

Article

A Square-Root Factor-Based Multi-Population Extension of the Mortality Laws

Petar Jevtić¹ and Luca Regis^{2,3*}

¹ School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ 85287, USA; petar.jevtic@asu.edu

² ESOMAS Department, University of Torino, Corso Unione Sovietica 218/bis, 10134 Torino, Italy

³ Collegio Carlo Alberto, Piazza Arbarello 8, 10128 Torino, Italy

* Correspondence: luca.regis@unito.it; Tel.: +39-011-670-6065

Abstract: In this paper, we present and calibrate a multi-population stochastic mortality model based on latent square-root affine factors of the Cox-Ingersoll and Ross type. The model considers a generalization of the traditional actuarial mortality laws to a stochastic, multi-population and time-varying setting. We calibrate the model to fit the mortality dynamics of UK males and females over the last 50 years. We estimate the optimal states and model parameters using quasi-maximum likelihood techniques.

Keywords: multi-population mortality; mortality surface; continuous-time stochastic mortality; quasi-linear Kalman filter estimation; square-root processes

JEL Classification: C13, C38, G22, J11



Citation: Jevtić, P.; Regis, L. A Square-Root Factor-Based Multi-Population Extension of the Mortality Laws. *Mathematics* **2021**, *9*, 2402. <https://doi.org/10.3390/math9192402>

Academic Editors: Alessandra Cretarola and Katia Colaneri

Received: 23 August 2021

Accepted: 19 September 2021

Published: 27 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Mortality laws are considered traditional actuarial models that describe the link between death probabilities (or the force of mortality) and the age of the individuals. However, it is very soon after continuous-time stochastic mortality models started developing that [1] proposed the use of latent factors to generalize mortality laws to a stochastic continuous-time setting. In the proposed framework, a set of time-varying stochastic components and parameters affected the evolution of mortality rates at different ages. In particular, [1] calibrated the one-year death probabilities of several ages of the Dutch population to 2- or 3-factor Gaussian models that extended the first and second Makeham's and Thiele's laws to a stochastic setting. The authors of [2] built on a similar setup to propose a general multi-population stochastic extension of mortality laws. There, the mortality intensities of several populations were affine functions of affine latent factors. Different populations and sub-populations within them had different mortality dynamics because they depended on some common factors—shared among population groups—and certain idiosyncratic factors, which were population-specific. Population-specific parameters interact with such factors to reproduce the age–mortality relation. While the setting described in [2] was very general, the implementation was restricted to the use of Gaussian factors. The Gaussian factors have the advantage of being easier to calibrate. However, Gaussian factors have the disadvantage of allowing for negative mortality intensities with non-negative probabilities. That is a relevant shortcoming, for instance, for annuity pricing purposes. In their work, [2] described the application of a 4-population 3-factor law, they estimated through the use of a sequential Kalman filter. In this paper, we advance the previous work and focus on the case in which factors are non-Gaussian, which has not been previously undertaken in the actuarial literature. While several other works have used factor models to represent the evolution of mortality intensity of multiple ages simultaneously in continuous time (see [3,4], for instance), only a few, however, have considered non-Gaussian factors [5].

In contrast to that paper, which calibrated a model for a restricted portion of the mortality curve of the two-population case, our work tries to parsimoniously capture the evolution of mortality of the entire curve or a large fraction of the ages. Further, this work implements a two-population model that lies within the general framework described in [2] and considers factors modeled as square-root processes à la [6]. Finally, we estimate a stochastic multi-population version of the second Makeham's law and a modification of Thiele's law to fit the observed mortality death rates of United Kingdom females and males from 1967 to 2017 by exploiting the state-space representation of our model. To achieve that, we take a centralized data fusion approach, assuming all observations come from the same source while at the same time assuming Gaussian measurement errors. In addition, the non-normality of the state process dynamics, which violates the assumptions under which the standard Kalman filter produces optimal state estimates and an exact log-likelihood for the model, was accounted for. Specifically, we followed [7] and modified the Kalman filter structure to introduce non-linearity and apply thus a Quasi-Maximum Likelihood approach. While producing a possibly conditionally biased estimate, the Kalman filter we implemented remains the optimal linear filter [8].

The paper is structured as follows. Section 2 draws a historical picture of mortality laws and provides some motivation for our work. Section 3 describes the modeling setup. Section 4 specifies the state-space form of our model. Finally, Section 5 presents our application of the stochastic multi-population version of the Second Makeham's and a 3-factor modified version of Thiele's mortality law to the UK male and female population.

2. Mortality Laws: Motivation

At an aggregate level, the most important covariate in explaining individual mortality is age. Indeed, mortality rates have an obvious relation with age. Historically, mortality rates have been high for infants due to the perils of labor and their fragility, increasing at older ages due to biological aging. Due to this persistent phenomenon, actuaries started describing the relation between mortality rates and age through mathematical models, starting from the XVIII century. These mathematical models are of the form:

$$\mu_x = f(x; \xi),$$

where μ_x denotes the mortality intensity of an age x and ξ is a set of parameters. The most famous laws were introduced in the XIX century. In what follows, we give an account of these models. The Gompertz law describes the mortality intensity μ_x at age x as a function of two parameters:

$$\mu_x = Bc^x.$$

It captures the phenomenon of so-called senescent mortality, since it assumes an exponential increment in the mortality intensity with age. Makeham laws, proposed in 1867 and 1890 respectively, have the following forms:

$$\mu_x = A + Bc^x;$$

$$\mu_x = A + Hx + Bc^x,$$

where, together with the exponential term as in the Gompertz law, a baseline mortality level (A) and a linear term (Hx) are added. The Thiele law, dating back to 1871, aimed at capturing the whole mortality-age curve, by considering, together with the exponential term à la Gompertz, a term that tries to capture the spike of the curve at age 0 associated with infant mortality and a term to reproduce the so-called accident hump:

$$\mu_x = Ae^{-Bx} + Ce^{-D(x-E)^2} + FG^x.$$

In the last century, several other laws have been proposed. They combine additional terms and different functional forms. Other prominent examples are Heligman–Pollard's laws and the Perks laws, which identify a functional relation between the death rate q

or the odds ratio $\frac{q}{p}$ (i.e., the ratio between the death rate and the survival probability p), and age. The interested reader may refer to the book [9] for a more detailed account.

The success of traditional mortality laws is due to their ability to capture parsimoniously in a deterministic fashion the relation between the mortality intensity and age. The functional forms used are able to capture the most important trends in such relation, such as the infant mortality pattern and the old-age increase in mortality rates. However, the age-patterns profiles have been changing in time. Thus, the fit of the mortality laws to the observed data of a certain population over different calendar years would produce different parameter estimates. To capture this issue and be able to describe mortality dynamics in a model, while maintaining the traditional actuarial approach of mortality laws, Schrager [1] suggested modeling some parameters in the mortality laws as stochastic factors. Indeed, the introduction of stochastic factors can capture the time trends in human mortality development and the uncertainty surrounding them. The authors of [2] extended that setup to describe the mortality intensity of multiple populations. In that framework, which will be reviewed more properly in the next section, the mortality intensity of individuals of different populations, at different ages and calendar years, is modeled as a function of population-specific parameters and factors. Such an extension allows to jointly describe the mortality intensity of several general populations or sub-groups (socio-economic groups), their evolution in time, and their uncertainty through traditional mortality laws.

3. The Model

In this section, we present a square-root factor-based approach to stochastic mortality modeling, building on the work in [2]. The mortality intensity of different ages of several populations is modeled parsimoniously as a function of some population-specific parameters and a set of latent factors described by a multivariate Cox, Ingersoll, and Ross (1985) process.

3.1. Mortality Intensity and Information

Our continuous-time modeling framework follows the approach first proposed by [10], in which the death process of an individual is described as the first jump time of a Poisson process with stochastic intensity. Mortality intensity is indeed such intensity. This modeling approach parallels the paradigm extensively used in credit risk modeling, in which firm default is similarly represented [11].

Let us introduce a probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, with $\mathbb{P}$ denoting a (historical) probability measure. The filtration $\{\mathcal{F}_t : 0 \leq t \leq T\}$ satisfies the usual properties of right-continuity and completeness and contains the information on the mortality rates, coming from the historical life tables, contained in sub-filtration $\mathcal{G}_t$, as well as the knowledge on whether the individual is alive at time t , contained in the sub-filtration $\mathcal{H}_t$. Hence, $\mathcal{F}_t = \mathcal{G}_t \cup \mathcal{H}_t$.

We assume there exist $J \in \mathbb{N}^+$ populations of interest, indexed by $j \in J = \{1, \dots, J\}$. Our interest is in modeling the evolution of the mortality intensities of sets of ages (cohorts of individuals sharing the same year of birth) belonging to these different populations. We define the set of ages of interest in population j as $X^j := \{x_{\min}^j, \dots, x_{\max}^j\}$, where integer ages are ordered from the smallest, $x_{\min}$, to the biggest, $x_{\max}$. In principle, we can consider different sets of ages for different populations. Let us define as $\mu_{x+t}^j(t)$ the set of predictable processes that represent the time t mortality intensity of an age $x+t$ belonging to the cohort age x at time t .

The mortality intensity of individuals belonging to population j depends on a population-specific set of parameters, γ^j and on a vector of unobservable stochastic factors $\mathbf{Y}^j(t) := [Y_1^j(t), \dots, Y_N^j(t)]^T$, $j \in J$, $N \in \mathbb{N}^{++}$. The value of these factors represents the mortality trend level, which decreases and represents improvements to longevity due to progress in medicine, food availability, health habits, living conditions, and individual well-being.

We consider that all J populations are affected by the same innovations, described by a vector $\mathbf{Z}(t)$ of Brownian noise. However, responses to such changes may differ across populations because factors can be population-specific and because these factors influence population mortality due to the population-specific parameters. Formally, the mortality intensity at time t of an age $x + t$, belonging to population j , $\mu_{x+t}^j(t)$, is described by a functional of the vector of state processes $\mathbf{Y}^j$:

$$\mu_{x+t}^j(t; x) := R^j(\mathbf{Y}^j(t), x, x + t, t, \gamma^j). \quad (1)$$

Notice that the functional R^j can in principle depend on the age $x + t$, the cohort x and calendar time t separately. We specify this functional further by assuming it has an affine dependence on the state-vector process. We define then two sets of population-specific functions $g_0^j(x + t, t, x)$ and

$$\mathbf{g}^j(x + t, t, x) := [g_1^j(x + t, t, x), \dots, g_N^j(x + t, t, x)]',$$

where $g_i^j : (\mathbb{R}^+)^3 \rightarrow \mathbb{R}^+, i \in \{0, 1, \dots, N\}$ are age $x + t$, time t and cohort x -dependent functions for the j -th population, so that

$$R^j(\mathbf{Y}^j(s); x + t, t, x) := g_0^j(x + t, t, x) + \mathbf{g}^j(x + t, t, x) \cdot \mathbf{Y}^j(t),$$

where $\cdot$ denotes the scalar product.

3.2. Stochastic Mortality Laws

We take advantage of the above setup to generalize the most common mortality laws we reviewed in Section 2:

- Gompertz's law:

$$\mu_{x+t}^j(t) := Y_1^j(t)(\gamma^j)^{x+t}.$$

In this case, $\mathbf{g}^j(x + t, t, x) := [(\gamma^j)^{x+t}]$, while $\mathbf{Y}^j \in \mathbb{R}$.

- First Makeham law:

$$\mu_{x+t}^j(t) := Y_1^j(t) + (\gamma^j)^{x+t} Y_2^j(t).$$

In this case, $\mathbf{g}^j(x + t, t, x) := [1, (\gamma^j)^{x+t}]$, while $\mathbf{Y}^j \in \mathbb{R}^2$.

- Second Makeham law:

$$\mu_{x+t}^j(t) := Y_1^j(t) + Y_2^j(t)(x + t) + (\gamma^j)^{x+t} Y_3^j(t).$$

In this case, $\mathbf{g}^j(x + t, t, x) := [1, x + t, (\gamma^j)^{x+t}]$, while $\mathbf{Y}^j \in \mathbb{R}^3$.

- Thiele's law:

$$\mu_{x+t}^j(t) := Y_1^j(t)e^{-\gamma_1^j(x+t)} + Y_2^j(t)e^{-\gamma_2^j(x+t-\gamma_3^j)^2} + Y_3^j(t)(\gamma_4^j)^{x+t}.$$

In this case, $\mathbf{g}^j(x + t, t, x) := [e^{-\gamma_1^j(x+t)}, e^{-\gamma_2^j(x+t-\gamma_3^j)^2}, (\gamma_4^j)^{x+t}]$, while $\mathbf{Y}^j \in \mathbb{R}^3$.

Additionally, we mention a slight modification of Thiele's law, which was proposed by [2] and which proved to fit the evolution of mortality better at younger ages:

$$\mu_{x+t}^j(t) := Y_1^j(t)e^{-\gamma_1^j(x+t)^{\frac{1}{\gamma_2^j}}} + Y_2^j(t)e^{-\gamma_3^j(x+t-\gamma_4^j)^2} + Y_3^j(t)(\gamma_5^j)^{x+t}.$$

We will fit this specification in our application. Notice that $g_0(x + t, t, x) = 0$ for all the above specifications.

3.3. The State Process Vector

We model the J state processes by means of continuous-time stochastic processes, belonging to the affine class. We focus on purely diffusive square-root processes of the Cox–Ingersoll–Ross type:

$$dY^j(t) = A^j(\theta^j - Y^j(t))dt + \Sigma^j \sqrt{V_t^j} dZ(t),$$

where $A^j = \text{diag}[a_1^j, \dots, a_N^j] \in \mathbb{R}_+^{N \times N}$ is a diagonal matrix, $\theta^j \in \mathbb{R}^N$ is a column vector, $\Sigma^j = \text{diag}[\sigma_1^j, \dots, \sigma_N^j] \in \mathbb{R}_+^{N \times N}$ is a diagonal matrix and

$$V_t^j = V_t^j(Y^j(t)) = \text{diag}[Y_1^j(t), \dots, Y_N^j(t)],$$

while $Z(t)$ is a vector of possibly correlated Brownian motions. To summarize, our state process vector captures the time-trend of mortality levels. The different processes impact different slices of the mortality curve, i.e., representing different mortality indexes for different classes of ages. θ^j denotes the vector of long-term values of such factors, while A^j determines the speed at which the current value of the factor returns to this long-term mean. The diffusion coefficient depends on the current value of the process. Overall, these factors, when the elements of A^j and θ^j are positive, are always non-negative. As a consequence, the mortality intensities can be guaranteed to be positive, provided that the elements in vectors γ^j are appropriately bounded. This choice avoids the issue of dealing with negative intensities with non-zero probability, which is instead a feature of all stochastic mortality models that assume Gaussian innovations.

3.4. Survival Probabilities

Because we are in an affine framework, we can compute the expectation of affine functionals of the state processes explicitly (see [12], for instance). In particular, we are able to obtain a closed-form expression for the survival probability ${}_{T-t}p_{x+t}^j(t, T)$, the probability that at calendar time t an age $x + t$ belonging to population j will survive until time $T > t$. The survival probability is indeed a functional of $\mu_{x+t}^j(t; x)$:

$$\begin{aligned} {}_{T-t}p_{x+t}^j(t, T) &:= \mathbb{E}^\mathbb{P}[\mathbb{1}_{\{\tau_x > T\}} \mid \mathcal{H}_t \cup \{\tau_x > t\}] = \mathbb{E}^\mathbb{P}[e^{-\int_t^T \mu_{x+s}^j(s; x) ds} \mid \mathcal{G}_t] = \\ &= \mathbb{E}^\mathbb{P}[e^{-\int_t^T R^j(Y^j(s), x+s, s, x) ds} \mid \mathcal{G}_t]. \end{aligned}$$

We can give the following Theorem (see [2,13]):

Theorem 1. Under the above setup, the time- t survival probability of an individual aged x at $t = 0$, belonging to population group j , up to time T is

$${}_{T-t}p_{x+t}^{j,k}(t) = e^{\alpha^j(x, t, T) + \beta^j(x, t, T) \cdot Y^j(t)}, \quad (2)$$

where $\alpha^j(x, t, T)$ and $\beta^j(x, t, T)$ solve the following system of ordinary differential equations:

$$\begin{aligned} \dot{\alpha}^j &= g_0^j(x + t, t, x) - A^j \theta^j \cdot \beta^j, \\ \dot{\beta}^j &= g^j(x + t, t, x) + (A^j)' \beta^j - \left[\frac{1}{2}(\beta_1^j)^2 \sigma_1^2, \dots, \frac{1}{2}(\beta_N^j)^2 \sigma_N^2 \right]', \end{aligned}$$

where $\alpha^j(x, T, T) = 0$ and $\beta^j(x, T, T) = \mathbf{0}$.

We refer the reader to [2] for a proof of this rather standard result based on the affine-matching principle (see also [14]).

4. The State-Space Form of the Model

To estimate our model, we write it in state-space form. From now on, we consider observations collected at discrete-time (e.g., annual) intervals. Without loss of generality, we assume that standard Brownian motions drive the processes in the state vector process. Given expression (2), it is easy to notice that the logarithms of survival probabilities are linear functions of the state process vector. Therefore, we consider the opposite of this quantity, which is positive, as our measurement. We denote the observations with $t \in \{1, \dots, T\}$, and with $\bar{q}^j(t)$ the elements of our observation vector, which are the opposite of the logarithm of the annual survival probabilities. We write:

$$\bar{q}^j(t) := -\ln(1 - \hat{q}^j(t)) = -\ln(\hat{p}^j(t)).$$

Equation (2) allows us to state that our observations are linear in the values of the factors. We still have to address how to treat the information coming from the life tables of the different populations. For simplicity, we adopt a centralized data fusion approach (see [15]). Hence, we gather all the information as part of a unique observational vector. We assume that each observation is a measurement of a quantity and the measurement entails some noise. We specify thus the following linear model:

$$\bar{q}_{x+t}^j(t) = -\alpha^j(x, t, t+1) - \beta^j(x, t, t+1)' Y^j(t) + \epsilon_x^j(t), \quad x \in X^j, \quad (3)$$

where $\epsilon_x^j(t)$ is the measurement noise. Following [1], we assume that the observation noise is Gaussian with mean zero and standard deviation proportional to the observation itself, i.e., $\text{Var}[\epsilon_x^j(t)] = s_{x,j}^2(t) = [k^j \bar{q}_x^j(t)]^2$, $k^j \in \mathbb{R}$. This assumption makes sense in our context because when age increases, the cohort size decreases. Observations relative to older ages, for which $\bar{q}_x$'s are larger, are likely to be less accurate than observations of younger ages. Notice that we allow for different measurement noises strengths k^j for each group to deal with the fact that measurements can come from different sources. We further assume that measurement errors are uncorrelated across ages and populations. Coupling this description of our data with the description of the state processes given in Section 3, our model has the following state-space representation:

$$\bar{Q}(t) = -\alpha(t) - \beta(t)Y(t) + \epsilon(t), \quad \epsilon(t) \sim \mathcal{N}(\mathbf{0}, R(t)) \quad (4)$$

$$Y(t+1) = DY(t) + \eta(t+1), \quad (5)$$

where $Y(t)$ is the M -dimensional vector of state process values that gathers all the state processes influencing the J populations of interest, $\bar{Q}(t)$ and $\epsilon(t)$ are the vectors that collect all the observations and corresponding measurement errors, respectively. $R(t)$ is a diagonal matrix, with elements $s_i^2(t)$ on the diagonal. The coefficients collected in $\alpha(t)$ and $\beta(t)$ are functions of both the parameters of the state-vector process and of the parameters of the mortality law. $Y(t)$ is the vector of state processes, while $\eta(t+1)$ collects its noise. Since $Y(t)$ is a vector of square-root diffusions, the noises are not Gaussian: the transition density of each element of Y follows a non-central χ^2 distribution. In this context, the Kalman filter, which is the traditional tool used to estimate Gaussian linear state-space models, is still the optimal linear filter (see [8]). However, although our measurement equation is linear, the standard Kalman filter estimates may be conditionally biased, because innovations in the state processes are not Gaussian (see, for instance, [16]). Several authors have discussed this issue and provided alternative solutions to this problem. We follow [7], who modified the standard Kalman filter to estimate the multi-factor Cox–Ingersoll–Ross model via a quasi-linear Kalman filter. Such estimator introduces modifications to the Kalman filter:

- it captures the dependence of the conditional variance on the current value of the state;
- it puts a non-negativity constraint on the value of the state process.

Indeed, the modified Kalman filter approximates the state-process transition density to a Gaussian density with the same conditional first two moments and introduces a non-

linearity through the non-negativity constraint. The resulting quasi-maximum likelihood is then maximized. In implementing the Kalman filter recursion, we adopt a sequential formulation (see [8,17]) which allows us to avoid matrix inversion. Indeed, we could have followed Monte Carlo approaches as in [18]. In particular, particle filtering was used in [19] in the context of interest rate modeling with CIR-type models. However, given the dimensionality of our problem, which includes several factors, a large number of observations and parameters to estimate, we preferred to apply Quasi-Maximum Likelihood methods, which turned out to be less computationally intensive. Our approach is justified in light of the evidence provided by several works [7,20,21] regarding the accuracy of the method for affine interest rate term-structure models.

4.1. Description of the Quasi-Linear Kalman Filter Algorithm

The Quasi-Linear Kalman Filter algorithm, as mentioned above, allows us to obtain an estimate of the conditional state process vector mean and variance covariance matrix, given the observations:

$$\hat{Y}(t) := \mathbb{E}[Y(t) \mid \mathcal{F}(t-1)] \quad \text{and} \\ P(t) := \text{Var}[Y(t) \mid \mathcal{F}(t-1)].$$

Additionally, it allows us to compute an approximate log-likelihood for our model, given the parameters. An optimization procedure then will explore the parameter space to maximize such log-likelihood. We report here the algorithm that implements the sequential quasi-linear Kalman filter equations for our model. Let us first denote our matrices as

$$\bar{Q}(t) = [\bar{q}_1(t), \dots, \bar{q}_p(t)]' \\ \epsilon(t) = [\epsilon_1(t), \dots, \epsilon_p(t)]' \\ \beta(t) = [\beta_1(t), \dots, \beta_p(t)]',$$

where $p = \sum_j X^j$ is the total number of observations considered for each calendar year, $\bar{q}_i(t)$ —with a slight abuse of notation—denotes the i -th observation, where i ranges over the set of integers which uniquely define the elements of the couple (x, j) . $\beta_i(t)$'s are row vectors. From now on, we also suppress dependence on x and $x + t$ for notational convenience. Since R is a diagonal matrix, the equation that describes the model for each single observation of interest is

$$\bar{q}_i(t) = -\alpha_i(t) - \beta_i(t)Y(t) + \epsilon_i(t). \quad (6)$$

The sequential Kalman filter algorithm runs then as follows:

1. Given a certain parameter set ξ , the filter is initialized using the unconditional mean and variance of the state-process vector,

$$\hat{Y}(1) = \mathbb{E}[Y(1) \mid \mathcal{F}(0)] = \mathbb{E}[Y(1)]. \\ P(1) = \text{Var}[Y(1) \mid \mathcal{F}(0)] = \text{Var}[Y(1)] = P,$$

2. For each time step t , the state process vector and its variance covariance matrix are updated:

$$\hat{Y}^-(t+1) = T(t)\hat{Y}(t) + \Theta(I_N - T(t))\hat{Y}(t) \\ P^-(t+1) = T(t)P(t)T'(t) + \Gamma(t+1),$$

where $T = \text{diag}[e^{-a_1}, \dots, e^{-a_M}]$, I_M denotes an M -dimensional identity matrix, $\Theta = \text{diag}[\theta_1^j, \dots, \theta_M^j]$, and $\Gamma(t+1)$ is the state-process conditional variance/covariance matrix. Notice that such variance/covariance matrix depends on $\hat{Y}(t)$, differently relative to the standard Gaussian case.

3. The estimates are initialized as

$$\begin{aligned}\hat{\mathbf{Y}}_0^+(t+1) &= \hat{\mathbf{Y}}^-(t) \\ \mathbf{P}_0^+(t+1) &= \mathbf{P}^-(t),\end{aligned}$$

and then updated after each observation $i = 1, \dots, p$, setting a floor to the value of the estimated mean to zero:

$$\begin{aligned}\hat{\mathbf{Y}}_i^+(t) &= \max(\hat{\mathbf{Y}}_{i-1}^+(t) + \mathbf{K}_i(t)v_i^+(t), 0) \\ \mathbf{P}_i^+(t) &= \mathbf{P}_{i-1}^+(t) - \mathbf{K}_i(t)\boldsymbol{\beta}_i(t)\mathbf{P}_{i-1}^+(t),\end{aligned}$$

where the measurement error prediction $v_i(t)$ and the Kalman gain $\mathbf{K}_i(t)$ are respectively:

$$\begin{aligned}v_i^+(t) &= \bar{q}_i(t) - (\boldsymbol{\alpha}_i(t) + \boldsymbol{\beta}_i(t)\hat{\mathbf{Y}}_{i-1}^+(t)) \\ \mathbf{K}_i(t) &= \frac{\mathbf{P}_{i-1}^+(t)\boldsymbol{\beta}_i'(t)}{s_i^2(t)}.\end{aligned}$$

Note that $v_i(t)$ and $F_i(t)$ are scalars and $\mathbf{K}_i(t)$ is a N-dimensional column vector.

4. Finally, when $i = p$, we set

$$\begin{aligned}\hat{\mathbf{Y}}^+(t) &= \hat{\mathbf{Y}}_p^+(t); \\ \mathbf{P}^+(t) &= \mathbf{P}_p^+(t).\end{aligned}$$

and we compute

$$\begin{aligned}F_i(t) &= \boldsymbol{\beta}_i(t)\mathbf{P}^+(t)\boldsymbol{\beta}_i'(t) + s_i^2(t), \\ v_i(t) &= \bar{q}_i(t) - (\boldsymbol{\alpha}_i(t) + \boldsymbol{\beta}_i(t)\hat{\mathbf{Y}}^+(t)).\end{aligned}$$

5. Using these last expressions for $F_i(t)$ and $v_i(t)$ we can compute the log-likelihood function:

$$\log L(\boldsymbol{\xi}) = \text{const} - \frac{1}{2} \sum_{t=1}^T \sum_{i=1}^p \log F_i(t) + v_i^2(t)F_i^{-1}(t),$$

which has to be maximized to obtain the optimal parameter set.

4.2. Log-Likelihood Maximisation

Log-likelihood maximization is performed using a global differential evolution algorithm [22]. With this algorithm, several sets of parameters (a population) are initially generated. Then, for each of these sets (individuals), the log-likelihood is computed. The best member of the population, i.e., the parameter set that provides the highest log-likelihood, survives to the next iteration. The other members may survive with a certain probability, but their genes (single parameters) can mutate, i.e., change their value to a different one, randomly picked. The process of population generation and mutation continues until a selected number of iterations is performed, or the best population member is unchanged after a given number of iterations.

5. Results

Since the UK annuity market represents the largest annuity market in the world, the UK's mortality data often represent a point of departure for the estimation of mortality models. This section applies our model to estimate two stochastic mortality laws to the UK males and females population. The UK aggregate mortality data are sourced from Scotland, England and Wales and Northern Ireland. While we do not consider the regional mortality experiences separately, our modelling framework could encompass the case

where region-specific parameters are accounted for, together with different measurement errors across regions. Specifically, our data sample is composed of one-year survival probabilities for ages 0–89 from 1967 to 2017, taken from the Human Mortality Database (www.mortality.org, accessed on 7 December 2021).

5.1. First Makeham's Multi-Population Law

We estimate the following model, where $j = 1, 2$ denote the male and female UK population, respectively:

$$\mu_{x+t}^j(t) := Y_1(t) + (\gamma^j)^{x+t} Y_2(t).$$

In this case, $g^j(x+t, t, x) := [1, (\gamma^j)^{x+t}]$, with $Y \in \mathbb{R}^2$. The model has two factors, which are the same across the populations. Indeed, we dropped dependence on j above. The coefficients $\alpha(\cdot)$ and $\beta(\cdot)^j = [\beta_1^j, \beta_2^j]$ can thus be obtained solving the following ODE system:

$$\dot{\alpha}^j = -A\theta \cdot \beta^j, \dot{\beta}^j = [1, (\gamma^j)^{x+t}]' + (A')' \beta^j - \left[\frac{1}{2} (\beta_1^j)^2 \sigma_1^2, \frac{1}{2} (\beta_2^j)^2 \sigma_2^2 \right]',$$

with final conditions $\alpha^j(x, T, T) = 0$ and $\beta^j(x, T, T) = \mathbf{0}$.

While affected by the same factors, they respond differently to the second factor. It is the only parameter γ^j that shapes the mortality–age relation for the different populations. Indeed, though very parsimonious, this specification is not rich enough to capture well the complexity of the whole mortality curve. We restrict our application to older ages only, ranging from 50 to 89. Table 1 reports the estimated parameters. The population-specific parameter, which captures the steepness of the exponential increase in mortality at old ages captured by Makeham's mortality law is not unexpectedly higher for UK males (0.1153) than for females (0.0944).

Table 1. This table collects the optimal parameters for our model. For each population, the “common” column collects the parameters relative to the factor processes and the measurement variance–covariance matrix s . The other two columns collect the parameters relative to the response functions of the two populations of males and females.

Maximized Log-Likelihood	22,409	Number of Parameters	9		
Optimal Parameters					
Common		Males		Females	
Parameter	Value	Parameter	Value	Parameter	Value
a_1	0.0233	γ^m	0.1153	γ^f	0.0944
a_2	0.0111				
σ_1	2.87×10^{-7}				
σ_2	0.0109				
θ_1	0.0014				
θ_2	2.37×10^{-6}				
s	0.1076				

Figure 1 displays the evolution of the values of the latent factors in time. The first factor, which acts as a baseline mortality intensity level, is estimated to decline rather smoothly over time. Notice that its initial value is higher than the long-term estimated average θ_1 (0.0014). This parameter can be interpreted as the estimated biological limit which the value of the factor can tend to. The speed of mean reversion a_1 (0.0233) can be interpreted as the speed at which the mortality baseline level tends to converge to such

long-term value. The second factor, which interacts with age to reproduce the exponential increase of mortality at older ages, also declines. Its volatility is higher because it influences mortality at older ages, reflecting the well-known fact that higher uncertainty surrounds older age mortality improvements. The current values of the factors (i.e., the estimates for the last calendar year) are still far from their long-term averages. Taking into account their exponential decay, we expect the current values (4×10^{-3} and 6×10^{-6} respectively) to halve in approximately 173 years, and 115,524 years respectively, and thus still be above their long-term average. The model produces a good fit of the mortality surface, despite its very simple form. Figure 2 reports the Mean Absolute Relative Errors of the single female and male ages (averaged over calendar years):

$$MARE(x, j) = \frac{1}{T} \sum_{t=1}^T \left| \frac{\hat{q}_x^j(t) - q_x^j(t)}{q_x^j(t)} \right|,$$

and of each calendar year (averaged over all the observations):

$$MARE(t) = \frac{1}{p} \sum_{i=1}^p \left| \frac{\hat{q}_i(t) - q_i(t)}{q_i(t)} \right|,$$

where $\hat{q}$ denotes the fitted one-year death probability. The model performance is good, given that 4080 observations are fitted with only nine parameters. Average MARE is 7.22%, and it is lower than 10% for almost all ages, apart from older males. This is in line with the two-factor model performance in [1], despite our two-population fit. Only older male figures seem to be more problematic to fit. Performance may be improved assuming that one state process (for instance, the first factor) differs across the two populations.

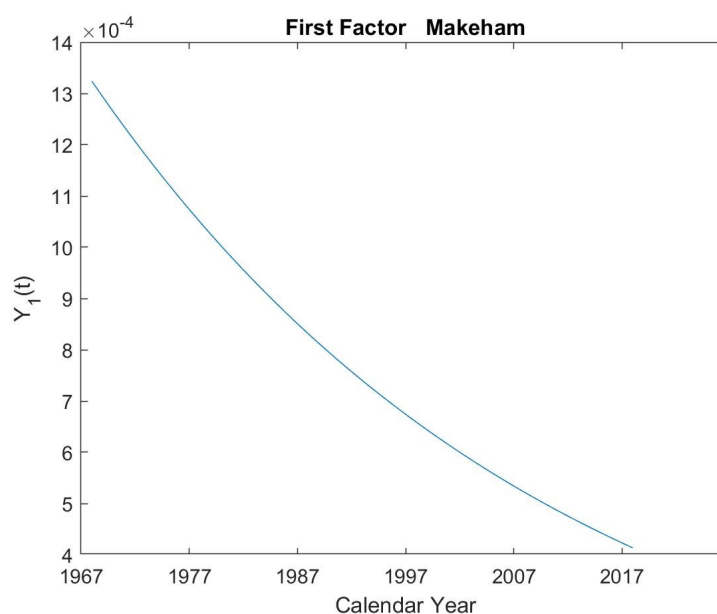


Figure 1. Cont.

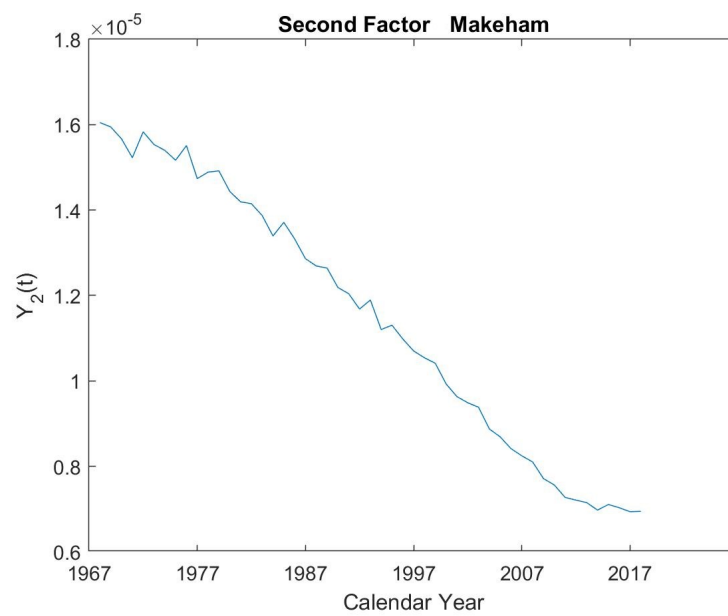


Figure 1. The figure plots the values of the two factors, Y_1, Y_2 .

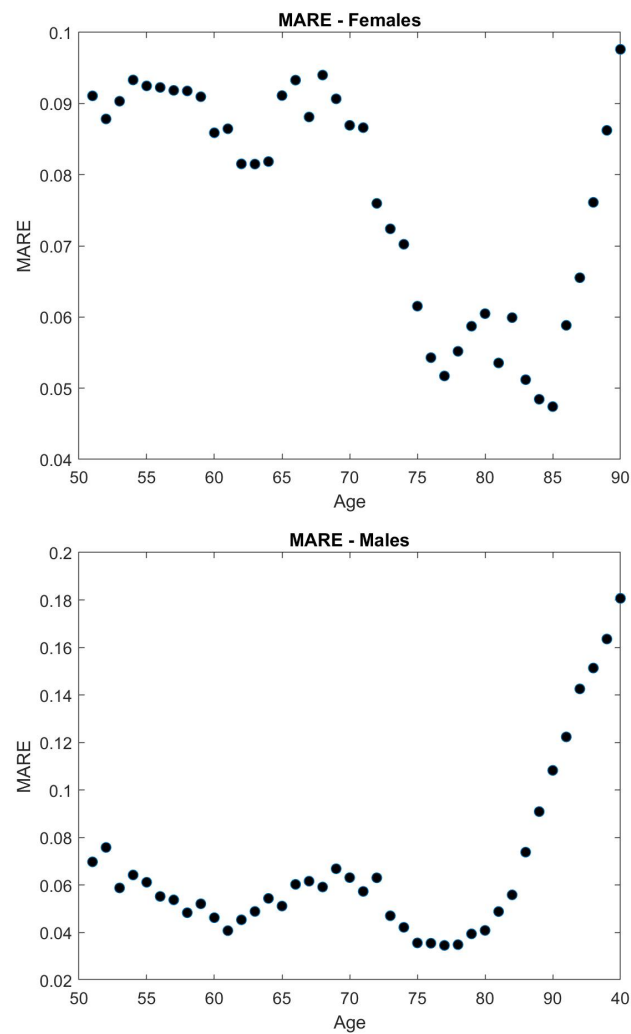


Figure 2. Cont.

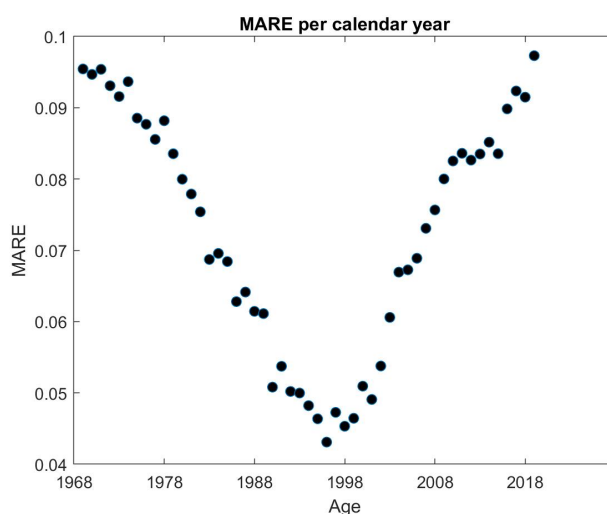


Figure 2. The figures plot the values of the MARE for female and male ages and averaged over calendar years.

5.2. Modified Thiele's Law

We estimate the following model, where $j = 1, 2$ denote the male and female UK population, respectively:

$$\mu_{x+t}^j(t) := Y_1(t)e^{-\gamma_1^j(x+t)^{\frac{1}{\gamma_2^j}}} + Y_2(t)e^{-\gamma_3^j(x+t-\gamma_4^j)^2} + Y_3(t)(\gamma_5^j)^{x+t}.$$

In this case, $g^j(x+t, t, x) := [e^{-\gamma_1^j(x+t)^{\frac{1}{\gamma_2^j}}}, e^{-\gamma_3^j(x+t-\gamma_4^j)^2}, (\gamma_5^j)^{x+t}]$ and $\mathbf{Y} \in \mathbb{R}^3$. This is a 3-factor model, where factors are the same across the populations. The coefficients $\alpha(\cdot)$ and $\beta(\cdot)^j = [\beta_1^j, \beta_2^j, \beta_3^j]$ can be obtained solving the system of ODEs:

$$\dot{\alpha}^j = -\mathbf{A}\boldsymbol{\theta} \cdot \boldsymbol{\beta}^j, \dot{\boldsymbol{\beta}}^j = [e^{-\gamma_1^j(x+t)^{\frac{1}{\gamma_2^j}}}, e^{-\gamma_3^j(x+t-\gamma_4^j)^2}, \gamma_5^j]^{x+t} + \mathbf{A}'\boldsymbol{\beta}^j - \left[\frac{1}{2}(\beta_1^j)^2\sigma_1^2, \frac{1}{2}(\beta_2^j)^2\sigma_2^2, \frac{1}{2}(\beta_3^j)^2\sigma_3^2 \right]',$$

with final conditions $\alpha^j(x, T, T) = 0$ and $\boldsymbol{\beta}^j(x, T, T) = \mathbf{0}$.

Table 2 reports the estimated parameters.

Table 2. This table collects the optimal parameters for our model and is divided in two panels, one for each population. For each population, the “common” column collects the parameters relative to the factor processes and the measurement variance covariance matrix s . The other two columns collect the parameters relative to the response functions of the two sub-populations of males and females.

Maximized log-Likelihood		Number of Parameters			
65,464		20			
Optimal Parameters					
Common		Males		Females	
Parameter	Value	Parameter	Value	Parameter	Value
a_1	2.95×10^{-18}	γ_1^m	14.8170	γ_1^f	14.4989
a_2	6.40×10^{-6}	γ_2^m	15.8046	γ_2^f	15.0810
a_3	2.76×10^{-16}	γ_3^m	0.0048	γ_3^f	0.0229
σ_1	0.8521	γ_4^m	88.2853	γ_4^f	23.1866
σ_2	2.21×10^{-6}	γ_5^m	0.0944	γ_5^f	0.1016
σ_3	2.10×10^{-13}				
θ_1	88.8547				
θ_2	2.08×10^{-8}				
θ_3	5.32×10^{-7}				
s	0.1496				

The factors have a constantly decreasing trend in time, apart from the second factor, which decreases over the last 15 calendar years only as depicted by Figure 3 (see Table 1). The first and the third factors decline very fast until the 2000s, to depict the extraordinary improvements in infant mortality and mortality at older ages. As for the Makeham’s law case, the factors are still way above their long-term averages θ and thus we expect further improvements of mortality in the long-run. Such long-term averages are very important for long-run projections of mortality rates. In unreported estimations, we test the changes of the values of these parameters to slight changes in the estimation time window, finding reasonable stability.

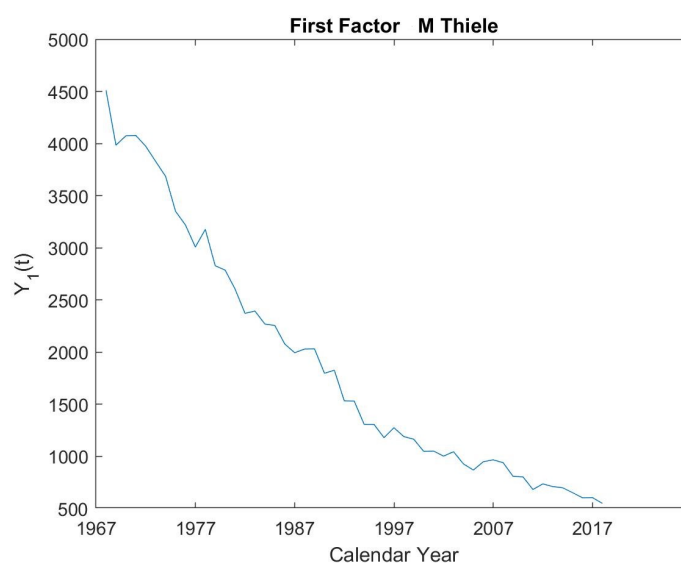


Figure 3. Cont.

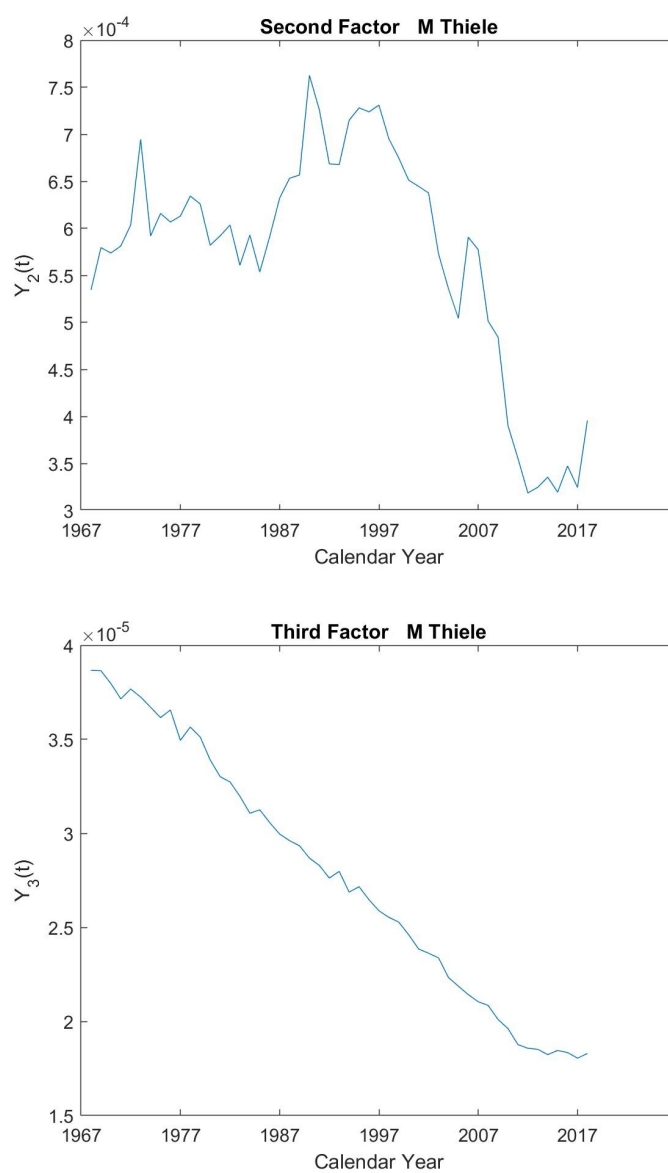


Figure 3. The figure plots the values of the three factors, Y_1 , Y_2 , Y_3 .

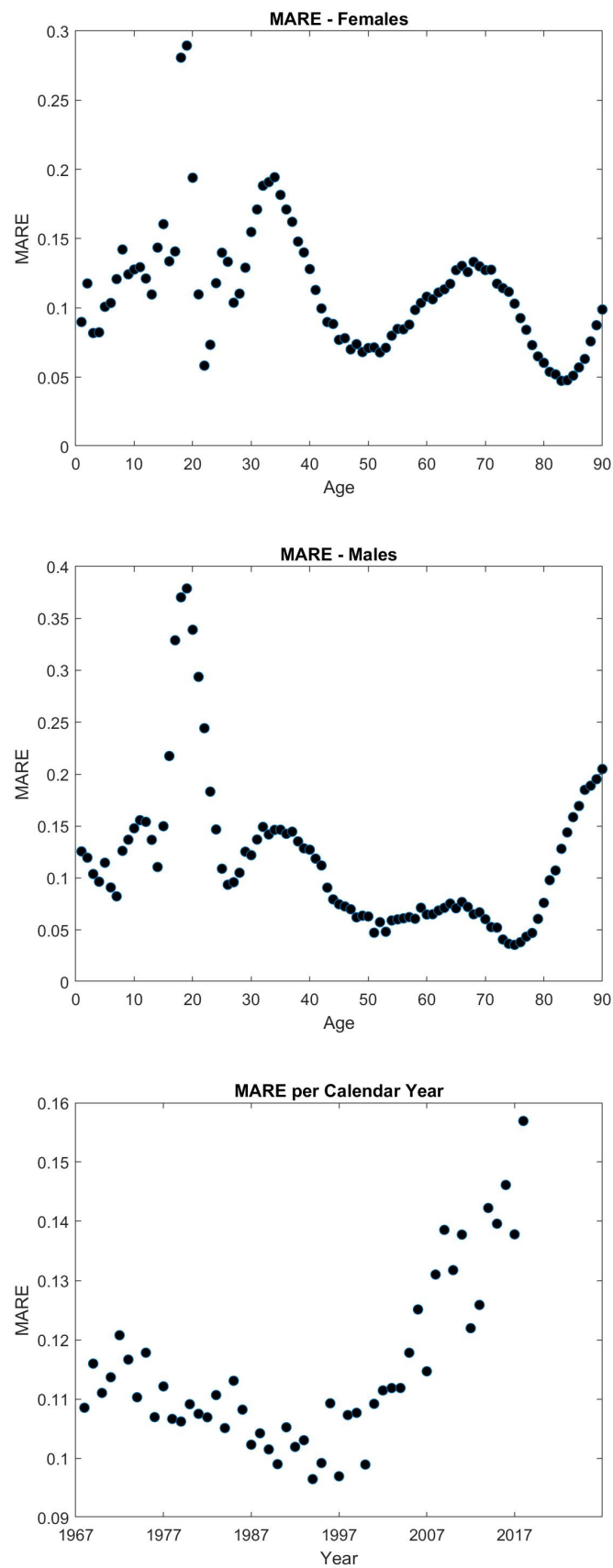


Figure 4. The figure plots the values of the MARE for the ages of female and male participants and averaged over calendar years.

In our estimation procedure, we have not constrained the value of the parameter that controls the age which is mostly affected by the second factor. Thus, the factor affects the different populations differently. For example, it influences female mortality at young adult ages to help to address the accident hump ($\gamma_4^f = 23.1866$). For males, it instead adds up to the third factor in influencing mortality at older ages ($\gamma_4^m = 88.2853$). As Figure 4 shows, the model's fit is very good for both female and male populations, with an average MARE across all observations of 11.47%. The first factor helps capture the infant mortality spike at age 0. Mortality at young adult ages and at older ages is slightly more difficult to capture for males, for which MARE is above the average between ages 18 and 23. Between ages 40 and 70, the curve's fit is better for males than females.

6. Conclusions

In this paper, we have focused on a stochastic and multi-population extension of traditional actuarial mortality laws. We place ourselves within the setup introduced by [2]. The mortality intensities of all (or a large set of) the ages of multiple populations are modeled within the affine framework as functions of population-specific parameters, which shape the age dependence and some unobservable factors that describe the evolution of mortality trends over time. Here, these latent factors are modeled as square-root processes and describe the evolution of mortality trends over time. The estimation procedure for such models entails dealing with a non-Gaussian state-space model. In this case, estimation via Kalman filter still provides us with the optimal linear state estimator. However, this may be sub-optimal due to the non-linearities introduced by the state-vector process density. We followed [7] and applied a Quasi-Linear Kalman Filter algorithm that approximates the state process transition density to a Gaussian density truncated at 0 with the same conditional first two moments. Although approximate, the procedure was proven accurate enough in a similar setup when fitting multi-factor CIR models for interest rates. Relative to previous similar multi-population models, the mortality intensities are thus guaranteed to be non-negative. This is indeed an advantage relative to Gaussian models, especially when the model needs to be applied to forecast intensities over long time periods. Obviously, this can be important when dealing with the pricing of long-term liabilities, such as annuities, or with risk management for pension funds. Inspired by the Makeham and the Thiele laws, our application provides the calibration of two stochastic multi-population mortality laws to the UK populations of males and females mortality rates jointly from 1968 to 2017. The models are parsimonious and are shown to provide a good fit of the mortality surface over time.

Author Contributions: Conceptualization, P.J. and L.R.; methodology, P.J. and L.R.; software, P.J. and L.R.; validation, P.J. and L.R.; formal analysis, P.J. and L.R.; investigation, P.J. and L.R.; data curation, P.J. and L.R.; writing—original draft preparation, P.J. and L.R.; writing—review and editing, P.J. and L.R.; visualization, P.J. and L.R.; supervision, P.J. and L.R.; project administration, P.J. and L.R.; funding acquisition, P.J. and L.R.. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the "Dipartimenti di Eccellenza 2018-2022" grant by the Italian Ministry of Education and Research.

Acknowledgments: Luca Regis acknowledges funding from the "Dipartimenti di Eccellenza 2018-2022" grant by the Italian Ministry of Education and Research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Schrager, D.F. Affine stochastic mortality. *Insur. Math. Econom.* **2006**, *38*, 81–97, doi:10.1016/j.insmatheco.2005.06.013.
2. Jevtić, P.; Regis, L. A continuous-time stochastic model for the mortality surface of multiple populations. *Insur. Math. Econom.* **2019**, *88*, 181–195, doi:10.1016/j.insmatheco.2019.07.001.
3. Jevtić, P.; Luciano, E.; Vigna, E. Mortality surface by means of continuous time cohort models. *Insur. Math. Econom.* **2013**, *53*, 122–133, doi:10.1016/j.insmatheco.2013.04.005.

4. Blackburn, C.; Sherris, M. Consistent dynamic affine mortality models for longevity risk applications. *Insur. Math. Econom.* **2013**, *53*, 64–73, doi:10.1016/j.insmatheco.2013.04.007.
5. De Rosa, C.; Luciano, E.; Regis, L. Geographical diversification and longevity risk mitigation in annuity portfolios. *ASTIN Bull. J. IAA* **2021**, *51*, 375–410.
6. Cox, J.C.; Ingersoll, J.E., Jr.; Ross, S.A. An intertemporal general equilibrium model of asset prices. *Econometrica* **1985**, *53*, 363–384, doi:10.2307/1911241.
7. Chen, R.R.; Scott, L. Multi-factor Cox-Ingersoll-Ross models of the term structure: Estimates and tests from a Kalman filter model. *J. Real Estate Financ. Econ.* **2003**, *27*, 143–172.
8. Simon, D. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*; John Wiley & Sons: Hoboken, NJ, USA, 2006.
9. Pitacco, E.; Denuit, M.; Haberman, S. *Modelling Longevity Dynamics for Pensions and Annuity Business*; Oxford University Press: Oxford, UK, 2009.
10. Milevsky, M.A.; Promislow, S.D. Mortality Derivatives and the Option to Annuity, *Insur. Math. Econom.* **2001**; Volume 29, pp. 299–318, doi:10.1016/S0167-6687(01)00093-2.
11. Lando, D. On Cox processes and credit risky securities. *Rev. Deriv. Res.* **1998**, *2*, 99–120.
12. Duffie, D.; Pan, J.; Singleton, K. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* **2000**, *68*, 1343–1376, doi:10.1111/1468-0262.00164.
13. Filipović, D. *Term-Structure Models: A Graduate Course*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2009.
14. Duffie, D.; Kan, R. A yield-factor model of interest rates. *Math. Financ.* **1996**, *6*, 379–406.
15. Raol, J.R. *Multi-Sensor Data Fusion with MATLAB®*; CRC Press: Boca Raton, FL, USA, 2009.
16. Duan, J.C.; Simonato, J.G. Estimating and testing exponential-affine term structure models by Kalman filter. *Rev. Quant. Financ. Account.* **1999**, *13*, 111–135.
17. Koopman, S.J.; Durbin, J. Fast filtering and smoothing for multivariate state space models. *J. Time Ser. Anal.* **2000**, *21*, 281–296, doi:10.1111/1467-9892.00186.
18. Kitagawa, G. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Stat.* **1996**, *5*, 1–25, doi:10.2307/1390750.
19. De Rossi, G. Maximum likelihood estimation of the Cox–Ingersoll–Ross model using particle filters. *Comput. Econ.* **2010**, *36*, 1–16.
20. Jong, F.d. Time series and cross-section information in affine term-structure models. *J. Bus. Econ. Stat.* **2000**, *18*, 300–314.
21. Chen, R.R.; Scott, L. Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates. *J. Fixed Income* **1993**, *3*, 14–31.
22. Storn, R.; Price, K. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J. Glob. Optim.* **1997**, *11*, 341–359.