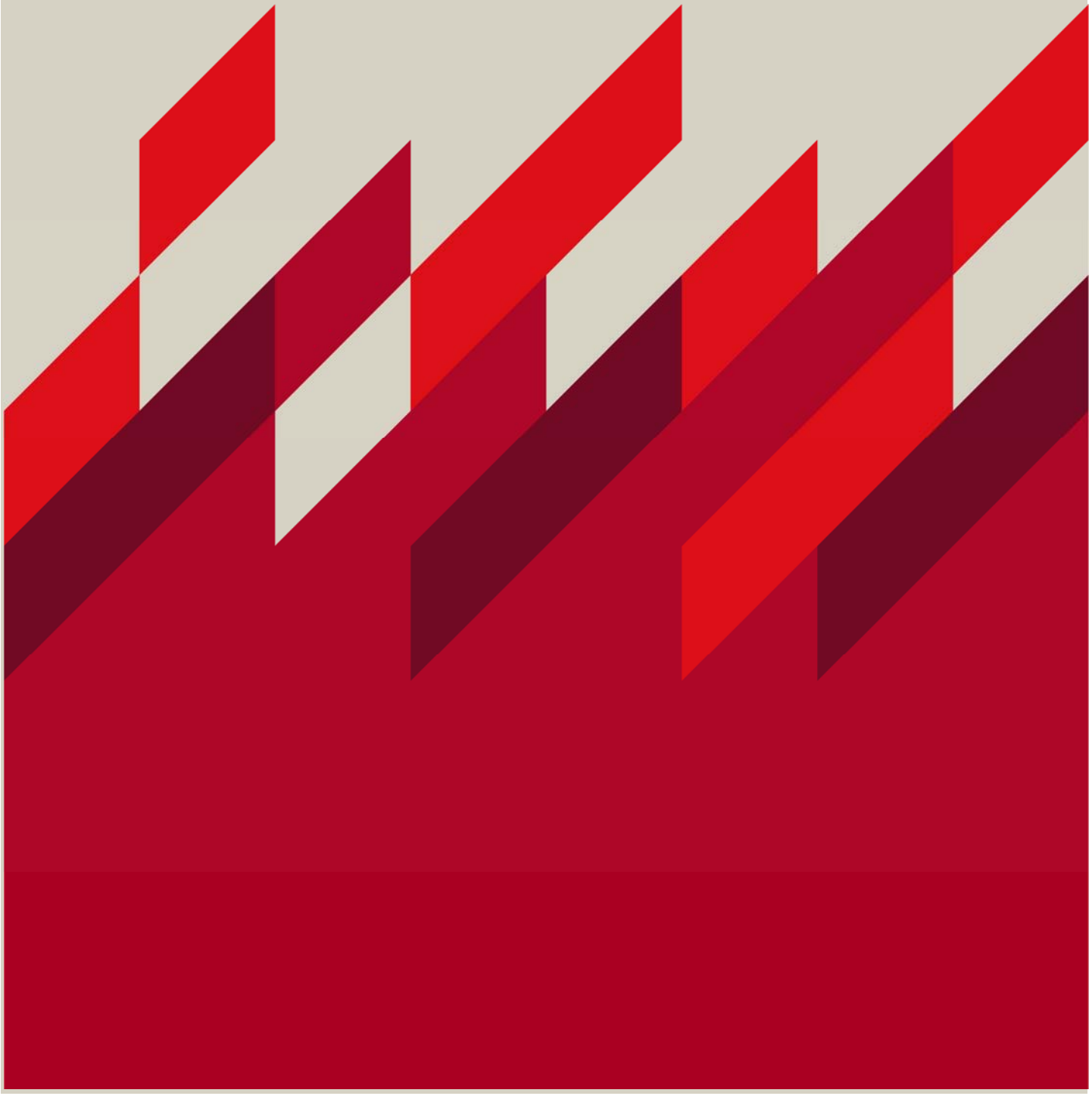
CENTRE FOR
FINANCIAL RISK
Faculty of Business and Economics

MACQUARIE
University

Cohort effects in mortality modelling: a Bayesian state-space approach

WORKING PAPER 18-03

Man Chung Fung, Gareth W. Peters, Pavel V. Shevchenko



Cohort effects in mortality modelling: a Bayesian state-space approach

Man Chung Fung[1], Gareth W. Peters[2]
Pavel V. Shevchenko[3*]

[1] Data61, CSIRO, Australia

[2] Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh, UK

[3] Department of Applied Finance and Actuarial Studies, Macquarie University, Australia

March 30, 2018

Abstract

Cohort effects are important factors in determining the evolution of human mortality for certain countries. Extensions of dynamic mortality models with cohort features have been proposed in the literature to account for these factors under the generalised linear modelling framework. In this paper we approach the problem of mortality modelling with cohort factors incorporated through a novel formulation under a state-space methodology. In the process we demonstrate that cohort factors can be formulated naturally under the state-space framework, despite the fact that cohort factors are indexed according to year-of-birth rather than year. Bayesian inference for cohort models in a state-space formulation is then developed based on an efficient Markov chain Monte Carlo sampler, allowing for the quantification of parameter uncertainty in cohort models and resulting mortality forecasts that are used for life expectancy and life table constructions. The effectiveness of our approach is examined through comprehensive empirical studies involving male and female populations from various countries. Our results show that cohort patterns are present for certain countries that we studied and the inclusion of cohort factors are crucial in capturing these phenomena, thus highlighting the benefits of introducing cohort models in the state-space framework. Forecasting of cohort models is also discussed in light of the projection of cohort factors.

Keywords: *mortality modelling, cohort features, state-space model, Bayesian inference, Markov chain Monte Carlo*

*Corresponding author

1 Introduction

The declining trend of mortality rates is generally observed in many developed countries. It is widely acknowledged among actuaries and demographers that dynamic mortality models are required to account for the uncertainty associated with the projection of mortality for different generations. From governments who are responsible for pension policy design to insurers who offer retirement income products, it is important to consider and incorporate different factors that would impact the projections of mortality rates. Arguably one of the most discussed and important factors is the so-called cohort effect, that is, the effect of year-of-birth on mortality rates.

It is perhaps not surprising that people born in different years or generations would undergo different mortality experiences. Willets (2004) finds evidence for the existence of cohort trends in the population of England and Wales. Murphy (2009) discusses the "golden generations" of the British population who were born in early 1930s and have experienced exceptionally rapid improvements in mortality rates. Willets (2004), Murphy (2009) and Murphy (2010) provide detailed discussions on cohort effects including their identification from mortality data and competing explanations. The aforementioned studies are not model-based however, but rely on empirical data analysis and qualitative analysis such as descriptive and graphical representations to conclude the significance of cohort effects on population mortality.

For actuarial applications such as longevity risk management (Cairns et al. (2008) and Barrieu et al. (2012)), one is often interested in building dynamic mortality models with stochastic cohort features. Renshaw and Haberman (2006) extend the well-known Lee-Carter mortality model (Lee and Carter (1992)) by introducing an age-modulated cohort effect. In a similar way, the Cairns-Blake-Dowd mortality model introduced in Cairns et al. (2006) is extended to incorporate cohort factors in Cairns et al. (2009). In contrast to studying cohort effects via qualitative arguments, mortality models offer a quantitative and statistical approach to identify and analyse cohort patterns exhibited in mortality data. Note that when using cohort models, the cohort patterns observed can be affected by the choice of identifiability constraints.

There are two main approaches for estimating mortality models. The first approach relies on least squares estimation based on singular value decomposition (SVD), pioneered by Lee and Carter (1992) in the mortality context, see also Renshaw and Haberman (2003), Yang et al. (2010) and Shang et al. (2011). The second approach employs regression-based methods to calibrate mortality models (Brouhns et al. (2002)). For some recent studies based on this approach, see O'Hare and Li (2012), van Berkum et al. (2016) and Enchev et al. (2016). The recent paper Currie (2016) provides a comprehensive summary on mortality modelling based on the generalised linear modelling framework. In particular, the paper points out that there are convergence and robustness problems, caused by identifiability issues, that need to be resolved for models with cohort features.

In this paper, we approach the problems of estimating and forecasting mortality with cohort features via state-space methodology, which is a natural extension of the framework recently described in Fung et al. (2017) to the cohort model formulation. The main

contribution of this paper is to demonstrate how mortality models with cohort effects can be formulated, estimated and forecasted under a Bayesian state-space framework. Other works using the state-space approach for mortality modelling include Pedroza (2006), De Jong and Tickle (2006), Kogure et al. (2009) and Liu and Li (2016b). In our view, the state-space approach has three major advantages.

First, the ability of modelling, estimating and forecasting mortality under a unified framework can avoid the potential pitfalls of the 2-step estimation procedure typically found in the literature. As discussed in details in Fung et al. (2017), a common practice of estimating mortality models consists of two steps:

Step 1 Obtain estimates of parameters including period (cohort) effects; the period (cohort) effects are treated as parameters without the assumptions of their dynamics.

Step 2 Assume a time series model, for example ARIMA models, for the period (cohort) effects; parameters of the time series model are then estimated by fitting the model to period (cohort) effects obtained from Step 1.

In a recent study in Leng and Peng (2016), the authors point out that the least-squares method for the 2-step estimation approach utilised in Lee and Carter (1992) will in general lead to inconsistent estimators unless restrictions are imposed on the possible range of time series models for the underlying dynamics. One may argue that it is more satisfactory to perform estimation and forecasting under a single universal framework, and one such example is the state-space methodology which is well-established in the statistics community.

Second, the ability to provide confidence or credible intervals for the estimated parameters, and hence the quantification of parameter risk in mortality projections, is a significant issue in valuing liabilities in life insurance portfolio and pension schemes. The impact of parameter uncertainty on forecasting mortality rates is documented in Czado et al. (2005), Koissi et al. (2006) and Kleinow and Richards (2016). A Bayesian approach to mortality modelling via credit risk plus methodology is considered in Shevchenko et al. (2015) and Hirz et al. (2017a,b). The rather short time series data used for calibration purpose typically assumed in the literature[1] further enforce the need to account for parameter uncertainty where confidence intervals are required. The state-space approach offers a particularly rich and flexible framework for Bayesian and frequentist estimation, where a range of techniques such as filtering, sequential Monte Carlo and Markov chain Monte Carlo (MCMC) methods can be employed (Fung et al. (2017)).

Third, the state-space approach allows for a wide range of mortality models to be considered while estimation and forecasting can still be performed efficiently. Fung et al. (2017) show that a majority of the popular mortality models can be cast in state-space formulation; in addition, stochastic volatility features can be introduced to dynamic mortality models where numerical filtering techniques are employed for model estimation.

[1]It is often the case that mortality rates data obtained before 1960 or even 1970 are not used when calibrating mortality models.

Applications of state-space formulation of mortality models in pricing and risk management of longevity-linked products are, for example, studied in Fung et al. (2015), Kogure and Kurachi (2010), Liu and Li (2016a) and Liu and Li (2016b).

Despite the advantages of the state-space method, the approach is still under explored in our view. A key element that is as yet missing from this literature, that we aim to address in this paper, is the consideration of cohort effects in a state-space modelling setting. Given the fact that cohort effects are known to be present in certain countries, the possibility of exploiting cohort features under a state-space framework will undoubtedly enhance an actuary's ability to analyse mortality data. The importance of incorporating cohort effects in state-space setting is also emphasized in Liu and Li (2016b), where the authors *"acknowledge that cohort effects are significant in certain populations, and that it is not trivial to incorporate cohort effects in a state-space representation in which the vector of hidden states evolve over time rather than year of birth"* (p.66). Therefore, in this paper we focus on addressing this missing piece of model formulation.

The paper is organised as follows. Section 2 provides an overview of different approaches to mortality modelling, leading to the introduction of the state-space methodology. State-space formulation of cohort models is discussed in details in Section 3. In Section 4, we develop Bayesian inference for cohort models under the state-space framework based on efficient MCMC sampling. Empirical studies for male and female population data from various countries using the state-space cohort models are conducted in Section 5. Finally, Section 6 concludes.

2 State-space approach to mortality modelling

In this section we provide an overview of different approaches to modelling mortality as well as their estimation methodologies. Our focus will be on single population mortality modelling, however the essential elements of the approaches and methods discussed can be carried over to multi-population settings, see for example Enchev et al. (2016).

2.1 Stochastic mortality models

The definitions here follow Dowd et al. (2010). We use $q_{x,t}$ to denote the true mortality rate, i.e., the probability of death between time t and $t+1$ for individuals aged x at time t. The true death rate, denoted by $m_{x,t}$, is related to the true mortality rate via

$$q_{x,t} = 1 - e^{-m_{x,t}}. \tag{1}$$

One can use the observed number of deaths D_{xt} and initial exposures E^i_{xt} data[2] to obtain the crude mortality rate as $\widetilde{q}_{x,t} = D_{xt}/E^i_{xt}$, which is a crude estimate of $q_{x,t}$. Similarly, the crude death rate $\widetilde{m}_{x,t}$ is defined as the ratio of the observed number of deaths D_{xt} to

[2]The initial exposure E^i_{xt} is the population size aged x at the beginning of year t.

the average population size E_{xt}, known as central exposures, ages x last birthday during year t.[3]

The work of Lee and Carter (1992) introduced the Lee-Carter (LC) model

$$\ln \widetilde{m}_{x,t} = \alpha_x + \beta_x \kappa_t + \varepsilon_{x,t}, \tag{2}$$

where the terms α_x and $\beta_x \kappa_t$ aim to capture the age and period effects respectively. Since they proposed to use singular value decomposition (SVD) to calibrate the model to mortality data, the noise term $\epsilon_{x,t}$ is included in addition to the age and period effects. Due to identification issue, it is typical to impose the following constraints

$$\sum_{x=x_1}^{x_p} \beta_x = 1, \quad \sum_{t=t_1}^{t_n} \kappa_t = 0, \tag{3}$$

so that the LC model becomes identifiable. Here p and n represent the number of ages and years respectively. Note that the crude death rate $\widetilde{m}_{x,t}$, which is obtained directly using mortality data, is used to obtain estimates of the parameters.

In contrast to the SVD approach where the models are fitted to crude death rates, Brouhns et al. (2002) considers an alternative approach where mortality models are fitted to the number of deaths instead. Using the number of observed deaths D_{xt} and central exposures E_{xt}, the model proposed by Lee and Carter can be reconsidered as

$$D_{xt} \sim \text{Poisson}(E_{xt}\, m_{x,t}), \quad \text{where } m_{x,t} = e^{\alpha_x + \beta_x \kappa_t}. \tag{4}$$

In this approach, the number of deaths plays an important role in calibrating the model and consequently the additive error structure in (2) is replaced by the Poisson error structure. Also notice that confusion may arise if one does not distinguish the true and crude death rate, for example it does not make sense to use $\widetilde{m}_{x,t}$ in (4).

Besides having a different statistical assumption on the error structure, the Poisson regression setting, which belongs to the class of models known as generalised linear/non-linear models, has the advantage that the cohort factor can be incorporated while estimation can still be performed without extra difficulty in contrast to the SVD approach. The Lee-Carter model in the Poisson setup can be enriched by adding a cohort factor γ_{t-x}, where $t - x$ refers to the year-of-birth, as follows:

$$D_{xt} \sim \text{Poisson}(E_{xt}\, m_{x,t}), \quad \text{where } m_{x,t} = e^{\alpha_x + \beta_x \kappa_t + \beta_x^{\gamma} \gamma_{t-x}}. \tag{5}$$

which was proposed in Renshaw and Haberman (2006). As t and x take values in $\{t_1, \ldots, t_n\}$ and $\{x_1, \ldots, x_p\}$ respectively, the cohort index $t - x$ takes values in $\{t_1 - x_p, \ldots, t_n - x_1\}$. The dimension of the cohort index is thus $n + p - 1$ which is different to the dimension of the period index n.

[3]The average population size is often determined approximately as the population size at the middle of the year.

The Renshaw-Haberman model (5) is suffering from an identification problem since the model is invariant to the following transformation

$$(\alpha_x, \beta_x, \kappa_t, \beta_x^\gamma, \gamma_{t-x}) \to \left(\alpha_x + c_1\beta_x + c_2\beta_x^\gamma, \frac{1}{c_3}\beta_x, c_3\left(\kappa_t - c_1\right), \frac{1}{c_4}\beta_x^\gamma, c_4\left(\gamma_{t-x} - c_2\right)\right), \quad (6)$$

where $c_3 \neq 0$ and $c_4 \neq 0$. One can impose the following parameter constraints

$$\sum_{x=x_1}^{x_p} \beta_x = 1, \quad \sum_{x=x_1}^{x_p} \beta_x^\gamma = 1, \quad \sum_{t=t_1}^{t_n} \kappa_t = 0, \quad \sum_{c=t_1-x_p}^{t_n-x_1} \gamma_c = 0, \quad (7)$$

to ensure a unique model structure is identified. Hunt and Villegas (2015) and Currie (2016) provide further discussions on the identifiability issues for cohort models.

Under the Lee-Carter original approach, one might consider modelling the crude death rate with cohort effects as follows:

$$\ln \widetilde{m}_{x,t} = \alpha_x + \beta_x\kappa_t + \beta_x^\gamma\gamma_{t-x} + \varepsilon_{x,t}. \quad (8)$$

However the dimension of the cohort index would cause difficulty for the SVD estimation approach. One of the goals of the present paper is to show that model (8) can be successfully estimated by considering it as a state-space model instead.

Model	Dynamics
Lee and Carter (1992)	$\ln m_{x,t} = \alpha_x + \beta_x\kappa_t$
Renshaw and Haberman (2003)	$\ln m_{x,t} = \alpha_x + \sum_{i=1}^k \beta_x^{(i)}\kappa_t^{(i)}$
Renshaw and Haberman (2006)	$\ln m_{x,t} = \alpha_x + \beta_x^{(1)}\kappa_t + \beta_x^{(2)}\gamma_{t-x}$
Currie (2009)	$\ln m_{x,t} = \alpha_x + \kappa_t + \gamma_{t-x}$
Cairns et al. (2006)	$\text{logit}(q_{x,t}) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x})$
Cairns et al. (2009)	$\text{logit}(q_{x,t}) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \gamma_{t-x}$
Plat (2009)	$\ln m_{x,t} = \alpha_x + \kappa_t^{(1)} + \kappa_t^{(2)}(\bar{x} - x) + \kappa_t^{(3)}(\bar{x} - x)^+ + \gamma_{t-x}$

Table 1: Examples of dynamic mortality models; here true death/mortality rates are being modelled.

Examples of popular mortality models are provided in Table 1. Renshaw and Haberman (2003) consider a multi-period ($\sum_{i=1}^k \beta_x^{(i)}\kappa_t^{(i)}$) extension of the LC model. Renshaw and Haberman (2006) introduce a cohort factor (γ_{t-x}) to LC model. A simplified version of the model in Renshaw and Haberman (2006) is studied in Currie (2009). Cairns et al. (2006) propose the so-called CBD model to model $\text{logit}(q_{x,t}) := \ln\left(q_{x,t}/(1 - q_{x,t})\right)$, where $\bar{x}$ is the average age of the sample range, i.e. $\bar{x} = \frac{1}{p}\sum_{i=1}^p x_i$, designed to capture the mature-age mortality dynamics. Cairns et al. (2009) extend the CBD model by incorporating a cohort factor. Identification constraints for the models in Table 1 are discussed in Villegas et al. (2015).

Remark 2.1 *An alternative approach to modelling cohort mortality is to consider the continuous-time modelling of the mortality intensity (force of mortality) $\mu_{x+t,t}$, with x and t taking continuous values. It is of particular relevance for financial and actuarial applications as pricing formulas can be derived and expressed via the mortality intensity. Continuous-time approaches are discussed in Cairns et al. (2008), and they have been studied extensively in the literature, see for example Biffis (2005), Dahl and Moller (2006), Luciano and Vigna (2008) and Fung et al. (2014).*

2.2 Generalised linear modelling framework

Estimation of stochastic mortality models such as those presented in Table 1 can be performed based on a flexible approach known as generalised linear modelling (GLM) framework (Villegas et al. (2015), Currie (2016)).

Given central exposures E_{xt}, it is typical to approximate initial exposures as $E^i_{xt} \approx E_{xt} + \frac{1}{2}D_{xt}$. Under the GLM framework, one is interested in modelling the number of deaths D_{xt} as random variables. Common examples include Poisson error structure

$$D_{xt} \sim \text{Poisson}(E_{xt}\, m_{x,t}) \tag{9}$$

and Binomial error structure

$$D_{xt} \sim \text{Binomial}(E^i_{xt},\, q_{x,t}). \tag{10}$$

Note that we have the expected values $\mathbb{E}\left[D_{xt}/E_{xt}\right] = m_{x,t}$ and $\mathbb{E}[D_{xt}/E^i_{xt}] = q_{x,t}$ for the Poisson and Binomial models respectively. Through the so-called link function g, one can associate the mean $\mathbb{E}\left[D_{xt}/E_{xt}\right]$ or $\mathbb{E}[D_{xt}/E^i_{xt}]$ with a predictor $\rho_{x,t}$ as

$$g\left(\mathbb{E}\left[\frac{D_{xt}}{\bar{E}_{xt}}\right]\right) = \rho_{x,t} \tag{11}$$

where $\bar{E}_{xt}$ can be initial or central exposures. Typical link functions for the Poisson and Binomial models are the log function and the logit function respectively.

The models in Table 1, from the viewpoint of GLM framework, provide a specification for the predictor. In particular, the last four models in Table 1 assume linear predictors while the first three models describe non-linear predictors since multiplicative terms of parameters such as $\beta_x\kappa_t$ are involved in the predictor. The latter models can be referred to as generalised non-linear models (Currie (2016)).

We note here that a clear advantage of the GLM framework is that it allows sophisticated error structures such as Poisson and Binomial distributions compared to the SVD approach. However, despite its flexibility, the framework involves a 2-step procedure for estimation which is in contrast to the state-space approach presented in the following sections where a joint estimation of model parameters and latent factors is performed.

2.3 State-space modelling framework

In this paper we extend previous work, see for example Pedroza (2006), Kogure and Kurachi (2010) and Fung et al. (2017), on using state-space techniques to model mortality dynamics with cohort features taken into consideration. We first present a brief review of mortality modelling via state-space representation.

A state-space model consists of two equations: the observation equation and the state equation which are given, respectively, by

$$\boldsymbol{z}_t = a(\boldsymbol{\phi}_t, \boldsymbol{u}_t), \tag{12a}$$
$$\boldsymbol{\phi}_t = b(\boldsymbol{\phi}_{t-1}, \boldsymbol{v}_t), \tag{12b}$$

where $\boldsymbol{z}_t$ represents an observed multi-dimensional time series and the state $\boldsymbol{\phi}_t$ represents a multi-dimensional hidden Markov process. Here, $\boldsymbol{u}_t$ and $\boldsymbol{v}_t$ are independent random noises and the functions a and b can be nonlinear in general. It is clear that each of the mortality models shown in Table 1, with $m_{x,t}$ replaced by $\widetilde{m}_{x,t}$ and an inclusion of a noise term, would specify the observation equation of a state-space model, where the period and cohort factors represent the hidden states. A time series model for the period effects will form the state equation, thus completing the description of a state-space system. As an example, the Lee-Carter model can be reformulated as

$$\boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}\kappa_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} \mathrm{N}(0, I_p\sigma_\varepsilon^2), \tag{13a}$$
$$\kappa_t = \kappa_{t-1} + \theta + \omega_t, \quad \omega_t \overset{iid}{\sim} \mathrm{N}(0, \sigma_\omega^2), \tag{13b}$$

where $y_{x,t} = \ln \widetilde{m}_{x,t}$, I_p is a p-dimensional identity matrix and $\mathrm{N}(a, b)$ denotes the normal distribution with mean a and covariance b.

Fung et al. (2017) study two generalisations of the state-space Lee-Carter system (13) to analyse long term mortality time series for the Danish population. The first generalisation is to incorporate heteroscedasticity into the model, i.e. the homogeneous covariance structure $I_p\sigma_\varepsilon^2$ in the observation equation in (13) which is replaced by a heterogeneous covariance matrix $I_p\sigma_{\varepsilon,x}^2$. Note that the product $I_p\sigma_{\epsilon,x}^2$ is a p by p diagonal matrix where the diagonal elements are $\sigma_{\epsilon,x_1}^2, \ldots, \sigma_{\epsilon,x_p}^2$. This feature turns out to be a major improvement to model fit for Danish mortality data. For simplicity, we assume constant variance in the observation equation for the models considered in the paper. Estimation for models with heteroscedasticity structure can also be performed as discussed in Fung et al. (2017).

The second generalisation is to consider stochastic volatility for the latent process, that is the period effect κ_t in the state equation in (13), to capture the observed characteristics of the long term time series data. Specifically, the following extended LC model with stochastic volatility feature is proposed:

$$\boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}\kappa_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} \mathrm{N}(\boldsymbol{0}, \sigma_\varepsilon^2 I_p), \tag{14a}$$
$$\kappa_t = \kappa_{t-1} + \theta + \omega_t, \quad \omega_t|\xi_t \sim \mathrm{N}(0, \exp\{\xi_t\}), \tag{14b}$$
$$\xi_t = \lambda_1\xi_{t-1} + \lambda_2 + \zeta_t, \quad \zeta_t \overset{iid}{\sim} \mathrm{N}(0, \sigma_\xi^2) \tag{14c}$$

where ξ_t captures the stochastic volatility for the period effect. Bayesian inference for the model is developed based on particle MCMC method (Andrieu et al. (2010)).

Despite the fact that cohort models are essential in mortality modelling as shown in Table 1, a formulation of mortality models with cohort factors incorporated is yet to be studied and analysed in the state-space setting in the literature. In the following we will present and demonstrate our approach for dealing with cohort models via state-space methodology.

3 Cohort effects: state-space formulation

This section presents a formulation of cohort models in state-space framework. We first describe how the cohort factor impacts the evolution of the age-specific death rates. The insight will give us a way to derive a state-space representation of the cohort effects. Bayesian inference for cohort models will be developed in Section 4.

3.1 Background

Renshaw and Haberman (2006) introduces a cohort factor γ_{t-x} to the Lee-Carter model together with an age-modulating coefficient β_x^γ as follows:

$$\ln m_{x,t} = \alpha_x + \beta_x \kappa_t + \beta_x^\gamma \gamma_{t-x}. \tag{15}$$

Here $t-x$ represents year-of-birth and hence γ_{t-x} is a factor created to capture the cohort effect.

Numerical estimation of the cohort model (15), however, is reported in Hunt and Villegas (2015) to produce mixed convergence results based on the regression setup. Robustness of the resulting regression-based cohort models is also questioned as the goodness of model fit is reported to be sensitive to the data being used and the fitting algorithm. These robustness and convergence problems are also noticed in Currie (2016) where the paper presents a comprehensive approach for mortality modelling based on generalised linear and non-linear models, see also Section 2.2 for a brief discussion.

Moreover, it is shown in Leng and Peng (2016) that the 2-step estimation approach utilised in Lee and Carter (1992) will in general lead to inconsistent estimators unless restrictions are imposed on the possible range of time series models for the underlying dynamics. The Bayesian state-space approach, in contrast, is not subject to such restrictions as one can perform estimation and forecasting under a single universal framework. We report our empirical findings for the estimation and forecasting of state-space cohort models in Section 5.

3.2 State-space formulation

In this paper we focus on the *full cohort model*

$$\ln \widetilde{m}_{x,t} = \alpha_x + \beta_x \kappa_t + \beta_x^\gamma \gamma_{t-x} + \varepsilon_{x,t}, \tag{16}$$

where a dynamics for the crude death rate is being modelled and a noise term $\varepsilon_{x,t}$ is included.

To aid in explaining how we derive a state-space representation of cohort models, we consider a matrix of cells where the row and column corresponds to age (x) and year (t) respectively, see Table 2. Here we assume $x = 1, \ldots, 3$ and $t = 1, \ldots, 4$ for illustration. The cohort factor γ_{t-x} is indexed by the year-of-birth $t - x$ and its value on each cell is displayed in Table 2. We first notice that the value γ_{t-x} is constant on the "cohort direction", that is on the cells (x, t), $(x + 1, t + 1)$ and so on.

age/year	$t = 1$	$t = 2$	$t = 3$	$t = 4$
$x = 1$	γ_0	γ_1	γ_2	γ_3
$x = 2$	γ_{-1}	γ_0	γ_1	γ_2
$x = 3$	γ_{-2}	γ_{-1}	γ_0	γ_1

Table 2: Values of the cohort factor γ_{t-x} on a matrix of cells (x, t).

Now consider the cohort model (16) and let $\gamma_t^x := \gamma_{t-x}$. The model can be expressed in matrix form as

$$\begin{pmatrix} \ln \widetilde{m}_{1,t} \\ \ln \widetilde{m}_{2,t} \\ \ln \widetilde{m}_{3,t} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} + \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \kappa_t + \begin{pmatrix} \beta_1^\gamma & 0 & 0 \\ 0 & \beta_2^\gamma & 0 \\ 0 & 0 & \beta_3^\gamma \end{pmatrix} \begin{pmatrix} \gamma_t^1 \\ \gamma_t^2 \\ \gamma_t^3 \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{pmatrix}. \tag{17}$$

As time flows from $t = 1$ to $t = 4$, the cohort vector $(\gamma_t^1, \gamma_t^2, \gamma_t^3)^\top$, which represents the cohort factor in matrix form, proceeds as

$$\begin{pmatrix} \gamma_1^1 (= \gamma_0) \\ \gamma_1^2 (= \gamma_{-1}) \\ \gamma_1^3 (= \gamma_{-2}) \end{pmatrix} \to \begin{pmatrix} \gamma_2^1 (= \gamma_1) \\ \gamma_2^2 (= \gamma_0) \\ \gamma_2^3 (= \gamma_{-1}) \end{pmatrix} \to \begin{pmatrix} \gamma_3^1 (= \gamma_2) \\ \gamma_3^2 (= \gamma_1) \\ \gamma_3^3 (= \gamma_0) \end{pmatrix} \to \begin{pmatrix} \gamma_4^1 (= \gamma_3) \\ \gamma_4^2 (= \gamma_2) \\ \gamma_4^3 (= \gamma_1) \end{pmatrix}. \tag{18}$$

The key observation here from (18) is that the first two elements of the cohort vector at time $t - 1$ will appear as the bottom two elements of the cohort vector at time t. The pattern can also be observed from Table 2. Therefore, the evolution of the cohort vector must satisfy

$$\begin{pmatrix} \gamma_t^1 \\ \gamma_t^2 \\ \gamma_t^3 \end{pmatrix} = \begin{pmatrix} * & * & * \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \gamma_{t-1}^1 \\ \gamma_{t-1}^2 \\ \gamma_{t-1}^3 \end{pmatrix} + \ldots, \tag{19}$$

which is in fact a result of the defining property of "cohort": $\gamma_{t-x} = \gamma_{(t-i)-(x-i)}$. Furthermore, it is obvious from (19) that one only needs to model the dynamics of γ_t^1 but not γ_t^2 and γ_t^3. We will use this observation to derive a state-space formulation of cohort models which is presented next.

Let $y_x = \ln \widetilde{m}_{x,t}$, in matrix notation we have (recall that $\gamma_t^x := \gamma_{t-x}$)

$$\begin{pmatrix} y_{x_1,t} \\ y_{x_2,t} \\ \vdots \\ y_{x_p,t} \end{pmatrix} = \begin{pmatrix} \alpha_{x_1} \\ \alpha_{x_2} \\ \vdots \\ \alpha_{x_p} \end{pmatrix} + \begin{pmatrix} \beta_{x_1} & \beta_{x_1}^{\gamma} & 0 & \cdots & 0 \\ \beta_{x_2} & 0 & \beta_{x_2}^{\gamma} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_{x_p} & 0 & 0 & \cdots & \beta_{x_p}^{\gamma} \end{pmatrix} \begin{pmatrix} \kappa_t \\ \gamma_t^{x_1} \\ \gamma_t^{x_2} \\ \vdots \\ \gamma_t^{x_p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{x_1,t} \\ \varepsilon_{x_2,t} \\ \vdots \\ \varepsilon_{x_p,t} \end{pmatrix}. \tag{20}$$

It is clear that, from (20), we have $y_{x_i,t} = \alpha_{x_i} + \beta_{x_i}\kappa_t + \beta_{x_i}^{\gamma}\gamma_t^{x_i} + \varepsilon_{x_i,t}$ which corresponds to (16) for $i \in \{1, \ldots, p\}$. Here $(\kappa_t, \gamma_t^{x_1}, \ldots, \gamma_t^{x_p})^\top$ is the $p+1$ dimensional state vector.

From (18)-(19), we can write the state equation in matrix notation as follows:

$$\begin{pmatrix} \kappa_t \\ \gamma_t^{x_1} \\ \gamma_t^{x_2} \\ \vdots \\ \gamma_t^{x_{p-1}} \\ \gamma_t^{x_p} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \kappa_{t-1} \\ \gamma_{t-1}^{x_1} \\ \gamma_{t-1}^{x_2} \\ \vdots \\ \gamma_{t-1}^{x_{p-1}} \\ \gamma_{t-1}^{x_p} \end{pmatrix} + \begin{pmatrix} \theta \\ \zeta \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \omega_t^{\kappa} \\ \omega_t^{\gamma} \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}. \tag{21}$$

Here we assume κ_t is a random walk with drift process (ARIMA$(0,1,0)$)

$$\kappa_t = \kappa_{t-1} + \theta + \omega_t^{\kappa}, \quad \omega_t^{\kappa} \overset{iid}{\sim} \mathrm{N}(0, \sigma_{\omega}^2), \tag{22}$$

and the dynamics of $\gamma_t^{x_1}$ is described by a stationary AR(1) process (ARIMA(1,0,0))

$$\gamma_t^{x_1} = \lambda\gamma_{t-1}^{x_1} + \zeta + \omega_t^{\gamma}, \quad \omega_t^{\gamma} \overset{iid}{\sim} \mathrm{N}(0, \sigma_{\gamma}^2), \tag{23}$$

where $|\lambda| < 1$. It should be note that one may consider non-stationary processes such as the random walk with drift process by fixing $\lambda = 1$. For illustration purpose we consider stationary processes for the cohort factor in this paper. We will provide further discussion on the modelling of cohort factor in Section 5.3.2. One may consider other dynamics for $\gamma_t^{x_1}$ by specifying the second row of the $p+1$ by $p+1$ matrix in (21). For example, one can consider generally the state equation as

$$\begin{pmatrix} \kappa_t \\ \gamma_t^{x_1} \\ \gamma_t^{x_2} \\ \vdots \\ \gamma_t^{x_{p-1}} \\ \gamma_t^{x_p} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_1 & \lambda_2 & \cdots & \lambda_{p-1} & \lambda_p \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \kappa_{t-1} \\ \gamma_{t-1}^{x_1} \\ \gamma_{t-1}^{x_2} \\ \vdots \\ \gamma_{t-1}^{x_{p-1}} \\ \gamma_{t-1}^{x_p} \end{pmatrix} + \begin{pmatrix} \theta \\ \zeta \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \omega_t^{\kappa} \\ \omega_t^{\gamma} \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \tag{24}$$

where $\gamma_t^{x_1} = \lambda_1\gamma_{t-1}^{x_1} + \lambda_2\gamma_{t-1}^{x_2} + \cdots + \lambda_{p-1}\gamma_{t-1}^{x_{p-1}} + \lambda_p\gamma_{t-1}^{x_p} + \zeta + \omega_t^{\gamma}$ which is an ARIMA(p,0,0) process since $\gamma_{t-1}^{x_i} = \gamma_{t-i}^{x_1}$, $i = 2, \ldots, p$. Note that in general the state-space modelling

approach can handle seasonal ARIMA models as well as non-linear models, see for instance Harvey (1989), Gourieroux and Monfort (1997) and Petris et al. (2009) for their representations in state-space form.

We can express the matrix form of (20)-(21) succinctly as

$$\boldsymbol{y}_t = \boldsymbol{\alpha} + B\boldsymbol{\varphi}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} \mathrm{N}(0, \sigma_\varepsilon^2 I_p), \tag{25a}$$

$$\boldsymbol{\varphi}_t = \Lambda\boldsymbol{\varphi}_{t-1} + \boldsymbol{\Theta} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \overset{iid}{\sim} \mathrm{N}(0, \Upsilon), \tag{25b}$$

where

$$B = \begin{pmatrix} \beta_{x_1} & \beta^{\gamma}_{x_1} & 0 & \cdots & 0 \\ \beta_{x_2} & 0 & \beta^{\gamma}_{x_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_{x_p} & 0 & 0 & \cdots & \beta^{\gamma}_{x_p} \end{pmatrix}, \quad \Lambda = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \quad \Theta = \begin{pmatrix} \theta \\ \zeta \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \tag{26}$$

and $\boldsymbol{\varphi}_t = (\kappa_t, \gamma_t^{x_1}, \ldots, \gamma_t^{x_p})^\top$, I_p the p-dimensional identity matrix and Υ is a $p+1$ by $p+1$ diagonal matrix with diagonal $(\sigma_\kappa^2, \sigma_\gamma^2, 0, \ldots, 0)$. For simplicity we assume homoscedasticity in the observation equation; heteroscedasticity can be considered as developed in Fung et al. (2017). Note that the set of constraints (30) are required for the identification of the cohort model (25).

Remark 3.1 *The observation and state equations* (20)-(21) *imply that the cohort model that we have formulated here belongs to the linear-Gaussian class of state-space models. As a result one can perform efficient maximum-likelihood or Bayesian estimation on fitting the model to data, see Fung et al. (2017). In this paper we focus on Bayesian inference so that mortality forecasts can take into account parameter uncertainty. It should also be note that in Fung et al. (2017) the formulation of cohort models in state-space framework is only mentioned briefly while their estimation and forecasting were not provided due to the focus and length of the paper. The current paper aims to fill this gap by providing details on how one may derive the formulation of cohort models under the state-space framework and to examine the resulting models comprehensively using a range of mortality data.*

3.3 A simplified cohort model

The cohort model (15) assumes that the impact of the cohort factor on age-specific death rates is modulated by the coefficient β_x^γ. Iteration-based estimation of the cohort model is reported in Cairns et al. (2009) and Hunt and Villegas (2015) to be suffering from convergence problems. As a result, Haberman and Renshaw (2011) consider to simplify the model structure to

$$\ln m_{x,t} = \alpha_x + \beta_x\kappa_t + \gamma_{t-x}. \tag{27}$$

That is, the modulating coefficient β_x^γ for the cohort factor is set to be one for all age x. It is suggested that the simplified model (27) exhibits better estimation convergence behaviour when fitting the model to mortality data.

Remark 3.2 *When fitting cohort models to data based on the Poisson or binomial regression setup via iteration-based estimation, it is typical to assume the starting values of the iteration scheme are coming from the estimates of other similar models, for example the LC model or APC model (Hunt and Villegas (2015), Currie (2016), Villegas et al. (2015)). Even doing so the convergence is not guaranteed. We will report in Section 5 that our approach based on Bayesian state-space framework do not require such an assumption and is able to successfully perform Bayesian estimation for various countries when cohort patterns are present in the data.*

In the following, we will also consider the model

$$\boldsymbol{y}_t = \boldsymbol{\alpha} + B^s \boldsymbol{\varphi}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} \mathrm{N}(0, \sigma_\varepsilon^2 I_p), \tag{28a}$$

$$\boldsymbol{\varphi}_t = \Lambda \boldsymbol{\varphi}_{t-1} + \boldsymbol{\Theta} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \overset{iid}{\sim} \mathrm{N}(0, \Upsilon), \tag{28b}$$

where

$$B^s = \begin{pmatrix} \beta_{x_1} & 1 & 0 & \cdots & 0 \\ \beta_{x_2} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_{x_p} & 0 & 0 & \cdots & 1 \end{pmatrix}. \tag{29}$$

It will be referred to as the *simplified cohort model*. In addition, the following set of parameter constraints

$$\sum_{x=x_1}^{x_p} \beta_x = 1, \quad \sum_{t=t_1}^{t_n} \kappa_t = 0, \quad \sum_{c=t_1-x_p}^{t_n-x_1} \gamma_c = 0, \tag{30}$$

is imposed to ensure the identifiability of the simplified cohort model.

Since our approach is able to estimate both the simplified and full cohort models, (27) and (15) respectively, for a range of mortality dataset, we include the simplified cohort model in our empirical analysis examining the performance of the simplified and full cohort models as a comparative study.

4 Bayesian inference for cohort models

In this section we begin by detailing the Bayesian estimation of the full cohort model in state-space formulation (25). Nested models including the simplified cohort model (28) and the LC model (13) will be discussed in Section 4.2. We first note that the cohort models and the LC model belong to the class of linear and Gaussian state-space models. As a result one can apply an efficient MCMC estimation algorithm based on Gibbs sampling with conjugate priors combined with forward-backward filtering as described in Fung et al. (2017), this forms a special case of the so-called collapsed Gibbs sampler framework of van Dyk and Park (2008).

4.1 Bayesian inference for the full cohort model

For the full cohort model (25), the target density is given by

$$\pi\left(\boldsymbol{\varphi}_{0:n}, \boldsymbol{\psi} | \boldsymbol{y}_{1:n}\right) \propto \pi(\boldsymbol{y}_{1:n} | \boldsymbol{\varphi}_{0:n}, \boldsymbol{\psi}) \pi(\boldsymbol{\varphi}_{0:n} | \boldsymbol{\psi}) \pi(\boldsymbol{\psi}) \tag{31}$$

$$= \prod_{k=1}^{n} \pi(\boldsymbol{y}_k | \boldsymbol{\varphi}_{0:k}, \boldsymbol{\psi}) \pi(\boldsymbol{\varphi}_k | \boldsymbol{\varphi}_{k-1}, \boldsymbol{\psi}) \pi(\boldsymbol{\varphi}_0) \pi(\boldsymbol{\psi}) \tag{32}$$

where $\boldsymbol{\varphi}_{0:n} := (\kappa_{0:n}, \gamma_{0:n}^{x_1}, \ldots, \gamma_{0:n}^{x_p})$ is the $p+1$ dimensional latent state vector for $t \in \{0, \ldots, n\}$[4] and $\boldsymbol{\psi} := (\boldsymbol{\beta}, \boldsymbol{\beta}^{\gamma}, \boldsymbol{\alpha}, \theta, \zeta, \lambda, \sigma_{\varepsilon}^2, \sigma_{\kappa}^2, \sigma_{\gamma}^2)$ is the $3p+6$ dimensional static parameter vector. In order to simplify the notation, we write $t = 1, \ldots, n$ instead of $t = t_1, \ldots, t_n$. We perform block sampling for the latent state via the so-called forward-filtering-backward-sampling (FFBS) algorithm (Carter and Kohn (1994)) and the posterior samples of the static parameters are obtained via conjugate priors. The sampling procedure is described in Algorithm 1, where M is the number of MCMC iterations performed. Note also that the notation $\boldsymbol{\psi}_{-\nu}^{(i)} := (\psi_{g_1}^{(i)}, \ldots, \psi_{g_{\nu-1}}^{(i)}, \psi_{g_{\nu+1}}^{(i-1)}, \ldots, \psi_{g_{3p+6}}^{(i-1)})$ is used in Algorithm 1 where the subscript g_1 refers to the first parameter and so on.[5] Moreover, the initial values $\boldsymbol{\psi}^{(0)}$ of the Markov chain can be set to any reasonable values and after discarding the initial iterations as burn-in period, the assumption of the initial values used will not impact the estimation results if the chain mixes well.

Remark 4.1 *Note that some elements of the state vector evolve deterministically as specified in* (21)*. Consequently the joint model decomposes as follows into product measures, which is the joint measures of the random observation and the latent state process, given explicitly by*

$$\begin{aligned} \pi\left(dy_{1:n}, d\kappa_{0:n}, d\gamma_{0:n}; \boldsymbol{\psi}\right) &= \prod_{t=1}^{n} \pi\left(dy_t | \kappa_t, \gamma_t; \boldsymbol{\psi}\right) \pi\left(d\kappa_t, d\gamma_t | \kappa_{t-1}, \gamma_{t-1}; \boldsymbol{\psi}\right) \\ &= \prod_{t=1}^{n} \pi\left(dy_t | \kappa_t, \gamma_t; \boldsymbol{\psi}\right) \pi\left(d\kappa_t, d\gamma_t^{x_1} | \kappa_{t-1}, \gamma_{t-1}^{x_1}; \boldsymbol{\psi}\right) \delta_{\gamma_{t-1}^{x_1}}\left(d\gamma_t^{x_2}\right) \ldots \delta_{\gamma_{t-1}^{x_{p-1}}}\left(d\gamma_t^{x_p}\right), \end{aligned} \tag{33}$$

where δ_w is the Dirac mass probability measure centered on point w. Then the MCMC sampling steps involved in Algorithm 1 can be carried out and justified using the Dirac mass probability measure for elements of the state vector that evolve deterministically. For examples of similar models in which Dirac measure is present in the likelihood specification and filtering equations, see e.g. Andrieu et al. (2010), Del Moral et al. (2013), Doucet et al. (2015) and Peters et al. (2017).

[4] That is, $\boldsymbol{\varphi}_{0:n} := (\kappa_0, \ldots, \kappa_n, \gamma_0^{x_1}, \ldots, \gamma_n^{x_1}, \gamma_0^{x_p}, \ldots, \gamma_n^{x_p})$. Throughout the paper, we use the notation that $X_{1:K} := (X_1, \ldots, X_K)$.

[5] It is typical to observe irregularity for the estimated pattern of $\boldsymbol{\beta}$. Smoothing procedures are proposed in Delwarde et al. (2007) for the SVD and Poisson regression approaches. For simplicity we do not consider this issue in the current paper.

In Algorithm 1, after obtaining $\tilde{\kappa}_{1:n}^{(i)}$ from line 3, one can impose the constraint $\sum_t \kappa_t^{(i)} = 0$ by setting $\kappa_t^{(i)} = \tilde{\kappa}_t^{(i)} - (1/n)\sum_{j=1}^{n} \tilde{\kappa}_j^{(i)}$ where $t = 1, \ldots, n$. Similarly, once $\tilde{\gamma}_c^{(i)}$ is obtained from $\boldsymbol{\varphi}_{0:n}^{(i)}$ in line 3, setting $\gamma_c^{(i)} = \tilde{\gamma}_c^{(i)} - (1/(n+p-1))\sum_{\ell=1-x_p}^{n-x_1} \tilde{\gamma}_\ell^{(i)}$, where $c = 1 - x_p, \ldots, n - x_1$, will ensure the constraint $\sum_c \gamma_c^{(i)} = 0$ is satisfied. To impose the constraint $\sum_x \beta_x^{(i)} = 1$, we set $\beta_x^{(i)} = \tilde{\beta}_x^{(i)} / \sum_{j=x_1}^{x_p} \tilde{\beta}_j^{(i)}$, where $x = x_1, \ldots, x_p$, once $\tilde{\boldsymbol{\beta}}^{(i)}$ is obtained from line 5-7. Constraint for $(\boldsymbol{\beta}^\gamma)^{(i)}$ can be imposed similarly.

Algorithm 1 MCMC sampling for $\pi(\boldsymbol{\varphi}_{0:n}, \boldsymbol{\psi} | \boldsymbol{y}_{1:n})$

1: Initialise: $\boldsymbol{\psi} = \boldsymbol{\psi}^{(0)}$.
2: **for** $i = 1, \ldots, M$ **do**
3: Sample $\boldsymbol{\varphi}_{0:n}^{(i)}$ from $\pi(\boldsymbol{\varphi}_{0:n} | \boldsymbol{\psi}^{(i-1)}, \boldsymbol{y}_{1:n})$ via FFBS (Section 4.1.1).
4: Impose the constraint $\sum_t \kappa_t^{(i)} = 0$ and $\sum_c \gamma_c^{(i)} = 0$.
5: **for** $h = 1, \ldots, p$ **do**
6: Sample $\beta_{x_h}^{(i)}$ from its posterior $\pi(\beta_{x_h} | \boldsymbol{\varphi}_{0:n}^{(i)}, \boldsymbol{\psi}_{-\beta_{x_h}}^{(i)}, \boldsymbol{y}_{1:n})$
7: **end for**
8: Impose the constraint $\sum_x \beta_x^{(i)} = 1$.
9: **for** $h = 1, \ldots, p$ **do**
10: Sample $(\beta_{x_h}^\gamma)^{(i)}$ from its posterior $\pi(\beta_{x_h}^\gamma | \boldsymbol{\varphi}_{0:n}^{(i)}, \boldsymbol{\psi}_{-\beta_{x_h}^\gamma}^{(i)}, \boldsymbol{y}_{1:n})$
11: **end for**
12: Impose the constraint $\sum_x (\beta_x^\gamma)^{(i)} = 1$.
13: **for** $h = 2p+1, \ldots, 3p+6$ **do**
14: Sample $\psi_h^{(i)}$ from $\pi(\psi_h | \boldsymbol{\varphi}_{0:n}^{(i)}, \boldsymbol{\psi}_{-h}^{(i)}, \boldsymbol{y}_{1:n})$
15: **end for**
16: **end for**

4.1.1 Forward-backward filtering for latent state dynamics

The FFBS procedure requires to carry out multivariate Kalman filtering forward in time and then sample backwardly using the obtained filtering distributions. For the full cohort model (20)-(21), the conditional distributions involved in the multivariate Kalman filtering recursions are given by

$$\boldsymbol{\varphi}_{t-1} | \boldsymbol{y}_{1:t-1} \sim \mathrm{N}(\boldsymbol{m}_{t-1}, C_{t-1}), \tag{34a}$$
$$\boldsymbol{\varphi}_t | \boldsymbol{y}_{1:t-1} \sim \mathrm{N}(\boldsymbol{a}_t, R_t), \tag{34b}$$
$$\boldsymbol{y}_t | \boldsymbol{y}_{1:t-1} \sim \mathrm{N}(\boldsymbol{f}_t, Q_t), \tag{34c}$$
$$\boldsymbol{\varphi}_t | \boldsymbol{y}_{1:t} \sim \mathrm{N}(\boldsymbol{m}_t, C_t), \tag{34d}$$

where

$$\boldsymbol{a}_t = \Lambda \boldsymbol{m}_{t-1} + \boldsymbol{\Theta}, \quad R_t = \Lambda C_{t-1} \Lambda^\top + \Upsilon, \tag{35a}$$
$$\boldsymbol{f}_t = \boldsymbol{\alpha} + B \boldsymbol{a}_t, \quad Q_t = B R_t B^\top + \sigma_\varepsilon^2 I_p, \tag{35b}$$

$$\boldsymbol{m}_t = \boldsymbol{a}_t + R_t B^\top Q_t^{-1}(\boldsymbol{y}_t - \boldsymbol{f}_t), \quad C_t = R_t - R_t B^\top Q_t^{-1} B R_t. \tag{35c}$$

for $t = 1, \ldots, n$. Since

$$\pi(\boldsymbol{\varphi}_{0:n}|\boldsymbol{\psi}, \boldsymbol{y}_{1:n}) = \prod_{t=0}^{n} \pi(\boldsymbol{\varphi}_t|\boldsymbol{\varphi}_{t+1:n}, \boldsymbol{\psi}, \boldsymbol{y}_{1:n}) = \prod_{t=0}^{n} \pi(\boldsymbol{\varphi}_t|\boldsymbol{\varphi}_{t+1}, \boldsymbol{\psi}, \boldsymbol{y}_{1:t}), \tag{36}$$

we see that for a block sampling of the latent state, one can first draw $\boldsymbol{\varphi}_n$ from $\mathrm{N}(\boldsymbol{m}_n, C_n)$ and then, for $t = n-1, \ldots, 1, 0$ (that is backward in time), draws a sample of $\boldsymbol{\varphi}_t|_{\boldsymbol{\varphi}_{t+1}, \boldsymbol{\psi}, \boldsymbol{y}_{1:t}}$ recursively given a sample of $\boldsymbol{\varphi}_{t+1}$. It turns out that $\boldsymbol{\varphi}_t|_{\boldsymbol{\varphi}_{t+1}, \boldsymbol{\psi}, \boldsymbol{y}_{1:t}} \sim \mathrm{N}(\boldsymbol{h}_t, H_t)$ where

$$\boldsymbol{h}_t = \boldsymbol{m}_t + C_t \Lambda^\top R_{t+1}^{-1}(\boldsymbol{\varphi}_{t+1} - \boldsymbol{a}_{t+1}), \tag{37a}$$

$$H_t = C_t - C_t \Lambda^\top R_{t+1}^{-1} \Lambda C_t, \tag{37b}$$

based on Kalman smoothing (Carter and Kohn (1994)).

4.1.2 Posteriors for static parameters

To sample the posterior distribution of the static parameters in Algorithm 1, we assume the following independent conjugate priors:

$$\alpha_x \sim \mathrm{N}(\tilde{\mu}_\alpha, \tilde{\sigma}_\alpha^2), \quad \beta_x \sim \mathrm{N}(\tilde{\mu}_\beta, \tilde{\sigma}_\beta^2), \quad \beta_x^\gamma \sim \mathrm{N}(\tilde{\mu}_{\beta^\gamma}, \tilde{\sigma}_{\beta^\gamma}^2), \tag{38a}$$

$$\theta \sim \mathrm{N}(\tilde{\mu}_\theta, \tilde{\sigma}_\theta^2), \quad \zeta \sim \mathrm{N}(\tilde{\mu}_\zeta, \tilde{\sigma}_\zeta^2), \quad \lambda \sim \mathrm{N}_{[-1,1]}(\tilde{\mu}_\lambda, \tilde{\sigma}_\lambda^2), \tag{38b}$$

$$\sigma_\varepsilon^2 \sim \mathrm{IG}(\tilde{a}_\varepsilon, \tilde{b}_\varepsilon), \quad \sigma_\kappa^2 \sim \mathrm{IG}(\tilde{a}_\kappa, \tilde{b}_\kappa), \quad \sigma_\gamma^2 \sim \mathrm{IG}(\tilde{a}_\gamma, \tilde{b}_\gamma), \tag{38c}$$

where $\mathrm{N}_{[-1,1]}$ denotes a truncated Gaussian with support $[-1, 1]$ and $\mathrm{IG}(\tilde{a}, \tilde{b})$ denotes an inverse-gamma distribution with mean $\tilde{b}/(\tilde{a}-1)$ and variance $\tilde{b}^2/((\tilde{a}-1)^2(\tilde{a}-2))$ for $\tilde{a} > 2$. The posteriors of the static parameters are then obtained as follows:[6]

$$\alpha_x|\boldsymbol{y}, \boldsymbol{\varphi}, \boldsymbol{\psi}_{-\alpha_x} \sim \mathrm{N}\left(\frac{\tilde{\sigma}_\alpha^2 \sum_{t=1}^n (y_{x,t} - \beta_x \kappa_t - \beta_x^\gamma \gamma_t^x) + \tilde{\mu}_\alpha \sigma_\varepsilon^2}{\tilde{\sigma}_\alpha^2 n + \sigma_\varepsilon^2}, \frac{\tilde{\sigma}_\alpha^2 \sigma_\varepsilon^2}{\tilde{\sigma}_\alpha^2 n + \sigma_\varepsilon^2}\right), \tag{39}$$

$$\beta_x|\boldsymbol{y}, \boldsymbol{\varphi}, \boldsymbol{\psi}_{-\beta_x} \sim \mathrm{N}\left(\frac{\tilde{\sigma}_\beta^2 \sum_{t=1}^n (y_{x,t} - (\alpha_x + \beta_x^\gamma \gamma_t^x))\kappa_t + \tilde{\mu}_\beta \sigma_\varepsilon^2}{\tilde{\sigma}_\beta^2 \sum_{t=1}^n \kappa_t^2 + \sigma_\varepsilon^2}, \frac{\tilde{\sigma}_\beta^2 \sigma_\varepsilon^2}{\tilde{\sigma}_\beta^2 \sum_{t=1}^n \kappa_t^2 + \sigma_\varepsilon^2}\right), \tag{40}$$

$$\beta_x^\gamma|\boldsymbol{y}, \boldsymbol{\varphi}, \boldsymbol{\psi}_{-\beta_x^\gamma} \sim \mathrm{N}\left(\frac{\tilde{\sigma}_{\beta^\gamma}^2 \sum_{t=1}^n (y_{x,t} - (\alpha_x + \beta_x \kappa_t))\gamma_t^x + \tilde{\mu}_{\beta^\gamma} \sigma_\varepsilon^2}{\tilde{\sigma}_{\beta^\gamma}^2 \sum_{t=1}^n (\gamma_t^x)^2 + \sigma_\varepsilon^2}, \frac{\tilde{\sigma}_{\beta^\gamma}^2 \sigma_\varepsilon^2}{\tilde{\sigma}_{\beta^\gamma}^2 \sum_{t=1}^n (\gamma_t^x)^2 + \sigma_\varepsilon^2}\right), \tag{41}$$

$$\theta|\boldsymbol{y}, \boldsymbol{\varphi}, \boldsymbol{\psi}_{-\theta} \sim \mathrm{N}\left(\frac{\tilde{\sigma}_\theta^2 \sum_{t=1}^n (\kappa_t - \kappa_{t-1}) + \tilde{\mu}_\theta \sigma_\omega^2}{\tilde{\sigma}_\theta^2 n + \sigma_\omega^2}, \frac{\tilde{\sigma}_\theta^2 \sigma_\omega^2}{\tilde{\sigma}_\theta^2 n + \sigma_\omega^2}\right), \tag{42}$$

$$\zeta|\boldsymbol{y}, \boldsymbol{\varphi}, \boldsymbol{\psi}_{-\theta} \sim \mathrm{N}\left(\frac{\tilde{\sigma}_\zeta^2 \sum_{t=1}^n (\gamma_t^{x_1} - \lambda \gamma_{t-1}^{x_1}) + \tilde{\mu}_\zeta \sigma_\gamma^2}{\tilde{\sigma}_\zeta^2 n + \sigma_\gamma^2}, \frac{\tilde{\sigma}_\zeta^2 \sigma_\gamma^2}{\tilde{\sigma}_\zeta^2 n + \sigma_\gamma^2}\right), \tag{43}$$

[6]For simplicity, we denote $\boldsymbol{y} = \boldsymbol{y}_{1:n}$, $\boldsymbol{\varphi} = \boldsymbol{\varphi}_{0:n}$ and $\boldsymbol{\psi}_{-h} = (\psi_1, \ldots, \psi_{h-1}, \psi_{h+1}, \ldots, \psi_{3p+6})$.

$$\lambda|\boldsymbol{y},\boldsymbol{\varphi},\boldsymbol{\psi}_{-\lambda} \sim \mathrm{N}_{[-1,1]}\left(\frac{\tilde{\sigma}_\lambda^2\sum_{t=1}^n((\gamma_t^{x_1}-\zeta)\gamma_{t-1}^{x_1})+\tilde{\mu}_\lambda\sigma_\gamma^2}{\tilde{\sigma}_\lambda^2\sum_{t=1}^n(\gamma_{t-1}^{x_1})^2+\sigma_\gamma^2}, \frac{\tilde{\sigma}_\lambda^2\sigma_\gamma^2}{\tilde{\sigma}_\lambda^2\sum_{t=1}^n(\gamma_{t-1}^{x_1})^2+\sigma_\gamma^2}\right), \quad (44)$$

$$\sigma_\varepsilon^2|\boldsymbol{y},\boldsymbol{\varphi},\boldsymbol{\psi}_{-\sigma_\varepsilon^2} \sim \mathrm{IG}\left(\tilde{a}_\varepsilon+\frac{np}{2}, \tilde{b}_\varepsilon+\frac{1}{2}\sum_{x=x_1}^{x_p}\sum_{t=1}^n\left(y_{x,t}-(\alpha_x+\beta_x\kappa_t+\beta_x^\gamma\gamma_t^x)\right)^2\right), \quad (45)$$

$$\sigma_\kappa^2|\boldsymbol{y},\boldsymbol{\varphi},\boldsymbol{\psi}_{-\sigma_\kappa^2} \sim \mathrm{IG}\left(\tilde{a}_\kappa+\frac{n}{2}, \tilde{b}_\kappa+\frac{1}{2}\sum_{t=1}^n\left(\kappa_t-(\kappa_{t-1}+\theta)\right)^2\right), \quad (46)$$

$$\sigma_\gamma^2|\boldsymbol{y},\boldsymbol{\varphi},\boldsymbol{\psi}_{-\sigma_\gamma^2} \sim \mathrm{IG}\left(\tilde{a}_\gamma+\frac{n}{2}, \tilde{b}_\gamma+\frac{1}{2}\sum_{t=1}^n\left(\gamma_t^{x_1}-\lambda\gamma_{t-1}^{x_1}\right)^2\right). \quad (47)$$

We note here that if one is interested in modelling the cohort factor as a random walk with drift process which is non-stationary, one may fix $\lambda = 1$ and hence there is no need to estimate the parameter λ.

4.2 Bayesian inference for nested models

The MCMC estimation for the full cohort model presented in Algorithm 1 can be applied to the nested models including the simplified cohort model and the LC model with only small adjustments.

The static parameter vector for the simplified cohort model is given by

$$\boldsymbol{\psi} := (\boldsymbol{\beta},\boldsymbol{\alpha},\theta,\zeta,\lambda,\sigma_\varepsilon^2,\sigma_\kappa^2,\sigma_\gamma^2), \quad (48)$$

where the sampling of the age-modulating coefficients β_x^γ for the cohort factor is not required. Consequently line 9-12 in Algorithm 1 can be removed in this case. Moreover, for the simplified cohort model, we set $\beta_x^\gamma = 1$ in the posterior distributions of α_x, β_x and σ_ε^2 in (39), (40) and (45) respectively.

The LC model can be viewed as a further nested model of the simplified cohort model with the cohort factor $\gamma_{t-x} = 0$. As a result the state equation is one-dimensional with $\boldsymbol{\varphi}_{0:n} := \kappa_{0:n}$ and the static parameter vector consists of

$$\boldsymbol{\psi} := (\boldsymbol{\beta},\boldsymbol{\alpha},\theta,\sigma_\varepsilon^2,\sigma_\kappa^2). \quad (49)$$

Recall the constraints (3) for the LC model. As a result the constraint $\sum_c \gamma_c^{(i)} = 0$ in line 4 and the sampling of β_x^γ in line 9-12 in Algorithm 1 are not required. We also set $\gamma_t^x = 0$ in the posterior distributions for $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and σ_ε^2 in (39), (40) and (45) respectively. Further details for a Bayesian estimation of the LC model can be found in Fung et al. (2017).

5 Empirical Studies

We analyse several sets of mortality data from different countries based on the cohort models formulated in the state-space framework. We consider both the full cohort model

and the simplified cohort model. In addition, we compare the cohort models against the LC model for model fitting as well as their forecasting properties. The countries that we consider includes England and Wales (UK), United States (US) and Italy (ITA). We perform our analysis on both male and female mortality data to investigate whether cohort effect within a country is shared for both genders. The data is obtained from the Human Mortality Database[7]. The year range is from year 1970 to 2010, and we restrict our attention to the age range 65-95. Consequently the range for the year-of-birth is 1875-1945.

5.1 Model estimation

We run the Markov chain sampler described in Section 4 for 30,000 iterations and the burn-in period is set to be 15,000 iterations to ensure that the chain has arrived to the stationary state; thus we are left with 15,000 posterior samples. For all Gaussian priors $\mathrm{N}(\tilde{\mu}, \tilde{\sigma}^2)$, including the truncated Gaussian $\mathrm{N}_{[-1,1]}(\tilde{\mu}, \tilde{\sigma}^2)$, we assume $\tilde{\mu} = 0$ and $\tilde{\sigma}^2 = 10$; while for the inverse-gamma priors $\mathrm{IG}(\tilde{a}, \tilde{b})$ we set $\tilde{a} = 2.01$ and $\tilde{b} = 0.01$. The hyperparameters are chosen and tested to ensure the priors are sufficiently vague. To start the Kalman filter, we assume $\boldsymbol{m}_0 = \mathbf{0}$ and C_0 is a diagonal covariance matrix with diagonal elements all equal to 10 (see (34a)). We refer $x = \{x_1, \ldots, x_p\}$ to age $= \{65, \ldots, 95\}$ and $t = \{t_1, \ldots, t_n\}$ to year $= \{1970, \ldots, 2010\}$ (i.e. $p = 31$ and $T = 41$).

As noted previously, no particular special initialization is required with this methodology, it seems relatively robust to the choice of starting points. To start the chain for the full cohort model, initial values of the static parameters are set as follows: $\alpha_x^{(0)} = (1/n)\sum_t y_{x,t}$, $\beta_x^{(0)} = (\beta_x^{\gamma})^{(0)} = 1/p$, $\theta^{(0)} = \zeta^{(0)} = -0.1$, $(\sigma_\varepsilon^2)^{(0)} = (\sigma_\omega^2)^{(0)} = (\sigma_\gamma^2)^{(0)} = 0.01$ and $\lambda^{(0)} = 0.5$ where $x \in \{x_1, \ldots, x_p\}$. The same set of values is also used for the simplified cohort model except that sampling of β_x^{γ} is not required.

5.1.1 Full cohort model

For the full cohort model, the posterior mean and 95% credible intervals for the parameters $\boldsymbol{\kappa}$, $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^{\gamma}$ for the UK, US and Italy male populations are shown in Figure 1.

Given the linear trend observed, the assumption that the period effect follows a random walk with drift process seems to be reasonable for the UK and Italy male populations, but is less appropriate for the US male population where structural changes seem to be present for the considered age range (see also Li et al. (2011) and van Berkum et al. (2016)). The estimated cohort factor shows a clear kink around the year-of-birth at 1920 for the UK and Italy data. For the England & Wales population, Cairns et al. (2016) finds evidence to suggest that the anomalies observed for the 1919 & 1920 cohorts are due to unusual pattern of births and the methodology employed to estimate mid-year population, which is related to the First World War. A mild kink also appears around the generation born in 1900 for the US population. The implication for these sudden changes of the cohort factor on model fit will be discussed in more details in Section 5.2 where

[7] www.mortality.org

a comparison of the cohort models with the LC model through residuals is presented. The heightened volatility surrounding the first several years of the cohort factor can be attributed to having limited data that are used to infer the cohort factor for the very first birth years, thus creating greater uncertainty.

Figure 2 displays the corresponding results for female populations. We also observe similar patterns where the kinks in the cohort factors for the female populations appear at almost exactly the same generations as in the male populations. It suggests that, at least for the countries that we presented here, cohort effect is not gender-specific but is a common phenomenon for people who belong to a particular generation. Such a finding in this case can perhaps be attributed to shared mortality experience in both genders arising from a significant global effect for Europe, corresponding to the First and Second World War.

Posterior statistics for the other parameters are displayed in Table 3.

	UK	US	Italy
	Males		
θ	-0.18 [-0.40, 0.02]	-0.20 [-0.35, -0.04]	-0.37 [-0.63, -0.10]
ζ	-0.57 [-0.79, -0.36]	-0.21 [-0.29, -0.14]	-0.55 [-0.98, -0.12]
λ	0.993 [0.977, 0.999]	0.990 [0.975, 0.999]	0.98 [0.94, 0.99]
σ_ε^2	0.00028 [0.00026, 0.00030]	0.00020 [0.00019, 0.00022]	0.00032 [0.00030, 0.00035]
σ_ω^2	0.46 [0.29, 0.72]	0.23 [0.14, 0.36]	0.71 [0.44, 1.10]
σ_γ^2	0.46 [0.28, 0.72]	0.019 [0.008, 0.03]	1.93 [1.22, 3.00]
	Females		
θ	-0.19 [-0.41, 0.02]	-0.51 [-0.70, -0.33]	-0.42 [-0.73, -0.11]
ζ	-0.37 [-0.56, -0.19]	0.38 [0.17, 0.61]	-0.51 [-0.88, -0.14]
λ	0.990 [0.966, 0.999]	0.89 [0.81, 0.96]	0.98 [0.94, 0.99]
σ_ε^2	0.00023 [0.00021, 0.00025]	0.00022 [0.00020, 0.00024]	0.00029 [0.00027, 0.00032]
σ_ω^2	0.52 [0.33, 0.81]	0.34 [0.22, 0.54]	0.98 [0.61, 1.54]
σ_γ^2	0.35 [0.22, 0.56]	0.07 [0.04, 0.13]	1.41 [0.90, 2.19]

Table 3: Estimated posterior mean of the static parameters for the full cohort model on male and female population data. [., .] next to the estimates represents 95% posterior credible interval.

5.1.2 Simplified cohort model

Figure 3 shows the posterior mean and 95% posterior credible intervals for the estimated parameters $\boldsymbol{\kappa}$, $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ where UK, US and Italy males mortality data are used to fit the simplified cohort model where it is assumed that $\beta_x^\gamma = 1$ for all age x.

The estimated figures from the simplified cohort model share very similar patterns to those obtained from the full cohort model for the UK and Italy data. Interestingly, the same may not be said for the US data, as the estimated $\boldsymbol{\kappa}$ and $\boldsymbol{\gamma}$ are quite different to the corresponding ones obtained from the full cohort model. As the only difference between

the simplified and the full cohort model is that β_x^γ is allowed to be flexible in the full cohort model, the rather substantial difference for the estimated $\boldsymbol{\kappa}$ and $\boldsymbol{\gamma}$ between the simplified and full cohort models might indicate that perhaps the cohort effect is so weak in the data that the model could not distinguish the period effect and cohort effect. We will further examine the possibility of the lack of cohort effect in US mortality data in Section 5.2. Note also that the estimated values of $\boldsymbol{\gamma}$ for the simplified model are significantly smaller than the full model. The reason for this is that the constraint $\sum_x \beta_x^\gamma = 1$ for the full model dictates that β_x^γ will take value in the order of $1/p$ which is substantially smaller than 1, where $p = 31$ is the number of ages considered in the data; while $\beta_x^\gamma = 1$ is assumed for the simplified model. One may consider fixing $\beta_x^\gamma = 1/p$ in the simplified model so that the estimated values of $\boldsymbol{\gamma}$ will be in the same order of magnitude for both the simplified and full model. However this should not affect the fitting and forecasting of the resulting model.

The corresponding estimation results for female populations are shown in Figure 4. We again observe that the estimates show similarity between the male and female populations for all the three countries, thus suggesting that mortality experiences for both genders share major characteristics.

Posterior estimates for other static parameters for the simplified cohort model on male and female mortality data are reported in Table 4.

	UK	US	Italy
	Males		
θ	-0.22 [-0.44, 0.007]	-0.50 [-0.67, -0.33]	-0.30 [-0.56, -0.03]
ζ	-0.022 [-0.034, -0.011]	-0.003 [-0.013, -0.007]	-0.019 [-0.034, -0.004]
λ	0.991 [0.970, 0.999]	0.973 [0.909, 0.999]	0.982 [0.944, 0.999]
σ_ε^2	0.00035 [0.00032, 0.00038]	0.00025 [0.00023, 0.00027]	0.00033 [0.00031, 0.00036]
σ_ω^2	0.46 [0.29, 0.73]	0.23 [0.14, 0.37]	0.68 [0.43, 1.06]
σ_γ^2	0.0012 [0.0008, 0.0019]	0.0007 [0.0004, 0.0010]	0.0023 [0.0015, 0.0036]
	Females		
θ	-0.40 [-0.64, -0.14]	-0.61 [-0.81, -0.41]	-0.39 [-0.70, -0.08]
ζ	-0.013 [-0.024, -0.002]	0.018 [0.0005, 0.043]	-0.019 [-0.033, -0.005]
λ	0.98 [0.95, 0.99]	0.89 [0.74, 0.99]	0.97 [0.93, 0.99]
σ_ε^2	0.00025 [0.00023, 0.00028]	0.00025 [0.00023, 0.00027]	0.00032 [0.00029, 0.00035]
σ_ω^2	0.60 [0.38, 0.94]	0.38 [0.24, 0.60]	0.95 [0.60, 1.50]
σ_γ^2	0.0012 [0.0008, 0.0019]	0.0006 [0.0004, 0.0010]	0.002 [0.001, 0.003]

Table 4: Estimated posterior mean of the static parameters for the simplified cohort model on male and female population data. [.,.] represents 95% posterior credible interval.

5.1.3 Robustness of the Bayesian state-space approach

To compare the robustness of the estimation performance of the Bayesian state-space approach with the GLM approach (Section 2.2), we have tested the estimation of cohort

models via the GLM approach using the R package called StMoMo (Version 0.4.0), see Villegas et al. (2015), Hunt and Villegas (2015).

We consider males mortality data from seven countries including US, UK, France, Australia, Denmark, Netherland and Italy. When fitting the cohort models in StMoMo (using the rh() function with log link), we employ the estimated parameters from the LC model as the initial values of the estimation scheme for the cohort models as suggested in Villegas et al. (2015). The age-range and year-range are 65-95 and 1970-2010 respectively.

For the simplified cohort model, the estimation scheme under the GLM approach fails to converge for data from US, France and Netherland. In contrast, the Bayesian state-space approach is able to estimate the model for all mortality data from the countries considered with sensible results produced.

For the full cohort model, the GLM approach is successful in estimating the model for data from France and Italy while unsuccessful for data from US, UK, Australia, Denmark and Netherland. On the other hand, the Bayesian state-space approach is able to estimate the model for all mortality data from the countries considered again with sensible results produced.

It should be note that the population data we have considered are not exhaustive but the results suggest that the Bayesian state-space approach is more robust compared to the GLM approach for the estimation of cohort models confirming the robustness issues found in Hunt and Villegas (2015).[8]

5.2 Model fitting: comparison with LC model

We examine the fitting of the cohort models via residual heatmap as well as model ranking via deviance information criterion (DIC). We emphasize the importance of a comparison for the cohort models with the LC model since if cohort patterns are present in mortality data, the ordinary LC model should not be able to capture this phenomenon but cohort models are designed to accomplish this.

5.2.1 Residual heatmap

Residuals under a state-space model are defined as the difference between the observed data $\boldsymbol{y}_t$ and the mean of the in-sample one-step-ahead model forecast given by (see (34c))

$$\mathbb{E}\left[\boldsymbol{y}_t|\boldsymbol{\psi}, \boldsymbol{y}_{1:t-1}\right] = \boldsymbol{f}_t, \tag{50}$$

where $t = 1, \ldots, n$. Explicitly, we have

$$\boldsymbol{e}_t := \boldsymbol{y}_t - \boldsymbol{f}_t \tag{51}$$

where $\boldsymbol{e}_t$ is the vector of residuals at time t. Using the posterior mean as point estimator for the static parameters, we obtain the mean of the in-sample one-step-ahead forecast $\boldsymbol{f}_t$ via Kalman filtering for the cohort models and the LC model.

[8]Results for the countries including France, Australia, Denmark and Netherland are available upon request.

Figure 5 shows residual heatmaps produced from the LC model, the simplified cohort model and the full cohort model for the UK, US and Italy male populations. The distinctive diagonal bands observed for the fitting of the LC model for the UK and Italy population, which correspond to the generations born around 1920, clearly suggest that cohort effects are strongly present in these countries and the LC model fails to account for these patterns. In contrast, the cohort models considered are capable of capturing these effects which is supported by the observation that the diagonal bands are removed from the residual plots. These results are consistent with the estimated cohort factors shown in Figure 1 and 2 where an apparent irregularity appears around the cohort born in 1920.

Residual heatmap from the LC model for the US male population, on the other hand, shows only small traces of diagonal bands around the cohorts 1900 and 1920 which are barely noticeable. The corresponding plots from the cohort models show that these bands are diminished even further. The very faint occurrence of the diagonal bands for the US males data is compatible with the remarks suggested in Section 5.1 that there is no clear evidence that the US male population exhibits certain cohort patterns.

Corresponding residual heatmaps for the female populations are displayed in Figure 6. The results follow very closely to the discussion for the male populations. It provides further evidence that major mortality characteristics including cohort effects are not gender-specific in the considered countries.

The residual plots shown here also suggest that the deciding factor for the presence of cohort patterns in a specific country depend on whether there are any abrupt irregularities observed in the estimated cohort factor; such non-smooth irregularities indicate that strong cohort patterns exist in the data where the LC model is not able to capture. On the other hand, if the estimated cohort factor is reasonably smooth as for the US data, cohort factors in mortality models may be not be required.

5.2.2 Deviance information criterion

We perform model ranking using deviance information criterion (DIC) which is designed specifically for Bayesian models taking into account the trade-off between model fits and complexity (Spiegelhalter et al. (2002)). There are several versions of DIC and here we will focus on the so-called conditional DIC where the latent states are treated as parameters when calculating the conditional likelihood (Celeux et al. (2006), Chan and Grant (2016)).

Specifically, the conditional DIC utilises the conditional log-likelihood which is given by

$$\ln \pi(\boldsymbol{y}_{1:n}|\boldsymbol{\psi}, \boldsymbol{\varphi}_{0:n}) = -\frac{1}{2}\sum_{x=x_1}^{x_p}\sum_{t=t_1}^{t_n}\left(\ln 2\pi\sigma_\varepsilon^2 + \left(\frac{y_{x,t} - (\alpha_x + \beta_x\kappa_t + \beta_x^\gamma\gamma_{t-x})}{\sigma_\varepsilon}\right)^2\right), \quad (52)$$

for the full cohort model; conditional log-likelihoods for the simplified cohort model and the LC model can be obtained similarly. Denote $\boldsymbol{\Psi} := (\boldsymbol{\psi}, \boldsymbol{\varphi}_{0:n})$, one defines

$$D(\boldsymbol{\Psi}) = -2\ln \pi(\boldsymbol{y}_{1:n}|\boldsymbol{\Psi}) + 2\ln h(\boldsymbol{y}_{1:n}), \quad (53)$$

as the deviance of the model. As $h(\boldsymbol{y}_{1:n})$ is independent to the models considered, it is typical to assume $h(\boldsymbol{y}_{1:n}) = 1$. The effective number of parameters p_D is defined to be

$$p_D = \bar{D}(\boldsymbol{\Psi}) - D(\bar{\boldsymbol{\Psi}}), \tag{54}$$

where $\bar{D}(\boldsymbol{\Psi})$ is the mean of $D(\boldsymbol{\Psi})$ while $\bar{\boldsymbol{\Psi}}$ is the posterior mean of $\boldsymbol{\Psi}$. One then defines the conditional DIC as

$$\text{DIC} := \bar{D}(\boldsymbol{\Psi}) + p_D = 2\bar{D}(\boldsymbol{\Psi}) - D(\bar{\boldsymbol{\Psi}}), \tag{55}$$

which can be calculated using the MCMC samples obtained as described in Section 4. Note that models with smaller DIC values are ranked higher then models with larger DIC values.

	LC model	Simplified cohort model	Full cohort model
		Males	
UK	-5418	-6376	**-6666**
US	-5575	-6836	**-7111**
ITA	-4758	-6433	**-6474**
		Females	
UK	-5053	-6794	**-6910**
US	-5395	-6824	**-6993**
ITA	-5098	-6485	**-6607**

Table 5: DIC values for the considered models on male and female population data.

Table 5 reports the estimated conditional DIC values obtained for the considered models on male and female population data. We observe that the improved fits are more than compensated for the complexity arising from the inclusion of cohort factors for the considered countries based on the estimated DIC values. The improvement of model fits is more significant for the simplified cohort model over the LC model than the full cohort model over the simplified cohort model. Interestingly, there is a distinct improvement in using cohort models over the LC model on the US mortality data, despite the lack of cohort patterns for the US data shown in Figure 5 and 6. It suggests that multi-factor models are preferred for the US population over a single period factor model such as the LC model. Whether a multi-period model or a period-cohort model is preferred for the US data is an interesting question which is, however, beyond the scope of the current paper.

5.3 Forecasting from cohort models

Forecasting of death rates for the cohort models and LC model is studied here. One of the advantages of considering Bayesian inference via MCMC sampling for state-space mortality models is that the forecasting distributions of death rates can be derived rigorously

and the samples of the forecasting distributions can be easily generated from the samples obtained from the MCMC estimation step.

Explicitly, we can express the forecasting distribution for the cohort models as

$$\pi(\boldsymbol{y}_{n+k}|\boldsymbol{y}_{1:n}) = \int \pi(\boldsymbol{y}_{n+k}|\boldsymbol{\varphi}_{n+k},\boldsymbol{\psi})\pi(\boldsymbol{\varphi}_{n+k}|\boldsymbol{\varphi}_{n+k-1},\boldsymbol{\psi})\ldots\pi(\boldsymbol{\varphi}_{n},\boldsymbol{\psi}|\boldsymbol{y}_{1:n})\,d\boldsymbol{\psi}d\boldsymbol{\varphi}_{n:n+k}, \quad (56)$$

where $\pi(\boldsymbol{y}_{n+k}|\boldsymbol{y}_{1:n})$ is the k-step ahead forecasting posterior predictive distribution. It means that one can simulate the forecasting samples of the dynamic factors to obtain recursively the forecasting samples of death rates as follow

$$\boldsymbol{\varphi}_{n+k}^{(\ell)} \sim \mathrm{N}\left(\Lambda^{(\ell)}\boldsymbol{\varphi}_{n+k-1}^{(\ell)} + \Theta^{(\ell)}, \Upsilon^{(\ell)}\right), \quad (57a)$$

$$\boldsymbol{y}_{n+k}^{(\ell)} \sim \mathrm{N}\left(\boldsymbol{\alpha}^{(\ell)} + B^{(\ell)}\boldsymbol{\varphi}_{n+k}^{(\ell)}, \left(\sigma_{\varepsilon}^{2}\right)^{(\ell)}\mathbf{1}_{p}\right), \quad (57b)$$

for the full cohort model, where $\ell = 1,\ldots,L$ and L is the number of MCMC samples after burn-in. For the simplified cohort model one simply replaces B by B^{s}. Forecasting for the LC model can be carried out by setting $\boldsymbol{\varphi} = \kappa$ in (56) and forecasting samples of death rates are obtained recursively as follow

$$\kappa_{n+k}^{(\ell)} \sim \mathrm{N}\left(\kappa_{n+k-1}^{(\ell)} + \theta^{(\ell)}, \left(\sigma_{\omega}^{2}\right)^{(\ell)}\right), \quad (58a)$$

$$\boldsymbol{y}_{n+k}^{(\ell)} \sim \mathrm{N}\left(\boldsymbol{\alpha}^{(\ell)} + \boldsymbol{\beta}^{(\ell)}\kappa_{n+k}^{(\ell)}, \left(\sigma_{\varepsilon}^{2}\right)^{(\ell)}\mathbf{1}_{p}\right), \quad (58b)$$

which is a special case of (57).

5.3.1 Projection of death rates

Figure 7 displays the mean and 95% posterior predictive forecast interval of the forecasted death rates from the full cohort model and LC model based on UK, US and Italy male mortality data for selected ages 65, 70, 75 and 80.

We observe that forecasts from the full cohort model show substantial difference to the forecasts from the LC model for the UK and Italy populations, while the difference is noticeably smaller for the US data. It is consistent with our results found in Section 5.1 and 5.2 that fitting of the cohort models suggest that cohort effect is strong for the UK and Italy populations but is weak for the US mortality data. As a result, it is expected that, from the forecasting perspective, cohort models will show greater difference to the LC model for the UK and Italy data but will be less so for the US data, which is confirmed by the forecasting results shown here.

Another important observation is that there is a clear change of trend for the forecasted death rates from the full cohort model. A closer inspection suggests that the change appears at the generation born in 1945. For example, the change is located in year 2020 for the forecasted death rates for age 75, which corresponds to the cohort with year-of-birth at 1945; the same conclusion can be drawn from the forecasted death rates for other ages. Interestingly, the most recent generation considered in our data is exactly the

year-of-birth at 1945.[9] This observation indicates that this change of trend behaviour is a consequence of the projection of the cohort factor which starts at 1945.[10] We provide further discussion on this issue in details in Section 5.3.2.

The corresponding forecasting distributions based on female mortality data are shown in Figure 8. The observations remarked above can also be applied in this case. As discussed in Section 5.1 and 5.2, we find that male and female populations in the studied countries show similar mortality features including cohort patterns in terms of model fitting. The result here signals that the same can be said from the forecasting standpoint.

A comparison of forecasting from the full cohort model and the simplified cohort model is shown in Figure 9 for the male populations and in Figure 10 for the female populations. It is clear that the full and simplified models generate similar forecasting distribution for the UK and Italy populations. However, the forecasting intervals produced by the simplified model are substantially wider than the intervals generated by the full cohort model for the US male population; the difference is even more pronounced for the female populations. These results suggest that if cohort patterns are strongly present in the data, then there is little difference between the full model and the simplified model in terms of model fitting and forecasting. On the contrary, if the data show only a weak presence of cohort patterns, the full cohort model can lead to significant different fitting and forecasting results to the simplified cohort model, which indicates that cohort models may not be required in this case.

5.3.2 Projection of period and cohort factors

The regular occurrence of trend changing behaviour of the forecasted death rates at year-of-birth 1945, which was discussed in Section 5.3.1, suggest that this phenomenon may originate from the projection of cohort factor starting at year-of-birth 1945.

To investigate this, we plot the projection of the period and cohort factors from the full cohort model in Figure 11 for the male populations and Figure 12 for the female populations in the considered countries.

We first notice that the projected period factor continues the linear trend as expected for both genders in UK, US and Italy populations. It reassures that the projection of the period factor behaves properly and therefore we should focus on the projection of the cohort factor to explain the aforementioned trend-changing behaviour.

Interestingly, we observe from the male populations that there is a slight change of direction of the projected cohort factor starting from the year 1945 compared to the trend during the calibration period 1920-1945 in the UK population. We observe essentially no change of direction for the US population and a significant change of direction for the Italy population. The same observation also applies to the female mortality data.

[9] Age range and year range for our data are 65-95 and 1970-2010 respectively; hence the most recent generation considered is the cohort born in $2010 - 65 = 1945$.

[10] Using the R package StMoMo (Villegas et al. (2015)), which perform estimation and forecasting of mortality models based on the GLM framework, we also observe the change of trend behaviour in the forecasted death rates from the cohort models.

These results strongly indicate that whether a change of trend will be observed for the forecasted death rates depend on the property of the projected cohort factor; namely it depends on whether there will be a change of direction from the estimated cohort factor when the projection starts. In the UK and Italy cases, the change of direction of the projected cohort factor at year 1945 is observed and hence results in the trend-changing behavior of the forecasted death rates shown in Figure 7 and 8, in contrast to the case of US data. Note also that these results are consistent with our discussion in previous sections that cohort patterns are strongly presence in the UK and Italy data while for the US data cohort patterns are not clear.

To investigate the consequences of using non-stationary process to model the cohort factor under the Bayesian state-space framework, we also considered the random walk with drift process to model the cohort factor by fixing $\lambda = 1$ in our estimation of the full cohort model. However, in our setup we found that there is essentially no difference of assuming $\lambda < 1$ (stationary process) and $\lambda = 1$ (non-stationary) in terms of DIC, see Table 6.

The estimated cohort factors shown in Figure 1-4 in fact raise questions about whether it is reasonable to assume that cohort factor dynamics can be captured by stationary ARIMA models which are commonly found in the literature. One may argue that cohort factors apparently exhibit trend changing behaviour. Recently there are growing interests in applying structural change dynamics for the period factor, see Li et al. (2011), van Berkum et al. (2016) and Liu and Li (2016b). The results suggest that this new type of models may be equally suitable for the cohort factor dynamics. However the investigation of this issue is out-of-scope of the current paper and will be left for future research.

	Full cohort model	
	Males ($\lambda < 1$)	Males ($\lambda = 1$)
UK	-6666	-6663
US	-7111	-7085
ITA	-6474	-6475
	Females ($\lambda < 1$)	Females ($\lambda = 1$)
UK	-6910	-6907
US	-6993	-7077
ITA	-6607	-6608

Table 6: DIC values for the full cohort models on male and female population data for cases with $\lambda < 1$ and $\lambda = 1$.

6 Conclusion

In this paper we investigate the formulation, estimation and forecasting of cohort models under the state-space approach. The state-space framework provides a unified environ-

ment where estimation and forecasting of dynamics mortality models can be carried out in a statistically rigorous manner. Continuing development of advanced statistical techniques in numerical filtering and model estimations suggest that state-space method can be an essential tool to handle the modelling of human mortality. The paper contributes to the literature in this area by showing that cohort models are compatible with the state-space framework.

We demonstrate in this paper that the problem of cohort factors being indexed according to year-of-birth instead of year can be overcome by considering a multi-dimensional state-space system. The defining property of the cohort factor imposes a restriction on the observation and state equations of the resulting state-space mortality model. Dynamics of the period and cohort factors are specified in the state equation which allows for a range of time series models to be considered.

By treating the period and cohort factors as the state dynamics of a state-space model, Bayesian inference for cohort models can be performed based on filtering and MCMC method. We develop an efficient MCMC sampler for the resulting model involving a combination of conjugate Gibbs sampling steps for the static parameters and a forward-backward Kalman filtering for the latent state dynamics. The overall algorithm can be applied naturally to the full cohort model as well as its nested models including the simplified cohort model and the LC model.

We apply the cohort models in state-space formulation to analyse male and female mortality data of the UK, US and Italy population. The Bayesian approach allows us to present estimation of the cohort models where parameter uncertainty is properly quantified and accounted for. Our empirical studies show that UK and Italy populations exhibit strong cohort patterns while cohort patterns for the US population are weak. We also find that both genders share common mortality characteristics including cohort effects for the considered countries. Examination of residual heatmaps produced from the cohort models and LC model suggest that cohort models are able to capture cohort effects while the LC model fails to do so. Using DIC for model ranking, we find that the full cohort model is preferred over the simplified cohort model, which in turn outperform the LC model for the considered countries. We show that forecasted death rates from cohort models display trend-changing behaviour at the year where cohort factors are projected, for countries that show strong presence of cohort patterns.

Our estimation and forecasting studies of state-space cohort models suggest that more sophisticated models such as structural change models may be required to capture the dynamics of cohort factors adequately. Moreover, the observed mortality characteristics shared by both genders indicate that one can generalise the approach proposed in this paper to multi-population setting where common period and cohort factors can be assumed for male and female populations within a country. Recently there are also growing interests in applying state-space method to deal with the problem of pricing and hedging of longevity risk (Liu and Li (2016a) and Liu and Li (2016b)). The approach presented in this paper can significantly enhance the capability of the state-space methodology to solving practical problems involving longevity risk since cohort models are shown to be compatible with the state-space framework. Applications of the state-space cohort models

studied in this paper to the key issues in longevity risk pricing and management will be left for future research.

Acknowledgements

This research was supported by the CSIRO-Monash Superannuation Research Cluster, a collaboration among CSIRO, Monash University, Griffith University, the University of Western Australia, the University of Warwick, and stakeholders of the retirement system in the interest of better outcomes for all. This research was also partially supported under the Australian Research Council's Discovery Projects funding scheme (project number: DP160103489). We acknowledge the grant of the interdisciplinary project "Research on Urban Intelligence" from Research Organization of Information and Systems in Japan.

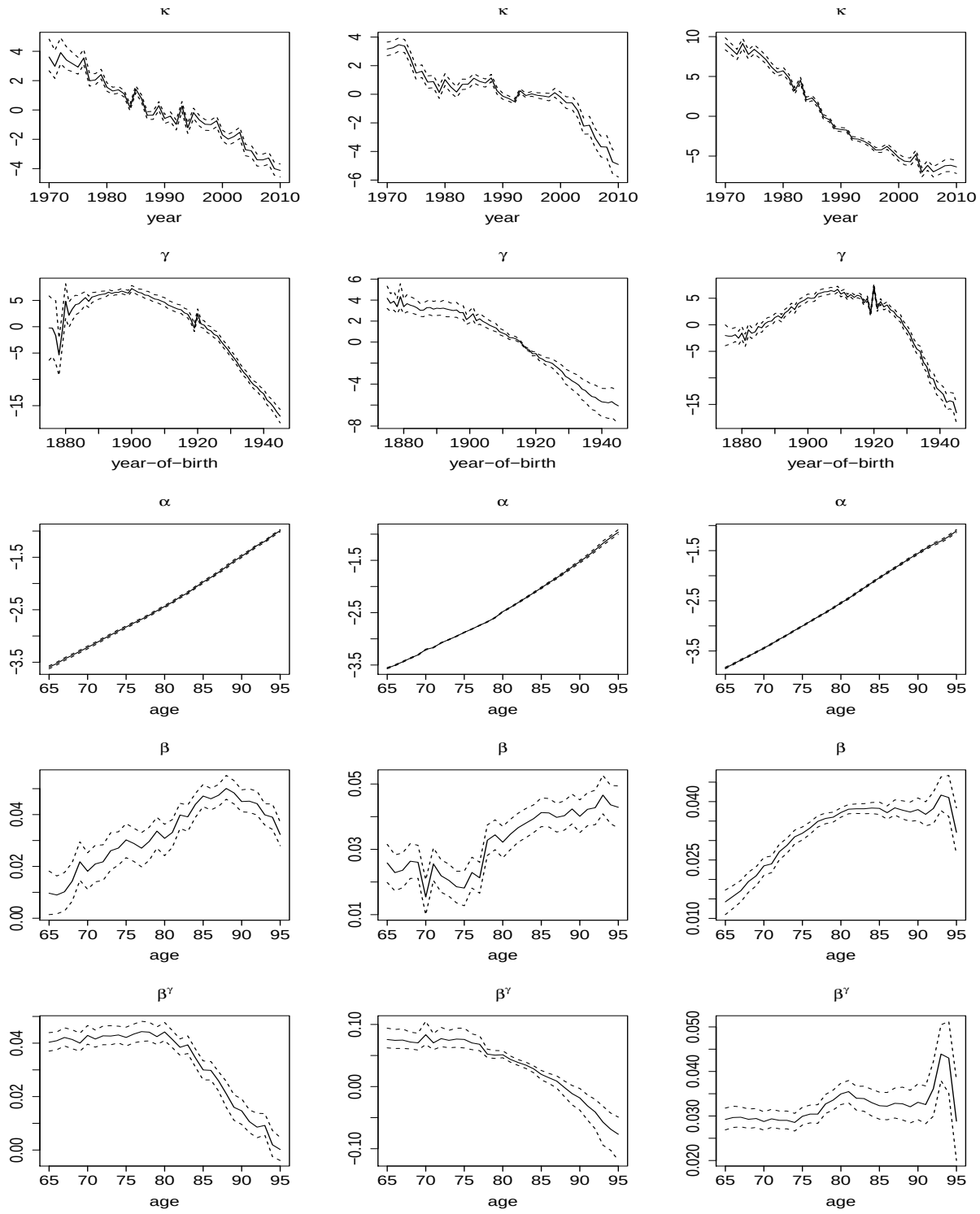


Figure 1: Full cohort model: estimated $\boldsymbol{\kappa}$, $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^{\gamma}$ for the (left column) UK, (middle column) US and (right column) Italy male populations. Mean (solid line) and 95% credible interval (dash lines) of the posterior distributions are plotted.

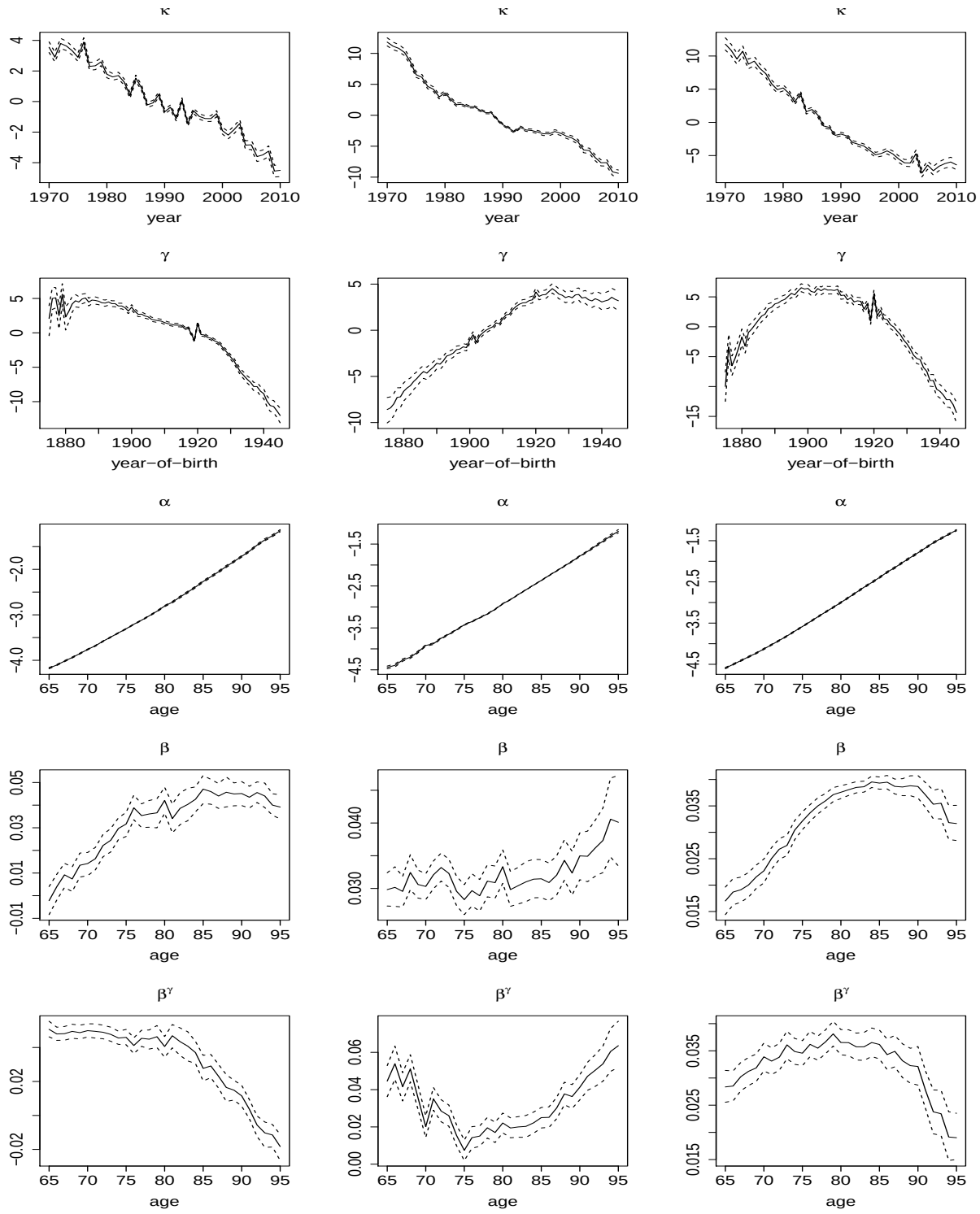


Figure 2: Full cohort model: estimated $\boldsymbol{\kappa}$, $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^{\gamma}$ for the (left column) UK, (middle column) US and (right column) Italy female populations. Mean (solid line) and 95% credible interval (dash lines) of the posterior distributions are plotted.

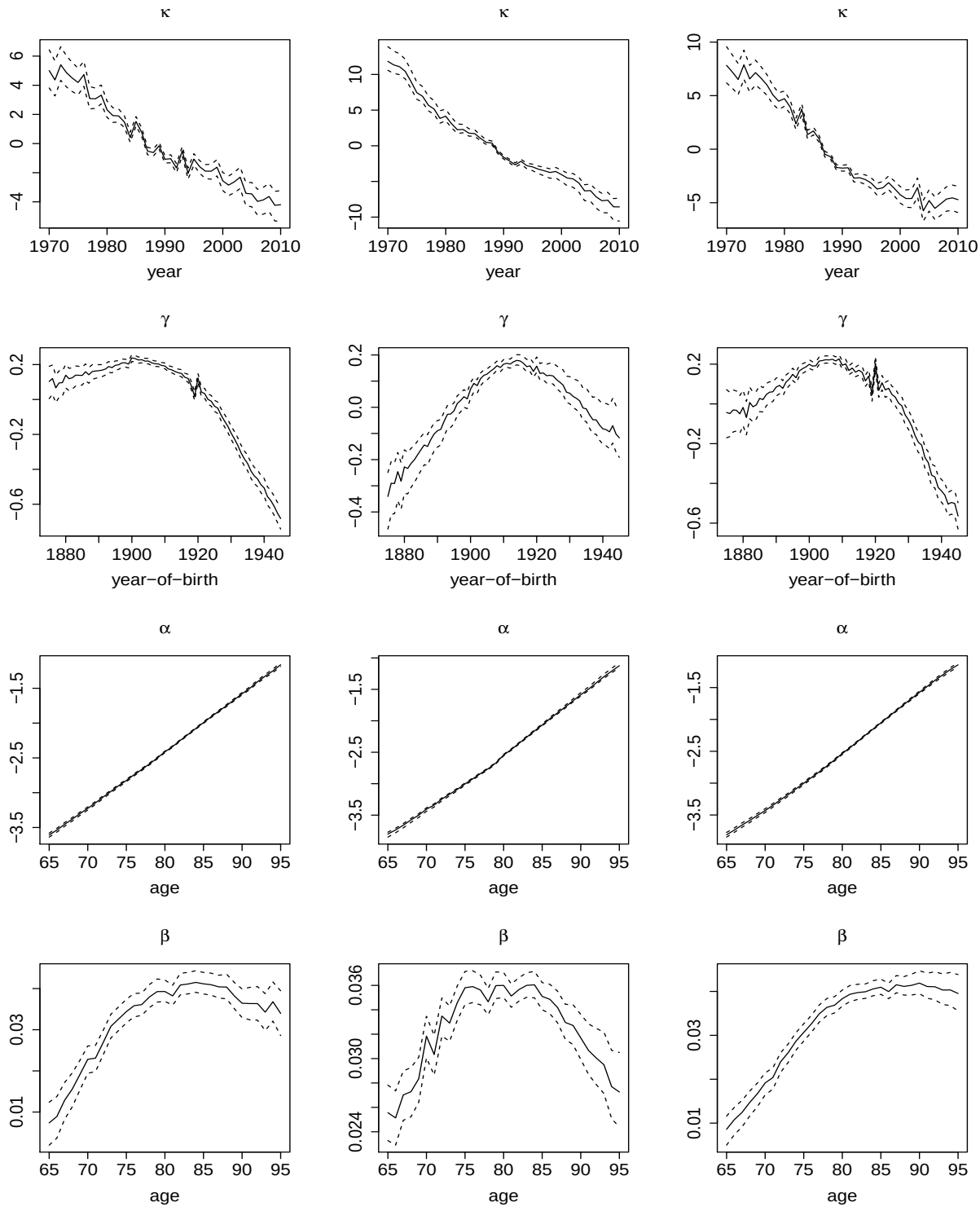


Figure 3: Simplified cohort model: estimated $\boldsymbol{\kappa}$, $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ for the (left column) UK, (middle column) US and (right column) Italy male populations. Mean (solid line) and 95% credible interval (dash lines) of the posterior distributions are plotted.

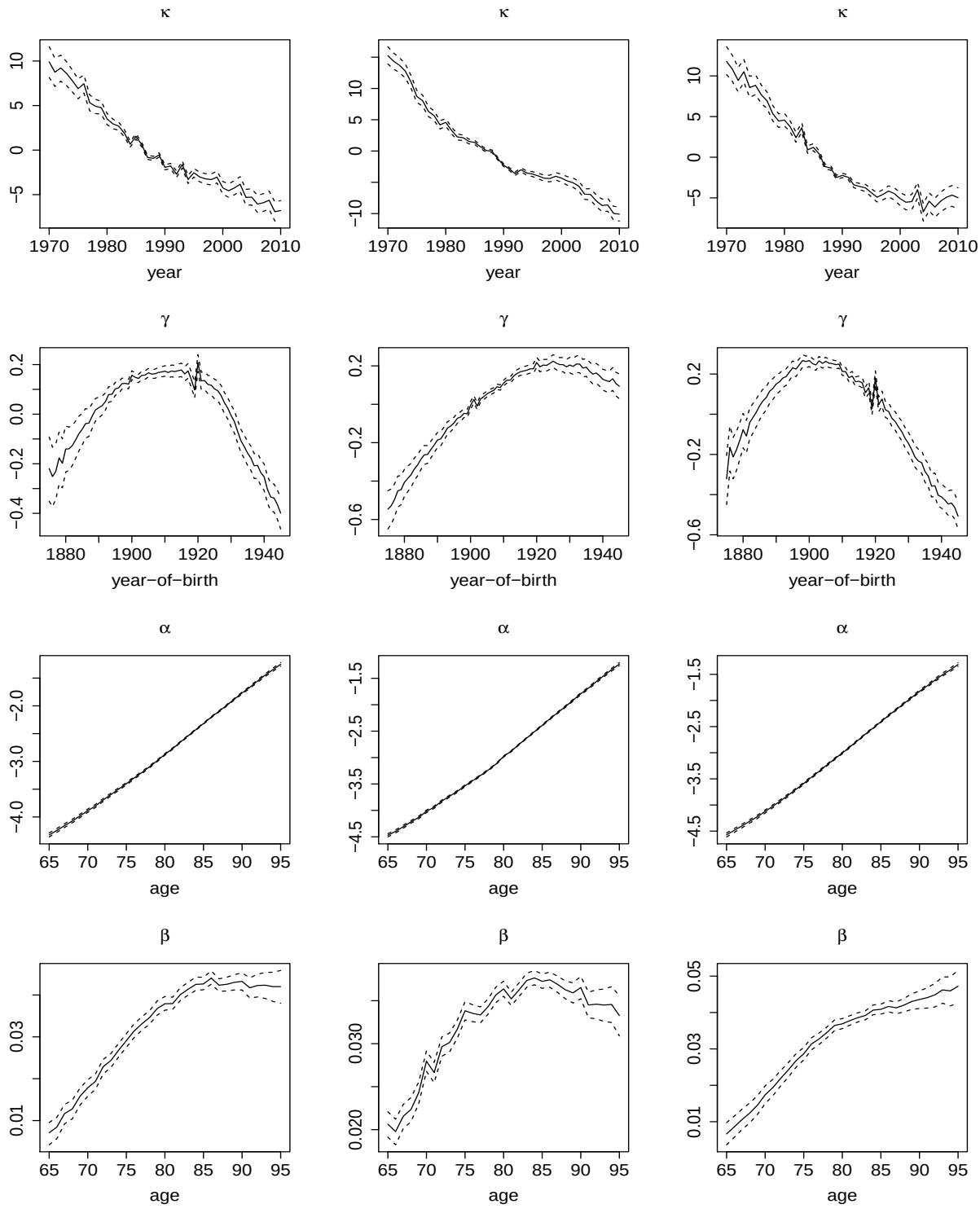


Figure 4: Simplified cohort model: estimated $\boldsymbol{\kappa}$, $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ for the (left column) UK, (middle column) US and (right column) Italy females populations. Mean (solid line) and 95% credible interval (dash lines) of the posterior distributions are plotted.

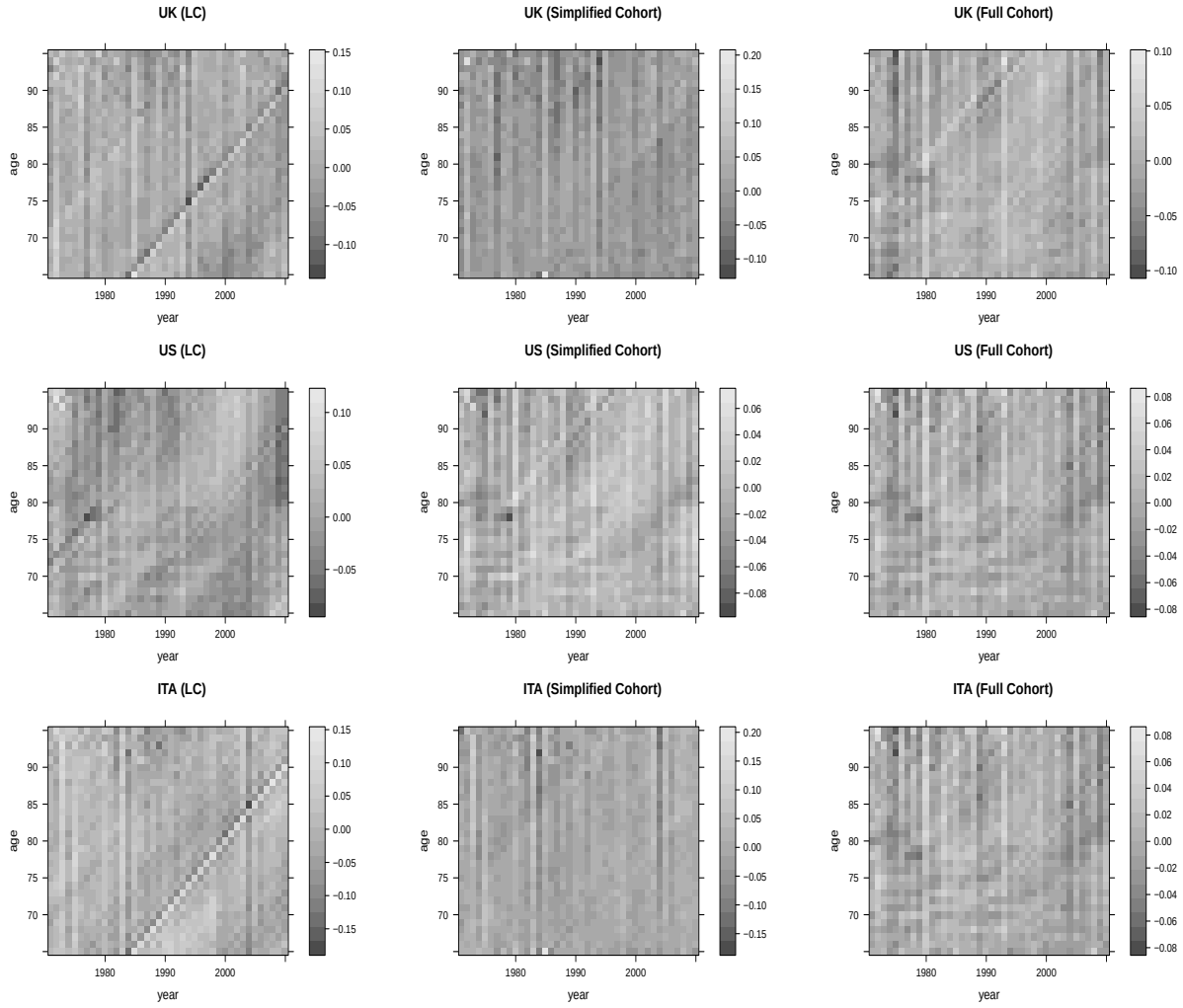


Figure 5: Residual heatmap produced from the LC model, the simplified cohort model and the full cohort model for the UK, US and Italy male populations.

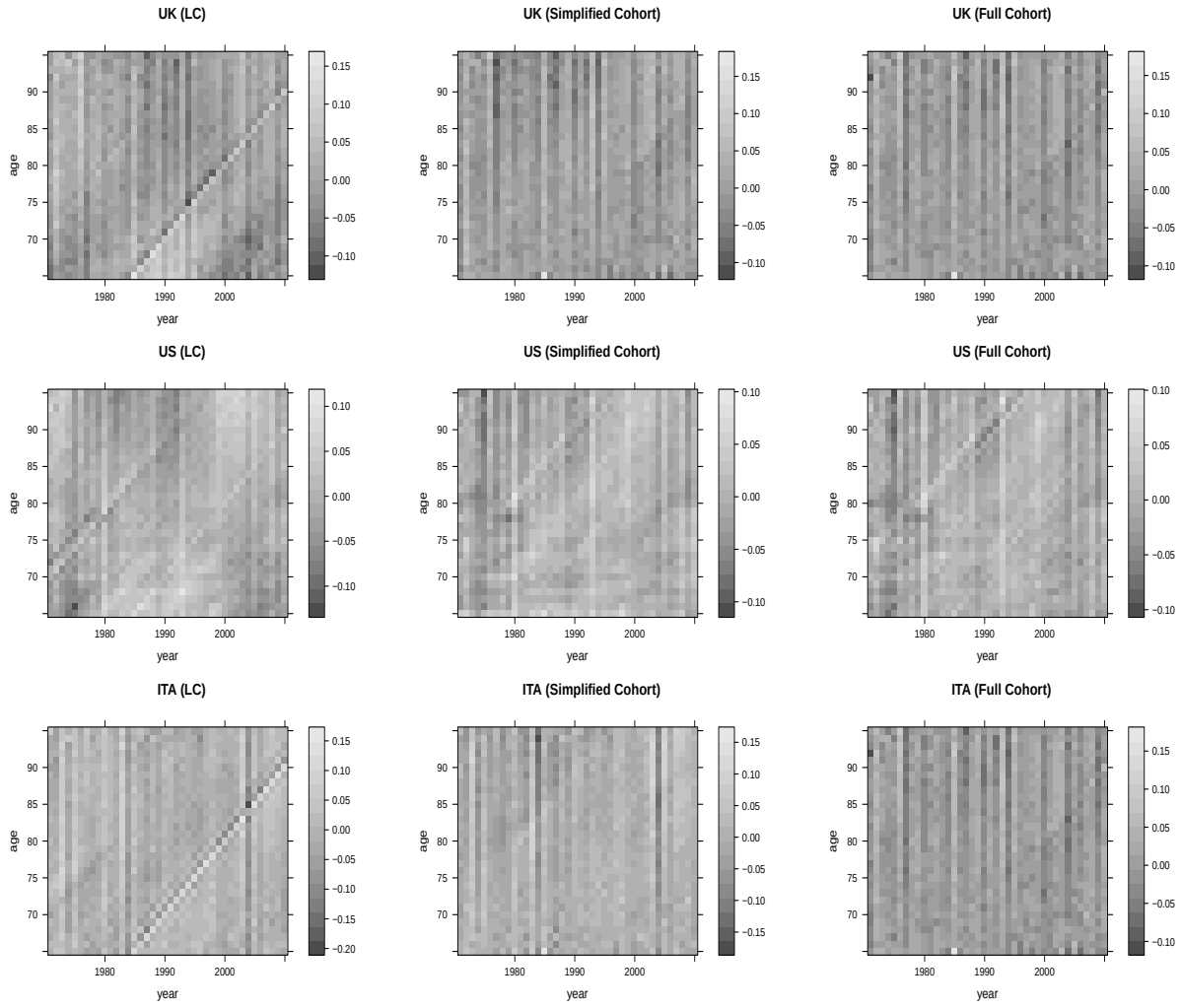


Figure 6: Residual heatmap produced from the LC model, the simplified cohort model and the full cohort model for the UK, US and Italy female populations.

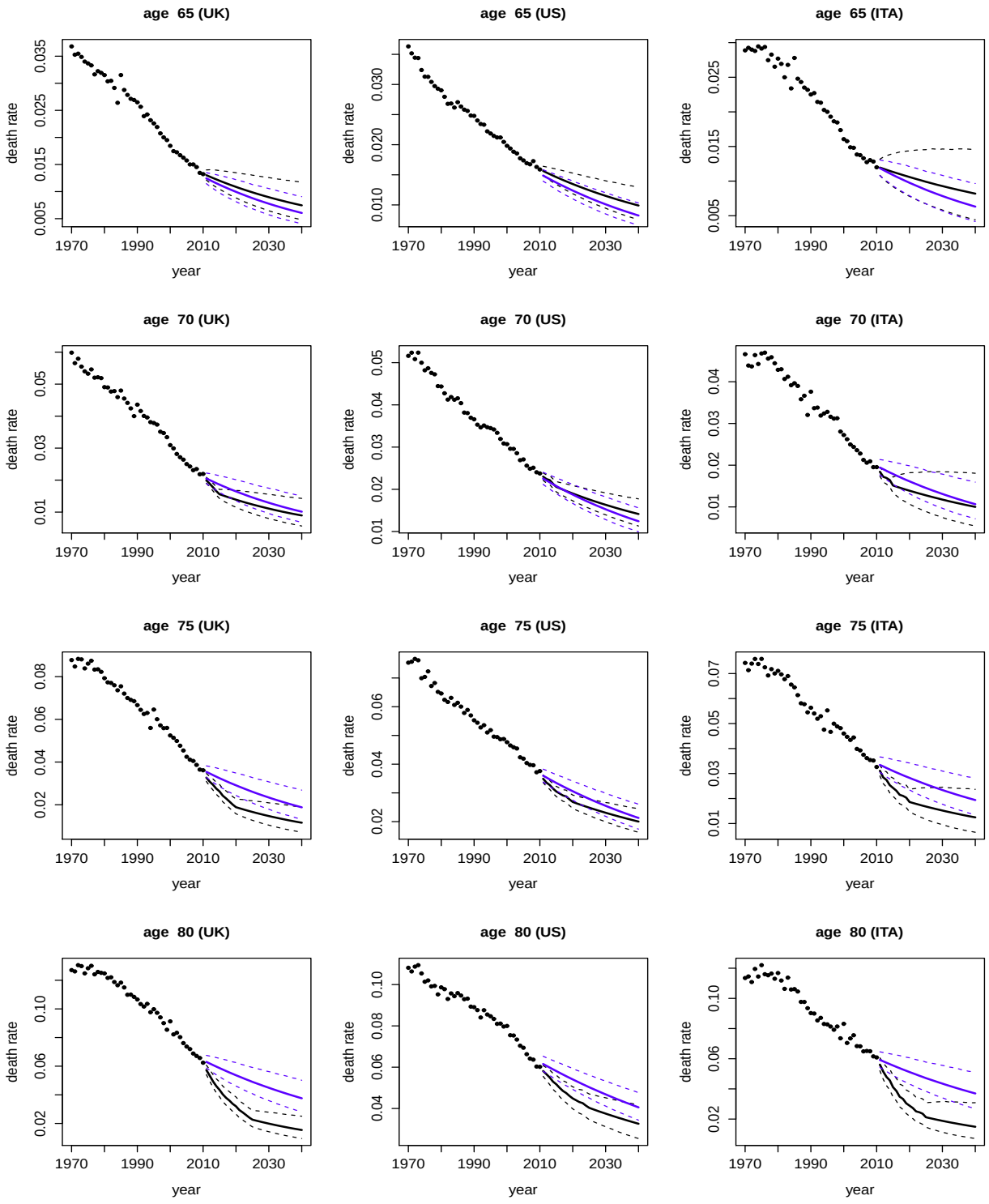


Figure 7: Forecast death rates for the UK, US and Italy male populations from the full cohort model (black lines) and the LC model (blue lines). Solid circles: observed data; solid lines: posterior mean; dash lines: 95% forecasting interval.

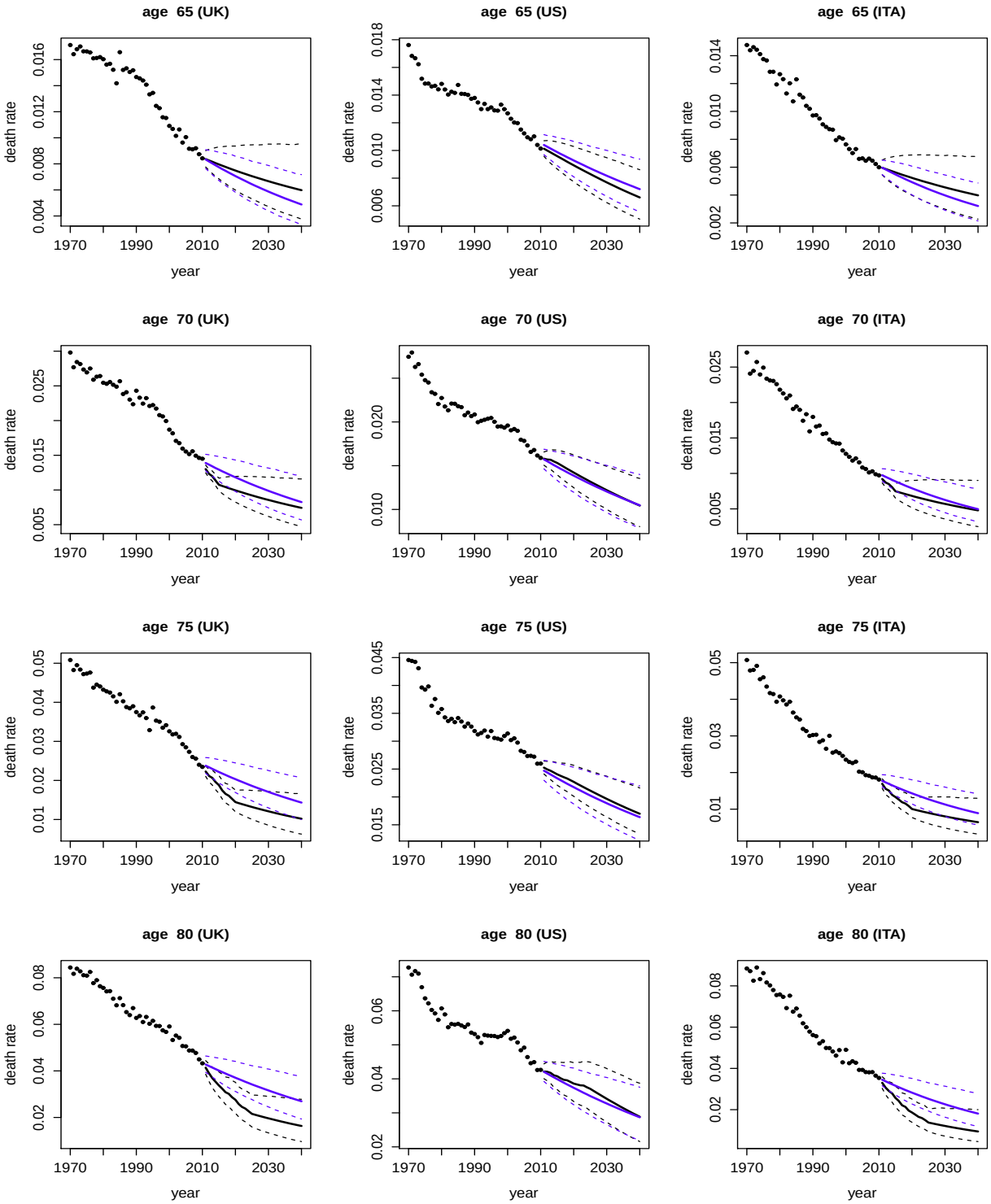


Figure 8: Forecast death rates for the UK, US and Italy female populations from the full cohort model (black lines) and the LC model (blue lines). Solid circles: observed data; solid lines: posterior mean; dash lines: 95% forecasting interval.

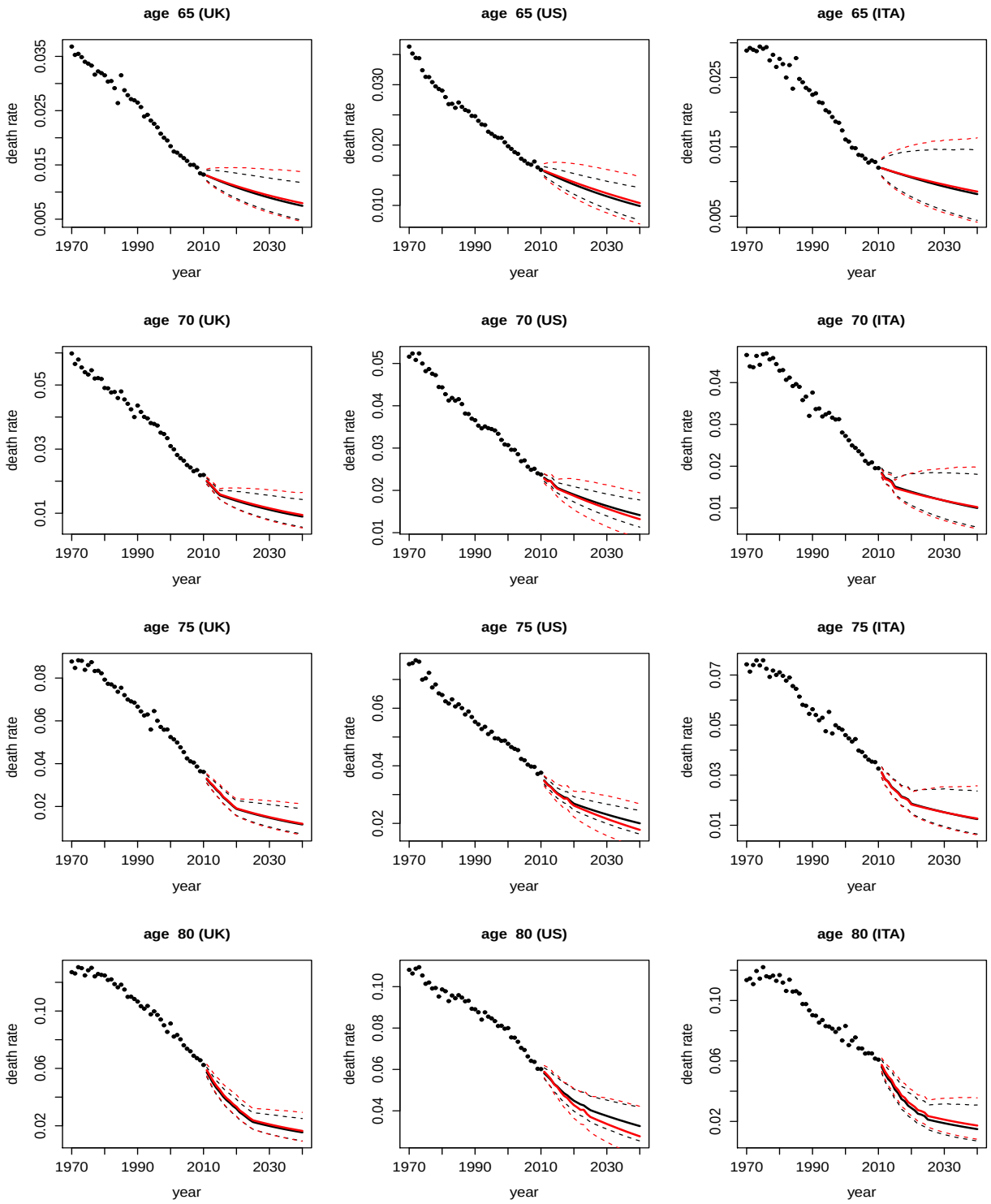


Figure 9: Forecast death rates for the UK, US and Italy male populations from the full cohort model (black lines) and the simplified cohort model (red lines). Solid circles: observed data; solid lines: posterior mean; dash lines: 95% forecasting interval.

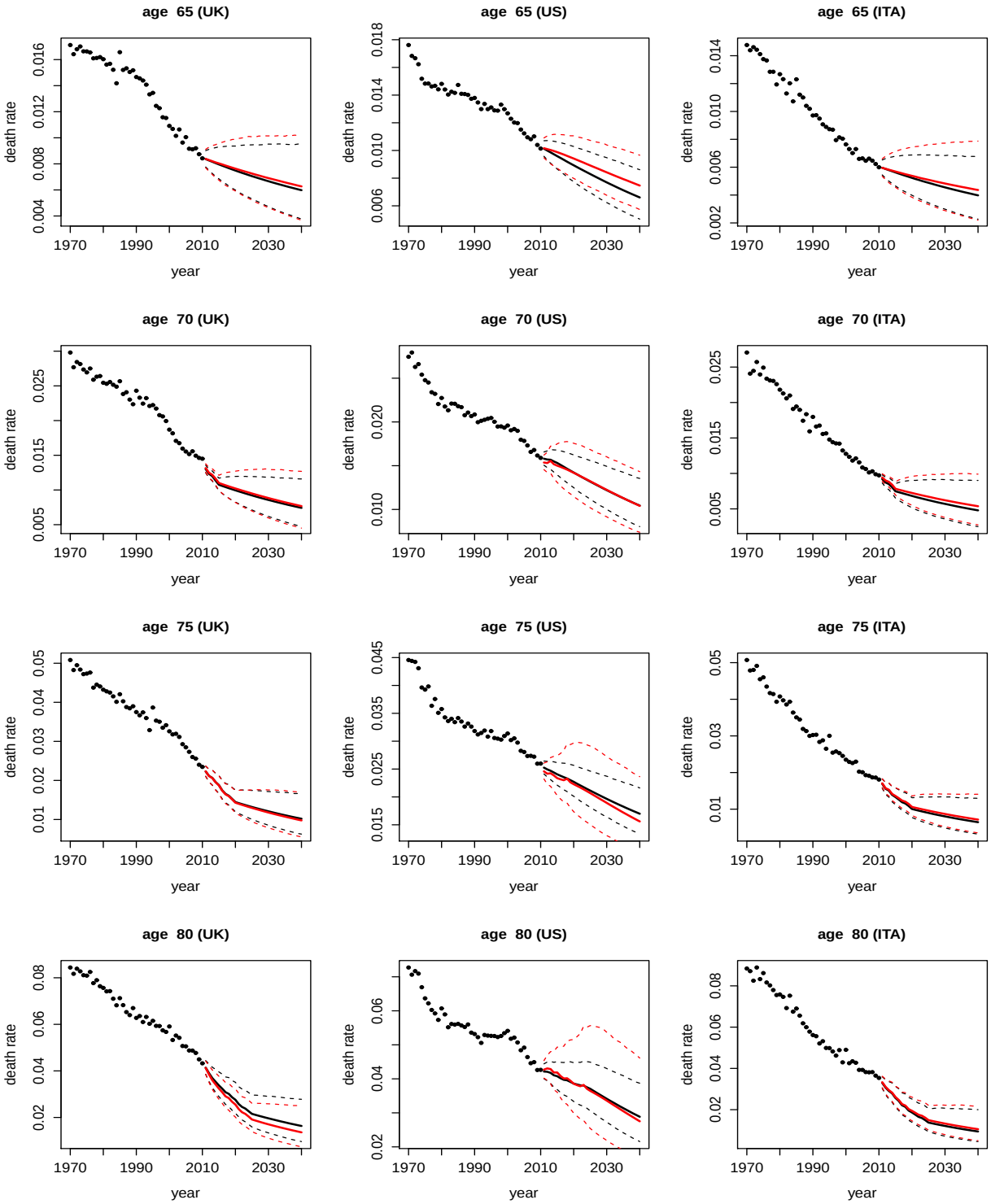


Figure 10: Forecast death rates for the UK, US and Italy female populations from the full cohort model (black lines) and the simplified cohort model (red lines). Solid circles: observed data; solid lines: posterior mean; dash lines: 95% forecasting interval.

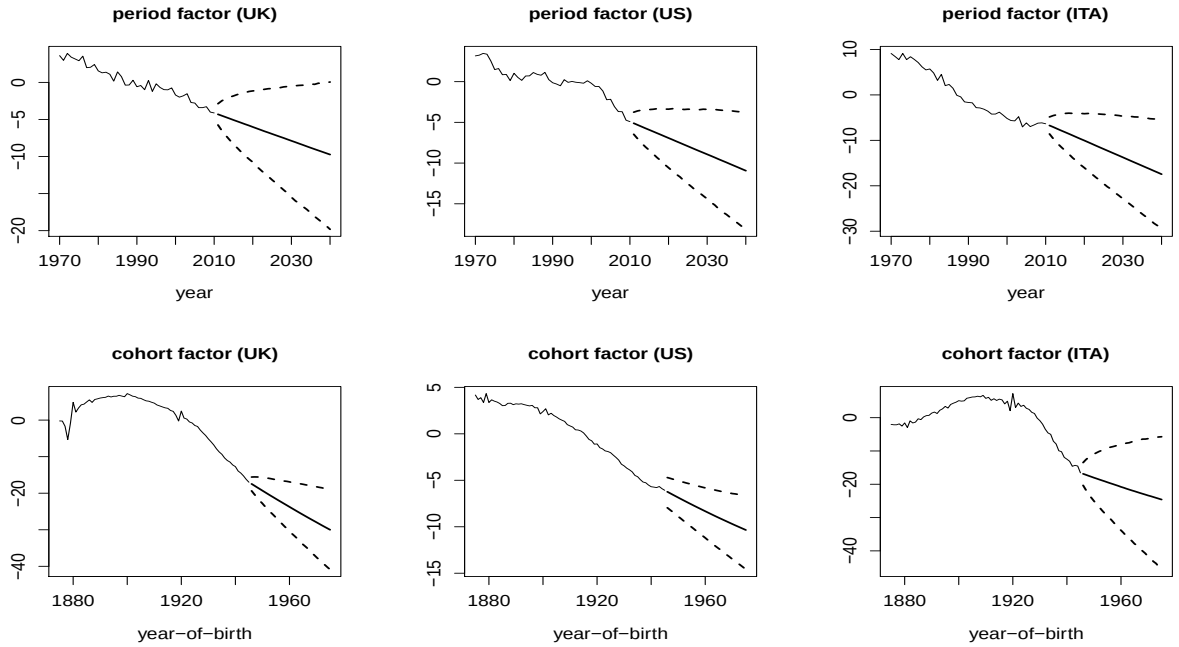


Figure 11: Projection of the period factor $\boldsymbol{\kappa}$ and cohort factor $\boldsymbol{\gamma}$ from the full cohort model for the UK, US and Italy male populations.

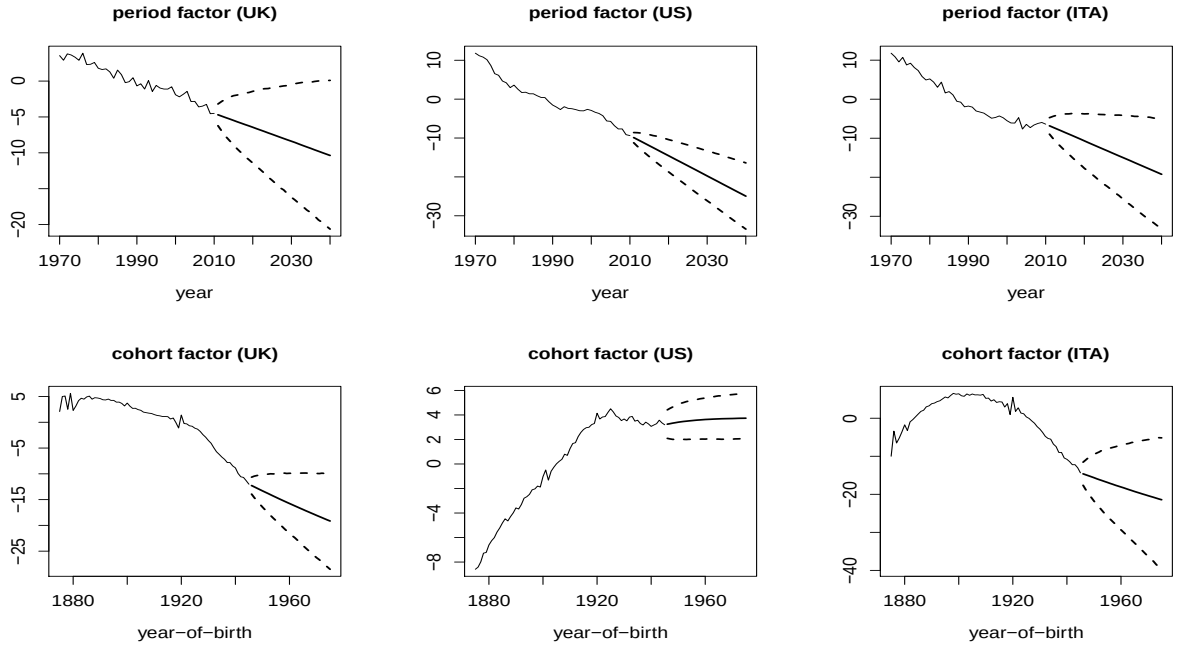


Figure 12: Projection of the period factor $\boldsymbol{\kappa}$ and cohort factor $\boldsymbol{\gamma}$ from the full cohort model for the UK, US and Italy female populations.

References

Andrieu, C., Doucet, A., Holenstein, R., 2010. Particle Markov Chain Monte Carlo Methods. Journal of Royal Statistical Society Series B 72, 269–342.

Barrieu, P., Bensusan, H., El Karoui, N., Hillairet, C., Loisel, S., Ravanelli, C., Salhi, Y., 2012. Understanding, modelling and managing longevity risk: key issues and main challenges. Scandinavian Actuarial Journal 3, 203–231.

Biffis, E., 2005. Affine processes for dynamic mortality and actuarial valuations. Insurance: Mathematics and Economics 37(3), 443–468.

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

Cairns, A., 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.

Cairns, A., Blake, D., Dowd, K., 2008. Modelling and management of mortality risk: A review. Scandinavian Actuarial Journal 2(3), 79–113.

Cairns, A., Blake, D., Dowd, K., Coughlan, G., 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.

Cairns, A., Blake, D., Dowd, K., Kessler, A. R., 2016. Phantoms never dies: living with unreliable population data. Journal of the Royal Statistical Society: Series A 179(4), 975–1005.

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

Celeux, C., Forbes, F., Robert, C. P., Titterington, D. M., 2006. Deviance information criteria for missing data models. Bayesian Analysis 1 (4), 651–674.

Chan, J. C. C., Grant, A. L., 2016. On the observed-data deviance information criterion for volatility modelling. Journal of Financial Econometrics 14(4), 772–802.

Currie, I. D., 2009. Smoothing and forecasting mortality rates with P-splines. http://www.ma.hw.ac.uk/ iain/research.talks.html.

Currie, I. D., 2016. On fitting generalized linear and non-linear models of mortality. Scandinavian Actuarial Journal 2016 (4), 356–383.

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

Dahl, M., Moller, T., 2006. Valuation and hedging of life insurance liabilities with systematic mortality risk. Insurance: Mathematics and Economics 39, 193–217.

De Jong, P., Tickle, L., 2006. Extending the Lee-Carter mortality forecasting. Mathematical Population Studies 13, 1–18.

Del Moral, P., Peters, G. W., C., V., 2013. An introduction to stochastic particle integration methods: With applications to risk and insurance. In: Dick J., Kuo, F., Peters, G., Sloan I. (eds) Monte Carlo and Quasi-Monte Carlo Methods 2012. Springer Proceedings in Mathematics & Statistics, vol 65. Springer, Berlin, Heidelberg. Available at SSRN: 2980466.

Delwarde, A., Denuit, M., Eilers, P. H. C., 2007. Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting: a penalized likelihood approach. Statistical Modelling 7, 29–48.

Doucet, A., Pitt, M. K., Deligiannidis, G., Kohn, R., 2015. Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. Biometrika 102 (2), 295–313.

Dowd, K., Cairns, A., Blake, D., Coughlan, G., Epstein, D., Khalaf-Allah, M., 2010. Evaluating the goodness of fit of stochastic mortality models. Insurance: Mathematics and Economics 47, 255–265.

Enchev, V., Kleinow, T., Cairns, A. J. G., 2016. Multi-population mortality models: fitting, forecasting and comparisons. Scandinavian Actuarial Journal.

Fung, M. C., Ignatieva, K., Sherris, M., 2014. Systematic mortality risk: An analysis of guaranteed lifetime withdrawal benefits in variable annuities. Insurance: Mathematics and Economics 58, 103–115.

Fung, M. C., Peters, G. W., Shevchenko, P. V., 2015. A state-space estimation of the Lee-Carter mortality model and implications for annuity pricing. In T. Weber, M. J. McPhee, and R. S. Anderssen (Eds.), MODSIM2015, 21st International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, pp. 952–958, www.mssanz.org.au/modsim2015/E1/fung.pdf.

Fung, M. C., Peters, G. W., Shevchenko, P. V., 2017. A unified approach to mortality modelling using state-space framework: characterisation, identification, estimation and forecasting. Annals of Actuarial Science 11(2), 343–389, https://doi.org/10.1017/S1748499517000069.

Gourieroux, C., Monfort, A., 1997. Time Series and Dynamic Models. Cambridge University Press.

Haberman, S., Renshaw, A., 2011. A comparative study of parametric mortality projection models. Insurance: Mathematics and Economics 48, 35–55.

Harvey, A. C., 1989. Forecasting: Structural Time Series Models and the Kalman Filter. Cambridge University Press.

Hirz, J., Schmock, U., Shevchenko, P. V., 2017a. Actuarial applications and estimation of extended CreditRisk$^+$. Risks 5 (2), 23:1–23:29.

Hirz, J., Schmock, U., Shevchenko, P. V., 2017b. Crunching mortality and life insurance portfolios with extended CreditRisk$^+$. Risk Magazine, January, 98–103.

Hunt, A., Villegas, A. M., 2015. Robustness and convergence in the Lee-Carter model with cohort effects. Insurance: Mathematics and Economics 64, 186–202.

Kleinow, T., Richards, S. J., 2016. Parameter risk in time-series mortality forecasts. Scandinavian Actuarial Journal. http://dx.doi.org/10.1080/03461238.2016.1255655.

Kogure, A., Kitsukawa, K., Kurachi, Y., 2009. A Bayesian comparison of models for changing mortalities toward evaluating longevity risk in japan. Asia-Pacific Journal of Risk and Insurance 3(2), 1–21.

Kogure, A., Kurachi, Y., 2010. A Bayesian approach to pricing longevity risk based on risk-neutral predictive distributions. Insurance: Mathematics and Economics 46, 162–172.

Koissi, M., Shapiro, A. F., Hognas, 2006. Evaluating and extending Lee-Carter model for mortality forecasting: Bootstrap confidence interval. Insurance: Mathematics and Economics 38, 1–20.

Lee, R. D., Carter, L. R., 1992. Modeling and forecasting U.S. mortality. Journal of the American Statistical Association 87, 659–675.

Leng, X., Peng, L., 2016. Inference pitfalls in Lee-Carter model for forecasting mortality. Insurance: Mathematics and Economics 70, 58–65.

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

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

Liu, Y., Li, J. S.-H., 2016b. The locally linear Cairns-Blake-Dowd model: A note on delta-nuga hedging of longevity risk. ASTIN Bulletin, 1–73.

Luciano, E., Vigna, E., 2008. Mortality risk via affine stochastic intensities: Calibration and empirical relevance. Belgian Actuarial Bulletin 8(1), 5–16.

Murphy, M., 2009. The "golden generations" in historical context. British Actuarial Journal 15, 151–184.

Murphy, M., 2010. Reexamining the dominance of birth cohort effects on mortality. Population and Development Review 36(2), 365–390.

O'Hare, C., Li, Y., 2012. Explaining young mortality. Insurance, Mathematics and Economics 50, 12–25.

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

Peters, G. W., Targino, R. S., Wuthrich, M. V., 2017. Bayesian modelling, Monte Carlo sampling and capital allocation of insurance risks. Available at SSRN: 2961888.

Petris, G., Petrone, S., Campagnoli, P., 2009. Dynamic Linear Models with R. Springer.

Plat, R., 2009. On stochastic mortality modeling. Insurance: Mathematics and Economics 45 (3), 393–404.

Renshaw, A., Haberman, S., 2003. Lee-Carter mortality forecasting with age-specific enhancement. Insurance: Mathematics and Economics 33, 255–272.

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

Shang, H. L., Booth, H., Hyndman, R. J., 2011. Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods. Demography 25 (5), 173–214.

Shevchenko, P. V., Hirz, J., Schmock, U., 2015. Forecasting leading death causes in australia using extended CreditRisk$^+$. In T. Weber, M. J. McPhee, and R. S. Anderssen (Eds.), MODSIM2015, 21st International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, pp. 966–972, www.mssanz.org.au/modsim2015/E1/shevchenko.pdf.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B 64, 583–639.

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

van Dyk, D. A., Park, T., 2008. Partially collapsed Gibbs samplers: theory and methods. Journal of the American Statistical Association 103(482), 790–796.

Villegas, A. M., Millossovich, P., Kaishev, V., 2015. StMoMo: An R package for stochastic mortality modelling. Available at SSRN: 2698729.

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

Yang, S. S., Yue, J. C., Huang, H., 2010. Modeling longevity risks using a principal component approach: A comparison with existing stochastic mortality models. Insurance: Mathematics and Economics 46(1), 254–270.