevity risk with mortality dependence. *Insurance: Mathematics & Economics*, *52*(2), 157–169.

Zhang, Y., & Li, X. R. (1996). Fixed-interval smoothing algorithm based on singular value decomposition. In *Proceeding of the 1996 Ieee international conference on control applications ieee international conference on control applications held together with ieee international symposium on intelligent control* (pp. 916–921). IEEE.