ARC Centre of Excellence in Population Ageing Research

Working Paper 2015/13

The Application of Affine Processes in Multi-Cohort Mortality Model.

Yajing Xu[1], Michael Sherris[2] and Jonathan Ziveyi[3]

[1]School of Risk and Actuarial Studies and ARC Centre of Excellence in Population Ageing Research (CEPAR), UNSW Business School, email: yajing.xu@unsw.edu.au
[2] School of Risk and Actuarial Studies, UNSW Business School and Chief Investigator, CEPAR, UNSW Australia, email: m.sherris@unsw.edu.au
[3]School of Risk and Actuarial Studies, UNSW Business School, email: j.ziveyi@unsw.edu.au

The Application of Affine Processes in Multi-Cohort Mortality Model

Yajing Xu*, Michael Sherris†and Jonathan Ziveyi‡

Working Paper: May 2015

Abstract

Cohort effects have been identified in many countries. However, some mortality models only consider the modelling and projection of age-period effects. Others, that incorporate cohort effects, do not consider cohort specific survival curves that are important for pricing and hedging purposes. In this paper, we consider modelling mortality development on a cohort basis, propose and assess a multi-cohort mortality model in an affine framework. We model the mortality intensity with common factors that affect all the cohorts as well as cohort specific factors that only affect specific cohorts, so that the correlations among cohorts are not perfect. In particular, we consider a three-factor case. The three-factor multi-cohort model is established using Danish male mortality data. The two common factors are extracted using a Kalman Filter algorithm and cohort specific factors are estimated by minimizing the residual calibration error. The calibration results demonstrate the need for cohort effects. The out-of-sample forecast performance of the proposed model, the RH model (age-period-cohort model developed of Renshaw and Haberman (2006)) and the CBD model (age-period model developed of Cairns et al. (2006)) are compared to actual mortality data. The results show that the proposed model produces more consistent estimates of cohort survival curves.

Keywords: Multi-cohort mortality model; Affine framework; Common factors; Cohort specific factors; Mortality projections.

JEL Classification Numbers: C13, C15, C51, C52, G22, G23, J11.

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

†School of Risk & Actuarial Studies, Australian Research Council Centre of Excellence in Population Ageing Research (CEPAR), UNSW Australia, email: m.sherris@unsw.edu.au

‡School of Risk & Actuarial Studies, UNSW Australia, email: j.ziveyi@unsw.edu.au

1 Introduction

A cohort is a group of individuals who are born in a particular year and share specific experience within a certain time period. This paper considers modelling mortality development on a cohort basis. One of the reasons why a cohort mortality model is important is mortality forecasting. Given differences in medicine, education and the like, it is obvious that mortality rates for different cohorts evolve in different patterns. The survival probability of a specific cohort depends on future mortality intensity movements which are affected by both common effects and cohort specific effects. Better forecasts of cohort survival curves provide a basis for accurate pricing and hedging of mortality-linked products issued by insurance companies and annuity providers. Life insurance companies and pension funds need to manage longevity risk. Their hedging strategies involve contracts that are contingent on future mortality evolution of specific cohorts. Mortality-linked derivative pricing and hedging provide another reason. For example, the future coupon payments of longevity bonds (also known as survivor bonds) depend on the percentage of chosen cohorts of retirement age (e.g. 65) on the issue date still alive on the future coupon payment dates. Even the price of more complicated securities, such as survivor swaps, q-forwards, and option-type longevity derivatives, requires assumptions on cohort survival curves.

The analysis of UK mortality data by Gallop (2008) shows that there are clear signs of cohort effects. Cohorts born in the 1920s and 1930s exhibit higher rates of mortality improvement than other cohorts. A number of possible reasons have been put forward including differences in education, medical advances and smoking patterns between cohorts (see Willets (2004) for an overview). Cohort effects have also been identified in other countries such as Denmark, Japan, France, Italy and Switzerland by Andreev and Vaupel (2005). However, some mortality models - that include Lee and Carter (1992), Brouhns et al. (2002) and Blackburn and Sherris (2013) - only consider the modelling and projection of age-period effects. Others incorporating cohort effects, for example Renshaw and Haberman (2006), do not consider cohort specific survival curves which are

important for pricing and hedging purposes. In this paper, we present and assess a cohort mortality model that can deliver tractable representations of cohort survival curves.

Dahl and Moller (2006) consider a single cohort mortality model with the mortality intensity

$$\mu(x,t) = \mu^o(x+t)\zeta(x,t), \tag{1}$$

where $(\mu^o(x+t))_{t\in[0,T]}$ is the initial mortality intensity at all ages and takes the form of a Gompertz-Makeham mortality law. Here, $\zeta(x,t)$ describes the change in the mortality from time 0 to t for a person aged $x+t$ and is modelled via a time-inhomogeneous CIR process.

We modify Dahl and Moller's mortality model for cohort i such that

$$\mu^i(x,t) = \mu_x^{o,i}(t) + \zeta^i(x,t), \tag{2}$$

so that this model can be calibrated in an affine framework. The affine framework is very well developed in interest rates modelling theory (see Duffie and Kan (1996), Dai and Singleton (2000)), the main advantage being its tractability and flexibility. Another advantage of the affine framework is its ease of application in pricing and risk management. Similar to the short rate models in interest rate theory, the pricing of life contingent contracts and risk analysis can be carried out under a risk-neutral measure using a correspondingly adjusted discount rate. Biffis (2005) considers a mortality model in a financial market and shows how to value several basic life insurance contracts in the affine framework.

We extend the single cohort model to multiple cohorts. As pointed out by Jevtic et al. (2013), a desirable model should capture the characteristics of the whole mortality surface. They propose a two-factor model that produces high correlations among cohorts. Their model fails to capture the dynamics of mortality intensity at older ages. We model the mortality intensity with common factors that affect all the cohorts as well as cohort specific factors that only affect specific cohorts, so that the correlations among cohorts

are not perfect. We estimate our model using Danish mortality data, and show that this model has good empirical fit.

The present paper is organized as follows. The next section outlines the general case of the model. Then, Section 3 discusses the special case of an affine three-factor model. Section 4 fits the model to Danish mortality data. Section 5 presents simulation results of our model. A comparison with Renshaw and Haberman's model and the CBD model is provided in Section 6. Concluding remarks are given in Section 7.

2 Model Specification

In this section, we present the general case of our cohort mortality model (Section 2.1) and illustrate how an affine framework can be extended to mortality rates modelling (Section 2.2).

2.1 The General Case

We consider the whole mortality surface comprising a number of cohorts. Each cohort is labelled by $i \in \mathbf{I} \subset \mathbb{R}$ indicating the year in which this cohort is born. We take as a starting point an initial curve for mortality intensity at all ages $(\mu_x^{o,i}(t))_{t\in[0,T]}$ for the cohort i aged $x+t$ at time t. Let $(\mu^i(x,t))_{t\in[0,T]}$ represent the mortality intensity for cohort i aged $x+t$ at time t. At time 0, we have $\mu^i(x,0) = \mu_x^{o,i}(0)$.

For the initial curve, we neglect the cohort specific effects and assume that the diffusion processes affecting each cohort are driven by some common factors. We then model the cohort specific effects via a diffusion process $\zeta^i = (\zeta^i(x,t))_{t\in[0,T]}$. The mortality intensity process for cohort i aged $x+t$ at time t is modelled via

$$\mu^i(x,t) = \mu_x^{o,i}(t) + \zeta^i(x,t). \tag{3}$$

The risk-neutral survival probability is defined as

$$S^i(x,t,T) = E^Q[\mathrm{e}^{-\int_t^T \mu^i(x,s)ds}|\mathcal{F}(t)] = E^Q[\mathrm{e}^{-\int_t^T (\mu_x^{o,i}(s)+\zeta^i(x,s))ds}|\mathcal{F}(t)], \tag{4}$$

where Q is the risk-neutral measure.

2.2 The Affine Mortality Framework

Equation (4) shows that the mortality model should have two components:

(a) the change of measure from data-generating measure P to risk-neutral measure Q;

(b) the dynamics of the mortality intensity μ under the risk-neutral measure Q.

A market for longevity risk does not exist at present. Several longevity-linked derivatives have been proposed in the literature and some transactions have been completed in practice involving longevity risk. These derivatives are not actively traded in the market and hence there is no market data to rely on. Since the market is incomplete, there is no unique risk-neutral measure Q. In the literature, some authors adopt a risk premium implied by longevity-linked products or institution standards like Cairns et al. (2009) and Biffis (2005). Cairns et al. (2009) propose to use the risk premium implied by the EIB bond, which is the first longevity bond but has not been actually traded. The EIB bond was designed to help financial institutions look for potential instruments to hedge their systematic longevity risks. It turned out that the EIB bond was not successful with investors, and that it did not generate sufficient demand to be launched. Biffis (2005) proposes to set the risk-neutral measure Q according to the International Accounting Standards Board's proposal. The risk-neutral measure Q restricted to financial market risks can be easily inferred from observable security prices in the market. As for the choice of Q restricted to mortality market risks, Biffis (2005) proposes to use the risk-adjusted discount rate in the embedded value method. Another common way is to adopt a zero market price assumption. One example is Schrager (2006), where the author fixes the market price of mortality risk to zero to price a guaranteed annuity option.

Since the focus of this paper lies on evaluating the empirical performance of the proposed multi-cohort mortality model and not on exploiting the market price of longevity risk, we follow the argumentation of Blackburn and Sherris (2013) and define this risk-neutral measure to be the best-estimate measure $\bar{Q}$ that is in accordance with historical mortality data. In our affine mortality model (b) is replaced by the following assumption:

(b)' the mortality intensity is an affine (constant-plus-linear) function, $R(\mathbf{Y})$, of factors where $\mathbf{Y} \in \mathbb{R}^N$ is a vector of factors, and the factors are affine diffusions (with both of the drift and the variance-covariance matrix being affine) under the best-estimate measure $\bar{Q}$.

The instantaneous mortality intensity $\mu^i(x,t)$ of a given cohort i and initial age x at time t is given by

$$\mu^i(x,t) = \delta_0 + \delta_1' \mathbf{Y}^i(t), \tag{5}$$

where $\delta_0 \in \mathbb{R}$ and $\delta_1 \in \mathbb{R}^N$. For the values of δ_0 and δ_1, a narrower but more usual assumption is $\delta_0 = 0$ and $\delta_1 = \mathbf{1}$ (see Blackburn and Sherris (2013), Jevtic et al. (2013)). $\mathbf{Y}^i(t)$ is a vector of N factors.

The process $\mathbf{Y}^i$ for a given cohort i is an affine diffusion process if $\mathbf{Y}^i$ solves the following stochastic differential equation

$$d\mathbf{Y}^i(t) = H(\mathbf{Y}^i(t))dt + G(\mathbf{Y}^i(t))d\mathbf{W}^{\bar{Q}}(t), \tag{6}$$

with

$$H(\mathbf{Y}^i(t)) = -\Phi \mathbf{Y}^i(t),$$
$$G(\mathbf{Y}^i(t)) = \Sigma s(\mathbf{Y}^i(t)),$$

where $s(\mathbf{Y}^i(t))$ is a diagonal $N \times N$ matrix with the j^{th} diagonal element $s_j(Y_j^i(t)) = \sqrt{s_{0j} + s_{1j} Y_j^i(t)}$, and $s_{0j}, s_{1j} \in \mathbb{R}$. $\Phi, \Sigma \in \mathbb{R}^{N \times N}$ are constants. $\mathbf{W}^{\bar{Q}}(t)$ is an N-dimensional standard Brownian motion under the best-estimate measure $\bar{Q}$.

Gaussian processes (when $s_{1j} = 0$, $j = 1, ..., N$) and square-root processes (when $s_{1j} \neq 0$, $j = 1, ..., N$) shown above are the best known examples of affine diffusions. The two classes differ with respect to their assumptions about the variance-covariance matrix $G(\mathbf{Y}^i(t))G(\mathbf{Y}^i(t))^T$. Gaussian processes have a constant variance-covariance matrix, which means $s_{1j} = 0$ for $j = 1, ..., N$ in our model. By setting $s_{0j} = 1$ (Σ is free), Equation (6) becomes a linear SDE

$$d\mathbf{Y}^i(t) = -\Phi\mathbf{Y}^i(t)dt + \Sigma d\mathbf{W}^{\bar{Q}}(t). \tag{7}$$

Inspired by interest rate theory, we define the average force of mortality curve for year t, initial age x as

$$\bar{\mu}^i(x, t, T) = -\frac{1}{T-t}\log[S^i(x, t, T)], \tag{8}$$

in the same way of defining yield curve. In affine mortality models the average force of mortality is also an affine function of the state vector $\mathbf{Y}^i(t)$

$$\bar{\mu}^i(x, t, T) = -\frac{A^i(t, T)}{T-t} - \frac{B^i(t, T)^T}{T-t}\mathbf{Y}^i(t), \tag{9}$$

where $A^i(t, T)$ and $B^i(t, T)^T$ are coefficients that only depend on $\tau = T-t$. The derivation is provided in the Appendix A.

3 The Affine Three-Factor Multi-Cohort Model

In this section, we provide a three-factor example of the proposed multi-cohort model. For simplicity, we assume $\delta_0 = 0$ and $\delta_1 = \mathbf{1}$. For cohort i ($i \in \mathbf{I} \subset \mathbb{R}$), $\mathbf{Y}^i(t)$ is a vector of three factors and can be divided into two parts

$$\mathbf{Y}^i(t) = \begin{pmatrix} \mathbf{X}(t) \\ Z^i(t) \end{pmatrix}, \tag{10}$$

where $\mathbf{X}(t)$ is a vector of two common factors that affect all the cohorts ($\mathbf{I} \subset \mathbb{R}$), and $Z^i(t)$ is the cohort specific factor that only affects the cohort i ($i \in \mathbf{I} \subset \mathbb{R}$). Thus the mortality intensity process for cohort i aged $x+t$ at time t is modelled via

$$\mu^i(x,t) = \mu_x^{o,i}(t) + \zeta^i(x,t), \tag{11}$$

where $\mu_x^{o,i}(t) = X_1(t) + X_2(t)$ and $\zeta^i(x,t) = Z^i(t)$.

We shall first illustrate the dynamics of initial mortality intensity (Section 3.1) and cohort mortality intensity (Section 3.2) separately, and then derive the correlations across cohorts (Section 3.3).

3.1 Initial Mortality Intensity Modelling

The common factors for all the cohorts are estimated using a Kalman Filter extracted from the mortality surface of age-period mortality data. We drop superscript i and subscript x in the remaining section.

The initial instantaneous mortality intensity $\mu^o(t)$ at time t is

$$\mu^o(t) = X_1(t) + X_2(t), \tag{12}$$

where the state variables $\mathbf{X}(t) = (X_1(t), X_2(t))$ have the dynamics under best-estimate measure $\bar{Q}$ expressed as

$$\begin{pmatrix} dX_1(t) \\ dX_2(t) \end{pmatrix} = -\begin{pmatrix} k_{11} & 0 \\ 0 & k_{22} \end{pmatrix} \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix} \begin{pmatrix} dW_1^{\bar{Q}}(t) \\ dW_2^{\bar{Q}}(t) \end{pmatrix}, \tag{13}$$

with $W_1^{\bar{Q}}(t)$ and $W_2^{\bar{Q}}(t)$ being two uncorrelated standard Wiener processes under $\bar{Q}$.

The corresponding initial survival probability is exponentially affine.

$$\begin{aligned} S^o(t,T) &= E^{\bar{Q}}[e^{-\int_t^T \mu^o(s)ds}|\mathcal{F}_t] \\ &= e^{B_1(t,T)X_1(t)+B_2(t,T)X_2(t)+C(t,T)}, \end{aligned} \tag{14}$$

where $B_1(t,T)$, $B_2(t,T)$ and $C(t,T)$ are solutions to a set of ordinary differential equations which can be represented as

$$\begin{aligned} B_1(t,T) &= -\frac{1-e^{-k_{11}(T-t)}}{k_{11}}, \\ B_2(t,T) &= -\frac{1-e^{-k_{22}(T-t)}}{k_{22}}, \\ C(t,T) &= \frac{1}{2}\sum_{i=1}^{2}\frac{\sigma_{ii}^2}{k_{ii}^3}[\frac{1}{2}(1-e^{-2k_{ii}(T-t)}) - 2(1-e^{-k_{ii}(T-t)}) + k_{ii}(T-t)]. \end{aligned} \tag{15}$$

Using the same algorithm as specified in Appendix A, we can prove this by substituting Equation (13) into Equation (46). Consequently, the initial average force of mortality curve is given by

$$\begin{aligned} \bar{\mu}^o(t,T) &= -\frac{1}{T-t}\log[S_x^o(t,T)] \\ &= \frac{1-e^{-k_{11}(T-t)}}{k_{11}(T-t)}X_1(t) + \frac{1-e^{-k_{22}(T-t)}}{k_{22}(T-t)}X_2(t) - \frac{C(t,T)}{T-t}. \end{aligned} \tag{16}$$

We define the P-measure to link state variables with age-period mortality data in each period. The dynamics of the two state variables under this measure can be represented as

$$\begin{pmatrix} dX_1(t) \\ dX_2(t) \end{pmatrix} = -\begin{pmatrix} k_{11}^P & 0 \\ 0 & k_{22}^P \end{pmatrix}\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix}dt + \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix}\begin{pmatrix} dW_1^P(t) \\ dW_2^P(t) \end{pmatrix}. \tag{17}$$

3.2 Cohort Mortality Intensity Modelling

The instantaneous mortality intensity $\mu^i(x,t)$ of a given cohort i aged $x+t$ at time t is

$$\mu^i(x,t) = X_1(t) + X_2(t) + Z^i(t), \tag{18}$$

where $X_1(t)$, $X_2(t)$ are common factors and $Z^i(t)$ is the cohort specific factor. The state variables $(X_1(t), X_2(t), Z^i(t))$ have the dynamics under best-estimate measure $\bar{Q}$ given as

$$\begin{pmatrix} dX_1(t) \\ dX_2(t) \\ dZ^i(t) \end{pmatrix} = - \begin{pmatrix} k_{11} & 0 & 0 \\ 0 & k_{22} & 0 \\ 0 & 0 & k^i_{33} \end{pmatrix} \begin{pmatrix} X_1(t) \\ X_2(t) \\ Z^i(t) \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma^i_{33} \end{pmatrix} \begin{pmatrix} dW_1^{\bar{Q}}(t) \\ dW_2^{\bar{Q}}(t) \\ dW_3^{i\bar{Q}}(t) \end{pmatrix}, \quad (19)$$

where $W_j^{\bar{Q}}(t)$ is a standard Wiener process under $\bar{Q}$ for $j = 1, 2, 3$.

The survival probabilities for year t, initial age x can then be represented as

$$\begin{aligned} S^i(x,t,T) &= E^{\bar{Q}}[e^{-\int_t^T \mu^i(s)ds} | \mathcal{F}_t] \\ &= e^{B_1(t,T)X_1(t) + B_2(t,T)X_2(t) + B_3(t,T)Z^i(t) + A(t,T)}, \end{aligned} \quad (20)$$

where $B_1(t,T)$, $B_2(t,T)$, $B_3(t,T)$ and $A(t,T)$ are governed by ordinary differential equations

$$\begin{pmatrix} \frac{dB_1(t,T)}{dt} \\ \frac{dB_2(t,T)}{dt} \\ \frac{dB_3(t,T)}{dt} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} k_{11} & 0 & 0 \\ 0 & k_{22} & 0 \\ 0 & 0 & k^i_{33} \end{pmatrix} \begin{pmatrix} B_1(t,T) \\ B_2(t,T) \\ B_3(t,T) \end{pmatrix},$$

$$\frac{dA(t,T)}{dt} = -\frac{1}{2} \sum_{j=1}^{3} (\Sigma' B(t,T) B(t,T)' \Sigma)_{j,j},$$

with boundary conditions $B_1(T,T) = B_2(T,T) = B_3(T,T) = A(T,T) = 0$.

The solution to the above system of ODEs is

$$\begin{aligned} B_1(t,T) &= -\frac{1 - e^{-k_{11}(T-t)}}{k_{11}}, \\ B_2(t,T) &= -\frac{1 - e^{-k_{22}(T-t)}}{k_{22}}, \\ B_3(t,T) &= -\frac{1 - e^{-k^i_{33}(T-t)}}{k^i_{33}}, \\ A(t,T) &= \frac{1}{2} \sum_{i=1}^{3} \frac{\sigma^2_{ii}}{k^3_{ii}} [\frac{1}{2}(1 - e^{-2k_{ii}(T-t)}) - 2(1 - e^{-k_{ii}(T-t)}) + k_{ii}(T-t)]. \end{aligned} \quad (21)$$

The above relationship can be proved by substituting Equation (19) in Equation (46).

The average force of mortality curve for year t, initial age x is given by

$$\begin{aligned}\bar{\mu}^i(x,t,T) &= -\frac{1}{T-t}\log[S^i(x,t,T)] \\ &= \frac{1-e^{-k_{11}(T-t)}}{k_{11}(T-t)}X_1(t) + \frac{1-e^{-k_{22}(T-t)}}{k_{22}(T-t)}X_2(t) + \frac{1-e^{-k^i_{33}(T-t)}}{k^i_{33}(T-t)}Z^i(t) - \frac{A(t,T)}{T-t}.\end{aligned} \tag{22}$$

3.3 Correlations Across Cohorts

The dynamics of the instantaneous mortality intensities for any two cohorts i and j ($i \neq j$ and $i, j \in \mathbf{I} \subset \mathbb{R}$) are governed by the stochastic differential equations such that

$$d\mu^i(x,t) = [-k_{11}X_1(t) - k_{22}X_2(t) - k^i_{33}Z_i(t)]dt + \sigma_{11}dW_1^{\bar{Q}}(t) + \sigma_{22}dW_2^{\bar{Q}}(t) + \sigma^i_{33}dW_3^{i\bar{Q}}(t), \tag{23}$$

$$d\mu^j(x,t) = [-k_{11}X_1(t) - k_{22}X_2(t) - k^j_{33}Z_j(t)]dt + \sigma_{11}dW_1^{\bar{Q}}(t) + \sigma_{22}dW_2^{\bar{Q}}(t) + \sigma^j_{33}dW_3^{j\bar{Q}}(t), \tag{24}$$

where $W_1^{\bar{Q}}(t)$, $W_2^{\bar{Q}}(t)$, $W_3^{i\bar{Q}}(t)$ and $W_3^{j\bar{Q}}(t)$ are independent Brownian motions. This can be directly seen from equation (18) and equation (19).

With the above two equations, we can easily deduce the instantaneous correlation between $\mu^i(x,t)$ and $\mu^j(x,t)$, which is given by

$$\text{Corr}(d\mu^i, d\mu^j)_t = \frac{\sigma^2_{11} + \sigma^2_{22}}{\sqrt{\sigma^2_{11} + \sigma^2_{22} + (\sigma^i_{33})^2}\sqrt{\sigma^2_{11} + \sigma^2_{22} + (\sigma^j_{33})^2}}. \tag{25}$$

The dynamics of mortality intensities are driven by both common factors and cohort specific factors. That has important implications for the correlations across cohorts. If the volatilities of cohort specific factors are small, the correlation between these two cohorts would be high because of the influence of common factors. If, on the other hand side, the correlation would be low as a result of high variations in cohort specific factors.

4 Calibration to Danish Male Mortality Data

In Section 3, we developed the requisite theory of a three-factor multi-cohort mortality model as an affine function of the underlying state variables. To estimate parameters and to assess the validity of this model as a tool for mortality modelling, in this section we calibrate our cohort model to Danish mortality data (Section 4.1). We use the Danish mortality data for males from the Human Mortality Database (HMD). We start with the estimation of the two common factors (Section 4.2), and then estimate cohort specific factors for 12 cohorts (Section 4.3).

4.1 Data Description

We test our model on Danish mortality data for two reasons. Firstly, Denmark has a long period of mortality data dating back to the seventeenth century, which ensures the quality of the data. Secondly, according to Andreev and Vaupel (2005), strong cohort effects have been found in Danish mortality data. In Section 6, we will compare our model with Renshaw and Haberman's model (Renshaw and Haberman (2006)) and the CBD model (Cairns et al. (2006)). Using Danish data allows us to see how these models perform in forecasting cohort survival curve when cohort effects exist.

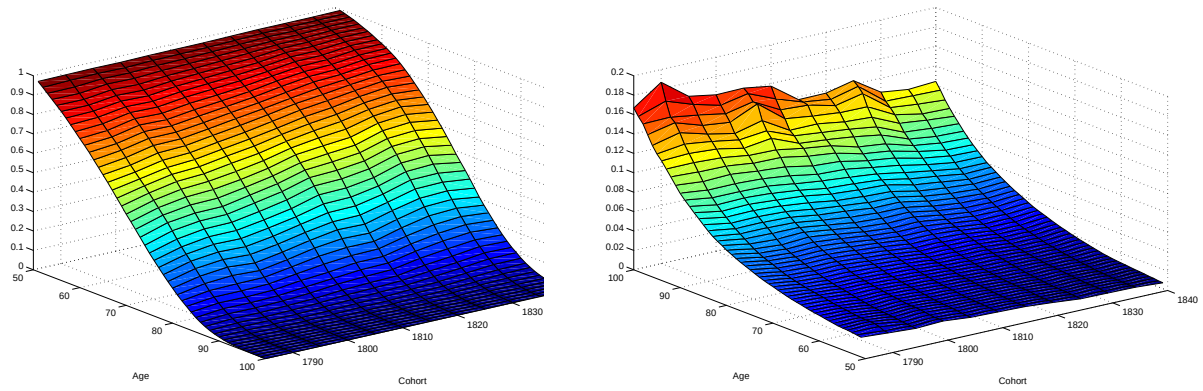
The Human Mortality Database includes death counts, exposure-to-risk and death rates constructed in both age-period form and age-cohort form. To be consistent we use the age-period data and extrapolate the cohort death rates $q^i_{50}(t)$ for males born between 1790 to 1909 with a 10-year step by taking the diagonal of the age-period data. Thus we use 12 cohorts with ages from 50 to 100. We focus on the older ages because we are interested in the application to longevity-linked derivative pricing and risk management. The actual survival probability and average force of mortality for cohort i aged x over

duration $\tau = T - t$ is calculated as follows

$$\tilde{S}^i(x,t,T) = \tilde{S}^i(x,\tau) = \prod_{s=1}^{\tau}(1-\tilde{q}^i(t+s-1,x+s-1)),$$
$$\tilde{\mu}^i(x,\tau) = -\frac{1}{\tau}\log\tilde{S}^i(x,\tau), \tag{26}$$

where $\tilde{q}^i(t,x)$ is the actual death rate at time t and age x.

The Danish cohort mortality data is shown in Figure 1. Figure 1a is the plot of survival probabilities for these 12 cohorts aged 50 to 100, while Figure 1b is the plot of corresponding average force of mortality.



(a) Survival probabilities (b) Average force of mortality

Figure 1: Danish cohort mortality data for males born from 1790 to 1909

4.2 Estimation of Common Factors

The common factors for all the cohorts are extracted from the mortality surface of age-period data and computed using a Kalman Filter. We begin with the state space equations which have two parts: the measurement equation which represents the observed affine relationship between the initial average force of mortality and the state variables, and the state transition equation which describes the unobserved dynamics of the state variables. We then use the Kalman Filter algorithm to recursively generate unobserved values of the state variables conditioning on the observed values of the initial average force of mortality.

Finally we choose the optimal parameter set which maximizes the log-likelihood function (specified in Equation (34)).

4.2.1 State Space Equations

The dynamic affine model can be represented in a state space form such that the observations (initial average force of mortality) and states are jointly Markovian. This allows the construction of an efficient filtering procedure to evaluate the likelihood.

We have defined the initial average force of mortality curve for year $t = i + x$ aged x in Equation (16), the measurement equation therefore is

$$\begin{pmatrix} \bar{\mu}^o(t,t+1) \\ \bar{\mu}^o(t,t+2) \\ \vdots \\ \bar{\mu}^o(t,t+n) \end{pmatrix} = \begin{pmatrix} \frac{1-e^{-k_{11}}}{k_{11}} & \frac{1-e^{-k_{22}}}{k_{22}} \\ \frac{1-e^{-2k_{11}}}{2k_{11}} & \frac{1-e^{-2k_{22}}}{2k_{22}} \\ \vdots & \vdots \\ \frac{1-e^{-nk_{11}}}{nk_{11}} & \frac{1-e^{-nk_{22}}}{nk_{22}} \end{pmatrix} \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} - \begin{pmatrix} \frac{1}{2}\sum_{i=1}^{2} \frac{\sigma_{ii}^2}{k_{ii}^3}[\frac{1}{2}(1-e^{-2k_{ii}}) - 2(1-e^{-k_{ii}}) + k_{ii}] \\ \frac{1}{2}\sum_{i=1}^{2} \frac{\sigma_{ii}^2}{2k_{ii}^3}[\frac{1}{2}(1-e^{-4k_{ii}}) - 2(1-e^{-2k_{ii}}) + 2k_{ii}] \\ \vdots \\ \frac{1}{2}\sum_{i=1}^{2} \frac{\sigma_{ii}^2}{nk_{ii}^3}[\frac{1}{2}(1-e^{-2nk_{ii}}) - 2(1-e^{-nk_{ii}}) + nk_{ii}] \end{pmatrix} + \begin{pmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \vdots \\ \varepsilon_n(t) \end{pmatrix}. \tag{27}$$

Equation (27) can be rewritten in the following identical representation

$$y_t = -BX_t - C + \varepsilon_t, \qquad \varepsilon_t \sim N(0, H), \tag{28}$$

where $B = -\begin{pmatrix} \frac{1-e^{-k_{11}}}{k_{11}} & \frac{1-e^{-k_{22}}}{k_{22}} \\ \frac{1-e^{-2k_{11}}}{2k_{11}} & \frac{1-e^{-2k_{22}}}{2k_{22}} \\ \vdots & \vdots \\ \frac{1-e^{-nk_{11}}}{nk_{11}} & \frac{1-e^{-nk_{22}}}{nk_{22}} \end{pmatrix}$, $C = \begin{pmatrix} \frac{1}{2}\sum_{i=1}^{2} \frac{\sigma_{ii}^2}{k_{ii}^3}[\frac{1}{2}(1-e^{-2k_{ii}}) - 2(1-e^{-k_{ii}}) + k_{ii}] \\ \frac{1}{2}\sum_{i=1}^{2} \frac{\sigma_{ii}^2}{2k_{ii}^3}[\frac{1}{2}(1-e^{-4k_{ii}}) - 2(1-e^{-2k_{ii}}) + 2k_{ii}] \\ \vdots \\ \frac{1}{2}\sum_{i=1}^{2} \frac{\sigma_{ii}^2}{nk_{ii}^3}[\frac{1}{2}(1-e^{-2nk_{ii}}) - 2(1-e^{-nk_{ii}}) + nk_{ii}] \end{pmatrix}$
and H is an n-dimensional diagonal matrix.

The state transition equation can be represented as

$$X_t = \Psi X_{t-1} + \eta_t, \qquad \eta_t \sim N(0, Q), \tag{29}$$

where $\Psi = \begin{pmatrix} e^{-k_{11}^P} & 0 \\ 0 & e^{-k_{22}^P} \end{pmatrix}$, $Q = \begin{pmatrix} \frac{\sigma_{11}^2}{2k_{11}^P}(1 - e^{-2k_{11}^P}) & 0 \\ 0 & \frac{\sigma_{22}^2}{2k_{22}^P}(1 - e^{-2k_{22}^P}) \end{pmatrix}$. The derivation is outlined in Appendix B.

4.2.2 Kalman Filter Algorithm

With the state space representation described by Equations (28) and (29), we now outline the Kalman filter algorithm to extract common factors. Denote the information available at time t by $Y_t = (y_1, y_2, ..., y_t)$, and the parameter set by ψ. In the forecasting step, we forecast unknown values of state variables conditioning on the information at time $t-1$ such that

$$X(t|t-1) = E^P[X_t|Y_{t-1}] = \Psi(\psi)X_{t-1}, \tag{30}$$

$$S(t|t-1) = \Psi(\psi)S(t-1|t-1)\Psi(\psi)' + Q_t(\psi). \tag{31}$$

We then use the information at time t to update our forecasts

$$X(t|t) = E^P[X_t|Y_t] = X(t|t-1) - S(t|t-1)B(\psi)'F_t^{-1}v_t, \tag{32}$$

$$S(t|t) = S(t|t-1) - S(t|t-1)B(\psi)'F_t^{-1}B(\psi)S(t|t-1), \tag{33}$$

where $v_t = y_t - E(y_t|Y_{t-1}) = y_t + C(\psi) + B(\psi)X(t|t-1)$, $F_t = cov(v_t) = B(\psi)S(t|t-1)B(\psi)' + H(\psi)$, $H(\psi) = diag(\sigma_\varepsilon^2(\tau_1), ..., \sigma_\varepsilon^2(\tau_N))$.

Every iteration will yield the value for the log-likelihood function shown below

$$\log l(y_1, ..., y_T; \psi) = \sum_{t=1}^{T}(-\frac{N}{2}\log(2\pi) - \frac{1}{2}\log(F_t) - \frac{1}{2}v_t'F_t^{-1}v_t), \tag{34}$$

where N is the number of observed average force of mortality. The estimated parameter

set is determined as the one which maximizes the log-likelihood function.

4.2.3 Estimation Results

Applying this estimation method to the data, we obtain the fitted parameters for Danish age-period data from year 1840 to 2011. Table 1 summarises the results.

Table 1: Kalman filter parameter estimates, log likelihood and RMSE

k_{11}^P	0.01914
k_{22}^P	0.00769
k_{11}	-0.03891
k_{22}	-0.10224
σ_{11}	0.00039
σ_{22}	0.00033
Log likelihood	-47016.82570
RMSE	0.00332

Figure 2 shows the two common factors obtained. We find a similar structural change as reported by Blackburn and Sherris (2013). When the authors calibrated their model to Swedish data, they found that one of the factors decreased between 1910 and 1960, and the downward trend stopped around 1960. As shown in Figure 2, the decreasing trend of the first factor in our model stops around 1960. Compared to the first factor, the second factor X_2 shows a slowly improving trend in mortality with a lower volatility. Figure 3 shows factor loadings of the two common factors. Both of these factor loadings are increasing with age, while the loading on the second factor mostly affects older ages.

4.3 Estimation of Cohort Specific Factors

We denote all the parameters of cohort specific factors by θ. θ^* is estimated by minimizing the calibration error. The optimization problem is shown below

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \sqrt{\sum_{\tau=1}^{50} (\tilde{\mu}^i(x,\tau) - \bar{\mu}^i(x,\tau))^2}. \tag{35}$$

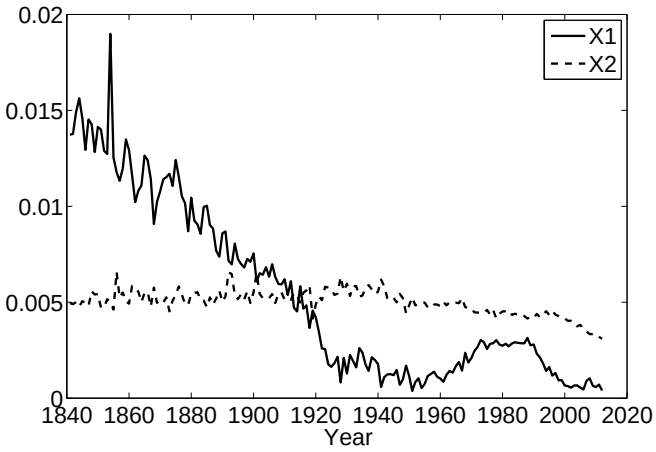


Figure 2: Common Factors

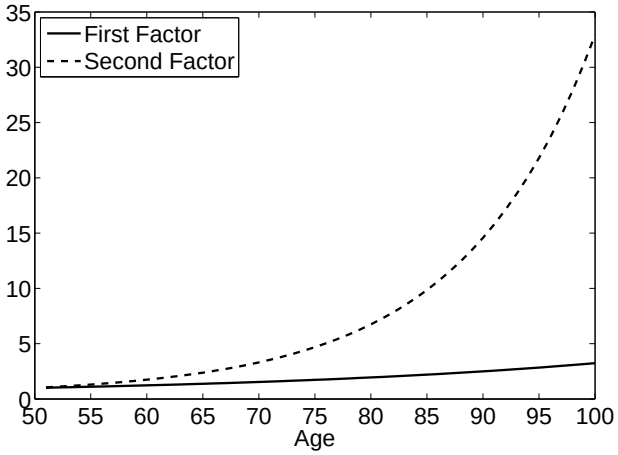


Figure 3: Factor Loadings of Common Factors

Table 2 summarises the estimation results for 12 cohorts born from 1790 to 1909 with a ten-year time step. We notice that σ^i_{33} has very low values for some cohorts. Up to five decimal places, the values of σ^i_{33} for two oldest cohorts and the 1890 cohort are zero. This issue is also reported by Luciano and Vigna (2005). The authors fit 16 UK cohorts with Feller processes, among which the volatility parameters for some cohorts are close to zero. We also observe an general increasing trend of initial values through cohorts. It indicates that though some cohorts perform differently than others, there is a general mortality improvement through cohorts.

Figure 4 shows the calibration residuals for the three-factor multi-cohort model. From the residuals we can see that the model provides a good fit. Except for several extreme points at very old ages, the remaining residuals are very small. Figure 5 shows the mean absolute

Table 2: Estimation Results for Cohort Specific Factors

	k^i_{33}	σ^i_{33}	$Z^i(t)$
1790	-0.15727	0.00000	-3.90E-05
1800	-0.01837	0.00000	-0.00183
1810	-0.07744	0.00090	-0.00179
1820	-0.09664	0.00034	-0.00031
1830	-0.07529	0.00060	-0.00106
1840	-0.07735	0.00070	-0.00127
1850	-0.09205	0.00074	-0.00173
1860	-0.12273	0.00018	-0.00015
1870	-0.09558	0.00027	-0.00055
1880	-0.11127	0.00027	-0.00046
1890	-0.12658	0.00000	-0.00073
1900	-0.06085	0.00225	0.00099

percentage error (MAPE) of the estimated survival probabilities for all the cohorts born from 1790 to 1900. We use a different scale for ages over 90. The MAPE is below 5% under the age of 90, and does not exceed 20% to the age of 100.

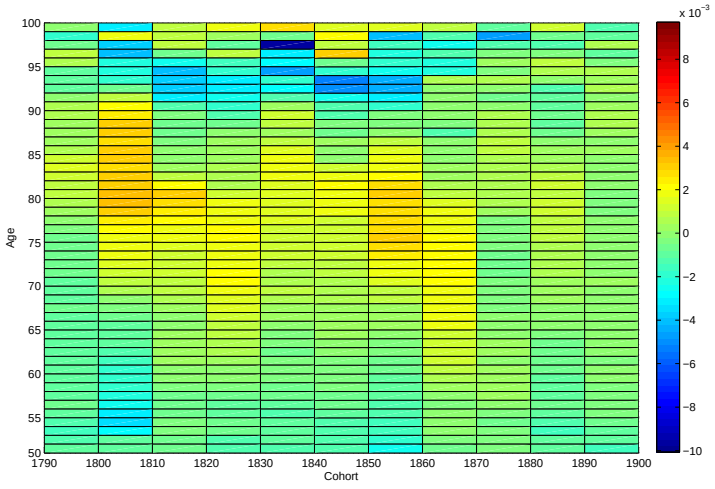


Figure 4: Calibration Residuals

5 Simulation Results for the Survival Curves

In Section 4, we have shown empirical evidence that the multi-cohort model has a satisfactory empirical fit to historical mortality data. This section presents how the calibrated model can be used to forecast the survival curve of the 1910 cohort. The choice of the

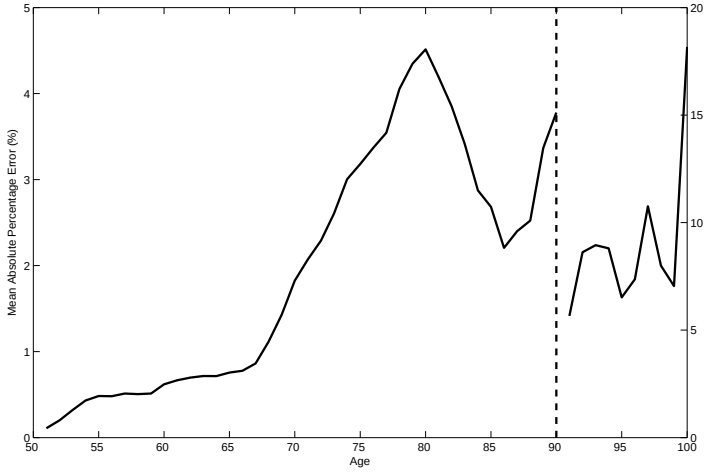


Figure 5: Mean Absolute Percentage Error

1910 cohort is made because it is the last cohort that the HMD provides death rates till an age of 100. This choice guarantees that we have the observed mortality survival curve as a comparison with our forecast survival curves.

The affine mortality model enables us to forecast cohort survival curves at future times. The states are simulated using

$$X_1(t+1) = e^{-k_{11}} X_1(t) + \sqrt{\frac{\sigma_{11}^2}{2k_{11}}(1 - e^{-2k_{11}})}\xi_1(t+1), \tag{36}$$

$$X_2(t+1) = e^{-k_{22}} X_2(t) + \sqrt{\frac{\sigma_{22}^2}{2k_{22}}(1 - e^{-2k_{22}})}\xi_2(t+1), \tag{37}$$

$$Z^i(t+1) = e^{-k_{33}} Z^i(t) + \sqrt{\frac{\sigma_{33}^2}{2k_{33}}(1 - e^{-2k_{33}})}\xi_3(t+1), \tag{38}$$

where $\xi_j(1)$, $\xi_j(2)$,..., $j = 1, 2, 3$, are independent standard normal random variables. This algorithm is specified in Appendix C.

Luciano and Vigna (2008) summarise two different ways in which the prediction can be carried out. The first one is forecasting within one cohort, the main idea of which is to forecast unknown future mortality improvements with the parameters calibrated using the observed mortality data of the same cohort. This method is adopted by Luciano and Vigna (2008) and Jevtic et al. (2013) in their forecasting process. The second one is to

forecast the mortality improvements using mortality trends by observing behaviours of other cohorts. In our multi-cohort mortality model, we focus on the cohort effects of different cohorts so we choose the second method.

To observe the mortality trends, we take the 1900 cohort as an example. Using the calibration results in Table 1 and 2, the calibrated correlation can be directly derived by Equation (25). We report the correlations between the 1900 cohort and other 5 cohorts in Table 3. We can observe that the 1900 cohort is generally more correlated with younger cohorts than older cohorts. We also note that the correlations across cohorts are not as high as the calibrated correlations reported by Jevtic et al. (2013), but are in line with the correlation matrices given in Chang and Sherris (2015). Chang and Sherris (2015) show that their correlation matrices are more realistic according to the historical experience.

Table 3: Calibrated correlations between the 1900 cohort and other cohorts

	1850	1860	1870	1880	1890
1900	0.13047	0.20919	0.19541	0.19634	0.22146

With the above observation, we model the cohort effect of the 1910 cohort using the calibration results of 10 younger cohorts, born from 1900 to 1909 with a one-year step. The estimation results are reported in Table 4. We consider three different parametrizations, Case I, Case II and Case III as shown in Table 5. Case III stands for the parametrization of the 10 younger cohorts by taking the average of the values of these 10 cohorts. Two alternative sets of simulation results, Case I and Case II, are presented to provide comparisons with Case III. The cohort specific effect as described in Case I has similar behaviour to cohorts born between 1840 and 1850, and the cohort specific effect as described in Case II has similar behaviour to cohorts born between 1870 and 1880.

Based on Table 5 and the formulas shown above ((36)-(38)), we perform 100,000 simulations for 50-year-old males born in 1910, the results are presented in Figure 6. Actual survival probabilities of the 1910 cohort are marked with asterisks. With the values of parameters specified in Case I, Case II and Case III, we show three correspondingly simulated survival curves. We can observe from the figure that there are no significant

Table 4: Estimation Results for Cohort 1910 to 1919

	k_{33}^i	σ_{33}^i	$Z^i(t)$
1900	-0.06085	0.00225	0.00099
1901	-0.05887	0.00186	0.00116
1902	-0.06585	0.00176	0.00089
1903	-0.06674	0.00180	0.00119
1904	-0.06363	0.00186	0.00113
1905	-0.07214	0.00129	0.00090
1906	-0.06445	0.00161	0.00116
1907	-0.06476	0.00179	0.00090
1908	-0.06707	0.00166	0.00079
1909	-0.06962	0.00157	0.00099

Table 5: Parametrization for Cohort Specific Factors

	k_{33}^i	σ_{33}^i	$Z^i(t)$
Case I	-0.0848	0.0007	-0.0015
Case II	-0.1034	0.0003	-0.0005
Case III	-0.0654	0.0017	0.0010

differences between the simulated curve based on the assumptions in Case III and the true survival curve, while the differences are distinguished with the assumptions specified in Case I and Case II.

6 Comparison of Cohort Survival Curves

In this section, we compare our cohort model with the age-period-cohort model developed by Renshaw and Haberman (2006) (RH model hereafter) and the age-period model developed by Cairns et al. (2006) (CBD model hereafter). We follow the notations used by Cairns et al. (2009). Cairns et al. (2009) have compared a range of mortality models qualitatively as well as quantitatively. They have evaluated properties such as the goodness of fit and the robustness of parameter estimates. In this research, we mainly focus on evaluating the ability of these three models in projecting cohort survival curves. To forecast future cohort mortality development, our cohort model uses a similar number of parameters to the CBD model but much fewer than the RH model.

We compare the simulated survival curves for the 1895, 1900, 1905 and 1910 cohorts. In

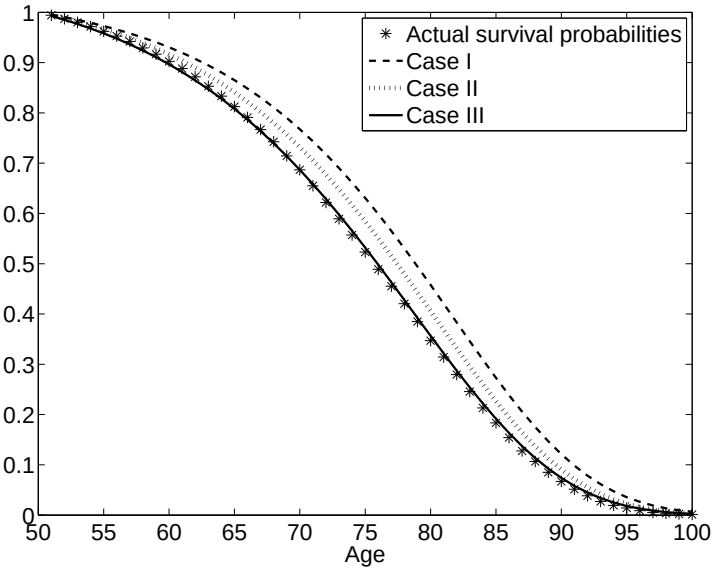


Figure 6: Actual and simulated survival probabilities for the 1910 cohort

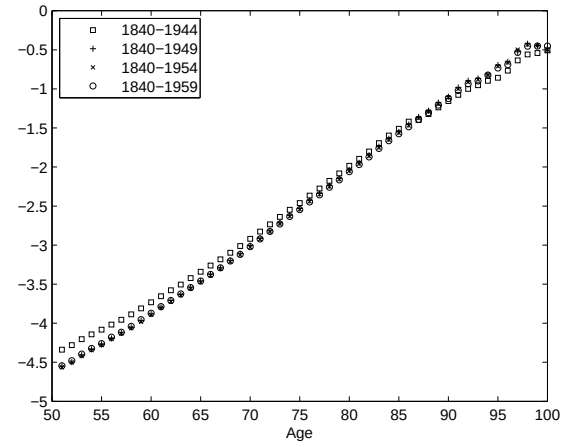
order to project cohort survival curve for the cohort born in 1895 (aged 50 in 1945), we fit all the models to the age-period mortality data of Denmark males from 1840 to 1944, aged 50 to 100. For each cohort born after 1895, we add in extra five-year age-period data and re-estimate the parameters. Then we do out-of-sample forecasting for the four cohorts and compare simulated survival probabilities with the true survival probabilities calculated using the actual death rates (which are available in HMD).

Renshaw and Haberman (2006) extend the original Lee-Carter model by including cohort effects. Following Cairns et al. (2009), in this section we define $q(t,x)$ to be the death rate at time t and age x. Then the RH model can be represented as

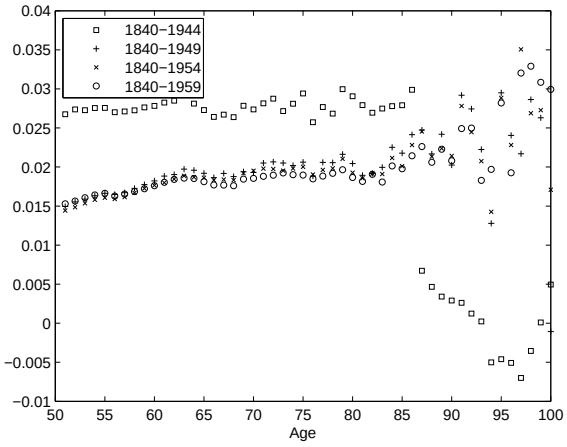
$$\log\ q(t,x) = \beta_1(x) + \beta_2(x)\kappa_2(t) + \beta_3(x)\gamma_3(t-x), \tag{39}$$

where $\beta_1(x)$ terms denote the main age effects, $\beta_2(x)\kappa_2(t)$ terms denote the age specific period effects and $\beta_3(x)\gamma_3(t-x)$ denote the age specific cohort effects. In Figure 7, we plot the fitted β_j, $j=1,2,3$, κ_2 and γ_3 based on data from 1840 to 1944, 1949, 1954 and 1959 respectively.

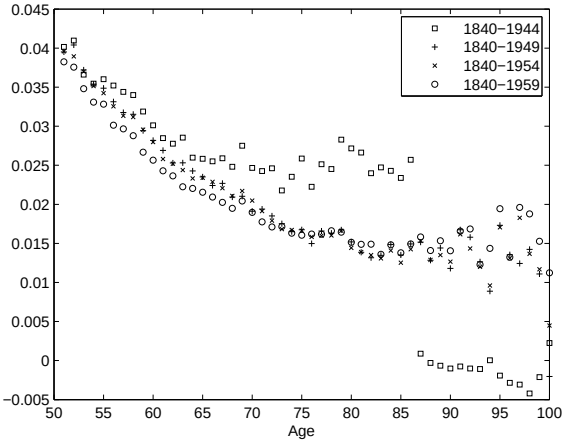
The forecasts of period effects and cohort effects are generated using autoregressive in-



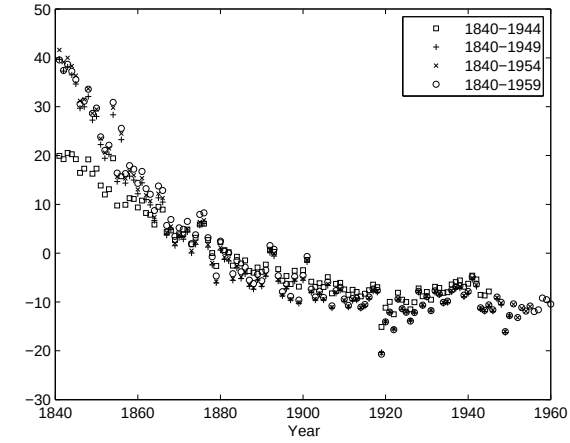
(a) $\beta_1(x)$



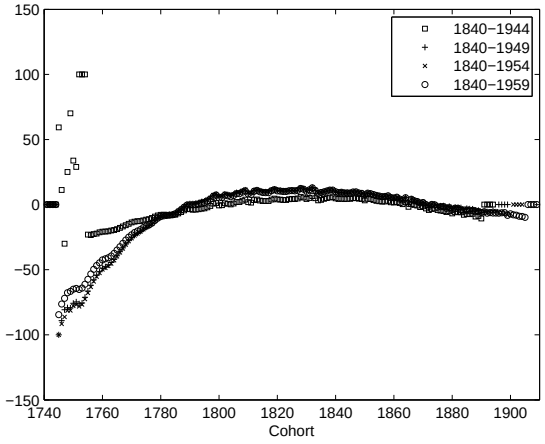
(b) $\beta_2(x)$



(c) $\beta_3(x)$



(d) $\kappa_2(t)$



(e) $\gamma_3(t-x)$

Figure 7: Parameter estimates in Denmark, males by the RH model

tegrated moving average (ARIMA) processes. In the context of Danish male mortality data, the forecasts of $\kappa_2(t)$ are generated using ARIMA(0,1,1) processes and the forecasts of $\gamma_3(t-x)$ are generated using ARIMA(0,1,0) processes.

We note the changes in the shapes of $\beta_2(x)$ and $\beta_3(x)$ in Figure 7b and 7c when the RH model is fitted to different periods of data. The lack of robustness in parameter patterns of the RH model is also noted by Cairns et al. (2009) and Haberman and Renshaw (2009) when U.S. and UK data are used. If the parameter estimates are strongly influenced by the length of historical period chosen, the reliability of the estimation may be questionable.

Cairns et al. (2006) develop a two-factor age-period model which is called the CBD model. Their model structure is

$$\text{logit } q(t,x) = \kappa_1(t) + \kappa_2(t)(x - \bar{x}), \tag{40}$$

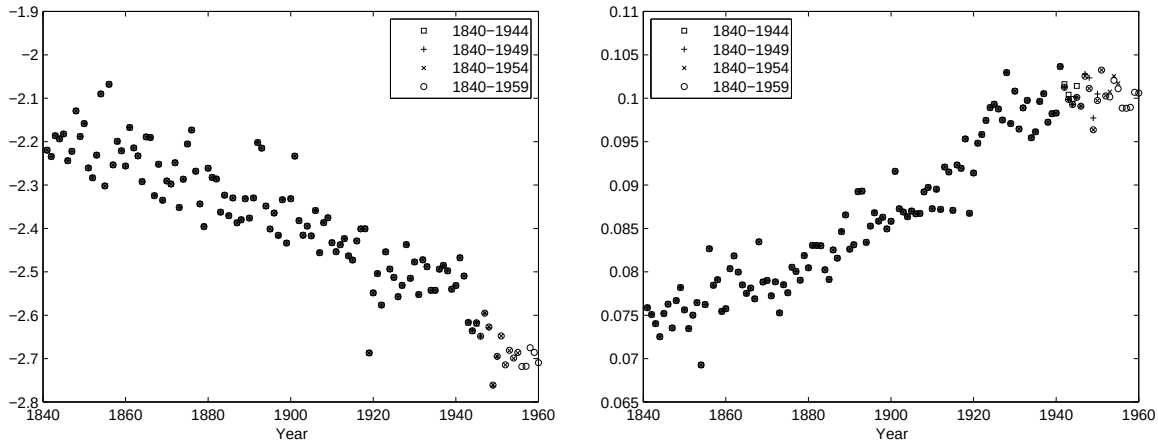
where $\bar{x}$ is the mean age in the sample range, $\kappa_1(t)$ and $\kappa_2(t)$ are time t-measurable stochastic processes. In Figure 8, we plot the fitted κ_1 and κ_2 based on data from 1840 to 1944, 1949, 1954 and 1959.

To make forecasts, the vector of factors, $\kappa(t) = \begin{pmatrix} \kappa_1(t) \\ \kappa_2(t) \end{pmatrix}$, is modelled as

$$\kappa(t+1) = \kappa(t) + \mu + C\varepsilon(t+1), \tag{41}$$

where μ is a constant 2×1 vector, C is a constant 2×2 upper triangular matrix and $\varepsilon(t)$ is a two-dimensional standard normal random variable. With the fitted values of κ_1 and κ_2 as shown in Figure 8, we have the estimated $\hat{\mu}$ and $\hat{C}$ for the four periods of data in Table 6. Both Figure 8 and Table 6 reveal that the parameter estimates for the CBD model are very robust. We can hardly distinguish the fitted κ_1 and κ_2 when we change the data period.

Both the RH model and the CBD model do not directly model mortality dynamics of



(a) $\kappa_1(t)$ (b) $\kappa_2(t)$

Figure 8: Parameter estimates in Denmark, males by the CBD model

Table 6: Estimated $\hat{\mu}$ and $\hat{C}$ for the CBD model

$\hat{\mu}$	$\hat{C}$
1840-1944	
$\begin{pmatrix} -0.00384 \\ 0.00025 \end{pmatrix}$	$\begin{pmatrix} 0.07846 & 0.00148 \\ 0 & 0.00265 \end{pmatrix}$
1840-1949	
$\begin{pmatrix} -0.00437 \\ 0.00023 \end{pmatrix}$	$\begin{pmatrix} 0.07822 & 0.00154 \\ 0 & 0.00261 \end{pmatrix}$
1840-1954	
$\begin{pmatrix} -0.00410 \\ 0.00023 \end{pmatrix}$	$\begin{pmatrix} 0.07659 & 0.00154 \\ 0 & 0.00258 \end{pmatrix}$
1840-1959	
$\begin{pmatrix} -0.00411 \\ 0.00021 \end{pmatrix}$	$\begin{pmatrix} 0.07552 & 0.00151 \\ 0 & 0.00254 \end{pmatrix}$

specific cohorts. When forecasting future mortality development using these two models, we will have a age-period surface of $q(t, x)$. Because our interest lies in cohort survival curves, we read off the diagonal of $q(t, x)$ for cohort specific mortality rates. Thus the survival probabilities of cohort i (in this case $i = t - x$) at age x over period τ can be calculated using the following equation

$$S^i(x, \tau) = \prod_{s=1}^{\tau} (1 - q(t + s - 1, x + s - 1)). \tag{42}$$

Figure 9 shows the comparison of the three simulated survival curves with the true survival curve for the four cohorts aged 50 to 100 on a logarithmic scale. In order to gain more precise insights into the characteristics of these survival curves, we use several measures as well. The results are reported in Table 7.

- The complete expectation of life at age 50 in our case is

$$\overset{o}{e}_{50} = \sum_{k=1}^{50} {}_kp_{50} + \frac{1}{2}. \tag{43}$$

- The entropy of the survival function at age 50 is defined as

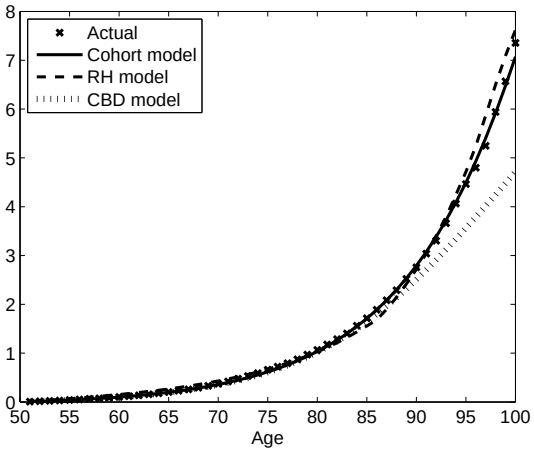
$$H(T_{50}) = -\frac{\int_{50}^{100} S(x) \ln S(x) dx}{\int_{50}^{100} S(x) dx}, \tag{44}$$

  where T_{50} denotes remaining lifetime of a person aged 50. This variability measure is used to describe the concentration of deaths. The value of H declines as deaths becoming more concentrated, and it reaches 0 if the survival curve has a perfectly rectangular shape.

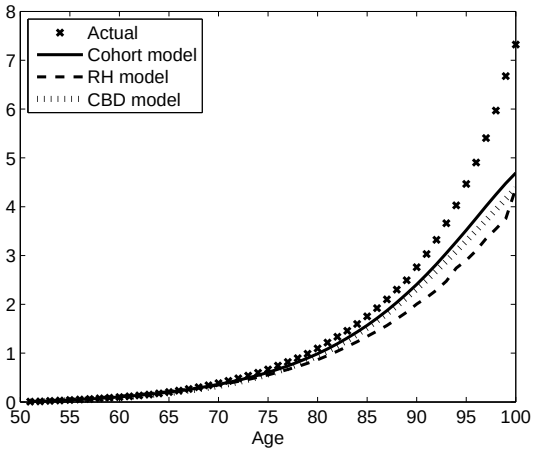
- The interquartile range concerning the distribution of T_{50} is defined as

$$IQR(T_{50}) = x'' - x', \tag{45}$$

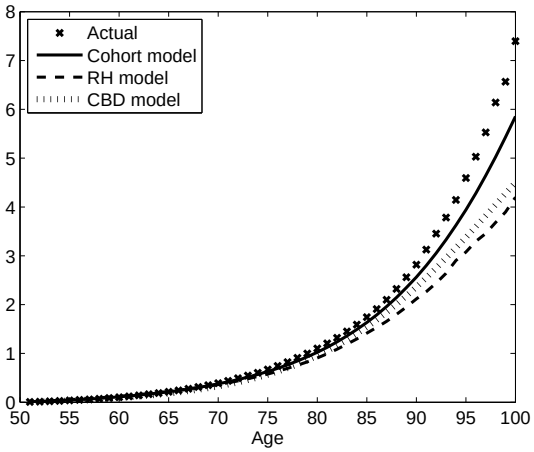
  where x' and x'' represent the 25-th percentile and 75-th percentile of the probability



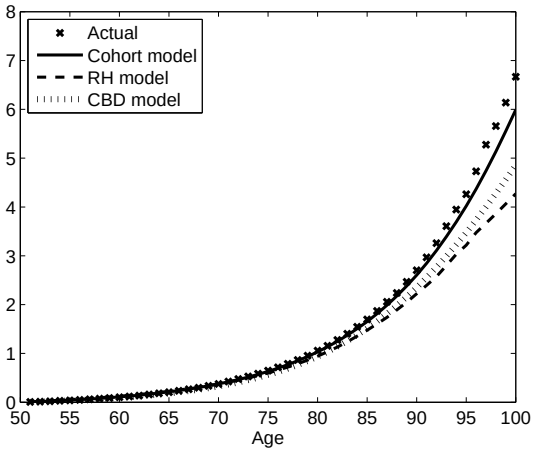
(a) The 1895 cohort



(b) The 1900 cohort



(c) The 1905 cohort



(d) The 1910 cohort

Figure 9: Actual and simulated curves for $-\log(S^i(50,\tau))$

distribution of T_{50} respectively (i.e. $S(x') = 0.75$, $S(x'') = 0.25$). The IQR decreases as the lifetime distribution becoming more concentrated.

Table 7: Demographic measures for true mortality data and three mortality models

Model	$\overset{o}{e}_{50}$	$H(T_{50})$	$IQR(T_{50})$
1895 cohort			
True data	24.92903	0.36930	15.01527
Cohort model	25.16629	0.36070	14.82586
CBD model	25.57882	0.37469	14.84059
RH model	24.30963	0.40874	17.84513
1900 cohort			
True data	24.78705	0.37090	14.84774
Cohort model	25.63367	0.38297	15.65102
CBD model	26.28802	0.37073	15.10790
RH model	26.54408	0.40059	17.44723
1905 cohort			
True data	24.68097	0.37249	15.18163
Cohort model	25.14916	0.38188	15.69336
CBD model	26.21475	0.36765	14.62593
RH model	26.08103	0.40418	17.17652
1910 cohort			
True data	24.98933	0.37159	15.18776
Cohort model	25.10262	0.37955	15.55184
CBD model	26.25764	0.43855	16.13864
RH model	25.69156	0.40562	16.97058

We can make following observations.

- The survival curve projected by the CBD model is less flexible than those projected by other models, thus the CBD model is less likely to capture the cohort effects than the mortality models taking account cohort effects. Moreover, the CBD model tends to overestimate the expectation of life, especially for younger cohorts. This tendency becomes more evident when this model is used to forecast survival curves for the 1915, 1920 and 1925 cohorts.

- While the RH model provides more flexibility in the shape of the survival curves, the estimation results for the RH model are very sensitive to the period of data employed as we have mentioned earlier. The concern that the survival curve projected by the

RH model may become unreasonable arises when we add in five-year data to forecast the survival curve for the 1915 cohort. In Figure 7, $\kappa_2(t)$ shows a declining trend up to the year of 1959. However, when we fit the model to the data from 1840 to 1964, $\kappa_2(t)$ keeps decreasing around the year of 1920, then shows an increasing trend afterwards. $\kappa_2(t)$ stands for the period effect in the RH model, a decreasing $\kappa_2(t)$ explains the improvement of mortality trend through time.

- Both Figure 9 and Table 7 indicate that the cohort model provides a better fit to the actual mortality data. It captures cohort specific effects so that it has the capability to generate more flexible and reasonable cohort survival curves.

7 Concluding Remarks

We investigate the theoretical properties and the empirical performance of a multi-cohort mortality model in the affine framework. In particular, we consider a three-factor case. The cohort mortality intensity is driven by common factors as well as cohort specific factors. Under the influence of cohort specific factors, our model allows for imperfect correlation across cohorts. A good feature of our model is that both cohort survival probabilities and correlation structure across cohorts are provided in closed-form. The three-factor multi-cohort model is then implemented on Danish male mortality data. We find that factor loadings of different cohorts have different trends, which indicates the existence of cohort effects.

Investigating cohort dynamics is crucial for pricing and risk management application. In this paper we compare cohort survival curves projected by our model with those projected by the RH model and the CBD model. Results from the comparison suggest that cohort survival curves projected by our cohort model is the closest to the actual mortality data.

An interesting question left for further research is the weighting of the latent factors, which is expressed in the value of δ_1. For simplicity we just follow Blackburn and Sherris (2013) and Jevtic et al. (2013) and assume $\delta_1 = \mathbf{1}$ for all the cohorts. A further refinement

is to allow the scalars to take different values for specific cohorts. That has important implications for the correlation structure across cohorts.

Appendix

A. Partial Differential Equation for Survival Probabilities

For simplicity, in the following we drop cohort indicator i to show the derivation algorithm. The process $\mathbf{Y}$ is an affine diffusion, which means that $\mathbf{Y}$ solves the following stochastic differential equation

$$d\mathbf{Y}(t) = H(\mathbf{Y}(t))dt + G(\mathbf{Y}(t))d\mathbf{W}^Q(t). \tag{46}$$

Survival probabilities can now be solved using the Feynman-Kac approach. We treat the conditional expected value $E^Q[e^{-\int_t^T \mu(s)ds}|\mathcal{F}_t]$ as the solution of the PDE for the survival probability $F(\mathbf{Y}, \tau)$, where $\tau = T - t$,

$$S(t,T) = E^Q[e^{-\int_t^T \mu(s)ds}|\mathcal{F}_t]. \tag{47}$$

From the above equation we have $F(\mathbf{Y}, 0) = 1$ and $F(\mathbf{Y}, \tau)$ is strictly positive. By Itô's Lemma, $F(\mathbf{Y}, \tau)$ is also an Itô process that

$$dF(\mathbf{Y}(t), \tau) = H_F(\mathbf{Y}(t), \tau)dt + F_Y(\mathbf{Y}(t), \tau)G(\mathbf{Y}(t))d\mathbf{W}^Q(t), \tag{48}$$

where

$$H_F(\mathbf{Y}, \tau) = -F_\tau(\mathbf{Y}, \tau) + F_Y(\mathbf{Y}, \tau)^T H(\mathbf{Y}) + \frac{1}{2}tr[G(\mathbf{Y})G(\mathbf{Y})^T F_{YY}(\mathbf{Y}, \tau)]. \tag{49}$$

F_τ, F_Y and F_{YY} are partial derivatives of F and tr denotes trace. We also have

$$H_F(\mathbf{Y}, \tau) - R(\mathbf{Y})F(\mathbf{Y}, \tau) = 0. \tag{50}$$

With our assumptions on $H(\mathbf{Y})$, $G(\mathbf{Y})$ and $R(\mathbf{Y})$ as

$$\begin{aligned} R(\mathbf{Y}) &= \delta_0 + \delta_1'\mathbf{Y}(t), \\ H(\mathbf{Y}(t)) &= -\Phi\mathbf{Y}(t), \\ G(\mathbf{Y}(t)) &= \Sigma s(\mathbf{Y}(t)). \end{aligned}$$

Duffie and Kan (1996) prove that the PDE can have a closed-form solution

$$F(\mathbf{Y}(t), \tau) = \exp(A(\tau) + B(\tau)^T\mathbf{Y}(t)). \tag{51}$$

From equation (49), we can deduce that

$$\frac{H_F(\mathbf{Y}, \tau)}{F(\mathbf{Y}, \tau)} = -A'(\tau) - B'(\tau)^T\mathbf{Y} + B(\tau)H(\mathbf{Y}) + \frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N B_i(\tau)B_j(\tau)G_i(\mathbf{Y})G_j(\mathbf{Y})^T. \tag{52}$$

We also know that $H_F(\mathbf{Y},\tau) = F(\mathbf{Y},\tau)R(\mathbf{Y})$. By matching the coefficients, $A(\tau) \in \mathbb{R}$, $B(\tau) \in \mathbb{R}^N$ solve the ODEs

$$A'(\tau) = -\delta_0 + \frac{1}{2}\sum_{j=1}^{N}(B(\tau)^T\Sigma)_j^2 s_{0j},$$

$$B'(\tau) = -\delta_1 - \Phi^T B(\tau) + \frac{1}{2}\sum_{j=1}^{N}(B(\tau)^T\Sigma)_j^2 s_{1j},$$

with $A(0) = B(0) = 0$.

B. Discretization of the Stochastic Differential Equations

We drop all the superscripts and subscripts. Just consider the following diffusion process

$$dX(t) = -kX(t)dt + \sigma dW(t). \tag{53}$$

Denote

$$f(X(t)) = -ke^{kt}X(t), \tag{54}$$

we have the following partial derivatives

$$\begin{aligned}\frac{\partial f(X(t))}{\partial t} &= -k^2e^{kt}X(t)\\ \frac{\partial f(X(t))}{\partial X(t)} &= -ke^{kt}\\ \frac{\partial f^2(X(t))}{\partial X(t)^2} &= 0.\end{aligned} \tag{55}$$

Applying Itô's theorem to $f(X(t))$, we have

$$\begin{aligned}f(X(t)) - f(X(0)) &= \int_0^t \frac{\partial f(X(s))}{\partial s}ds + \int_0^t \frac{\partial f(X(s))}{\partial X(s)}dX(s) + \frac{1}{2}\int_0^t \frac{\partial f^2(X(s))}{\partial X(s)^2}d(X(s))^2\\ &= \int_0^t -k^2e^{ks}X(s)ds + \int_0^t -ke^{ks}dX(s)\\ &= \int_0^t -k^2e^{ks}X(s)ds + \int_0^t -ke^{ks}(-kX(s)ds + \sigma dW(s))\\ &= -\int_0^t ke^{ks}\sigma dW(s).\end{aligned} \tag{56}$$

Plug equation (54) in equation (56)

$$\begin{aligned} -ke^{kt}X(t) + ke^{k*0}X(0) &= -\int_0^t ke^{ks}\sigma dW(s) \\ -ke^{kt}X(t) &= -kX(0) - \int_0^t ke^{ks}\sigma dW(s) \\ X(t) &= e^{-kt}X(0) + \int_0^t e^{-k(t-s)}\sigma dW(s). \end{aligned} \tag{57}$$

It is easy to obtain

$$X(t) = e^{-k}X(t-1) + \sigma \int_{t-1}^t e^{-k(t-s)}dW(s). \tag{58}$$

Therefore, given $X(t-1)$, $X(t)$ is normally distributed with mean $e^{-k}X(t-1)$ and variance $\sigma^2 \int_{t-1}^t e^{-2k(t-s)}ds = \frac{\sigma^2}{2k}(1-e^{-2k})$. We can also write

$$X_t = e^{-k}X_{t-1} + \eta_t, \tag{59}$$

where $\eta_t \sim N(0, \frac{\sigma^2}{2k}(1-e^{-2k}))$.

C. Generating Sample Paths for Multi-factor Models

Gaussian processes in $\mathbb{R}^N$ have the form

$$d\mathbf{Y}^i(t) = -\Phi^i\mathbf{Y}^i(t)dt + \Sigma^i d\mathbf{W}^i(t), \qquad i \in \mathbf{I} \subset \mathbb{R}. \tag{60}$$

It follows that the components of $(Y_1^i, ..., Y_N^i)$ satisfy

$$dY_j^i(t) = -k_j^i Y_j^i(t)dt + \sigma_j^i dW_j^i(t), \qquad j = 1, 2, ..., N, \quad i \in \mathbf{I} \subset \mathbb{R}. \tag{61}$$

The above SDE has solution

$$Y_j^i(t) = e^{-k_j^i t}Y_j^i(0) + \sigma_j^i \int_0^t e^{-k_j^i(t-s)}dW_j^i(s). \tag{62}$$

From this it follows that, for any $0 < u < t$,

$$Y_j^i(t) = e^{-k_j^i(t-u)}Y_j^i(u) + \sigma_j^i \int_u^t e^{-k_j^i(t-s)}dW_j^i(s). \tag{63}$$

We can conclude that given $Y_j^i(u)$, $Y_j^i(t)$ is normally distributed with mean $e^{-k_j^i(t-u)}Y_j^i(u)$, and variance $(\sigma_j^i)^2 \int_u^t e^{-2k_j^i(t-s)}ds = \frac{(\sigma_j^i)^2}{2k_j^i}(1-e^{-2k_j^i(t-u)})$.

Given $Y_j^i(t)$, to simulate Y_j^i at time $t+1$ we can therefore set

$$Y_j^i(t+1) = e^{-k_j^i}Y_j^i(t) + \sqrt{\frac{(\sigma_j^i)^2}{2k_j^i}(1-e^{-2k_j^i})}\xi_j^i(t+1), \tag{64}$$

where $\xi_j^i(1)$, $\xi_j^i(2)$, ... are independent standard normal random variables.

The algorithm we chosen is an exact simulation method. For simplicity, one can also use the Euler scheme

$$Y_j^i(t+1) = (1-k_j^i)Y_j^i(t) + \sigma_j^i \xi_j^i(t+1) \tag{65}$$

to simulate Y_j^i at time $t+1$.

Acknowledgements

The authors acknowledge support from Australian Research Council Centre of Excellence in Population Ageing Research (CEPAR) in providing resources and a scholarship for this research.

References

Andreev, K. and Vaupel, J. (2005). Patterns of mortality improvement over age and time in developed countries: Estimation, presentation and implications for mortality forecasting. In *Paper presented PAA annual meetings*, Philadelphia.

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

Blackburn, C. and Sherris, M. (2013). Consistent dynamic affine mortality models for longevity risk applications. *Insurance: Mathematics and Economics*, 53:64–73.

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

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

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from england and wales and the united states. *North American Actuarial Journal*, 13:1–35.

Chang, Y. and Sherris, M. (2015). A value based cohort index for longevity risk management. Working Paper No. 05/2015, ARC Centre of Excellence in Population Ageing Research.

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

Dai, Q. and Singleton, K. (2000). Specification analysis of affine term structure models. *Journal of Finance*, 55:1943–1978.

Duffie, D. and Kan, R. (1996). A yield-factor model of interest rates. *Mathematical Finance*, 6:379–406.

Gallop, A. (2008). Mortality projections in the United Kingdom. In *Presented at the Living to 100 and Beyond Symposium*, Orlando, Fla.

Haberman, S. and Renshaw, A. (2009). On age-period-cohort parametric mortality rate projections. *Insurance: Mathematics and Economics*, 45(2):255–270.

Jevtic, P., Luciano, E., and Vigna, E. (2013). Mortality surface by means of continuous time cohort models. *Insurance: Mathematics and Economics*, 53:122–133.

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

Luciano, E. and Vigna, E. (2005). Non mean reverting affine processes for stochastic mortality. Working Paper No. 4/2005, International Centre for Economic Research.

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

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

Schrager, D. F. (2006). Affine stochastic mortality. *Insurance: Mathematics and Economics*, 38:81–97.

Willets, R. C. (2004). The cohort effect: Insights and explanations. *British Actuarial Journal*, 10:833–877.