ARC Centre of Excellence in Population Ageing Research

Working Paper 2023/09

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

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

---

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

Yuxin Zhou*, Len Patrick Garces†, Yang Shen‡, Michael Sherris§, Jonathan Ziveyi¶

May, 2023

Abstract

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 CIR-type models for each category of age-dependent models. Both categories of age-dependent models involve age and calendar time, which in turn specifies 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 regularise the parameter estimation by adding a penalty term which penalises 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. We calibrate the mortality models to data from multiple countries which include Australia, Denmark, UK, and USA to assess and compare in-sample fit and forecasting performance. 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 regularisation 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.

Keywords: Mortality model, Age-dependent, Multi-cohort, Cohort correlation, Incomplete cohort, Affine, Regularisation

JEL Classifications: G22, C13, C22, C52, C53, J11

*School of Risk & Actuarial Studies, Australian Research Council Centre of Excellence in Population Ageing Research (CEPAR), UNSW Sydney, email: yuxin.zhou@unsw.edu.au

†CEPAR, UNSW Sydney and School of Mathematical and Physical Sciences, University of Technology Sydney, email: LenPatrickDominic.Garces@uts.edu.au

‡School of Risk & Actuarial Studies, CEPAR, UNSW Sydney, email: y.shen@unsw.edu.au

§School of Risk & Actuarial Studies, CEPAR, UNSW Sydney, email: m.sherris@unsw.edu.au

¶School of Risk & Actuarial Studies, CEPAR, UNSW Sydney, email: j.ziveyi@unsw.edu.au

1 Introduction

Mortality modelling is 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 model 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 are needed to fully explain the evolution of mortality rates. Following Schrager (2006) and Njenga and Sherris (2011), 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) at old ages. 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 modelled 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. However, very few models in the literature consider age dependence. Therefore, we include age dependence in the model and find improvements in the goodness of fit, as evidenced by the analysis in this paper.

Another issue is that previous work 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. On one hand, Huang et al. (2022) and Ungolo et al. (2023) use age-cohort data to calibrate the mortality models as age-cohort mortality models are more suitable for pricing insurance products which are typically issued to individuals belonging to different cohorts. On the other hand, 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 this issue, we propose age-dependent affine mortality models, 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. Wills and Sherris (2008) and Chang and Sherris (2018) make initial attempts to

include age dependence in affine mortality models. 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. Our improved models avoid this issue by introducing dependence on both age and calendar time at the same time.

We also improve the estimation of cohort correlation by including regularisation in Kalman filtering. The difference between the empirical cohort correlation and the estimated cohort correlation is included in the objective function as a penalty term. We test different scales of the penalty parameter to decide the most suitable one that can balance the goodness of fit to the empirical mortality data and empirical cohort correlation matrix.

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 modelling (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, utilising incomplete cohort data leads to more accurate out-of-sample forecasts. To address this issue, we improve the Kalman filtering algorithm to incorporate incomplete cohort data in the calibration process. This allows us to start our forecast from more recent cohorts as well as provide better forecasting results.

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

2 Age-Dependent Affine Mortality Models

We propose two categories of age-dependent affine mortality models. The first category of models makes the coefficients of factors age-dependent, namely the age-dependent coefficient (ADC) model. Meanwhile, the other category of models makes the factors themselves age-dependent through allowing age-dependent drift and volatility terms in the stochastic differential equation (SDE), thus called age-dependent factor (ADF) model. Define a probability space (Ω, $\mathcal{F}$, $\mathcal{F}_t$, Q) in an arbitrage-free market, where Ω is the sample space, $\mathcal{F}$ is the σ-algebra, $\mathcal{F}_t$ is the σ-algebra generated by the information up to time t and satisfies the usual conditions in Williams (1991), and the risk-neutral measure Q is the measure such that the survival probability for the cohort born in year c and aged $x = t - c$ at time t to survive to time T is represented as:

$$\begin{aligned} S_x(t,T) =& E^Q\left[e^{-\int_t^T \mu_{s-c}(s)\mathrm{d}s} \mid \mathcal{F}_t\right] \\ =& \mathrm{e}^{B_1(t,T)Z_1(t)+B_2(t,T)Z_2(t)+B_3(t,T)Z_3(t)+A(t,T)}. \end{aligned} \tag{1}$$

The general frameworks of the two categories of models are described as follows:

Definition 1. For age-dependent coefficient (ADC) models, the force of mortality for cohort

born in year c and aged $x = t - c$ at time t is defined as $\mu_x(t) = \mu_{t-c}(t)$:

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

where the coefficient $g_i(t-c)$ is dependent on the age $x = t - c$ that is increasing with time t for a given cohort born in year c.

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

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

where the factors $Z_i(t, x)$ are dependent on the age x through the age-dependent drift and volatility terms in the SDE. Unlike ADC models, we separate the data into different age groups, and the age x remains constant within one age group, which we will explain in detail in the following sections.

2.1 ADC Models

The force of mortality $\mu_x(t)$ of 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:

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

where a_1, a_2, a_3, and b_2 are positive constants, so that each factor affects different ages differently. Therefore, the coefficient $g_1(t-c)$ affects more of the younger ages within the age range, $g_2(t-c)$ affects more of the middle ages, and $g_3(t-c)$ affects more of the older ages. For the factors $\mathrm{Z}_i(t)$, we allow them to be either a Gaussian process or a CIR process.

2.1.1 ADC Gaussian Model

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

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

where for $i = 1, 2, 3$, k_i^Q are the transition parameters, θ_i^Q are the mean-reverting parameters, σ_i are the volatility parameters, and $W_i^Q(t)$ are standard Brownian motions under the risk-neutral measure Q.

Using the Feynman-Kac formula on Equation (1) with S as the short-hand notation for $S_{t-c}(t, T)$

yields:

$$\frac{\partial S}{\partial t}+\sum_{i=1}^{3}\frac{\partial S}{\partial Z_i}(k_i^Q(\theta_i^Q-Z_i))+\frac{1}{2}\sum_{i=1}^{3}\sigma_i^2\frac{\partial^2 S}{\partial Z_i^2}-\mu_{t-c}(t)S=0. \tag{4}$$

Meanwhile, from Equation (1) we also have:

$$\begin{aligned}\frac{\partial S}{\partial t}=&\left[\frac{\partial A}{\partial t}+\sum_{i=1}^{3}\frac{\partial B_i}{\partial t}Z_i\right]S,\\ \frac{\partial S}{\partial Z_i}=&B_iS,\\ \frac{\partial^2 S}{\partial Z_i^2}=&B_i^2S.\end{aligned}$$

Substituting into Equation (4) yields:

$$\frac{\partial A}{\partial t}+\sum_{i=1}^{3}B_ik_i^Q\theta_i^Q+\frac{1}{2}\sum_{i=1}^{3}\sigma_i^2B_i^2+\sum_{i=1}^{3}\left[\frac{\partial B_i}{\partial t}-B_ik_i^Q-g_i(t-c)\right]Z_i=0. \tag{5}$$

We want the coefficients of Z_i and the constant term to be 0 for Equation (5) to hold. Therefore, we derive the linear system of ODEs to solve for $A(t,T)$ and $B(t,T)$:

$$\begin{aligned}\frac{\partial A}{\partial t}=&-\sum_{i=1}^{3}B_ik_i^Q\theta_i^Q-\frac{1}{2}\sum_{i=1}^{3}\sigma_i^2B_i^2,\\ \frac{\partial B_i}{\partial t}=&B_ik_i^Q+g_i(t-c)\quad(\text{for } i=1,2,3).\end{aligned} \tag{6}$$

The terminal conditions are $A(T,T)=0$ and $B_i(T,T)=0$ for all models in this paper as in Christensen et al. (2011) and Blackburn and Sherris (2013). Due to the existence 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 using inbuilt ODE solvers which comes with a variety of software applications, from which 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:

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

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

Similar to the derivation in Subsection 2.1.1, using the Feynman-Kac formula on Equation (1),

we obtain the following Riccati system of ODEs to solve for $A(t,T)$ and $B(t,T)$:

$$\begin{aligned}\frac{\partial A}{\partial t} &= -\sum_{i=1}^{3} B_i k_i^Q \theta_i^Q, \\ \frac{\partial B_i}{\partial t} &= -\frac{1}{2}\sigma_i^2 B_i^2 + B_i k_i^Q + g_i(t-c) \quad \text{(for i=1, 2, 3).}\end{aligned} \tag{7}$$

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 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 (3). We let $\mathbf{Z}(t,x)$ be a 3×1 vector of the age-dependent factors $Z_i(t,x)$. The SDE of vector $\mathbf{Z}(t,x)$ under the Q measure is:

$$\mathrm{d}\mathbf{Z}(t,x) = \Lambda^Q\left(t,x,\mathbf{Z}(t,x)\right)\mathrm{d}t + \Sigma\left(t,x,\mathbf{Z}(t,x)\right)\mathrm{d}W_t^Q,$$

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 age x, and W_t^Q is a 3×1 standard Brownian motion. We also allow the factors $Z_i(t,x)$ 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:

$$\Lambda^Q\left(t,x,\mathbf{Z}(t,x)\right) = \begin{pmatrix} a_1^Q + b_1^Q h_x & 0 & 0 \\ 0 & a_2^Q + b_2^Q h_x & 0 \\ 0 & 0 & a_3^Q + b_3^Q h_x \end{pmatrix} \left[\begin{pmatrix} \theta_1^Q \\ \theta_2^Q \\ \theta_3^Q \end{pmatrix} - \begin{pmatrix} Z_1(t,x) \\ Z_2(t,x) \\ Z_3(t,x) \end{pmatrix} \right],$$

$$\Sigma\left(t,x,\mathbf{Z}(t,x)\right) = \begin{pmatrix} e^{c_1+d_1 h_x} & 0 & 0 \\ 0 & e^{c_2+d_2 h_x} & 0 \\ 0 & 0 & e^{c_3+d_3 h_x} \end{pmatrix}.$$

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:

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

where $k_{i,x}^Q = a_i^Q + b_i^Q h_x$ for $i = 1, 2, 3$.

2.2.2 ADF CIR Model

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

$$\Lambda^Q\left(t,x,\mathbf{Z}(t,x)\right) = \begin{pmatrix} a_1^Q+b_1^Q h_x & 0 & 0 \\ 0 & a_2^Q+b_2^Q h_x & 0 \\ 0 & 0 & a_3^Q+b_3^Q h_x \end{pmatrix}\left[\begin{pmatrix} Z_1(t,x) \\ Z_2(t,x) \\ Z_3(t,x) \end{pmatrix} - \begin{pmatrix} \theta_1^Q \\ \theta_2^Q \\ \theta_3^Q \end{pmatrix}\right],$$

$$\Sigma\left(t,x,\mathbf{Z}(t,x)\right)$$
$$= \begin{pmatrix} e^{c_1+d_1h_x} & 0 & 0 \\ 0 & e^{c_2+d_2h_x} & 0 \\ 0 & 0 & e^{c_3+d_3h_x} \end{pmatrix}\begin{pmatrix} \sqrt{Z_1(t,x)} & 0 & 0 \\ 0 & \sqrt{Z_2(t,x)} & 0 \\ 0 & 0 & \sqrt{Z_3(t,x)} \end{pmatrix}.$$

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:

$$B_i(t,T,x) = -\frac{2\left(e^{\gamma_{i,x}(T-t)}-1\right)}{\left(k_{i,x}^Q+\gamma_{i,x}\right)\left(e^{\gamma_{i,x}(T-t)}-1\right)+2\gamma_{i,x}},$$
$$A(t,T,x) = \sum_{i=1}^{3}\frac{2k_{i,x}^Q\theta_i^Q}{(e^{c_i+d_ih_x})^2}\ln\left[\frac{2\gamma_{i,x}\exp\left(\frac{\left(k_{i,x}^Q+\gamma_{i,x}\right)(T-t)}{2}\right)}{\left(k_{i,x}^Q+\gamma_{i,x}\right)\left(e^{\gamma_{i,x}(T-t)}-1\right)+2\gamma_{i,x}}\right],$$

where $\gamma_{i,x} = \sqrt{(k_{i,x}^Q)^2 + 2(e^{c_i+d_ih_x})^2}$ 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:

$$\mathrm{d}W_t^Q = \mathrm{d}W_t^P + \Lambda_t \mathrm{d}t,$$

where W_t^P is a 3×1 vector of three standard Brownian motions under the real-world measure P, and Λ_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:

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

where $\Lambda_t \in \mathbb{R}^{3\times 1}, \lambda^0 \in \mathbb{R}^{3\times 1}$ and $\lambda^1 \in \mathbb{R}^{3\times 3}$.

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

$$d\boldsymbol{Z}(t) = K^P \left[\Theta^P - \boldsymbol{Z}(t)\right] \mathrm{d}t + \Sigma \mathrm{d}W_t^P,$$

and for ADC CIR model, the SDE after change of measure can be represented as:

$$d\boldsymbol{Z}(t) = K^P \left[\Theta^P - \boldsymbol{Z}(t)\right] \mathrm{d}t + \Sigma D\left(\boldsymbol{Z}(t), t\right) \mathrm{d}W_t^P.$$

where K^P is a 3×3 diagonal matrix, Θ^P is a 3×1 vector, and W_t^P 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 Appendix 1.

We assume both Θ^Q and Θ^P are zero vectors so that the Gaussian factor processes are mean-reverting to zero, not mean-reverting, or random walks (Blackburn and Sherris, 2013), and the CIR factor processes are mean-reverting to zero.

Proof. See Appendix 1. □

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 ADF Gaussian model, the SDE of the factor $\boldsymbol{Z}(t, x)$ after change of measure can be represented as:*

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

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

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

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:

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

Proof. See Appendix 1. □

2.4 Cohort Correlation

Cohort correlation can be easily modelled 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 ADC Gaussian Model

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

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

$$\mathrm{d}\mu_x(t) = \sum_{i=1}^{3} g_i(t-c)\left[-\left(k_i^P - \frac{g_i'(t-c)}{g_i(t-c)}\right) Z_i(t)\mathrm{d}t + \sigma_i \mathrm{d}W_i^P(t)\right],$$

and we can set $\tilde{\lambda}_i(t-c) = \left(k_i^P - \frac{g_i'(t-c)}{g_i(t-c)}\right)$ *for notational convenience.*

Proof. See Appendix 2. □

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

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

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:

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

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

2.4.2 ADC CIR Model

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

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

Proof. See Appendix 2. □

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

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

where

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

Proof. Calculate the correlation between the SDEs 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 ADF Gaussian Model

Lemma 3. *For 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:*

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

Proof. Use Ito's Lemma on Equation (3). □

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

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

Proof. Calculate the correlation between the SDEs 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 ADF CIR Model

Lemma 4. *For 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:*

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

Proof. Use Ito's Lemma on Equation (3). □

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

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

where

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

Proof. Same as the proof in Proposition 5. □

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, UK, and USA obtained from the Human Mortality Database (2022) (HMD), will be 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 UK, and 1922-2020 for USA based on data availability.

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:

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

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.

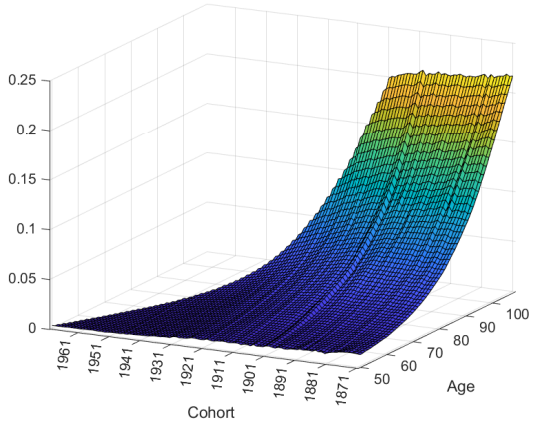


Figure 1: Australian cohort average force of mortality for males born between 1871 and 1970, from age 50 to 109.

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 length of the time range, $N_x = x_{max} - x_0 + 1$ the length of the age range, and $N_c = c_{max} - c_0 + 1$ the length of the cohort range.

Figure 1 is an 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 triangle-shaped incomplete cohort data after the complete cohorts.

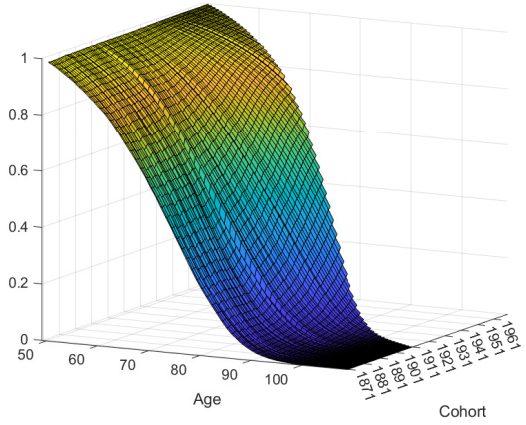


Figure 2: Australian cohort survival probabilities for males born between 1871 and 1970, from age 50 to 109.

Figure 2 shows the full cohort data of survival probabilities that includes the triangle incomplete cohort data. From the figure, we observe that the mortality has improved significantly for cohorts born after 1911, especially around age 60-80. However, for these cohorts, we only have incomplete observations. So 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 power of forecasting. 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 proposed models. Kalman filtering can be used to estimate the parameters in the model by sequentially updating the parameters and maximising 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 model's forecasting ability. We also add regularisation 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 regularisation 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

Transforming Equation (1) into vector representation, the average force of mortality for all proposed models can be represented as:

$$\bar{\mu}_x(t,T) = -\frac{1}{T-t}\log\left[S_x(t,T)\right] = \mathbf{B}'\mathbf{Z}_t + \mathbf{A}, \tag{8}$$

where $\mathbf{B} = -\frac{1}{T-t}[B_1(t,T), B_2(t,T), B_3(t,T)]'$, $\mathbf{A} = -\frac{1}{T-t}A(t,T)$, and $\mathbf{Z}_t = [Z_1(t), Z_2(t), Z_3(t)]'$.

We introduce measurement error ε_t into Equation (8) as in Blackburn and Sherris (2013) to obtain the measurement equation, represented as:

$$\bar{\mu}_x(t,T) = \mathbf{B}'\mathbf{Z}_t + \mathbf{A} + \varepsilon_t.$$

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:

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

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

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

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

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

$$Q = \int_0^{\Delta t} e^{-K^P s}\Sigma\Sigma' e^{-(K^P)' s} ds,$$

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

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

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

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 are utilising 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, we use three parameters r_1, r_2, and r_c to construct the equation below to represent the increasing measurement error with age

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

where r_1, r_2 and r_c only take positive values.

4.2 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:

$$\begin{aligned}\mathbf{Z}_{t|t-1} &= \mathbb{E}\left[\mathbf{Z}_t \mid Y_{t-1}\right] = \Phi(\psi)\mathbf{Z}_{t-1},\\ \Sigma_{t|t-1} &= \Phi(\psi)\Sigma_{t-1}\Phi(\psi)' + Q(\psi).\end{aligned}$$

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

$$\begin{aligned}\mathbf{Z}_t &= \mathbb{E}\left[\mathbf{Z}_t \mid Y_t\right] = \mathbf{Z}_{t|t-1} + \Sigma_{t|t-1}\mathbf{B}(\psi)' F_t^{-1} v_t,\\ \Sigma_t &= \Sigma_{t|t-1} - \Sigma_{t|t-1}\mathbf{B}(\psi)' F_t^{-1}\mathbf{B}(\psi)\Sigma_{t|t-1},\end{aligned}$$

where

$$\begin{aligned}v_t &= y_t - \mathbb{E}\left[y_t \mid Y_{t-1}\right] = y_t - \mathbf{A}(\psi) - \mathbf{B}(\psi)\mathbf{Z}_{t|t-1},\\ F_t &= \operatorname{cov}(v_t) = \mathbf{B}(\psi)\Sigma_{t|t-1}\mathbf{B}(\psi)' + R(\psi),\end{aligned}$$

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

$$\log l\left(y_1,\ldots,y_T;\psi\right) = \sum_{t=1}^{T}\left(-\frac{N}{2}\log(2\pi) - \frac{1}{2}\log\left(F_t\right) - \frac{1}{2}v_t' F_t^{-1} v_t\right),$$

and the optimal parameter set is the one that maximises the log-likelihood function. The built-in MATLAB optimisation toolbox will be used to maximise 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 need to be sufficient 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 error (MAPE) starts exploding from age 80 in our experiments. 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 a reasonable number.

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 optimisation. The objective function is represented as:

$$\sum_{g\in G} \log l\left(y_1, \ldots, y_T; \psi\right) = \sum_{g\in G}\sum_{t=1}^{T} w_g \left(-\frac{N}{2}\log(2\pi) - \frac{1}{2}\log\left(F_t\right) - \frac{1}{2}v_t' F_t^{-1} v_t\right),$$

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 filtering to maximise the aggregate likelihood function which yields a set of constant a_i, b_i, c_i, d_i for all age groups.

One constraint imposed in Kalman filtering is $0 < a_i^P + b_i^P h_x < 1$ 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.

4.3.1 Kalman Filtering with Regularisation

Regularisation of Cohort Correlation
The objective function is represented as:

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

where $\hat{f}(i, j)$ is the estimated cohort correlation, $f(i, j)$ is the empirical cohort correlation, and λ is the regularisation parameter controlling the weight of the penalty.

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 horizon 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 red lines in Figure 3.

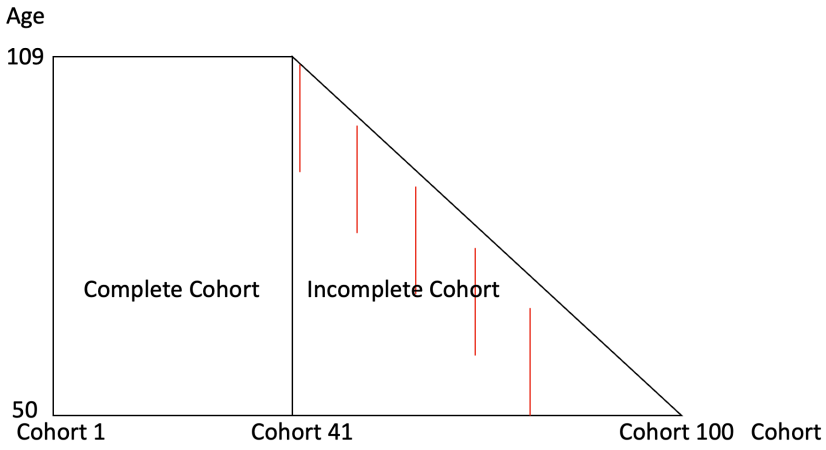


Figure 3: Illustration of calculating empirical cohort correlation: Red lines represent the selected differenced mortality rates for calculating the correlation.

Regularisation 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 originated in Bohlmann (1899), Whittaker (1922) and Henderson (1924), we add an additional penalty term to the objective function of ADF models as the following:

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

where ζ controls the weight for the smoothness regularisation penalty, ∇ represents the backward differencing operator $\nabla \hat{S}_x(t,T) = \hat{S}_x(t,T) - \hat{S}_x(t,T-1)$ and $\nabla^2 \hat{S}_x(t,T) = \nabla \hat{S}_x(t,T) - \nabla \hat{S}_x(t,T-1)$, where $\hat{S}_x(t,T)$ is the estimated survival probability with the parameter set. We select $\zeta = 100,000$ because the smoothness term is around the scale of 10^{-2}.

5 Estimation Results

5.1 Model Estimation with Complete-Cohort Data

Firstly, we calibrate the proposed models only with the complete cohort data to compare with the existing models in the literature. We use mean absolute percentage error (MAPE) to evaluate the goodness of fit of the model. 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 calibrated models using Australian male data are shown below:

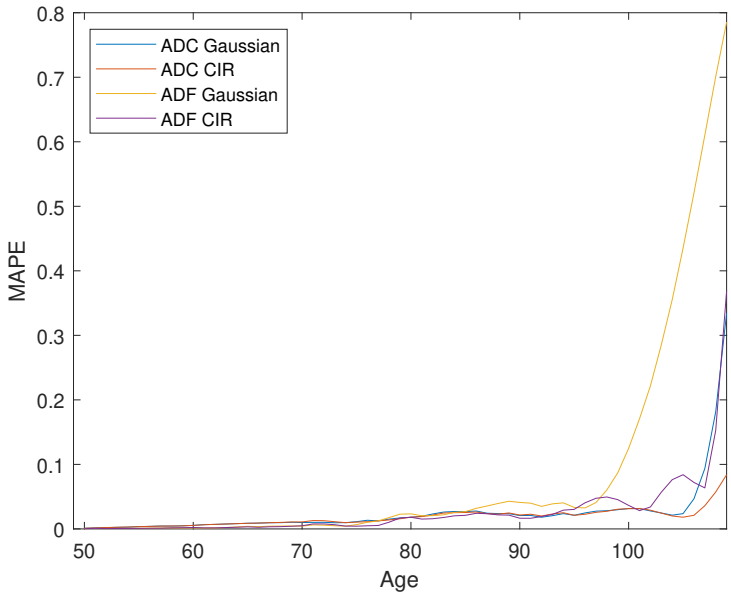
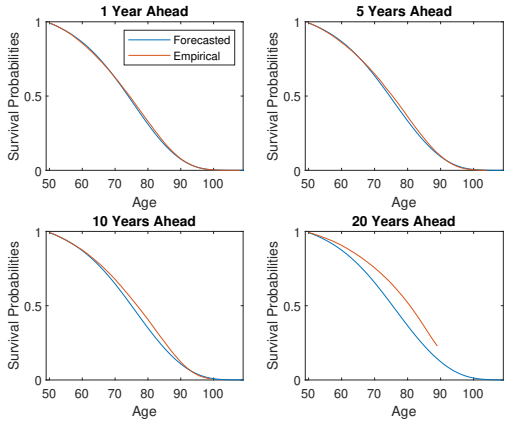


Figure 4: MAPE for Australian male survival probabilities: Models estimated with complete cohort data.

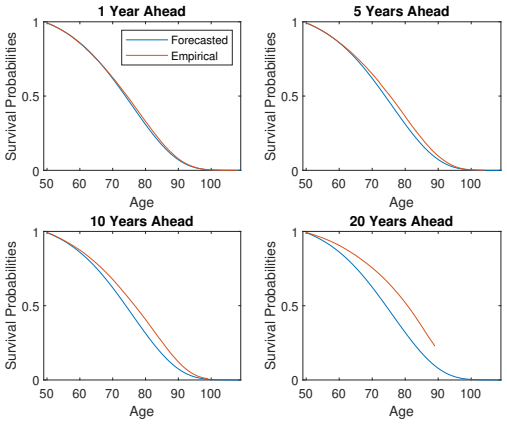
We can see that the MAPE at age 100 is around 3% for the ADC Gaussian, ADC CIR, and ADF CIR models, and around 12% for the ADF Gaussian model. We compare our MAPEs at age 100 to the models in the literature whose maximum age is 100 and do not consider age dependence, and we find that the MAPEs in our age-dependent models are much lower. 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 at age 109 are higher 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 causing a large relative 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

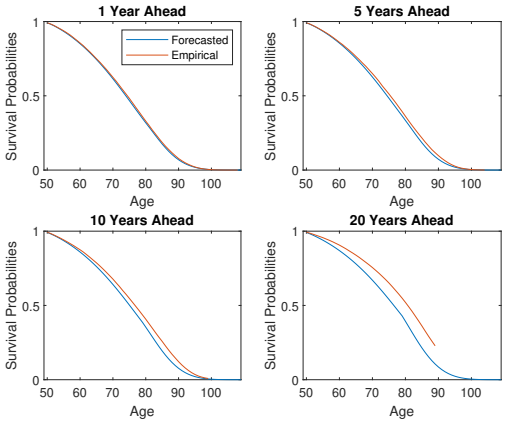
We perform the out-of-sample test by simulating via the transition equation. This test is performed for all models calibrated to Australian data. For each test, we perform the $1, 5, 10$, and 20 years ahead forecast, and compare the forecast with the incomplete empirical data. The results are shown in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



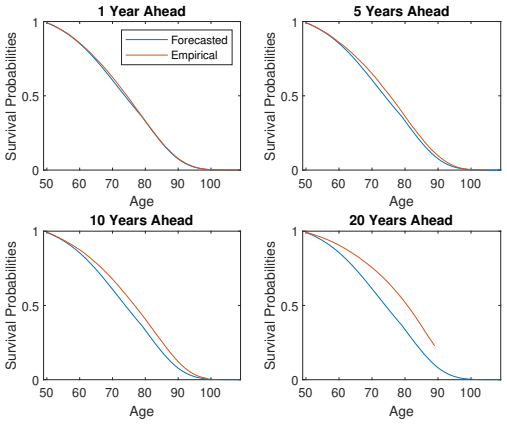
(a) ADC Gaussian



(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.

From the above comparisons, we note that the models have good out-of-sample performance in a short period horizon (that is 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 that has 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 model, 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 perform the calibration, 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.

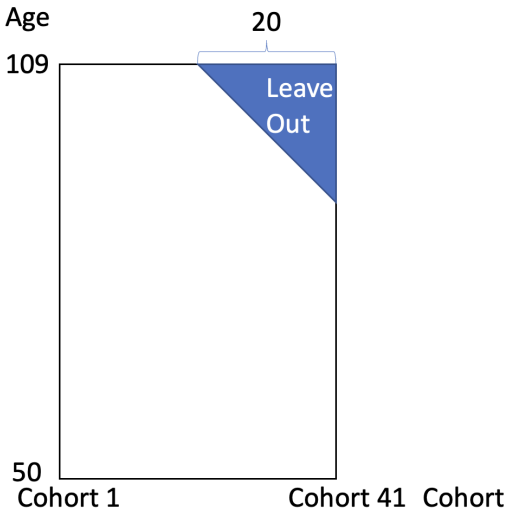
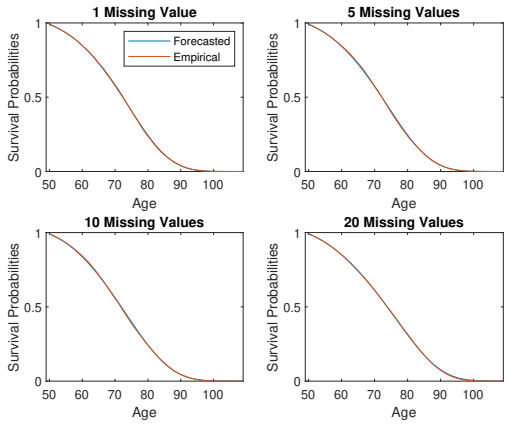
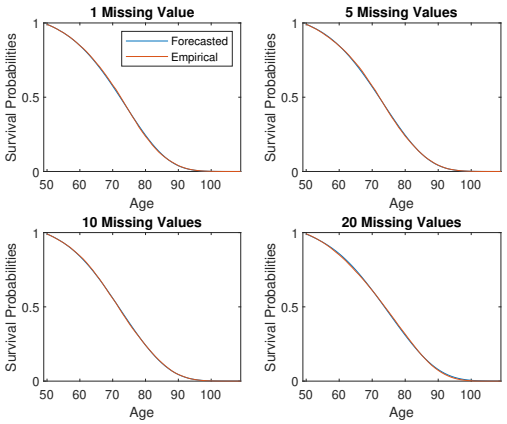


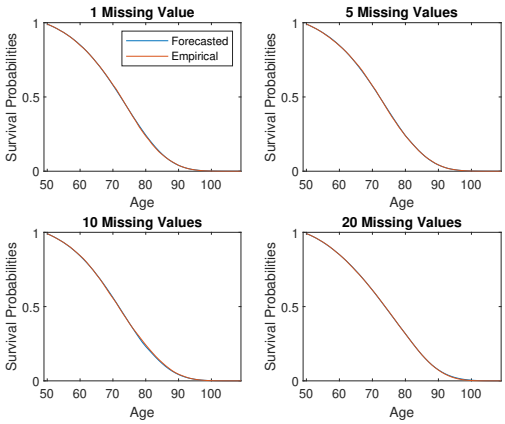
Figure 6: Leaving out a 20 by 20 triangle from complete cohort data to perform out-of-sample test.



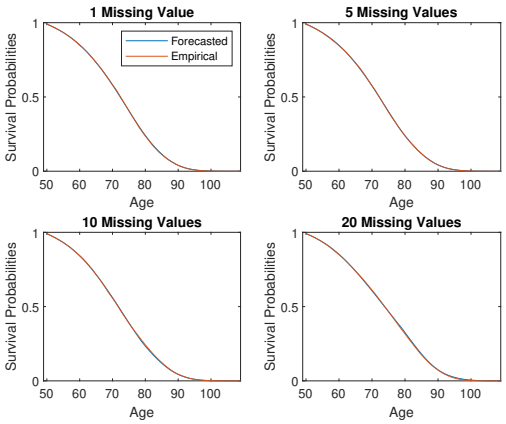
(a) ADC Gaussian



(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.

Figure 7 shows the comparison between the estimation results from the model 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 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. 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 $\lambda = 1,000$ and $\zeta = 100,000$ and calibrated with Australian male mortality data are shown in Figure 8.

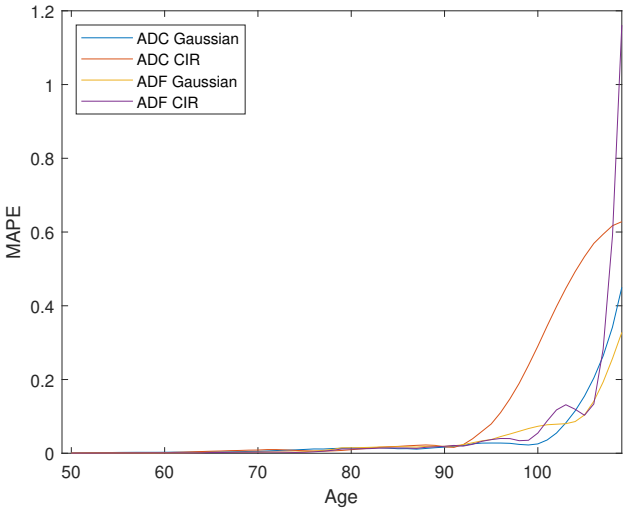


Figure 8: MAPE for Australian male survival probabilities: Incomplete cohort.

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. But the MAPEs for all four models are still at a low level, indicating nice goodness of fit.

5.3.1 Forecast Out-of-Sample 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. The idea is illustrated in Figure 9:

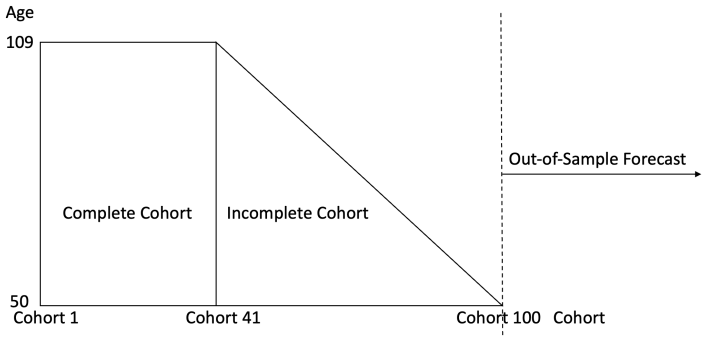
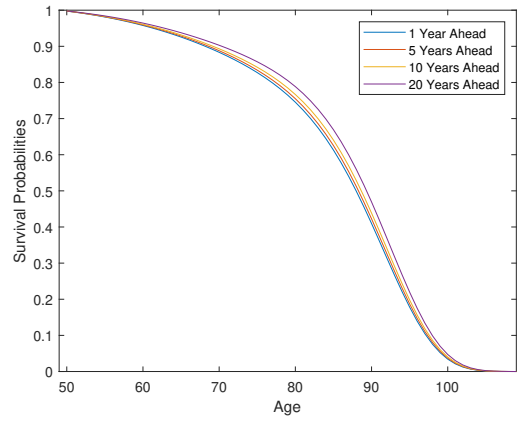
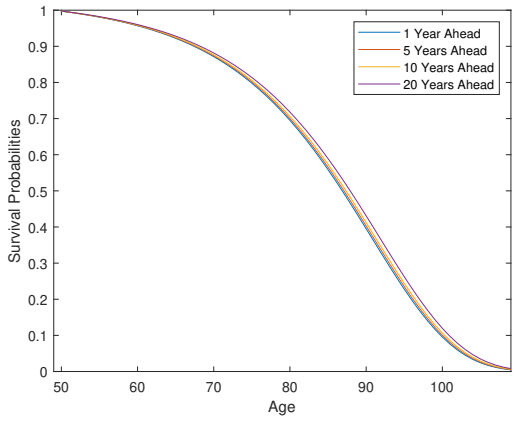


Figure 9: Incomplete cohort data and illustration of out-of-sample forecast.

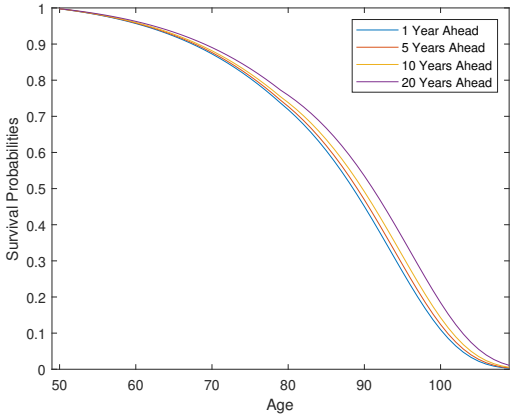
The forecasting results are shown in Figure 10 which presents $1, 5, 10$, and 20 years ahead out-of-sample forecasts. For all four models, we observe the mortality improvement since the survival probabilities are increasing over time. It is observed that the Gaussian models predict higher survival probabilities 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.



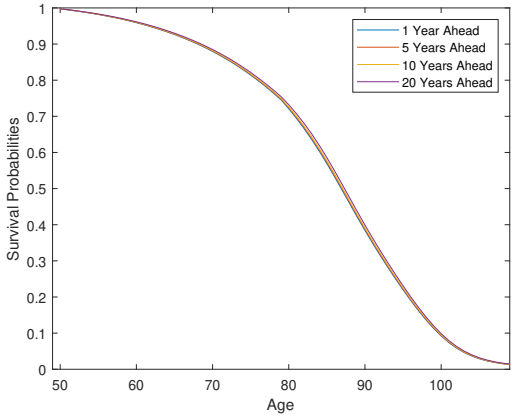
(a) ADC Gaussian



(b) ADC CIR



(c) ADF Gaussian



(d) ADF CIR

Figure 10: Out-of-sample forecast of Australian male survival probabilities: Models estimated with all available incomplete cohort data.

6 Analysis on Cohort Correlation

With our estimated parameters and using the equations in Section 2.4, we can calculate the estimated correlation between cohorts.

For the ADC Gaussian model, the cohort correlations for the cohorts aged 50, 60, 70, 80, and 90 in 2020 without regularisation are displayed in Table 1

Table 1: Estimated Australian male cohort correlations between ages 50 and 90 from ADC Gaussian model, $\lambda = 0$.

	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				

Next we wish to add regularisation and make the estimated correlation matrix move toward the empirical one. Table 2 shows the results in different countries when we increase the regularisation parameter λ. We measure the goodness of fit to the empirical cohort correlation matrix by Frobenius norm, which is the square root of the sum of squared absolute errors of all elements.

Table 2: Frobenius norm between empirical and estimated cohort correlations for different λ 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

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 under 100 is too small to force the estimated correlation matrix to move towards the empirical one. Meanwhile, $\lambda = 10,000$ is too high and will worsen the goodness of fit to the mortality data very much. Therefore, we conclude that λ at the scale of $1,000$ is the most suitable and does help with reducing the Frobenius norm. We also find that the ADC Gaussian model produces the lowest Frobenius norm for three of the four countries except Denmark. We also note that the model that gives the highest Frobenius norm differs for different countries.

Table 3-6 show the cohort correlation matrices of the four models in calendar year 2000 using Australian male data when we set $\lambda = 1,000$.

Table 3: Estimated Australian male cohort correlations between age 50 and 90 from ADC Gaussian Model, $\lambda = 1,000$.

	50	60	70	80	90
50	1	0.999301	0.939419	0.448329	0.162052
60	0.999301	1	0.951232	0.480029	0.19714
70	0.939419	0.951232	1	0.727035	0.48961
80	0.448329	0.480029	0.727035	1	0.954633
90	0.162052	0.19714	0.48961	0.954633	1
norm	1.5021				

Table 4: Estimated Australian male cohort correlations between age 50 and 90 from ADC CIR Model, $\lambda = 1,000$.

	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 5: Estimated Australian male cohort correlations between age 50 and 90 from ADF Gaussian Model, $\lambda = 1,000$.

	50	60	70	80	90
50	1	0.983868	0.917261	0.79657	0.660012
60	0.983868	1	0.973715	0.891865	0.783748
70	0.917261	0.973715	1	0.971443	0.904608
80	0.79657	0.891865	0.971443	1	0.979911
90	0.660012	0.783748	0.904608	0.979911	1
norm	2.2573				

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 the ADC CIR model at 0.8303.

Table 6: Estimated Australian male cohort correlations between age 50 and 90 from ADF CIR Model, $\lambda = 1,000$.

	50	60	70	80	90
50	1	0.975172	0.795162	0.52568	0.301748
60	0.975172	1	0.909699	0.700172	0.505138
70	0.795162	0.909699	1	0.932021	0.817651
80	0.52568	0.700172	0.932021	1	0.968788
90	0.301748	0.505138	0.817651	0.968788	1
norm	1.7233				

7 Conclusions

This paper 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 age-dependent coefficient (ADC) and age-dependent factor (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, Denmark, UK, and USA 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 papers utilising age-cohort data to calibrate the model only use 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 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 thus are truly cohort-specific. To better capture empirical cohort correlation, we regularise the parameter estimation method by penalising the difference between the estimated and empirical cohort correlations. We find that adding regularisation into 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. The results of this study can be used for mutual mortality-sharing products that have multiple cohorts in the risk pool. This should be presented in our future work.

8 Acknowledgement

The authors acknowledge financial support from the Society of Actuaries [Center of Actuarial Excellence Research Grant 2016: Longevity Risk: Actuarial and Predictive Models, Retirement

Product Innovation, and Risk Management Strategies] and the Australian Research Council [Centre of Excellence in Population Ageing Research (CEPAR) project number CE170100005].

References

Biffis, E. (2005). 'Affine processes for dynamic mortality and actuarial valuations'. *Insurance: Mathematics and Economics*, vol. 37, no. 3, pp. 443–468.

Blackburn, C. and Sherris, M. (2013). 'Consistent dynamic affine mortality models for longevity risk applications'. *Insurance: Mathematics and Economics*, vol. 53, no. 1, pp. 64–73.

Bohlmann, G. (1899). 'Ein ausgleichungsproblem'. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, vol. 1899, no. 3, pp. 260–271.

Chang, Y. and Sherris, M. (2018). 'Longevity risk management and the development of a value-based longevity index'. *Risks*, vol. 6, no. 1, p. 10.

Chen, R.-R. and Scott, L. (2003). 'Multi-factor Cox-Ingersoll-Ross models of the term structure: Estimates and tests from a Kalman filter model'. *The Journal of Real Estate Finance and Economics*, vol. 27, no. 2, pp. 143–172.

Christensen, J., Diebold, F. and Rudebusch, G. (2011). 'The affine arbitrage-free class of Nelson–Siegel term structure models'. *Journal of Econometrics*, vol. 164, no. 1, pp. 4–20.

Cox, J.C., Ingersoll Jr, J.E. and Ross, S.A. (1985). 'A theory of the term structure of interest rates'. *Econometrica*, vol. 53, no. 2, pp. 385–407.

Dahl, M. (2004). 'Stochastic mortality in life insurance: Market reserves and mortality-linked insurance contracts'. *Insurance: Mathematics and Economics*, vol. 35, no. 1, pp. 113–136.

Geyer, A.L. and Pichler, S. (1999). 'A state-space approach to estimate and test multifactor Cox-Ingersoll-Ross models of the term structure'. *Journal of Financial Research*, vol. 22, no. 1, pp. 107–130.

Henderson, R. (1924). 'A new method of graduation'. *Transactions of the Actuarial Society of America*, vol. 25, pp. 29–40.

Huang, Z., Sherris, M., Villegas, A.M. and Ziveyi, J. (2022). 'Modelling USA age-cohort mortality: A comparison of multi-factor affine mortality models'. *Risks*, vol. 10, no. 9, p. 183.

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). Available at: www.mortality.org (data downloaded in 2022).

Jevtić, P., Luciano, E. and Vigna, E. (2013). 'Mortality surface by means of continuous time cohort models'. *Insurance: Mathematics and Economics*, vol. 53, no. 1, pp. 122–133.

Kalman, R.E. (1960). 'A new approach to linear filtering and prediction problems'. *Journal of Fluids Engineering*, vol. 82(1), pp. 35–45.

Luciano, E. and Vigna, E. (2008). 'Mortality risk via affine stochastic intensities: Calibration and empirical relevance'. Working Paper. Available at: https://mpra.ub.uni-muenchen.de/59627/.

Njenga, C.N. and Sherris, M. (2011). 'Longevity risk and the econometric analysis of mortality trends and volatility'. *Asia-Pacific Journal of Risk and Insurance*, vol. 5, no. 2, article 2.

Piggott, J., Valdez, E.A. and Detzel, B. (2005). 'The simple analytics of a pooled annuity fund'. *Journal of Risk and Insurance*, vol. 72, no. 3, pp. 497–520.

Schrager, D.F. (2006). 'Affine stochastic mortality'. *Insurance: Mathematics and Economics*, vol. 38, no. 1, pp. 81–97.

Ungolo, F., Garces, L.P., Sherris, M. and Zhou, Y. (2023). 'Estimation, comparison and projection of multi-factor age-cohort affine mortality models'. Working Paper. Available at: SSRN 3912981.

Whittaker, E.T. (1922). 'On a new method of graduation'. *Proceedings of the Edinburgh Mathematical Society*, vol. 41, pp. 63–75.

Williams, D. (1991). *Probability with martingales*. Cambridge University Press.

Wills, S. and Sherris, M. (2008). 'Integrating financial and demographic longevity risk models: an Australian model for financial applications'. Working Paper. Available at: SSRN 1139724.

Xu, Y., Sherris, M. and Ziveyi, J. (2020). 'Continuous-time multi-cohort mortality modelling with affine processes'. *Scandinavian Actuarial Journal*, vol. 2020, no. 6, pp. 526–552.

Appendix 1. Change of Measure

ADC Gaussian Model

Following Blackburn and Sherris (2013), we define the market price of risk Λ_t to be:

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

where $\Lambda_t \in \mathbb{R}^{3\times 1}, \lambda^0 \in \mathbb{R}^{3\times 1}$ and $\lambda^1 \in \mathbb{R}^{3\times 3}$. Therefore, by Girsanov's theorem, we have the following relationship between the risk-neutral measure Q and the real-world measure P:

$$\mathrm{d}W_t^Q = \mathrm{d}W_t^P + \Lambda_t \mathrm{d}t,$$

where W_t^P 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:

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

where $K^P = (K^Q - \Sigma\lambda_1)$, and $\Theta^P = (K^Q - \Sigma\lambda_1)^{-1}(K^Q\Theta^Q + \Sigma\lambda_0)$. 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.

ADC CIR Model

Following Huang et al. (2022), the market price of risk Λ_t for the CIR model is:

$$\begin{aligned}
\Lambda_t &= D\left(\mathbf{Z}(t), t\right)\lambda^0, \\
\mathrm{d}W_t^Q &= \mathrm{d}W_t^P + \Lambda_t\mathrm{d}t,
\end{aligned}$$

where $\lambda^0 = \begin{pmatrix} \lambda_1^0 \\ \lambda_2^0 \\ \lambda_3^0 \end{pmatrix} \in \mathbb{R}^{3\times 1}$.

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

where

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

ADF Gaussian Model

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

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

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

ADF CIR Model

Similar to the ADC CIR model, we set

$$\begin{aligned} \Lambda_t =& D\left(\mathbf{Z}(t,x),t\right)\lambda_x^0, \\ \mathrm{d}W_t^Q =& \mathrm{d}W_t^P + \Lambda_t\mathrm{d}t, \end{aligned}$$

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's chosen to make Λ_t indifferent between the two age groups.

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

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

where

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

For the same reason as in the previous models, we assume both Θ^Q and Θ^P are equal to zero.

Appendix 2. Derive SDEs of the Force of Mortality for ADC models

ADC Gaussian Model

For ADC Gaussian model, using Ito's formula, 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:

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

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

Therefore, the SDE is simplified to be:

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

ADC CIR Model

Similarly, for the CIR model, the SDE becomes

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

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

Therefore, the SDE is simplified to be:

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

Appendix 3. Estimated Parameters with All Available Australian Incomplete-Cohort Data

The estimated parameters of the four models with all incomplete Australian male mortality data, $\lambda = 1,000$, and $\zeta = 100,000$ for ADF models are in the table below.

Table 7: Estimated Parameters with Australian Data

AUS	ADC Gaussian	ADC CIR		ADF Gaussian	ADF CIR
a_1	0.0730	5.10E-06	a_1^P	-3.66E-05	5.57E-03
a_2	0.0029	0.0014	a_2^P	1.88E-02	9.24E-05
a_3	1.36E-05	0.0318	a_3^P	1.77E-03	3.14E-03
b_2	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

For ADC Gaussian model, b_2 is between 50 and 109, indicating that the second factor affects more of the middle age. While 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. The probabilities of negative 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.