

North American Actuarial Journal

ISSN: 1092-0277 (Print) 2325-0453 (Online) Journal homepage: www.tandfonline.com/journals/uaaj20

Age-Dependent Multi-Cohort Affine Mortality Models with Cohort Correlation

Yuxin Zhou, Len Patrick Garces, Yang Shen, Michael Sherris & Jonathan Ziveyi

To cite this article: Yuxin Zhou, Len Patrick Garces, Yang Shen, Michael Sherris & Jonathan Ziveyi (07 Jul 2026): Age-Dependent Multi-Cohort Affine Mortality Models with Cohort Correlation, North American Actuarial Journal, DOI: 10.1080/10920277.2026.2682217

To link to this article: https://doi.org/10.1080/10920277.2026.2682217

© 2026 The Author(s). Published with license by Taylor & Francis Group, LLC

Published online: 07 Jul 2026.

Submit your article to this journal

Article views: 384

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=uaaj20

North American Actuarial Journal, 0(0), 1–31, 2026
© 2026 The Author(s). Published with license by Taylor & Francis Group, LLC
ISSN: 1092-0277 print / 2325-0453 online
DOI: 10.1080/10920277.2026.2682217

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Age-Dependent Multi-Cohort Affine Mortality Models with Cohort Correlation

Yuxin Zhou,[1,2] Len Patrick Garces,[2,3]* Yang Shen,[1,2] Michael Sherris,[1,2] and Jonathan Ziveyi[1,2]

[1]*School of Risk and Actuarial Studies, University of New South Wales, Kensington, New South Wales, Australia*
[2]*ARC Centre of Excellence in Population Ageing Research, University of New South Wales, Kensington, New South Wales, Australia*
[3]*School of Mathematical and Physical Sciences, University of Technology Sydney, Ultimo, New South Wales, Australia*

Continuous-time affine mortality models are useful in the analysis of age-cohort mortality rates as they yield a closed-form expression for survival curves, which are consistent with the dynamics of latent factors driving mortality and are well-suited for finance and insurance applications. We extend and improve these mortality models by introducing age dependence of mortality rates and correlation between cohorts. We propose and compare two classes of age-dependent mortality models, namely the age-dependent coefficient model and the age-dependent factor model. Specifically, we assess both Gaussian and Cox–Ingersoll–Ross (CIR)–type models for each category of age-dependent models. Both categories of age-dependent models involve age and calendar time, which in turn specify the cohort. Thus, our models admit an analytical form for the instantaneous correlation between mortality rates of different cohorts. Moreover, we propose two improvements to the parameter estimation process. First, to improve the estimation of cohort correlations, we regularize the parameter estimation by adding a penalty term which penalizes larger differences between empirical and estimated correlations. Second, we develop and assess a method to include incomplete cohorts into the Kalman filtering algorithm for parameter estimation. By incorporating age dependence and using incomplete cohort mortality data, we improve the goodness of fit and produce more reasonable out-of-sample forecasts of survival probabilities. We also show that regularization produces more realistic correlations between cohorts for varying age differences. Our results show that, under most circumstances, the correlation between cohorts decreases as the age difference increases.

1. INTRODUCTION

Mortality models are essential in the pricing and risk management of mortality-linked products. The literature of continuous-time affine mortality models has been growing since the early work of Dahl (2004), Biffis (2005), and Luciano and Vigna (2008) who use the one-factor affine processes to model the mortality rates of a fixed cohort of individuals. The continuous-time models have the advantage of being able to model the trend and volatility in the mortality rates by representing the evolution of mortality rates through a stochastic differential equation (SDE). Moreover, survival probabilities under the affine framework have a closed-form expression. Further developments allow the continuous-time mortality models to describe the entire mortality surface of multiple cohorts, thereby capturing mortality improvements. Schrager (2006) proposes a general framework of multi-factor affine mortality models, which captures the dynamics of the mortality surface using age-period data. This is supported by the principal component analysis (PCA) in Njenga and Sherris (2011) who find that multiple factors

Address correspondence to Yuxin Zhou, School of Risk and Actuarial Studies, University of New South Wales, Kensington, NSW, Australia. E-mail: yuxin.zhou@unsw.edu.au

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

*Current affiliation: School of Risk and Actuarial Studies, University of New South Wales, Kensington, New South Wales, Australia.

are needed to fully explain the evolution of mortality rates. Following Schrager (2006), Njenga and Sherris (2011), and Blackburn and Sherris (2013) develop a three-factor model for survival probabilities defined under the risk-neutral measure.

However, there are still issues that remain to be addressed, the first being high mean squared error (MSE) observed at old ages in the fitted mortality models. Blackburn and Sherris (2013) compare the two-factor and three-factor mortality models and find that adding a third factor better captures mortality trends at older ages. Another possible reason for the high MSE at old ages is that the volatility structure in the mortality rates has not been modeled carefully. Age-dependent trend and volatility in the mortality data have been observed in the literature, which motivates this research. Both Piggott et al. (2005) and Chang and Sherris (2018) find that the volatility of mortality rates increases with age.

Another issue to be addressed is the more accurate estimation of cohort correlations. Previous works on continuous-time affine mortality models, for example Blackburn and Sherris (2013) and Huang et al. (2022), assume that the factors equally affect all ages or, equivalently, all cohorts at a given time. This results in a perfect instantaneous correlation in the mortality rates of different cohorts, which does not hold in view of historical data. To address the perfect cohort correlation issue, scholars have proposed models that use age-cohort data or models where each cohort has its own cohort-specific factor to allow imperfect cohort correlation. For example, Jevtić et al. (2013) first introduce cohort-specific factors and calibrate a two-factor model. This has been extended in Xu et al. (2020) whose model has two-common factors and one cohort-specific factor. Under these assumptions, the correlations of the instantaneous mortality rates between different cohorts are available in analytical form. However, this type of model has one limitation, that is, the cohort-specific factors are estimated as free parameters separately from the estimation of the common factors. This implies that cohort-specific factors for out-of-sample cohorts that are not included in the estimation cannot be determined using data on preceding cohorts. Therefore, the models using cohort-specific factors have a limited ability to forecast future survival probabilities.

To address these issues, age-dependent affine mortality models have been proposed in the literature, which are naturally specific to certain cohorts. Age-dependent models can capture the imperfect instantaneous cohort correlation while maintaining the ability to consistently predict mortality outcomes of further cohorts. Within the affine model framework, age-dependence can be incorporated via age-dependent factors or coefficients. Dahl (2004) introduces a framework for one-factor age-dependent short-rate models and discusses in detail the implications of omitting age dependence from the drift and volatility specifications in subsequent applications. Extending this framework, Cairns et al. (2006) propose a generalized n-factor age-dependent short-rate model, although no explicit functional form is specified for the age-dependent drift and volatility components. Subsequent studies have begun to specify these structures more concretely. Building on Renshaw and Haberman (2006), Cairns et al. (2009), and Wills and Sherris (2008) propose a one-factor mortality model in which the drift term is a linear function of the current age, and the percentage volatility term is a constant. Chang and Sherris (2018) find that the drift and volatility of the difference in mortality intensity both increase with age. They propose a two-factor model with one cohort-specific factor, where both the drift and volatility terms are increasing functions of age. Chang and Sherris (2018) find that including age dependence improves the goodness of fit at old ages compared to Jevtić et al. (2013) who use constant drift and volatility terms. However, Wills and Sherris (2008) and Chang and Sherris (2018) only use one and two factors respectively with Chang and Sherris (2018) having the issue of using cohort-specific factors discussed above, which limits their forecasting ability. The models used in this article avoid this issue by introducing dependence on both age and calendar time at the same time.

Meanwhile, age-dependence can be incorporated via factor coefficients. Jevtić and Regis (2019) propose a general form of the multi-population model in which the coefficients are functions of time, initial age, and age increasing with time, which makes the mortality intensity naturally specific to a cohort. The age-dependent and thus cohort-specific models allow a closed-form expression of cohort correlation, which motivates us to study age-dependent coefficient (ADC) and also age-dependent factor (ADF) models. In this article, we consider an alternative approach of parameter estimation where we incorporate empirical cohort correlation with the goal of improving the accuracy of cohort correlation estimations.

Recent advances in the literature have enriched mortality modeling by introducing more flexible dynamics. For example, Zhu and Bauer (2022) model both spot and forward mortality rates within an affine framework, showing that forward mortality rates are Markovian while spot rates are not. To capture long-term persistence in mortality improvements, Zhou et al. (2022) introduce long-range dependence via fractional Brownian motion (fBM), and Chiu et al. (2026) further extend this framework by incorporating cointegration across multiple populations. These contributions primarily enhance the temporal dependence structure of mortality dynamics. Meanwhile, the focus of the present article is to address cross-cohort correlation. The proposed ADF models retain a Markovian affine structure, which preserves analytical tractability and enables estimation via Kalman filtering in the presence of incomplete cohort data. Extensions that combine age-dependent factor dynamics with richer temporal dependence structures, such as long-range dependence, represent a natural direction for future research.

Moreover, most of the existing models in the literature that use age-cohort data only use complete-cohort data to calibrate the model. A complete cohort refers to a cohort for which we have a full observation of the age range we are modeling (that is,

50 to 109). Therefore, individuals in these complete cohorts are over age 109, which means most of them have already passed away. Meanwhile, for younger cohorts still alive, we have only incomplete observations. Continuous-time mortality models in the literature often disregard these cohorts in the estimation procedure. However, these incomplete-cohort data contain more recent information on mortality evolution. Thus, utilizing incomplete cohort data leads to more accurate out-of-sample forecasts.

This article contributes to the literature by developing age-dependent and cohort-specific affine mortality models that address the key issues identified above. First, our models incorporate both age and calendar time dependence to improve fit and forecasting performance, especially at older ages. Second, we propose a new regularization method within the Kalman filtering framework to improve estimation of cohort correlations. By including a penalty term for deviations from the empirical cohort correlation matrix in the objective function, we enhance the ability of the model to estimate cohort correlations. Third, we develop a modified Kalman filtering algorithm to incorporate incomplete-cohort data into the estimation process. Unlike previous studies that rely solely on complete cohorts, which are typically from older generations, our method enables the use of more recent, partially observed cohort data, leading to improved forecasting for younger and future cohorts.

This article is structured as follows. Section 2 presents the framework of the proposed age-dependent affine mortality models. Section 3 summarizes the construction of incomplete age-cohort mortality data. Section 4 explains the estimation technique, including how we incorporate incomplete cohort data and add regularization. Section 5 discusses the goodness of fit and forecasting performance of the proposed mortality models. Section 6 analyses the estimated cohort correlation after incorporating regularization. Section 7 concludes the article.

2. AGE-DEPENDENT AFFINE MORTALITY MODELS

We study two categories of age-dependent affine mortality models. The first category of models assumes that the coefficients of the factors are age-dependent, namely the ADC models. The other category of models assumes that the factors themselves are age-dependent by allowing age-dependent drift and volatility terms in the SDE, thus called the ADF models.

Consider a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, Q)$, where Ω is the sample space, $\mathcal{F}$ is a σ-algebra on Ω, Q is a probability measure on $(\Omega, \mathcal{F})$ interpreted as the risk-neutral measure, and $\{\mathcal{F}_t\}_{t\geq 0} := \mathbb{F}$ is a right-continuous and Q-complete filtration.

We adopt a three-factor specification, as it has been shown to deliver strong empirical performance while maintaining manageable model complexity (Blackburn and Sherris 2013; Xu et al. 2020; Huang et al. 2022; Ungolo et al. 2024). Suppose that W_1^Q, W_2^Q, and W_3^Q are three independent standard Brownian motions defined on the probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, Q)$. Furthermore, we assume that the filtration $\{\mathcal{F}_t\}_{t\geq 0}$ is the natural filtration generated by the Brownian motions:

$$\mathcal{F}_t := \sigma\left(\left(W_1^Q(s), W_2^Q(s), W_3^Q(s)\right)^\top, 0 \leq s \leq t\right), \quad t \geq 0.$$

We first develop the affine mortality model under the Q measure and then change to the real-world measure P in Subsection 2.3 for parameter estimation. Let $\mathbf{Z}(t,x) = (Z_1(t,x), Z_2(t,x), Z_3(t,x))^\top$ be a 3×1 vector of the factors $Z_i(t,x), i = 1,2,3$ dependent on age x, which follows the SDE below:

$$\mathrm{d}\mathbf{Z}(t,x) = \Lambda^Q(x, \mathbf{Z}(t,x))\mathrm{d}t + \Sigma(x, \mathbf{Z}(t,x))\mathrm{d}W^Q(t), \tag{1}$$

where the drift $\Lambda^Q(\cdot) \in \mathbb{R}^{3\times 1}$ and the volatility $\Sigma(\cdot) \in \mathbb{R}^{3\times 3}$ are functions that depend on time t, age x, and $W^Q(t) = (W_1^Q(t), W_2^Q(t), W_3^Q(t))^\top$ is a 3×1 vector of $W_i^Q(t)$ for $i = 1,2,3$.

The general frameworks of the two categories of three-factor age-dependent models are described in terms of the force of mortality as follows.

Definition 1. *For ADC models, the force of mortality for a cohort born in year c and aged $x = t - c$ at time t is defined as $\mu_x(t) \equiv \mu_{t-c}(t)$:*

$$\mu_x(t) = \sum_{i=1}^{3} g_i(t-c) Z_i(t), \tag{2}$$

where the coefficients $g_i(t-c)$ depend on the age $x = t - c$ which increases with time t for a given cohort born in year c, the factors $Z_i(t)$ depend on time t, and $i = 1,2,3$ represent the three factors. The factors $Z_i(t)$ follow a simplified dynamics in which the latent factors do not depend on age:

$$d\mathbf{Z}(t) = \Lambda^{Q}(\mathbf{Z}(t))dt + \Sigma(\mathbf{Z}(t))dW^{Q}(t), \tag{3}$$

which corresponds to a special case of Equation (1) with age-independent drift and volatility.

Definition 2. *For ADF models, the force of mortality $\mu_x(t)$ of cohort aged x at time t is defined as*

$$\mu_x(t) = \sum_{i=1}^{3} Z_i(t,x), \tag{4}$$

where the factors $Z_i(t,x)$ depend on the age x through the age-dependent drift and volatility terms in the SDE, and $i = 1,2,3$ represent the three factors.

The formulation of the ADF model is motivated by the literature on age-dependent drift and volatility and by empirical evidence that mortality dynamics vary across age, particularly at advanced ages where both the drift and volatility of mortality improvements differ markedly (Dahl 2004; Cairns et al. 2006; Wills and Sherris 2008; Chang and Sherris 2018).

Unlike ADC models, when specifying the ADF models, we separate the data into different age groups. The factors are then made dependent on age via a representative age in each given age group. The specifications will be explained in the following sections.

Following the standard results of affine models (Duffie et al. 2000), the survival probability $S_x(t,T)$ for the cohort born in year c and aged $x = t - c$ at time t to survive to time T is represented as

$$\begin{aligned} S_x(t,T) &= E^{Q}\left[e^{-\int_t^T \mu_{s-c}(s)ds} | \mathcal{F}_t\right] \\ &= e^{B_1(t,T,x)Z_1(t,x)+B_2(t,T,x)Z_2(t,x)+B_3(t,T,x)Z_3(t,x)+A(t,T,x)}, \end{aligned} \tag{5}$$

where $Z_i(t,x)$ and $B_i(t,T,x)$, $i = 1,2,3$ are the three factors and their factor loadings, respectively, and $A(t,T,x)$ is an adjustment term. The factor loadings $B_i(t,T,x)$ and adjustment term $A(t,T,x)$ are obtained as solutions to a system of ordinary differential equations (ODEs) that depend on the specification of Equation (1) and Equation (2) or (4).

2.1. ADC Models

The force of mortality $\mu_x(t)$ of a cohort born in year c and aged $x = t - c$ at time t is decomposed as shown in Equation (2). The construction of the coefficient functions $g_i(t-c)$ follows one example in Schrager (2006) as

$$\begin{cases} g_1(t-c) = e^{-a_1^g(t-c)}, \\ g_2(t-c) = e^{-a_2^g(t-c-b_2)^2}, \\ g_3(t-c) = e^{a_3^g(t-c)}, \end{cases}$$

where a_1^g, a_2^g, a_3^g, and b_2^g are positive constants. Thus, each factor affects different ages differently since $g_1(t-c)$ decreases with age $(t-c)$, $g_2(t-c)$ is a concave function of age $(t-c)$, and $g_3(t-c)$ increases with age $(t-c)$. Therefore, the coefficient $g_1(t-c)$ targets the younger ages within the age range, $g_2(t-c)$ targets the middle ages, and $g_3(t-c)$ targets the older ages. For the factors $Z_i(t)$, we allow them to be either a Gaussian process or a CIR process. The ADC Gaussian and CIR models fall within the framework in Jevtić and Regis (2019). However, Jevtić and Regis (2019) consider only the Gaussian specification, whereas we examine both Gaussian and Cox–Ingersoll–Ross (CIR) models in this article, which allows a comparison of distributional assumptions within a common structural framework.

2.1.1. ADC Gaussian Model

In the ADC Gaussian model, we assume the factors $Z_i(t)$ follow a Gaussian process. We let $\mathbf{Z}(t)$ be 3×1 vector of the three factors $Z_i(t)$, and a general form of the SDE is first defined under the risk-neutral measure Q:

$$
\begin{aligned}
\mathrm{d}\mathbf{Z}(t) &= K^{Q}(\Theta^{Q} - \mathbf{Z}(t))\mathrm{d}t + \Sigma \mathrm{d}W^{Q}(t) \\
&= \begin{pmatrix} k_1^Q & 0 & 0 \\ 0 & k_2^Q & 0 \\ 0 & 0 & k_3^Q \end{pmatrix} \left[\begin{pmatrix} \theta_1^Q \\ \theta_2^Q \\ \theta_3^Q \end{pmatrix} - \begin{pmatrix} Z_1(t) \\ Z_2(t) \\ Z_3(t) \end{pmatrix} \right] \mathrm{d}t + \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \begin{pmatrix} \mathrm{d}W_1^Q(t) \\ \mathrm{d}W_2^Q(t) \\ \mathrm{d}W_3^Q(t) \end{pmatrix},
\end{aligned}
$$

where for $i = 1, 2, 3$, k_i^{Q} are the rate of mean-reversion parameters, θ_i^{Q} are the long-term means, σ_i are the volatility parameters, and $W_i^{\mathrm{Q}}(t)$ are standard Brownian motions under the risk-neutral measure Q.

Using the Feynman-Kac formula on Equation (5) with S as the short-hand notation for $S_{t-c}(t, T)$ yields

$$
\frac{\partial S}{\partial t} + \sum_{i=1}^{3} \frac{\partial S}{\partial Z_i} \left(k_i^{\mathrm{Q}}(\theta_i^{\mathrm{Q}} - Z_i) \right) + \frac{1}{2} \sum_{i=1}^{3} \sigma_i^2 \frac{\partial^2 S}{\partial Z_i^2} - \mu_{t-c}(t) S = 0. \tag{6}
$$

Meanwhile, from Equation (5) we also have

$$
\frac{\partial S}{\partial t} = \left[\frac{\partial A}{\partial t} + \sum_{i=1}^{3} \frac{\partial B_i}{\partial t} Z_i \right] S, \tag{7}
$$

$$
\frac{\partial S}{\partial Z_i} = B_i S, \tag{8}
$$

$$
\frac{\partial^2 S}{\partial Z_i^2} = B_i^2 S. \tag{9}
$$

Substituting Equations (7)–(9) into Equation (6) yields

$$
\frac{\partial A}{\partial t} + \sum_{i=1}^{3} B_i k_i^{\mathrm{Q}} \theta_i^{\mathrm{Q}} + \frac{1}{2} \sum_{i=1}^{3} \sigma_i^2 B_i^2 + \sum_{i=1}^{3} \left[\frac{\partial B_i}{\partial t} - B_i k_i^{\mathrm{Q}} - g_i(t-c) \right] Z_i = 0. \tag{10}
$$

We match the coefficients of Z_i and the constant term to zero in Equation (10). Therefore, we derive the linear system of ODEs to solve for $A(t, T)$ and $B_i(t, T)$:

$$
\begin{aligned}
\frac{\partial A}{\partial t} &= -\sum_{i=1}^{3} B_i k_i^{\mathrm{Q}} \theta_i^{\mathrm{Q}} - \frac{1}{2} \sum_{i=1}^{3} \sigma_i^2 B_i^2, \\
\frac{\partial B_i}{\partial t} &= B_i k_i^{\mathrm{Q}} + g_i(t-c) \quad (\text{for } i = 1, 2, 3).
\end{aligned} \tag{11}
$$

The terminal conditions are $A(T, T) = 0$ and $B_i(T, T) = 0$ for all models in this article as in Christensen et al. (2011) and Blackburn and Sherris (2013). Due to the presence of the B_i^2 term and the $g_2(t-c)$ term, in general, an analytical solution to the above ODEs is not available. Therefore, we solve the system of ODEs numerically.[1]

[1]We select the "ode45()" function in MATLAB(R2021a) that enables us to obtain numerical solutions within a second.

2.1.2. ADC CIR Model

The SDE of $\mathbf{Z}(t)$ is

$$\begin{aligned} d\mathbf{Z}(t) &= \begin{pmatrix} k_1^Q & 0 & 0 \\ 0 & k_2^Q & 0 \\ 0 & 0 & k_3^Q \end{pmatrix} \left[\begin{pmatrix} \theta_1^Q \\ \theta_2^Q \\ \theta_3^Q \end{pmatrix} - \begin{pmatrix} Z_1(t) \\ Z_2(t) \\ Z_3(t) \end{pmatrix} \right] dt + \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \begin{pmatrix} \sqrt{Z_1(t)} & 0 & 0 \\ 0 & \sqrt{Z_2(t)} & 0 \\ 0 & 0 & \sqrt{Z_3(t)} \end{pmatrix} \begin{pmatrix} dW_1^Q(t) \\ dW_2^Q(t) \\ dW_3^Q(t) \end{pmatrix} \\ &= K^Q \left[\Theta^Q - \mathbf{Z}(t) \right] dt + \Sigma D(\mathbf{Z}(t)) dW^Q(t), \end{aligned}$$

where $D(\mathbf{Z}(t))$ is a 3×3 diagonal matrix with $\sqrt{Z_i(t)}$ on the i^{th} column and row.

Using the Feynman-Kac formula in Equation (5), following the derivation in Subsection 2.1.1, we obtain the following Riccati system of ODEs to solve for $A(t,T)$ and $B_i(t,T)$:

$$\begin{aligned} \frac{\partial A}{\partial t} &= -\sum_{i=1}^{3} B_i k_i^Q \theta_i^Q, \\ \frac{\partial B_i}{\partial t} &= -\frac{1}{2}\sigma_i^2 B_i^2 + B_i k_i^Q + g_i(t-c) \quad (\text{for i } = 1,\ 2,\ 3). \end{aligned} \tag{12}$$

For the same reason, an analytical solution is not available due to the existence of the B_i^2 term and the $g_2(t-c)$ term. Therefore, we also need to solve the system of ODEs numerically.

2.2. ADF Models

The force of mortality $\mu_x(t)$ of cohort aged x at time t is decomposed into the sum of age-dependent factors $Z_i(t,x)$ as shown in Equation (4). The vector $\mathbf{Z}(t,x)$ of age-dependent factors follow the SDE in Equation (1). We allow the factors $Z_i(t,x), i = 1,2,3$ to follow a Gaussian process or a CIR process.

2.2.1. ADF Gaussian Model

We extend the structure by allowing the drift and volatility terms to be age-dependent on the age index h_x. For ADF models, we separate the data into two age groups, and the age index h_x remains constant within one age group and is equal to the starting age of this age group. For example, if we split the data into two age groups $[50,79]$ and $[80,109]$, then $h_x = 50$ when $x \in [50,79]$, and $h_x = 80$ when $x \in [80,109]$. The drift $\Lambda^Q(\cdot)$ has a linear relationship with the age index h_x, and the volatility changes exponentially with the age index h_x and remains positive with the exponential function:

$$\begin{aligned} \Lambda^Q(x, \mathbf{Z}(t,x)) &= \begin{pmatrix} a_1^Q + b_1^Q h_x & 0 & 0 \\ 0 & a_2^Q + b_2^Q h_x & 0 \\ 0 & 0 & a_3^Q + b_3^Q h_x \end{pmatrix} \left[\begin{pmatrix} \theta_1^Q \\ \theta_2^Q \\ \theta_3^Q \end{pmatrix} - \begin{pmatrix} Z_1(t,x) \\ Z_2(t,x) \\ Z_3(t,x) \end{pmatrix} \right], \\ \Sigma(x, \mathbf{Z}(t,x)) &= \begin{pmatrix} e^{c_1 + d_1 h_x} & 0 & 0 \\ 0 & e^{c_2 + d_2 h_x} & 0 \\ 0 & 0 & e^{c_3 + d_3 h_x} \end{pmatrix}. \end{aligned}$$

For simplicity, we set

$$\begin{pmatrix} a_1^Q + b_1^Q h_x & 0 & 0 \\ 0 & a_2^Q + b_2^Q h_x & 0 \\ 0 & 0 & a_3^Q + b_3^Q h_x \end{pmatrix} = \begin{pmatrix} k_{1,x}^Q & 0 & 0 \\ 0 & k_{2,x}^Q & 0 \\ 0 & 0 & k_{3,x}^Q \end{pmatrix} = K_x^Q,$$

where $k_{i,x}^Q,\ i = 1,2,3$ represent the transition parameters. Since we keep the age within one age group to be a constant h_x, we can use the results for the factor loadings in Blackburn and Sherris (2013) and replace the constant parameters K^Q and σ with our age-dependent parameters:

$$B_i(t,T,x) = \frac{1 - e^{-k^Q_{i,x}(T-t)}}{-k^Q_{i,x}},$$
$$A(t,T,x) = \frac{1}{2}\sum_{i=1}^{3} \frac{\left(e^{c_i + d_i h_x}\right)^2}{\left(k^Q_{i,x}\right)^3}\left[\frac{1}{2}\left(1 - \mathrm{e}^{-2k^Q_{i,x}(T-t)}\right) - 2\left(1 - \mathrm{e}^{-k^Q_{i,x}(T-t)}\right) + k^Q_{i,x}(T-t)\right],$$

where $k^Q_{i,x} = a^{\mathrm{Q}}_i + b^{\mathrm{Q}}_i h_x$ for $i = 1, 2, 3$.

2.2.2. ADF CIR Model

In the case of the ADF CIR model, the drift and volatility terms are as follows:

$$\Lambda^Q(x, \mathbf{Z}(t,x)) = \begin{pmatrix} a^Q_1 + b^Q_1 h_x & 0 & 0 \\ 0 & a^Q_2 + b^Q_2 h_x & 0 \\ 0 & 0 & a^Q_3 + b^Q_3 h_x \end{pmatrix} \left[\begin{pmatrix} Z_1(t,x) \\ Z_2(t,x) \\ Z_3(t,x) \end{pmatrix} - \begin{pmatrix} \theta^Q_1 \\ \theta^Q_2 \\ \theta^Q_3 \end{pmatrix} \right],$$
$$\Sigma(x, \mathbf{Z}(t,x)) = \begin{pmatrix} e^{c_1 + d_1 h_x} & 0 & 0 \\ 0 & e^{c_2 + d_2 h_x} & 0 \\ 0 & 0 & e^{c_3 + d_3 h_x} \end{pmatrix} \begin{pmatrix} \sqrt{Z_1(t,x)} & 0 & 0 \\ 0 & \sqrt{Z_2(t,x)} & 0 \\ 0 & 0 & \sqrt{Z_3(t,x)} \end{pmatrix}.$$

Within each age group with the same age index h_x, we can also replace the constant parameters in Geyer and Pichler (1999); Chen and Scott (2003); and Huang et al. (2022) with our age-dependent parameters, which yields the factor loadings below:

$$B_i(t,T,x) = -\frac{2\left(e^{\gamma_{i,x}(T-t)} - 1\right)}{\left(k^Q_{i,x} + \gamma_{i,x}\right)\left(e^{\gamma_{i,x}(T-t)} - 1\right) + 2\gamma_{i,x}},$$
$$A(t,T,x) = \sum_{i=1}^{3} \frac{2k^Q_{i,x}\theta^Q_i}{\left(e^{c_i + d_i h_x}\right)^2} \ln\left[\frac{2\gamma_{i,x}\exp\left(\frac{\left(k^Q_{i,x} + \gamma_{i,x}\right)(T-t)}{2}\right)}{\left(k^Q_{i,x} + \gamma_{i,x}\right)\left(e^{\gamma_{i,x}(T-t)} - 1\right) + 2\gamma_{i,x}}\right],$$

where $\gamma_{i,x} = \sqrt{\left(k^Q_{i,x}\right)^2 + 2\left(e^{c_i + d_i h_x}\right)^2}$ for $i = 1, 2, 3$.

2.3. Change of Measure

We apply Girsanov's theorem to change from the risk-neutral measure Q to the real-world measure P, thereby accommodating historical mortality data over time. We have the following relationship between the risk-neutral measure Q and the real-world measure P:

$$\mathrm{d}W^Q(t) = \mathrm{d}W^P(t) + \Lambda^M(t)\mathrm{d}t,$$

where $W^P(t)$ is a 3×1 vector of three standard Brownian motions under the real-world measure P, and $\Lambda^M(t)$ is the market price of risk. Following Blackburn and Sherris (2013) and Huang et al. (2022), we define the mortality risk premium to be as follows:

$$\Lambda^M(t) = \begin{cases} \lambda^0 + \lambda^1 \mathbf{Z}(t) & \text{for Gaussian models,} \\ D(\mathbf{Z}(t))\lambda^0 & \text{for CIR models,} \end{cases}$$

where $\Lambda^M(t) \in \mathbb{R}^{3\times 1}$, $\lambda^0 \in \mathbb{R}^{3\times 1}$, and $\lambda^1 \in \mathbb{R}^{3\times 3}$.

Proposition 1. *For the ADC Gaussian model, after the change of measure, the SDE of the factor $\mathbf{Z}(t)$ can be represented as*

$$d\mathbf{Z}(t) = K^P\left[\Theta^P - \mathbf{Z}(t)\right]\mathrm{d}t + \Sigma\mathrm{d}W^P(t), \tag{13}$$

and for the ADC CIR model, the SDE after the change of measure is represented as

$$d\mathbf{Z}(t) = K^P\left[\Theta^P - \mathbf{Z}(t)\right]\mathrm{d}t + \Sigma D(\mathbf{Z}(t))\mathrm{d}W^P(t). \tag{14}$$

where K^P is a 3×3 diagonal matrix of the rate of mean-reversion, Θ^P is a 3×1 vector of the long-term mean, and $W^P(t)$ is a 3×1 vector of three standard Brownian motions under the real-world measure P. The detailed definitions of K^P and Θ^P can be found in Appendices A.1 and A.2.

We assume both Θ^Q and Θ^P are zero vectors for practical and identificational reasons. We note that mortality rates are exponential rather than mean-reverting processes, consistent with Luciano and Vigna (2008) and Blackburn and Sherris (2013). Allowing for non-zero long-term means introduces six additional parameters and therefore increases model complexity. However, these parameters are weakly identified and lead to less robust calibration, regardless of whether the Feller condition is imposed in the CIR specification as a calibration constraint, without providing material improvements in either in-sample fit or out-of-sample forecasting performance. Furthermore, as displayed in Table 9, the estimated mean-reversion speeds are extremely small, so the latent factors behave effectively as non-mean-reverting over empirically relevant horizons. As a result, imposing zero long-term mean vectors has a negligible practical impact while improving estimation robustness.

Proof. See Appendix A.1 and A.2. ■

Proposition 2. *For ADF models, we have more than one age group, and for cohorts in the same age group that start with age x. For the ADF Gaussian model, the SDE of the factor $\mathbf{Z}(t,x)$ after change of measure can be represented as*

$$d\mathbf{Z}(t,x) = K_x^{\mathrm{P}}\left[\Theta^P - \mathbf{Z}(t,x)\right]\mathrm{d}t + \Sigma_x \mathrm{d}W^P(t), \tag{15}$$

and for the ADF CIR model, the SDE after the change of measure is represented as

$$d\mathbf{Z}(t,x) = K_x^{\mathrm{P}}\left[\Theta^P - \mathbf{Z}(t,x)\right]\mathrm{d}t + \Sigma_x D(\mathbf{Z}(t,x))\mathrm{d}W^P(t). \tag{16}$$

For the same reason as in ADC models, we assume both Θ^Q and Θ^P are zero vectors. Since we only have two age groups, we are able to represent the K^P in terms of a_i^{P} and b_i^{P} as follows*:*

$$K_x^{\mathrm{P}} = \begin{pmatrix} a_1^P + b_1^P h_x & 0 & 0 \\ 0 & a_2^P + b_2^P h_x & 0 \\ 0 & 0 & a_3^P + b_3^P h_x \end{pmatrix}.$$

Proof. See Appendix A.3 and A.4. ■

2.4. The Cohort Correlation

Cohort correlation can be easily modeled under the age dependence framework. We calculate the correlation between the infinitesimal change $\mathrm{d}\mu_x(t)$ of different cohorts to obtain the cohort correlation between the change in the mortality rates at time t.

2.4.1. The ADC Gaussian Model

For ADC models, we need to first use Itô's formula to obtain the SDE for the mortality intensities $\mathrm{d}\mu_x(t)$.

Lemma 1. *For the ADC Gaussian model, the SDE of the mortality intensity is represented as*

$$\mathrm{d}\mu_x(t) = \sum_{i=1}^{3} g_i(t-c)\Big[-\tilde{\lambda}_i(t-c)Z_i(t)\mathrm{d}t + \sigma_i \mathrm{d}W_i^{\mathrm{P}}(t)\Big],$$

where $\tilde{\lambda}_i(t-c) = (k_i^{\mathrm{P}} - \frac{g_i'(t-c)}{g_i(t-c)})$ *for notational convenience, and* $g_i'(t-c)$ *represents* $\frac{\partial g_i(t-c)}{\partial t}$.

Proof. See Appendix B.1. ■

Proposition 3. *For the ADC Gaussian model, the instantaneous cohort correlation between cohort* m *and* n *conditional on the information up to time* t *is*

$$\rho(\mathrm{d}\mu_m(t), \mathrm{d}\mu_n(t)|\mathcal{F}_t) = \frac{\sigma_1^2 g_1(m)g_1(n) + \sigma_2^2 g_2(m)g_2(n) + \sigma_3^2 g_3(m)g_3(n)}{\sqrt{\sigma_1^2 g_1(m)^2 + \sigma_2^2 g_2(m)^2 + \sigma_3^2 g_3(m)^2}\sqrt{\sigma_1^2 g_1(n)^2 + \sigma_2^2 g_2(n)^2 + \sigma_3^2 g_3(n)^2}}.$$

The cohort correlation is the same under the risk-neutral measure Q, since the change of measure does not affect the correlation, and this applies to all four models.

Proof. From Lemma 1, the SDEs for the mortality intensities of two cohorts aged m and n at the same time t are

$$\begin{aligned}
\mathrm{d}\mu_m(t) &= \sum_{i=1}^{3} g_i(m)\Big[-\tilde{\lambda}_i(t-c)Z_i(t)\mathrm{d}t + \sigma_i \mathrm{d}W_i^{\mathrm{P}}(t)\Big],\\
\mathrm{d}\mu_n(t) &= \sum_{i=1}^{3} g_i(n)\Big[-\tilde{\lambda}_i(t-c)Z_i(t)\mathrm{d}t + \sigma_i \mathrm{d}W_i^{\mathrm{P}}(t)\Big],
\end{aligned}$$

from which we obtain the cohort correlation by calculating the correlation between the two SDEs

$$\begin{aligned}
&\rho(\mathrm{d}\mu_m(t), \mathrm{d}\mu_n(t)|\mathcal{F}_t)\\
&= \frac{Cov(\mathrm{d}\mu_m(t), \mathrm{d}\mu_n(t)|\mathcal{F}_t)}{\sqrt{Var(\mathrm{d}\mu_m(t)|\mathcal{F}_t)}\sqrt{Var(\mathrm{d}\mu_n(t)|\mathcal{F}_t)}}\\
&= \frac{\sigma_1^2 g_1(m)g_1(n) + \sigma_2^2 g_2(m)g_2(n) + \sigma_3^2 g_3(m)g_3(n)}{\sqrt{\sigma_1^2 g_1(m)^2 + \sigma_2^2 g_2(m)^2 + \sigma_3^2 g_3(m)^2}\sqrt{\sigma_1^2 g_1(n)^2 + \sigma_2^2 g_2(n)^2 + \sigma_3^2 g_3(n)^2}}.
\end{aligned}$$

■

2.4.2. The ADC CIR Model

Lemma 2. *For the ADC CIR model, the SDEs for the mortality intensities of two cohorts aged* m *and* n *at the same time* t *are shownas*

$$\begin{aligned}
\mathrm{d}\mu_m(t) &= \sum_{i=1}^{3} g_i(m)\Big[-\tilde{\lambda}_i(t-c)Z_i(t)\mathrm{d}t + \sigma_i\sqrt{Z_i(t)}\mathrm{d}W_i^{\mathrm{P}}(t)\Big],\\
\mathrm{d}\mu_n(t) &= \sum_{i=1}^{3} g_i(n)\Big[-\tilde{\lambda}_i(t-c)Z_i(t)\mathrm{d}t + \sigma_i\sqrt{Z_i(t)}\mathrm{d}W_i^{\mathrm{P}}(t)\Big].
\end{aligned}$$

Proof. See Appendix B.2. ■

Proposition 4. *For the ADC CIR model, the cohort correlation between cohorts aged* m *and* n *conditional on the information up to time* t *is*

$$\rho(\mathrm{d}\mu_m(t), \mathrm{d}\mu_n(t)|\mathcal{F}_t) = \frac{J(t,m,n)}{K(t,m,n)},$$

where

$$\begin{aligned} J(t,m,n) &= \sigma_1^2 g_1(m)g_1(n)Z_1(t) + \sigma_2^2 g_2(m)g_2(n)Z_2(t) + \sigma_3^2 g_3(m)g_3(n)Z_3(t), \\ K(t,m,n) &= \sqrt{\sigma_1^2 g_1(m)^2 Z_1(t) + \sigma_2^2 g_2(m)^2 Z_2(t) + \sigma_3^2 g_3(m)^2 Z_3(t)} \\ &\quad \times \sqrt{\sigma_1^2 g_1(n)^2 Z_1(t) + \sigma_2^2 g_2(n)^2 Z_2(t) + \sigma_3^2 g_3(n)^2 Z_3(t)}. \end{aligned}$$

Proof. The calculations are similar to those presented in the proof of Proposition 3, using the SDEs for the mortality intensities presented in Lemma 2. ■

Compared with the Gaussian model, the CIR model has the advantage that the correlations are truly specific to cohorts as the term $Z_i(t)$ enters into the correlation equation and brings in the dimension of time t. Therefore, for CIR models, we have specified both ages m, n and the time t, making the cohort correlation expression time-inhomogeneous and specific to the cohorts. Meanwhile, for the Gaussian model, the cohort correlation is time-homogeneous and thus age-difference correlation.

2.4.3. The ADF Gaussian Model

Lemma 3. *For the ADF Gaussian model, the SDEs for the mortality intensities of cohort aged* m *and* n *at time* t *that belong to age groups with age indexes* h_m *and* h_n *are*

$$\begin{aligned} \mathrm{d}\mu_m(t) &= \sum_{i=1}^{3}\Big[-\big(a_i^{\mathrm{P}} + b_i^{\mathrm{P}} h_m\big)Z_i(t,h_m)\mathrm{d}t + e^{c_i + d_i h_m}\mathrm{d}W_i^{\mathrm{P}}(t)\Big], \\ \mathrm{d}\mu_n(t) &= \sum_{i=1}^{3}\Big[-\big(a_i^{\mathrm{P}} + b_i^{\mathrm{P}} h_n\big)Z_i(t,h_n)\mathrm{d}t + e^{c_i + d_i h_n}\mathrm{d}W_i^{\mathrm{P}}(t)\Big]. \end{aligned}$$

Proof. Using Itô's Lemma on Equation (4), where $Z_i(t,h_x)$ satisfies the SDEs in Equation (15)

$$\begin{aligned} \mathrm{d}\mu_x(t) &= \sum_{i=1}^{3}\mathrm{d}Z_i(t,h_x) \\ &= \sum_{i=1}^{3} -\big(a_i^{\mathrm{P}} + b_i^{\mathrm{P}} h_x\big)Z_i(t,h_x)\mathrm{d}t + e^{c_i + d_i h_x}\mathrm{d}W_i^{\mathrm{P}}(t). \end{aligned}$$

■

Proposition 5. *For the ADF Gaussian model, the correlation between cohorts aged* m *and* n *conditional on the information up to time* t *is*

$$\rho(\mathrm{d}\mu_m(t), \mathrm{d}\mu_n(t)|\mathcal{F}_t) = \frac{e^{2c_1 + d_1(m+n)} + e^{2c_2 + d_2(m+n)} + e^{2c_3 + d_3(m+n)}}{\sqrt{e^{2(c_1 + d_1 m)} + e^{2(c_2 + d_2 m)} + e^{2(c_3 + d_3 m)}}\sqrt{e^{2(c_1 + d_1 n)} + e^{2(c_2 + d_2 n)} + e^{2(c_3 + d_3 n)}}}.$$

Proof. The calculations are similar to those presented in the proof of Proposition 3, using the SDEs for the mortality intensities presented in Lemma 3. We replace the age index h_x with the actual age x to smooth the inter-group cohort correlation. ■

2.4.4. The ADF CIR Model

Lemma 4. *For the ADF CIR model, the SDEs for the mortality intensities of cohorts aged* m *and* n *at time* t *that belong to age groups with age indexes h_m and h_n are*

$$\begin{aligned}
\mathrm{d}\mu_m(t) &= \sum_{i=1}^{3}\Big[-\big(a_i^{\mathrm{P}} + b_i^{\mathrm{P}}h_m\big)Z_i(t,h_m)\mathrm{d}t + e^{c_i+d_ih_m}\sqrt{Z_i(t,h_m)}\mathrm{d}W_i^{\mathrm{P}}(t)\Big],\\
\mathrm{d}\mu_n(t) &= \sum_{i=1}^{3}\Big[-\big(a_i^{\mathrm{P}} + b_i^{\mathrm{P}}h_n\big)Z_i(t,h_n)\mathrm{d}t + e^{c_i+d_ih_n}\sqrt{Z_i(t,h_n)}\mathrm{d}W_i^{\mathrm{P}}(t)\Big].
\end{aligned}$$

Proof. Use Itô's Lemma on Equation (4), where $Z_i(t,h_x)$ satisfies the SDEs in Equation (16). ■

Proposition 6. *For the ADF CIR model, the correlation between cohorts aged* m *and* n *conditional on the information up to time* t *is*

$$\rho(\mathrm{d}\mu_m(t),\mathrm{d}\mu_n(t)|\mathcal{F}_t) = \frac{J(t,m,n)}{K(t,m,n)},$$

where

$$\begin{aligned}
J(t,m,n) &= e^{2c_1+d_1(m+n)}\sqrt{Z_1(t,m)Z_1(t,n)} + e^{2c_2+d_2(m+n)}\sqrt{Z_2(t,m)Z_2(t,n)}\\
&\quad + e^{2c_3+d_3(m+n)}\sqrt{Z_3(t,m)Z_3(t,n)},\\
K(t,m,n) &= \sqrt{e^{2(c_1+d_1m)}Z_1(t,m) + e^{2(c_2+d_2m)}Z_2(t,m) + e^{2(c_3+d_3m)}Z_3(t,m)}\\
&\quad \times\sqrt{e^{2(c_1+d_1n)}Z_1(t,n) + e^{2(c_2+d_2n)}Z_2(t,n) + e^{2(c_3+d_3n)}Z_3(t,n)}.
\end{aligned}$$

Proof. The calculations are similar to those presented in the proof of Proposition 3, using the SDEs for the mortality intensities presented in Lemma 4. We replace the age index h_x with the actual age x to smooth the inter-group cohort correlation. ■

The previous argument on the time-inhomogeneous cohort correlation of the CIR model holds under the setting of ADF models.

3. DATA: CONSTRUCTION OF INCOMPLETE AGE-COHORT DATA

Mortality data for males from four countries: Australia, Denmark, the United Kingdom (UK), and the United States of America (USA) obtained from the Human Mortality Database (2022) (HMD) are used in this research. The age range is 50–109, and the ranges of calendar years are 1921–2020 for Australia, 1920–2021 for Denmark, 1933–2020 for the UK, and 1922–2020 for the USA depending on data availability.[2]

We first obtain the age-period-based mortality data from the HMD, but age-cohort data is required since we aim to model the survival probability for each cohort. Since the age of a cohort i increases by 1 as the time moves forward by 1, we need to take along the diagonal of the mortality table to obtain the age-cohort data. The survival probability $S_x(t,T)$ for a cohort aged x at time t to survive to time T is

$$\begin{aligned}
S_x(t,T) &= \prod_{s=1}^{T-t}[1-q(x+s-1,t+s-1)],\\
\bar{\mu}_x(t,T) &= -\frac{1}{T-t}\log\left[S_x(t,T)\right],
\end{aligned}$$

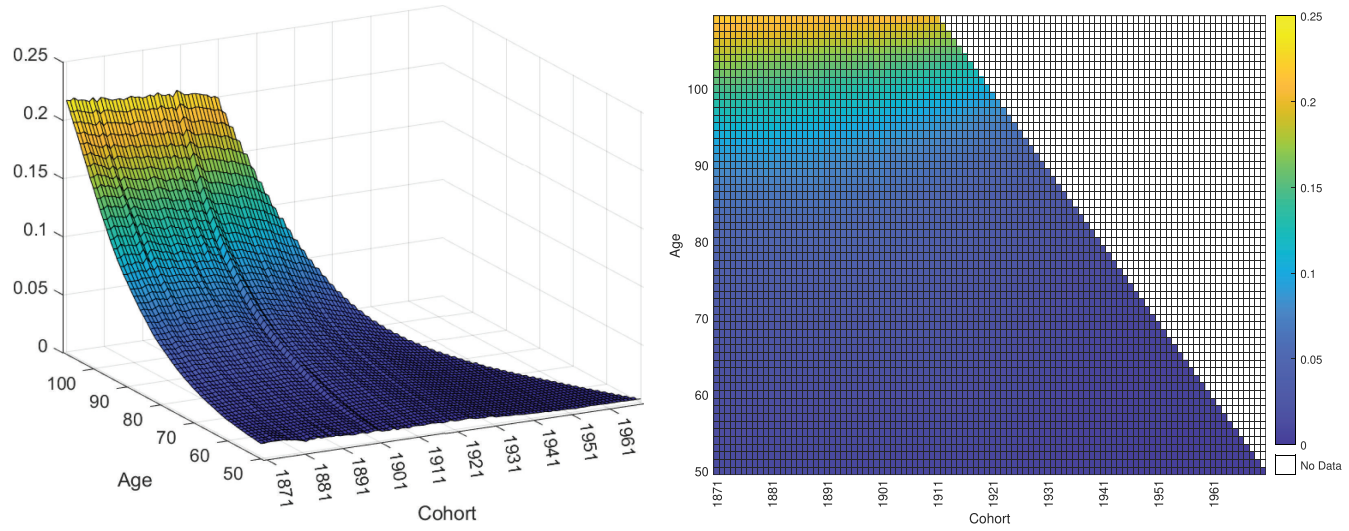
[2]The slight inconsistency in the calendar year ranges across countries reflects differences in when data became available for each country.

where $q(x,t)$ is the probability of dying within one year for a cohort aged x at time t, and $\bar{\mu}_x(t,T)$ is the average force of mortality of the cohort from time t to T.

We denote the range of the year of birth of cohorts as $[c_0, c_{\max}]$, the range of age as $[x_0, x_{\max}]$, and thus the range of calendar years as $[t_0, t_{\max}]$, where $t_0 = c_0 + x_0$, $t_{\max} = c_{\max} + x_{\max}$ for complete-cohort data only, and $t_{\max} = c_{\max} + x_0$ including the incomplete-cohort data. We also denote $N_t = t_{\max} - t_0 + 1$ the number of time points, $N_x = x_{\max} - x_0 + 1$ the number of ages, and $N_c = c_{\max} - c_0 + 1$ the number of cohort years of birth.

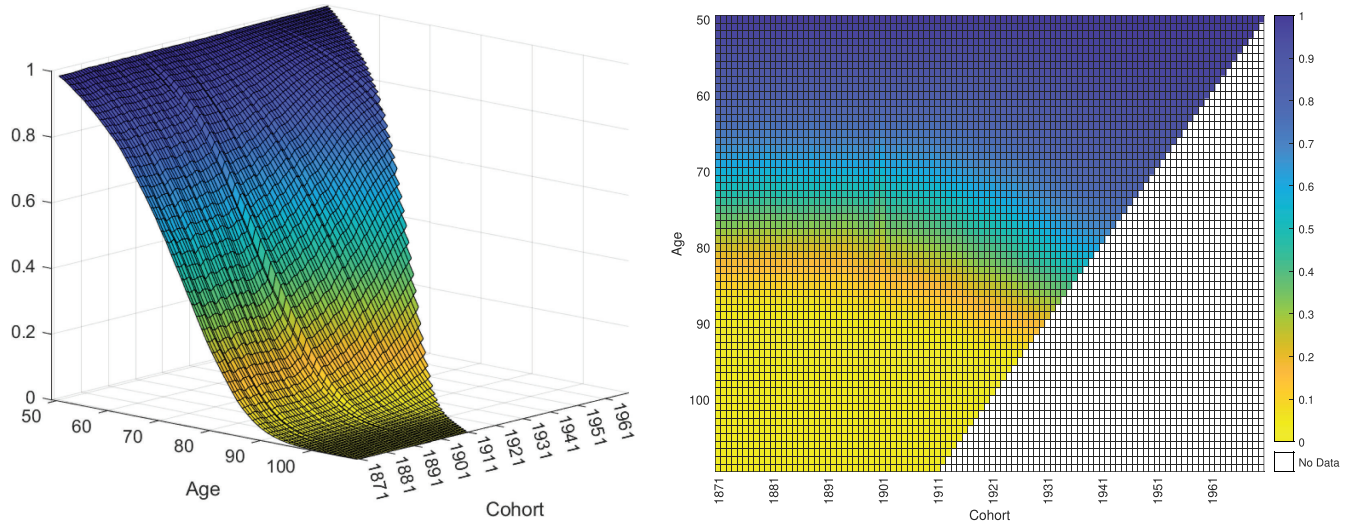
Figure 1 shows the example of the Australian cohort average force of mortality for males. As we can see, the cohort axis represents the year of birth of the cohorts. From how we construct cohort data, we know that the recent cohorts, for example, the cohort born in 1951 and aged 69 in 2020, only have incomplete observation because they have not experienced the age 70–109. Therefore, in Figure 1, there are triangular incomplete-cohort data after the complete cohorts.

Figure 2 shows the full cohort data of survival probabilities from age 50 that includes the triangular incomplete-cohort data.



(a) Surface. (b) Heatmap.

FIGURE 1. Australian Cohort Average Force of Mortality for Males Born between 1871 and 1970, from Age 50 to 109: Surface Plot and Heatmap.



(a) Surface. (b) Heatmap.

FIGURE 2. Australian Cohort Survival Probabilities for Males Born between 1871 and 1970, from Age 50 to 109: Surface Plot and Heatmap.

From Figure 2(b), we observe that the mortality has improved significantly for cohorts born after 1911, especially around age 60–80, as evidenced by the darker shade of blue. However, for these cohorts, we only have incomplete observations. Hence, if we only use complete-cohort data to calibrate the model, we will miss some important features in the recent mortality data and thus reduce the forecasting power of the model. This leads us to use incomplete-cohort data to calibrate the model, and the detailed algorithm will be introduced in Section 4.

4. ESTIMATION TECHNIQUE

We use the Kalman filter (Kalman 1960) to estimate the parameters of the mortality models. Kalman filtering can be used to estimate the model parameters by sequentially updating the parameters and maximizing the log-likelihood function, and this technique has been applied in Christensen et al. (2011); Blackburn and Sherris (2013); Xu et al. (2020); and Huang et al. (2022) to estimate the parameters under the affine framework. We propose an extension of this technique which can accommodate incomplete-cohort data to improve the forecasting ability of the model. We also add regularization to improve the goodness of fit to empirical cohort correlation. Moreover, for ADF models, we separate two age groups and sum their log-likelihood, and we include another regularization term to improve the smoothness of survival curves at the joint of the two age groups. The procedure is described below.

4.1. General Framework

By transforming Equation (5) into vector form, the average force of mortality for all age-dependent models can be represented as

$$\bar{\mu}_x(t,T) = -\frac{1}{T-t}\log\left[S_x(t,T)\right] = \mathbf{B}^\top \mathbf{Z}_t + \mathbf{A}, \tag{17}$$

where $\mathbf{B} = -\frac{1}{T-t}[B_1(t,T), B_2(t,T), B_3(t,T)]^\top$, $\mathbf{A} = -\frac{1}{T-t}A(t,T)$, and $\mathbf{Z}_t = [Z_1(t), Z_2(t), Z_3(t)]^\top$.

We introduce measurement error ε_t into Equation (17) as in Blackburn and Sherris (2013) to obtain the measurement equation, represented as

$$\bar{\mu}_x(t,T) = \mathbf{B}^\top \mathbf{Z}_t + \mathbf{A} + \varepsilon_t.$$

Meanwhile, we discretise SDEs of $\mathbf{Z}_t$ in Propositions 1 and 2 and introduce transition error η_t. Since we have $\Theta^P = \mathbf{0}$, the state transition equation is represented as

$$\mathbf{Z}_t = \Phi \mathbf{Z}_{t-1} + \eta_t, \tag{18}$$

where $\Phi = e^{-K^P}$. The error structure is represented as

$$\begin{pmatrix} \eta_t \\ \varepsilon_t \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} Q^e & 0 \\ 0 & R^e \end{pmatrix}\right],$$

where Q^e and R^e are the covariance matrices of the transition error and the measurement error, respectively.

According to Christensen et al. (2011), the transition error matrix Q^e has the following representation for Gaussian models:

$$Q^e = \int_0^{\Delta t} e^{-K^P s}\Sigma\Sigma^\top e^{-(K^P)^\top s} ds,$$

where we set $\Delta t = 1$ representing one year.

Meanwhile, for CIR models, the transition error matrix Q^e is a 3×3 time-dependent diagonal matrix $Q^e(t)$, and we set the i^{th} element on the diagonal $Q^e_{i,i}(t)$ to be

$$Q^e_{i,i}(t) = Z_i(t-1)\frac{\sigma_i^2}{k_i^{\mathrm{P}}}\left(e^{-k_i^{\mathrm{P}}\Delta t} - e^{-2k_i^{\mathrm{P}}\Delta t}\right),$$

where we also set $\Delta t = 1$. The above representation for CIR models is an approximation because the distribution of $Z_i(t)$ conditional on $Z_i(t-1)$ is a χ^2 distribution (Cox et al. 1985). The approach we use approximates the exact distribution with a Gaussian distribution that matches the first and second moments (Geyer and Pichler 1999; Chen and Scott 2003; Huang et al. 2022).

For the measurement error matrix R^e, we use three parameters r_1, r_2, and r_c to construct the following equation representing the increasing measurement error with age

$$R^e(t,T) = \frac{1}{T-t}\sum_{i=1}^{T-t}\left[r_c + r_1 e^{r_2 i}\right],$$

where r_1, r_2, and r_c only take positive values.

4.2. The Kalman Filtering with Incomplete-Cohort Data

Let Y_t denote the information up to time t, y_t the vector of the average force of mortality for the cohort aged x at time t for all integer values of $T-t$, N the number of observations, and ψ the set of parameters. The best-estimated state factors and volatility matrix at time t conditional on the information at time $t-1$ are

$$\begin{aligned}\mathbf{Z}_{t|t-1} &= \mathbb{E}[\mathbf{Z}_t|Y_{t-1}] = \Phi(\psi)\mathbf{Z}_{t-1},\\ \Sigma_{t|t-1} &= \Phi(\psi)\Sigma_{t-1}\Phi(\psi)^\top + Q^e(\psi).\end{aligned}$$

Then, use the information at time t to update the forecast

$$\begin{aligned}\mathbf{Z}_t &= \mathbb{E}[\mathbf{Z}_t|Y_t] = \mathbf{Z}_{t|t-1} + \Sigma_{t|t-1}\mathbf{B}(\psi)^\top F_t^{-1} v_t,\\ \Sigma_t &= \Sigma_{t|t-1} - \Sigma_{t|t-1}\mathbf{B}(\psi)^\top F_t^{-1}\mathbf{B}(\psi)\Sigma_{t|t-1},\end{aligned}$$

where

$$\begin{aligned}v_t &= y_t - \mathbb{E}[y_t|Y_{t-1}] = y_t - \mathbf{A}(\psi) - \mathbf{B}(\psi)\mathbf{Z}_{t|t-1},\\ F_t &= \mathrm{cov}(v_t) = \mathbf{B}(\psi)\Sigma_{t|t-1}\mathbf{B}(\psi)^\top + R^e(\psi),\end{aligned}$$

$\Phi(\psi) = e^{-K^P}$, $\mathbf{B}(\psi)$, and $\mathbf{A}(\psi)$ are the factor loadings with the set of parameters ψ, and $Q^e(\psi)$ and $R^e(\psi)$ are the transition error and measurement error matrices with the set of parameters ψ.

If at some time t we only have incomplete observation of y_t of length m, then we shrink the size of $\mathbf{A}(\psi)$, $\mathbf{B}(\psi)$, v_t, F_t to $m\times 1$, $m\times 3$, $m\times 1$, $m\times m$, respectively, by dropping the elements in the last rows or columns and take the above updating step.

The updating procedure gives the log-likelihood function

$$\log l(y_1,...,y_T;\psi) = \sum_{t=1}^{T}\left(-\frac{N}{2}\log(2\pi) - \frac{1}{2}\log(F_t) - \frac{1}{2}v_t^\top F_t^{-1} v_t\right),$$

and the optimal parameter set is the one that maximizes the log-likelihood function. The built-in MATLAB optimization toolbox is used to maximize the objective function.

4.3. Separate Age Groups for ADF Models

For ADF models, the factors $a_i + b_i h_x$ and $e^{c_i + d_i h_x}$ are constant if we do not separate the age groups. We can set the age to be $t-c$ for the ADC models to have the age increasing as we integrate, but this implies that we need to recalculate the survival probability for each 1-year horizon with a different initial age, which will make the number of factors unreasonably high.

To this end, we consider ADF models where the factor dynamics depend on age groups, rather than individual ages. The idea of age grouping is to form the ages with a similar level of error into an age group to be estimated, and thus reduce

the error and at the same time estimate the age dependence relationship. Meanwhile, the number of age groups needs to reflect the age dependence, but also not too large to avoid introducing too many factors.

We decide to choose two age groups: age $[50-79]$, and age $[80-109]$ because the mean absolute percentage errors (MAPEs) for survival probabilities start exploding from age 80 in our experiments and also as observed in the literature (Blackburn and Sherris 2013; Xu et al. 2020; Huang et al. 2022; Ungolo et al. 2024). Within each age group, the age index h_x is assumed to remain constant as the starting age of this age group. For example, a cohort aged $x = 65$ at time t belongs to the age group $[50-79]$ and thus $h_{65} = 50$. With two age groups, we have three factors for each age group so a total of six factors, which is considered a reasonable number since the number of factors for a single-age-group model is normally between one to five (Dahl 2004; Biffis 2005; Blackburn and Sherris 2013; Ungolo et al. 2024). While a larger number of age groups could provide a finer presentation of age dependence, in the context of multi-factor affine models, such refinements come at a substantial cost in terms of the number of parameters and estimation reliability. Increasing the number of age groups to three leads to nine latent factors and weakly identified parameters in calibration. For lower-dimensional specifications (for example, one or two-factor models), introducing additional age groups may be more feasible, and we acknowledge this as a potential direction for future research.

Moreover, the most pronounced change in the drift and volatility structure of mortality dynamics occurs around age 80, which is consistent with the distinction between "young old" and "old old" (Smith et al. 2002). In line with Chang and Sherris (2018), the volatility of mortality rates appears to increase exponentially with age while remaining relatively stable up to around age 80. The two-age-group specification is therefore intended to capture this dominant structural difference while maintaining model parsimony. Beyond discrete age grouping, alternative formulations could also be considered within the ADF framework. One possibility is to allow the drift and volatility parameters of the latent factors to vary smoothly with age. Another approach is to introduce age-dependent regime-switching dynamics, in which the latent factors follow different stochastic laws across age-dependent regimes. While these alternatives would offer greater flexibility in capturing age heterogeneity in mortality dynamics, they would substantially complicate the affine structure and the associated state-space representation, and are therefore beyond the scope of the present article.

Then, Kalman filtering can be used to estimate the parameters for each age group. If we implement the Kalman filter for each age group separately, it will result in different values of a_i, b_i, c_i, d_i across the age groups. Since our goal is to estimate the same values of a_i, b_i, c_i, d_i for both age groups, we propose the following approach.

We can set the objective function to be the sum of the likelihood functions for all age groups and run the optimization. The objective function is represented as

$$\sum_{g\in G} \log l(y_1, ..., y_T; \psi) = \sum_{g\in G} \sum_{t=1}^{T} w_g \left(-\frac{N}{2}\log(2\pi) - \frac{1}{2}\log(F_t) - \frac{1}{2} v_t^\top F_t^{-1} v_t \right), \tag{19}$$

where $g \in G = \{1, 2\}$ is the notation for the two age groups, and w_g represents the weight of the log-likelihood for age group g. In this work, we set equal weight $w_1 = w_2 = 1$. The next step involves running the Kalman filter to maximize the aggregate likelihood, which yields a set of constants a_i, b_i, c_i, d_i for all age groups.

One constraint imposed during Kalman filtering is $0 < a_i^{\mathrm{P}} + b_i^{\mathrm{P}} h_x < 1$ for $h_x = 50$ and 80 for ADF models, $i = 1, 2, 3$, and $0 < k^P < 1$ for ADC models, so that the processes of $\mathbf{Z}(t, x)$ are finite almost surely over the horizon considered.

4.4. Kalman Filtering with Regularization

4.4.1. Regularization of the Cohort Correlation

The objective function is represented as

$$\text{obj} = \log l(y_1, ..., y_T; \psi) - \lambda^r \sqrt{\sum_{i,j} (\hat{f}(i,j) - f(i,j))^2},$$

where $\hat{f}(i,j)$ is the estimated cohort correlation for a chosen calendar year which is 2000 in this article, $f(i,j)$ is the empirical cohort correlation for the chosen calendar year, $i, j = 1, 2, 3, 4, 5$ represent the selected cohorts aged $50, 60, 70, 80, 90$ in year 2000 as presented in Figure 3, and λ^r is the regularization parameter controlling the weight of the penalty. The discussion on the choice of λ^r can be found in Section 6. The idea of regularizing the distance between the estimated and empirical cohort correlation is to obtain the estimated parameters which better estimate the cohort correlation.

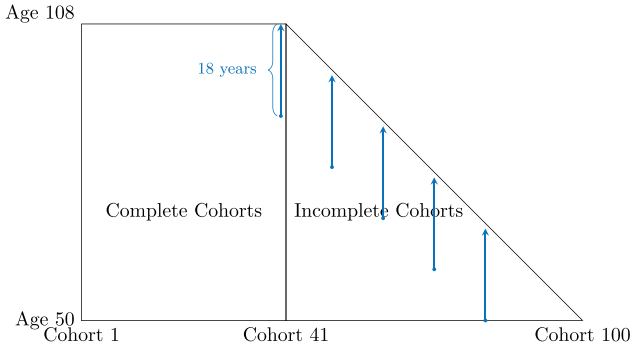


FIGURE 3. Differenced Force of Mortality $\Delta\mu_x^i(t)$. Blue Arrows from Left to Right Represent the Differenced Force of Mortality of Selected Cohorts Aged 90, 80, 70, 60, and 50 Respectively in Year 2000 for the Calculation of Cohort Correlation. For Each Cohort, an 18-Year Upward Horizon in Age is Selected to Calculate the Cohort Correlation.

For each combination of cohort and age, we only have one point observation of the instantaneous force of mortality. Therefore, we first use the one-year probability of death q_x at time t to approximate the force of mortality $\mu_x^i(t)$ of cohort i aged x at time t. Then we take the difference to approximate the dynamics, and the change in the force of mortality is $\Delta\mu_x^i(t) = \mu_{x+1}^i(t+1) - \mu_x^i(t)$. Finally, we use an 18-year upward horizon of age and approximate the empirical instantaneous cohort correlation as the figure below illustrates. The cohorts selected to build the empirical correlation matrix are those aged $50, 60, 70, 80$ and 90 in the calendar year 2000, and the selected cohorts along with their horizons are represented as the blue arrows in Figure 3. We choose an 18-year upward horizon because the empirical cohort correlation is calculated in year 2000, while the latest data is in year 2020 or 2021 depending on the country. This is reflected in Figure 3 that the arrows are close to the edge of the maximum available data. The motivation for this choice is to anchor the model-implied cohort correlation structure to the latest available mortality dynamics that are most relevant for forecasting, rather than to historical regimes that may no longer be representative of current mortality patterns. To assess the robustness of this choice, we repeat the regularized estimation using cohort correlations computed from alternative calendar years, and the results from these robustness checks can be found in Appendix D.

4.4.2. Regularization of Survival Curve Smoothness for ADF Models

For ADF models, we are interested in improving the smoothness of the survival curves at the joint point of the two age groups. Following the idea of Whittaker-Henderson smoothing (Bohlmann 1899; Whittaker 1922; Henderson 1924), we add an additional penalty term to the objective function of ADF models as follows:

$$\text{obj} = \log l(y_1, ..., y_T; \psi) - \lambda^r \sqrt{\sum_{i,j} (\hat{f}(i,j) - f(i,j))^2} - \zeta^r \sum_{t=t_0}^{t_0+N_c-1} \sum_{T=t+3}^{t+N_x-1} \left| \nabla^2 \hat{S}_{x_0}(t,T) \right|^2, \tag{20}$$

where ζ^r controls the weight for the smoothness regularization penalty, $\hat{S}_x(t,T)$ is the estimated survival probability with the parameter set, and ∇ represents the backward differencing operator $\nabla \hat{S}_x(t,T) = \hat{S}_x(t,T) - \hat{S}_x(t,T-1)$ and $\nabla^2 \hat{S}_x(t,T) = \nabla \hat{S}_x(t,T) - \nabla \hat{S}_x(t,T-1)$. We select $\zeta^r = 100,000$ because the smoothness term is around the scale of 10^{-2}. Therefore, the product of ζ^r and the smoothness term is around the scale of 10^3, while the log-likelihood function is around the scale of 10^4. This means the smoothness penalty impacts the likelihood function to a reasonable extent so that it is neither negligible nor outweighing.

5. ESTIMATION RESULTS

5.1. Model Estimation with Complete-Cohort Data

First, we calibrate the four models using only the complete-cohort data to compare with the existing models in the literature. We use the mean absolute percentage error (MAPE) to evaluate the goodness of fit of the models. In Blackburn and Sherris (2013), Huang et al. (2022), and Xu et al. (2020), the independent Gaussian affine mortality models that do not consider age dependence usually have MAPE of around 20–40% at age 100, which we aim to reduce by incorporating age dependence. The MAPEs for the survival probabilities of the four models calibrated using Australian male data are shown in Figure 4.

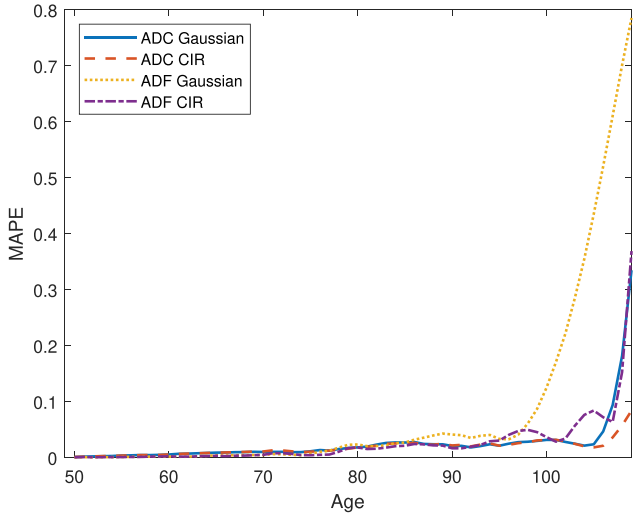


FIGURE 4. MAPE for Australian Male Survival Probabilities: Models Estimated with Complete-Cohort Data.

We can see that the MAPE at age 100 is around 3% for the ADC Gaussian, ADC CIR, and ADF CIR models, and around 12% for the ADF Gaussian model. We compare our MAPEs at age 100 to the models in the literature whose maximum age is 100 and do not consider age dependence (Blackburn and Sherris 2013; Xu et al. 2020; Huang et al. 2022; Ungolo et al. 2024), and we find that the MAPEs in our age-dependent models are significantly lower in the vast majority of cases. The ADC CIR model performs the best in the MAPE at age 109, which yields MAPE at around 8%, followed by ADC Gaussian and ADF CIR models at around 35%, and by ADF Gaussian at around 80%. The MAPEs are higher at age 109 than at age 100 because the survival probabilities to age 109 are almost equal to zero (that is 10^{-5}), making a tiny absolute difference in the estimated survival probability cause a large percentage error. From these results, we illustrate that adding age dependence to the mortality model can improve the goodness of fit at very old ages.

5.1.1. Forecast Out-of-Sample Survival Probabilities from Complete-Cohort Data

We perform and out-of-sample test by simulating via the transition equation given in Equation (18). This test is performed for all models calibrated to Australian data. For each test, we perform the 1, 5, 10, and 20 years ahead forecast, and compare the forecast with the incomplete empirical data. The results are shown in Figure 5. The red dashed lines are the forecasted survival probabilities, and the blue solid lines are the empirical ones. As we move forward in time, we enter the incomplete cohort portion of the data. Thus, the blue solid lines have missing values.

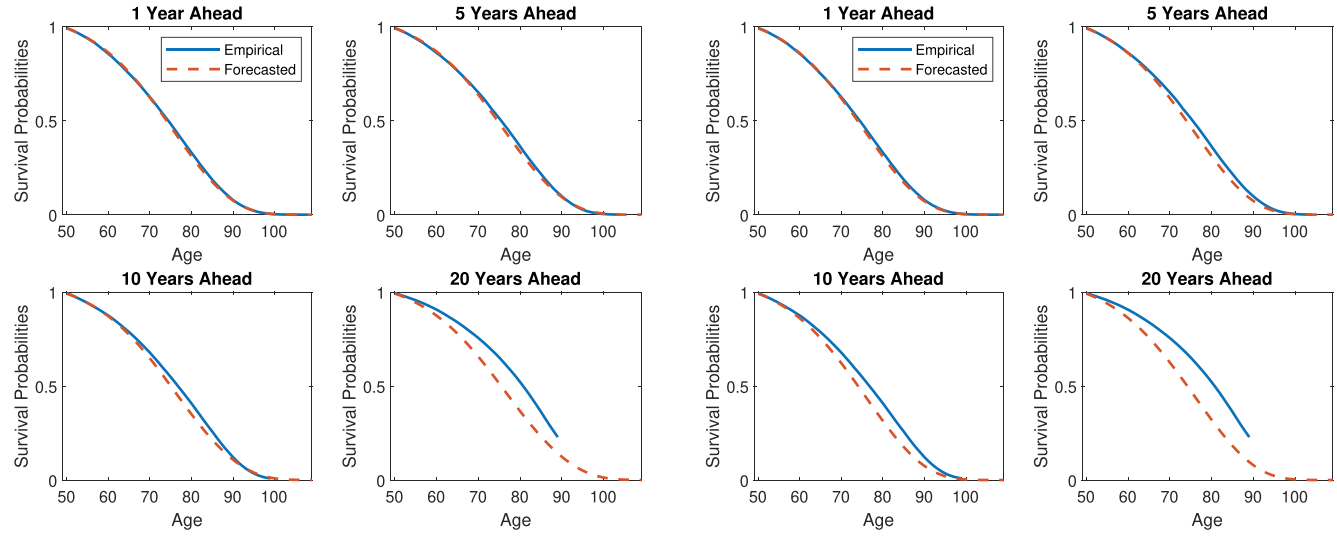
Table 1 displays the root mean squared error (RMSE) for different years ahead of forecast under different models. We observe from Table 1 that the RMSE increases as the forecasting time horizon increases. The RMSEs are below 0.01 for the 1-year ahead forecast, while they increase to around 0.1 for the 20-years ahead forecast. From Figure 5 and Table 1, we note that the models have relatively good out-of-sample performance in a short period horizon (that is, 1 and 5 years ahead). However, all models underestimate the survival probability to some extent, especially around age 60–80 in the 20-year ahead forecast. The reason for the underestimation is that the youngest cohort with complete observation was born 109 years ago, while mortality has improved significantly in the last 109 years, especially around age 60–80, as illustrated in Figure 2. This leads us to use incomplete-cohort data to calibrate the model, as we also discussed in Section 3.

5.2. Model Estimation with Manually Made Incomplete-Cohort Data

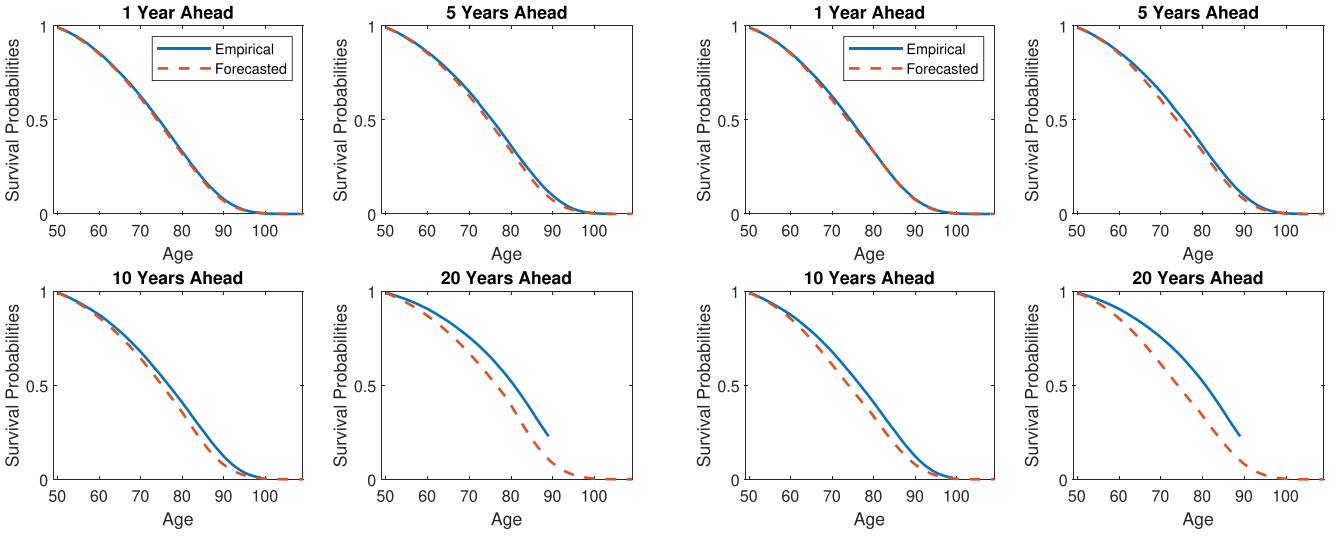
Before we use all the available incomplete-cohort data to fit the models, we are interested in testing the performance of the proposed method that shrinks the vector length in Kalman filtering to incorporate incomplete-cohort data. To achieve this, we intentionally make the complete-cohort data incomplete and calibrate the model, then compare the estimation results with the empirical data to see how well the missing data can be replicated.

We perform this test with Australian data and intentionally remove a 20×20 triangle in the complete data, as shown in Figure 6.

Figure 7 shows the comparison between the estimation results from the models calibrated with the intentionally made incomplete data and the empirical complete data. From Figure 7, we can see that our proposed approach yields a very good



(a) ADC Gaussian (b) ADC CIR



(c) ADF Gaussian (d) ADF CIR

FIGURE 5. Forecasted vs Empirical Australian Male Survival Probabilities of out-of-Sample Cohorts: Models Estimated with Complete-Cohort Data.

TABLE 1
RMSE of Forecasted Australian Male Survival Probabilities of Out-of-Sample Cohorts: Models Estimated with Complete-Cohort Data

Model	1 Year Ahead	5 Years Ahead	10 Years Ahead	20 Years Ahead
ADC Gaussian	0.0083	0.0140	0.0294	0.0985
ADC CIR	0.0098	0.0273	0.0531	0.1308
ADF Gaussian	0.0079	0.0190	0.0363	0.0925
ADF CIR	0.0084	0.0250	0.0506	0.1284

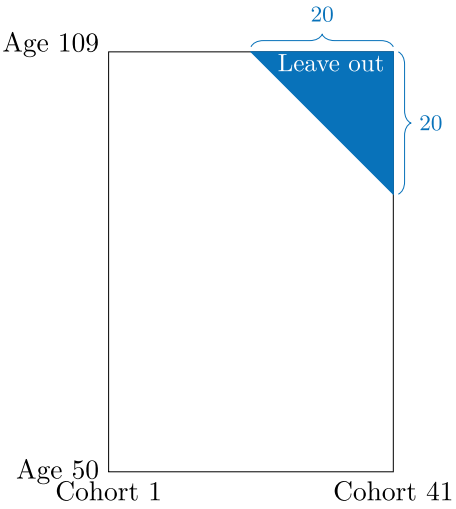


FIGURE 6. Leaving out a 20×20 Triangle from Complete-Cohort Data to Perform out-of-Sample Test.

TABLE 2
RMSE of Australian Male Survival Probabilities: Models Estimated with Manually Made Incomplete-Cohort Data

Model	1 Missing Value	5 Missing Values	10 Missing Values	20 Missing Values
ADC Gaussian	0.0031	0.0039	0.0040	0.0027
ADC CIR	0.0038	0.0037	0.0023	0.0051
ADF Gaussian	0.0026	0.0019	0.0038	0.0016
ADF CIR	0.0019	0.0017	0.0030	0.0037

estimate of the missing values. The four sub-figures are the cases when we have $1, 5, 10,$ and 20 intentionally made missing values. Even in the case when we have 20 missing values, our approach still replicates the missing data well. Table 2 shows the RMSEs for different models with different missing values, from which we can see the RMSEs are all below 0.01. Therefore, our approach provides a good way to fill up the missing values in the triangle of the incomplete-cohort data.

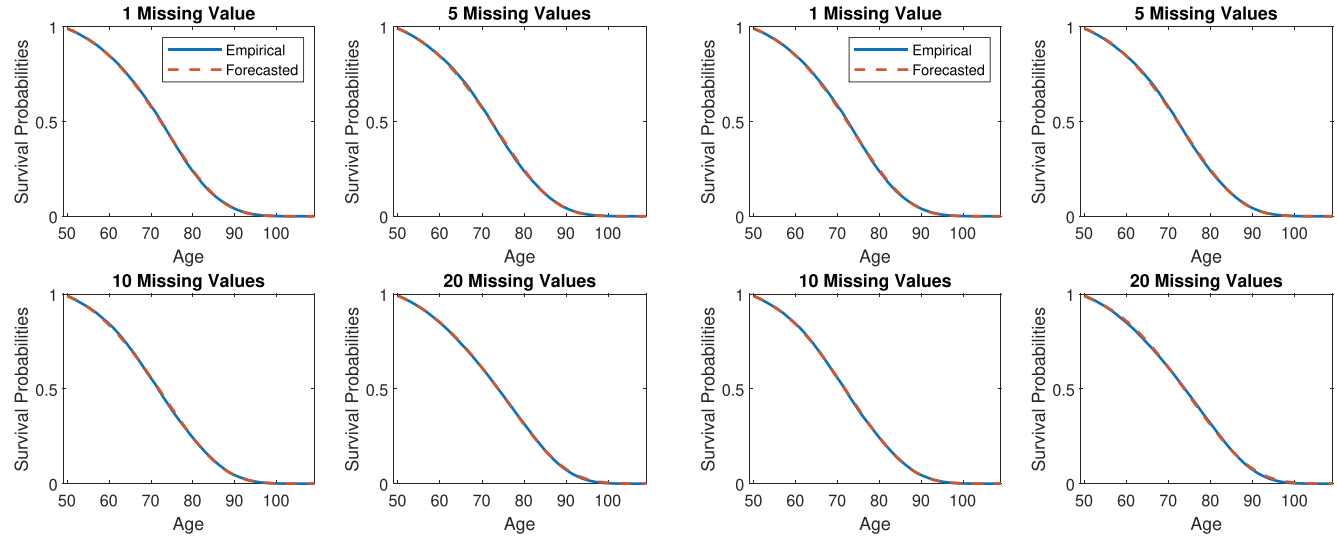
5.3. Model Estimation with All Available Incomplete-Cohort Data

Having illustrated the effectiveness of our approach to include incomplete cohorts, we now use all available incomplete-cohort data to calibrate the model. The MAPEs of the four models, when we set $\lambda^r = 1,000$ and $\zeta^r = 100,000$ and calibrated with Australian male mortality data, are shown in Figure 8. The discussion on the choice of λ^r and ζ^r can be found in Section 6 and Subsection 4.4 respectively, and the estimated parameters can be found in Appendix C.

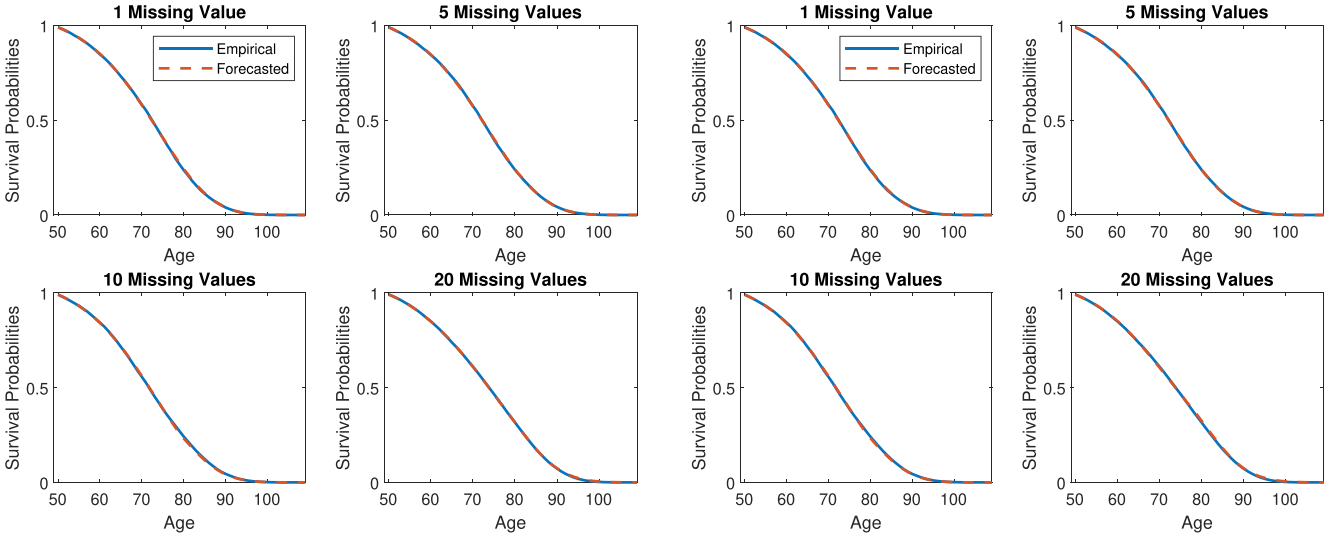
We can see that all models have very low MAPEs before age 90. At age 100, the ADC Gaussian, ADF Gaussian, and ADF CIR models have very low MAPEs of around 8%. The ADC CIR model has MAPE of around 30% at age 100, as opposed to being the best-performing model using all the complete-cohort data, but still indicates a good fit. At age 109, the ADC Gaussian and ADF Gaussian models perform the best, while the ADC CIR and ADF CIR models have higher MAPEs. But since the survival probabilities to age 109 are almost equal to zero (at the level of 10^{-5}) for our in-sample data, a slightly higher percentage error at age 109 does not mean much deviation from the estimation. Generally speaking, the MAPEs at old ages using incomplete-cohort data are higher than those using complete-cohort data because we are including the more recent incomplete cohorts that have missing data at old ages. Despite this, the MAPEs for all four models are still at a low level, indicating a nice goodness of fit.

5.3.1. Forecast Survival Probabilities from All Available Incomplete-Cohort Data

Having illustrated the goodness of fit of the models calibrated to all available incomplete-cohort data, we then use the estimated parameters to forecast the future survival probabilities from the youngest cohort we have, which is a much younger cohort compared to when the model is only calibrated with complete-cohort data. The idea is illustrated in Figure 9. We start



(a) ADC Gaussian (b) ADC CIR



(c) ADF Gaussian (d) ADF CIR

FIGURE 7. Estimated vs Empirical Australian Male Survival Probabilities: Models Estimated with Manually Made Incomplete-Cohort Data.

the forecast from Cohort 100, while models in the literature calibrated with complete-cohort data can only start the forecast from Cohort 41. It has been illustrated in Figure 5 that models calibrated with only complete-cohort data underestimate the survival probabilities for incomplete cohorts, while Figure 7 illustrates that our approach can effectively fill up the missing values for the incomplete cohorts.

The forecasting results are shown in Figure 10 which presents $1, 5, 10,$ and 20 years ahead forecasts. For all four models, we observe the mortality improvement since the survival probabilities are increasing over time. It is observed that the Gaussian models predict higher survival probabilities at the middle ages around 80 to 90, while the CIR models have a slightly higher tail at age 109. Note that for a 20 year ahead of forecast, the corresponding cohorts are people aged 30 now. The results in Figure 10 are obtained using regularization based on the cohort correlation in calendar year 2000, as shown in Figure 3. For robustness check, comparisons of 20-year ahead forecasts using models calibrated with cohort correlations calculated in calendar years 1980, 1990, and 2000 are reported in Appendix D.

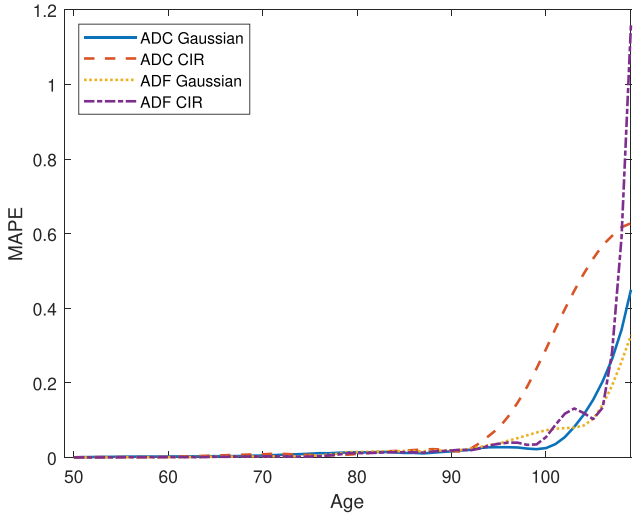


FIGURE 8. MAPE for Australian Male Survival Probabilities: Incomplete Cohort.

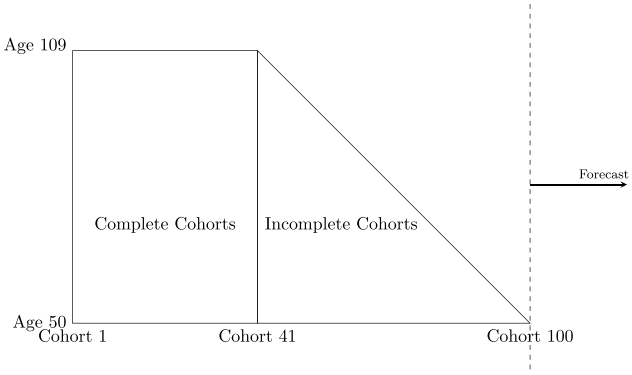


FIGURE 9. Incomplete-Cohort Data and Illustration of Forecast.

6. ANALYSIS OF COHORT CORRELATION

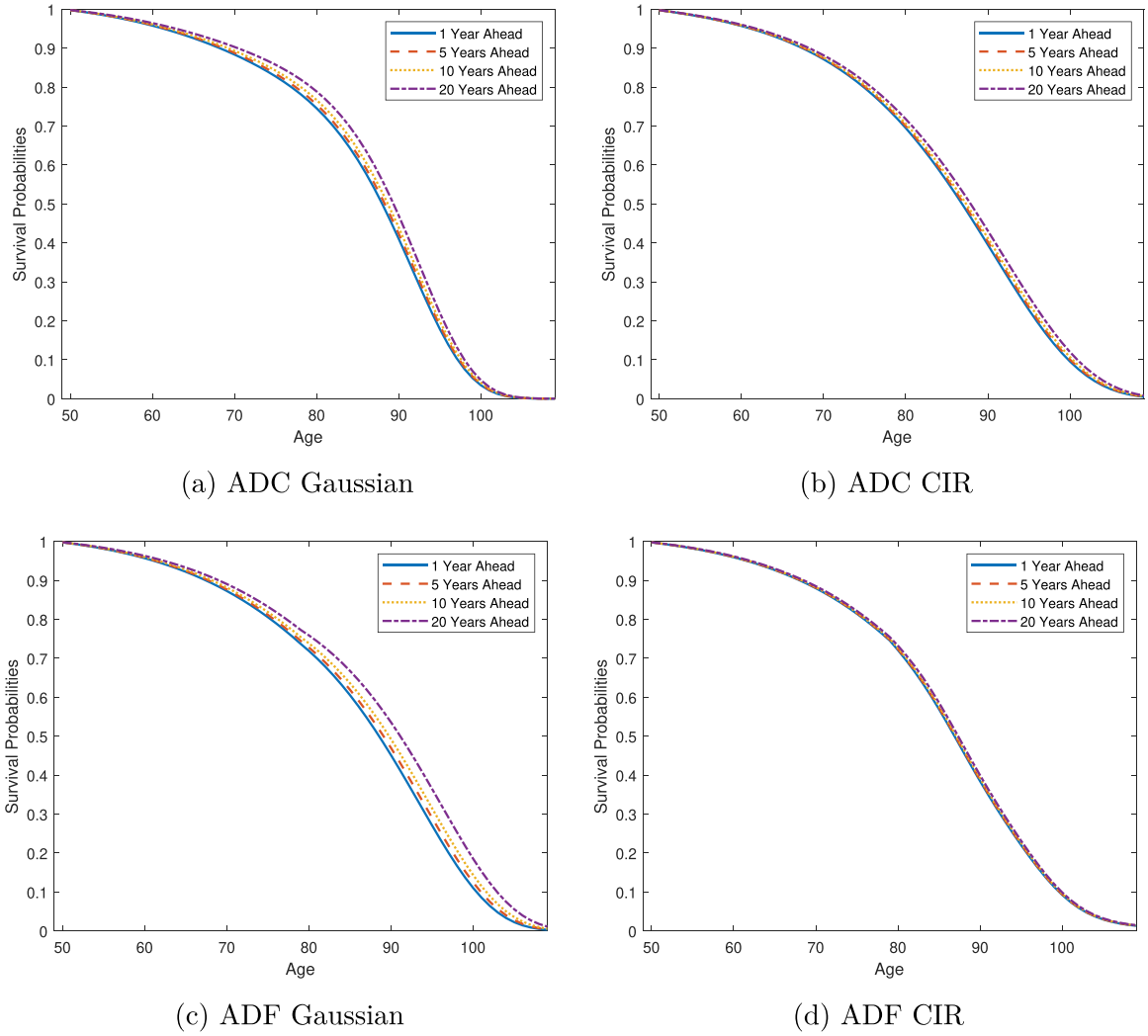
With our estimated parameters and using the equations in Subsection 2.4, we can calculate the estimated correlation in 2000 between cohorts aged 50, 60, 70, 80, and 90, which are the ones born in 1950, 1940, 1930, 1920, and 1910, respectively.

For the ADC Gaussian model, the cohort correlations for the cohorts aged 50, 60, 70, 80, and 90 in 2000 without regularization are displayed in Table 3.

Next, we wish to add regularization and make the estimated correlation matrix match the empirical one. We examine our models and find them to be applicable to other countries. Table 4 shows the results in Australia, Denmark, the UK, and the USA when we increase the regularization parameter λ^r. We measure the goodness of fit to the empirical cohort correlation matrix by the Frobenius norm, which is the square root of the sum of squared absolute errors of all elements.

For all four countries, the scale of the likelihood is around $20,000$, and the scale of the Frobenius norm is between 1 and 3. Therefore, we find that λ^r under 100 is too small to force the estimated correlation matrix to move toward the empirical one. Meanwhile, $\lambda^r = 10,000$ is too high and will worsen the goodness of fit to the mortality data very much. Therefore, we conclude that λ^r at the scale of $1,000$ is the most suitable and does help with reducing the Frobenius norm. Through experiments, $\lambda^r = 1,000$ achieves a balance between these two, allowing for a substantial reduction in the Frobenius norm without overly sacrificing goodness of fit. We also find that the ADC Gaussian model produces the lowest Frobenius norm for three of the four countries, except for Denmark. We also note that the model that gives the highest Frobenius norm differs for different countries.

Tables 5–8 show the cohort correlation matrices of the four models in calendar year 2000 using Australian male data when we set $\lambda^r = 1,000$. The results of all four models support that the cohort correlation reduces as age difference increases. The smallest correlation for a 40-year age difference is given by the ADC Gaussian model at 0.1621, while the largest is given by



(a) ADC Gaussian (b) ADC CIR

(c) ADF Gaussian (d) ADF CIR

FIGURE 10. Forecasted Australian Male Survival Probabilities: Models Estimated with All Available Incomplete-Cohort Data.

TABLE 3
Estimated Australian Male Cohort Correlations between Ages 50 and 90 from the ADC Gaussian Model, $\lambda^r = 0$

	50	60	70	80	90
50	1	0.995924	0.924646	0.692869	0.529981
60	0.995924	1	0.955199	0.754992	0.604218
70	0.924646	0.955199	1	0.915237	0.812978
80	0.692869	0.754992	0.915237	1	0.978672
90	0.529981	0.604218	0.812978	0.978672	1
norm	1.9730				

the ADC CIR model at 0.8303. Comparing the estimated cohort correlations with those in Jevtić et al. (2013), Chang and Sherris (2018), and Xu et al. (2020), it can be found that the estimated cohort correlations in Jevtić et al. (2013) are quite high and mostly above 0.95 for age differences from 0 to 50 so almost perfectly correlated, while in Xu et al. (2020) the estimated cohort correlations are mostly below 0.25. In Chang and Sherris (2018), the estimated cohort correlations fail to capture the

TABLE 4
Frobenius Norm between Empirical and Estimated Cohort Correlations for Different λ^r in Australia, Denmark, UK, and USA

ADC Gaussian	0	100	1,000
Australia	1.9730	1.5357	1.5021
Denmark	3.0147	2.8969	2.7325
UK	1.5770	1.5525	1.2703
USA	1.4447	1.3655	1.2740
ADC CIR	0	100	1,000
Australia	2.7650	2.7126	2.5312
Denmark	3.2246	3.1914	2.5928
UK	2.5116	2.1969	1.5992
USA	2.5840	2.5744	1.4328
ADF Gaussian	0	100	1,000
Australia	2.2861	2.2835	2.2573
Denmark	2.6526	2.5515	2.4479
UK	2.4338	2.3418	1.4221
USA	2.2764	2.1555	2.1300
ADF CIR	0	100	1,000
Australia	1.9761	1.9704	1.7233
Denmark	3.0622	3.0081	2.9776
UK	1.9470	1.9318	1.8995
USA	1.9583	1.9106	1.5886

TABLE 5
Estimated Australian Male Cohort Correlations between Age 50 and 90 from the ADC Gaussian Model, $\lambda^r = 1,000$

	50	60	70	80	90
50	1	0.999301	0.939419	0.448329	0.162052
60	0.999301	1	0.951232	0.480029	0.197140
70	0.939419	0.951232	1	0.727035	0.489610
80	0.448329	0.480029	0.727035	1	0.954633
90	0.162052	0.197140	0.489610	0.954633	1
norm	1.5021				

TABLE 6
Estimated Australian Male Cohort Correlations between Age 50 and 90 from the ADC CIR Model, $\lambda^r = 1,000$

	50	60	70	80	90
50	1	0.992099	0.962156	0.906185	0.830281
60	0.992099	1	0.988714	0.952054	0.893644
70	0.962156	0.988714	1	0.987122	0.950626
80	0.906185	0.952054	0.987122	1	0.987979
90	0.830281	0.893644	0.950626	0.987979	1
norm	2.5312				

TABLE 7
Estimated Australian Male Cohort Correlations between Age 50 and 90 from the ADF Gaussian Model, $\lambda^r = 1,000$

	50	60	70	80	90
50	1	0.983868	0.917261	0.796570	0.660012
60	0.983868	1	0.973715	0.891865	0.783748
70	0.917261	0.973715	1	0.971443	0.904608
80	0.796570	0.891865	0.971443	1	0.979911
90	0.660012	0.783748	0.904608	0.979911	1
Norm	2.2573				

TABLE 8
Estimated Australian Male Cohort Correlations between Age 50 and 90 from the ADF CIR Model, $\lambda^r = 1,000$

	50	60	70	80	90
50	1	0.975172	0.795162	0.525680	0.301748
60	0.975172	1	0.909699	0.700172	0.505138
70	0.795162	0.909699	1	0.932021	0.817651
80	0.525680	0.700172	0.932021	1	0.968788
90	0.301748	0.505138	0.817651	0.968788	1
norm	1.7233				

trend that they should decrease as the age difference increases, as reflected in historical data. Therefore, it can be concluded that the approach proposed in this article that incorporates cohort correlation in parameter estimation provides estimated cohort correlations that are the closest to historical values and captures the trend that cohort correlations decrease as the age difference increases.

7. CONCLUSIONS

This article improves the affine mortality models by introducing age dependence into the model, with both Gaussian and CIR settings. We propose two categories of models, namely ADC and ADF models that incorporate age dependence from the factor coefficients or the drift and volatility in the SDEs of the factors, respectively. We derive the SDEs of the factor loadings for the ADC models, and we obtain the analytical solutions to the factor loadings for the ADF models. Our models are calibrated to Australian male mortality data from age 50 to age 109, and our results show that adding age dependence into the model improves the goodness of fit of the models, especially at old ages.

Age-cohort mortality models are useful in actuarial applications since insurance products are issued to certain cohorts and are calculated on a cohort basis. However, existing literature utilizing age-cohort data to calibrate the model use only cohorts that have complete observation over the age horizon, making a great amount of data from the most recent cohorts unused. Our research extends the Kalman filtering algorithm to incorporate the more recent incomplete-cohort data in our model calibration, and we show that the out-of-sample forecasts improve compared to only using the complete cohort data. In particular, it reduces the underestimation of the survival probabilities around ages 70–80 compared with when only complete age-cohort data is used. Furthermore, our estimated parameters for the ADF models also provide evidence that both the drift and volatilities increase with age.

Moreover, our age-dependent models allow us to derive closed-form solutions to the instantaneous cohort correlation for all four models because both the age and the calendar times are specified. Also, the CIR models have the advantage of being time-inhomogeneous and are therefore truly cohort-specific. To better capture empirical cohort correlation, we regularize the parameter estimation method by penalizing the difference between the estimated and empirical cohort correlations. We find that adding regularization to parameter estimation produces more realistic estimated cohort correlations. Furthermore, the multi-country analysis shows that in general, the cohort correlation reduces as age difference increases. Future research could

extend this framework by allowing both the coefficients and factors to be age-dependent, thereby enhancing the ability to capture age-specific mortality dynamics. Additionally, the insights from this study on cohort correlations can be extended to mutual mortality-sharing products involving multiple cohorts within a single risk pool.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

FUNDING

The authors gratefully acknowledge financial support from the Australian Research Council (ARC) Centre of Excellence in Population Ageing Research (CEPAR) project number CE170100005, ARC Discovery Early Career Researcher Award (DECRA) project number DE200101266, and ARC Discovery Projects project number DP210101195.

ORCID

Yuxin Zhou http://orcid.org/0000-0001-9990-8622
Len Patrick Garces http://orcid.org/0000-0002-2737-7348
Yang Shen http://orcid.org/0000-0002-0559-2231
Michael Sherris http://orcid.org/0000-0002-8077-3649
Jonathan Ziveyi http://orcid.org/0000-0001-9029-8039

REFERENCES

Biffis E. 2005. Affine processes for dynamic mortality and actuarial valuations. Insur Math Econ. 37 (3):443–468. https://doi.org/10.1016/j.insmatheco.2005.05.003

Blackburn C, Sherris M. 2013. Consistent dynamic affine mortality models for longevity risk applications. Insur Math Econ. 53 (1):64–73. https://doi.org/10.1016/j.insmatheco.2013.04.007

Bohlmann G. 1899. Ein ausgleichungsproblem. Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse. 1899 (3):260–271.

Cairns AJ, Blake D, Dowd K. 2006. Pricing death: frameworks for the valuation and securitization of mortality risk. ASTIN Bull. 36 (1):79–120. https://doi.org/10.2143/AST.36.1.2014145

Cairns AJ et al. 2009. A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. North Am Actuar J. 13 (1):1–35. https://doi.org/10.1080/10920277.2009.10597538

Chang Y, Sherris M. 2018. Longevity risk management and the development of a value-based longevity index. Risks. 6 (1):10. https://doi.org/10.3390/risks6010010

Chen R-R, Scott L. 2003. Multi-factor Cox-Ingersoll-Ross models of the term structure: estimates and tests from a Kalman filter model. J Real Estate Financ Econ (Dordr). 27 (2):143–172. https://doi.org/10.1023/A:1024736903090

Chiu MC, Wang L, Wong HY. 2026. Long-range dependent mortality modeling with cointegration. Scand Actuar J. 2026 (1):60–93. https://doi.org/10.1080/03461238.2025.2503290

Christensen J, Diebold F, Rudebusch G. 2011. The affine arbitrage-free class of Nelson–Siegel term structure models. J Econom. 164 (1):4–20. https://doi.org/10.1016/j.jeconom.2011.02.011

Cox JC, Ingersoll JE, Jr, Ross SA. 1985. A theory of the term structure of interest rates. Econometrica. 53(2):385–407. https://doi.org/10.2307/1911242

Dahl M. 2004. Stochastic mortality in life insurance: market reserves and mortality-linked insurance contracts. Insur Math Econ. 35 (1):113–136. https://doi.org/10.1016/j.insmatheco.2004.05.003

Duffie D, Pan J, Singleton K. 2000. Transform analysis and asset pricing for affine jump-diffusions. Econometrica. 68 (6):1343–1376. https://doi.org/10.1111/1468-0262.00164

Geyer AL, Pichler S. 1999. A state-space approach to estimate and test multifactor Cox-Ingersoll-Ross models of the term structure. J Financ Res. 22 (1):107–130. https://doi.org/10.1111/j.1475-6803.1999.tb00717.x

Henderson R. 1924. A new method of graduation. Trans Actuar Soc Am. 25:29–40.

Huang Z, Sherris M, Villegas AM, Ziveyi J. 2022. Modelling USA age-cohort mortality: a comparison of multi-factor affine mortality models. Risks. 10 (9):183. https://doi.org/10.3390/risks10090183

Human Mortality Database. 2022. Australia, Denmark, U.K., and U.S.A. life tables. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). www.mortality.org (data downloaded in 2022).

Jevtić P, Luciano E, Vigna E. 2013. Mortality surface by means of continuous time cohort models. Insur Math Econ. 53 (1):122–133. https://doi.org/10.1016/j.insmatheco.2013.04.005

Jevtić P, Regis L. 2019. A continuous-time stochastic model for the mortality surface of multiple populations. Insur Math Econ. 88:181–195. https://doi.org/10.1016/j.insmatheco.2013.04.005

Kalman RE. 1960. A new approach to linear filtering and prediction problems. J Fluids Eng. 82(1):35–45. https://doi.org/10.1115/1.3662552

Luciano E, Vigna E. 2005. Non mean reverting affine processes for stochastic mortality. Working Paper. Available at: SSRN 724706.
Luciano E, Vigna E. 2008. Mortality risk via affine stochastic intensities: calibration and empirical relevance. Working Paper. Available at: MPRA 59627.
Njenga CN, Sherris M. 2011. Longevity risk and the econometric analysis of mortality trends and volatility. Asia-Pacific J Risk Insur. 5(2):2. https://doi.org/10.2202/2153-3792.1115
Piggott J, Valdez EA, Detzel B. 2005. The simple analytics of a pooled annuity fund. J Risk Insur. 72(3):497–520. https://doi.org/10.1111/j.1539-6975.2005.00134.x
Renshaw AE, Haberman S. 2006. A cohort-based extension to the Lee-Carter model for mortality reduction factors. Insur Math Econ. 38 (3):556–570. https://doi.org/10.1016/j.insmatheco.2005.12.001
Schrager DF. 2006. Affine stochastic mortality. Insur Math Econ. 38(1):81–97. https://doi.org/10.1016/j.insmatheco.2005.06.013
Smith J, Borchelt M, Maier H, Jopp D. 2002. Health and well–being in the young old and oldest old. J Soc Issues. 58 (4):715–732. https://doi.org/10.1111/1540-4560.00286
Ungolo F, Garces LPDM, Sherris M, Zhou Y. 2024. Estimation, comparison, and projection of multifactor age–cohort affine mortality models. North Am Actuar J. 28 (3):570–592. https://doi.org/10.1080/10920277.2023.2238793
Whittaker ET. 1922. On a new method of graduation. Proc Edinburgh Math Soc. 41:63–75. https://doi.org/10.1017/S0013091500077853
Wills S, Sherris M. 2008. Integrating financial and demographic longevity risk models: an Australian model for financial applications. UNSW Australian School of Business Research Paper, no. 2008ACTL05.
Xu Y, Sherris M, Ziveyi J. 2020. Continuous-time multi-cohort mortality modelling with affine processes. Scand Actuar J. 2020 (6):526–552. https://doi.org/10.1080/03461238.2019.1696223
Zhou H, Zhou KQ, Li X. 2022. Stochastic mortality dynamics driven by mixed frac-tional brownian motion. Insur Math Econ. 106:218–238. https://doi.org/10.1016/j.insmatheco.2022.07.006
Zhu N, Bauer D. 2022. Modeling the risk in mortality projections. Oper Res. 70 (4):2069–2084. https://doi.org/10.1287/opre.2021.2255

APPENDIX A. CHANGE OF MEASURE

A.1. Proof of Proposition 1: ADC Gaussian Model

Following Blackburn and Sherris (2013), we define the market price of risk $\Lambda^M(t)$ for the Gaussian model to be:

$$\Lambda^M(t) = \lambda^0 + \lambda^1 \mathbf{Z}(t),$$

where $\Lambda^M(t) \in \mathbb{R}^{3\times 1}, \lambda^0 \in \mathbb{R}^{3\times 1}$ and $\lambda^1 \in \mathbb{R}^{3\times 3}$. We assume that Novikov's condition is satisfied such that $E^P\left[\exp\left(\frac{1}{2}\int_0^T \Lambda^M(s)^2 ds\right)\right] < \infty$. By Girsanov's theorem, we have the following relationship between the risk-neutral measure Q and the real-world measure P:

$$\mathrm{d}W^Q(t) = \mathrm{d}W^P(t) + \Lambda^M(t)\mathrm{d}t,$$

where $W^P(t)$ is a 3×1 vector of three standard Brownian motions under the real-world measure P.
Then, the SDE of the factor $\mathbf{Z}(t)$ can be represented as:

$$\begin{aligned} d\mathbf{Z}(t) &= \left[K^Q\Theta^Q - K^Q\mathbf{Z}(t)\right]\mathrm{d}t + \Sigma[\lambda_0 + \lambda_1\mathbf{Z}(t)]\mathrm{d}t + \Sigma\mathrm{d}W^P(t) \\ &= \left[K^Q\Theta^Q + \Sigma\lambda_0\right]\mathrm{d}t - \left[K^Q - \Sigma\lambda_1\right]\mathbf{Z}(t)\mathrm{d}t + \Sigma\mathrm{d}W^P(t) \\ &= (K^Q - \Sigma\lambda_1)\left[(K^Q - \Sigma\lambda_1)^{-1}(K^Q\Theta^Q + \Sigma\lambda_0) - \mathbf{Z}(t)\right]\mathrm{d}t + \Sigma\mathrm{d}W^P(t) \\ &= K^P\left[\Theta^P - \mathbf{Z}(t)\right]\mathrm{d}t + \Sigma\mathrm{d}W^P(t), \end{aligned}$$

where $K^P = (K^Q - \Sigma\lambda_1)$, $\Theta^P = (K^Q - \Sigma\lambda_1)^{-1}(K^Q\Theta^Q + \Sigma\lambda_0)$, and $(K^Q - \Sigma\lambda_1)$ is assumed to be invertible. The above structure allows us to assume both Θ^Q and Θ^P are equal to zero when we set λ_0 to be a zero vector.

A.2. Proof of Proposition 1: ADC CIR Model

Following Huang et al. (2022), the market price of risk $\Lambda^M(t)$ for the CIR model is:

$$\begin{aligned}\Lambda^M(t) &= D(\mathbf{Z}(t))\lambda^0,\\ \mathrm{d}W^Q(t) &= \mathrm{d}W^P(t) + \Lambda^M(t)\mathrm{d}t,\end{aligned}$$

where $\lambda^0 = \begin{pmatrix}\lambda_1^0\\ \lambda_2^0\\ \lambda_3^0\end{pmatrix} \in \mathbb{R}^{3\times 1}$, and Novikov's condition $E^P\left[\exp\left(\frac{1}{2}\int_0^T \Lambda^M(s)^2 ds\right)\right] < \infty$ is assumed to be satisfied.

$$\begin{aligned}d\mathbf{Z}(t) &= K^Q\left[\Theta^Q - \mathbf{Z}(t)\right]\mathrm{d}t + \Sigma D(\mathbf{Z}(t))\left[\mathrm{d}W^P(t) + \Lambda^M(t)\mathrm{d}t\right]\\ &= K^Q\left[\Theta^Q - \mathbf{Z}(t)\right]\mathrm{d}t + \Sigma D(\mathbf{Z}(t))\left[D(\mathbf{Z}(t))\lambda^0\mathrm{d}t + \mathrm{d}W^P(t)\right]\\ &= \left[K^Q\Theta^Q - K^Q\mathbf{Z}(t) + \Sigma D^2(\mathbf{Z}(t))\lambda^0\right]\mathrm{d}t + \Sigma D(\mathbf{Z}(t))\mathrm{d}W^P(t)\\ &= \left[K^Q\Theta^Q - (K^Q - \Sigma\Lambda^0)\mathbf{Z}(t)\right]\mathrm{d}t + \Sigma D(\mathbf{Z}(t))\mathrm{d}W^P(t)\\ &= (K^Q - \Sigma\Lambda^0)\left[(K^Q - \Sigma\Lambda^0)^{-1}K^Q\Theta^Q - \mathbf{Z}(t)\right]\mathrm{d}t + \Sigma D(\mathbf{Z}(t))\mathrm{d}W^P(t)\\ &= K^P\left[\Theta^P - \mathbf{Z}(t)\right]\mathrm{d}t + \Sigma D(\mathbf{Z}(t))\mathrm{d}W^P(t),\end{aligned}$$

where

$$K^P = K^Q - \Sigma\Lambda^0, \quad \Theta^P = (K^Q - \Sigma\Lambda^0)^{-1}K^Q\Theta^Q = (K^P)^{-1}K^Q\Theta^Q, \quad \Lambda^0 = \begin{pmatrix}\lambda_1^0 & 0 & 0\\ 0 & \lambda_2^0 & 0\\ 0 & 0 & \lambda_3^0\end{pmatrix},$$

and $K^P = K^Q - \Sigma\Lambda^0$ is assumed to be invertible.

A.3. Proof of Proposition 2: ADF Gaussian Model

For cohorts in the same age group starting with age x, using the results in the ADC Gaussian model, the SDE of the factor $\mathbf{Z}(t,x)$ can be represented as:

$$\begin{aligned}d\mathbf{Z}(t,x) &= \left[K_x^{\mathrm{Q}}\Theta^Q - K_x^{\mathrm{Q}}\mathbf{Z}(t,x)\right]\mathrm{d}t + \Sigma_x\left[\lambda^0 + \lambda^1\mathbf{Z}(t,x)\right]\mathrm{d}t + \Sigma_x\mathrm{d}W^P(t)\\ &= \left(K_x^{\mathrm{Q}} - \Sigma_x\lambda^1\right)\left[(K_x^{\mathrm{Q}} - \Sigma_x\lambda^1)^{-1}(K_x^{\mathrm{Q}}\Theta^Q + \Sigma_x\lambda^0) - \mathbf{Z}(t,x)\right]\mathrm{d}t + \Sigma_x\mathrm{d}W^P(t)\\ &= K_x^{\mathrm{P}}\left[\Theta^P - \mathbf{Z}(t,x)\right]\mathrm{d}t + \Sigma_x\mathrm{d}W^P(t).\end{aligned}$$

where $\Theta^P = (K_x^{\mathrm{Q}} - \Sigma_x\lambda^1)^{-1}(K_x^{\mathrm{Q}}\Theta^Q + \Sigma_x\lambda^0)$, and $(K_x^{\mathrm{Q}} - \Sigma_x\lambda^1)$ is assumed to be invertible. The above structure allows us to assume both Θ^Q and Θ^P are equal to zero when we set $\lambda^0 = 0$.

A.4. Proof of Proposition 2: ADF CIR Model

Similar to the ADC CIR model, we set

$$\begin{aligned}\Lambda^M(t) &= D(\mathbf{Z}(t,x))\lambda_x^0,\\ \mathrm{d}W^Q(t) &= \mathrm{d}W^P(t) + \Lambda^M(t)\mathrm{d}t,\end{aligned}$$

where $\lambda_x^0 = \begin{pmatrix}\lambda_{1,x}^0\\ \lambda_{2,x}^0\\ \lambda_{3,x}^0\end{pmatrix} \in \mathbb{R}^{3\times 1}$, and it is chosen to make $\Lambda^M(t)$ indifferent between the two age groups.

Then, for cohorts in the same age group that start with age x,

$$\begin{aligned} d\mathbf{Z}(t,x) &= K_x^{\mathrm{Q}}\big[\Theta^{Q} - \mathbf{Z}(t,x)\big]\mathrm{d}t + \Sigma_x D(\mathbf{Z}(t,x))\big[\mathrm{d}W^{P}(t) + \Lambda^{M}(t)\mathrm{d}t\big] \\ &= \Big(K_x^{\mathrm{Q}} - \Sigma_x\Lambda_x^0\Big)\Big[(K_x^{\mathrm{Q}} - \Sigma_x\Lambda_x^0)^{-1}K_x^{\mathrm{Q}}\Theta^{Q} - \mathbf{Z}(t,x)\Big]\mathrm{d}t + \Sigma_x D(\mathbf{Z}(t,x))\mathrm{d}W^{P}(t) \\ &= K_x^{\mathrm{P}}\big[\Theta^{P} - \mathbf{Z}(t,x)\big]\mathrm{d}t + \Sigma_x D(\mathbf{Z}(t,x))\mathrm{d}W^{P}(t), \end{aligned}$$

where

$$K_x^{\mathrm{P}} = K_x^{\mathrm{Q}} - \Sigma_x\Lambda_x^0, \quad \Theta^{P} = (K_x^{\mathrm{P}})^{-1}K_x^{\mathrm{Q}}\Theta^{Q}, \quad \Lambda_x^0 = \begin{pmatrix} \lambda_{1,x}^0 & 0 & 0 \\ 0 & \lambda_{2,x}^0 & 0 \\ 0 & 0 & \lambda_{3,x}^0 \end{pmatrix},$$

and $K^{P} = K^{Q} - \Sigma\Lambda^0$ is assumed to be invertible. For the same reason as in the previous models, we assume both Θ^{Q} and Θ^{P} are equal to zero.

APPENDIX B. SDES OF THE FORCE OF MORTALITY FOR ADC MODELS

B.1. ADC Gaussian Model

For ADC Gaussian model, using Ito's lemma, the SDE for the mortality intensities of cohort aged $t-c$ at time t is:

$$\mathrm{d}\mu_{t-c}(t) = \frac{\partial\mu}{\partial t}\mathrm{d}t + \sum_{i=1}^{3}\left[\frac{\partial\mu}{\partial Z_i}\mathrm{d}Z_i(t) + \frac{1}{2}\frac{\partial^2\mu}{\partial Z_i^2}(\mathrm{d}Z_i(t))^2\right],$$

where

$$\begin{aligned} &\frac{\partial\mu}{\partial t} = \sum_{i=1}^{3}\frac{\partial g_i(t-c)}{\partial t}Z_i(t) \quad \text{(later we use } g_i'(t-c) \text{ to represent } \frac{\partial g_i(t-c)}{\partial t}\text{)}, \\ &\frac{\partial\mu}{\partial Z_i} = g_i(t-c), \\ &\frac{\partial^2\mu}{\partial Z_i^2} = 0. \end{aligned}$$

Therefore, the SDE of $\mathrm{d}\mu_{t-c}(t)$ becomes:

$$\begin{aligned} \mathrm{d}\mu_{t-c}(t) &= \sum_{i=1}^{3}\frac{\partial g_i(t-c)}{\partial t}Z_i(t)\mathrm{d}t + \sum_{i=1}^{3}g_i(t-c)\mathrm{d}Z_i(t) \\ &= \sum_{i=1}^{3}g_i'(t-c)Z_i(t)\mathrm{d}t + g_i(t-c)\Big[k_i^{\mathrm{P}}(\theta_{iP}^{\mathrm{P}} - Z_i(t))\Big]\mathrm{d}t + g_i(t-c)\sigma_i\mathrm{d}W_i^{\mathrm{P}}(t) \\ &= \sum_{i=1}^{3}g_i(t-c)\left[\left[\frac{g_i'(t-c)}{g_i(t-c)}Z_i(t) + \Big[k_i^{\mathrm{P}}(\theta_i^{\mathrm{P}} - Z_i(t))\Big]\right]\mathrm{d}t + \sigma_i\mathrm{d}W_i^{\mathrm{P}}(t)\right] \\ &= \sum_{i=1}^{3}g_i(t-c)\left[\left(k_i^{\mathrm{P}} - \frac{g_i'(t-c)}{g_i(t-c)}\right)\left[\frac{k_i^{\mathrm{P}}\theta_i^{\mathrm{P}}}{\left(k_i^{\mathrm{P}} - \frac{g_i'(t-c)}{g_i(t-c)}\right)} - Z_i(t)\right]\mathrm{d}t + \sigma_i\mathrm{d}W_i^{\mathrm{P}}(t)\right] \\ &= \sum_{i=1}^{3}g_i(t-c)\left[-\left(k_i^{\mathrm{P}} - \frac{g_i'(t-c)}{g_i(t-c)}\right)Z_i(t)\mathrm{d}t + \sigma_i\mathrm{d}W_i^{\mathrm{P}}(t)\right], \end{aligned}$$

where we can think $\frac{k_i^{\mathrm{P}}\theta_i^{\mathrm{P}}}{(k_i^{\mathrm{P}}-\frac{g_i'(t-c)}{g_i(t-c)})}$ as a mean reversion parameter, and it is 0 because we set $\theta_i^{\mathrm{P}}=0$, and we can set $\tilde{\lambda}_i(t-c)=\left(k_i^{\mathrm{P}}-\frac{g_i'(t-c)}{g_i(t-c)}\right)$ to simplify the equations.
Therefore, the SDE is simplified to be:

$$\mathrm{d}\mu_x(t)=\sum_{i=1}^{3}g_i(x)\left[-\tilde{\lambda}_i(t-c)Z_i(t)\mathrm{d}t+\sigma_i\mathrm{d}W_i^{\mathrm{P}}(t)\right].$$

B.2. ADC CIR Model

Similarly, for the CIR model, the SDE becomes

$$\begin{aligned}
\mathrm{d}\mu_{t-c}(t)&=\sum_{i=1}^{3}\frac{\partial g_i(t-c)}{\partial t}Z_i(t)\mathrm{d}t+\sum_{i=1}^{3}g_i(t-c)\mathrm{d}Z_i(t)\\
&=\sum_{i=1}^{3}g_i'(t-c)Z_i(t)\mathrm{d}t+g_i(t-c)\left[k_i^{\mathrm{P}}(\theta_i^{\mathrm{P}}-Z_i(t))\right]\mathrm{d}t+g_i(t-c)\sigma_i\sqrt{Z_i(t)}\mathrm{d}W_i^{\mathrm{P}}(t)\\
&=\sum_{i=1}^{3}g_i(t-c)\left[\left[\frac{g_i'(t-c)}{g_i(t-c)}Z_i(t)+\left[k_i^{\mathrm{P}}(\theta_i^{\mathrm{P}}-Z_i(t))\right]\right]\mathrm{d}t+\sigma_i\sqrt{Z_i(t)}\mathrm{d}W_i^{\mathrm{P}}(t)\right]\\
&=\sum_{i=1}^{3}g_i(t-c)\left[\left(k_i^{\mathrm{P}}-\frac{g_i'(t-c)}{g_i(t-c)}\right)\left[\frac{k_i^{\mathrm{P}}\theta_i^{\mathrm{P}}}{\left(k_i^{\mathrm{P}}-\frac{g_i'(t-c)}{g_i(t-c)}\right)}-Z_i(t)\right]\mathrm{d}t+\sigma_i\sqrt{Z_i(t)}\mathrm{d}W_i^{\mathrm{P}}(t)\right]\\
&=\sum_{i=1}^{3}g_i(t-c)\left[-\left(k_i^{\mathrm{P}}-\frac{g_i'(t-c)}{g_i(t-c)}\right)Z_i(t)\mathrm{d}t+\sigma_i\sqrt{Z_i(t)}\mathrm{d}W_i^{\mathrm{P}}(t)\right],
\end{aligned}$$

where we can think $\frac{k_i^{\mathrm{P}}\theta_i^{\mathrm{P}}}{(k_i^{\mathrm{P}}-\frac{g_i'(t-c)}{g_i(t-c)})}$ as a mean reversion parameter, and it is 0 because we set $\theta_i^{\mathrm{P}}=0$. We can also set $\tilde{\lambda}_i(t-c)=(k_i^{\mathrm{P}}-\frac{g_i'(t-c)}{g_i(t-c)})$ to simplify the equations.
Therefore, the SDE is simplified to be:

$$\mathrm{d}\mu_x(t)=\sum_{i=1}^{3}g_i(x)\left[-\tilde{\lambda}_i(t-c)Z_i(t)\mathrm{d}t+\sigma_i\sqrt{Z_i(t)}\mathrm{d}W_i^{\mathrm{P}}(t)\right].$$

APPENDIX C. ESTIMATED PARAMETERS WITH ALL AVAILABLE AUSTRALIAN INCOMPLETE-COHORT DATA

The estimated parameters of the four models with all incomplete Australian male mortality data, $\lambda^r=1,000$, and $\zeta^r=100,000$ for ADF models are displayed in Table 9. From Table 9, it can be noted that for ADC Gaussian model, b_2 is between 50 and 109, indicating that the second factor affects more of the middle age. Meanwhile, for ADC CIR model, b_2 is very close to 50, indicating that the second factor affects more of the younger age. For both ADF models, all the d_i are positive values, indicating that the volatility increases with age. We also have five out of six b_i positive, while a positive b_i means that the drift increases as age increases. We impose parameter constraints during calibration, which is capable of mitigating the issue of negative mortality rates as detailed in Jevtić et al. (2013); Jevtić and Regis (2019). Specifically, for ADC models, we restrict $k^P\in[0,1]$ and $a^g,b^g,\sigma,r\in[0,\infty)$. Meanwhile, for ADF models, we impose $k^P=a^P+b^Ph_x\in[0,1]$ for each age group, together with $r\in[0,\infty]$. These constraints improve numerical stability and reduce the risk of economically implausible mortality dynamics during estimation. The probabilities of negative

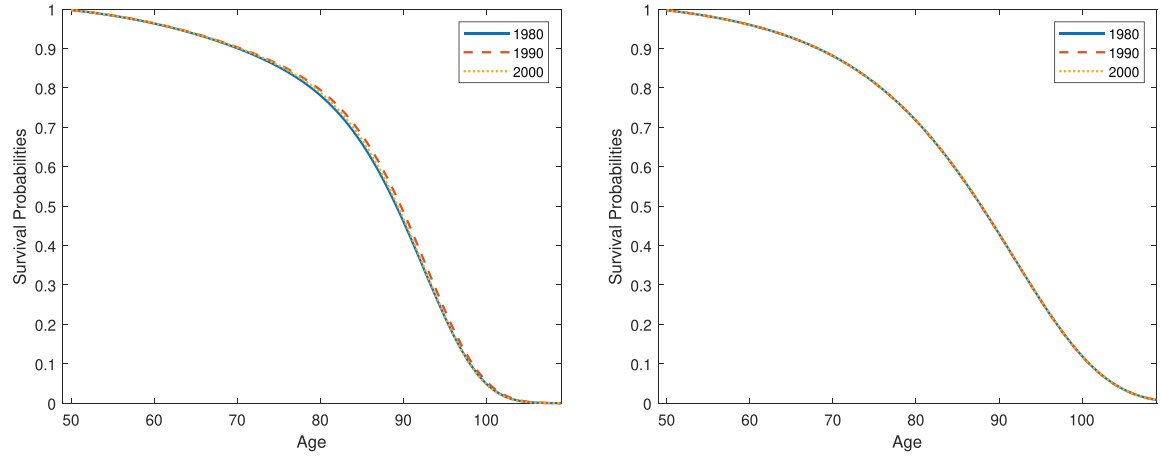
TABLE 9
Estimated Parameters with Australian Data

AUS	ADC Gaussian	ADC CIR		ADF Gaussian	ADF CIR
a_1^g	0.0730	5.10E-06	a_1^P	−3.66E-05	5.57E-03
a_2^g	0.0029	0.0014	a_2^P	1.88E-02	9.24E-05
a_3^g	1.36E-05	0.0318	a_3^P	1.77E-03	3.14E-03
b_2^g	91.8980	50.0911	b_1^P	2.93E-04	1.05E-04
k_1^P	1.07E-05	1.13E-03	b_2^P	2.17E-06	7.75E-06
k_2^P	2.88E-05	6.58E-04	b_3^P	1.33E-04	−2.58E-05
k_3^P	1.15E-02	5.53E-03	c_1	−9.5987	−10.6575
k_1^Q	−0.1989	−0.0157	c_2	−6.1007	−3.4776
k_2^Q	−0.0249	−0.1309	c_3	−7.7135	−5.1581
k_3^Q	−0.0835	−0.0648	d_1	0.0512	0.1128
σ_1	1.51E-03	0.0040	d_2	0.0022	0.0151
σ_2	1.33E-05	0.0030	d_3	0.0045	0.0170
σ_3	4.52E-07	0.0007	r_1	4.44E-09	2.28E-09
r_1	8.09E-17	1.86E-17	r_2	0.2253	0.2368
r_2	0.5105	0.6184	r_c	4.37E-08	4.71E-08
r_c	3.29E-07	2.43E-07	a_1^Q	−2.98E-05	−1.99E-02
			a_2^Q	4.16E-02	3.44E-02
			a_3^Q	1.69E-03	3.53E-02
			b_1^Q	−7.83E-04	1.20E-03
			b_2^Q	1.82E-03	−3.41E-03
			b_3^Q	−1.62E-03	−3.05E-03

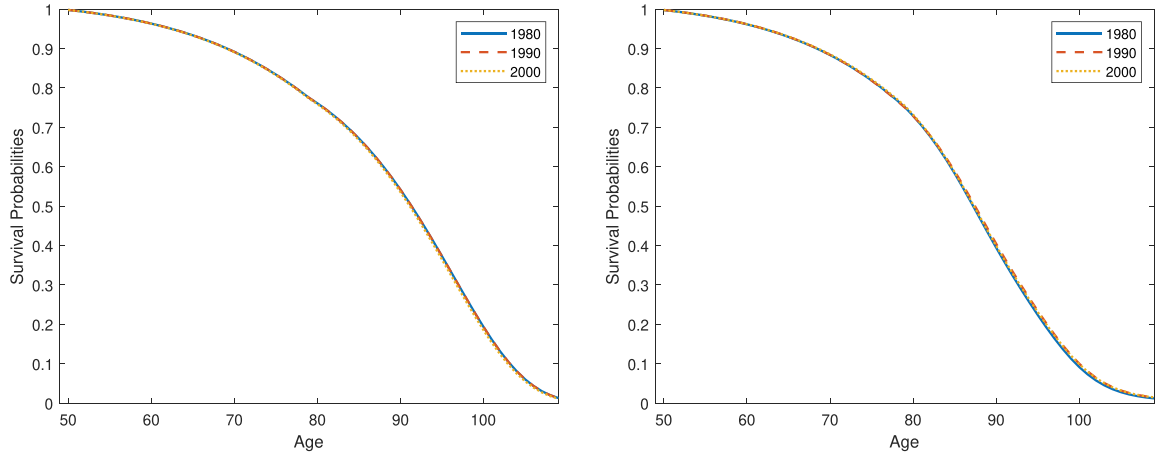
mortality rates from 10^6 simulations of the Australian male cohort born in 1971 are 0 and 0.0027 for ADC Gaussian and ADF Gaussian models respectively, which is negligible and is consistent with models with Gaussian specification (Luciano and Vigna 2005; Jevtić et al. 2013; Jevtić and Regis 2019; Xu et al. 2020; Zhu and Bauer 2022). Moreover, we refer to the magnitude of the estimated parameter values in the literature (Blackburn and Sherris 2013; Jevtić and Regis 2019; Huang et al. 2022; Ungolo et al. 2024) as the initial values of the estimation to improve robustness.

APPENDIX D. COMPARISON OF 20-YEAR AHEAD FORECASTS FOR CALENDAR YEARS 1980, 1990, AND 2000

To examine the robustness of model performance with respect to the choice of calendar year used to compute the cohort correlation in Equations (19) and (20), we compare 20-year ahead forecasted survival probabilities based on cohort correlations calculated using calendar years 1980, 1990, and 2000. As shown in Figure 11, the differences in the forecasted survival probabilities is negligible, indicating that the results are robust to the choice of calendar year used to compute the cohort correlation.



(a) ADC Gaussian (b) ADC CIR



(c) ADF Gaussian (d) ADF CIR

FIGURE 11. Comparison of 20-Year Ahead Forecasted Survival Probabilities for Calendar Years 1980, 1990, and 2000 Used to Calculate Cohort Correlation.