ARC Centre of Excellence in Population Ageing Research

Working Paper 2021/27

AffineMortality: An R package for estimation, analysis and projection of affine mortality models

Francesco Ungolo, Len Patrick Dominic M. Garces, Michael Sherris, and Yuxin Zhou

`AffineMortality`: An R package for estimation, analysis and projection of affine mortality models

Francesco Ungolo[1,2,3], Len Patrick Dominic M. Garces[2], Michael Sherris[1,2], and Yuxin Zhou[1,2]

[1] *School of Risk and Actuarial Studies, University of New South Wales, Kensington NSW 2052, Australia*
[2] *ARC Centre of Excellence in Population Ageing Research, University of New South Wales, Kensington NSW 2052, Australia*
[3] *Chair of Mathematical Finance, Technical University of Munich, 85748 Garching bei München, Germany*

March 30, 2023

This paper presents the `AffineMortality` R package which performs parameter estimation, goodness of fit analysis, simulation and projection of future mortality rates for a set of affine mortality models for use in pricing and reserving. The computational routines build on the univariate Kalman Filtering approach of Koopman & Durbin (2000) along other numerical methods to enhance the robustness of the results. This paper provides a discussion of how the package works in order to effectively estimate and project the models, and describes the available functions. Illustration of the package for mortality analysis of the US HMD dataset is provided.

Keywords: Longevity Risk, Kalman Filter, State-space models, Affine mortality

1. Introduction

The analysis of mortality rates is fundamental for actuaries as these used to develop and set the premium for life insurance products, to estimate liabilities, and to develop the corresponding risk management strategies.

It is widely acknowledged in practice how the development of future mortality rates is the outcome of a stochastic process (Cairns et al. 2006*a*). The seminal work of Lee & Carter (1992) attracted significant attention in the last three decades towards the development and the extension of stochastic mortality models, which, to different degrees, are capable of capturing the different features of the mortality surface, such as the presence of the cohort effect (Willets 2004). These models have been originally developed for the analysis of data in discrete time, such as integer ages and integer calendar (or birth) year. Some examples include the Cairns-Blake-Dowd model (Cairns et al. 2006*b*), the Poisson log-bilinear approach of Brouhns et al. (2002), the Plat model (Plat 2009) and the functional approach of Hyndman & Shahid Ullah (2007).

More recently, models developed using the tools of financial mathematics gained attention in the literature, following the work of Milevsky & Promislow (2001). For example, the papers of Schrager (2006), Biffis (2005), Dahl (2004), Blackburn & Sherris (2013), Jevtić et al. (2013), Jevtić & Regis (2019) and Jevtić & Regis (2021) propose the use of the affine interest rate modelling framework developed by Duffie & Kan (1996) for the analysis of mortality dynamics. In this way, it is possible to obtain closed-form formula for the survival curves, for use in pricing longevity-related securities and devise risk management strategies for these products. These models assume that the mortality intensity process is driven by a set of latent variables, whose dynamics are characterized by a stochastic differential equation with mean reversion. This implies that the closed-form survival curve is an exponentially affine function of the latent variables.

An advantage of the continuous-time approach for mortality modelling is the use of financial pricing techniques which are familiar to market practitioners. Specifically, a no-arbitrage valuation framework can be used for pricing life-contingent products and developing appropriate risk management strategies for these products using the analytical results for affine processes (Biffis 2005).

This paper describes the R package `AffineMortality`, which supports an extensive analysis of continuous time affine mortality models in the spirit of Blackburn & Sherris (2013), Huang et al. (2022) and Ungolo et al. (2023). More precisely, the package estimates the parameters of these models using mortality data collected at discrete time points and ages. In addition, the package facilitates the analysis of model fit and the simulation and projection of future mortality rates. These tasks require us to discretize the continuous-time models and to recast the inferential problem in a state-space form.

Schrager (2006) and Ungolo et al. (2023) describe how the estimation of these models using the base Kalman filter can be problematic due to the numerical issues which follow from the multiplication and inversion of large-dimensional matrices. Hence, they advocate the use of the univariate Kalman filter of Koopman & Durbin (2000). For this reason, the computational routines in `AffineMortality` are implemented using the univariate Kalman filtering approach, together with other numerical tricks described in Ungolo et al. (2023) which make the Kalman filtering procedure more stable.

The package allows to implement and assess among others, the Blackburn-Sherris model with independent factors, the Blackburn-Sherris model with two and three dependent factors, the Arbitrage-Free Nelson-Siegel mortality model with independent and dependent factors, and the Cox-Ingersoll-Ross mortality model. The plan is to further expand the library of models available for analysis, and to extend the existing ones by accounting for cohort or period specific factors.

To the best of our knowledge, currently available software focus on the analysis of mortality models in discrete time. For example, the R package `StMoMo` (Villegas et al. (2018)) allows for the analysis, among others, of the Lee-Carter (Lee & Carter (1992)), CBD (Cairns et al. (2006*b*)) and the age-period-cohort model by Renshaw & Haberman (2006). In a similar fashion, `StanMoMo` (Barigou et al. (2023)) performs a Bayesian analysis of stochastic mortality models using Stan. The R package `demography` (Hyndman et al. 2022) instead provides functions for demographic analysis including: lifetable calculations; Lee-Carter modelling; functional data analysis of mortality rates, fertility rates, net migration numbers, and stochastic population forecasting.

Furthermore, there are `R` packages for the analysis of state-space models, such as `dse` (Gilbert 2009), `sspir` (Dethlefsen et al. 2022), `dlm` (Petris 2010), `FKF` (Luethi et al. 2022) and `KFAS` (Helske 2017) (see also Tusell (2011) for a comprehensive review). However, these packages do not readily adapt to the state-space model used for the analysis of affine mortality models. This is because the system matrices of the affine mortality models follow as the solution of an ordinary differential equation, as we briefly illustrate in Section 3, and its subsequent discretization.

R and RStudio may be subject to crashes. Since the optimization process may take a long time to perform, the functions in `AffineMortality` which require a lot of time to execute (e.g. `affine_fit()` for estimating the model parameters) allow the user to stop the process without losing the computations performed thus far. The package deals with fault tolerance by allowing the user to input a directory where the work can be saved as an `.Rdata` file.

`AffineMortality` can be installed by using the following commands[1]:

```
library(devtools)
install_github("ungolof/AffineMortality")
library(AffineMortality)
```

The source code of `AffineMortality` is available through the Github repository `https://github.com/ungolof/AffineMortality`.

The paper develops as follows: Section 2 introduces the data to be used as input for the analysis, Section 3 summarizes the affine mortality modelling framework and briefly describes the mortality models supported by the package, and Section 4 describes their parameter estimation procedure. Section 5 describes how the package can be used to perform the goodness-of-fit analysis and compare affine models, and Section 6 describes the function `affine_project()`, which can be called to project future cohort survival curves. The package provides two methods to analyse parameter uncertainty, described in Section 7: the first estimates the covariance of the parameter estimates by using the bootstrap, while the other implements a multiple imputation based method. The step-by-step illustration of the package is provided in Section 8. Section 9 describes other functions in `AffineMortality` and Section 10 concludes.

---

[1] The required R package `devtools` should be already installed on the machine.

2. Input data

Let $\mu_{x,t}$ denote the force of mortality for an individual aged x last birthday in calendar year t. Without loss of generality $\mu_{x,t}$ is approximated as the central death rate $m_{x,t}$, when we assume that the force of mortality is constant between each integer age x and $x+1$, and between each calendar year t and $t+1$. The central death rate $m_{x,t}$ is empirically estimated as the ratio between the observed number of deaths $d_{x,t}$ and the central exposure at risk years $E^c_{x,t}$.

The survival probability at time t of an individual aged x at time t until age $x+T-t$ is given by:

$$S_x(t,T) = \prod_{j=1}^{T-t} \exp\left(-\mu_{x+j-1,t}\right) = \exp\left(-\sum_{j=1}^{T-t} \mu_{x+j-1,t}\right) \tag{2.1}$$

The estimation of affine mortality models uses the average force of mortality denoted for each calendar year t and each age as:

$$\bar{\mu}_x(t,T) = \frac{1}{T-t}\sum_{j=1}^{T-t} \mu_{x+j-1,t} \approx \frac{1}{T-t}\sum_{j=1}^{T-t} m_{x+j-1,t} \tag{2.2}$$

Here, x is fixed and denotes the smallest age in the age-range of interest (for example, equal to 50 in Ungolo et al. (2023) and Huang et al. (2022)). The dataset for the analysis is a matrix of dimension $N \times K$, where $N = (T-t)$ is the number of ages in the age-range of interest and K is the number of calendar years for the analysis.

Conversely, given the average force of mortality, we can obtain μ, by means of the following recursion, starting from $\mu_{x,t} = \bar{\mu}_x(t, t+1)$:

$$\mu_{x+i,t} = i\bar{\mu}_x(t,t+i) - (i-1)\,\bar{\mu}_x(t,t+i-1) \tag{2.3}$$

for $i = 2, \ldots, N$. Furthermore, let us define the column vector $\bar{\mu}_t = [\bar{\mu}_x(t,t+1), \ldots, \bar{\mu}_x(t,t+N)]'$ (we drop the reference on x for practical reasons).

When processing the force of mortality data $\mu_{x,t}$, the function `rates2avg()` in `AffineMortality` can be used to transform the $N \times K$-dimensional matrix of $\mu_{x,t}$ rates into the corresponding matrix of $\bar{\mu}$ rates. The inverse operation can be carried out using the function `avg2rates()`.

The matrix of average forces of mortality is the key input for the estimation of the affine mortality models. The use of the average forces of mortality yields smoother data which renders the estimation process more stable. This is the approach adopted within the interest rate literature (see Christensen et al. (2011)), as well as in the analysis of affine mortality models (Blackburn & Sherris (2013), Huang et al. (2022) and Jevtić & Regis (2019)).

A similar data set can be set up for the analysis of age-cohort mortality rates, as discussed in Huang et al. (2022) and Ungolo et al. (2023).

3. Affine Mortality Models

In this section, we provide a brief overview of the affine mortality modelling framework, and describe the different mortality models supported by `AffineMortality`. We present the framework in a risk-neutral probability space $(\Omega, \mathcal{F}, Q)$ where Q is an appropriate pricing measure for both the financial and insurance markets. The Q measure ensures a consistent pricing property for the fitted survival probabilities. In Section 3.5, we apply the Girsanov's theorem to specify the model dynamics under the historical probability measure in order to estimate the model parameters. We consider a filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0}$ which can be decomposed as $\mathbb{F} = \mathbb{G} \vee \mathbb{H}$, where $\mathbb{G}$ is a filtration containing all financial and actuarial information, except the time of death, which is contained in $\mathbb{H}$. See, for example, Biffis (2005) and Blackburn & Sherris (2013) for further details.

The random time of death of an individual, currently aged x, is assumed to be a doubly-stochastic stopping time with intensity process $\mu_x(t)$, which is modelled as an affine function of an M-dimensional, $\mathbb{F}$-adapted, latent factor process $X(t)$; that is, there exist $\rho_0^{(x)} \in \mathbb{R}$ and $\rho_1^{(x)} \in \mathbb{R}^M$, possibly dependent on the base age x, such that

$$\mu_x(t) = \rho_0^{(x)} + (\rho_1^{(x)})' X(t), \qquad t \geq 0.$$

The process $X(t)$ is assumed to be a solution of the (vector) stochastic differential equation

$$dX(t) = \Delta(\theta^Q - X(t))dt + \Sigma D(X(t), t) dW^Q(t), \qquad X(0) = x_0 \in \mathbb{R}^M, \tag{3.1}$$

where $\Delta \in \mathbb{R}^{M\times M}$, $\theta^Q \in \mathbb{R}^M$, and $\Sigma \in \mathbb{R}^{M\times M}$. $D(X(t), t)$ is an M-dimensional diagonal matrix with diagonal elements $d_{ii}(X(t), t)$ is given by

$$d_{ii}(X(t), t) = \sqrt{\alpha^i(t) + \beta_1^i(t) X_1(t) + \cdots + \beta_M^i(t) X_M(t)}, \qquad i = 1, \ldots, M,$$

where α^i and $\beta^i := (\beta_1^i, \ldots, \beta_M^i)'$ are bounded and continuous functions, and W^Q is a standard M-dimensional Q-Brownian motion. The quantities Δ, θ^Q, and Σ represent the rate of mean reversion, the long-run mean, and the volatility of $X(t)$ respectively.

Let $S_x(t, T) := E^Q[\exp\{-\int_t^T \mu_x(s) ds\} | \mathcal{F}_t]$ denote the (risk-neutral) probability that an individual aged x at time t, conditional on being alive at time t, survives up to time T. Hence, following the affine framework set in Duffie & Kan (1996) and Duffie et al. (2000), $S_x(t, T)$ is an exponentially affine function of $X(t)$:

$$S_x(t, T) = \exp\{A_x(t, T) + B_x(t, T)' X(t)\},$$

where A_x and B_x are solutions of the system of ODEs

$$\begin{aligned} \frac{dB_x(t, T)}{dt} &= \rho_1^{(x)} + \Delta' B_x(t, T) - \frac{1}{2} \sum_{k=1}^{M} [\Sigma' B_x(t, T) B_x(t, T)' \Sigma]_{k,k} (\beta^k(t))' \\ \frac{dA_x(t, T)}{dt} &= \rho_0^{(x)} + B_x(t, T)' \Delta \theta^Q - \frac{1}{2} \sum_{k=1}^{M} [\Sigma' B_x(t, T) B_x(t, T)' \Sigma]_{k,k} \alpha^k(t) \end{aligned} \tag{3.2}$$

with terminal condition $A(T,T) = 0$ and $B(T,T) = \mathbf{0}$.

Therefore, the average force of mortality over the period $[t,T]$, defined as $\bar{\mu}_x(t,T) := -\frac{1}{T-t}\log S_x(t,T)$ is an affine function of the latent state process $X(t)$, i.e.

$$\bar{\mu}_x(t,T) = -\frac{A_x(t,T)}{T-t} - \frac{B_x(t,T)'}{T-t}X(t).$$

In correspondence to equation (2.2), $\bar{\mu}_x(t,T)$ represents the average force of mortality for an individual aged x at time t from ages x to $x + (T-t)$.

The affine representation of the average force of mortality allows us to cast the parameter estimation problem into that for a state-space model where $\bar{\mu}_x(t,T)$ are the observations and $X(t)$ is the unknown state process.

In the following sections, we discuss the affine mortality models we implement in `AffineMortality`. Specifically, we state the corresponding SDE for $X(t)$ and the functional form of $\mu_x(t)$. For each model, the factor loadings $A_x(t,T)$ and $B_x(t,T)$ are available in closed form and are functions of only $T-t$, the survival time from time t to time T. For all models, except the Gompertz-Makeham law, the mortality intensity is not dependent on x, hence we drop it from the notation.

3.1. Blackburn-Sherris model

The Blackburn & Sherris (BS) model assumes that

$$\mu(t) = X_1(t) + \cdots + X_M(t),$$

i.e. $\rho_0 = 0$ and $\rho_1 = (1,\ldots,1)'$, where $X(t) = (X_1(t),\ldots,X_M(t))'$ with dynamics

$$dX(t) = -\Delta X(t)dt + \Sigma dW^Q(t). \tag{3.3}$$

The components of $X(t)$ can be assumed to be independent by specifying Δ and Σ as diagonal matrices, i.e. $\Delta = \operatorname{diag}(\delta_1,\ldots,\delta_M)$ and $\Sigma = \operatorname{diag}(\sigma_1,\ldots,\sigma_M)$. Dependence among the components of $X(t)$ can be induced by setting

$$\Delta = \begin{pmatrix} \delta_1 & 0 & 0 & \ldots & 0 \\ \delta_{12} & \delta_2 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \delta_{1M} & \delta_{2M} & \delta_{3M} & \ldots & \delta_M \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 & \ldots & 0 \\ \sigma_{12} & \sigma_2 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{1M} & \sigma_{2M} & \sigma_{3M} & \ldots & \sigma_M \end{pmatrix}.$$

In `AffineMortality`, the user can specify as many latent factors as desired in the independent factor case. The factor loading expressions for the independent factor case can be found in Blackburn & Sherris (2013). However, while the factor loadings are still available in closed form in the dependent factor case, one cannot obtain a scalable nested expression for the factor loadings. `AffineMortality` currently supports the two-factor and three-factor Blackburn-Sherris model with dependent factors. The factor loading expressions for the dependent case can be found in Huang et al. (2022, Appendix A).

3.2. Gompertz-Makeham model

The Gompertz-Makeham model (see Schrager 2006) introduces age-dependence in the mortality intensity process. Specifically, the model assumes that $X(t)$ has the same dynamics outlined in equation (3.3), and

$$\mu_x(t) = X_1(t) + e^{\gamma x} X_2(t)$$

for some $\gamma > 0$. The functional specification of $\mu_x(t)$ implies that the mortality intensity increases exponentially in the base age x and it is $X_2(t)$ which drives the stochasticity in the intensity process at older ages. In this case, we have $\rho_0^{(x)} = \rho_0 = 0$ and $\rho_1^{(x)} = (1, e^{\gamma x})'$. As before, dependence between the latent factors can be introduced by replacing Δ and Σ by lower-triangular matrices. The factor loading expressions for both the independent and dependent factor cases can be found in Appendix B.1.

3.3. Arbitrage-Free Nelson-Siegel model

The Arbitrage-Free Nelson-Siegel model (AFNS), proposed by Christensen et al. (2011) for modelling the term structure of interest rates, assumes that there are three latent factors, identified as Level (L), Slope (S) and Curvature (C), with risk-neutral dynamics:

$$\begin{pmatrix} dL(t) \\ dS(t) \\ dC(t) \end{pmatrix} = -\begin{pmatrix} 0 & 0 & 0 \\ 0 & \delta & -\delta \\ 0 & 0 & \delta \end{pmatrix} \begin{pmatrix} L(t) \\ S(t) \\ C(t) \end{pmatrix} dt + \begin{pmatrix} \sigma_L & 0 & 0 \\ 0 & \sigma_S & 0 \\ 0 & 0 & \sigma_C \end{pmatrix} \begin{pmatrix} dW_L^Q(t) \\ dW_S^Q(t) \\ dW_C^Q(t) \end{pmatrix}. \tag{3.4}$$

The AFNS model assumes that the mortality intensity is the sum of the level and slope factors,

$$\mu(t) = L(t) + S(t).$$

The structure of Δ implies that the factor loadings B_L, B_S, and B_C control for the shape (i.e. level, slope, and curvature, respectively) of the average force of mortality, with randomness driven by the dynamics of L, S, and C, see Christensen et al. (2011, Proposition 1 and Section 2.3) for the closed-form expressions for the factor loadings.

A key feature of the AFNS model is the connection between the latent factors and the shape of the average force of mortality curve through the structure of Δ. We can induce factor dependence through the diffusion matrix Σ. Specifically, we replace Σ by a lower triangular matrix in the dependent factor case. As such, B_L, B_S, and B_C remain the same in the dependent factor case; see Christensen et al. (2011, Appendix B) for the formula of $A(t, T)$.

3.3.1. Arbitrage-Free Generalized Nelson-Siegel model

The Arbitrage-Free Generalized Nelson-Siegel (AFGNS) model is an extension of the AFNS model proposed by Christensen et al. (2009) which includes an additional slope and curvature factor. The AFGNS model was proposed as an arbitrage-free version of the four-factor Nelson-Siegel-Svensson model, which extends the AFNS model by adding a second curvature factor. The AFGNS model is thus a five-factor model whose latent factors, in the independent factor case, satisfy the SDE

$$\begin{pmatrix} \mathrm{d}L\,(t) \\ \mathrm{d}S_1\,(t) \\ \mathrm{d}S_2\,(t) \\ \mathrm{d}C_1\,(t) \\ \mathrm{d}C_2\,(t) \end{pmatrix} = - \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \delta_1 & 0 & -\delta_1 & 0 \\ 0 & 0 & \delta_2 & 0 & -\delta_2 \\ 0 & 0 & 0 & \delta_1 & 0 \\ 0 & 0 & 0 & 0 & \delta_2 \end{pmatrix} \begin{pmatrix} L\,(t) \\ S_1\,(t) \\ S_2\,(t) \\ C_1\,(t) \\ C_2\,(t) \end{pmatrix} \mathrm{d}t + \begin{pmatrix} \sigma_L & 0 & 0 & 0 & 0 \\ 0 & \sigma_{S_1} & 0 & 0 & 0 \\ 0 & 0 & \sigma_{S_2} & 0 & 0 \\ 0 & 0 & 0 & \sigma_{C_1} & 0 \\ 0 & 0 & 0 & 0 & \sigma_{C_2} \end{pmatrix} \begin{pmatrix} \mathrm{d}W_L^Q\,(t) \\ \mathrm{d}W_{S_1}^Q\,(t) \\ \mathrm{d}W_{S_2}^Q\,(t) \\ \mathrm{d}W_{C_1}^Q\,(t) \\ \mathrm{d}W_{C_2}^Q\,(t) \end{pmatrix},$$

where $\delta_1 \neq \delta_2$. As in the AFNS model, the dependent factor case consists of replacing a diagonal diffusion matrix with a lower triangular one. As in the AFNS model, under the AFGNS model, the mortality intensity is modelled as the sum of the level and the two slope factors,

$$\mu(t) = L(t) + S_1(t) + S_2(t).$$

Factor loading expressions for the independent factor case can be found in Christensen et al. (2009, Proposition 3.1). In the dependent factor case, only the form of $A(t, T)$ changes; this can be found in Christensen et al. (2009, Appendix).

3.3.2. Arbitrage-Free Reduced Nelson-Siegel model (AFRNS)

We introduce a version of the AFNS model without the curvature factor and call it the Arbitrage-Free Reduced Nelson-Siegel (AFRNS) model. This model has been included in the package in order to assess the effect of the curvature on the resulting model. Experiments we conducted on the mortality data of several countries showed that the presence of the curvature factor may produce negative mortality rates in some cases when projected 25 years ahead, despite the better in-sample performance of the AFNS and AFGNS models.

We consider two latent factors, with dynamics

$$\begin{pmatrix} dL(t) \\ dS(t) \end{pmatrix} = - \begin{pmatrix} 0 & 0 \\ 0 & \delta \end{pmatrix} \begin{pmatrix} L(t) \\ S(t) \end{pmatrix} dt + \begin{pmatrix} \sigma_L & 0 \\ 0 & \sigma_S \end{pmatrix} \begin{pmatrix} dW_L^Q(t) \\ dW_S^Q(t) \end{pmatrix},$$

for the independent factor case. As before, we replace the volatility matrix with a lower triangular matrix in the dependent factor case. The resulting factor loadings for both the independent and dependent factor cases can be found in Appendix B.2.

3.3.3. Arbitrage-Free Unrestricted Nelson-Siegel model

We also introduce a variation of the AFNS model where the elements of the drift coefficient matrix Δ have possibly unequal values. We call this model the Arbitrage-Free Unrestricted Nelson-Siegel (AFUNS) model. The latent factor dynamics under the AFUNS model are given by

$$\begin{pmatrix} \mathrm{d}L(t) \\ \mathrm{d}S(t) \\ \mathrm{d}C(t) \end{pmatrix} = - \begin{pmatrix} 0 & 0 & 0 \\ 0 & \delta_1 & \delta_2 \\ 0 & 0 & \delta_3 \end{pmatrix} \begin{pmatrix} L(t) \\ S(t) \\ C(t) \end{pmatrix} \mathrm{d}t + \begin{pmatrix} \sigma_L & 0 & 0 \\ 0 & \sigma_S & 0 \\ 0 & 0 & \sigma_C \end{pmatrix} \begin{pmatrix} \mathrm{d}W_L^Q(t) \\ \mathrm{d}W_S^Q(t) \\ \mathrm{d}W_C^Q(t) \end{pmatrix}. \qquad (3.5)$$

As before, the mortality intensity is modelled as the sum of the level and slope factors,

$$\mu(t) = L(t) + S(t).$$

We recover the AFNS model by setting $\delta_2 = -\delta_1$ and in the limit as $\delta_3 \to \delta_1$. The dependent factor case is obtained by replacing the volatility matrix by a lower triangular matrix.

3.4. Cox-Ingersoll-Ross Model

Under the CIR model, the mortality intensity is modelled as the sum of the components of the latent factor process $X(t)$,

$$\mu(t) = X_1(t) + \cdots + X_M(t),$$

where each $X_i(t)$ is a square-root diffusion process given by

$$dX_i(t) = \delta_i(\theta_i^Q - X_i(t))dt + \sigma_i\sqrt{X_i(t)}dW_i^Q(t).$$

This implies that each component of $X(t)$ is nonnegative Q-almost surely and is strictly positive Q-almost surely if $X_i(0) > 0$ and $2\delta_i\theta_i^Q \geq \sigma_i^2$. Each $X_i(t)$ is asymptotically Gamma distributed (as $t \to \infty$) (Cox et al. 1985), hence the CIR mortality model can capture the heterogeneity of mortality rates at older ages (Pitacco 2016). The factor loadings for the CIR model are given by

$$\begin{aligned}
B_i(t,T) &= -\frac{2(e^{\vartheta_i(T-t)} - 1)}{(\delta_i + \vartheta_i)(e^{\vartheta_i(T-t)} - 1) + 2\vartheta_i}, \qquad i = 1, 2, \ldots, M \\
A(t,T) &= \sum_{i=1}^{M} \frac{2\delta_i\theta_i^Q}{\sigma_i^2} \log\left[\frac{2\vartheta_i e^{\frac{1}{2}(\delta_i+\vartheta_i)(T-t)}}{(\delta_i + \vartheta_i)(e^{\vartheta_i(T-t)} - 1) + 2\vartheta_i}\right],
\end{aligned}$$

where $\vartheta_i = \sqrt{\delta_i^2 + 2\sigma_i^2}$. Their application in the analysis of HMD national mortality data can be found in Huang et al. (2022) and Ungolo et al. (2023).

3.5. Risk-free to real world dynamics

The mortality models illustrated so far were characterized under the risk-neutral measure Q. However, parameter estimation must be performed using historical data on mortality rates. We thus need to specify the dynamics of the latent factors under the real-world probability measure P.

To this end, we follow Duffee (2002) and adopt the *essentially affine* specification of the market price of longevity risk or the longevity risk premium. This allows us to preserve the affine structure of the latent factor dynamics when changing from Q to P and vice versa. By Girsanov's theorem, the latent factor dynamics under the P measure is given by the SDE

$$dX(t) = \boldsymbol{\kappa}(\theta^P - X(t))dt + \Sigma D(X(t),t)dW^P(t), \qquad X(0) = x_0,$$

where $\boldsymbol{\kappa} \in \mathbb{R}^{M \times M}$, $\theta^P \in \mathbb{R}^M$, and W^P is a standard M-dimensional Brownian motion under P. See Blackburn & Sherris (2013, Definition 5) and Huang et al. (2022, Section 2.3) for further details. We note that the diffusion term does not change when the change of measure is parameterized by the essentially affine specification.

Another advantage of the flexibility of the essentially affine specification of the market price of risk is that we are free to choose $\boldsymbol{\kappa}$ and θ^P while preserving any special structures in the Q measure. For example, this is especially relevant for the class of AFNS models for which the level, slope, and curvature interpretations of the latent factors are directly related to the specification of Δ in the Q-dynamics. Thus, we set $\theta^P = 0$ for all models except CIR, and assume that $\boldsymbol{\kappa}$ is a diagonal matrix $\boldsymbol{\kappa} = \mathrm{diag}(\kappa_1, \dots, \kappa_M)$ for all models. Further details on the application of the essentially affine specification in the context of affine mortality models can be found in Ungolo et al. (2023) and Huang et al. (2022).

4. Parameter estimation

The model parameters are estimated using data collected in discrete time, as illustrated in Section 2. By discretizing the stochastic differential equation of each model, and using the affine representation of the average force of mortality, we obtain the following equally time-spaced state-space formulation:

$$X(t) = \Phi_t X(t-j) + \eta_t \qquad \eta_t \sim N(0, R_t) \tag{4.1}$$

$$\bar{\mu}_t = A_t + B_t X(t) + \epsilon_t \qquad \epsilon_t \sim N(0, H) \qquad t = 1, \dots, T \tag{4.2}$$

where $A_t = [A(t, t+1), \dots, A(t, t+N)]'$ and $B_t = [B(t, t+1), \dots, B(t, t+N)]'$. The age subscript has been omitted for notational convenience.

The *state equation* (4.1) describes the dynamics of the factor as an autoregressive process of order 1 with system matrix $\Phi_t = e^{-\boldsymbol{\kappa} j}$ and stochastic noise $\eta_t \sim N(0, R_t)$. The *measurement equation* (4.2) describes $\bar{\mu}_t \in \mathbb{R}^N$ as an affine function of the latent variable $X(t)$ with error term $\epsilon_t \sim N(0, H)$. We assume that η_t and ϵ_t are independently distributed.

For the models so far implemented in `AffineMortality`, A, B, H and $\Phi = e^{-\boldsymbol{\kappa} j}$ do not depend on t. For Gaussian models, such as the Blackburn-Sherris and the AFNS models, we have $R_t = R$:

$$R = \left[I - e^{-\boldsymbol{\kappa} j}\right] \Sigma \Sigma' \left[I - e^{-\boldsymbol{\kappa} j}\right]'. \tag{4.3}$$

For the CIR model, because of the independence between the factors, R_t is a diagonal matrix with kth diagonal $r_{t,k}$ equal to:

$$r_{t,k} = \sigma^2 \left(\frac{1 - e^{-\kappa_k}}{\kappa_k} \right) \left(\frac{1}{2} \theta^P \left(1 - e^{-\kappa_k}\right) + e^{-\kappa_k} X(t) \right).$$

Here, Φ, A, B, R_t and H depend on the parameters that we estimate based on the statistical inference for the dynamics of the mortality rates. Furthermore, H is a diagonal matrix, where the diagonal elements ω_i^2 are equal to

$$\omega_i^2 = r_c + r_1 \sum_{k=1}^{i} \exp\left(r_2 k\right)/i \tag{4.4}$$

for $i = 1, \dots, N$. In this way, the measurement equation (4.2) accounts for the increasing variation in the mortality rates at older ages.

The parameters are estimated using maximum likelihood. Let ψ denote the vector of parameters to be estimated. The likelihood function is readily obtained from the univariate Kalman Filter recursion (see Koopman & Durbin (2000) and Ungolo et al. (2023) for its implementation in the context of affine mortality models), given the observed average mortality rates $\bar{\mu}_{1:T}$:

$$\log L\left(\psi \mid \bar{\mu}_{1:T}\right) = -\frac{TN}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\log F_{t,i} + \nu_{t,i}^2 F_{t,i}^{-1}\right), \tag{4.5}$$

where $\nu_{t,i} = \bar{\mu}_{t,i} - a_i - b_i \widehat{x}_{t,i}$ is the measurement error, and $F_{t,i} = b_i \widehat{\Sigma}_{t,i} b_i' + \omega_i^2$ the covariance of $\bar{\mu}_{t,i}$ (denoting $\mu_x\left(t, t+i\right)$) taking into account the uncertainty about the latent state $X\left(t\right)$. Here, a_i denotes the ith element of A_t, b_i the corresponding row of the matrix B_t, and $\widehat{x}_{t,i}$ and $\widehat{\Sigma}_{t,i}$ are the univariate Kalman Filter updates of the moments of $X\left(t\right)$ (see Appendix A for additional details).

For the CIR mortality model, the estimated parameter vector $\widehat{\psi}$ corresponds to the quasi-maximum likelihood estimator. See Chen & Scott (2003) and Jevtić & Regis (2021) for additional details.

4.1. Implementation

The parameter estimation task includes the initial state variable $X\left(0\right)$ among the set of unknown parameters. Other numerical tricks to foster reasonable parameter estimates, e.g. ensuring the positive-definiteness of the covariance matrix, are described in Ungolo et al. (2023).

We recommend the use of multiple starting values due to the high non-linearity of the log-likelihood function, which may have multiple local maxima. Some initial values are provided in `AffineMortality` though the list object `sv_default`, which we briefly illustrate in Section 8. These are based on previous analysis on country mortality rates. When fitting dependent factor models we recommend the use as starting values of the parameter estimates obtained from the correspondent independent factor models.

4.2. `affine_fit()`

The function `affine_fit()` of the package `AffineMortality` allows to carry out the parameter estimation task. The log-likelihood function of equation (4.5) is optimized sequentially by group of parameters (Coordinate Ascent) by using the gradient-free simplex Nelder-Mead method as recommended by Christensen et al. (2011). This routine is readily available in `R` within the function `optim`.

The function `affine_fit()` takes the following input:

- `model = c("BS", "AFNS", "AFGNS", "AFUNS", "AFRNS", "CIR", "GMk")`, to select one of the model family to be fitted. Its default value is `BS`;
- `fact_dep=c(FALSE, TRUE)`, to select whether the model accounts for factor dependence (default `FALSE`);
- `n_factors`, to select the number of factors (only for the BS and the CIR models; default set to 3);
- `data`, the rectangular data set of $\bar{\mu}$ rates used for the analysis;
- `st_val`, corresponding to the set of starting values for the parameters. These must be supplied as a list of parameters, e.g.

  ```
  st_val=list(x0=c(6.960591e-03, 9.017154e-03, 5.091784e-03),
      delta=c(0.04268782, -0.03122758, -0.08573677),
      kappa=c(1.162624e-02, 6.787268e-02, 5.061539e-03),
      sigma=exp(c(-6.806310, -6.790270, -7.559145)),
      r1=exp(-3.327060e+01), r2=exp(-6.086479e-01), rc=exp(-1.553156e+01))
  ```

  for the Blackburn-Sherris model with three independent factors. For dependent factor models we instead supply `sigma_dg`, which is the parameter denoting the standard deviation, and `Sigma_cov` indicating the elements of the off-diagonal elements of the covariance matrix (generally a vector of zero, as suggested in Section 4.1);
- `max_iter`: maximum number of iterations for the coordinate ascent algorithm (default 200);
- `tolerance`: maximum log-likelihood value increase between iterations such that the optimizer can stop (default 0.1);
- `wd`: working directory to save the intermediate values of the parameters throughout iterations;

This function returns a list with

- `model`: same as input;
- `fit`: list with:
  - `par_est`: list of parameter estimates;
  - `log_lik`: value of the log-likelihood function;
  - `CA_par`: Table listing the value of the parameters throughout the coordinate ascent algorithm iterations;
- `n.parameters`: total number of estimated parameters;

- `AIC`: value of Akaike Information Criterion (see Section XX - Goodness of fit);
- `BIC`: value of Bayesian Information Criterion;

`affine_fit()` can be run as follows:

```
affine_fit(model="BS", fact_dep=FALSE, n_factors=3, data=mu_bar_USA,
st_val=st_val, max_iter=200, tolerance=0.1)
```

During its execution the console shows the value of the parameters, the log-likelihood function and the iteration number:

```
[1] "X(0)_1  0.005" "X(0)_2 0.006" "X(0)_3 0.004"
[1] "delta_1 0.053"  "delta_2 -0.018" "delta_3 -0.087"
[1] "kappa_1 0.032"  "kappa_2 0.007"  "kappa_3 -0.003"
[1] "sigma_1 0.001" "sigma_2 0.001" "sigma_3 0"
[1] "r1 0"     "r2 0.548" "rc 0"
[1] "log_lik 10538.37"
[1] "1 % iteration"
[1] "--------------------------------"
[1] "X(0)_1 0"     "X(0)_2 0.007" "X(0)_3 0.007"
[1] "delta_1 0.043"  "delta_2 -0.022" "delta_3 -0.086"
[1] "kappa_1 0.017" "kappa_2 0.002" "kappa_3 0.01"
[1] "sigma_1 0.001" "sigma_2 0.001" "sigma_3 0"
[1] "r1 0"     "r2 0.556" "rc 0"
[1] "log_lik 10623.92"
[1] "2 % iteration"
```

5. Goodness of fit

The fitted rates, denoted as $\widehat{\bar{\mu}}_t$ (which are N-dimensional vectors, for $t = 1, \ldots, K$), of use for the analysis of the goodness of fit of each model with respect to historical data, can be obtained using the function `mubar_hat()`, which takes the following as input: `model`, `fact_dep`, `n_factors`, `parameters`, `data`. Again, the input `parameters` is supplied as a list, similar to the starting values of `affine_fit()`. The resulting fitted rates can then be used to analyze the goodness-of-fit of each model with respect to historical data.

5.1. Numerical measures

`AffineMortality` considers four goodness-of-fit measures, following Blackburn & Sherris (2013) and Huang et al. (2022):

- Akaike Information Criterion (AIC, Akaike (1974)):

$$\text{AIC} = -2 \log L\left(\widehat{\psi} \mid \bar{\mu}_{1:K}\right) + 2k \tag{5.1}$$

- Bayesian Information Criterion (BIC, Schwarz (1978)):

$$\text{BIC} = -2 \log L\left(\widehat{\psi} \mid \bar{\mu}_{1:K}\right) + 2kKN \tag{5.2}$$

- Root Mean Squared Error (RMSE):

$$\text{RMSE} = \frac{1}{KN}\sum_x \sum_t \left(\bar{\mu}_{x,t} - \widehat{\bar{\mu}}_{x,t}\right)^2 \tag{5.3}$$

- Mean Absolute Percentage Error (MAPE, by age x):

$$\text{MAPE}_x = \frac{1}{K}\sum_{t=1}^{K} \frac{|\bar{\mu}_{x,t} - \widehat{\bar{\mu}}_{x,t}|}{\bar{\mu}_{x,t}} \tag{5.4}$$

where k is the number of parameters, $t = 1, \dots, K$ and N is the number of ages considered in the analysis.

The AIC and BIC can be obtained as the output of `affine_fit()` (see Section 4.2). The RMSE is obtained by running `RMSE(fitted, observed)`, where `fitted` is the $N \times K$-dimensional matrix of the fitted rates obtained using the function `mubar_hat()`, and `observed` is the corresponding matrix of the $\bar{\mu}$-rates used for parameter estimation. The function `MAPE_row(fitted, observed)` yields an N-dimensional vector, which can be used to assess how the model fits at every age in the range of interest.

A desirable characteristic of affine mortality models is that their parameters ensure that the probability of negative rates is negligible. This is a potential limitation of Gaussian models, since $X(t)$ can assume any real value.

At this purpose, the function `prob_neg_mu()` yields an N-dimensional vector with the probability of negative mortality rates at each age (based on the input data set) for a specific h-year ahead projection, based on the simulated values of $X(t)$. `prob_neg_mu()` takes as input `model`, `fact_dep`, `n_factors`, `parameters`, `data`, `years_proj`, `n_simulations` (default value set to 100,000).

5.2. Residuals

The function `std_res()` returns an $N \times K$-dimensional matrix of the standardized residuals for the model of interest. These are computed as N-dimensional vectors for each year $t = 1, \dots, K$ as:

$$r_t = \left(\sqrt{\widehat{\mathbb{V}\left(\bar{\mu}_t\right)}}\right)^{-1} \left(\bar{\mu}_t - \widehat{\bar{\mu}}_t\right), \tag{5.5}$$

Further details about this formula can be found in Ungolo et al. (2023). A heatmap of the standardized residuals by age and year (Figure 5.1) can be generated by using the function `heatmap_res()`:

```
std_resid <- std_res(model="BS", fact_dep=TRUE, n_factors=3,
             parameters=par_list, data=mu_bar)
heatmap_res(residuals=std_resid, color=FALSE)
```

In this way, we can visually detect the presence of period effects (if analysing age-cohort data) or of cohort effects (if analysing age-period data).

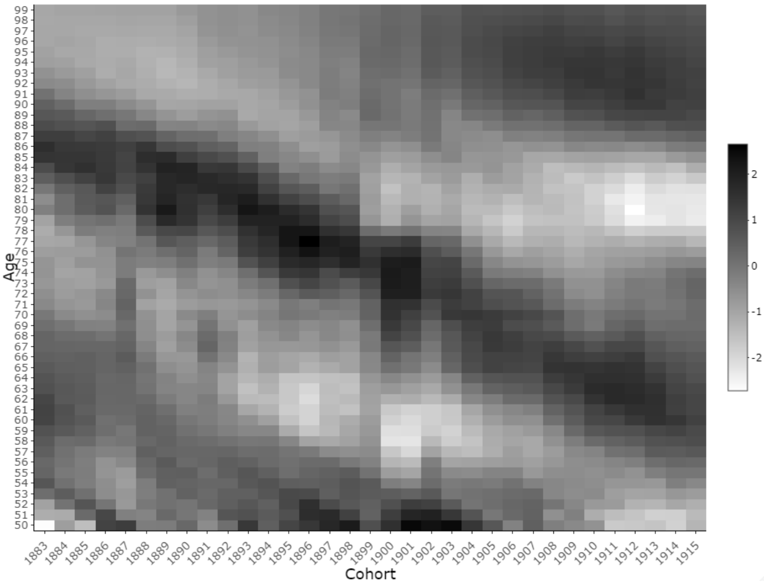

Figure 5.1: Heatmap of the standardized residuals for the Blackburn-Sherris model with three independent factors. Source: Ungolo et al. (2023)

6. Projection

The function `affine_project()` returns the projected survival curve for h time periods ahead (cohort or calendar year, depending on the analysed data set), based on the optimal forecast of the average force of mortality under the quadratic loss (Christensen et al. (2011)):

$$\bar{\mu}_x\left(t+h, T+h\right) = -\frac{B_x\left(t,T\right)}{T-t}\mathbb{E}\left[X\left(t+h\right) \mid X\left(t\right)\right] - \frac{A_x\left(t,T\right)}{T-t}, \qquad (6.1)$$

The corresponding survival probability is given by

$$S_x\left(t+h, T+h\right) = \exp\left(B_x\left(t,T\right)' \mathbb{E}\left[X\left(t+h\right) \mid X\left(t\right)\right] + A_x\left(t,T\right)\right). \qquad (6.2)$$

The function `affine_project()` is illustrated in Section 8 when analysing the Blackburn-Sherris model with three dependent factors for the US dataset. Its input structure is similar to `affine_fit()`, with the additional argument `years_proj` corresponding to the h time periods ahead for the projection.

7. Parameter uncertainty

A further source of risk when projecting cohort survival curves is the uncertainty inherited from the parameter estimation process.

The function `par_cov()`, returns the variance-covariance matrix of the parameter estimates and their corresponding standard errors. It allows the user to choose between two methods for estimating parameter uncertainty, namely multiple imputation and bootstrap.

The first method, described in Ungolo et al. (2023), can be chosen by setting `method="MI"` within `par_cov()`. It is recommended for Gaussian affine mortality models. Briefly, it consists of a procedure which randomly imputes a value of the latent state variable $X(t)$ sampled from the smoothing distribution at the value of the parameter estimates. In this way, we obtain a set of "completed" data sets such that parameters are then re-estimated. The number of completed data sets can be specified through the argument `D_se`. This method turns out useful, because on one hand, it may not be possible to numerically compute the Information matrix from the optimization process of the likelihood function due to its very flat surface. On the other hand, the alternative bootstrap method (briefly described later) may be computationally expensive if carried out hundreds of times, as recommended in practice. The downside of multiple imputation, is that unlike the bootstrap, it may tend to underestimate the standard errors, since it is a delta method. From a computational perspective, a potential downside is the need to invert a Hessian matrix of larger dimensions, although this task is simpler than the inversion of the Hessian matrix from the estimation procedure (whose likelihood is marginalized with respect to the latent states). This method is not recommended, nor implemented for the CIR model, due to the lower truncation of the latent variable $X(t)$.

When parameter uncertainty is assessed by multiple imputation, the function `par_cov` returns a list with two elements: `Cov_par_est`, the variance-covariance matrix of the parameters and `St_err` which is a list of the standard error of the parameters.

The bootstrap method draws on the work of Stoffer & Wall (2009), and was used by Blackburn & Sherris (2013) in the context of affine mortality models. It can be implemented in `AffineMortality` by specifying the argument `method="BS"` in the function `par_cov()`. In few words, this method consists of an iterative procedure which first computes the standardized innovations from the measurement error, in order to obtain a bootstrapped dataset of average mortality rates. These are used to obtain a new set of parameter estimates. Hence, once `n_BS` parameter estimates are obtained, their variance-covariance is computed. In this case, the function `par_cov()` provides an additional element given by the parameter estimates over the bootstrapped samples. The argument `t_excl` (default value set to 4 (Stoffer & Wall 2009)) sets the number of the oldest residuals in terms of t to adjust for the Kalman Filter startup irregularities.

Both multiple imputation and bootstrap algorithms are initialized at the parameter estimates in order to better explore the likelihood surface in the neighborhood $\widehat{\psi}$. The use of both methods is illustrated in Section 8.

We recommend to use the bootstrap for models with a large number of parameters, such as the AFGNS and for the CIR model with any number of factors.

8. Illustration

Data preparation

`AffineMortality` provides the data set of the average mortality rates for the US males aged 50-99 born in the years 1883-1915 analysed in Ungolo et al. (2023), which can be loaded as follows:

```
data(mu_bar)
```

This dataset will be used for illustrating the analysis of the Blackburn-Sherris model with three dependent factors.

If we want for example we want to analyse the data from the Human Mortality Database, we can use the package **HMDHFDplus** (Riffe 2015) to source the death counts and the exposure-at-risk years for the computation of the μ and $\bar{\mu}$ rates for the set of ages and periods/cohorts of interest:

```
deaths <- readHMDweb(CNTRY = "DNK", item = "Deaths_1x1", username,
            password, fixup = TRUE)
exposures <- readHMDweb(CNTRY = "DNK", item = "Exposures_1x1", username,
               password, fixup = TRUE)
```

Deaths and exposure-at-risk years data are then used to calculate the $\mu_{x,t}$ rates, which are then transformed into the $\bar{\mu}$-rates matrix using the function `rates2avg()` (see Section 2).

```
mu_bar <- rates2avg(mu_xt_matrix)
```

Appendix C shows an example of R code which can be used to obtain the matrix of $\mu_{x,t}$-rates `mu_xt_matrix`.

Parameter estimation

A critical aspect of the analysis of affine mortality models is the specification of the starting values for the algorithm. As emphasized in Ungolo et al. (2023) and Blackburn & Sherris (2013) the likelihood function can have multiple local optima, hence the fitting algorithm should be initialised several times. In `AffineMortality` we provide a set of starting values (`sv_default` in an R list format), which can be used by the researcher for a first exploration of the models.

For example, to get the default starting values for the Blackburn-Sherris model with dependent factors, then we run the following code:

```
starting_values <- sv_default$BSd
```

For the other supported models we can use `sv_default$`

- `BSi` for the Blackburn-Sherris model up to four factors;
- `AFNSi` and `AFNSd` for the AFNS model with independent and dependent factor respectively;

- `GMki` and `GMkd` for the Gompertz-Makeham model with independent and dependent factor respectively;
- `AFGNSi` and `AFGNSd` for the AFGNS model with independent and dependent factor respectively. Similar for the AFRNS and AFUNS models;
- `CIR` for the Cox-Ingersoll-Ross model up to four factors;

As highlighted in Section 4.1, a general recommendation when analysing models with dependent factors is to initialize `affine_fit()` with the parameter estimates of the corresponding independent factor models, and to set the starting values of the additional parameters (such as the off-diagonal elements of the covariance matrix of $X(t)$ and of the mean reversion matrix Δ) to zero.

We can thus estimate the parameters with `affine_fit()` as follows:

```
pe_BSd_3F <- affine_fit(model="BS", fact_dep=TRUE, n_factors=3,
data=mu_bar, st_val=starting_values, max_iter=5, tolerance=0.1,
wd="working_folder_directory")
```

In practice, we run the fitting process for a larger number of iterations. For example, in Ungolo et al. (2023) the authors set `max_iter=200`.

Goodness of fit analysis

Once we obtain the object `pe_BSd_3F`, we run the following lines to obtain the AIC and the BIC

```
> pe_BSd_3F$AIC
-21405.58

> pe_BSd_3F$BIC
-21292
```

This output can be used to compare between different models. Suppose we want to compare it with the AFNS model with independent factor, then we run the following lines:

```
pe_AFNSi <- affine_fit(model="AFNS", fact_dep=FALSE, st_val=sv_default$AFNSi,
data=mu_bar, max_iter=5, tolerance=0.1)

> pe_AFNSi$AIC
-20842.76

> pe_AFNSi$BIC
-20772.45
```

Since the Blackburn-Sherris model with dependent factors has a smaller value of both AIC and BIC compared to the AFNS model with independent factors, then we conclude that the former shows a better in-sample fit.

The fitted average mortality rates can be then obtained by using `mubar_hat()`

```
fitted_BSd <- mubar_hat(model="BS", fact_dep=TRUE, n_factors=3,
              parameters=pe_BSd_3F$fit$par_est, data=mu_bar)
```

The fitted average mortality rates obtained using `mubar_hat()` can be used to calculate the RMSE and the MAPE for each age in the range of interest:

```
> RMSE(mu_bar, fitted_BSd)
[1] 0.002011956

MAPE_age(mu_bar, fitted_BSd)
```

which can be similarly used for comparing the models.

Furthermore, we can obtain the fitted μ-rates by using the function `avg2rates()` described in Section 2:

```
avg2rates(fitted_BSd)
```

In order to ensure that the probability of negative rates for the projected 10-years ahead cohort is negligible for the model under analysis, we run the function `prob_neg_mu()`

```
> prob_neg_mu(model="BS", fact_dep=TRUE, n_factors=3,
parameters=pe_BSd_3F$fit$par_est, data=mu_bar, years_proj = 10,
n_simulations=1000)

 [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

This shows vector of zeros indicating that the fitted model has zero probability of yielding negative rates for the cohort born in 1925.

Projection

The vector of the 1-year ahead (hence for the US males born in 1916) survival rates for the age-range of interest can be obtained by using the `affine_project()` function and then plotted as follows:

```
BSd_3F_proj <- affine_project(model="BS", fact_dep=TRUE, n_factors=3,
parameters=pe_BSd_3F$fit$par_est, data=mu_bar, years_proj = 1)

plot(rownames(mu_bar), BSd_3F_proj, type="l", ylab = "S(t)", xlab = "Age")
```

Parameter uncertainty

The estimation of the parameter uncertainty by multiple imputation through the function `par_cov()` described in Section 7 can be carried out as follows:

```
par_unc_MI <- par_cov(method="MI", model="BS", fact_dep=TRUE, n_factors=3,
parameters=pe_BSd_3F$fit$par_est, data=mu_bar, D_se=5, max_iter=10,
tolerance=0.1, wd=0)
```

We recommend to set the number of imputations `D_se` at least equal to 50 to obtain a more robust estimate of the parameter uncertainty. In this illustration, we have the following set of standard errors of the parameters as an R list object:

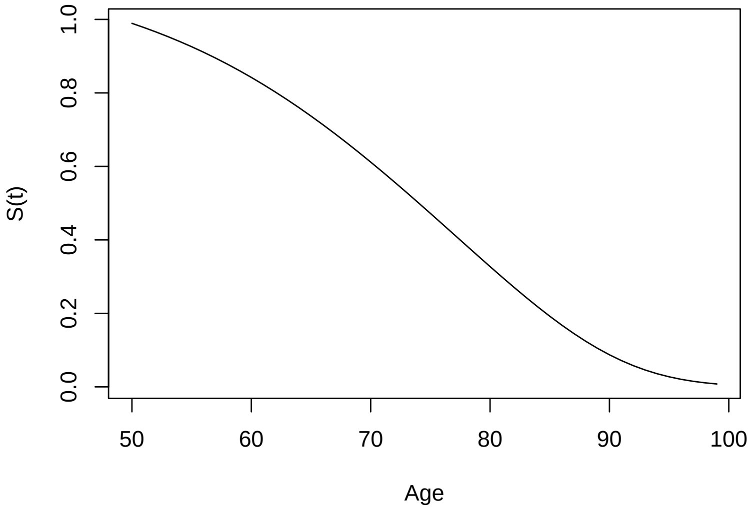

Figure 8.1: 1-year ahead projected survival curve for the Blackburn-Sherris model with three dependent factors

```
> par_unc_MI$St_err
$delta
[1] 0.0038258337 0.0738215285 0.0024289977 0.0529585515 0.0020585681 0.0004005845

$kappa
[1] 0.041057554 0.010218757 0.004028687

$sigma_dg
[1] 1.850683e-06 5.515050e-05 4.927546e-05

$Sigma_cov
[1] 0.000446849 0.015596337 0.005928911

$r1
[1] 8.016628e-16

$r2
[1] 0.005309632

$rc
[1] 1.330768e-09
```

Similarly, we can also estimate the parameter uncertainty by using the bootstrap method:

```
par_unc_Bts <- par_cov(method="Bootstrap", model="BS", fact_dep=TRUE,
n_factors=3, parameters=par_list, data=mu_bar, t_excl=4, BS_s=5, max_iter=3,
tolerance=10, wd=0)
```

When using the bootstrap in practice, Blackburn & Sherris (2013) set the number of bootstrap samples `BS_s` equal to 500, while the default value of the number of residuals to be excluded `t_excl=4` follows from Stoffer & Wall (2009). As for the function `affine_fit()`, the arguments `max_iter` and `tolerance` are usually set to 200 and 0.1 respectively, although the user can set any desired value.

9. Other functions

The function `xfilter` has the same input structure as `mubar_hat` and returns a list of conditional mean and covariance of the filtering distribution of the latent variable as obtained from the application of the univariate Kalman-Filter of Koopman & Durbin (2000). More precisely, the matrix `X_t` returns the value of $\mathbb{E}\left[X\left(t\right) \mid \bar{\mu}_{1:t}\right]$ (time-update step), for $t = 0, \ldots, K$ while the matrix `X_t_c` returns the value of $\mathbb{E}\left[X\left(t\right) \mid \bar{\mu}_{1:t-1}\right]$ (forecasting step). `S_t` and `S_t_c` are the corresponding covariance matrix of $X\left(t\right)$. For a brief illustration, this can be run as follows:

```
X_filtered <- xfilter(model="BS", fact_dep=TRUE, n_factors=3,
                parameters=pe_BSd_3F$fit$par_est, data=mu_bar)
```

The function `xsmooth` implements the Rauch-Tung-Striebel (Rauch et al. (1965)) smoothing procedure to obtain the conditional mean and covariance matrix of the distribution of $X\left(t\right)$ conditional to $\bar{\mu}_{1:T}$, that is, the entire time-series of the observations. It uses as input the results from the `xfilter` function and the value of the parameter **kappa** driving the dynamics of the SDE illustrated in Section 3 under the real-world probability measure.

```
X_smoothed <- xsmooth(filterobject=X_filtered,
                kappa=pe_BSd_3F$fit$par_est$kappa)
```

10. Conclusion and further developments

This paper describes the `AffineMortality` R package, which allows the user to estimate, compare, project and assess the uncertainty of affine mortality models. These models can be analysed from an age-period as well as from an age-cohort perspective. The package can be used to support researcher about a wide range of questions involving stochastic mortality, including pricing of mortality contingent securities (Xu et al. (2020*a*) and Xu et al. (2020*b*)), risk management of mortality contingent products, assessment of the natural hedging of life insurance policies and life annuities (Blackburn et al. (2017) and Sherris et al. (2020)) as well as the design of innovative mortality pooling products.

The authors plan to further expand the range of models which can be fitted, such as the Squared Gaussian model used in interest rates modelling (Leippold & Wu (2002)), and incorporate other features, such as cohort specific factors and other age-dependent models. Further additions will encompass the possibility to account for incomplete cohort data and the inclusion of models whose mean-reversion parameter is non zero.

Another strand of future developments of `AffineMortality` include the possibility of using alternative optimization methods, such as the Subplex algorithm of Rowan (1990), which is available in R through the package `nloptr` (Johnson (2020)).

11. Acknowledgments

Francesco Ungolo acknowledges financial support from ERGO Center of Excellence in Insurance. Michael Sherris and Yuxin Zhou acknowledge financial support from the Society of Actuaries Center of Actuarial Excellence Research Grant 2017-2020: Longevity Risk: Actuarial and Predictive Models, Retirement Product Innovation, and Risk Management Strategies; and from the Australian Research Council Discovery Grant DP170102275 Retirement Product Innovation. MS, Len Patrick Garces, and YZ also acknowledge support provided by the Australian Research Council Centre of Excellence in Population Ageing Research (CEPAR) project number CE170100005.

The authors are grateful to Dr. Patrick J. Laub for his valuable help and insights in making the package available.

A. Univariate Kalman Filtering

Following the outline in Ungolo et al. (2023), the ith element of the vector $\bar{\mu}_t$ can be written as:

$$\bar{\mu}_{t,i} = a_i + b_i x_{t,i} + \epsilon_{t,i}, \quad \epsilon_{t,i} \sim N\left(0, \omega_i\right), \tag{A.1}$$

Hence, the state equation corresponding to each observation $\bar{\mu}_{t,i}$ is:

$$\begin{aligned} x_{t+1,1} &= \Phi x_{t,N} + \eta_t, \\ x_{t,i+1} &= x_{t,i} \end{aligned} \tag{A.2}$$

for $i = 1, \dots N-1$ and $t = 1, \dots, K$, given initial state $x_{0,N} = X(0)$. Let $\bar{\mu}_{1:t} = [\bar{\mu}_1, \dots, \bar{\mu}_t]$ and $\bar{\mu}_{t,1:i} = [\bar{\mu}_{t,1}, \dots, \bar{\mu}_{t,i}]$.

Given initial state $x_{0,N} := X(0)$ and initial conditional covariance $\Sigma_{0,N} = \text{diag}\left(10^{-10}, \dots, 10^{-10}\right)$:

1. Forecasting ($i = 1$ only):

$$\begin{aligned} \widehat{x}_{t,1} &= \mathbb{E}\left(x_{t,1} \mid \bar{\mu}_{1:t-1}\right) = \Phi \widehat{x}_{t-1,N}, \\ \widehat{\Sigma}_{t,1} &= \mathbb{V}\left(x_{t,1} \mid \bar{\mu}_{1:t-1}\right) = \Phi \widehat{\Sigma}_{t-1,N} \Phi' + R; \end{aligned} \tag{A.3}$$

2. Time-update ($i = 1, \dots, N-1$ on the left-hand side):

$$\begin{aligned} \widehat{x}_{t,i+1} &= \mathbb{E}\left(x_{t,i+1} \mid \bar{\mu}_{1:t-1}, \bar{\mu}_{t,1:i}\right) = \widehat{x}_{t,i} + K_{t,i}\nu_{t,i}, \\ \widehat{\Sigma}_{t,i+1} &= \mathbb{V}\left(x_{t,i+1} \mid \bar{\mu}_{1:t-1}, \bar{\mu}_{t,1:i}\right) = \widehat{\Sigma}_{t,i} - K_{t,i}F_{t,i}K'_{t,i} \\ &= \left(I - K_{t,i}b_i\right)\widehat{\Sigma}_{t,i}\left(I - K_{t,i}b_i\right)' + K_{t,i}\omega_i^2 K'_{t,i}, \end{aligned} \tag{A.4}$$

where the scalar quantities $\nu_{t,i}$ and $F_{t,i}$, and the $M \times 1$-dimensional vector $K_{t,i}$ are given by

$$\begin{aligned} \nu_{t,i} &= \bar{\mu}_{t,i} - a_i - b_i \widehat{x}_{t,i}, \\ F_{t,i} &= b_i \widehat{\Sigma}_{t,i} b_i' + \omega_i^2, \\ K_{t,i} &= \widehat{\Sigma}_{t,i} b_i' F_{t,i}^{-1}. \end{aligned} \tag{A.5}$$

B. Factor Loading Expressions

B.1. Gompertz-Makeham Model

In the independent-factor Gompertz-Makeham model, the factor loadings are given by

$$\begin{aligned} B_1(t,T) &= -\frac{1 - e^{-\delta_1 (T-t)}}{\delta_1} \\ B_{2,x}(t,T) &= -e^{\gamma x} \frac{e^{(\gamma - \delta_2)(T-t)} - 1}{\delta_2 - \gamma} \\ A_x(t,T) &= -\sigma_1^2 \left[\frac{T-t}{2\delta_1^2} - \frac{1 - e^{-\delta_1 (T-t)}}{\delta_1^3} + \frac{1 - e^{-2\delta_1 (T-t)}}{4\delta_1^3} \right] \\ &\quad - \frac{\sigma_2^2 e^{2\gamma x (T-t)}}{(\delta_2 - \gamma)^2} \left[\frac{1 - e^{2\gamma (T-t)}}{2\gamma} - 2\frac{1 - e^{(\gamma - \delta_2)(T-t)}}{\delta_2 + \gamma} + \frac{1 - e^{2\delta_2 (T-t)}}{2\delta_2} \right]. \end{aligned}$$

In the dependent factor case, we have

$$\begin{aligned}
B_1(t,T) &= (a_1+a_2)e^{-\delta_1(T-t)} - (a_1 + a_2 e^{-\delta_2(T-t)}) \\
B_2(t,T) &= -\frac{e^{\gamma x}}{\delta_2}(1-e^{-\delta_2}(t-T)) \\
A(t,T) &= -\frac{3}{2\delta_2}\left[\frac{\sigma_2^2 e^{2\gamma x}}{2\delta_2^2}+\frac{\sigma_{12}^2 e^{2\gamma x}}{2\delta_2^3}\right] \\
&\quad + (T-t)\left[\frac{\sigma_2^2 e^{2\gamma x}}{2\delta_2^2}+\frac{\sigma_{12}^2 e^{2\gamma x}}{2\delta_2^3}+\frac{a_1\sigma_1\sigma_{12}e^{\gamma x}}{2\delta_2}+\frac{a_1\sigma_1}{2}+\frac{a_1\sigma_1\sigma_{12}e^{\gamma x}}{4\delta_2}\right] \\
&\quad - (1-e^{-\delta_1(T-t)})\left[\frac{3\sigma_1\sigma_{12}e^{\gamma x}(a_1+a_2)}{4\delta_1\delta_2}+\frac{a_1\sigma_1^2(a_1+a_2)}{\delta_1}\right] \\
&\quad + (1-e^{-\delta_2(T-t)})\left[\frac{(a_1-1)\sigma_1\sigma_{12}e^{\gamma x}}{2\delta_2^2}+\frac{a_1a_2(\sigma_1^2+2)}{2\delta_2}-\frac{\sigma_1\sigma_{12}e^{\gamma x}(a_1-a_2)}{4\delta_2^2}\right] \\
&\quad + (1-e^{-2\delta_1(T-t)})\frac{\sigma_1^2(a_1+a_2)^2}{4\delta_1} \\
&\quad + (1-e^{-2\delta_2(T-t)})\left[\frac{a_2^2}{2\delta_2}-\frac{a_2\sigma_1\sigma_{12}e^{\gamma x}}{2\delta_2^2}-\frac{\sigma_1\sigma_{12}e^{\gamma x}}{4\delta_2^2}\right] \\
&\quad - (1-e^{-\delta_1(T-t)-\delta_2(T-t)})\left[\frac{2a_2^2+a_2^2\sigma_1^2+2a_1a_2+a_1a_2\sigma_1}{2(\delta_1+\delta_2)}+\frac{\sigma_1\sigma_{12}e^{\gamma x}(a_1+a_2)}{4\delta_2(\delta_1+\delta_2)}\right] \\
&\quad + e^{-2\delta_2(T-t)}(4e^{\delta_2(T-t)}-1)\left[\frac{\sigma_2^2 e^{2\gamma x}}{4\delta_2^3}+\frac{\sigma_{12}^2 e^{2\gamma x}}{4\delta_2^4}\right],
\end{aligned}$$

where, for convenience, we define

$$a_1 = \frac{1}{\delta_1} - \frac{\delta_{12}e^{\gamma x}}{\delta_1\delta_2}, \qquad a_2 = -\frac{\delta_{12}e^{\gamma x}}{\delta_2^2-\delta_1\delta_2}.$$

B.2. Arbitrage-Free Reduced Nelson-Siegel (AFRNS) Model

In the independent factor case, the factor loadings are given by

$$\begin{aligned}
B_L(t,T) &= -(T-t) \\
B_S(t,T) &= -\frac{1-e^{-\delta(T-t)}}{\delta} \\
A(t,T) &= \frac{\sigma_1^2}{6}(T-t)^3 + \sigma_2^2(T-t)\left[\frac{1}{2\delta^2}-\frac{1-e^{-\delta(T-t)}}{(T-t)\delta^3}+\frac{1}{4}\frac{1-e^{-2\delta(T-t)}}{\delta^3}\right]
\end{aligned}$$

On the other hand, in the dependent factor case, we have

$$
\begin{aligned}
B_L(t,T) &= -(T-t)\\
B_S(t,T) &= -\frac{1-e^{-\delta(T-t)}}{\delta}\\
A(t,T) &= \frac{1}{12\delta^3}\Big[6\sigma_L\sigma_{LS}(e^{-\delta(T-t)}-1)+6\delta\sigma_L\sigma_{LS}e^{-\delta(T-t)}(T-t)\\
&\qquad\quad +2\delta^3\sigma_L^2(T-t)^3+3\delta^3\sigma_L\sigma_{LS}(T-t)^2\Big]\\
&\quad -\frac{1}{4\delta^3}\Big[2\sigma_L\sigma_{LS}+3(\sigma_{LS}^2+\sigma_S^2)+\sigma_S^2e^{-2\delta(T-t)}-2(\sigma_L\sigma_{LS}-2(\sigma_{LS}^2+\sigma_S^2))e^{-\delta(T-t)}\\
&\qquad\quad -2\delta(\sigma_{LS}^2+\sigma_S^2)(T-t)-\delta^2\sigma_L\sigma_{LS}(T-t)^2-2\delta\sigma_L\sigma_{LS}e^{-\delta(T-t)}(T-t)\Big].
\end{aligned}
$$

B.3. Arbitrage-Free Unrestricted Nelson-Siegel (AFUNS) Model

In the independent factor case, we have the following expressions for the factor loadings

$$
\begin{aligned}
B_L(t,T) &= -(T-t)\\
B_S(t,T) &= -\frac{1-e^{-\delta_1(T-t)}}{\delta_1}\\
B_C(t,T) &= -\delta_2(T-t)\frac{1-e^{-\delta_1(T-t)}}{\delta_1}+\frac{1-e^{-\delta_3(T-t)}}{\delta_3(\delta_1-\delta_3)}\\
A(t,T) &= \frac{1}{6}\sigma_1^2(T-t)^3+\frac{\sigma_2^2}{2\delta_1^2}(T-t)-\frac{1-e^{-\delta_1(T-t)}}{\delta_1^3}+\frac{1-e^{-2\delta_1(T-t)}}{4\delta_1^3}\\
&\quad +\frac{T-t}{2}\left(\frac{\sigma_3\delta_3}{\delta_1-\delta_3}\right)^2\left\{\left[1-\frac{2(1-e^{-\delta_1(T-t)})}{\delta_1(T-t)}+\frac{1-e^{-2\delta_1(T-t)}}{2\delta_1(T-t)}\right]\frac{1}{\delta_1^2}\right.\\
&\qquad\qquad +\left[1-\frac{2(1-e^{-\delta_3(T-t)})}{\delta_3(T-t)}+\frac{1-e^{-2\delta_3(T-t)}}{\delta_3(T-t)}\right]\frac{1}{\delta_3^2}\\
&\qquad\qquad \left.+2\left[1-\frac{1-e^{-\delta_1(T-t)}}{\delta_1(T-t)}-\frac{1-e^{-\delta_3(T-t)}}{\delta_3(T-t)}+\frac{1-e^{-(\delta_1+\delta_3)(T-t)}}{(\delta_1+\delta_3)(T-t)}\right]\frac{1}{\delta_1\delta_3}\right\}
\end{aligned}
$$

Meanwhile, the expression for $A(t,T)$ in the dependent factor case are given by

$$\begin{aligned}
A(t,T) = {} & \frac{1}{6}\sigma_L^2(T-t)^3 + \frac{1}{2}\left[\frac{c_1}{\delta_1} + \frac{c_2c_3}{\delta_1} + \frac{c_2c_3}{\delta_3}\right](T-t)^2 \\
& + \left[\frac{c_4 + c_3c_5 + 2c_3}{2\delta_1^2} + \frac{c_3c_5 + 2c_3 + 2c_3c_6}{2\delta_1\delta_3} + \left(\frac{1}{\delta_1} + \frac{1}{\delta_3}\right)\left(\frac{c_3^2c_6 - c_3c_5}{2}\right)\right](T-t) \\
& + \left[\frac{c_1 + 2c_2c_3}{2\delta_1^3}\right]\left(e^{-\delta_1(T-t)}\delta_1(t-T) - (1 - e^{-\delta_1(T-t)})\right) \\
& + \frac{c_2c_3}{\delta_3^3}\left(e^{-\delta_3(T-t)}\delta_3(t-T) - (1 - e^{-\delta_3(T-t)})\right) \\
& + \left[\frac{c_3c_5 - c_4 + 2c_3 + c_3c_5}{\delta_1^3} + \frac{c_3 - c_3c_5 + c_3^2c_6}{\delta_1^2\delta_3} - \frac{c_3c_5}{2\delta_1^2}\right]\left(e^{-\delta_1(T-t)} - 1\right) \\
& + \left[\frac{2c_3^2c_6 - c_3c_5}{2\delta_3^2} + \frac{c_3 + c_3^2c_6}{\delta_1\delta_3^2}\right]\left(e^{-\delta_3(T-t)} - 1\right) \\
& + \left[\frac{c_3c_5 - c_4 - 2c_3 + c_3^2c_6}{4\delta_1^3} - \frac{c_3c_5}{4\delta_1^2\delta_3}\right]\left(e^{-2\delta_1(T-t)} - 1\right) \\
& - \frac{c_3^2c_6}{4\delta_3^3}\left(e^{-2\delta_3(T-t)} - 1\right) + \frac{c_3 + c_3^2c_6}{\delta_1\delta_3(\delta_1 + \delta_3)}\left(e^{-\delta_1(T-t)-\delta_3(t-T)} - 1\right)
\end{aligned}$$

where we have

$$\begin{aligned}
c_1 &= \sigma_L\sigma_{LS} & \qquad c_4 &= \sigma_S^2 + \sigma_{LS}^2 \\
c_2 &= \sigma_L\sigma_{LC} & \qquad c_5 &= \sigma_S\sigma_{SC} + \sigma_{LC}\sigma_{LS} \\
c_3 &= \frac{\delta_2}{\delta_1 - \delta_3} & \qquad c_6 &= \sigma_C^2 + \sigma_{LC}^2 + \sigma_{SC}^2.
\end{aligned}$$

C. R code example of data processing

First we select the age-range and the calendar year data to analyse:

```
AgeRange <- c(50:99)
PeriodRange <- seq(1933, 2015)
```

Hence, we extract the matrix of Death counts (`D_xt`) and of exposure-at-risk years (`E_xt`) for the male population

```
D_xt <- matrix(NA, length(AgeRange),length(PeriodRange))
E_xt <- matrix(NA, length(AgeRange),length(PeriodRange))

col <- 1
for(t in PeriodRange){
  row <- 1
  for(x in AgeRange){
    D_xt[row,col] <- deaths[((deaths$Age==x) & (deaths$Year==t)),"Male"]
    E_xt[row,col] <- exposures[((exposures$Age==x) & (exposures$Year==t)),"Male"]
    row <- row + 1
```

```
  }
  col <- col + 1
}
```

The matrix of $\mu_{x,t}$ rates is obtained as the ration between `D_xt` and `E_xt`:

```
mu_xt_matrix <- D_xt / E_xt
```

References

Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **19**(6), 716–723.

Barigou, K., Goffard, P.-O., Loisel, S. & Salhi, Y. (2023), 'Bayesian model averaging for mortality forecasting using leave-future-out validation', *International Journal of Forecasting* **39**(2), 674–690.
URL: *https://www.sciencedirect.com/science/article/pii/S0169207022000243*

Biffis, E. (2005), 'Affine processes for dynamic mortality and actuarial valuations', *Insurance: Mathematics and Economics* **37**(3), 443–468.
URL: *https://www.sciencedirect.com/science/article/pii/S0167668705000600*

Blackburn, C., Hanewald, K., Olivieri, A. & Sherris, M. (2017), 'Longevity risk management and shareholder value for a life annuity business', *ASTIN Bulletin: The Journal of the IAA* **47**(1), 43–77.

Blackburn, C. & Sherris, M. (2013), 'Consistent dynamic affine mortality models for longevity risk applications', *Insurance: Mathematics and Economics* **53**(1), 64 – 73.
URL: *http://www.sciencedirect.com/science/article/pii/S0167668713000632*

Brouhns, N., Denuit, M. & Vermunt, J. (2002), 'A Poisson Log-Bilinear Regression Approach to the Construction of Projected Life Tables', *Insurance: Mathematics and Economics* **31**, 373–393.

Cairns, A. J., Blake, D. & Dowd, K. (2006*a*), 'Pricing Death: Frameworks for the Valuation and Securitization of Mortality Risk', *ASTIN Bulletin* **36**(1), 79–120.

Cairns, A. J. G., Blake, D. & Dowd, K. (2006*b*), 'A Two-Factor Model for Stochastic Mortality With Parameter Uncertainty: Theory and Calibration', *Journal of Risk and Insurance* **73**, 687–718.

Chen, R.-R. & Scott, L. (2003), 'Multi-Factor Cox-Ingersoll-Ross Models of the Term Structure: Estimates and Tests from a Kalman Filter Model', *The Journal of Real Estate Finance and Economics* **27**(488), 143–172.

Christensen, J. H., Diebold, F. X. & Rudebusch, G. D. (2011), '"the affine arbitrage-free class of Nelson—Siegel term structure models"', *Journal of Econometrics* **164**(1), 4 – 20. Annals Issue on Forecasting.
URL: *http://www.sciencedirect.com/science/article/pii/S0304407611000388*

Christensen, J. H. E., Diebold, F. X. & Rudebusch, G. D. (2009), 'An arbitrage-free generalized Nelson—Siegel term structure model', *The Econometrics Journal* **12**(3), C33–C64.
URL: *http://www.jstor.org/stable/23116046*

Cox, J. C., Ingersoll, J. E. & Ross, S. A. (1985), 'A Theory of the Term Structure of Interest Rates', *Econometrica* **53**(2), 385–407.
URL: *http://www.jstor.org/stable/1911242*

Dahl, M. (2004), 'Stochastic mortality in life insurance: market reserves and mortality-linked insurance contracts', *Insurance: Mathematics and Economics* **35**(1), 113–136.
URL: *https://www.sciencedirect.com/science/article/pii/S0167668704000642*

Dethlefsen, C., Lundbye-Christensen, S. & Christensen, A. L. (2022), *sspir: State Space Models in R*, R Foundation for Statistical Computing.
URL: *http://CRAN.R-project.org/package=sspir*

Duffee, G. R. (2002), 'Term Premia and Interest Rate Forecasts in Affine Models', *The Journal of Finance* **57**(1), 405–443.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/1540-6261.00426*

Duffie, D. & Kan, R. (1996), 'A yield-factor model of interest rates', *Mathematical Finance* **6**(4), 379–406.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9965.1996.tb00123.x*

Duffie, D., Pan, J. & Singleton, K. (2000), 'Transform analysis and asset pricing for affine jump-diffusions', *Econometrica* **68**(6), 1343–1376.

Gilbert, P. D. (2009), *Brief User's Guide: Dynamic Systems Estimation*, R Foundation for Statistical Computing.
URL: *http://CRAN.R-project.org/package=dse*

Helske, J. (2017), 'Kfas: Exponential family state space models in r', *Journal of Statistical Software* **78**(10), 1–39.
URL: *https://www.jstatsoft.org/index.php/jss/article/view/v078i10*

Huang, Z., Sherris, M., Villegas, A. M. & Ziveyi, J. (2022), 'Modelling USA age-cohort mortality: A comparison of multi-factor affine mortality models', *Risks* **10**(9).
URL: *https://www.mdpi.com/2227-9091/10/9/183*

Hyndman, R. J., Booth, H., Tickle, L. & Maindonald, J. (2022), *demography: Forecasting Mortality, Fertility, Migration and Population Data*, R Foundation for Statistical Computing.
URL: *https://www.R-project.org/*

Hyndman, R. J. & Shahid Ullah, M. (2007), 'Robust forecasting of mortality and fertility rates: A functional data approach', *Computational Statistics and Data Analysis* **51**(10), 4942–4956.
URL: *https://www.sciencedirect.com/science/article/pii/S0167947306002453*

Jevtić, P., Luciano, E. & Vigna, E. (2013), 'Mortality surface by means of continuous time cohort models', *Insurance: Mathematics and Economics* **53**(1), 122–133.
URL: *https://www.sciencedirect.com/science/article/pii/S0167668713000619*

Jevtić, P. & Regis, L. (2019), 'A continuous-time stochastic model for the mortality surface of multiple populations', *Insurance: Mathematics and Economics* **88**, 181–195.
URL: *https://www.sciencedirect.com/science/article/pii/S0167668716302906*

Jevtić, P. & Regis, L. (2021), 'A square-root factor-based multi-population extension of the mortality laws', *Mathematics* **9**(19).
URL: *https://www.mdpi.com/2227-7390/9/19/2402*

Johnson, S. G. (2020), *The NLopt nonlinear-optimization package.*
URL: *http://github.com/stevengj/nlopt*

Koopman, S. J. & Durbin, J. (2000), 'Fast Filtering and Smoothing for Multivariate State Space Models', *Journal of Time Series Analysis* **21**(3), 281–296.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9892.00186*

Lee, R. D. & Carter, L. R. (1992), 'Modeling and Forecasting U. S. Mortality', *Journal of the American Statistical Association* **87**(419), 659–671.
URL: *http://www.jstor.org/stable/2290201*

Leippold, M. & Wu, L. (2002), 'Asset pricing under the quadratic class', *The Journal of Financial and Quantitative Analysis* **37**(2), 271–295.
URL: *http://www.jstor.org/stable/3595006*

Luethi, D., Erb, P. & Otziger, S. (2022), *FKF: Fast Kalman Filter*, R Foundation for Statistical Computing.
URL: *http://CRAN.R-project.org/package=FKF*

Milevsky, M. & Promislow, S. (2001), 'Mortality Derivatives and the Option to Annuitize', *Insurance: Mathematics and Economics* **29**, 299–318.

Petris, G. (2010), 'An r package for dynamic linear models', *Journal of Statistical Software* **36**(12), 1–16.
URL: *https://www.jstatsoft.org/index.php/jss/article/view/v036i12*

Pitacco, E. (2016), 'High Age Mortality and Frailty. Some remarks and hints for Actuarial Modeling', *Available at CEPAR* .
URL: *https://cepar.edu.au/sites/default/files/High_Age_Mortality_and_Frailty.pdf*

Plat, R. (2009), 'On stochastic mortality modeling', *Insurance: Mathematics and Economics* **45**(3), 393–404.
URL: *https://www.sciencedirect.com/science/article/pii/S0167668709000973*

Rauch, H. E., Tung, F. & Striebel, C. T. (1965), 'Maximum likelihood estimates of linear dynamic systems', *AIAA Journal* **3**(8), 1445–1450.
URL: *https://doi.org/10.2514/3.3166*

Renshaw, A. & Haberman, S. (2006), 'A Cohort-Based Extension to the Lee-Carter Model for Mortality Reduction Factors', *Insurance: Mathematics and Economics* **38**, 556–570.

Riffe, T. (2015), HMDHFDplus: Reading Human Fertility Database and Human Mortality Database data into R, Technical Report TR-2015-004, MPIDR.
URL: *http://www.demogr.mpg.de/en/projects_publications/publications_1904/mpidr_technical_reports/re*

Rowan, T. H. (1990), Functional Stability Analysis of Numerical Algorithms, PhD thesis, University of Texas at Austin, USA.

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

Schwarz, G. (1978), 'Estimating the Dimension of a Model', *Ann. Statist.* **6**(2), 461–464.

Sherris, M., Xu, Y. & Ziveyi, J. (2020), 'Cohort and value-based multi-country longevity risk management', *Scandinavian Actuarial Journal* **2020**(7), 650–676.
URL: *https://doi.org/10.1080/03461238.2019.1711450*

Stoffer, D. & Wall, K. (2009), 'Resampling in State Space Models', *State Space and Unobserved Component Models: Theory and Applications* .

Tusell, F. (2011), 'Kalman filtering in r', *Journal of Statistical Software* **39**(2), 1–27.
URL: *https://www.jstatsoft.org/index.php/jss/article/view/v039i02*

Ungolo, F., Garces, L. P. D. M., Sherris, M. & Zhou, Y. (2023), 'Estimation, comparison, and projection of multi-factor age-cohort affine mortality models', *Submitted* .

Villegas, A. M., Kaishev, V. K. & Millossovich, P. (2018), 'StMoMo: An R package for stochastic mortality modeling', *Journal of Statistical Software* **84**(3), 1–38.

Willets, R. C. (2004), 'The cohort effect: Insights and explanations', *British Actuarial Journal* **10**(4), 833–877.

Xu, Y., Sherris, M. & Ziveyi, J. (2020*a*), 'Continuous-time multi-cohort mortality modelling with affine processes', *Scandinavian Actuarial Journal* **2020**(6), 526–552.
URL: *https://doi.org/10.1080/03461238.2019.1696223*

Xu, Y., Sherris, M. & Ziveyi, J. (2020*b*), 'Market price of longevity risk for a multi-cohort mortality model with application to longevity bond option pricing', *Journal of Risk and Insurance* **87**(3), 571–595.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/jori.12273*