

Article



Predicting Human Mortality: Quantitative Evaluation of Four Stochastic Models

Anastasia Novokreshchenova

Dipartimento di Statistica e Matematica Applicata, corso Unione Sovietica 218 bis, Torino 10134, Italy; anastasia.novokreshchenova@carloalberto.org; Tel.: +39-011-670-5742

Academic Editor: Qihe Tang Received: 29 August 2016; Accepted: 25 November 2016; Published: 2 December 2016

Abstract: In this paper, we quantitatively compare the forecasts from four different mortality models. We consider one discrete-time model proposed by Lee and Carter (1992) and three continuous-time models: the Wills and Sherris (2011) model, the Feller process and the Ornstein-Uhlenbeck (OU) process. The first two models estimate the whole surface of mortality simultaneously, while in the latter two, each generation is modelled and calibrated separately. We calibrate the models to UK and Australian population data. We find that all the models show relatively similar absolute total error for a given dataset, except the Lee-Carter model, whose performance differs significantly. To evaluate the forecasted and the actual mortality rates and the percentage of actual mortality rates which fall within a prediction interval. In terms of the prediction intervals, the results are more divergent since each model implies a different structure for the variance of mortality rates. According to our experiments, the Wills and Sherris model produces superior results in terms of the prediction intervals. However, in terms of the mean absolute error, the OU and the Feller processes perform better. The forecasting performance of the Lee Carter model is mostly dependent on the choice of the dataset.

Keywords: stochastic mortality; affine processes; survival probability modelling; mortality probability calibration

1. Introduction

One of the main issues facing financial and governmental institutions, within the current economic climate, is the forecasting of mortality among an elderly population. Within a vast list of effected parties are public pension policies, private pension funds and life insurance businesses. They face the greatest risk, due to an increasing life expectancy across developed countries.

Over the last few decades it has become widely accepted that mortality can be more accurately measured by the use of stochastic models (see [1]), since they are better able to capture the uncertainty inherent within the problem. For any given individual, the probability of death naturally increases with age, however, as life expectancy increases over time, we observe improvements in mortality rates. Due to these effects, "dynamic mortality" has been introduced to produce models with age and time dependence. One of the seminal works, which became a benchmark within the industry, is the model of Lee and Carter [2] who model the central death rate as a two variable function. Since the publication of their work, several extensions of the Lee-Carter model have been proposed. For example, Renshaw-Haberman [3] considered a model that allows for a cohort effect and Blake and Dowd [4] proposed a two-factor model for mortality rates. Traditionally mortality models are used for forecasting mortality for older generations (ages over 50) since these mostly affect the uncertainty in the value of financial instruments offered by pension funds due to improvements in mortality and longer life expectancy (phenomena referred to in the literature as *longevity risk*). However, Plat [5] has

recently suggested a model that can fit mortality to a wider range of ages (20–89). In [6] this model has been extended to fit even younger ages (5–89).

A fairly recent stream of actuarial literature has dealt with the phenomenon of stochastic mortality by modelling the instantaneous mortality intensity as a stochastic process. Recent works include Milevsky and Promislow [7], Dahl [8], Biffis [9], Denuit and Devolder [10], Luciano and Vigna [11], Shrager [12]. The mathematical framework in these models has been adapted from the credit risk literature to value securities subject to risk to default. Similarities between the time to default and remaining lifetime and between short-term interest rate and the force of mortality are exploited in this approach. Moreover, if the intensity process is affine, then the survival function for an individual can be derived in a closed form. This is extremely useful when pricing mortality-linked financial products, such as endowments, annuities, variable annuities and other forms of mortality-linked financial securities.

Luciano and Vigna [11] have studied the applicability of the affine processes, such as Ornstein-Uhlenbeck and Feller, for modelling mortality intensities. The approach is focused on fitting the survival curve for which closed-form solutions are available. The future projections for survival probabilities are made, their closeness to the historical values is discussed, but not evaluated quantitatively.

Another continuous-time mortality model we consider in this work is the one proposed by Wills and Sherris (2011) [13] for the Australian population. As with the Lee and Carter (1992) [2] model, it is able to capture the whole "mortality surface" across age and period. Moreover, it takes account of the correlation structure between different generations. This is important for life offices portfolios which often have contracts written on individuals from different cohorts. The authors have shown that the multiple risk factors implied by the model reflect the actual correlation structure between generations inferred from the data and that the model is suitable for pricing financial instruments (see Wills and Sherris [13,14]).

The advantages of continuous time mortality models mean that it is important to study how well continuous time processes can predict future mortality. There are numerous papers comparing the performance of mortality models—[5,6,15–18] are among them. Nevertheless, most of them have focused on discrete stochastic mortality models. For example, Cairns et al. [15] examined the in-sample fits of eight different discrete time stochastic mortality models. However, as noted in Dowd et al. [16], it is quite possible for a model to provide a good in-sample fit to historical data and produce forecasts that appear plausible *ex ante*, but still produce poor *ex-post* forecasts, that is, forecasts that differ significantly from the subsequently realised outcomes. Consequently, a "good" model should produce forecasts that perform well out-of-sample which can be evaluated using backtesting methods.

Lee and Miller [19] evaluated the performance of the Lee-Carter model by examining the behaviour of forecast errors and plots of "percentile error distributions", although they did not report any formal test results. In contrast, Dowd et al. [16] formally evaluate the forecasting performance of six different stochastic mortality models applied to male mortality data for England and Wales. They use a backtesting procedure to test the stability of forecasts over different time horizons and conclude that the investigated models perform adequately, and that there is little difference between them.

The framework for backtesting stochastic mortality models in Dowd et al. [16] is a very general one. The "backtests" might involve the use of plots whose goodness of fit is interpreted informally, as well as formal statistical tests of predictions. The evaluation can be done for different metrics (the forecasted variable) of interest – possible metrics include mortality rates, life expectancy, future survival rates, the prices of annuities and other life-contingent financial instruments.

This paper focuses on the forecasting performance of several continuous-time models, making a novel contribution to the literature. More specifically, we concentrate on the following continuous-time mortality models: the Ornstein-Uhlenbeck process, the Feller process and the Wills and Sherris model.

To compare the performance of these models to a benchmark, we also include the Lee and Carter model in our experiments. We evaluate the in-sample goodness of fit by using statistical techniques

including the BIC criteria and an analysis of the fitted residuals. To assess the forecasting performance of each model we employ out-of-sample back-testing methods using mortality rates as the metric. This is done first by computing the relative error between the forecast and actual mortality rates and then by looking at the percentage of observed mortality rates which fall within a prediction interval. However, the same backtesting procedure using different metrics might be relevant for different purposes.

For our analysis we employ the data of the British and Australian population as they are among the countries where the market for mortality derivatives has started to emerge. According to [20], the annuity markets are relatively well developed in the UK and US. Some product innovations, such as variable annuities with guaranteed withdrawal lifetime benefits have been introduced in Australia, Japan and Europe. Multiple mortality and longevity derivatives (such as q-forwards, s-forwards, longevity and survivor bonds and swaps) have been suggested in the literature as well, see [14,21]. In [14,20] the authors study the securitisation of longevity risk for the Australian pension industry. In [22] natural hedging of longevity risk with application to the UK population is analysed.

This paper is organised as follows: in Section 2 we present some notation and description of the data that will be used in the subsequent analysis. In Section 3 we provide an overview of the Lee-Carter model, which we will use as a benchmark for our comparisons. Section 4 provides the Wills and Sherris model setup. Section 5 describes time-homogeneous affine processes. Section 6 calibrates the four models to the UK female dataset and Section 7 compares the results of this calibration for the four models. Section 8 discusses the robustness of the simulation results on the male and female datasets for the British and Australian populations. Section 9 concludes.

2. Notation and Data Description

Throughout the paper we use the following notation. Define m(x, t) to be the *observed central death rate* in year *t* for lives initially aged *x* as a number of deaths divided by the population exposure:

$$m(x,t) = \frac{D(x,t)}{E(x,t)},\tag{1}$$

Here E(t, x) is the average size of the population aged x last birthday during year t and D(t, x) is the number of deaths during year t recorded as age x last birthday at the date of death. The observed central death rate can be calculated directly from the data.

Another measure of mortality is the *force of mortality* $\mu(x, t)$. It is interpreted as the instantaneous death rate at exact time *t* for individuals aged exactly *x* at time *t*. The probability of death between *t* and t + dt for small *t* is then approximately $\mu(x, t) \times dt$. Thus, assuming that the force of mortality remains constant over a year: $\mu(x + s, t + u) = \mu(x, t)$ for $0 \le s, u < 1$, we can approximate the force of mortality rate m(x, t).

A typical dataset consists of a number of deaths, D(x,t), and the corresponding exposures, E(x,t), over a range of years t and ages x. The data for the UK we use in this study contains the number of deaths and the population exposure. It was taken from the Human Mortality Database (Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany)). We consider female population aged 50–99 (which is relevant to the pension fund industry) during the years 1970–2009.

3. Lee-Carter Model

In this section we describe general characteristics of the famous Lee Carter model [2] and its estimation process and forecasting technique. Lee-Carter mortality model is used widely in academia, as well as industry. It has been proposed by Lee and Carter in 1992 specifically for US mortality data covering years 1933–1987. However it has been used as a benchmark model to mortality data from many countries and time-periods. It has been shown (see [23]) that the Lee-Carter model is a special type of multivariate random walk with a drift (RWD), in which the covariance matrix depends on the

drift vector. For estimation of the model parameters, principle component analysis (PCA) with a single component is applied to the census data.

Let m_{xt} denote the log of the mortality rate in an age group x (x = 1, ..., A) and time t (t = 1, ..., T) for one country. The mortality rate is modelled as follows:

$$m_{xt} = \alpha_x + \beta_x \kappa_t + \epsilon_{xt} \tag{2}$$

where ϵ_{xt} is a set of random disturbances and α_x , β_x and κ_t are parameters to be estimated:

- α_x is the average mortality curve across ages;
- β_x is a set of parameters representing the sensitivity of the mortality rate at age x to changes in κ_t ;
- $-\kappa_t$ is a time-varying parameter representing a common risk factor;
- ϵ_{xt} is a zero mean Gaussian error $N(0, \sigma^2)$.

The parametrisation in (2) is not unique, since the likelihood function associated with the model above has an infinite number of equivalent maxima, each of which would produce identical forecasts, see Lee and Carter [2]. In practice, model identification implies imposing constraints. Lee and Carter adopt the constraints $\sum_{t} \kappa_t = 0$ and $\sum_{x} \beta_x = 1$.

The constraint $\sum_t \kappa_t = 0$ implies that the parameter α_x is simply the empirical average over time of the age profile in age group a: $\alpha_x = \bar{m}_x$. We can therefore rewrite the model in terms of the mean centered log-mortality rate, $\bar{m}_{xt} = m_{xt} - \bar{m}_x$. Thus, we can rewrite Equation (2) as a multiplicative fixed effects model for the centered age profile:

$$\tilde{m}_{xt} \sim N(\bar{\mu}_{xt}, \sigma^2),$$

$$E(\tilde{m}_{xt}) = \bar{\mu}_{xt} = \beta_x \kappa_t.$$
(3)

As a result, we use A + T parameters (with A and T being the total number ages and the total number of years considered) to approximate the $A \times T$ elements of the mortality matrix, where each row represents the age of the population and each column represents the year of the observation, with the age-specific parameter β_x which is fixed over time for all x and the time-specific parameter κ_t which is fixed over age groups for all t.

The parameters β_x and κ_t in the model can be found easily using singular value decomposition (SVD) of the matrix of centered age profiles, $\tilde{m} = BLU' = Z$, which we denote by Z. Then the estimate for β_x is the first column of B, $\mathbf{b_1}$ (normalised eigenvector of the matrix ZZ') and the estimate for κ_t is $\lambda_1 \mathbf{u_1}$, where $\mathbf{u_1}$ is the first column of the matrix U (normalised eigenvector of the matrix Z'Z) and λ_1 is the first element of the diagonal matrix L (the largest eigenvalue corresponding to the eigenvectors). Typically, for low-mortality populations, the approximation $Z \approx \lambda_1 \mathbf{b_1} \mathbf{u'_1}$ accounts for more than 90% of the variance of m_{xt} , see [23].

To forecast future mortality, Lee and Carter assume that α_x and β_x remain constant over time and the time factor γ_t is viewed as a stochastic process. They find that a random walk with drift is the most appropriate model for their data:

$$\begin{aligned} \hat{\kappa}_t &= \hat{\kappa}_{t-1} + \theta + \xi_t; \\ \xi_t &\sim N(0, \sigma_{rw}^2), \end{aligned} \tag{4}$$

where θ is known as the drift parameter and its maximum likelihood estimate is simply $\hat{\theta} = (\hat{\kappa}_T - \hat{\kappa}_1)/(T-1)$, which only depends on the first and last components of the κ_t vector.

We can forecast $\hat{\kappa}_t$ at time T + h with data available up to period T, as follow:

$$\hat{\kappa}_{T+h} = \hat{\kappa}_T + h\hat{\theta} + \sum_{l=1}^h \xi_{T+l-1}.$$
(5)

From this model, we can obtain forecast point estimates, which follow a straight line as a function of *h* with slope $\hat{\theta}$:

$$E[\hat{\kappa}_{T+h}|\hat{\kappa}_1,\ldots,\hat{\kappa}_T] = \hat{\kappa}_T + h\hat{\theta}.$$
(6)

To make a point estimate forecast for log-mortality we plug the obtained expression for $\hat{\kappa}_{T+h}$ into the vectorised version of expression (3):

$$\mu_{T+(\Delta t)} = \bar{m} + \hat{\beta}_x \hat{\kappa}_{T+h} = \bar{m} + \hat{\beta}_x [\hat{\kappa}_T + h\hat{\theta}], \tag{7}$$

where $\hat{\beta}_x = \mathbf{b_1}$ and $\hat{\kappa}_T = \lambda_1 \mathbf{u_1}$ are the estimates of β_x and κ_T respectfully obtained using SVD.

4. Wills and Sherris Model

Wills and Sherris suggested a stochastic longevity model where the force of mortality for age x at time t has the following dynamics (see [13,14]):

$$d\mu(x,t) = (a(x+t)+b)\mu(x,t)dt + \sigma\mu(x,t)dW(x,t),$$

$$0 < x < \omega, \quad 0 < t < \omega - x.$$
(8)

In the above expression the drift parameter is an affine function of the current age (x + t), while volatility function is a constant. Applying the Ito's lemma, we find the solution to the SDE (8):

$$\mu(x,t) = \mu(x,0) \exp\left[\frac{a}{2}t^2 + (ax+b-\frac{1}{2}\sigma^2)t + \sigma W(x,t)\right],$$

which can be written as follows:

$$\ln\left[\frac{\mu(x,t)}{\mu(x,0)}\right] = \left[\frac{a}{2}t + ax + b - \frac{1}{2}\sigma^2\right]t + \sigma W(x,t).$$

For all ages x_1, \ldots, x_N , we consider a multivariate random vector of mortality rates:

$$\underline{\mu}(x,t) = \begin{bmatrix} \mu(x_1,t) \\ \vdots \\ \mu(x_N,t) \end{bmatrix}$$

The dynamics $\underline{d\mu}(x,t)$ are assumed to be driven by the multivariate Wiener process $\underline{dW}(x,t)$, with mean zero and the instantaneous correlation matrix given by:

$$D = \begin{bmatrix} \delta_{11} & \dots & \delta_{1N} \\ \vdots & \ddots & \vdots \\ \delta_{N1} & \dots & \delta_{NN} \end{bmatrix}$$

This means that the Wiener processes are independent between time periods, but correlated between ages and the multivariate Wiener process $\underline{dW}(x,t)$ can be expressed in terms of independent Wiener process $\underline{dZ}(x,t) = [dZ_1(t), \dots, dZ_N(t)]'$ as dW(x,t) = DdZ(x,t).

Thus, the model described by Equation (8), becomes a system of equations where the dependence between the ages is captured by the $\delta_{x,i}$ term:

$$d\mu(x,t) = (a(x+t)+b)\mu(x,t)dt + \sigma\mu(x,t)\sum_{i=1}^N \delta_{x,i}dZ_i(t) \quad \forall x = x_1,\ldots,x_N.$$

Using the fact that the distribution of the changes in the force of mortality follows a normal distribution, we can find the parameters \hat{a} , \hat{b} and $\hat{\sigma}$ by means of maximum likelihood estimation. In particular,

$$\Delta \mu(x,t) \sim N((a(x+t)+b)\mu,\sigma\mu)$$

To estimate the covariance matrix of $\underline{dW}(x, t)$, we apply Principle Component Analysis (PCA) to the standardised residuals of the model. For each year, they are the realisations of the random vector $\underline{dW}(x, t)$:

$$r(x,t) = \frac{\Delta\hat{\mu}(x,t)/\hat{\mu}(x,t) - (\hat{a}(x+t) + \hat{b})}{\hat{\sigma}}$$

5. Time-Homogeneous Affine Processes

Mortality intensity since recently has been modelled as a stochastic process, (see Cairns [1]). In this field, an important stream of literature focuses on describing death arrival as the first jump time of a Poisson process with stochastic intensity. This approach is named doubly stochastic. Milevsky and Promislow [7] have used a stochastic force of mortality, whose expectation at any future date has a Gompertz specification. Dahl [8], Biffis [9], Denuit and Devolder [10] and Schrager [12] in modelling the stochastic force of mortality have applied the same mathematical tools used in the credit risk literature to model the time to default. Under this setting, the remaining lifetime of an individual, τ , is a doubly stochastic stopping time with intensity λ .

Let the mortality process $\mu_x(t)$ represent the mortality intensity of an individual belonging to the generation x at (calendar) time t and τ be the time of death of an individual of generation x. Then the survival probability from time t to time $T \ge t$ is defined as a function of τ , $S_x(t, T)$, under the probability measure P, conditional on the survivorship up to time t:

$$S_x(t,T) = \P(\tau \ge T | \tau > t), \tag{9}$$

A doubly stochastic stopping time is the analogue of the first jump time of a Poisson process, where the intensity is a stochastic process. If τ is the first jump time of a Poisson process with parameter μ , then

$$P(\tau > t) = e^{-\mu t} \tag{10}$$

Similarly, if τ is doubly stochastic with intensity μ , then the individual's survival function $S_x(t)$ is given by

$$S_x(t,T) = P(\tau > t|F_s) = E\left(e^{-\int_s^t \mu(u)du}|F_s\right)$$
(11)

where F_s describes the information at time s.

In general, the expectation in (11) is not easy to calculate. However, if the intensity process is affine (see Duffie, Filipovic and Schachermayer [24]), then it is possible to provide the closed form for the survival probability:

$$S_{x}(t,T) = e^{\alpha(T-t) + \beta(T-t)\mu_{x}(t)}.$$
(12)

where the functions $\alpha(\cdot)$ and $\beta(\cdot)$ satisfy generalised Riccati ODEs, which can be solved analytically or at least numerically. The closed-from expression of survival probabilities (12) in affine framework allows to price financial instruments written on the underlying population, such as endowments, annuities, variable annuities and other forms of mortality-linked financial securities. Due to this result, in applications the processes selected for the mortality intensity are typically affine.

Luciano and Vigna [11] proposed and tested time-homogeneous non-mean reverting affine processes for the intensity of mortality, which are natural generalisation of the Gompertz law of mortality. They consider Ornstein Uhlenbeck process, Ornstein Uhlenbeck process with jumps and the Feller process. They provide the analytical solutions for survival function (12) for these processes and discuss the appropriateness of using them in modelling mortality. Calibrations on historical data show

that despite their simple form, these processes fit mortality intensity dynamics very well. Another study shows how to use these processes to delta-gamma hedge mortality and interest rate risk, see Luciano, Regis and Vigna [22].

5.1. The Ornstein-Uhlenbeck Processes

The SDE for the Ornstein-Uhlenbeck (OU) process without mean-reversion, with the associated solutions of the Riccati ODE $\alpha(\cdot)$ and $\beta(\cdot)$, is the following:

$$d\mu_{x}(t) = a\mu_{x}(t)dt + \sigma dW_{x}(t),$$

$$\alpha(t) = \frac{\sigma^{2}}{2a^{2}}t - \frac{\sigma^{2}}{a^{3}}e^{at} + \frac{\sigma^{2}}{4a^{3}}e^{2at} + \frac{3\sigma^{2}}{4a^{3}},$$

$$\beta(t) = \frac{1}{a}(1 - e^{at}),$$
(13)

where a > 0, $\sigma > 0$.

We calibrate the parameters of the OU process by means of Maximum Likelihood method applied to the mortality intensities.

Assume that the dynamics of the mortality intensity is described by the OU process without mean reversion as given by SDE (13). Then, the conditional probability density of an observation μ_{i+1} , given a previous observation μ_i (with a δ time step between them), has a form (here we omit x which symbolises a certain generation):

$$f(\boldsymbol{\mu}_{i+1}|\boldsymbol{\mu}_i;a,\hat{\boldsymbol{\sigma}}) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}}e^{-\frac{(\boldsymbol{\mu}_{i+1}-\boldsymbol{\mu}_ie^{a\hat{\boldsymbol{\sigma}}})^2}{2\hat{\sigma}^2}},$$

where $\hat{\sigma}^2 = \sigma^2 \frac{1 - e^{2a\delta}}{2a}$.

The log-likelihood function of a set of observations $\bar{\mu} = (\mu_1, \mu_2, \dots, \mu_n)$ can be derived from the conditional density function:

$$L(\bar{\mu}; a, \hat{\sigma}) = \sum_{i=1}^{n} \ln f(\mu_{i+1} | \mu_i; a, \hat{\sigma}) =$$

= $-\frac{n}{2} \ln(2\pi) - n \ln(\hat{\sigma}) - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^{n} (\mu_{i+1} - \mu_i e^{a\delta})^2.$ (14)

From the Maximum Likelihood conditions we find the following equations for the parameters:

$$a = \frac{1}{\delta} \frac{\sum_{i=1}^{n} \mu_{i+1} \mu_{i}}{\sum_{i=1}^{n} \mu_{i}^{2}}$$

$$\hat{\sigma}^{2} = \frac{\sum_{i=1}^{n} (\mu_{i+1} - \mu_{i} e^{-a\delta})^{2}}{n}$$
(15)

The OU process in general can produce negative paths. The probability of λ_x turning negative is

$$P(\mu_{x}(t) \leq 0) = \Phi\left(-\frac{\mu_{x}(0)e^{at}}{\sigma\sqrt{\frac{e^{2at}-1}{2a}}}\right) = \Phi\left(\zeta(\sigma, a)\right),\tag{16}$$

where Φ is the distribution function of a standard normal. In fact, the function $\zeta(\sigma, a) = -\frac{\mu_x(0)e^{at}}{\sigma\sqrt{\frac{e^{2at}-1}{2a}}}$ is increasing in σ and decreasing in a, as well as the probability of negative values of μ . In mortality modelling applications the probability that $\mu(t)$ takes negative values is very small, because in practice the obtained value of σ is small enough and the value of *a*, on the contrary, is high enough. In our calibration we check that the values of the obtained

parameters are such that there are no negative intensities. Otherwise, to keep mortality intensity positive, it is possible to impose a restriction using Equation (16) during the parameter search, such that probability (16) is negligible (see Luciano and Vigna [11] for more details).

5.2. The Feller Process

The fourth model is the Feller process which is described by the following SDE with the associated solutions of the Riccati ODEs $\alpha(\cdot)$ and $\beta(\cdot)$:

$$d\mu(t) = a\mu(t)dt + \sigma \sqrt{\mu(t)}dW(t),$$

$$\alpha(t) = 0,$$

$$\beta(t) = \frac{1 - e^{bt}}{c + de^{bt}},$$
(17)

where a > 0, $\sigma \ge 0$, the boundary conditions are $\alpha(0) = 0$ and $\beta(0) = 0$, and the coefficients are:

$$b = -\sqrt{a^2 + 2\sigma^2}$$

$$c = \frac{b+a}{2},$$

$$d = \frac{b-a}{2}.$$
(18)

The solution to the Equation (17) has the form:

$$\mu(t) = \mu(0)e^{at} + \sigma \int_0^t e^{a(t-u)} dW(u)$$
(19)

The Feller process is a type of the Cox, Ingersoll, Ross (1985) process [25] without mean reversion. It was proposed as a model of a short rate for financial market, referred to as the CIR model. This model is described by the following SDE:

$$dr(t) = a(b - r(t))dt + \sigma\sqrt{r(t)}dW(t),$$
(20)

where b > 0 is the mean-reversion level. Thus, the model suggests that the r(t) is pulled towards b at a speed controlled by a. If condition $2ab > \sigma^2$ holds and r(0) > 0, then the CIR process remains strictly positive, almost surely, and the state (marginal) distribution of the process is steady. The marginal density is gamma-distributed. The maximum likelihood estimation of the parameter vector $\theta = (a, b, \sigma)$ is based on the transition density. Given r_t at time t the density of $r_{t+\Delta t}$ at time $t + \Delta t$ is

$$p(r_{t+\Delta t}|r_t;\theta) = ce^{-u-v} \left(\frac{v}{u}\right)^{\frac{q}{2}} I_q(2\sqrt{uv}),\tag{21}$$

where

$$c = \frac{2a}{\sigma^2(1 - e^{-a\Delta t})}$$

$$u = cr_t e^{-a\Delta t},$$

$$v = cr_{t+\Delta t},$$

$$q = \frac{2ab}{\sigma^2} - 1,$$
(22)

and $I_q(2\sqrt{uv})$ is modified Bessel function of the first kind and of order q. Then the likelihood function of the time series $(r_1 \dots r_N)$ with the time between two observations $\Delta t = 1$ is

$$L(\theta) = \prod_{i=1}^{N-1} p(r_{i+1}|r_i;\theta),$$
(23)

from which the log-likelihood function of the CIR process is derived:

$$\ln L(\theta) = (N-1)\ln c + \sum_{i=1}^{N-1} \left[-u_i - v_{i+1} + 0.5q\ln\left(\frac{v_{i+1}}{u_i}\right) + \ln(I_q(2\sqrt{u_iv_{i+1}})) \right],$$
(24)

where $u_i = cr_i e^{-a\Delta t}$ and $v_{i+1} = cr_{i+1}$.

There are two approaches to calibrating affine mortality processes to the historical data. One is to match the survival function (Equation (12)) using the solutions of the Riccati ODEs for the OU (Equation (13)) and the Feller (Equation (17)) processes to the set of observed survival probabilities. Another approach is to maximise the likelihood function of the transition density. In this work we employ the second approach as described in [26] since both for the OU process and the Feller process the transition density is known in closed-form.

6. Models Calibration

In this section we work with the UK female dataset which describes the mortality in population aged 50–99 for the years 1970–2009. First, we divide the data in two data sets: the estimation data set, containing 30 years of observations, from 1970 until 1999; and the backtesting data set containing the last 10 years of observations, from 2000 till 2009. First we estimate the model parameters on the estimation data set, then we make 10-years predictions of mortality rates and calculate how well the forecast is compared with actual mortality rates for the period 2000–2009.

For the Lee and Carter and the Wills and Sherris models we use the whole surface of mortality to calibrate the models. Then, to compare the performance of the models between each other, we chose 19 generations. To have reliable estimation results and to make the comparison between the models which simulate the whole mortality surface (the Lee Carter and the Wills and Sherris models) and the ones which model each generation separately (the OU and the Feller processes) fairer, we take all possible generations from the data, which satisfy the criteria that the length of the backtesting period would not be less than 10 years. This results in 19 generations—aged 42–60 in the year 1970. We obtain the mortality rates for corresponding generations from the surface by taking the relative diagonal of the matrix. For the OU and the Feller processes, however, we calibrate the parameters for each of the 19 generations separately. We calculate the parameters on the estimation time period (1970–1999) and then use them to make forecasts of mortality for the next 10 years. Thus, we build forecasts for these generations and compute the relative error of prediction, as well as the percentage value of the actual mortality rates which fall within the prediction interval in the test period 2000–2009.

6.1. Calibration of the Lee-Carter Model

First of all, we compute the average of the log mortality m_{xt} for every age over time period 1970-1999 for the estimation dataset and subtract it in order to obtain mean centered log-mortality rates, $\tilde{m}_{xt} = m_{xt} - \bar{m}_x$. The average of the log mortality for the whole dataset is shown in Figure 1. Then, we perform SVD on \tilde{m}_{xt} matrix and obtain estimates for parameters – two vectors $\hat{\beta}_x$ and $\hat{\kappa}_t$. The actual centered mortality and its SVD approximation are illustrated in Figure 2.

The obtained ML estimates for the drift and the variance of the innovations are $\hat{\theta}_{ML} = -0.5992$ and $\hat{\sigma}_{rw}^2 = 0.9154$, respectively. Using these parameters we can compute the forecast for κ_t as given by Equation (4) and its forecast point estimate as described in Equation (6). In Figure 3a the estimated vector of κ_t and its forecast obtained for the next 10 years (in red) are shown. Then, we calculate the forecast for log-mortality as given in Equation (7). Figure 3b shows the mortality for the 10 years forecasted by the Lee-Carter. The forecast corresponds well to the observed mortality rates for the UK female population presented in Figure 1b. However, we can see that the cohort effect (diagonal trends in the data present in Figure 1b) is not captured by the Lee-Carter forecast of mortality.



Figure 1. Observed mortality rates for the UK female population. (a) Estimation data set, 1970–1999; (b) Backtesting data set, 2000–2009.



Figure 2. Actual centered mortality and its approximation. (**a**) Mean centered log-mortality rate; (**b**) Approximation by 1-factor SDV.



Figure 3. Results of the Lee-Carter model. (a) Estimation and forecast of κ_t ; (b) Lee-Carter forecast for 2000–2009.

6.2. Calibration of the Wills and Sherris Model

The analysis of the fit is based on the assumption that the residuals are independent and identically distributed normal variables with mean 0 and standard deviation 1. Figure 4a shows the graph of residuals for the UK female population aged 50–99, years 1970–1999 (the estimation dataset). The plot, together with the residuals descriptive statistics in Table 1, supports the hypothesis that the residuals follow a standard normal distribution with mean close to zero and standard deviation very close to one. The table also contains the value of the log likelihood function, the BIC criteria and the value of the χ -square statistics.

The Bayesian information criteria (BIC) is defined as

$$BIC = -2\ln(\hat{L}) + k\ln(n),$$

where:

n – the number of observations (sample size);

k – the number of free parameters to be estimated.

 \hat{L} – the maximized value of the likelihood function of the model.

Pearson's chi-square statistics, defined as $\chi^2 = \sum_{observations} \frac{(O_i - E_i)^2}{E_i}$, allows us to evaluate the goodness of fit by testing wether or not an observed frequency distribution differs from the theoretical one. We compare weather the computed value of χ^2 with the critical value of the statistic with degree of freedom defined as

df = number of observations – number of independent parameters – 1.

The obtained value of the χ^2 is 1.5835. This is compared to the chi-square distribution with 217 degrees of freedom (49 * 29 - (49²/2 + 3) - 1). Higher values of the χ^2 statistic suggest a poorer fit. Since the calculated value is very low, the test confirms a very good fit to the data.



Figure 4. Results of the Wills and Sherris model. (**a**) UK female 1970-1999 fitted residual; (**b**) Mortality forecast for the test period.

To capture the correlation structure between ages we calculate eigenvectors of the matrix of the obtained residuals using Principal Component Analysis. Table 2 summarises the percentage of the observed variation explained by these vectors. The observed age-correlation matrix has a total of 49 eigenvalues. In our experiments we take first 30 eigenvectors to approximate the correlation matrix as they account for 98.9% of the variation.

Parameter Estimates					
a	0.0007032				
b	0.0850				
σ	0.0385				
Log-likelihood	7246.4				
BIC-criteria	-14480				
Residual Descrip	Residual Descriptive Statistics				
mean	1.3192×10^{-15}				
Minimum	-5.1384				
Maximum	2.9886				
Standard Deviation	1.0004				
Standard Error	0.0125				
Confidence Level	0.0003				
χ^2	1.5835				

Table 1. Parameter estimates and residual descriptive statistics for the Wills and Sherris model fit toUK female mortality rates 1970–1999.

Table 2. Percentage of the observed variation in residuals explained by the eigenvectors using PCA.

Number of Eigenvectors	% of Observed Variation
1	28.1
5	55.8
10	75.4
15	86.5
20	93.1
25	96.8
30	98.9
35	100

Figure 4b shows the mortality surface for the test period built with the Wills and Sherris model using the parameters obtained on the estimation dataset. By comparing the forecast with the actual mortality rates (Figure 1b), we can see that the model gives projections which are similar to the real data, although we see more variation in the simulated mortality intensities. In order to obtain a reliable prediction of mortality rates for a particular generation, we perform Monte Carlo simulations of the mortality surface for the test period and extract from the surface a diagonal corresponding to a specific generation. Then we estimate the mean of the Monte Carlo simulations for a given generation, together with the 90% prediction interval.

6.3. Calibration of the OU-Process

We calibrate the model on 19 generations and evaluate the goodness of fit by means of the BIC criteria and analysis of the residuals. For each generation x, having a series of length N we use n = N - 10 observations (first n years of the sample) to estimate the parameters a and σ and last 10 observation for backtesting the results. For instance, for the UK data, if we consider individuals who were 50 in the year 1970, and we have the data until the year 2009, we have 40 years of observations. Then the first thirty years of observations (1970–1999) is the estimation data and the last ten years of observations (2000–2009) is the backtesting data.

After obtaining the parameters we use the following simulation equation to generate paths of the mortality intensities. This expression is an exact solution of the SDE (13):

$$\mu_{i-1} = \mu_i e^{a\delta} + \sigma \sqrt{\frac{1 - e^{2a\delta}}{-2a}} N_{0,1}.$$
(25)

From the mortality intensities one can easily obtain the survival probabilities by means of the analytical formula 12 with $t = T_{j+1} - T_j = 1$. In this way the expression for a one-year survival curve with α and β being constants is:

$$S_{x}(T_{j+1}, T_{j}) = e^{\alpha + \beta \mu_{x_{j}}},$$

$$\alpha = \frac{\sigma^{2}}{2a^{2}} - \frac{\sigma^{2}}{a^{3}}e^{a} + \frac{\sigma^{2}}{4a^{3}}e^{2a} + \frac{3\sigma^{2}}{4a^{3}},$$

$$\beta = \frac{1}{a}(1 - e^{a}),$$
(26)

However, in our study we focus on mortality intensities. We obtain the parameters using the estimation dataset as described above, after which we use them to generate paths and to forecast mortality intensities. Finally, we calculate the error between the forecasted mortality curve and the actual mortality rates.

The residuals of the model are the realisations of the random component dW(t) which should follow the standard normal distribution if the parameters are estimated correctly:

$$\frac{\Delta\mu-a\mu}{\sigma}\sim N(0,1).$$

We use Kolmogorov-Smirnov statistic to test hypothesis that the errors come from a standard normal distribution.

Taking the mortality intensities for the 7 generations we obtain the parameters presented

In Table 3 we report the obtained parameters for selected 7 generations. As expected, the *a* parameter is increasing with age, which means that the average mortality intensity is larger for older generations. The σ parameter is also growing with age. This proves the fact that there is more uncertainty in mortality rates for older ages.

Table 3. ML parameters of the OU-process and maximised log-likelihood.



Figure 5. Results of the OU-process for UK female generation aged 51 in the year 1970. (**a**) Fitted residuals of the model (1971–2009); (**b**) Historical mortality and simulated paths (1971–2009).

Figure 5a represents the residuals of the model. According to the Kolmogorov-Smirnov test, the errors of the model are confirmed to be standard normal at 5% significance level. Figure 5b illustrates historical mortality intensities (in blue), 1000 MC simulations (in yellow), average among simulations (in red) and confidence intervals (in green) for the entire period (1970–2009). This graph is done for the generation aged 51 in the year 1970.

6.4. Calibration of the Feller Process

We have maximised the log-likelihood function as it is stated in Equation (24) assuming that the mean-reversion parameter is zero. Table 4 reports the obtained parameters and the value of the maximazed log-likelihood function for each generation. The obtained parameter values correspond well with the previous work, such as in Luciano and Vigna [11].

Generation Age in 1970	а	σ	MaxLogLikelihood	BIC
60	0.0984	0.0073	147.1586	-287.5148
57	0.0955	0.0061	156.6162	-306.4299
54	0.0922	0.0043	171.9529	-337.1034
51	0.0869	0.0046	172.7420	-338.6817
48	0.0833	0.0036	170.9023	-335.1403
45	0.0826	0.0034	155.2175	-303.9972
42	0.0801	0.0031	139.2860	-272.3900

Table 4. ML parameters of the Feller process and maximised log-likelihood.

The simulation of the future mortality is performed by discretising Equation (17) with time step equal to one year:

$$\mu_{t+1} = \mu_t + a\mu_t + \sigma\sqrt{\mu_t}N_{(0,1)}.$$
(27)

7. Comparison of the Four Models

To compare the performance of the models for the 19 generations based on their age in 1970. For each, we forecast mortality rates in the period 2000–2009 and compute the relative error of prediction, as well as the percentage of the observed mortality rates in the test period which fall within the prediction interval. The forecast and the prediction bounds are obtained using 15,000 Monte Carlo simulations.

In this section we define the tests of the mean relative error and the prediction intervals. A model can perform very well with respect to the percentage of the mortality rates which fall within the prediction bounds, while at the same time having a high relative error, if its variance grows rapidly and, therefore, the model produces wide prediction bounds. We say that a model is *precise* if its forecasts of mortality are consistent with respect to the prediction interval and that a model is *accurate* if its mean absolute forecast errors are small. Of course, it is desirable for a good model to be both accurate and precise. To interpret the results of the experiments, it is important to understand the form of the variance implied by each model which we discuss in this section as well.

7.1. Relative Error

For each *x*, *t* the *relative error* is defined as follows:

$$error_{x}(t) = \frac{\mu_{x}^{predicted}(t) - \mu_{x}^{actual}(t)}{\mu_{x}^{actual}(t)} \quad \forall x, t.$$

Since the longevity risk corresponds to lower-than-expected mortality rates, we define the error so that it is positive if the forecast of mortality exceeds the historical values (actual values are lower-than-expected), and negative in the opposite case.

We compute the relative error for 19 generations—they are female aged 42–60 in the base year 1970. Thus, in the test period, for which the graphs of error are plotted, they are 30 years older—72–90 years, respectively.

The results of the experiments are presented in Figures 6 and 7 and Tables 5 and 6. The graphs of the mean absolute errors in Figure 7 illustrate the results shown in Tables 5 and 6. We can see that most of the errors fall in the range [-0.1; 0.1]. The exception is the OU process for the generation aged 60 in 1970, especially for later years of projections. The error for this generation in the Feller process forecast is also large – its absolute mean for generation aged 60 in 1970 is 0.0427 (Table 5). This increases the mean absolute errors for this generation for these processes shown in Table 5. Figure 7b shows the relative absolute error for each year in the test period average over 19 generations. We see that the error is smaller for younger ages . All the models show a high error for the generation aged 50 in the year 1970. This might be due to the cohort effect which is generally present in the UK data. The biggest error for this generation is produced by the Lee-Carter model. The graph of the relative absolute errors averaged over generations by year (Figure 7a) shows an increasing trend for all 4 models, especially for the Lee and Carter model (red line). This effect is due to the fact that the variances of the projected mortality rates increase with projection time. However, this does not happen at the same rate in different models.



Figure 6. Relative error of each model for every generation (**a**)–(**d**). (**a**) Lee-Carte model; (**b**) OU-process; (**c**) Wills-Sherris model; (**d**) Feller process.

The errors of the Lee-Carter model are mostly positive (Figure 6a) and we can observe a relative increase of the errors in time for each generation, indicating that the Lee-Carter model has tended to predict mortality rates that are too high. The errors of the Wills and Sherris model exhibit two patterns for different generations (Figure 6c). They are negative for the older generations (aged 60, 57, 54, 51 and 48 in 1970) and positive for the two youngest ones (45, 42 in 1970). We note that the

older generations belong to the lower diagonal of the initial mortality matrix, while the younger two belong – to the upper diagonal of the matrix. Thus, it may be that the Wills and Sherris model has a tendency to overestimate the mortality for younger generations and underestimates the mortality for the older ones.



Figure 7. Mean relative absolute error of each model, UK female data. (a) Relative absolute error, average by year; (b) Relative absolute error, average by generation.

Age in 1970	Wills and Sherris	Lee-Carter	OU-Process	Feller
60	0.0383	0.0387	0.0837	0.0428
59	0.0346	0.0245	0.0543	0.0251
58	0.0396	0.0232	0.0562	0.0248
57	0.0457	0.0256	0.0430	0.0269
56	0.0374	0.0313	0.0573	0.0218
55	0.0420	0.0611	0.0209	0.0313
54	0.0515	0.0599	0.0220	0.0376
53	0.0379	0.0965	0.0160	0.0383
52	0.0362	0.0696	0.0260	0.0320
51	0.0320	0.0337	0.0169	0.0404
50	0.1008	0.1630	0.0919	0.1164
49	0.0388	0.0249	0.0256	0.0405
48	0.0353	0.0441	0.0383	0.0588
47	0.0258	0.0406	0.0257	0.0393
46	0.0212	0.0258	0.0251	0.0409
45	0.0279	0.0296	0.0263	0.0298
44	0.0267	0.0299	0.0193	0.0228
43	0.0312	0.0509	0.0221	0.0285
42	0.0334	0.0611	0.0232	0.0220
Mean (Rank)	0.0388 (3)	0.0492 (4)	0.0365 (1)	0.0379 (2)

Table 5. Mean (over 10 years) of the absolute errors for each generation, UK female data.

According to the Tables 5 and 6, the OU process exhibits the lowest mean absolute error, followed by the Feller process, the Wills and Sherris model and the Lee and Carter model for the UK female data.

Year (Rank)	Wills and Sherris	Lee-Carter	OU-Process	Feller
2000	0.0314	0.0380	0.0379	0.0318
2001	0.0248	0.0417	0.0367	0.0247
2002	0.0267	0.0320	0.0191	0.0187
2003	0.0493	0.0316	0.0241	0.0462
2004	0.0332	0.0473	0.0321	0.0204
2005	0.0409	0.0401	0.0254	0.0311
2006	0.0326	0.0642	0.0405	0.0281
2007	0.0457	0.0537	0.0366	0.0475
2008	0.0626	0.0479	0.0483	0.0690
2009	0.0403	0.0951	0.0645	0.0614
Mean (Rank)	0.0388 (3)	0.0492 (4)	0.0365 (1)	0.0379 (2)

Table 6. Mean (over 19 generations) of the absolute errors for each year, UK female data.

7.2. Discussion on the Variances

As it has been stated in the description of the model, the variance of the mortality intensity $\mu(t)$, conditional on time 0, in the OU specification has a form $\sigma^2 \frac{e^{2at}-1}{2a}$, where *t* is the time elapsed. For the Feller process, when intensity $\mu(t)$ is specified by the CIR process of the form:

$$d\mu(t) = (b + a\mu(t))dt + \sigma\sqrt{\mu(t)}dW(t),$$

with a > 0, b > 0, $\sigma > 0$, the conditional distribution of the mortality intensity at time *t*, conditional on time 0 is given by a non-central chi-square distribution:

$$\mu(t) \sim \frac{\sigma^2(e^{at}-1)}{4a} \chi_d^2(\nu),$$

where $\chi_d^2(\nu)$ denotes the density of a non-central chi-square random variable with *d* degrees of freedom:

$$d=\frac{4b}{\sigma^2},$$

and the non-centrality parameter ν is

$$\nu = \frac{4ae^{at}}{\sigma^2(e^{at}-1)}\mu(0).$$

The $\chi_d^2(\nu)$ distribution, has a variance $Var_{\chi_d^2(\nu)} = 2(d+2\nu)$. Thus, intensity $\mu(t)$ has a variance:

$$\frac{\sigma^2(e^{at}-1)}{2a}\left(\frac{4b}{\sigma^2}+\frac{8ae^{at}}{\sigma^2(e^{at}-1)}\mu(0)\right)$$

In the Feller specification, parameter *b* is not defined and, hence, the number of degrees of freedom *d* is not defined either. However, we can see that, other parameters being equal, the variance of the OU process should grow faster in time than the variance of the CIR process, as it has e^{2at} term rather than e^{at} .

The Wills and Sherris model assumes that the distribution of the changes in the force of mortality follows a normal distribution:

$$\Delta \mu(x,t) \sim N((a(x+t)+b)\mu,\sigma\mu).$$

Thus, the variance of the mortality intensity grows in time as at each time installment it is multiplied by the mortality rate from the year before.

For the Lee Carter model, the variance of the logarithm of the mortality rate at time *t* for each age *x*, is $\beta_x \hat{\sigma}_{rw}^2 \sqrt{t}$, where $\hat{\sigma}_{rw}$ is the variance of the random walk process κ_t , in our case equal to 0.9154, and *t* is the time passed.

To understand better the variance of the distribution of mortality rates at each point in time we show the graphs of the standard deviation in Figure 8. The variance of each model grows with time, however, for the Wills and Sharris model it grows substantially faster than for other three models. The second fastest growing variance is of the Feller process, then the OU-process and, finally, the Lee-Carter model shows the slowest growth in variance.



Figure 8. Standard deviation of each model for every generation (**a**–**d**). (**a**) Lee-Carte model; (**b**) OU-process; (**c**) Wills-Sherris model; (**d**) Feller process.

Regarding the covariance/correlation across generations and ages, all models employ a different structure. Lee-Carter is a one-factor model, which results in mortality improvements at all ages being perfectly correlated. Wills and Sherris model is designed to capture correlation between the ages. In practice it amplifies the effect of the variance growth over time since in reality the correlation increases with age (see Wills and Sherris [13]). In fact, in the simulation procedure the Wiener process is multiplied by the instantaneous correlation matrix D which describes the correlation structure between ages. The OU and the Feller processes in the current study do not take into account the correlation between generations. However, they can also be extended to the case of multiple generations. In [27] it is described how the OU process can be extended to the case of two generations, whose changes in mortality intensities are correlated with an instantaneous correlation coefficient.

7.3. Discussion on the Number of Parameters

The number of estimated parameters is different for each procedure. The Wills and Sherris model estimates only 3 parameters for the whole dataset plus eigenvectors to approximate the correlation matrix (in our case we take 30 eigenvectors), while both the OU and the Feller processes fit 2 parameters for each generation. To calibrate the Lee-Carter model, we have to estimate A + T = 50 + 30 = 80 parameters. To predict mortality for each generation in the dataset for which the size of the estimation part would be not be less that 10 years, we have to estimate the OU and the Feller processes for 19 generations resulting in 38 parameters each, for the Wills and Sherris model we have used 3 + 30 = 33 parameters.

7.4. Prediction Intervals

The results of the experiments based on prediction intervals are presented in the Figure 9 and Table 7. Figure 9 illustrates the 90% prediction intervals built for each model with 15000 MC simulations for generation aged 57 in 1970. We can see that the width of the intervals is different for each model—it is the smallest for the OU-process; medium for the Lee-Carter model and the Feller process, and it is the widest for the Wills and Sherris model, especially for the older ages.



Figure 9. Actual mortality rate and 90% prediction intervals for generation aged 57 for each model. (a) Lee-Carte model; (b) OU-process; (c) Wills-Sherris model; (d) Feller process.

According to Table 7, the Wills and Sherris model shows the best results in terms of the percentage of the actual future mortality which appear within the confidence bounds—for 15 out of 19 generations the percentage reaches 100%. The Lee-Carter model and the Feller process results are comparably

good, while the OU process shows the worst result with the average percentage of the actual mortality within the prediction intervals being 53% (Table 7).

Note, that although the variances of the Lee Carter model is smaller than the ones produced by the OU-process and the Feller process, the first model shows better results.

Table 7. Percentage of the actual mortality rates which falls within a 90% prediction interval, UK female data.

Age in 1970	Wills and Sherris	Lee-Carter	OU-Process	Feller
60	1.0000	0.5000	0.2000	0.7000
59	1.0000	0.9000	0.3000	0.8000
58	1.0000	0.9000	0.3000	0.8000
57	1.0000	0.9000	0.5000	0.8000
56	1.0000	0.8000	0.1000	0.9000
55	1.0000	0.3000	0.6000	0.8000
54	1.0000	0.7000	0.8000	0.7000
53	1.0000	0.2000	0.8000	0.7000
52	1.0000	0.6000	0.5000	0.7000
51	1.0000	0.9000	0.8000	0.7000
50	0.5000	0	0	0
49	0.9000	1.0000	0.7000	0.5000
48	1.0000	0.7000	0.4000	0.5000
47	1.0000	0.9000	0.8000	0.8000
46	1.0000	1.0000	0.4000	0.5000
45	1.0000	1.0000	0.6000	0.9000
44	1.0000	1.0000	0.7000	1.0000
43	0.9000	0.8000	0.7000	0.8000
42	0.9000	0.7000	0.9000	0.9000
Mean (Rank)	0.9579 (1)	0.7263 (2)	0.5316 (4)	0.7105 (3)

8. Robustness of Simulation Results

Here we evaluate the performance of the approach described above on the 4 datasets. They are:

- 1. UK Females
- 2. UK Males
- 3. Australian Females
- 4. Australian Males

The experiments in this section are made using the same time and the age periods—1970–2009 and 50–99. The generations aged 42–60 in the year 1970 are chosen in the same manner as in Sections 6 and 7.

The results of the estimation are presented in Tables 8 and 9. More detailed results of the estimation for the 4 datasets are included in the Appendix as tables and plots of the errors for each of the 19 generations. According to the Tables 8 and 9, the results for the UK males data with regard to accuracy are the same as for the UK females—the OU and the Feller processes produce the smallest error, while the Lee-Carter and the Wills and Sherris models show the largest error. However, the mean of the absolute error for the Lee-Carter model in this case is 3 times larger in comparison to the error of the UK females estimation. This model also shows very bad result according to precision (with the percentage of the actual mortality within prediction interval being only 4.21%).

Model; Dataset	UK, Females	UK, Males	Australia, Females	Australia, Males
Wills and Sherris	0.0388 (3)	0.0580 (3)	0.1368 (4)	0.0787 (3)
Lee-Carter	0.0492 (4)	0.1252 (4)	0.0380(1)	0.0712 (1)
OU-process	0.0365 (1)	0.0470 (2)	0.1046 (2)	0.0767 (2)
Feller	0.0379 (2)	0.0391 (1)	0.1326 (3)	0.0840 (4)

Table 8. Mean of the absolute errors for each dataset over 19 generations (rank of accuracy).

Table 9. Percentage within a 90% prediction interval for each dataset (rank of precision).

Model; Dataset	UK, Females	UK, Males	Australia, Females	Australia, Males
Wills and Sherris	0.9579 (1)	0.9105 (1)	0.7789 (2)	1.0000 (1)
Lee-Carter	0.7263 (2)	0.0421 (4)	0.8368 (1)	0.3053 (4)
OU-process	0.7105 (3)	0.4000 (3)	0.2947 (4)	0.4421 (3)
Feller	0.5316 (4)	0.7158 (2)	0.3474 (3)	0.5737 (2)

Australian females data is the only dataset which shows good results using the Lee Carter model, both according to precision and accuracy. The OU and the Feller processes, on the contrary, produce large errors for this dataset, especially for the generations aged 45–55. This may be explained by the fact that mortality in Australia is lower for people in their 40s and 50s in comparison to their UK counterparts, and, as a consequence, mortality intensities are larger for older ages. This can be seen from the plots of mortality curves for generations aged 51 and 54 in the year 1970 (Figure 10). More prominent convex form of the mortality curves for Australian population makes the error (which is calculated for the last 10 years of the observations) larger as the prediction of mortality underestimates the actual mortality intensity. We would suggest that the inclusion of the correlation coefficient for the COU and the Feller processes to describe the dependence between the generations could improve the calibration results for these procedures by taking into account the fact that if mortality of the generations aged 45–55 is rather low, it would imply an increase in the mortality intensity for the older ages.



Figure 10. Observed mortality curves for the UK and Australian generations aged 51 (**a**) and 54 (**b**). (**a**) Generation aged 51 in the year 1970; (**b**) Generation aged 54 in the year 1970.

It is worth noting that for all datasets the errors produced by the Wills and Sherris model, the OU process and the Feller process exhibit similar patterns (Figures 7b, 11b, 12b and 13b), while the errors produced by the Lee-Carter model have a different pattern. This may be explained by the fact that the first three procedures model the advances (changes) in mortality intensity for a cohort, while the Lee-Carter models the central mortality rate itself.



Figure 11. Mean relative absolute error of each model, UK male data. (a) Relative absolute error, average by year; (b) Relative absolute error, average by generation.



Figure 12. Mean relative absolute error of each model, Australian female data. (**a**) Relative absolute error, average by year; (**b**) Relative absolute error, average by generation.



Figure 13. Mean relative absolute error of each model, Australian male data. (**a**) Relative absolute error, average by year; (**b**) Relative absolute error, average by generation.

On the whole, we can say that the results are data dependent. However, from the estimation results on the four datasets, we can conclude that the Wills and Sherris model performs best in terms of precision, but it is one of the worst in terms of accuracy. The Lee Carter model shows better fit to the Australian population dataset rather that to the British one, both for males and females. The OU process and the Feller process provide rather good results in terms of accuracy, while they have often get low ranks in terms of precision.

9. Conclusions

In this study we have calibrated 4 mortality models to the UK and Australian populations and have quantitatively compared their accuracy and precision in forecasting mortality rates. To evaluate this we have used two measures – first, we looked at the relative errors between the forecasted and the observed mortality rates and second, we investigated the percentage of the observed mortality rates which fell within the projected prediction intervals. Our experiments compare one discrete-time model, proposed by Lee and Carter, and three continuous-time models—the Wills and Sherris model, the Ornstein-Uhlenbeck process and the Feller process. The first two models estimate the whole surface of mortality across ages and years simultaneously, while the latter two model each generation separately. One major advantage of the OU and the Feller processes is that they belong to the affine class of mortality models and so allow closed-form expressions for survival probabilities, which is useful for pricing many financial securities. On the other hand, the Wills and Sherris model allows the dependencies between generations to be captured, which may be useful for life offices who have portfolios written on multiple cohorts.

The choice of the model may depend on the goal and the data available. As a result of our experiments with the UK female, the Wills and Sherris model performs best in terms of the prediction interval, followed by the Lee-Carter model. In terms of the mean absolute error, the OU and the Feller processes are better. Thus, for the UK data models which capture the whole mortality surface are more precise, meaning that their forecast prediction intervals are more likely to include the observed mortality rates. Models for a single generation, on the other hand, tend to be more accurate, meaning that their mean absolute errors between the forecast and observed mortality are smaller. For the UK male data the results are rather similar—the main difference here is that the LC model in this case provides much worst result both in terms of precision and accuracy.

However, the results are different for the Australian dataset. In this case, the Lee-Carter model and the OU process are the best in terms of accuracy, both for males and females. The Wills and Sherris model shows good result with respect to the precision measure for Australia as well, followed by the LC for the females and the Feller process for the males.

Based on our experiments, different models appear to be preferred for specific generations and years. We believe that our analysis and the results discussed in this paper are useful for the insurance industry. In particular, we provide potentially useful insights into different mortality modelling frameworks and allow practitioners to chose a model that suits their specific needs.

Acknowledgments: The author thanks Luca Regis, assistant professor at the IMT Luca for assistance with the experiments setup. I am also grateful to the participants of the ASF 2016 and ICASQF 2016 conferences, especially to Anrew Hunt, research actuary at Pacific Life Re and Lewis Ramsden, postgraduate research student at the University of Liverpool for their comments on the earlier version of the manuscript.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

Age in 1970	Wills and Sherris	Lee-Carter	OU-Process	Feller
60	0.0494	0.0612	0.0758	0.0560
59	0.0445	0.0634	0.0737	0.0506
58	0.0227	0.0739	0.0320	0.0222
57	0.0288	0.0941	0.0342	0.0257
56	0.0360	0.1148	0.0425	0.0315
55	0.0602	0.1601	0.0670	0.0506
54	0.0511	0.1746	0.0549	0.0423
53	0.0583	0.2275	0.0399	0.0340
52	0.0527	0.2031	0.0368	0.0344
51	0.0536	0.1455	0.0143	0.0188
50	0.0356	0.0676	0.0249	0.0228
49	0.0408	0.1205	0.0204	0.0178
48	0.0579	0.1350	0.0347	0.0306
47	0.0585	0.1369	0.0280	0.0260
46	0.0738	0.1086	0.0456	0.0410
45	0.0680	0.1169	0.0274	0.0278
44	0.0924	0.1065	0.0659	0.0598
43	0.0976	0.1291	0.0735	0.0614
42	0.1196	0.1404	0.1007	0.0904
Mean (Rank)	0.0580 (3)	0.1252 (4)	0.0470 (2)	0.0391 (1)

Table A1. Mean (over 10 years) of the absolute errors for each generation, UK male data.

Table A2. Mean (over 19 generations) of the absolute errors for each year, UK male data.

Year (Rank)	Wills and Sherris	Lee-Carter	OU-Process	Feller
2000	0.0318	0.0643	0.0283	0.0268
2001	0.0418	0.0790	0.0342	0.0312
2002	0.0354	0.0793	0.0268	0.0226
2003	0.0378	0.0735	0.0267	0.0271
2004	0.0621	0.1237	0.0485	0.0413
2005	0.0576	0.1248	0.0449	0.0356
2006	0.0759	0.1575	0.0611	0.0503
2007	0.0713	0.1646	0.0566	0.0449
2008	0.0632	0.1632	0.0494	0.0349
2009	0.1027	0.2224	0.0932	0.0768
Mean (Rank)	0.0580 (4)	0.1252 (1)	0.0470 (3)	0.0391 (2)

Age in 1970	Wills and Sherris	Lee-Carter	OU-Process	Feller
60	1.0000	0.1000	0.1000	0.5000
59	1.0000	0.1000	0.1000	0.6000
58	1.0000	0	0.6000	0.9000
57	1.0000	0	0.7000	1.0000
56	1.0000	0	0.3000	0.7000
55	1.0000	0	0	0.5000
54	1.0000	0	0.4000	0.7000
53	1.0000	0	0.6000	0.9000
52	1.0000	0	0.4000	0.7000
51	1.0000	0.1000	1.0000	1.0000
50	1.0000	0.4000	0.8000	0.9000
49	1.0000	0	0.7000	0.9000
48	1.0000	0	0.5000	1.0000
47	1.0000	0	0.5000	0.6000
46	0.9000	0	0.2000	0.6000
45	1.0000	0	0.6000	1.0000
44	0.8000	0.1000	0.1000	0.6000
43	0.5000	0	0	0.5000
42	0.1000	0	0	0
Mean (Rank)	0.9105 (1)	0.0421 (4)	0.4000 (3)	0.7158 (2)

Table A3. Percentage of the actual mortality rates which falls within a 90% prediction interval, UK male data.

 Table A4. Mean (over 10 years) of the absolute errors for each generation, Australian female data.

Age in 1970	Wills and Sherris	Lee-Carter	OU-Process	Feller
60	0.1215	0.0697	0.0746	0.0983
59	0.1113	0.0386	0.0461	0.0834
58	0.1191	0.0573	0.0347	0.0777
57	0.1201	0.0368	0.0372	0.0869
56	0.1405	0.0591	0.0561	0.1049
55	0.1270	0.0198	0.0570	0.1018
54	0.1637	0.0215	0.1123	0.1490
53	0.1533	0.0229	0.1158	0.1469
52	0.1713	0.0269	0.1402	0.1709
51	0.1736	0.0253	0.1633	0.1791
50	0.1955	0.1026	0.1802	0.2019
49	0.1438	0.0255	0.1269	0.1533
48	0.1434	0.0300	0.1030	0.1390
47	0.1622	0.0344	0.1735	0.1814
46	0.1551	0.0311	0.1665	0.1839
45	0.1227	0.0151	0.1331	0.1453
44	0.0986	0.0356	0.1044	0.1206
43	0.0905	0.0331	0.0833	0.1017
42	0.0850	0.0373	0.0787	0.0944
Mean (Rank)	0.1368 (3)	0.0380 (4)	0.1046 (1)	0.1326 (2)

Year (Rank)	Wills and Sherris	Lee-Carter	OU-Process	Feller
2000	0.0322	0.0400	0.0303	0.0308
2001	0.0436	0.0253	0.0354	0.0391
2002	0.0978	0.0394	0.0728	0.0887
2003	0.1041	0.0309	0.0710	0.0941
2004	0.1207	0.0316	0.0825	0.1110
2005	0.1284	0.0322	0.0877	0.1205
2006	0.1664	0.0358	0.1231	0.1612
2007	0.2025	0.0412	0.1588	0.2007
2008	0.2421	0.0606	0.1992	0.2441
2009	0.2296	0.0434	0.1849	0.2361
Mean (Rank)	0.1368 (4)	0.0380(1)	0.1046 (2)	0.1326 (3)

Table A5. Mean (over 19 generations) of the absolute errors for each year, Australian female data.

Table A6. Percentage of the actual mortality rates which falls within a 90% prediction interval, Australian female data.

Age in 1970	Wills and Sherris	Lee-Carter	OU-Process	Feller
60	1.0000	0.4000	0.3000	0.4000
59	1.0000	0.7000	0.4000	0.4000
58	1.0000	0.4000	0.6000	0.4000
57	1.0000	0.7000	0.6000	0.5000
56	1.0000	0.6000	0.2000	0.2000
55	1.0000	1.0000	0.4000	0.3000
54	0.7000	1.0000	0.1000	0.2000
53	0.7000	0.9000	0.2000	0.2000
52	0.5000	1.0000	0.1000	0.1000
51	0.5000	1.0000	0.1000	0.2000
50	0.3000	0.4000	0	0.1000
49	0.7000	0.9000	0.1000	0.2000
48	0.6000	1.0000	0.2000	0.5000
47	0.6000	0.9000	0.2000	0.2000
46	0.6000	1.0000	0.1000	0.2000
45	0.8000	1.0000	0.2000	0.4000
44	0.8000	1.0000	0.4000	0.5000
43	1.0000	1.0000	0.7000	0.7000
42	1.0000	1.0000	0.7000	0.9000
Mean (Rank)	0.7789 (2)	0.8368 (1)	0.2947 (4)	0.3474 (3)

 Table A7. Mean (over 10 years) of the absolute errors for each generation, Australian male data.

Age in 1970	Wills and Sherris	Lee-Carter	OU-Process	Feller
60	0.0940	0.0930	0.0801	0.0945
59	0.0439	0.0534	0.0386	0.0346
58	0.1171	0.0553	0.0980	0.1133
57	0.0879	0.0376	0.0586	0.0783
56	0.0974	0.0231	0.0657	0.0867
55	0.1238	0.0522	0.1385	0.1357
54	0.0949	0.0622	0.0805	0.0958
53	0.1144	0.0737	0.1163	0.1257
52	0.0928	0.0860	0.0817	0.0955
51	0.0900	0.0806	0.0880	0.0968
50	0.1338	0.0179	0.1543	0.1531
49	0.0725	0.1036	0.0916	0.0906
48	0.0693	0.0807	0.0541	0.0658
47	0.0581	0.0967	0.0578	0.0663
46	0.0499	0.0874	0.0562	0.0617
45	0.0460	0.0784	0.0458	0.0500
44	0.0387	0.0860	0.0470	0.0486
43	0.0356	0.1111	0.0628	0.0591
42	0.0357	0.0742	0.0422	0.0432
Mean (Rank)	0.0787 (3)	0.0712 (1)	0.0767 (2)	0.0840 (4)

Year (Rank)	Wills and Sherris	Lee-Carter	OU-Process	Feller
2000	0.0279	0.0365	0.0287	0.0285
2001	0.0362	0.0516	0.0370	0.0367
2002	0.0623	0.0404	0.0590	0.0615
2003	0.0536	0.0609	0.0521	0.0525
2004	0.0615	0.0597	0.0580	0.0640
2005	0.0561	0.0970	0.0478	0.0578
2006	0.0858	0.0913	0.0825	0.0926
2007	0.1132	0.0866	0.1076	0.1239
2008	0.1526	0.0814	0.1497	0.1652
2009	0.1380	0.1067	0.1448	0.1571
Mean (Rank)	0.0787 (3)	0.0712 (1)	0.0767 (2)	0.0840 (4)

Table A8. Mean (over 19 generations) of the absolute errors for each year, Australian male data.

Table A9. Percentage of the actual mortality rates which falls within a 90% prediction interval, Australian male data.

Age in 1970	Wills and Sherris	Lee-Carter	OU-process	Feller
60	1.0000	0.2000	0.4000	0.6000
59	1.0000	0.4000	0.7000	1.0000
58	1.0000	0.2000	0.2000	0.3000
57	1.0000	0.3000	0.4000	0.8000
56	1.0000	0.6000	0.4000	0.5000
55	1.0000	0.1000	0	0
54	1.0000	0.2000	0.4000	0.5000
53	1.0000	0.1000	0.2000	0.3000
52	1.0000	0	0.3000	0.4000
51	1.0000	0.2000	0.3000	0.3000
50	1.0000	1.0000	0	0.2000
49	1.0000	0	0.4000	0.7000
48	1.0000	0.2000	0.6000	0.8000
47	1.0000	0.1000	0.4000	0.6000
46	1.0000	0.5000	0.7000	0.7000
45	1.0000	0.6000	0.7000	0.8000
44	1.0000	0.4000	0.7000	0.8000
43	1.0000	0.1000	0.7000	0.7000
42	1.0000	0.6000	0.9000	0.9000
Mean (Rank)	1.0000 (1)	0.3053 (4)	0.4421 (3)	0.5737 (2)

References

- 1. Cairns, A.J.G.; Blake, D.; Dowd, K. Pricing death: Framework for the valuation and securitization of mortality risk. *ASTIN Bull.* **2005**, *36*, 79–120.
- 2. Lee R.D.; Carter, L.R. Modelling and forecasting U.S. mortality. J. Am. Stat. Assoc. 1992, 87, 659–671.
- 3. Renshaw, A.E.; Haberman, S. A cohort-based extension to the Lee- Carter model for mortality reduction factors. *Insur. Math. Econ.* **2006**, *38*, 556–570.
- 4. Cairns, A.J.G.; Blake, D.; Dowd, K. A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *J. Risk Insur.* **2006**, *73*, 687–718.
- 5. Plat, R. On stochastic mortality modelling. *Insur. Math. Econ.* 2009, 45, 393–404.
- 6. O'Hare, C.; Li, Y. Explaining young mortality. Insur. Math. Econ. 2012, 50, 12–25.
- 7. Milevsky, M.A.; Promislow, D. Mortality Derivatives and the Option to Annuitize. *Insur. Math. Econ.* **2001**, 29, 299–318.
- 8. Dahl, M. Stochastic mortality in life insurance: Market reserves and mortality-linked insurance contracts. *Insur. Math. Econ.* **2004**, *35*, 113–136.
- 9. Biffis, E. Affine processes for dynamic mortality and actuarial valuations. *Insur. Math. Econ.* 2005, 37, 443–468.
- 10. Denuit, M.; Devolder, P. *Continuous Time Stochastic Mortality and Securitization of Longevity Risk*; Louvain-la-Neuve Working Paper 06-02; Institut des Sciences Actuarielles, Universite' Catholique de Louvain: Louvain-la-Neuve, Belgium, 2006.

- 11. Luciano, E.; Vigna, E. Mortality risk via affine stochastic intensities: Calibration and empirical relevance. *Belgian Actuar. Bull.* **2008**, *8*, 5–16.
- 12. Schrager, D. Affine stochastic mortality. Insur. Math. Econ. 2006, 38, 81–97.
- Wills, S.; Sherris, M. Integrating Financial and Demographic Longevity Risk Models: An Australian Model for Financial Applications; UNSW Australian School of Business Research Paper No. 2008ACTL05; UNSW Australian School: Kensington NSW, Australia, 2011.
- 14. Wills, S.; Sherris, M. Securitization, Structuring and Pricing of Longevity Risk. *Insur. Math. Econ.* **2010**, *46*, 173–185.
- 15. Cairns, A.J.G.; Blake, D.; Dowd, K.; Coughlan, G.D.; Epstein, D.; Ong, A.; Balevich, I. A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *N. Am. Actuar. J.* **2009**, *13*, 1–35.
- 16. Dowd, K.; Cairns, A.J.G.; Blake, D.; Coughlan, G.D.; Epstein, D.; Khalaf-Allah, M. Evaluating the goodness of fit of stochastic mortality models. *Insur. Math. Econ.* **2010**, *47*, 255–265.
- Dowd, K.; Cairns, A.J.G.; Blake, D.; Coughlan, G.D.; Epstein, D.; Khalaf-Allah, M. Backtesting stochastic mortality models: An ex-post evaluation of multi-period-ahead density forecasts. *N. Am. Actuar. J.* 2010, 14, 281–298.
- Giacometti, R.; Bertocchi, M.; Rachev, S.T.; Fabozzi, F.J. A comparison of the Lee-Carter model and AR-ARCH model for forecasting mortality rates. *Insur. Math. Econ.* 2012, *50*, 85–93.
- 19. Lee, R.D.; Miller, T. Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography* **2001**, *38*, 537–549.
- 20. Ngai, N.; Sherris, M. Longevity risk management for life and variable annuities: The effectiveness of static hedging using longevity bonds and derivatives. *Insur. Math. Econ.* **2011**, 49, 100–114.
- 21. Blake, D.; Burrows, W. Survivor bonds: Helping to hedge mortality risk. J. Risk Insur. 2001, 68, 339–348.
- 22. Luciano, E.; Regis, L.; Vigna, E. Delta-Gamma hedging of mortality and interest rate risk. *Insur. Math. Econ.* **2012**, *50*, 402–412.
- 23. Girosi, F.; King, G. Understanding the Lee-Carter Mortality Forecasting Method. Available online: http://gking.harvard.edu/files/abs/lc-abs.shtml (accessed on 29 November 2016).
- 24. Duffie, D.; Filipovic, D.; Schachermayer, W. Affine processes and applications in finance. *Ann. Appl. Probab.* **2003**, *13*, 984–1053.
- 25. Cox, J.C.; Ingersoll, J.E.; Ross, S.A. A Theory of the Term Structure of Interest Rates. *Econometrica* **1985**, *53*, 385–407.
- 26. Kladivko, K. Maximum Likelihood Estimation of the Cox-Ingersoll-Ross Process: The Matlab Implementation. In Proceedings of the Technical Computing Conference, Prague, Czech Republic, 19–24 July 2007.
- 27. Jevtic, P.; Luciano, E.; Vigna, E. Mortality surface by means of continuous time cohort models. *Insur. Math. Econ.* **2013**, *53*, 122–133.



© 2016 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).