# ARC Centre of Excellence in Population Ageing Research

# Working Paper 2021/26

## Estimation, Comparison and Projection of Multi-factor Age-Cohort Affine Mortality Models

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

---

# Estimation, Comparison and Projection of Multi-factor Age-Cohort 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*

February 28, 2023

Affine mortality models, developed in continuous time, are well suited to longevity risk applications including pricing and capital management. A major advantage of this mortality modelling approach is the availability of closed-form cohort survival curves, consistent with the assumed time dynamics of mortality rates. This paper makes new contributions to the estimation of multi-factor continuous-time affine models including the canonical Blackburn-Sherris, the AFNS and the CIR mortality models. We discuss and address numerical issues with model estimation. We apply the estimation methods to age-cohort mortality data from five different countries, providing insights into the dynamics of mortality rates and the fitting performance of the models. We show how the use of maximum likelihood with the univariate Kalman filter turns out to be faster and more robust compared to traditional estimation methods which heavily use large matrix multiplication and inversion. We present graphical and numerical goodness-of-fit results, and assess model robustness. We project cohort survival curves and assess the out-of-sample performance of the models for the five countries. We confirm previous results, by showing that, across these countries, although the CIR mortality model fits the historical mortality data well, particularly at older ages, the canonical and AFNS affine mortality models provide better out-of-sample performance. We also show how these affine mortality models are robust with respect to the set of age-cohort data used for parameter estimation. R code is provided.

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

## 1. Introduction

The dynamics of mortality rates across countries have been modelled assuming many different stochastic processes (see Cairns et al. (2006*b*) and references therein). Following the seminal work of Lee & Carter (1992), many stochastic mortality models have been proposed (Cairns et al. (2006*a*), Plat (2009), Hyndman & Shahid Ullah (2007) to name a few). These discrete-time models have mostly been fitted to age-period data. Several extensions have been proposed to account for the cohort effect (see for example Renshaw & Haberman (2006) and Currie (2006)).

In contrast, we directly focus on age-cohort data because in actuarial applications age-cohort survival curves are required for pricing longevity-linked securities. Cohort effects are persistent throughout adult life, as justified in McCarthy (2021), unlike period effects. Period effects in age-cohort models impact all ages in a period to a greater-or-lesser extent. In contrast cohort effects in age-period models are less easy to rationalize.

Following Huang et al. (2022) we focus on affine mortality models, developed in continuous time, which build on the affine framework in Duffie & Kan (1996) applied to modelling interest rates. Amongst many others Milevsky & Promislow (2001), Dahl (2004) and Biffis (2005) are early examples of mortality models applying the modelling approach originally developed for interest rates. Our approach models the cohort survival curve as an exponentially affine function of a vector of latent variables driving the mortality dynamics through time. Previous research focused on models with two and three factors as in Jevtić et al. (2013), Blackburn & Sherris (2013), Xu et al. (2020), Jevtić & Regis (2019), Huang et al. (2022) and Jevtić & Regis (2021).

We contribute new insights to modelling the dynamics of mortality rates and the estimation of mortality models in a number of important respects:

- We shed light on the parameter estimation process using the univariate Kalman filter maximum likelihood initially proposed by Koopman & Durbin (2000) by addressing the numerical issues in its use. In addition, we provide researchers and practitioners with access to our robust and transparent methods implemented in open source R code including parameter estimation and goodness-of-fit methods used in the paper. Compared to the works of Jevtić & Regis (2019) and Jevtić & Regis (2021) who similarly use the univariate Kalman filter, we enhance the robustness of the algorithm with additional numerical tricks (e.g. the Joseph stabilized update for the conditional covariance of the latent states and the estimation of the diffusion matrix for dependent factor models);

- We propose a faster and efficient method to estimate the parameter uncertainty by means of a multiple imputation-based procedure as opposed to the computationally expensive bootstrap method used in Blackburn & Sherris (2013) and Jevtić & Regis (2019);

- We expand the range of affine specifications, and discuss the quality of fit and of the projection for models with more than three factors as thus far analysed in the current literature;

- We show the wider applicability of affine mortality models by analysing the mortality rates of five different countries with different population size, mortality experience and length of the time series of available data. Specifically, we compare the results in terms of goodness of fit measures, as well as using graphical tools to assess in-sample and out-of-sample performance of the models. In addition, we assess the model robustness with respect to the set of cohorts used for parameter estimation. To the best of our knowledge, this is the first paper analysing these aspects of affine mortality models.

This approach to mortality modelling is of relevance to practitioners since the underlying model approach based on interest rate and credit risk modelling is familiar to financial market practitioners. The modelling approach has attractive features including analytical tractability with closed form survival curves for the affine mortality class that we extend in our research, consistency between the mortality dynamics and the functional form of the survival curve, stability of parameter estimates as we show in our research, natural extensions to the multi-factor models, capturing differing trends, volatility and correlations by age and cohort that we present and assess in our research, and an arbitrage-free model setting along with real world dynamics to allow calibration of prices of risk for financial and actuarial applications.

The paper is structured as follows. Section 2 presents an overview of the mathematical framework of continuous-time affine mortality models. Section 3 summarizes the mortality data used for calibration from the Human Mortality Database (HMD). Section 4 outlines the parameter estimation procedure, based on the univariate formulation of the Kalman filter procedure to estimate the latent state variables, and to obtain a more tractable log-likelihood function. Section 5 analyses the results about the estimated parameter values. Section 6 considers the goodness-of-fit of the affine mortality models for the five countries. Section 7 assesses robustness of the fitted models with respect to the set of cohorts used for calibration. Section 8 considers and compares the projection performance of each mortality model, and Section 9 outlines a possible extension to account for period effects. Section 10 concludes. The supplementary material provides more details on parameter estimation including information for the R code.

# 2. Continuous-Time Affine Mortality Models

## 2.1. Framework

The affine mortality modelling approach is based on a consistent and arbitrage-free multi-factor model for the term structure of interest rates developed by, among others, Duffie & Kan (1996) and Dai & Singleton (2000). In particular, we assume the existence of an instantaneous mortality intensity process $\mu$, which is modelled as an affine function of an $M$-dimensional latent factor process $X$. See, for example, Biffis (2005), Dahl (2004), Schrager (2006) for an in-depth theoretical treatment.

We fix a probability space $(\Omega, \mathcal{F}, Q)$. Since we are interested in financial and actuarial pricing applications, we suppose $(\Omega, \mathcal{F}, Q)$ corresponds to an arbitrage-free market setting, where $Q$ is a risk-neutral probability measure. Furthermore, we equip $(\Omega, \mathcal{F}, Q)$ with a right-continuous and $Q$-complete filtration $\mathbb{F} := \{\mathcal{F}_t\}_{t\geq 0}$, where $\mathbb{F} := \mathbb{G} \vee \mathbb{H}$, with $\mathbb{G} := \{\mathcal{G}_t\}_{t\geq 0}$ the filtration containing all financial and actuarial information except the time of death and $\mathbb{H} := \{\mathcal{H}_t\}_{t\geq 0} = \{\sigma(\mathbf{1}_{\{\tau\leq s\}} : 0 \leq s \leq t)\}_{t\geq 0}$ the smallest filtration under which the time of death $\tau$ is a stopping time (i.e. $\mathbb{H}$ is the minimal filtration containing information about the time of death).

The latent factor process $X := \{X(t)\}_{t\geq 0} = \{(X_1(t), \ldots, X_M(t))'\}_{t\geq 0}$ ($'$ indicating matrix transpose) is modelled as an $\mathbb{F}$-Markov process whose dynamics are given by the (vector) stochastic differential equation (SDE)

$$dX(t) = \Delta(\theta^Q - X(t))dt + \Sigma D(X(t), t)dW^Q(t), \qquad X(0) = x_0 \in \mathbb{R}^M, \tag{2.1}$$

where $\Delta \in \mathbb{R}^{M\times M}$ is the mean reversion matrix, $\theta^Q \in \mathbb{R}^M$ is the long-run mean of $X$, $\Sigma \in \mathbb{R}^{M\times M}$ is the volatility matrix, $D(X(t), t)$ is an $M \times M$ diagonal matrix whose $i$th diagonal entry $d_{ii}(X(t), t)$ is given by $d_{ii}(X(t), t) = \sqrt{\alpha^i(t) + (\beta^i(t))'X(t)}$, for $i = 1, \ldots, M$, where $\alpha^i$ and $\beta^i := (\beta^i_1, \ldots, \beta^i_M)'$ are bounded and continuous functions, and $W^Q$ is a standard $M$-dimensional $Q$-Brownian motion.

The instantaneous mortality intensity $\mu := \{\mu(t)\}_{t\geq 0}$ is assumed to be an affine function of $X$, i.e. for some $\rho_0 \in \mathbb{R}$ and $\rho_1 \in \mathbb{R}^M$, we have $\mu(t) = \rho_0 + \rho_1' X(t)$. Then, using standard results for affine processes (see e.g. Duffie et al. 2000), without loss of generality, the (risk-neutral) probability that a newborn at time $t$ survives up to time $T$ is given by

$$S(t, T) := \mathbb{E}^Q\left[\exp\left\{-\int_t^T \mu(s)ds\right\}\middle|\mathcal{F}_t\right] = \exp\{A(t, T) + B(t, T)'X(t)\},$$

where $A(t, T)$ and $B(t, T)$ are solutions of a system of ordinary differential equations dependent on the coefficients of the model. We refer to $A(t, T)$ and $B(t, T)$ as the factor loadings. Consequently, the average force of mortality from $t$ to $T$ is also an affine function of the latent state process,

$$\bar{\mu}(t, T) := -\frac{1}{T-t}\log S(t, T) = -\frac{B(t, T)'}{T-t}X(t) - \frac{A(t, T)}{T-t}. \tag{2.2}$$

In our setting, $\bar{\mu}(t, T)$ and the factor loadings only depend on $T - t$.

## 2.2. Mortality Model Specification

The mortality models we analyze in this paper are specified by the dynamics of $X$, given by (2.1), and the coefficients $\rho_0$ and $\rho_1$ of $\mu$. In particular, we focus on several mortality models for which $A(t,T)$ and $B(t,T)$ are available in closed form. In all models of interest, we set $\rho_0 = 0$. In particular, we analyze the multi-factor Blackburn-Sherris model (Blackburn & Sherris 2013), the arbitrage-free Nelson-Siegel (AFNS) model (Christensen et al. 2011), the arbitrage-free generalized Nelson-Siegel (AFGNS) model (Christensen et al. 2009), and the multi-factor Cox-Ingersoll-Ross model (Cox et al. 1985, Chen & Scott 2003, Geyer & Pichler 1999).

### 2.2.1. Multi-Factor Blackburn-Sherris Model

In the three-factor version of the model, the mortality intensity is modelled as $\mu(t) = X_1(t) + X_2(t) + X_3(t)$, i.e. $\rho_1 = (1,1,1)'$, where the latent process $X = (X_1, X_2, X_3)'$ satisfies the SDE

$$\begin{pmatrix} \mathrm{d}X_1(t) \\ \mathrm{d}X_2(t) \\ \mathrm{d}X_3(t) \end{pmatrix} = -\begin{pmatrix} \delta_{1,1} & 0 & 0 \\ 0 & \delta_{2,2} & 0 \\ 0 & 0 & \delta_{3,3} \end{pmatrix} \begin{pmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \end{pmatrix} \mathrm{d}t + \begin{pmatrix} \sigma_{1,1} & 0 & 0 \\ 0 & \sigma_{2,2} & 0 \\ 0 & 0 & \sigma_{3,3} \end{pmatrix} \begin{pmatrix} \mathrm{d}W_1^Q(t) \\ \mathrm{d}W_2^Q(t) \\ \mathrm{d}W_3^Q(t) \end{pmatrix} \tag{2.3}$$

Here, we assume that $\theta^Q = \mathbf{0}$ and, for each $i, k = 1, \ldots, 3$, $\alpha^i(t) = 1$ and $\beta_k^i(t) = 0$ in the general equation (2.1). We assume that the latent factors are independent via the diagonal specification of the volatility matrix.

Given the parameterization of equation (2.3), the factor loadings $A(t,T)$ and $B(t,T)$ can be obtained in closed-form; see Section 1 of the supplementary material. In this model, $B_k(t,T)$, for $k = 1,2,3$, all have the same structure and, based on the value of $\delta_{k,k}$, they measure the sensitivity of the cohort mortality curve with respect to $X_k(t)$ for differing ages. More precisely, the larger the value of $\delta_{k,k}$, the larger the impact of $X_k(t)$ at older ages.

Dependence among the latent factors is induced by replacing $\Delta$ and $\Sigma$ with lower triangular matrices. The corresponding factor loading equations for the three-factor dependent case can be found in Huang et al. (2022, Appendix A). The extension to the four-factor case is straightforward. In the four-factor case, we only consider the case of independent factors.

### 2.2.2. Arbitrage-Free Nelson-Siegel (AFNS) and Arbitrage-Free Generalized Nelson-Siegel (AFGNS) Models

Proposed by Christensen et al. (2009), the AFGNS model is an extension of the AFNS model (Christensen et al. 2011) and is, in turn, a dynamic, arbitrage-free version of the Svensson (1995) extension of the Nelson-Siegel model. The Svensson model is a four-factor model with a level, slope, and two curvature factors to improve the fit at longer maturities. However, Christensen et al. (2009) show that an arbitrage-free version of the Svensson model can only be obtained by pairing the additional curvature factor with a corresponding slope factor. Therefore, the AFGNS model is a five-factor model with one level factor $L(t)$, two slope factors $S_1(t)$ and $S_2(t)$, and two curvature factors $C_1(t)$ and $C_2(t)$, under which the instantaneous mortality intensity is modelled as $\mu(t) = L(t) + S_1(t) + S_2(t)$, i.e. $X = (L, S_1, S_2, C_1, C_2)'$ and $\rho_1 = (1, 1, 1, 0, 0)'$.

Under the AFGNS model with independent factors, the latent factor process $X$ satisfies equation

$$\begin{pmatrix} \mathrm{d}L(t) \\ \mathrm{d}S_1(t) \\ \mathrm{d}S_2(t) \\ \mathrm{d}C_1(t) \\ \mathrm{d}C_2(t) \end{pmatrix} = -\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \delta_1 & 0 & -\delta_1 & 0 \\ 0 & 0 & \delta_2 & 0 & -\delta_2 \\ 0 & 0 & 0 & \delta_1 & 0 \\ 0 & 0 & 0 & 0 & \delta_2 \end{pmatrix} \begin{pmatrix} L(t) \\ S_1(t) \\ S_2(t) \\ C_1(t) \\ C_2(t) \end{pmatrix} \mathrm{d}t + \begin{pmatrix} \sigma_{1,1} & 0 & 0 & 0 & 0 \\ 0 & \sigma_{2,2} & 0 & 0 & 0 \\ 0 & 0 & \sigma_{3,3} & 0 & 0 \\ 0 & 0 & 0 & 0\sigma_{4,4} & 0 \\ 0 & 0 & 0 & 0 & \sigma_{5,5} \end{pmatrix} \begin{pmatrix} \mathrm{d}W_1^Q(t) \\ \mathrm{d}W_2^Q(t) \\ \mathrm{d}W_3^Q(t) \\ \mathrm{d}W_4^Q(t) \\ \mathrm{d}W_5^Q(t) \end{pmatrix}. \tag{2.4}$$

We assume that $\theta^Q = \mathbf{0}$ and, for each $i, k = 1, \ldots, 5$, $\alpha^i(t) = 1$ and $\beta_k^i(t) = 0$. The special structure of $\Delta$ above ensures that we preserve the factor loading structure inherent in the Nelson-Siegel and the AFNS models. We assume that $\delta_1 \neq \delta_2$, which is a non-binding restriction due to symmetry.

The factor loadings for the AFGNS model are available in closed form (Christensen et al. 2009, Proposition 3.1). See Section 1 of the Supplementary Material. The loading $B_L(t, T)$ affects all ages in the same way, $B_{S_\ell}(t, T)$ is increasing with the value of $T - t$, impacting older ages more than younger ages, while $B_{C_\ell}(t, T)$ is decreasing with respect to $\delta_\ell$, measuring the extent of the curvature of mortality rates. Since $\delta_1 > \delta_2$, the factor loadings for $S_1$ and $C_1$ decay faster than those for $S_2$ and $C_2$, allowing us to better capture the dynamics of the mortality intensity at older ages.

Dependence among the latent factors can be induced by replacing $\Sigma$ in (2.4) by a lower triangular matrix In this case, $B(t, T)$ is still given by (**??**). We refer to Christensen et al. (2009, Appendix) for the solution $A(t, T)$ in the dependent case. To preserve the special structure of the factor loadings and the geometric interpretation of the latent factors, we assume the same form for $\Delta$ in the dependent-factor case.

The AFNS model is a special case of the AFGNS model in which we do not have a second slope and curvature factor. That is, the instantaneous mortality intensity is characterized as $\mu(t) = L(t) + S(t)$, where the latent factor process $X = (L, S, C)'$ satisfies the equation

$$\begin{pmatrix} \mathrm{d}L(t) \\ \mathrm{d}S(t) \\ \mathrm{d}C(t) \end{pmatrix} = -\begin{pmatrix} 0 & 0 & 0 \\ 0 & \delta & -\delta \\ 0 & 0 & \delta \end{pmatrix} \begin{pmatrix} L(t) \\ S(t) \\ C(t) \end{pmatrix} \mathrm{d}t + \begin{pmatrix} \sigma_{1,1} & 0 & 0 \\ 0 & \sigma_{2,2} & 0 \\ 0 & 0 & \sigma_{3,3} \end{pmatrix} \begin{pmatrix} \mathrm{d}W_1^Q(t) \\ \mathrm{d}W_2^Q(t) \\ \mathrm{d}W_3^Q(t) \end{pmatrix}. \tag{2.5}$$

The corresponding factor loadings $A(t,T)$ and $B(t,T)$ under the AFNS model can be found in the Supplementary material.

As before, dependence among the latent factors can be introduced by replacing $\Sigma$ with a lower triangular matrix. The expressions for $B_L$, $B_S$, and $B_C$ are the same in the dependent case. The equation for $A(t,T)$ in the dependent factor case can be found in Christensen et al. (2011, Appendix B).

### 2.2.3. Multi-Factor Cox-Ingersoll-Ross Model

In the CIR model, the instantaneous mortality intensity is modelled as the sum of the components of $X = (X_1, X_2, X_3)'$, where the latent factor process $X$ satisfies the SDE

$$\begin{pmatrix} \mathrm{d}X_1(t) \\ \mathrm{d}X_2(t) \\ \mathrm{d}X_3(t) \end{pmatrix} = \begin{pmatrix} \delta_{1,1} & 0 & 0 \\ 0 & \delta_{2,2} & 0 \\ 0 & 0 & \delta_{3,3} \end{pmatrix} \left[ \begin{pmatrix} \theta_1^Q \\ \theta_2^Q \\ \theta_3^Q \end{pmatrix} - \begin{pmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \end{pmatrix} \right] \mathrm{d}t$$
$$+ \begin{pmatrix} \sigma_{1,1} & 0 & 0 \\ 0 & \sigma_{2,2} & 0 \\ 0 & 0 & \sigma_{3,3} \end{pmatrix} \begin{pmatrix} \sqrt{X_1(t)} & 0 & 0 \\ 0 & \sqrt{X_2(t)} & 0 \\ 0 & 0 & \sqrt{X_3(t)} \end{pmatrix} \begin{pmatrix} \mathrm{d}W_1^Q(t) \\ \mathrm{d}W_2^Q(t) \\ \mathrm{d}W_3^Q(t) \end{pmatrix}. \tag{2.6}$$

The presence of a square-root term in the diffusion term implies that each component of $X$ is nonnegative $Q$-a.s.; in fact, if $2\delta_{k,k}\theta_k^Q \geq \sigma_{k,k}^2$, then $X_k$ is strictly positive $Q$-a.s. Furthermore, since $X_k(s)$ is asymptotically gamma distributed (as $t \to \infty$) (Cox et al. 1985), the CIR mortality model is able to capture the heterogeneity of mortality rates at older ages; see Pitacco (2016) and the references therein.

Under this model, the factor loadings are available in closed form; see Appendix **??**. Here, the factor loadings may have a non-monotonic relationship with respect to $T - t$.

The four-factor model is obtained by increasing the dimension of (2.6) accordingly. The corresponding factor loading expressions are given by equation (1.4) in the Supplementary Material.

## 2.3. Change of Measure

Thus far, the mortality models have been developed in an arbitrage-free setting under the risk-neutral probability measure $Q$. However, to estimate the parameters of the model using historical data, we must recast the models under the historical probability measure $P$. This can be achieved by specifying an $\mathbb{R}^M$-valued $\mathbb{F}$-adapted risk premium process $\Lambda := \{\Lambda(t)\}_{t\in[0,T]}$, interpreted in our setting as the market price of mortality risk, and assuming that

$$\left.\frac{\mathrm{d}Q}{\mathrm{d}P}\right|_{\mathcal{F}_T} = M(T) := \exp\left\{ -\int_0^T \Lambda'(t)\mathrm{d}W^P(t) - \frac{1}{2}\int_0^T |\Lambda(t)|^2 \mathrm{d}t \right\},$$

where $W^P$ is an $M$-dimensional Brownian motion under $P$. Provided that $\Lambda$ is chosen such that $E_P[M(T)] = 1$, by Girsanov's Theorem, we can conclude that $Q$ and $P$ are equivalent probability measures on $(\Omega, \mathcal{F}_T)$ and the $Q$-Brownian motion $W^Q$ and the $P$-Brownian motion $W^P$ are connected via the equation $\mathrm{d}W^Q(t) = \Lambda(t)\mathrm{d}t + \mathrm{d}W^P(t)$, (see e.g. Björk 2020, Section 12.2 for further details) Consequently, the $P$-dynamics of $X$, whose risk-neutral dynamics given by (2.1), is given by

$$\mathrm{d}X(t) = (\Delta\theta^Q - \Delta X(t) + \Sigma D(X(t), t)\Lambda(t))\mathrm{d}t + \Sigma D(X(t), t)\mathrm{d}W^P(t). \tag{2.7}$$

We note that only the drift term changes due to the change of measure.

Following Duffee (2002), we specify $\Lambda$ as $\Lambda(t) = \lambda_0 + \lambda_1 X(t)$ if $X$ is given by either the Blackburn-Sherris, the AFNS, or the AFGNS model, or as

$$\Lambda(t) = \mathrm{diag}\left(\sqrt{X_1(t)}, \sqrt{X_2(t)}, \sqrt{X_3(t)}\right)\lambda_0$$

if $X$ is given by the three-factor CIR model (with obvious adjustments for the four-factor case). In both cases, $\lambda_0$ and $\lambda_1$ are, respectively, a constant column vector and a square matrix of appropriate dimensions.

This specification of the risk premium preserves the affine structure of the dynamics of $X$ under both $P$ and $Q$; that is, for some $K \in \mathbb{R}^{M\times M}$ and $\theta^P \in \mathbb{R}^M$, we can write

$$\mathrm{d}X(t) = K(\theta^P - X(t))\mathrm{d}t + \Sigma D(X(t), t)\mathrm{d}W^P(t).$$

Specifically, for the Blackburn-Sherris, the AFNS, and the AFGNS models, we have

$$\mathrm{d}X(t) = K(\theta^P - X(t))\mathrm{d}t + \Sigma\mathrm{d}W^P(t), \tag{2.8}$$

where $\theta^P = \Delta^{-1}(\Delta\theta^Q + \Sigma\lambda_0)$ and $K = \Delta - \Sigma\lambda_1$. Similarly, for the CIR model, we have

$$\mathrm{d}X(t) = K(\theta^P - X(t))\mathrm{d}t + \Sigma D(X(t), t)\mathrm{d}W^P(t), \tag{2.9}$$

where $K = \Delta - \Sigma\mathrm{diag}(\lambda_0)$ and $\theta^P = K^{-1}\Delta\theta^Q$.

The flexibility in choosing $\lambda_0$ and $\lambda_1$ implies that we are free to choose $K$ and $\theta^P$ (Duffee 2002). Thus, for the Blackburn-Sherris, ANFS, and AFGNS models, we set $\theta^P = 0$, consistent with their specification under $Q$ which also assumes a zero mean reversion level. For all models, we assume that $K = \mathrm{diag}(\kappa_1, \ldots, \kappa_M)$, where $M = 3, 4, 5$ depending on how many latent factors are involved.

## 3. Data

We analyse yearly cohort mortality data for males aged 50 to 99 for USA (cohorts born from 1883 to 1915), Australia (1872-1916), England & Wales (1795-1914), Denmark (1790-1914) and Japan (1887-1916). In this way, we consider countries having different population size, mortality experience where for example E&W has more pronounced cohort effects, and with differing lengths of the time series of available data.

Central exposure-at-risk years $E^c_{x,t}$ and number of deaths $D_{x,t}$ for each age $x$ and cohort year $t$ are sourced from the Human Mortality Database[1]. We estimate the central rate

[1] Human Mortality Database, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 15 September 2020.

of mortality with $m_{x,t} = D_{x,t}/E^c_{x,t}$. These are equal to the force of mortality under the assumption that the latter are constant between integer ages.

The probability of an individual aged $x$ at time $x+t$ to survive until calendar year $T$ (hence, to survive until age $T-t$) is given by:

$$S_x(t,T) = \prod_{j=1}^{T-t-x} e^{-m_{x+j-1,t}}. \tag{3.1}$$

It follows that the corresponding average force of mortality for an individual between age $x$ and $T-t$ is:

$$\bar{\mu}_x(t,T) = \frac{1}{T-t-x} \sum_{j=1}^{T-t-x} m_{x+j-1,t}, \tag{3.2}$$

which is analogous to equation (2.2), which referred to a new born.

In this work we set $x = 50$ and write $\bar{\mu}_t = [\bar{\mu}_{50}(t,t+1), \ldots, \bar{\mu}_{50}(t,t+N)]'$. Table 3.1 provides a stylized representation of the age-cohort dataset of average mortality rates used for the analysis. The focus on the average mortality rates is motivated by their smoothness compared to the $\mu$ rates widely used in stochastic mortality modelling. This aspect turns out valuable when dealing with parameter estimation.

Table 3.1: Dataset for the analysis of the average mortality rates between age 50 and 99 of $Y$ cohort data.

| | Cohort birth year | | | | | |
|---|---|---|---|---|---|---|
| **Age** | **1** | **2** | ... | **t** | ... | **Y** |
| 50 | $\bar{\mu}_{50}(1,52)$ | $\bar{\mu}_{50}(2,53)$ | ... | $\bar{\mu}_{50}(t,t+51)$ | ... | $\bar{\mu}_{50}(Y,Y+51)$ |
| 51 | $\bar{\mu}_{50}(1,53)$ | $\bar{\mu}_{50}(2,54)$ | ... | $\bar{\mu}_{50}(t,t+52)$ | ... | $\bar{\mu}_{50}(Y,Y+52)$ |
| ... | ... | ... | ... | ... | ... | ... |
| i | $\bar{\mu}_{50}(1,i+2)$ | $\bar{\mu}_{50}(2,i+3)$ | ... | $\bar{\mu}_{50}(t,t+i+1)$ | ... | $\bar{\mu}_{50}(Y,Y+i+1)$ |
| ... | ... | ... | ... | ... | ... | ... |
| 99 | $\bar{\mu}_{50}(1,101)$ | $\bar{\mu}_{50}(2,102)$ | ... | $\bar{\mu}_{50}(t,t+100)$ | ... | $\bar{\mu}_{50}(Y,Y+100)$ |

# 4. Inference

## 4.1. State-space representation

Although the models are developed in continuous time, their parameter estimation process is carried out using data collected in discrete time. This is analogous to the Kalman filter based estimation approaches used among the others by Blackburn & Sherris (2013) and Jevtić & Regis (2021). From the affine representation of the average force of mortality (2.2) and the discretized solution of the SDE of equation (2.8) or (2.9), we obtain the state-space model:

$$X(t) = \Phi_t X(t-j) + \eta_t, \tag{4.1}$$
$$\bar{\mu}_t = A_t + B_t X(t) + \epsilon_t \tag{4.2}$$

for $t = j, 2j, \ldots, T$, $A_t = \left[-\frac{A(t,t+1)}{1}, \ldots, -\frac{A(t,t+N)}{N}\right]'$ and $B_t = \left[-\frac{B(t,t+1)}{1}, \ldots, -\frac{B(t,t+N)}{N}\right]'$.

The *state equation* (4.1) describes the dynamics of the factor, or latent variable, $X(t) \in \mathbb{R}^M$ driven by the stochastic noise $\eta_t \sim N(0, R_t)$, $\eta_t \in \mathbb{R}^M$, $R_t \in \mathbb{R}^{M\times M}$ and the system matrix $\Phi_t \in \mathbb{R}^{M\times M}$, while the *measurement equation* (4.2) represents the linear relationship between the observed measurement $\bar{\mu}_t \in \mathbb{R}^N$ and the latent $X(t)$. $A_t \in \mathbb{R}^N$ denotes the deterministic factor underlying the dynamics of $\bar{\mu}_t$, $B_t \in \mathbb{R}^{N\times M}$ represents the factor loading and finally $\epsilon_t \sim N(0, H_t)$ is the error term of the measurement equation ($\epsilon_t \in \mathbb{R}^N$, $H_t \in \mathbb{R}^{N\times N}$). We assume that $\eta_t$ and $\epsilon_t$ are independently distributed.

For a single cohort, $A$, $B$, $H$ and $\Phi = e^{-Kj}$ do not depend on $t$, which is fixed at the initial year for the cohort data. For yearly cohort mortality rates, we set $j = 1$. For Gaussian models, such as the Blackburn-Sherris and the AFNS models, $R_t = R$ has the following structure:

$$R = \mathbb{E}\left[\left(\int_t^{t+j} e^{-K(s-t)}\Sigma \mathrm{d}W^P(s)\right)^2\right] = \left[I - e^{-Kj}\right]\Sigma\Sigma'\left[I - e^{-Kj}\right]'. \tag{4.3}$$

For the CIR model, because of the independence between the factors, $R_t$ is a diagonal matrix where each element along the diagonal $R_{t,k}$ is given by:

$$\begin{aligned} R_{t+j,k} &= \mathbb{E}\left[\left(\int_t^{t+j} e^{-\kappa_k(s-t)}\sigma_{k,k}\sqrt{X_k(s)}\mathrm{d}W_k^P(s)\right)^2\right] \\ &= \sigma_{k,k}^2\left(\frac{1-e^{-\kappa_k j}}{\kappa_k}\right)\left(\frac{1}{2}\theta_k^P\left(1-e^{-\kappa_k j}\right) + e^{-\kappa_k j}X_k(t)\right). \end{aligned} \tag{4.4}$$

The state-space representation for the CIR model represents an approximation, since $X(t)$, conditional on $X(t-1)$, has a noncentral $\chi^2$ distribution (Cox et al. 1985). However, for the parameter estimation methodology we propose, we approximate the conditional distribution by a normal distribution with mean and covariance identified by matching it to the first two moments of the exact distribution (Chen & Scott 2003, Geyer & Pichler 1999).

We remark how $\Phi$, $A$, $B$, $R_t$ and $H$ are function of the parameters indexing the state-space model.

## 4.2. Univariate Kalman filter (KF)

The Kalman filter provides the least squares estimator of the distribution of the latent state variable $X(t)$, conditional on the measurement $\bar{\mu}_t$, when $X(t)$ and $\bar{\mu}_t$ are normally distributed. This is not the case for the CIR mortality model, for the reasons explained in Section 4.1, the Kalman filter provides only an approximation in this case.

The Kalman filter consists of two recursive steps, called *forecasting* and *time-update*. A description of the maximum likelihood estimation of continuous-time affine mortality models can be found in more details in Blackburn & Sherris (2013) and Xu et al. (2020). However, when implemented on a computer, these two steps may produce a *round-off error*, which likely occurs when multiplying and inverting matrices of large dimensions, and these errors propagate at each iteration. In order to avoid the need of inverting large matrices, we employ the univariate treatment of the Kalman filter proposed by Koopman & Durbin (2000), which yields a log-likelihood function which is numerically stable for optimization.

For simplicity, we let $H$ be a diagonal matrix with diagonal elements $\left(\omega_1^2, \ldots, \omega_N^2\right)$, implying that measurements are independent across all ages. Thus, the measurement equation can be written as:

$$\bar{\mu}_{t,i} = a_i + b_i x_{t,i} + \epsilon_{t,i}, \quad \epsilon_{t,i} \sim N\left(0, \omega_i\right), \tag{4.5}$$

where $\bar{\mu}_{t,i}$ is the $i$th element of $\bar{\mu}_t$, $a_i$ is the $i$th element of the vector $A_t$, and $b_i$ is the $i$th row of the matrix $B_t$ and $x_{t,i} = X\left(t\right)$.

In the univariate Kalman filter, the state equation is written as

$$\begin{aligned} x_{t+1,1} &= \Phi x_{t,N} + \eta_t, \\ x_{t,i+1} &= x_{t,i} \end{aligned} \tag{4.6}$$

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

Koopman & Durbin (2000) derived the following recursions, given initial state $x_{0,N}$ and initial conditional covariance $\Sigma_{0,N}$:

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

$$\begin{aligned} \widehat{x}_{t,1} &= \mathbb{E}\left(x_{t,1} \mid \bar{\mu}_{1:t-1}\right) = \Phi \widehat{x}_{t-1,N}, \\ \widehat{\Sigma}_{t,1} &= \mathbb{V}\left(x_{t,1} \mid \bar{\mu}_{1:t-1}\right) = \Phi \widehat{\Sigma}_{t-1,N} \Phi' + R; \end{aligned} \tag{4.7}$$

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

$$\begin{aligned} \widehat{x}_{t,i+1} &= \mathbb{E}\left(x_{t,i+1} \mid \bar{\mu}_{1:t-1}, \bar{\mu}_{t,1:i}\right) = \widehat{x}_{t,i} + K_{t,i}\nu_{t,i}, \\ \widehat{\Sigma}_{t,i+1} &= \mathbb{V}\left(x_{t,i+1} \mid \bar{\mu}_{1:t-1}, \bar{\mu}_{t,1:i}\right) = \widehat{\Sigma}_{t,i} - K_{t,i}F_{t,i}K_{t,i}' \\ &= \left(I - K_{t,i}b_i\right)\widehat{\Sigma}_{t,i}\left(I - K_{t,i}b_i\right)' + K_{t,i}\omega_i^2 K_{t,i}', \end{aligned} \tag{4.8}$$

   where the scalar quantities $\nu_{t,i}$ and $F_{t,i}$, and the $3 \times 1$-dimensional vector $K_{t,i}$ are given by

$$\begin{aligned} \nu_{t,i} &= \bar{\mu}_{t,i} - a_i - b_i \widehat{x}_{t,i}, \\ F_{t,i} &= b_i \widehat{\Sigma}_{t,i} b_i' + \omega_i^2, \\ K_{t,i} &= \widehat{\Sigma}_{t,i} b_i' F_{t,i}^{-1}. \end{aligned} \tag{4.9}$$

When forecasting $\Sigma$, we use the Joseph stabilized form (Bucy & Joseph 1968), which is extended here to the case of univariate KF. In this way, we ensure that $\Sigma_{t,i+1}$ will be positive semidefinite, since the third line of equation (4.8) is a quadratic form.

The value of $\widehat{X}(t) = \mathbb{E}\left(X(t) \mid \bar{\mu}_{1:t}\right)$, $\widehat{X}(t \mid t-1) = \mathbb{E}\left(X(t) \mid \bar{\mu}_{1:t-1}\right)$, $\widehat{\Sigma}(t) = \mathbb{V}\left(X\left(t\right) \mid \bar{\mu}_{1:t}\right)$, and $\widehat{\Sigma}(t \mid t-1) = \mathbb{V}\left(X\left(t\right) \mid \bar{\mu}_{1:t-1}\right)$ are obtained as $\widehat{x}_{t,N}$, $\widehat{x}_{t,1}$, $\widehat{\Sigma}_{t,N}$ and $\widehat{\Sigma}_{t,1}$, respectively.

Several alternatives have been proposed to handle or to reduce the extent of the round-off error (see Chapter 7 of Grewal & Andrews (2014) for a review). Among others, we mention the square root covariance filtering, which updates the matrix $G_t$ defined such that $\Sigma_t = G_t G_t'$. For example, Wang et al. (1992) and Zhang & Li (1996) characterize the Kalman filter recursions in terms of the singular value decomposition (SVD) of $\Sigma$. In addition, Kulikova & Tsyganova (2017) provide an analytic formulation for the log-likelihood function. Our derivation of the univariate treatment of the SVD-based Kalman filter yielded the same results as the Koopman-Durbin implementation, while being much slower in the optimization of the log-likelihood function. In case the round-off error persists, an approach is to replace the matrix $\Sigma$ with its nearest symmetric positive semidefinite matrix (see Higham (1988) for further details).

Other parameter estimation methods include the use of the EM algorithm (see Särkkä (2013)). However, in our implementation this turned out to be particularly unstable as the log-likelihood function tended to diverge towards very large values.

## 4.3. Parameter estimation

The parameters of the state-space system are estimated using the maximum likelihood estimator (MLE). The likelihood function of the parameter vector $\psi$ given the observed average mortality rates $\bar{\mu}_{1:T}$ given by:

$$\log L\left(\psi \mid \bar{\mu}_{1:T}\right) = -\frac{TN}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\log F_{t,i} + \nu_{t,i}^2 F_{t,i}^{-1}\right). \tag{4.10}$$

For the CIR model, the estimated parameter vector $\widehat{\psi}$ corresponds to the quasi-MLE, as described in more detail by Chen & Scott (2003).

## 4.4. Implementation

To account for the increasing variation in the mortality rates at older ages, we specify the following parametric form for the variance of the measurement error terms:

$$\omega_i^2 = r_c + r_1 \sum_{k=1}^{i} \exp\left(r_2 k\right)/i \tag{4.11}$$

for $i = 1, \ldots, N$ (Blackburn & Sherris 2013, Huang et al. 2022). In order to foster convergence and avoid the possibility of negative values of $\omega_i^2$, we optimize the likelihood over $r_c^* = \log r_c$, $r_1^* = \log r_1$ and $r_2^* = \log r_2$. We assume that $r_1$ and $r_2$ are positive in order to capture the increasing variability in mortality rates for older individuals, where mortality data are scanty.

The initial state $x_{0,N}$ is included among the parameters to be estimated, while $\Sigma_{0,N}$ is set as a diagonal matrix with diagonal elements equal to $10^{-10}$. For the CIR model, in order we optimize over $x^*_{0,N} = \log x_{0,N}$ to ensure $x_{0,N} > 0$.[2] Similarly, we optimize over $\kappa^* = \log \kappa$ and $\theta^{P*} = \log \theta^P$. Geyer & Pichler (1999) recommend bounding $X$ below during the KF forecasting and time update steps. In our implementation we set a lower bound of $10^{-10}$.

The diffusion term for the models with dependent factors is given by the Cholesky factor of the covariance matrix of the Wiener process underlying the SDE of the latent variable $X(t)$. The positive definiteness of this covariance matrix is ensured by using the log-Cholesky parametrization described in Section A.1. In this way, we can optimize the log-likelihood function without imposing any constraint over the parameters.

The log-likelihood function we seek to optimize (eq. (4.10)) is highly non-linear with several local maxima, and may diverge in some areas of the parameter space. Hence, we need to carefully choose the starting values.

The estimation is carried out by using the coordinate ascent algorithm, where the log-likelihood function is iteratively optimized by groups of parameters, instead of taking all of them simultaneously. We find that this is more useful and robust for the estimation of largely parametrized models such as the Blackburn-Sherris model with three dependent factors and the CIR model. The gradient-free simplex method Nelder-Mead, which is readily available in `R` within the function `optim`, is used for optimization. At each iteration we use the estimates from the previous iterations as starting values.

The R code for estimating these models, is available from the Github repository `https://github.com/ungolof/affine_mortality.git` (Ungolo et al. (2021)). The resources in the repository allow also to estimate parameter uncertainty and carry out further analysis as covered in the remainder of this paper. Section 7 of the Supplementary material shows an example of how to use the code for fitting the models to mortality data.

## 4.5. Estimation of the standard errors by multiple imputation

We estimate the standard errors of the parameters of Gaussian models (Blackburn-Sherris, AFNS, and AFGNS) by using multiple imputation (Rubin 1978, Little & Rubin 2019) for the values of the latent state variables, based on the parameter estimates $\hat{\psi}$. The idea is to "complete" the joint density $f(y_{1:T}, x_{1:T} \mid \psi)$ by drawing $D$ values $x^{(1)}_{1:T}, \ldots, x^{(D)}_{1:T}$ from the smoothing distribution $f(x_{1:T} \mid y_{1:T}, \psi)$ (Rauch et al. 1965). This speeds up the optimization process and reduces the multimodality of the log-likelihood function.

For the CIR model, we found that this procedure encounters numerical problems due to the positivity constraint placed on the latent state variables, following its estimation by quasi-MLE. Thus, the CIR parameters standard error are estimated by using the bootstrap procedure of Stoffer & Wall (2009) and utilized in this context by Blackburn

[2] Geyer & Pichler (1999) recommend to initialize the filter at the long-term value of the state variable $\theta$.

& Sherris (2013).

We implement the following procedure, drawing on the framework of Little & Rubin (2019), for estimating the standard errors of the unconstrained parameter vector $\psi^*$:

**Step 1** Calculate the smoothing distribution of the states based on the parameter estimate $\widehat{\psi}^*$;

For $d = 1, \ldots, D$, repeat the following Step 2 and Step 3:

**Step 2** Draw a value for $x_1^{(d)}, x_2^{(d)}, \ldots, x_T^{(d)}$, sampled from the density $p\left(X_{1:T} \mid y_{1:T}, \widehat{\psi}^*\right)$ (the smoothing distribution);

**Step 3** Estimate the parameters of the probability distribution $p\left(x_{1:T}^{(d)}, y_{1:T}; \psi^*\right)$ hence obtaining $\widehat{\psi}^{*(d)}$:

$$\widehat{\psi}^{*(d)} = \arg\max_{\psi^*} \log p\left(x_{1:T}^{(d)}, y_{1:T}; \psi^*\right) \tag{4.12}$$

where

$$\log p\left(x_{1:T}^{(d)}, y_{1:T}; \psi^*\right) = \sum_{t=1}^{T} \left[\log p\left(y_t \mid x_t^{(d)}; \psi^*\right) + \log p\left(x_t^{(d)} \mid x_{t-1}^{(d)}; \psi^*\right)\right] \tag{4.13}$$

**Step 4** Estimation of standard errors:

**Step 4.1** Calculate $\bar{\psi}^*$:

$$\bar{\psi}^* = \frac{1}{D} \sum_{d=1}^{D} \widehat{\psi}^{*(d)} \tag{4.14}$$

**Step 4.2** Estimate the covariance matrix of the parameter estimates $\mathbb{V}\left(\widehat{\psi}^*\right)$ as follows:

$$\widehat{\mathbb{V}}\left(\widehat{\psi}^*\right) = \frac{1}{D} \sum_{d=1}^{D} \mathbb{V}\left(\widehat{\psi}^{*(d)}\right) + \left(1 + D^{-1}\right) \left[\frac{1}{D-1} \sum_{d=1}^{D} \left(\widehat{\psi}^{*(d)} - \bar{\psi}^*\right)\right] \tag{4.15}$$

where $\mathbb{V}\left(\widehat{\psi}^{*(d)}\right)$ is the covariance matrix obtained from the inverse of the Hessian matrix from the optimization process of equation (4.12). The factor $\left(1 + D^{-1}\right)$ is a correction to account for smaller samples.

**Step 4.3** The standard errors of the parameters are easily obtained as the square root of the elements on the diagonal of $\mathbb{V}\left(\widehat{\psi}^*\right)$.

The covariance matrix of the original parameter vector $\psi$ is obtained by applying the multivariate delta method to the unconstrained parameter $\psi^*$. Let $g(\cdot)$ define a function such that $g(\psi^*) = \psi$ and and $\Sigma^*_{\psi^*}$ the covariance matrix of $\psi^*$ as obtained from the multiple imputation optimization process.

The computation of the covariance matrix for $\psi = g^{-1}(\psi^*)$, denoted as $\Sigma_\psi$ requires the calculation of the Jacobian matrix $\nabla g(\psi^*)$:

$$\Sigma_\psi = \nabla g(\psi^*)^T \Sigma^*_{\psi^*} \nabla g(\psi^*) \tag{4.16}$$

For independent factor models, the only transformations are the logarithms of some parameters in order to ensure they are positive, as described in Section 4.4. Hence, $\nabla g(\psi^*)$ is a diagonal matrix with the vector of first derivatives of the inverse of the log-transformation (the exponential function) along the main diagonal. The parameters which have not been transformed, such as the $\delta$'s, will have derivative equal to 1. For the dependent factor models, we numerically compute the Jacobian for the parameters of the factor covariance matrix. This procedure is implemented more efficiently if we take $x_0$ from the smoothing distribution, allowing us to implement the optimization process in a lower number of variables.

This methodology turns out useful, because on one hand, it may not be possible to numerically compute the information matrix coming from the optimization of the likelihood function of equation (4.10) due to its very flat surface. On the other hand, bootstrap methods, as used in Blackburn & Sherris (2013), are computationally expensive if carried out hundreds of times.

The downside of this methodology, is that unlike the bootstrap, it tends to underestimate the standard error, since it is a delta method. From a computational perspective, a potential downside is given by the need to invert a Hessian matrix of larger dimensions. However, this task is simpler than inverting the Hessian matrix from the estimation procedure, where likelihood is integrated out the latent states.

# 5. Parameter estimates

Tables 2.1-2.5 of the Supplementary Material show the parameter estimates along their standard errors for each model fitted on the USA, Australia, Denmark, England&Wales and Japan male population. For all countries and all models, the $\delta$ parameters are statistically significant at 95% (i.e. $\left|\frac{\widehat{\delta}}{\text{std.err.}(\widehat{\delta})}\right| > 2$). More precisely, the $\delta$ parameters of the AFNS and AFGNS models are always negative, and the same evidence is observed for most of the diagonal elements of $\Delta$ for the CIR and Blackburn-Sherris models.

The estimated values of the $\kappa$ parameters, driving the dynamics of $X(t)$ under the real probability measure, are always such that its stochastic process is stationary[3] for the Blackburn-Sherris and the AFNS models (this condition is imposed for the CIR model). For the AFGNS models, this holds only for the Australian and for the Japanese mortality datasets. However, the $\kappa$ parameters are estimated with higher uncertainty, as

[3]For a diagonal matrix $K$ this is ensured when $\kappa_1 + \ldots + \kappa_M > 0$.

shown by their larger standard errors. This evidence has also been noted in Blackburn & Sherris (2013) when analysing two and three independent factor models for the analysis of the Swedish male population.

The standard deviations of the diffusion process parameters are always statistically significant. In most cases, their magnitude is positively associated with the magnitude of the $\kappa$ parameters. Conversely, the covariance parameter estimates are never statistically significant, despite their improvement in the goodness-of-fit compared to the independent factor models as shown in Section 6.

The $\theta^P$ parameters of the CIR models are generally very small, with the exception of $\theta_1^P$ of the CIR model with four factors for the E&W dataset and $\theta_3^P$ of the CIR model with three factors for the Japanese dataset. We observed in these cases that this evidence paired with an extremely small value of the corresponding $\kappa$ parameter, indicating that the factor follows a random walk.

Finally, the parameters $r_1$, $r_2$ and $r_c$ are always statistically significant. In any case, their small magnitude ensures that the uncertainty around the measurement is very low, ensuring a good in-sample fit of each model.

# 6. Goodness of Fit

## 6.1. AIC, BIC and RMSE

We assess the goodness of fit of the models models by means of a range of measures: Akaike Information Criterion (AIC, Akaike (1974)), Bayesian Information Criterion (BIC, Schwarz (1978)) and Root Mean Squared Error (RMSE), which we seek to minimize, as in Blackburn & Sherris (2013) and Huang et al. (2022):

$$\text{AIC} = -2\log L\left(\widehat{\psi} \mid \bar{\mu}_{1:T}\right) + 2k,$$
$$\text{BIC} = -2\log L\left(\widehat{\psi} \mid \bar{\mu}_{1:T}\right) + 2kTN,$$
$$\text{RMSE} = \frac{1}{TN}\sum_x\sum_t\left(\bar{\mu}_{t,x} - \widehat{\bar{\mu}}_{t,x}\right)^2,$$

where $k$ is the number of parameters, $t = 1, \dots, T$ and $x = 50, \dots, 99$, hence $N = 50$. When counting the number of parameters, we take into account the additional number of initial states $X(0)$ we estimate, as described in Section 4.4. We show the results for each model and for each country in Table 6.1.

The use of additional factors sensibly improves the in sample performance in terms of the AIC, BIC and RMSE for all countries, compared to the three-factor models as analysed by Huang et al. (2022) for the US dataset. In particular, the CIR mortality model with four factors shows the best performance in terms of AIC and BIC for all countries, except E&W. In this latter case the best performance is achieved under the AFGNS model with five dependent factors. Additional factors improve the in-sample fit in terms of RMSE for all countries with exception of the Japan. We must assume this is due to the smaller dataset which is likely to make parameter estimate lesser stable. The evidence in this case is mixed: the RMSE is smallest when using the Blackburn-Sherris model with four independent factors for the Australian dataset, the AFGNS model with independent factors for the US data, the CIR model with four factors when analysing the Danish dataset and the AFGNS model with dependent variable for the E&W data. The Blackburn-Sherris model with three dependent factors shows the smallest RMSE for the Japan mortality data.

We also considered versions of the Gaussian models with a nonzero long-run mean $\theta^P$. However, we observed that the quality of the fit does not improve significantly over the case with $\theta^P = 0$. Similar results are obtained when analysing the two-Gaussian factor Makeham-Gompertz model presented in Schrager (2006) which yields age-dependent $A(t,T)$ and $B(t,T)$[4].

The results for the US data are consistent with those of Huang et al. (2022), who also found that the CIR model with three factors is the best performing model in terms of AIC, BIC and RMSE. However, the log-likelihood function turns out to be higher, as well as the RMSE. The parameter estimates obtained are consistent with Huang et al. (2022).

All mortality models, with parameter estimates shown in the Supplementary material, except the AFGNS model with independent factors have negligible probability of negative mortality rates at all ages when projected over the next cohort year. For the US dataset, also the AFGNS model with dependent factors shows a probability of negative rates around 5% after age 90.

We also assess the in-sample model performance by comparing the empirically observedcohort survival curves with those fitted for each mortality model. We also calculate the Mean Absolute Percentage Error (MAPE) for each age, across all cohorts to compare the estimated survival curves. The results are plotted in Figure 6.1 for Australia and England&Wales[5], where we use a separate scale to account for the larger magnitude of the MAPE above age 85.

When considering MAPE for the survival curves, we see that the additional factors generally improve the MAPE over the whole age span, with the only exception given by the AFGNS models when fitted to the older ages for the Japanese dataset.

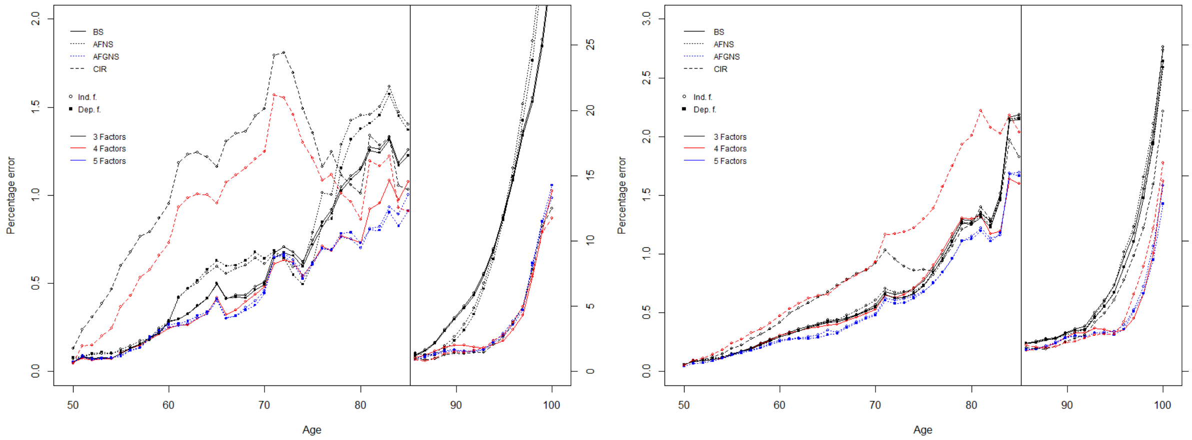
Furthermore, for all countries the CIR models show the worst fit at younger ages, although they provide a good fit at older ages. We hypothesise that this is due to the Gamma distribution of the factors, which are better at capturing increased variability arising from heterogeneity at older ages. This is more pronounced for the USA and the Australian data. The other Gaussian models have similar fits to the survival curve data.

[4] Full results are available upon request to the authors.

[5] The plots for the other countries are available upon request.

Table 6.1: Comparison of Affine Models (the best results in terms of AIC, BIC and RMSE are underlined)

| | Model | M | USA | Aus | Dnk | E&W | Jap |
|---|---|---|---|---|---|---|---|
| | BS ind. | 3 | −21252.67 | −28283.02 | −77165.58 | −74206.37 | −13998.13 |
| | BS ind. | 4 | −22726.41 | −29476.88 | −78544.46 | −75941.66 | −14063.30 |
| | BS dep. | 3 | −21442.46 | −28414.46 | −77629.61 | −74685.19 | −14016.87 |
| | AFNS ind. | 3 | −20849.12 | −27883.92 | −77116.23 | −74171.88 | −13984.60 |
| **AIC** | AFNS dep. | 3 | −21368.86 | −28129.14 | −77430.76 | −74667.22 | −14003.06 |
| | AFGNS ind. | 5 | −22792.23 | −29286.10 | −77696.70 | −75354.73 | −14951.12 |
| | AFGNS dep. | 5 | −22875.43 | −29831.39 | −78850.31 | <u>−76618.15</u> | −15031.00 |
| | CIR | 3 | −21440.65 | −28646.51 | −78456.13 | −75265.79 | −14253.31 |
| | CIR | 4 | <u>−23804.88</u> | <u>−30008.43</u> | <u>−79056.70</u> | −76520.06 | <u>−15032.41</u> |
| | BS ind. | 3 | −21187.77 | −28214.40 | −77084.70 | −74125.97 | −13939.24 |
| | BS ind. | 4 | −22639.87 | −29385.38 | −78436.61 | −75834.47 | −13984.78 |
| | BS dep. | 3 | −21345.10 | −28311.52 | −77508.28 | −74564.60 | −13928.53 |
| | AFNS ind. | 3 | −20795.04 | −27826.73 | −77048.83 | −74104.88 | −13935.52 |
| **BIC** | AFNS dep. | 3 | −21298.55 | −28054.80 | −77343.13 | −74580.13 | −13939.25 |
| | AFGNS ind. | 5 | −22700.29 | −29188.88 | −77582.12 | −75240.84 | −14867.69 |
| | AFGNS dep. | 5 | −22729.40 | −29676.99 | −78668.33 | <u>−76437.26</u> | −14898.49 |
| | CIR | 3 | −21359.52 | −28560.73 | −78355.02 | −75165.30 | −14179.69 |
| | CIR | 4 | <u>−23696.71</u> | <u>−29894.06</u> | <u>−78921.90</u> | −76386.07 | <u>−14934.25</u> |
| | BS ind. | 3 | 0.00212 | 0.00187 | 0.00253 | 0.00160 | 0.000298 |
| | BS ind. | 4 | 0.00059 | <u>0.00078</u> | 0.00252 | 0.00086 | 0.000305 |
| | BS dep. | 3 | 0.00202 | 0.00183 | 0.00251 | 0.00148 | <u>0.000289</u> |
| | AFNS ind. | 3 | 0.00287 | 0.00205 | 0.00261 | 0.00162 | 0.000302 |
| **RMSE** | AFNS dep. | 3 | 0.00212 | 0.00193 | 0.00248 | 0.00147 | 0.000297 |
| | AFGNS ind. | 5 | <u>0.00035</u> | 0.00081 | 0.00260 | 0.00086 | 0.001833 |
| | AFGNS dep. | 5 | 0.00052 | 0.00082 | 0.00247 | <u>0.00081</u> | 0.001812 |
| | CIR | 3 | 0.00092 | 0.00112 | 0.00228 | 0.00128 | 0.000378 |
| | CIR | 4 | 0.00110 | 0.00095 | <u>0.00217</u> | 0.00105 | 0.000513 |




(a) Australia

(b) England&Wales

Figure 6.1: Mean Absolute Percentage Error by age for each country and model.

## 6.2. Residuals

We also assess model adequacy by considering the quality of fit visually based on the standardized residuals. These are given in vector form for each cohort by $r_t$, calculated as

$$r_t = \left(\sqrt{\widehat{\mathbb{V}\left(\bar{\mu}_t\right)}}\right)^{-1} \left(\bar{\mu}_t - \widehat{\bar{\mu}}_t\right), \tag{6.1}$$

where from the measurement equation (4.2) it follows that

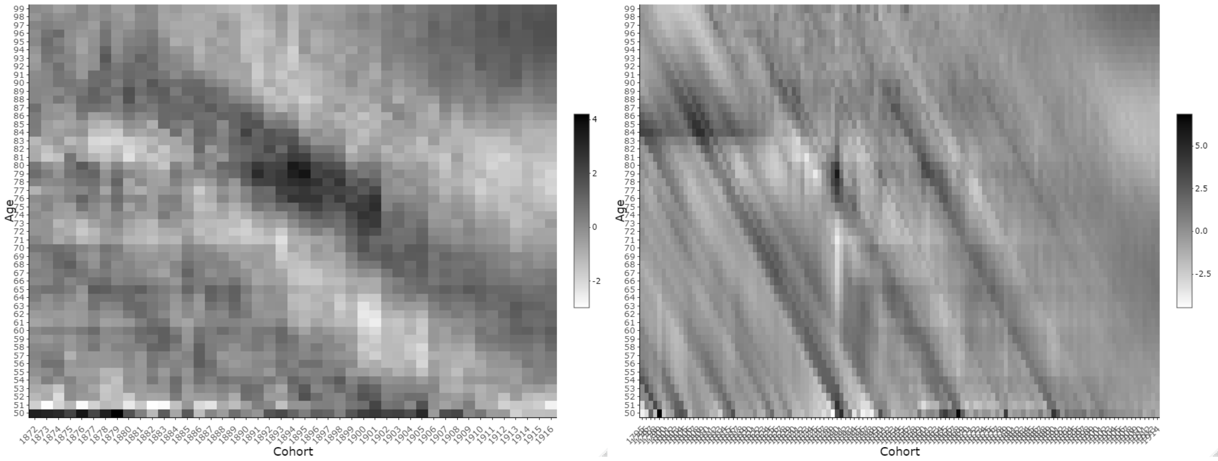
$$\begin{aligned} \mathbb{V}\left(\bar{\mu}_t\right) &= \mathbb{E}\left(\mathbb{V}\left(\bar{\mu}_t \mid X_t\right)\right) + \mathbb{V}\left(\mathbb{E}\left(\bar{\mu}_t \mid X_t\right)\right) \\ &= H + BR_tB' \end{aligned} \tag{6.2}$$

Figure 6.2 shows the standardized residuals for the Blackburn-Sherris model with three dependent factors for the Australian mortality dataset (left) and for the AFNS independent-factor model for England&Wales (right). The plots for the other models and countries are shown in Section 2 of the Supplementary material.

Since we analyse age-cohort data, period effects appear as diagonals in the residuals. We observe that for USA, Australia, and Denmark there appears to be a period effect corresponding to calendar years around 1970. The same holds for Japan, although to a lesser extent. This finding is consistent with those of Huang et al. (2022) who, in the context of USA data, attribute this to a period of mortality improvements around 1970 likely reflecting the impact of reductions in smoking.

Similar evidence for the presence of a period effect is found for Denmark as well as E&W around 1890 and 1920, the latter corresponding to the global influenza pandemic of 1918-1920. Another period effect is observed around 1940 at the beginning of the

Second World War. We see no significant period effects for Japan, since the standardized residuals are relatively small.




(a) Blackburn-Sherris three dependent factors for USA

(b) AFNS with independent factors for E&W

Figure 6.2: Plots of standardized residuals by age and cohort.

# 7. Robustness

We analyse the robustness of the parameter estimates with respect to the set of cohorts included in the estimation of the models. A model is robust when its parameter estimates are not sensitive with respect to the set of data used for inference.

The estimation process is repeated using the Danish mortality data restricted to the cohorts born from 1810, 1830 and 1850, and for the England & Wales cohorts born from 1815, 1835 and 1855. We report the results only for these two countries, due to the length of available time series of mortality rates.

The Blackburn-Sherris model with independent factors (Figure 4.1 in the Supplementary Material7.1) is robust, except for the parameter $\delta_{11}$ of the three factor model, which affects the factor loading $B_1(t,T)$ for the Danish male population. For the other parameters, we observe that the factor loadings and the yield-adjustment term $A(t,T)$ maintain the same relationship with respect to the age. Furthermore, all factors $X(t)$ follow the same behavior across the cohorts, except for a parallel shift, which depends on the estimated value of the initial state value, which is estimated among the parameters of the model. We see that a larger initial value of one factor is compensated by a smaller value of another of the others. We also see that the parameters of the mortality dynamics are stable over time. This evidence is consistent with the results of Blackburn & Sherris (2013). When we account for factor dependence in the three-factor Blackburn-Sherris model, we have a similar result compared to the independent factor model. In

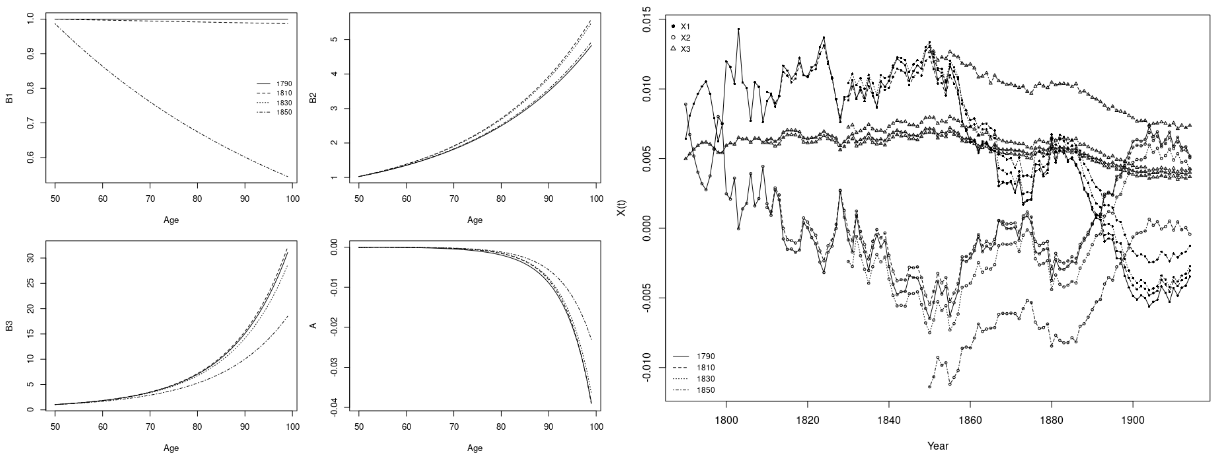


Figure 7.1: Factor loadings (left) and factor values $X(t)$ (right) for the Blackbun-Sherris model with three independent factors for Denmark.

this case, we see that for the E&W dataset the model lacks some robustness when using the smallest set of cohorts (born from 1855).

The two AFNS models are also robust with respect to the set of cohorts used for their calibration. Once again, the parallel shifts observed in the values of level, slope and curvature are due to a different estimate of the initial value of the initial state $X(0)$. The values of the factor loadings $B_2(t)$ and $B_3(t)$, which depend uniquely on the parameter $\delta$, are not affected by the length of the time series of the restricted datasets.

The factor loadings of the two AFGNS models have a consistent relationship with respect to the age, irrespective of the cohort range used in the estimation. However, a closer inspection of the level, the two slope and the two curvature factors shows how the use of younger cohorts yields steeper trajectories for the latent states, showing a relative lack of robustness of these models, especially for their use in projecting future mortality rates.

The CIR model is robust with respect to the cohorts used in the analysis. The only exception seems the estimated parameter $\delta_2$ for the model with three factor in the E&W dataset and of the four factor model for the Danish dataset. In both cases, this problem occurs only when considering the smallest dataset. Nevertheless, this seems not to affect the estimated value of the latent states $X(t)$, which seems consistent across the datasets used for parameters estimation. The only issue we observe is that for the E&W dataset, one of the factors hits the lower bound for the youngest cohorts.

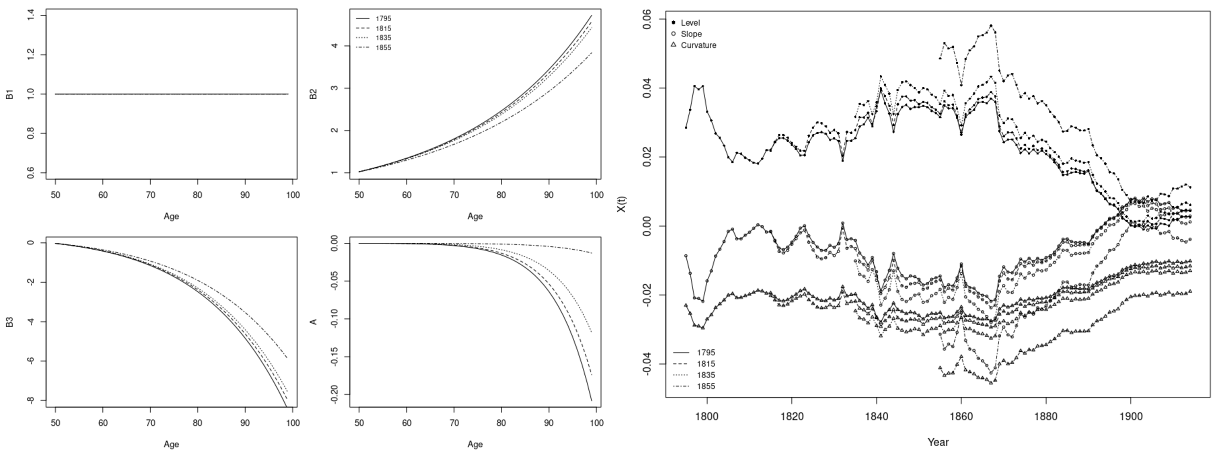



Figure 7.2: Factor loadings (left) and factor values $X(t)$ (right) for the AFNS model with dependent factors for the E&W dataset.

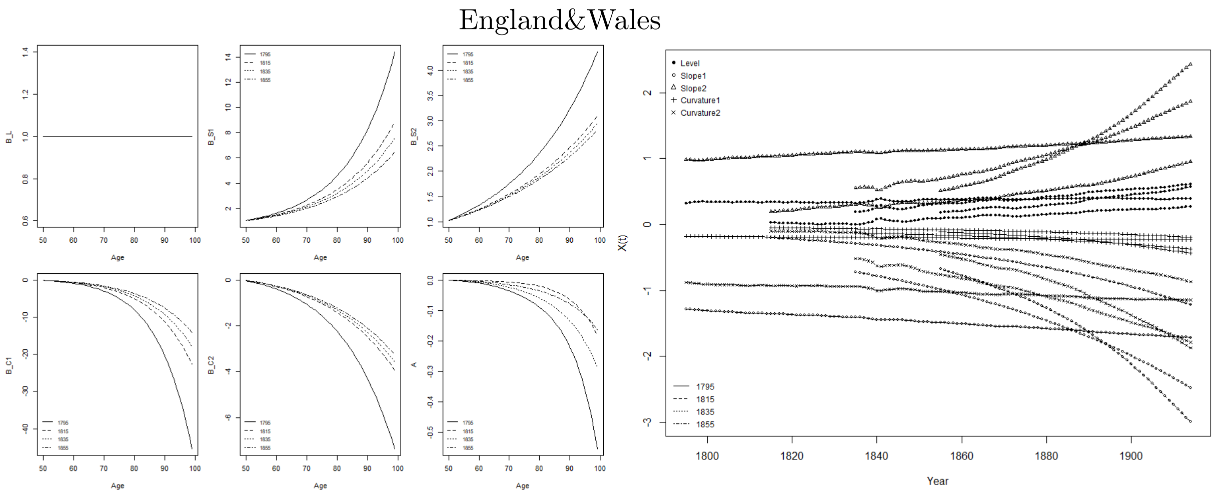



Figure 7.3: Factor loadings (left) and factor values $X(t)$ (right) for the AFGNS model with independent factors for the E&W dataset.

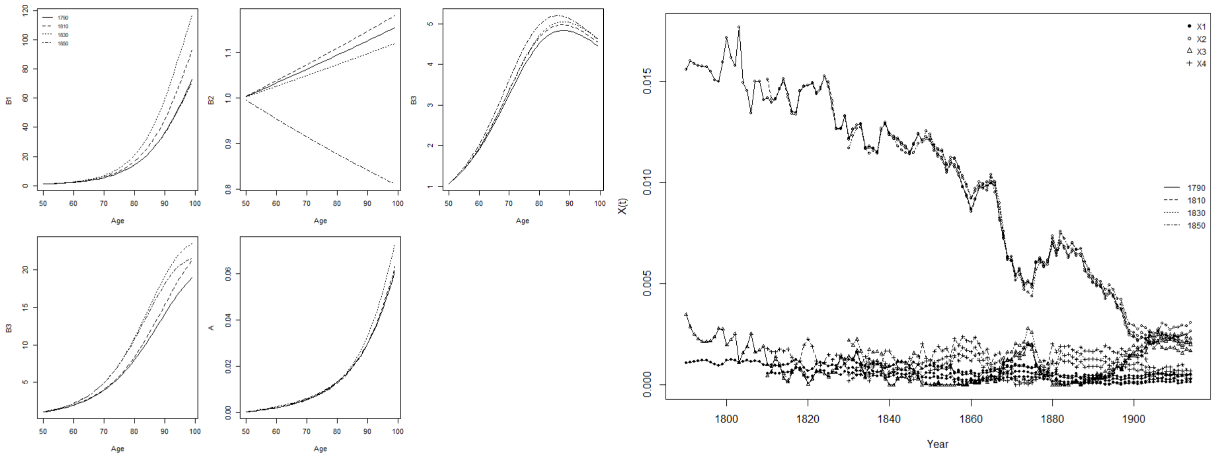



Figure 7.4: Factor loadings (left) and factor values $X(t)$ (right) for the CIR model with four factors for Denmark.

# 8. Forecasts of Cohort Survival Curve

We project survival curves for each affine mortality model for the US cohort born in 1916, the Australian and Japanese cohorts born in 1917 and the E&W and Danish cohort born in 1915.

The forecast is constructed using the conditional expectation of $X(t+h)$, given $X(t)$. That is, forecasts of the average force of mortality and the survival probability are calculated as

$$\bar{\mu}_x(t+h, T+h) = -\frac{B(t,T)}{T-t}\mathbb{E}\left[X(t+h) \mid X(t)\right] - \frac{A(t,T)}{T-t}, \tag{8.1}$$

and

$$S_x(t+h, T+h) = \exp\left(B(t,T)' \, \mathbb{E}\left[X(t+h) \mid X(t)\right] + A(t,T)\right). \tag{8.2}$$

This is the optimal forecast of the average force of mortality under quadratic loss (see Christensen et al. (2011)). Table 8.1 shows the RMSE of the survival curve forecasts.

The AFGNS model with independent factors has the best one-year out of sample performance in terms of RMSE for Australia and Denmark, whereas the AFGNS model with dependent factors best forecasts the survival curve for USA. The Blackburn-Sherris model with four independent factors yields the best performance for England & Wales, and, for Japan, the Blackburn-Sherris model with three independent factors performs best.

We also plot the survival curves and the MAPEs with respect to realized survival probabilities for the cohorts of interest in Figures 8.1 to 8.2, and Figures 5.1-5.3 in the Supplementary Material.

Table 8.1: RMSE for comparing Actual and Best-Estimate average mortality rates for projected cohorts.

| | M | USA 1916 | Australia 1917 | Denmark 1915 | E&W 1915 | Japan 1917 |
|---|---|---|---|---|---|---|
| BS ind. | 3 | 0.00200 | 0.00366 | 0.00234 | 0.00308 | 0.00086 |
| BS ind. | 4 | 0.00248 | 0.00051 | 0.00153 | 0.00036 | 0.00210 |
| BS dep. | 3 | 0.00204 | 0.00389 | 0.00169 | 0.00264 | 0.00181 |
| AFNS ind. | 4 | 0.00351 | 0.00406 | 0.00182 | 0.00314 | 0.00120 |
| AFNS dep. | 3 | 0.00135 | 0.00303 | 0.00144 | 0.00249 | 0.00129 |
| AFGNS ind. | 5 | 0.00139 | 0.00044 | 0.00119 | 0.00074 | 0.00556 |
| AFGNS dep. | 5 | 0.00054 | 0.00069 | 0.00255 | 0.00155 | 0.00573 |
| CIR | 3 | 0.00152 | 0.00063 | 0.00128 | 0.00145 | 0.00308 |
| CIR | 4 | 0.00168 | 0.00098 | 0.00164 | 0.00143 | 0.00087 |

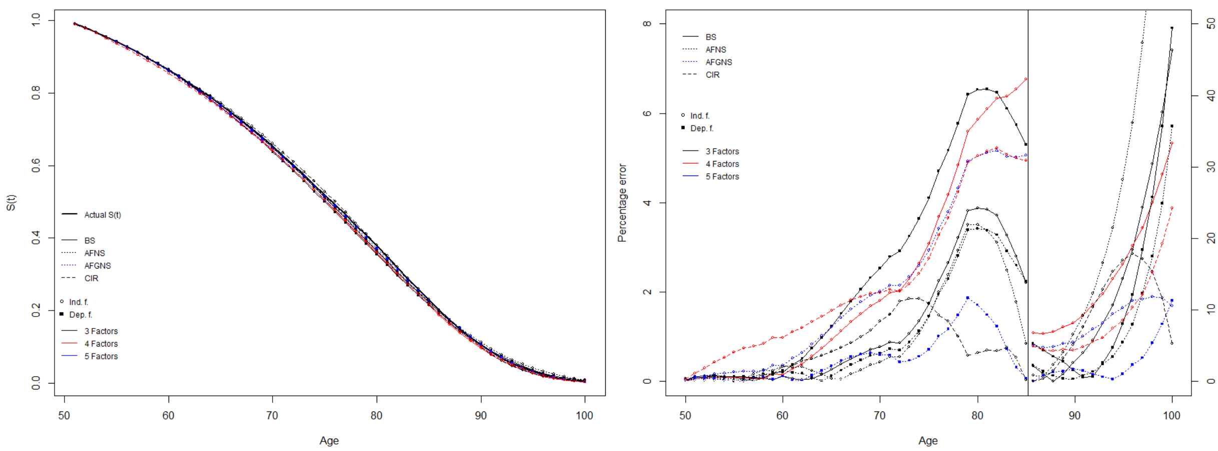



Figure 8.1: Projection of the survival curves for the USA male cohort born in 1916 under each model (left) and their MAPE (right).

For USA, we note that all mortality models tend to underestimate mortality rates at older ages, except the CIR with three factors. The opposite holds between age 65 and 85, since their survival curves lie above the empirical curve. The MAPE considerably increases at older ages for all models, except the CIR model with three factors and the two AFGNS models.

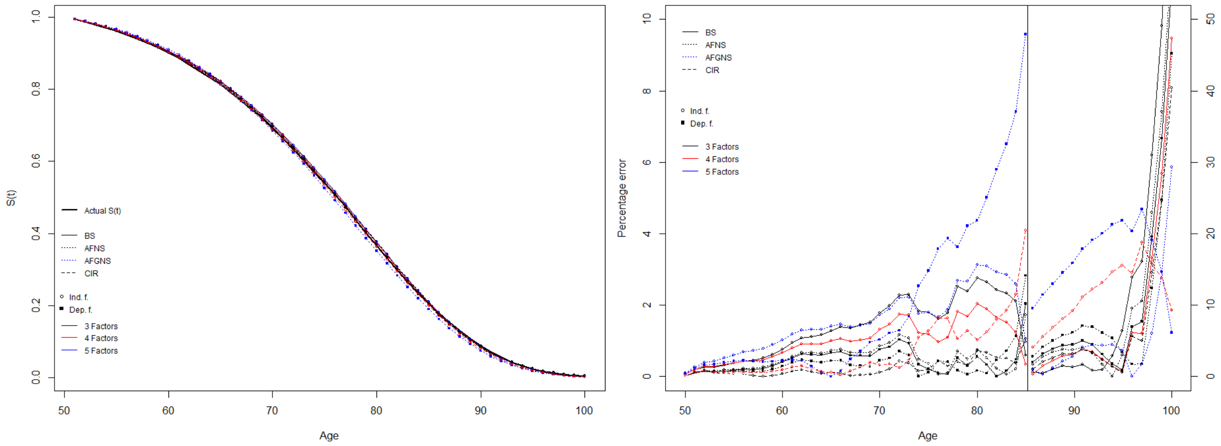



Figure 8.2: Projection of the survival curves for the Danish male cohort born in 1915 under each model (left) and their MAPE (right).

For Australia, the Blackburn-Sherris model with three independent factors, the AFGNS model with dependent factors, and the CIR model with three dependent factors slightly overestimate the mortality rates between ages 70 and 85. Mortality rates are underestimated for ages over 90 under the two Blackburn-Sherris models with three factors and the AFNS model with dependent factors. The AFNS model with independent factors tends to underestimate mortality rates under age 70. Survival curves projected using the Blackburn-Sherris and the CIR models with four factors and the AFGNS model with independent factors are very close to observed survival curve.

The models we consider yield a larger RMSE because of poorer forecasting performance at older ages, where mortality rates are underestimated. Under age 95, the MAPE is below 10%, except for the CIR model with four factors and the AFGNS model with dependent factors. The CIR mortality model has the lowest MAPE at very old ages, but underestimates mortality rates for Australia across all other ages.

The AFGNS mortality model with dependent factors shows a relatively poor performance when predicting the survival curve of the 1915 Danish cohort, particularly until age 95, where mortality rates are overestimated. For this dataset, the additional factor for the Blackburn-Sherris model yielded projected mortality rates closer to those empirically observed compared to their three-factor counterparts. The other models yield survival curves close to the empirically observed one. The MAPE is below 10% for all models until age 95, except the AFGNS with dependent factor and the CIR model with four factors.

For the England&Wales, the CIR models tend to underestimate mortality rates at younger ages, as their projected survival curve lies above the empirically observed one. All other models show a good out-of-sample performance at least until age 92. For very older people, the two CIR, the two AFGNS and the Blackburn-Sherris with four factor models improve the out-of-sample performance.

For the Japanese 1917 cohort, we see that the two AFGNS models show the worst out-of-sample performance, as seen from a large MAPE and a survival curve which overestimates mortality rates until age 95 and overestimate these at very old ages. At a lesser extent, the same holds for the CIR model with three facors. Conversely, the Blackburn-Sherris model with four factors, tends to underestimate mortality rates for the projected survival curve. All other models yield survival curves which closely overlap the one observed for the 1917 cohort, as also shown by a relatively small MAPE.

## 9. Period effects

The age-cohort affine models do not include factors accounting for any period effects, which would capture changes in mortality due to wars, pandemics, and other effects that impact all ages to a greater or lesser extent over specified time periods.

In Section 6.2 we discuss potential period effects in the model residuals. An approach to capture period effects is to use period-related factor values. For example, for the Blackburn-Sherris and CIR type models, we can specify $\mu'(t) = \mu(t) + 1_{[p=1]}X^P(t)$, where $X^P(t)$ is a period specific effect and $p = 1$ if there is a period effect component and 0 otherwise. This specification reflects the approach used in Jevtić et al. (2013) and Xu et al. (2020), where the period effect was the main time dimension and then a cohort specific factor was added in this specification. Again, $X^P(t)$ has dynamics characterised by the same SDE of equation (2.1), allowing for mean-reversion. For the AFNS, it is important to consider period-specific slope and curvature factors as the two cannot be separated (see Christensen et al. (2009)).

Including period effects in this way would allow the application of our modelling approach to forecast future cohort survival curves and to assess the impact of stress scenarios from period effects such as pandemics. A challenge is to identify the years that a period effect occurs and to separate these from age-cohort effects that differ across ages. This aspect is not considered here and is left for future research.

## 10. Conclusions

We contribute to the analysis and application of multi-factor affine mortality models by providing a new estimation methodology and insights from the analysis of mortality data across different countries.

We developed and implemented an improved estimation method for continuous-time affine models using age-cohort data. In this way, we can easily estimate models with several factors in a reasonable amount of time. The approach accounts for the increasing variation at older ages in the measurement equation and can be extended to include additional cohort- and period-specific factors and parameters. The models can also be easily extended to account for a non-zero correlation in the average force of mortality rates across ages (Koopman & Durbin (2000)).

We fitted the affine mortality models to the mortality rates of five different countries with differing length of the time series of available age-cohort data and compared them to observed mortality rates. We show how additional factors enhance the in-sample fit performance as well as a better out of sample performance. The four factor CIR mortality model shows a better in-sample performance in terms of information criteria, while all other models with more than three factors show a reduced RMSE. The CIR mortality models had better in-sample performance for fitting mortality rates at older ages, reflecting its ability to better capture possible mortality heterogeneity. The affine mortality models based on Gaussian distributed factors fit well, particularly at younger ages and perform well in forecasting future cohort survival curves. We also show that affine mortality models are generally robust with respect to the range of cohorts used for their calibration, with exception of the AFGNS models.

To assist other researchers and practitioners, we provide R code for the models estimation and assessment.

## 11. Acknowledgements

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.

# A. Appendix

### A.1. Optimization with respect to the covariance matrix

The estimation process through optimization of any $M \times M$ covariance matrix $\mathbf{\Lambda}$ must ensure that the resulting estimate is positive semidefinite. If the covariance matrix is diagonal, that is $\mathbf{\Lambda} = \text{diag}\left(\lambda_1^2, \dots, \lambda_N^2\right)$, then it is sufficient that $\lambda_i > 0$. In order to remove the constraint, we optimize over $\lambda_i^* = \log \lambda_i$.

In the general case of nonzero covariances, setting constraints in order to keep $\mathbf{\Lambda}$ positive semidefinite can be very cumbersome, as discussed by Pinheiro & Bates (1996). In our implementation we remove the constraint by using the log-Cholesky parametrization analysed in Pinheiro & Bates (1996).

Let $L$ be a lower triangular matrix, denoting the Cholesky factor of $\mathbf{\Lambda}$, such that $\mathbf{\Lambda} = LL^T$. In order to ensure that the Cholesky factor is unique, we constrain the diagonal elements of $L$ to be positive.

The parameter vector $\theta$ used for optimization is the following:

$$\theta = (\log l_{11}, l_{21}, \log l_{22}, l_{31}, l_{32}, \log l_{33}, \dots, \log l_{MM}) . \tag{A.1}$$

The log-Cholesky factorization turns out to be computationally simple and stable, but on the other hand, it lacks a direct interpretation of the parameters in terms of $\mathbf{\Lambda}$.

## References

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

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*

Björk, T. (2020), *Arbitrage Theory in Continuous Time*, 4th edn, Oxford University Press, Oxford, UK.

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*

Bucy, R. & Joseph, P. (1968), *Filtering for Stochastic Processes with Applications to Guidance*, AMS Chelsea Publishing Series, AMS Chelsea Pub.

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*

Currie, I. D. (2006), 'Smoothing and Forecasting Mortality Rates with P-Splines', *London: Institute and Faculty of Actuaries* .
**URL:** *https://www.actuaries.org.uk/system/files/documents/pdf/currie.pdf*

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*

Dai, Q. & Singleton, K. J. (2000), 'Specfication analysis of affine term structure models', *Journal of Finance* **55**(5), 1943–1978.

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.

Geyer, A. L. J. & Pichler, S. (1999), 'A state-space approach to estimate and test multifactor Cox-Ingersoll-Ross models of the term structure', *Journal of Financial Research* **22**(1), 107–130.
**URL:** *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-6803.1999.tb00717.x*

Grewal, M. & Andrews, A. (2014), *Kalman filtering: theory and practice using MATLAB*, New York: John Wiley and Sons.

Higham, N. J. (1988), 'Computing a nearest symmetric positive semidefinite matrix', *Linear Algebra and its Applications* **103**, 103 – 118.
**URL:** *http://www.sciencedirect.com/science/article/pii/0024379588902236*

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. & 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*

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*

Kulikova, M. & Tsyganova, J. (2017), 'Improved Discrete-Time Kalman Filtering within Singular Value Decomposition', *IET Control Theory and Applications* **11**, 2412–2418.

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*

Little, R. J. A. & Rubin, D. B. (2019), *Statistical Analysis with Missing Data, 3rd ed.*, Wiley, USA.

McCarthy, D. (2021), '80 will be the new 70: Old-age mortality postponement in the United States and its likely effect on the finances of the OASI program', *Journal of Risk and Insurance* **88**(2), 381–412.
**URL:** *https://onlinelibrary.wiley.com/doi/abs/10.1111/jori.12321*

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

Pinheiro, J. C. & Bates, D. M. (1996), 'Unconstrained Parameterizations for Variance-Covariance Matrices', *Statistics and Computing* **6**, 289–296.

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.

Rubin, D. (1978), Multiple imputations in sample surveys-a phenomenological Bayesian approach to nonresponse, *in* 'Proceedings of the Survey Research Methods Section of the American Statistical Association', American Statistical Association, VA, USA, pp. 20–34.

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.

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

Svensson, L. E. O. (1995), 'Estimating forward interest rates with the extended Nelson–Siegel method', *Sveriges Riksbank Quarterly Review* **1995**(3), 13–26.

Särkkä, S. (2013), *Bayesian Filtering and Smoothing*, Institute of Mathematical Statistics Textbooks, Cambridge University Press.

Ungolo, F., Garces, L. P. D. M., Sherris, M. & Zhou, Y. (2021), '`affine_mortality`: A github repository for estimation and analysis of affine mortality models', *Working paper* .

Wang, L., Libert, G. & Manneback, P. (1992), Kalman filter algorithm based on singular value decomposition, *in* '[1992] Proceedings of the 31st IEEE Conference on Decision and Control', pp. 1224–1229 vol.1.

Xu, Y., Sherris, M. & Ziveyi, J. (2020), '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*

Zhang, Y. & Li, X. R. (1996), Fixed-interval smoothing algorithm based on singular value decomposition, *in* 'Proceeding of the 1996 IEEE International Conference on Control Applications IEEE International Conference on Control Applications held together with IEEE International Symposium on Intelligent Contro', pp. 916–921.

# Supplemetary material to Estimation, Comparison and Projection of Multi-factor Age-Cohort 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*

February 28, 2023

## 1 Factor Loadings for the Mortality Models

### 1.1 Multi-Factor Blackburn-Sherris Model

The factor loadings for the three-factor Blackburn-Sherris model with independent factors are given by

$$\begin{aligned} A(t,T) &= \frac{1}{2}\sum_{k=1}^{3}\frac{\sigma_{k,k}^2}{\delta_{k,k}^3}\left[\frac{1}{2}(1-e^{-2\delta_{k,k}(T-t)})-2(1-e^{-\delta_{k,k}(T-t)})+\delta_{k,k}(T-t)\right] \\ B_k(t,T) &= -\frac{1-e^{-\delta_{k,k}(T-t)}}{\delta_{k,k}}, \qquad k=1,2,3. \end{aligned} \tag{1.1}$$

The factor loadings in the four-factor case are given by the same expressions for $k = 1,2,3,4$.

For the three-factor Blackburn-Sherris model with dependent factors, we refer to Huang et al. (2022, Appendix A) for the corresponding factor loading expressions.

## 1.2 Arbitrage-Free Generalized Nelson-Siegel Model

The factor loadings for the AFGNS model with independent factors are given by

$$
\begin{aligned}
A(t,T) &= \sigma_{1,1}^2 \frac{(T-t)^3}{6} + \sigma_{2,2}^2 (T-t) \left[ \frac{1}{2\delta_1^2} + \frac{1}{\delta_1^3} \frac{1-e^{-\delta_1(T-t)}}{T-t} + \frac{1}{4\delta_1^3} \frac{1-e^{-2\delta_1(T-t)}}{T-t} \right] \\
&\quad + \sigma_{3,3}^2 (T-t) \left[ \frac{1}{2\delta_2^2} + \frac{1}{\delta_2^3} \frac{1-e^{-\delta_2(T-t)}}{T-t} + \frac{1}{4\delta_2^3} \frac{1-e^{-2\delta_2(T-t)}}{T-t} \right] \\
&\quad + \sigma_{4,4}^2 (T-t) \left[ \frac{1}{2\delta_1^2} + \frac{1}{\delta_1^2} e^{-\delta_1(T-t)} - \frac{1}{4\delta_1}(T-t) e^{-2\delta_1(T-t)} - \frac{3}{4\delta_1^2} e^{-2\delta_1(T-t)} \right. \\
&\qquad \left. - \frac{2}{\delta_1^3} \frac{1-e^{-\delta_1(T-t)}}{T-t} + \frac{5}{8\delta_1^3} \frac{1-e^{-2\delta_1(T-t)}}{T-t} \right] \\
&\quad + \sigma_{5,5}^2 (T-t) \left[ \frac{1}{2\delta_2^2} + \frac{1}{\delta_2^2} e^{-\delta_2(T-t)} - \frac{1}{4\delta_2}(T-t) e^{-2\delta_2(T-t)} - \frac{3}{4\delta_2^2} e^{-2\delta_2(T-t)} \right. \\
&\qquad \left. - \frac{2}{\delta_2^3} \frac{1-e^{-\delta_2(T-t)}}{T-t} + \frac{5}{8\delta_2^3} \frac{1-e^{-2\delta_2(T-t)}}{T-t} \right] \\
B_L(t,T) &= -(T-t) \\
B_{S_\ell}(t,T) &= -\frac{1-e^{-\delta_\ell(T-t)}}{\delta_\ell}, \qquad \ell = 1,2 \\
B_{C_\ell}(t,T) &= (T-t)e^{-\delta_\ell(T-t)} - \frac{1-e^{-\delta_\ell(T-t)}}{\delta_\ell}, \qquad \ell = 1,2.
\end{aligned}
\tag{1.2}
$$

In the dependent-factor case, the expressions for $B(t,T)$ are the same as that of the independent-factor case since $\Delta$ is unchanged. However, the expression for $A(t,T)$ becomes considerably more complicated; we refer to Christensen et al. (2009, Appendix).

## 1.3 Arbitrage-Free Nelson-Siegel Model

The factor loadings for the AFNS model with independent factors are given by

$$
\begin{aligned}
A(t,T) &= \sigma_{1,1}^2 \frac{(T-t)^3}{6} + \sigma_{2,2}^2 (T-t)\left[\frac{1}{2\delta^2} - \frac{1}{\delta^3}\frac{1-e^{-\delta(T-t)}}{T-t} + \frac{1}{4\delta^3}\frac{1-e^{-2\delta(T-t)}}{T-t}\right] \\
&\quad + \sigma_{3,3}^2 (T-t)\left[\frac{1}{2\delta^2} + \frac{1}{\delta^2}e^{-\delta(T-t)} - \frac{1}{4\delta}(T-t)\,e^{-2\delta(T-t)} - \frac{3}{4\delta^2}e^{-2\delta(T-t)} \right. \\
&\qquad \left. - \frac{2}{\delta^3}\frac{1-e^{-\delta(T-t)}}{T-t} + \frac{5}{8\delta^3}\frac{1-e^{-2\delta(T-t)}}{T-t}\right], \\
B_L(t,T) &= -(T-t), \\
B_S(t,T) &= -\frac{1-e^{-\delta(T-t)}}{\delta}, \\
B_C(t,T) &= (T-t)\,e^{-\delta(T-t)} - \frac{1-e^{-\delta(T-t)}}{\delta}.
\end{aligned}
\tag{1.3}
$$

In the dependent-factor case, the expressions for $B(t,T)$ are the same as that of the independent-factor case since $\Delta$ is unchanged. However, the expression for $A(t,T)$ becomes considerably more complicated; we refer to Christensen et al. (2011, Appendix B).

## 1.4 Multi-Factor Cox-Ingersoll-Ross Model

The factor loadings for the three-factor CIR model are given by

$$
\begin{aligned}
A(t,T) &= \sum_{k=1}^{3} \frac{2\delta_{k,k}\theta_k^Q}{\sigma_{k,k}^2} \log\left[\frac{2\gamma_k e^{\frac{1}{2}(\delta_{k,k}+\gamma_k)(T-t)}}{(\delta_{k,k}+\gamma_k)(e^{\gamma_k(T-t)}-1)+2\gamma_k}\right] \\
B_k(t,T) &= -\frac{2(e^{\gamma_k(T-t)}-1)}{(\delta_{k,k}+\gamma_k)(e^{\gamma_k(T-t)}-1)+2\gamma_k}, \qquad k=1,2,3,
\end{aligned}
\tag{1.4}
$$

where $\gamma_k := \sqrt{\delta_{k,k}^2 + 2\sigma_{k,k}^2}$. The extension to the four-factor case is straightforward. We do not consider the dependent-factor case in this paper.

# 2 Parameter estimates

Table 2.1: Parameter estimates for the USA dataset

| **Par.** | **Blackburn-Sherris** | | | **AFNS** | | **AFGNS** | | **CIR** | |
|---|---|---|---|---|---|---|---|---|---|
| | **Indep. fact.** | | **Dep. fact.** | **Indep. fact.** | **Dep. fact.** | **Indep. fact.** | **Dep. fact.** | | |
| | 3 | 4 | 3 | 3 | 3 | 5 | 5 | 3 | 4 |
| $\delta_{11}$ | 0.04269 | $-0.01246$ | $-0.01101$ | $-0.06922$ | $-0.06337$ | $-0.07117$ | $-0.08304$ | $-0.22347$ | $-0.13199$ |
| | (0.00127) | (0.00010) | (0.00456) | (0.00003) | (0.00005) | (0.00003) | (0.00024) | (0.00383) | (0.00148) |
| $\delta_{21}$ | – | – | 1.16407 | – | – | – | – | – | – |
| | | | (0.02188) | | | | | | |
| $\delta_{22}$ | -0.03123 | -0.07528 | -0.00518 | – | – | -0.07338 | -0.04983 | 0.23036 | 0.29486 |
| | (0.00031) | (0.00008) | (0.00324) | | | (0.00002) | (0.00040) | (0.01571) | (0.01939) |
| $\delta_{31}$ | – | – | $-0.78085$ | – | – | – | – | – | – |
| | | | (0.01799) | | | | | | |
| $\delta_{32}$ | – | – | $-0.03675$ | – | – | – | – | – | – |
| | | | (0.00231) | | | | | | |
| $\delta_{33}$ | -0.08574 | -0.12354 | -0.07178 | – | – | – | – | -0.13107 | -0.09322 |
| | (8.843e-05) | (0.00012) | (0.00010) | | | | | (0.00286) | (0.00178) |
| $\delta_{44}$ | – | $-0.05468$ | – | – | – | – | – | – | $-0.07894$ |
| | | (0.00012) | | | | | | | (0.00166) |
| $\kappa_1$ | 0.01475 | 0.08000 | 5.20880 | 0.09672 | 0.01784 | 0.08596 | 0.00188 | 0.00111 | 0.07241 |
| | (0.02132) | (0.00183) | (0.06356) | (0.02529) | (0.06373) | (0.00078) | (0.00323) | (6.450e-05) | (0.01194) |
| $\kappa_2$ | -0.00004 | 0.06723 | -0.03927 | -0.00183 | 0.00969 | 0.01794 | -0.00268 | 0.40895 | 0.40756 |
| | (0.00876) | (0.00103) | (0.00765) | (0.00401) | (0.00272) | (0.00002) | (0.00047) | (0.01099) | (0.01757) |
| $\kappa_3$ | 0.01156 | -0.00988 | 0.00320 | 0.08407 | 0.00220 | 0.01058 | -0.00307 | 0.12902 | 1.404e-06 |
| | (0.00229) | (0.00221) | (0.00413) | (0.01348) | (0.02201) | (0.00010) | (0.00075) | (0.03335) | (0.00041) |
| $\kappa_4$ | – | 0.10149 | – | – | – | 0.01040 | -0.00281 | – | 0.01439 |
| | | (0.00181) | | | | (0.00019) | (0.00047) | | (0.00212) |
| $\kappa_5$ | – | – | – | – | – | 0.03536 | -0.00138 | – | – |
| | | | | | | (0.00098) | (0.00267) | | |
| $\sigma_{11}$ | 0.00079 | 0.00170 | 0.00138 | 0.00064 | 0.00889 | 0.00093 | 0.00175 | 0.00260 | 0.00039 |
| | (7.845e-05) | (0.00007) | (6.509e-06) | (4.367e-05) | (9.189e-06) | (0.00006) | (0.00014) | (0.00011) | (7.564e-05) |
| $\sigma_{21}$ | – | – | $-4.466$e-07 | – | $-7.720$e-05 | – | 3.422e-06 | – | – |
| | | | (0.00009) | | (0.01040) | | (0.00008) | | |
| $\sigma_{22}$ | 0.00067 | 0.00180 | 0.00054 | 0.00035 | 0.00869 | 6.651e-07 | 0.00196 | 0.00307 | 0.00258 |
| | (5.845e-05) | (0.00002) | (4.817e-05) | (1.926e-05) | (2.776e-06) | (9.763e-08) | (4.423e-08) | (0.00071) | (0.00034) |
| $\sigma_{31}$ | – | – | 2.592e-07 | – | $-3.173$e-05 | – | $-6.187$e-06 | – | – |
| | | | (0.00053) | | (0.03393) | | (1.793e-03) | | |
| $\sigma_{32}$ | – | – | $-2.226$e-07 | – | 3.101e-05 | – | $-6.937$e-06 | – | – |
| | | | (0.00107) | | (0.00105) | | (0.00004) | | |
| $\sigma_{33}$ | 0.00009 | 0.00009 | 0.00042 | 0.00012 | 0.00357 | 0.00576 | 0.00354 | 0.02146 | 0.01381 |
| | (1.149e-05) | (3.432e-06) | (3.798e-06) | (7.517e-06) | (5.969e-06) | (2.533e-06) | (6.505e-07) | (0.00069) | (0.00042) |
| $\sigma_{44}$ | – | 0.00737 | – | – | – | 0.00002 | 0.00034 | – | 0.03150 |
| | | (1.932e-05) | | | | (8.122e-07) | (7.827e-07) | | (0.00053) |
| $\sigma_{55}$ | – | – | – | – | – | 0.00030 | 0.00841 | – | – |
| | | | | | | (5.587e-06) | (9.331e-06) | | |
| $r_1$ | 2.157e-15 | 1.410e-32 | 3.547e-15 | 2.458e-15 | 7.151e-16 | 1.140e-40 | 4.901e-30 | 7.606e-22 | 1.574e-28 |
| | (5.827e-16) | (1.910e-32) | (1.192e-15) | (1.068e-15) | (2.363e-16) | (3.137e-40) | (8.485e-30) | (8.249e-20) | (2.850e-23) |
| $r_2$ | 0.55467 | 1.31053 | 0.54388 | 0.56463 | 0.57985 | 0.61608 | 1.185 | 0.82391 | 1.119 |
| | (0.00602) | (0.0288) | (0.00748) | (0.00998) | (0.00765) | (0.06063) | (0.03664) | (0.05771) | (0.06026) |
| $r_c$ | 9.494e-08 | 6.591e-08 | 8.081e-08 | 1.044e-07 | 8.611e-08 | 7.554e-08 | 6.627e-08 | 1.499e-07 | 2.611e-08 |
| | (3.377e-09) | (7.919e-10) | (1.689e-09) | (3.699e-09) | (1.764e-09) | (8.256e-10) | (9.911e-10) | (1.304e-08) | (2.136e-09) |
| **CIR $\theta^{\mathbf{P}}$** | $\theta_1^P$ | $\theta_2^P$ | $\theta_3^P$ | $\theta_4^P$ | | | | | |
| 3 Fact. | 0.00519 | 0.00713 | 5.912e-09 | – | | | | | |
| | (0.00031) | (0.00030) | (6.024e-05) | | | | | | |
| 4 Fact. | 0.00035 | 0.00789 | 1.502e-07 | 0.01049 | | | | | |
| | (2.573e-05) | (7.439e-05) | (2.69463) | (0.00094) | | | | | |
| **AFGNS dep.** | $\sigma_{41}$ | $\sigma_{42}$ | $\sigma_{43}$ | $\sigma_{51}$ | $\sigma_{52}$ | $\sigma_{53}$ | $\sigma_{54}$ | | |
| | 5.710e-07 | 6.412e-07 | $-1.163$e-06 | 9.362e-07 | 1.151e-06 | $-1.153$e-06 | $-1.169$e-09 | | |
| | (3.602e-05) | (3.863e-06) | (8.615e-06) | (5.672e-06) | (1.756e-05) | (2.739e-05) | (8.233e-06) | | |

Table 2.2: Parameter estimates for the Australian dataset

| **Par.** | **Blackburn-Sherris** | | | **AFNS** | | **AFGNS** | | **CIR** | |
|---|---|---|---|---|---|---|---|---|---|
| | **Indep. fact.** | | **Dep. fact.** | **Indep. fact.** | **Dep. fact.** | **Indep. fact.** | **Dep. fact.** | | |
| | 3 | 4 | 3 | 3 | 3 | 5 | 5 | 3 | 4 |
| $\delta_{11}$ | 0.11204 | 0.06071 | 0.05432 | −0.04991 | −0.04891 | −0.08269 | −0.08313 | −0.13785 | −0.11850 |
| | (0.00189) | (0.00098) | (0.00325) | (0.00004) | (0.00004) | (0.00005) | (0.00018) | (0.00268) | (0.00280) |
| $\delta_{21}$ | – | – | 0.53855 | – | – | – | – | – | – |
| | | | (0.02205) | | | | | | |
| $\delta_{22}$ | 0.01030 | −0.05963 | 0.05291 | – | – | −0.04968 | −0.04982 | −0.09152 | −0.11543 |
| | (0.00050) | (0.00008) | (0.00194) | | | (0.00003) | (0.00030) | (0.00387) | (0.00609) |
| $\delta_{31}$ | – | – | −0.27095 | – | – | – | – | – | – |
| | | | (0.01759) | | | | | | |
| $\delta_{32}$ | – | – | −0.07973 | – | – | – | – | – | – |
| | | | (0.00153) | | | | | | |
| $\delta_{33}$ | −0.07157 | −0.14399 | −0.07234 | – | – | – | – | −0.08292 | −0.08675 |
| | (0.00004) | (0.00017) | (0.00010) | | | | | (0.00115) | (0.00323) |
| $\delta_{44}$ | – | −0.02189 | – | – | – | – | – | – | 0.35470 |
| | | (0.00025) | | | | | | | (0.03923) |
| $\kappa_1$ | 0.06129 | 0.04099 | 0.07122 | 0.04773 | 0.02331 | 0.00672 | 0.00616 | 0.01266 | 0.13478 |
| | (0.01698) | (0.01468) | (0.01880) | (0.00907) | (0.01867) | (0.00015) | (0.00574) | (0.11629) | (0.03541) |
| $\kappa_2$ | 0.03814 | 0.01583 | 0.03390 | 0.05301 | 0.03381 | 0.00014 | 0.00010 | 0.00186 | 0.00323 |
| | (0.01020) | (0.00383) | (0.01125) | (0.01129) | (0.03420) | (0.00013) | (0.00009) | (0.00079) | (0.00054) |
| $\kappa_3$ | 0.00965 | -0.00757 | 0.02488 | 0.02019 | 0.01095 | −0.00115 | −0.00115 | 7.870e-07 | 0.08643 |
| | (0.00185) | (0.01979) | (0.00691) | (0.00229) | (0.00719) | (0.00030) | (0.00097) | (0.00846) | (0.02999) |
| $\kappa_4$ | – | 0.02641 | – | – | – | −0.00034 | −0.00029 | – | 0.36586 |
| | | (0.00770) | | | | (0.00004) | (0.00033) | | (0.03478) |
| $\kappa_5$ | – | – | – | – | – | 0.00162 | 0.00173 | – | – |
| | | | | | | (0.00006) | (0.00131) | | |
| $\sigma_{11}$ | 0.00105 | 0.00130 | 0.00068 | 0.00096 | 0.00269 | 0.01368 | 0.00486 | 0.00127 | 0.00367 |
| | (0.00009) | (0.00008) | (0.00004) | (0.00005) | (0.00013) | (0.00002) | (0.00071) | (0.00073) | (0.00051) |
| $\sigma_{21}$ | – | – | −1.612e-06 | – | −7.138e-06 | – | −1.478e-06 | – | – |
| | | | (0.00046) | | (0.00478) | | (0.00009) | | |
| $\sigma_{22}$ | 0.00092 | 0.00092 | 0.00254 | 0.00078 | 0.00269 | 0.00145 | 0.00046 | 0.05282 | 0.05607 |
| | (0.00008) | (0.00002) | (0.00003) | (0.00004) | (0.00007) | (0.00004) | (0.00004) | (0.00216) | (0.00218) |
| $\sigma_{31}$ | – | – | 1.263e-06 | – | −3.051e-06 | – | -0.00002 | – | – |
| | | | (0.00272) | | (0.00318) | | (0.00557) | | |
| $\sigma_{32}$ | – | – | −4.873e-06 | – | 3.074e-06 | – | 1.105e-06 | – | – |
| | | | (0.00079) | | (0.00037) | | (0.00080) | | |
| $\sigma_{33}$ | 0.00024 | 0.00003 | 0.00193 | 0.00030 | 0.00115 | 0.00285 | 0.00438 | 0.01044 | 0.01246 |
| | (0.00002) | (9.924e-07) | (0.00001) | (0.00002) | (0.00002) | (0.00003) | (0.00001) | (0.00084) | (0.00109) |
| $\sigma_{44}$ | – | 0.00184 | – | – | – | 0.00001 | 0.00026 | – | 0.01371 |
| | | (0.00004) | | | | (1.215e-06) | (0.00680) | | (0.00220) |
| $\sigma_{55}$ | – | – | – | – | – | 0.00554 | 0.00397 | – | – |
| | | | | | | (0.00004) | (0.00004) | | |
| $r_1$ | 4.672e-13 | 1.165e-12 | 4.711e-13 | 1.215e-15 | 1.189e-15 | 1.454e-14 | 1.768e-15 | 4.303e-19 | 9.634e-14 |
| | (1.235e-13) | (1.092e-12) | (1.469e-13) | (4.425e-16) | (3.885e-16) | (1.654e-14) | (2.425e-15) | (2.732e-17) | (4.158e-12) |
| $r_2$ | 0.43094 | 0.35870 | 0.42991 | 0.56201 | 0.56248 | 0.45856 | 0.50646 | 0.68340 | 0.41813 |
| | (0.00621) | (0.02018) | (0.00727) | (0.00815) | (0.00716) | (0.02480) | (0.02907) | (0.10433) | (0.03920) |
| $r_c$ | 1.298e-07 | 1.124e-07 | 1.272e-07 | 1.985e-07 | 1.833e-07 | 1.078e-07 | 1.070e-07 | 1.904e-07 | 8.183e-08 |
| | (4.213e-09) | (3.622e-09) | (4.305e-09) | (5.654e-09) | (5.428e-09) | (1.821e-09) | (1.788e-09) | (1.545e-08) | (4.992e-09) |
| **CIR** $\theta^{\mathbf{P}}$ | $\theta_1^P$ | $\theta_2^P$ | $\theta_3^P$ | $\theta_4^P$ | | | | | |
| 3 Fact. | 1.276e-09 | 1.814e-02 | 1.080e-08 | – | | | | | |
| | (6.278e-05) | (0.00297) | (8.039e-05) | | | | | | |
| 4 Fact. | 00044 | 0.04216 | 0.00014 | 0.00250 | | | | | |
| | (6.780e-05) | (0.00674) | (0.00096) | (0.00022) | | | | | |
| **AFGNS dep.** | $\sigma_{41}$ | $\sigma_{42}$ | $\sigma_{43}$ | $\sigma_{51}$ | $\sigma_{52}$ | $\sigma_{53}$ | $\sigma_{54}$ | | |
| | 1.140e-06 | −8.819e-08 | −1.010505e-06 | −1.865e-05 | 1.434e-06 | 1.629e-05 | −9.999e-07 | | |
| | (0.28756) | (0.06038) | (0.11084) | (16.91549) | (3.43681) | (2.68759) | (0.46707) | | |

## Table 2.3: Parameter estimates for the Danish dataset

| **Par.** | **Blackburn-Sherris** | | | **AFNS** | | **AFGNS** | | **CIR** | |
|---|---|---|---|---|---|---|---|---|---|
| | **Indep. fact.** | | **Dep. fact.** | **Indep. fact.** | **Dep. fact.** | **Indep. fact.** | **Dep. fact.** | | |
| | 3 | 4 | 3 | 3 | 3 | 5 | 5 | 3 | 4 |
| $\delta_{11}$ | 8.364e-07 | 0.23551 | −0.05330 | −0.08301 | −0.07491 | −0.08248 | −0.08271 | −0.12329 | −0.13324 |
| | (2.286e-06) | (0.00218) | (0.00110) | (0.00003) | (0.00004) | (0.00004) | (0.00006) | (0.00150) | (0.00190) |
| $\delta_{21}$ | – | – | 0.43602 | – | – | – | – | – | – |
| | | | (0.00452) | | | | | | |
| $\delta_{22}$ | −0.05220 | −0.03288 | −0.05259 | – | – | -0.04718 | −0.04934 | 0.00797 | −0.00584 |
| | (0.00004) | (0.00009) | (0.00053) | | | (0.00002) | (0.00015) | (0.00244) | (0.00342) |
| $\delta_{31}$ | – | – | -0.21455 | – | – | – | – | – | – |
| | | | (0.00358) | | | | | | |
| $\delta_{32}$ | – | – | 0.01697 | – | – | – | – | – | – |
| | | | (0.00038) | | | | | | |
| $\delta_{33}$ | −0.10132 | −0.09814 | −0.06072 | – | – | – | – | −0.08836 | −0.11211 |
| | (0.00001) | (0.00005) | (0.00006) | | | | | (0.00300) | (0.00316) |
| $\delta_{44}$ | – | 0.04056 | – | – | – | – | – | – | −0.11386 |
| | | (0.00025) | | | | | | | (0.00257) |
| $\kappa_1$ | 0.01146 | 0.11510 | 0.03309 | 0.00915 | 0.01385 | 0.00046 | −0.00993 | 0.03829 | 0.11431 |
| | (0.00843) | (0.03052) | (0.00896) | (0.00495) | (0.00961) | (0.00024) | (0.00204) | (0.01242) | (0.03884) |
| $\kappa_2$ | 0.06877 | 0.06832 | 0.01035 | 0.01068 | 0.00351 | −0.00222 | −0.00317 | 0.04984 | 0.02932 |
| | (0.01224) | (0.01720) | (0.00319) | (0.00591) | (0.01528) | (0.00026) | (0.00065) | (0.00958) | (0.00318) |
| $\kappa_3$ | 0.00498 | 0.00495 | 0.00970 | 0.00744 | 0.00302 | −0.00179 | −0.00115 | 0.04859 | 0.02569 |
| | (0.00559) | (0.00529) | (0.00273) | (0.00690) | (0.01207) | (0.00062) | (0.00128) | (0.00725) | (0.00552) |
| $\kappa_4$ | – | 0.02618 | – | – | – | −0.00185 | −0.00067 | – | 0.00002 |
| | | (0.01396) | | | | (0.00027) | (0.00077) | | (2.033e-05) |
| $\kappa_5$ | – | – | – | – | – | -0.00238 | -0.00997 | – | – |
| | | | | | | (0.00053) | (0.00280) | | |
| $\sigma_{11}$ | 0.00109 | 0.00204 | 0.00070 | 0.00066 | 0.00319 | 0.01031 | 0.00955 | 0.00535 | 0.00532 |
| | (0.00003) | (0.00010) | (0.00006) | (0.00003) | (0.00012) | (0.00004) | (0.00089) | (0.00015) | (0.00015) |
| $\sigma_{21}$ | – | – | −1.330e-06 | – | −8.198e-06 | – | −7.790e-06 | – | – |
| | | | (0.00078) | | (0.00082) | | (0.00010) | | |
| $\sigma_{22}$ | 0.00112 | 0.00142 | 0.00208 | 0.00053 | 0.00261 | 0.00137 | 0.00262 | 0.00424 | 0.00369 |
| | (0.00001) | (0.00004) | (0.00007) | (0.00002) | (0.00001) | (0.00001) | (0.00016) | (0.00052) | (0.00054) |
| $\sigma_{31}$ | – | – | 1.232e-06 | – | −3.630e-06 | – | -0.00008 | – | – |
| | | | (0.00337) | | (0.00112) | | (0.00167) | | |
| $\sigma_{32}$ | – | – | -3.834e-06 | – | 2.994e-06 | – | 1.083e-06 | – | – |
| | | | (0.00105) | | (0.00016) | | (0.00047) | | |
| $\sigma_{33}$ | 0.00052 | 0.00055 | 0.00189 | 0.00021 | 0.00116 | 0.00377 | 0.00929 | 0.02590 | 0.03122 |
| | (2.914e-06) | (3.398e-06) | (0.00007) | (5.212e-06) | (7.761e-06) | (0.00003) | (0.00002) | (0.00219) | (0.00155) |
| $\sigma_{44}$ | – | 0.00218 | – | – | – | 0.00018 | 0.00097 | – | 0.01261 |
| | | (0.00007) | | | | (4.185e-06) | (0.00011) | | (0.00046) |
| $\sigma_{55}$ | – | – | – | – | – | 0.00545 | 0.01162 | – | – |
| | | | | | | (0.00013) | (1.200e-07) | | |
| $r_1$ | 3.716e-15 | 2.350e-13 | 3.098e-15 | 2.416e-15 | 3.385e-15 | 4.468e-15 | 1.044e-14 | 4.887e-16 | 1.975e-15 |
| | (1.099e-15) | (5.560e-14) | (8.814e-16) | (7.644e-16) | (9.980e-16) | (1.631e-15) | (2.957e-15) | (1.860e-15) | (3.942e-14) |
| $r_2$ | 0.54312 | 0.45143 | 0.54693 | 0.55369 | 0.54440 | 0.53604 | 0.51683 | 0.58006 | 0.54856 |
| | (0.00665) | (0.00553) | (0.00622) | (0.00698) | (0.00651) | (0.00781) | (0.00635) | (0.03173) | (0.03167) |
| $r_c$ | 1.796e-07 | 9.973e-08 | 1.729e-07 | 1.877e-07 | 1.816e-07 | 1.288e-07 | 1.176e-07 | 1.469e-07 | 1.279e-07 |
| | (3.365e-09) | (2.044e-09) | (1.887e-09) | (3.544e-09) | (3.264e-09) | (1.444e-09) | (1.223e-09) | (1.593e-08) | (1.395e-08) |
| **CIR $\theta^{\mathbf{P}}$** | $\theta_1^P$ | $\theta_2^P$ | $\theta_3^P$ | $\theta_4^P$ | | | | | |
| **3 Fact.** | 4.915e-18 | 0.00510 | 0.00332 | – | | | | | |
| | (2.099e-08) | (0.00056) | (0.00046) | | | | | | |
| **4 Fact.** | 0.00065 | 0.00527 | 0.00221 | 2.701e-12 | | | | | |
| | (0.00011) | (0.00049) | (0.00051) | (6.478e-07) | | | | | |
| **AFGNS.dep.** | $\sigma_{41}$ | $\sigma_{42}$ | $\sigma_{43}$ | $\sigma_{51}$ | $\sigma_{52}$ | $\sigma_{53}$ | $\sigma_{54}$ | | |
| | 6.915e-06 | 2.941e-07 | −8.019e-06 | −9.898e-05 | 1.068e-05 | 9.694e-05 | −9.968e-06 | | |
| | (0.00280) | (0.00137) | (0.00190) | (6.48895) | (0.55735) | (3.10011) | (0.92826) | | |

Table 2.4: Parameter estimates for the E&W civilian population dataset

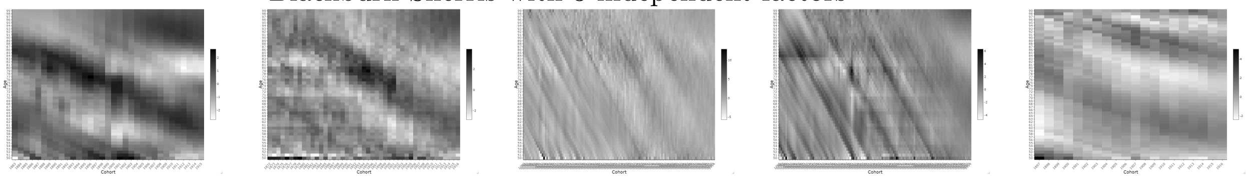
| Par. | Blackburn-Sherris | | | AFNS | | AFGNS | | CIR | |
|---|---|---|---|---|---|---|---|---|---|
| | Indep. fact. | | Dep. fact. | Indep. fact. | Dep. fact. | Indep. fact. | Dep. fact. | | |
| | 3 | 4 | 3 | 3 | 3 | 5 | 5 | 3 | 4 |
| $\delta_{11}$ | 0.04017 | 0.01264 | -0.02247 | -0.06154 | -0.05159 | -0.08184 | -0.08337 | -0.11253 | -0.11122 |
| | (0.00037) | (0.00020) | (0.00160) | (0.00003) | (0.00003) | (0.00002) | (0.00008) | (0.00135) | (0.00125) |
| $\delta_{21}$ | – | – | 0.76411 | – | – | – | – | – | – |
| | | | (0.00706) | | | | | | |
| $\delta_{22}$ | -0.03819 | -0.08277 | -0.01988 | – | – | -0.04924 | -0.04927 | -0.00708 | 0.00988 |
| | (0.00013) | (0.00002) | (0.00065) | | | (0.00001) | (0.00012) | (0.00325) | (0.00122) |
| $\delta_{31}$ | – | – | -0.35646 | – | – | – | – | – | – |
| | | | (0.00553) | | | | | | |
| $\delta_{32}$ | – | – | -0.00214 | – | – | – | – | – | – |
| | | | (0.00046) | | | | | | |
| $\delta_{33}$ | -0.08624 | -0.18177 | -0.05992 | – | – | – | – | -0.09758 | -0.07582 |
| | (0.00003) | (0.00014) | (0.00008) | | | | | (0.00458) | (0.00268) |
| $\delta_{44}$ | – | -0.04412 | – | – | – | – | – | – | -0.13049 |
| | | (0.00008) | | | | | | | (0.00239) |
| $\kappa_1$ | 0.01804 | 0.01406 | 0.02721 | 0.01174 | 0.00529 | 0.00101 | 0.00709 | 0.05367 | 1.040e-08 |
| | (0.00903) | (0.00630) | (0.01005) | (0.00477) | (0.00608) | (0.00008) | (0.00440) | (0.00335) | (4.370e-10) |
| $\kappa_2$ | 0.03410 | 0.00833 | 0.00687 | 0.02245 | -0.00008 | -0.00231 | -0.00740 | 0.01871 | 0.01670 |
| | (0.01608) | (0.00175) | (0.00319) | (0.01326) | (0.02362) | (0.00009) | (0.00065) | (0.00786) | (0.00360) |
| $\kappa_3$ | 0.00622 | 0.03690 | 0.00592 | 0.00743 | 0.00167 | -0.00268 | -0.00903 | 0.39703 | 0.26070 |
| | (0.00521) | (0.00732) | (0.00171) | (0.00440) | (0.00888) | (0.00020) | (0.00103) | (0.09355) | (0.02845) |
| $\kappa_4$ | – | 0.01777 | – | – | – | -0.00233 | -0.00670 | – | 2.387e-15 |
| | | (0.00751) | | | | (0.00009) | (0.00063) | | (6.928e-10) |
| $\kappa_5$ | – | – | – | – | – | -0.00185 | -0.00488 | – | – |
| | | | | | | (0.00003) | (0.00229) | | |
| $\sigma_{11}$ | 0.00097 | 0.00104 | 0.00072 | 0.00115 | 0.00699 | 0.02058 | 0.00552 | 0.00953 | 0.00829 |
| | (0.00005) | (0.00003) | (0.00005) | (0.00003) | (0.00029) | (0.00003) | (0.00042) | (0.00022) | (7.934e-05) |
| $\sigma_{21}$ | – | – | -2.321e-06 | – | -0.00005 | – | -2.301e-06 | – | – |
| | | | (0.01263) | | (0.02047) | | (0.00006) | | |
| $\sigma_{22}$ | 0.00101 | 0.00040 | 0.00338 | 0.00100 | 0.00757 | 0.00156 | 0.00186 | 0.00524 | 0.01004 |
| | (0.00004) | (0.00001) | (0.00055) | (0.00003) | (0.00007) | (0.00004) | (0.00013) | (0.00073) | (0.00062) |
| $\sigma_{31}$ | – | – | 1.629e-06 | – | -0.00003 | – | -0.00003 | – | – |
| | | | (0.00587) | | (0.01558) | | (0.00101) | | |
| $\sigma_{32}$ | – | – | -7.883e-06 | – | 0.00003 | – | -1.970e-06 | – | – |
| | | | (0.00135) | | (0.00183) | | (0.00045) | | |
| $\sigma_{33}$ | 0.00094 | 3.523e-06 | 0.00235 | 0.00074 | 0.00452 | 0.00295 | 0.00549 | 0.03227 | 0.03254 |
| | (3.445e-06) | (1.387e-07) | (0.00005) | (9.124e-06) | (0.00004) | (0.00004) | (2.303e-06) | (0.001923) | (0.00077) |
| $\sigma_{44}$ | – | 0.00125 | – | – | – | 0.00004 | 0.00062 | – | 0.01426 |
| | | (0.00002) | | | | (1.945e-06) | (0.00018) | | (0.00044) |
| $\sigma_{55}$ | – | – | – | – | – | 0.00574 | 0.00897 | – | – |
| | | | | | | (0.00011) | (0.00128) | | |
| $r_1$ | 3.159e-11 | 4.832e-10 | 3.514e-11 | 1.894e-11 | 3.685e-11 | 3.878e-10 | 6.518e-10 | 1.867e-11 | 3.887e-12 |
| | (6.681e-12) | (8.392e-11) | (8.382e-12) | (4.372e-12) | (7.303e-12) | (9.728e-11) | (1.289e-10) | (1.890e-11) | (3.881e-12) |
| $r_2$ | 0.32370 | 0.22870 | 0.31842 | 0.33537 | 0.31709 | 0.23723 | 0.22176 | 0.32445 | 0.34321 |
| | (5.046e-03) | (4.113e-03) | (5.680e-03) | (5.273e-03) | (4.753e-03) | (5.749e-03) | (4.767e-03) | (0.03006) | (0.03513) |
| $r_c$ | 1.656e-07 | 1.538e-07 | 1.608e-07 | 1.728e-07 | 1.622e-07 | 1.276e-07 | 1.167e-07 | 1.550e-07 | 1.449e-07 |
| | (3.506e-11) | (3.157e-11) | (3.484e-11) | (3.752e-11) | (3.629e-11) | (1.712e-11) | (1.754e-11) | (1.164e-08) | (6.903e-09) |
| **CIR $\theta^{\mathbf{P}}$** | $\theta_1^P$ | $\theta_2^P$ | $\theta_3^P$ | $\theta_4^P$ | | | | | |
| 3 Fact. | 0.00518 | 0.00151 | 0.00009 | – | | | | | |
| | (0.00043) | (0.00143) | (0.00073) | | | | | | |
| 4 Fact. | 23075.34809 | 0.00480 | 0.00051 | 6.516e-30 | | | | | |
| | (110.22) | (0.00080) | (4.474e-05) | (0.14078) | | | | | |
| **AFGNS.dep.** | $\sigma_{41}$ | $\sigma_{42}$ | $\sigma_{43}$ | $\sigma_{51}$ | $\sigma_{52}$ | $\sigma_{53}$ | $\sigma_{54}$ | | |
| | 2.092e-06 | 3.924e-07 | -2.702e-06 | -3.829e-05 | 3.823e-06 | 0.00003 | -4.607e-06 | | |
| | (0.00302) | (0.00241) | (0.00278) | (103.02721) | (7.63012) | (82.22063) | (17.49291) | | |

Table 2.5: Parameter estimates for the Japanese dataset

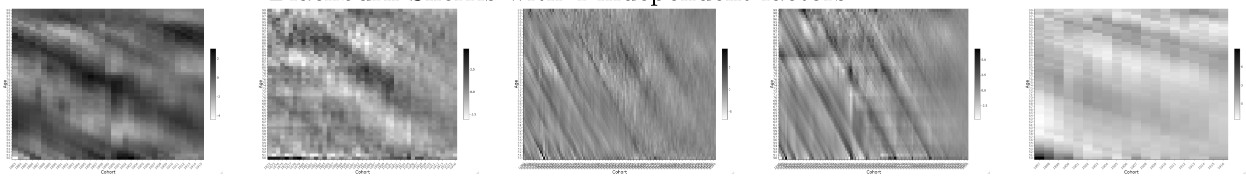
| **Par.** | **Blackburn-Sherris** | | | **AFNS** | | **AFGNS** | | **CIR** | |
|---|---|---|---|---|---|---|---|---|---|
| | **Indep. fact.** | | **Dep. fact.** | **Indep. fact.** | **Dep. fact.** | **Indep. fact.** | **Dep. fact.** | | |
| | 3 | 4 | 3 | 3 | 3 | 5 | 5 | 3 | 4 |
| $\delta_{11}$ | -0.07423 | 0.01713 | -0.07882 | -0.09489 | -0.09444 | -0.08245 | -0.08284 | -0.11439 | -0.11832 |
| | (0.00002) | (0.00079) | (0.00065) | (0.00002) | (0.00001) | (0.00019) | (0.00223) | (0.00291) | (0.00120) |
| $\delta_{21}$ | – | – | 0.02280 | – | – | – | – | – | – |
| | | | (0.00133) | | | | | | |
| $\delta_{22}$ | 0.25703 | -0.03913 | -3.417e-07 | – | – | -0.04973 | -0.04994 | 0.62391 | 1.28506 |
| | (0.01170) | (0.00014) | (0.00111) | | | (7.596e-06) | (0.00134) | (0.03265) | (0.00659) |
| $\delta_{31}$ | – | – | -0.00684 | – | – | – | – | – | – |
| | | | (0.00043) | | | | | | |
| $\delta_{32}$ | – | – | -0.04925 | – | – | – | – | – | – |
| | | | (0.00053) | | | | | | |
| $\delta_{33}$ | -0.15143 | -0.07665 | -0.12597 | – | – | – | – | -0.10646 | -0.08796 |
| | (0.00013) | (0.00004) | (0.00025) | | | | | (0.00769) | (0.00066) |
| $\delta_{44}$ | – | 0.00598 | – | – | – | – | – | – | -0.03500 |
| | | (0.00060) | | | | | | | (0.00288) |
| $\kappa_1$ | 0.02986 | -0.08308 | 0.03705 | -0.00373 | 0.00145 | 0.12492 | 0.16089 | 0.00034 | 0.02815 |
| | (0.00248) | (0.00041) | (0.00029) | (0.00918) | (0.00284) | (0.00020) | (0.00372) | (2.797e-05) | (0.00539) |
| $\kappa_2$ | 0.09056 | -0.07310 | -0.03760 | 0.03908 | 0.03975 | 0.05296 | 0.07584 | 0.63792 | 1.29611 |
| | (0.04171) | (0.00019) | (0.01139) | (0.00333) | (0.00274) | (0.00000) | (0.00110) | (0.01032) | (0.01618) |
| $\kappa_3$ | 0.02603 | -0.01406 | 0.03802 | 0.07654 | 0.09926 | 0.04126 | 0.05894 | 0.00000 | 0.18922 |
| | (0.00268) | (0.00825) | (0.03373) | (0.01613) | (0.00266) | (0.00003) | (0.00147) | (5.002e-11) | (0.01207) |
| $\kappa_4$ | – | -0.08047 | – | – | – | 0.04273 | 0.06016 | – | 0.11300 |
| | | (0.00007) | | | | (0.00010) | (0.00143) | | (0.00960) |
| $\kappa_5$ | – | – | – | – | – | 0.07282 | 0.10161 | – | – |
| | | | | | | (0.00004) | (0.00105) | | |
| $\sigma_{11}$ | 0.00013 | 0.00021 | 0.00047 | 0.00017 | 0.00038 | 0.00115 | 0.00237 | 0.00809 | 0.00521 |
| | (0.00001) | (0.00002) | (0.00004) | (0.00002) | (0.00010) | (0.00008) | (0.00014) | (0.00102) | (8.203e-05) |
| $\sigma_{21}$ | – | – | -2.184e-07 | – | -9.503e-08 | – | -8.148e-07 | – | – |
| | | | (0.00025) | | (0.00006) | | (1.998e-06) | | |
| $\sigma_{22}$ | 0.00030 | 3.106e-07 | 0.00053 | 0.00013 | 0.00029 | 1.492e-07 | 0.00389 | 0.00922 | 0.02903 |
| | (0.00003) | (7.783e-08) | (0.00003) | (5.965e-06) | (4.024e-06) | (3.233e-08) | (0.00001) | (0.00114) | (0.00242) |
| $\sigma_{31}$ | – | – | 1.222e-07 | – | -4.693e-08 | – | -1.345e-06 | – | – |
| | | | (0.00071) | | (0.00061) | | (0.00003) | | |
| $\sigma_{32}$ | – | – | -1.489e-07 | – | 3.843700e-08 | – | -0.00001 | – | – |
| | | | (0.00028) | | (0.00014) | | (0.00019) | | |
| $\sigma_{33}$ | 0.00004 | 0.00077 | 0.00028 | 0.00006 | 0.00013 | 0.00334 | 0.00436 | 0.03237 | 0.00631 |
| | (2.962e-07) | (7.455e-06) | (1.835e-06) | (1.922e-06) | (3.100e-06) | (0.00001) | (0.00004) | (0.00102) | (0.00016) |
| $\sigma_{44}$ | – | 8.740e-07 | – | – | – | 0.00006 | 0.00079 | – | 0.04087 |
| | | (2.028e-07) | | | | (7.377e-06) | (1.905e-07) | | (0.00124) |
| $\sigma_{55}$ | – | – | – | – | – | 0.00580 | 0.00208 | – | – |
| | | | | | | (5.054e-06) | (1.369e-06) | | |
| $r_1$ | 2.649e-34 | 3.791e-34 | 5.661e-34 | 1.455e-34 | 1.533e-33 | 2.746e-21 | 1.892e-21 | 1.676e-23 | 6.430e-11 |
| | (2.083e-33) | (1.610e-33) | (5.352e-33) | (1.067e-33) | (1.258e-31) | (2.155e-21) | (2.025e-21) | (2.149e-13) | (1.492e-11) |
| $r_2$ | 1.32529 | 1.32896 | 1.29393 | 1.33389 | 1.37580 | 0.82927 | 0.83461 | 0.86033 | 0.23442 |
| | (0.15181) | (0.08752) | (0.16549) | (0.15077) | (0.12502) | (0.01716) | (0.02337) | (0.24213) | (0.00771) |
| $r_c$ | 8.04133 | 7.59980 | 8.20018 | 8.44065 | 8.38675 | 0.70797 | 0.73568 | 4.56694 | 0.64562 |
| | (2.784e-09) | (8.670e-10) | (8.847e-10) | (2.936e-09) | (9.206e-10) | (1.199e-10) | (1.605e-10) | (6.046e-09) | (3.598e-10) |
| **CIR** $\theta^{\mathbf{P}}$ | $\theta_1^P$ | $\theta_2^P$ | $\theta_3^P$ | $\theta_4^P$ | | | | | |
| 3 Fact. | 0.68899 | 0.00731 | 13.28658 | – | | | | | |
| | (0.02920) | (0.00017) | (9.845e+05) | | | | | | |
| 4 Fact. | 0.00079 | 0.00550 | 0.00000 | 0.00122 | | | | | |
| | (9.287e-05) | (1.750e-05) | (2.894e-11) | (6.804e-05) | | | | | |
| AFGNS.dep. | $\sigma_{41}$ | $\sigma_{42}$ | $\sigma_{43}$ | $\sigma_{51}$ | $\sigma_{52}$ | $\sigma_{53}$ | $\sigma_{54}$ | | |
| | 2.450e-07 | 3.007e-06 | -2.957e-06 | -4.284e-06 | 3.611e-06 | -1.744e-07 | 3.993e-07 | | |
| | (0.00049) | (0.00369) | (0.00015) | (0.15452) | (0.06615) | (0.05625) | (0.01435) | | |

# 3 Standardized Residuals

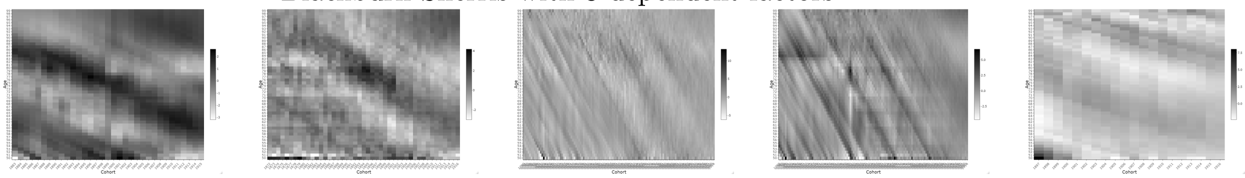
Blackburn-Sherris with 3 independent factors



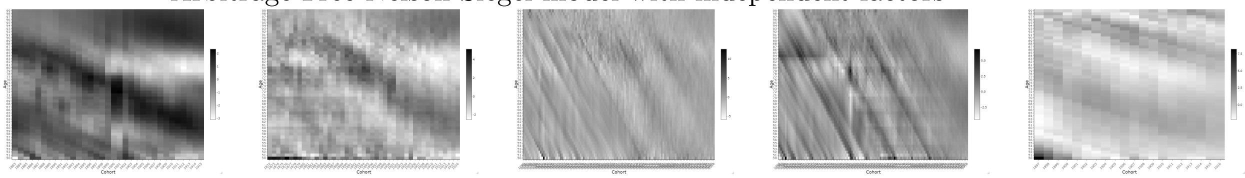
Blackburn-Sherris with 4 independent factors



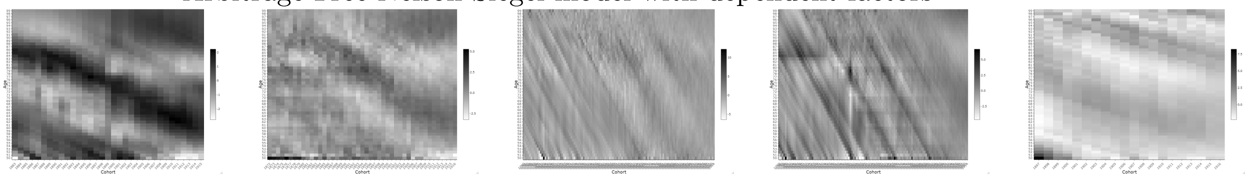
Blackburn-Sherris with 3 dependent factors



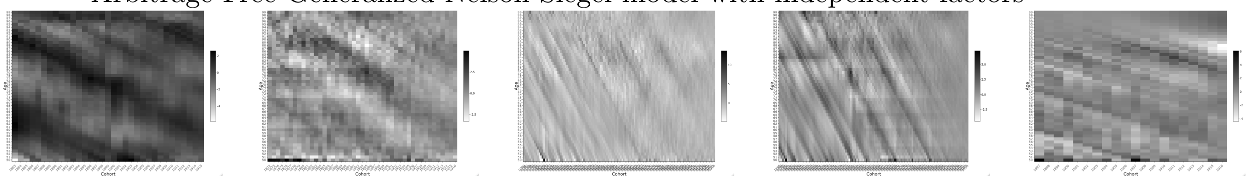
Arbitrage-Free Nelson-Siegel model with independent factors



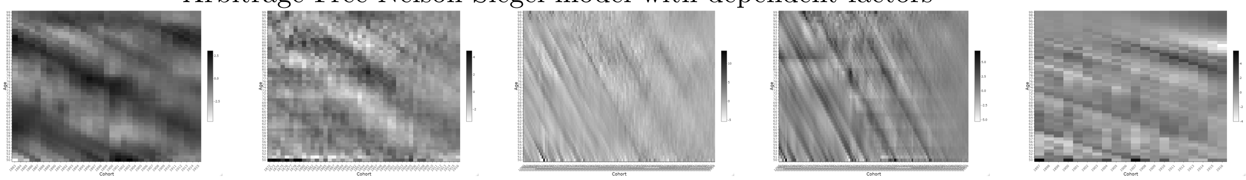
Arbitrage-Free Nelson-Siegel model with dependent factors



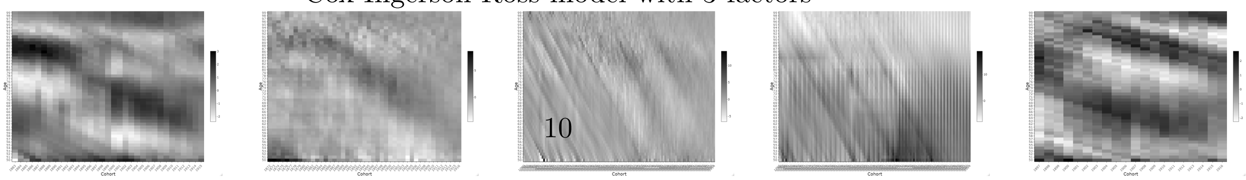
Arbitrage-Free Generalized Nelson-Siegel model with independent factors



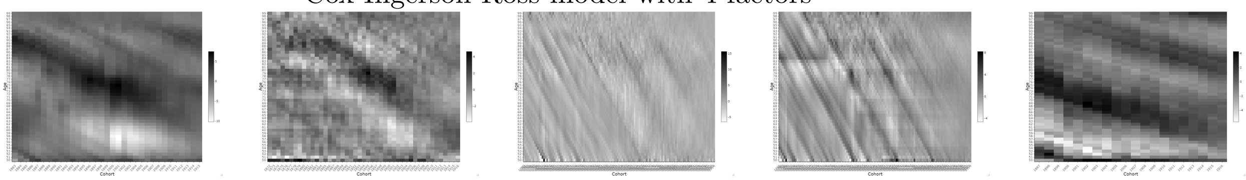
Arbitrage-Free Nelson-Siegel model with dependent factors



Cox-Ingersoll-Ross model with 3 factors



Cox-Ingersoll-Ross model with 4 factors



(a) USA (b) Austr. (c) Denm. (d) E&W (e) Japan

## 4 Plots of factor loadings and of $X(t)$ for the robustness analysis of affine mortality models.

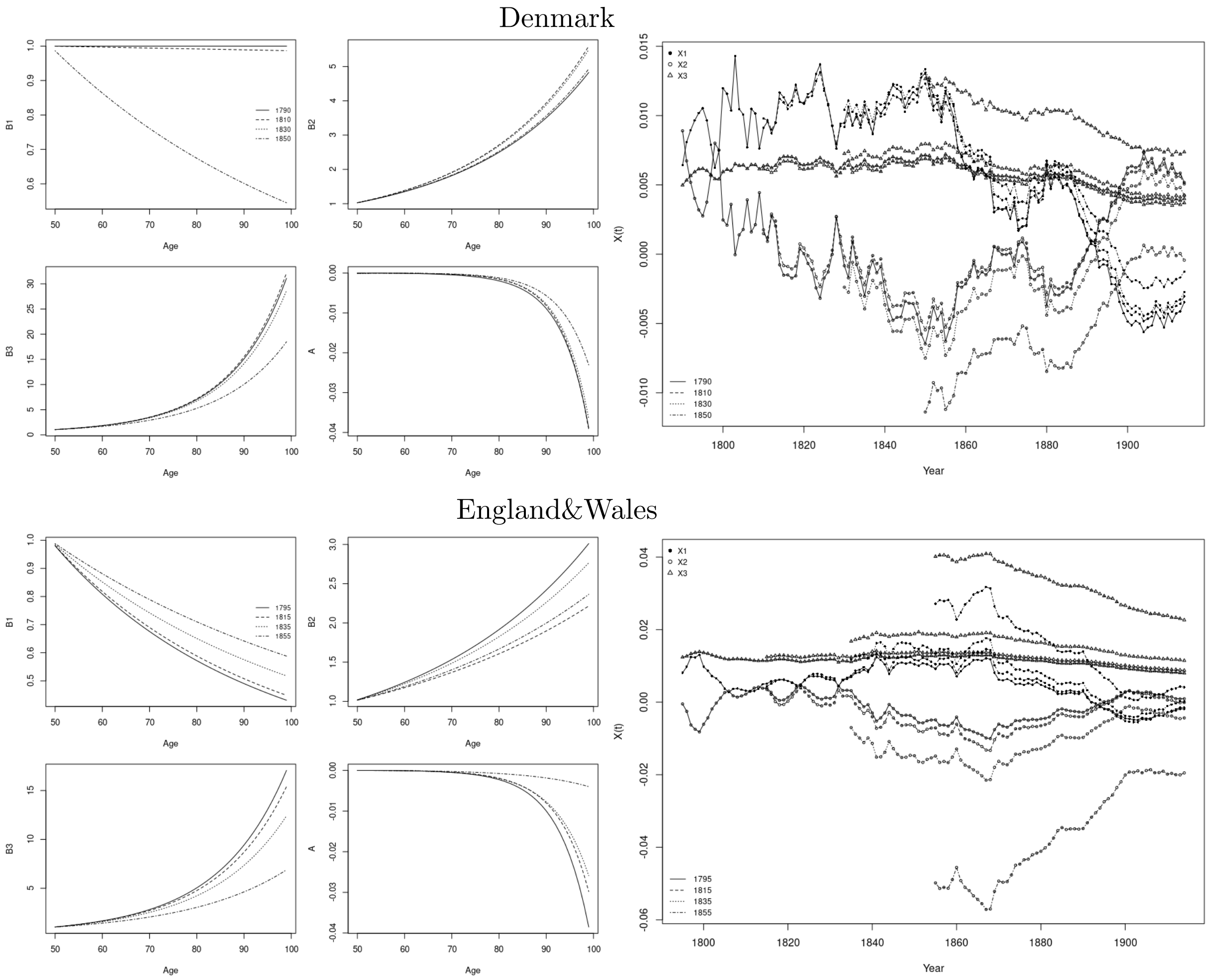



Figure 4.1: Factor loadings (left) and factor values $X(t)$ (right) for the Blackbun-Sherris model with three independent factors for the Denmark (top) and the E&W datasets (bottom).

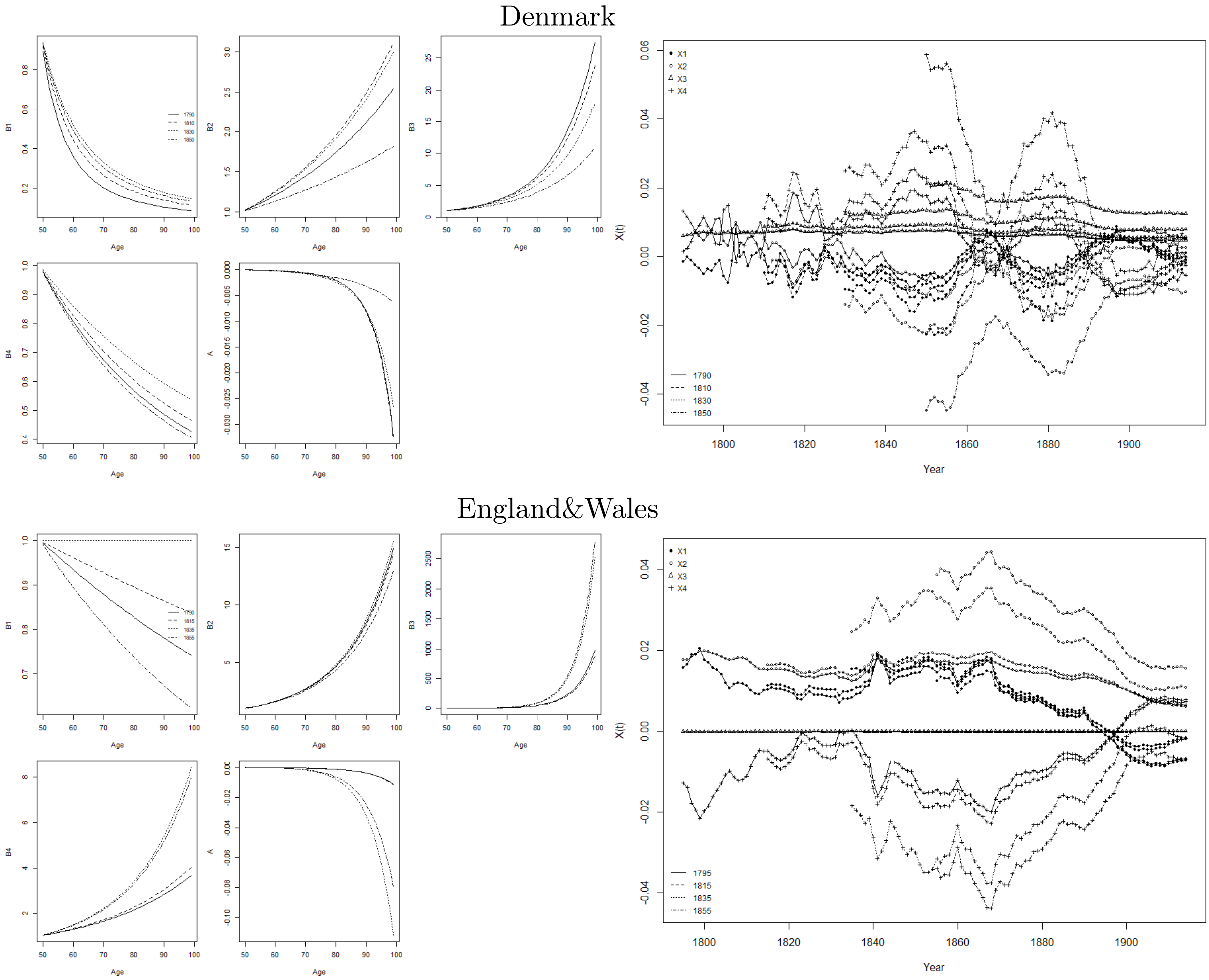



Figure 4.2: Factor loadings (left) and factor values $X(t)$ (right) for the Blackbun-Sherris model with four independent factors for the Denmark (top) and the E&W datasets (bottom).

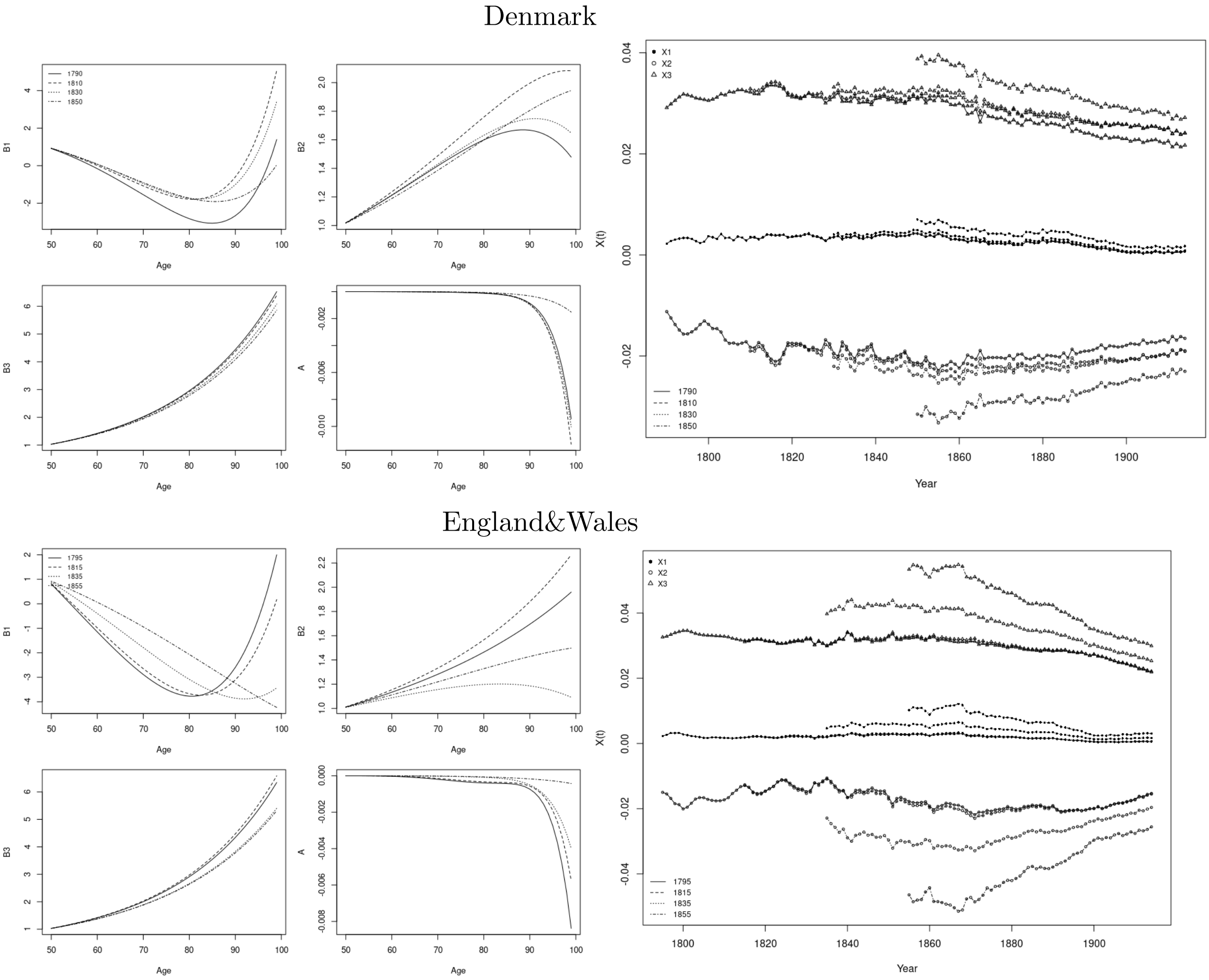



Figure 4.3: Factor loadings (left) and factor values $X(t)$ (right) for the Blackbun-Sherris model with dependent factors for the Denmark (top) and the E&W datasets (bottom).

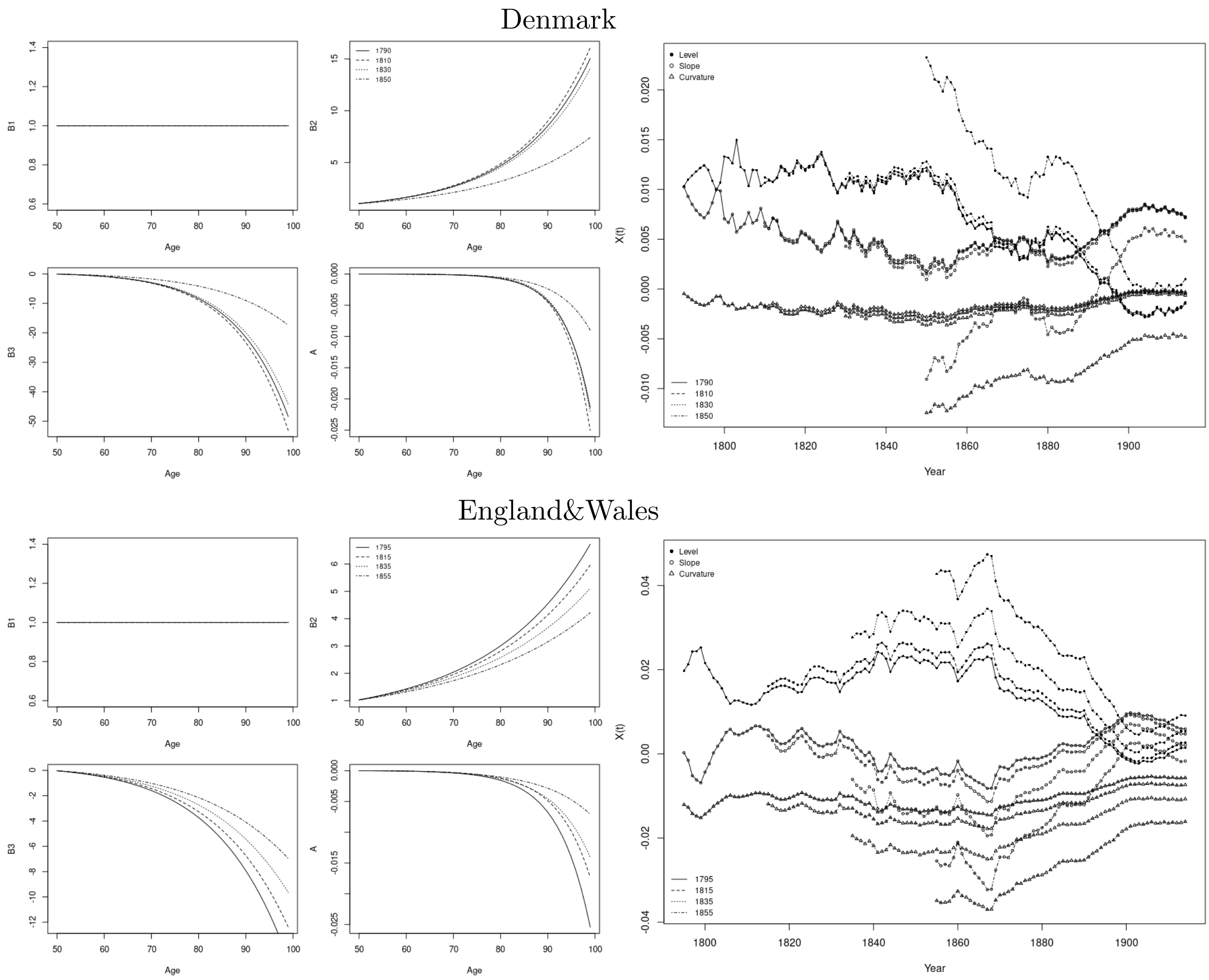



Figure 4.4: Factor loadings (left) and factor values $X(t)$ (right) for the AFNS model with independent factors for the Denmark (top) and the E&W datasets (bottom).

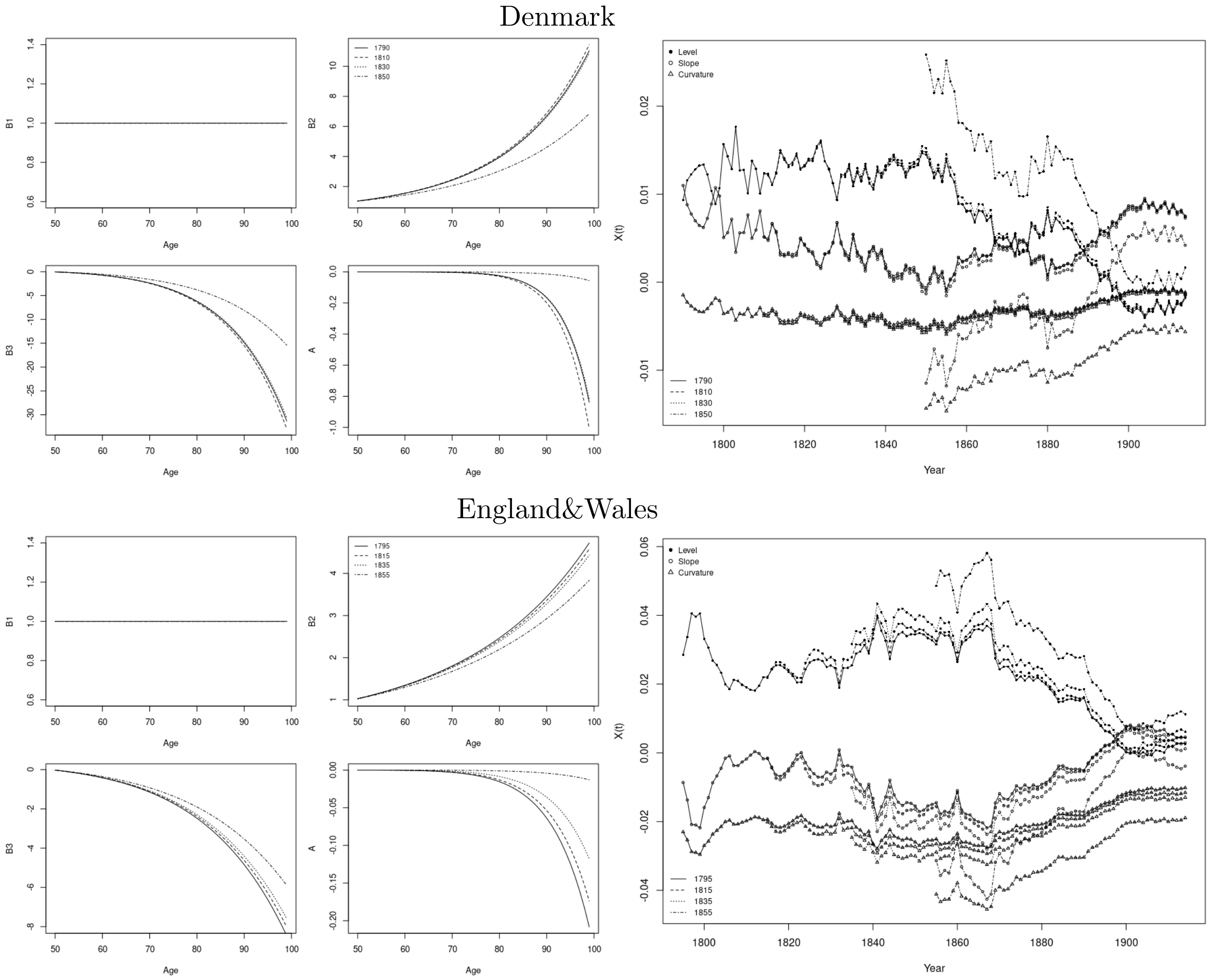



Figure 4.5: Factor loadings (left) and factor values $X(t)$ (right) for the AFNS model with dependent factors for the Denmark (top) and the E&W datasets (bottom).

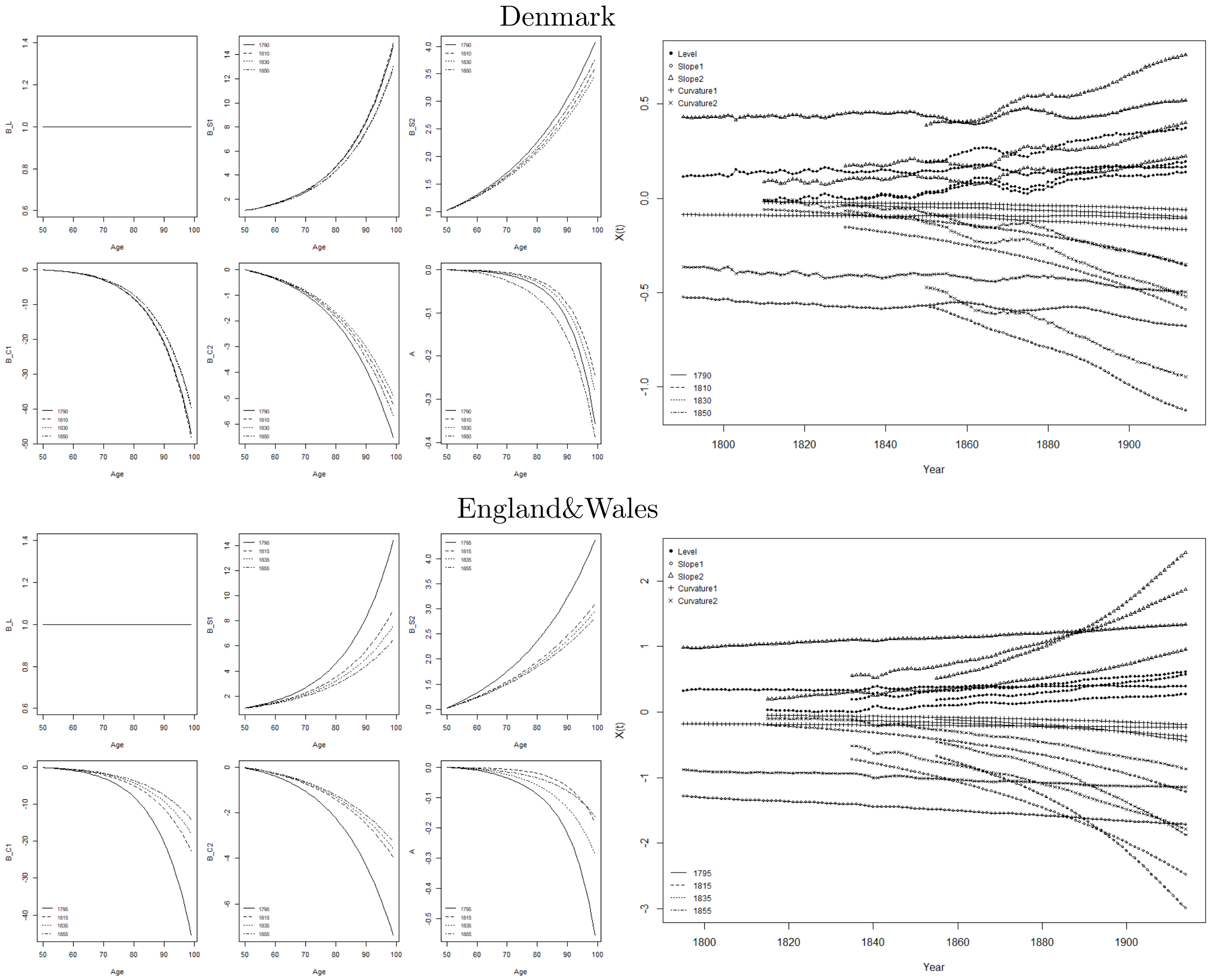



Figure 4.6: Factor loadings (left) and factor values $X(t)$ (right) for the AFGNS model with independent factors for the Denmark (top) and the E&W datasets (bottom).

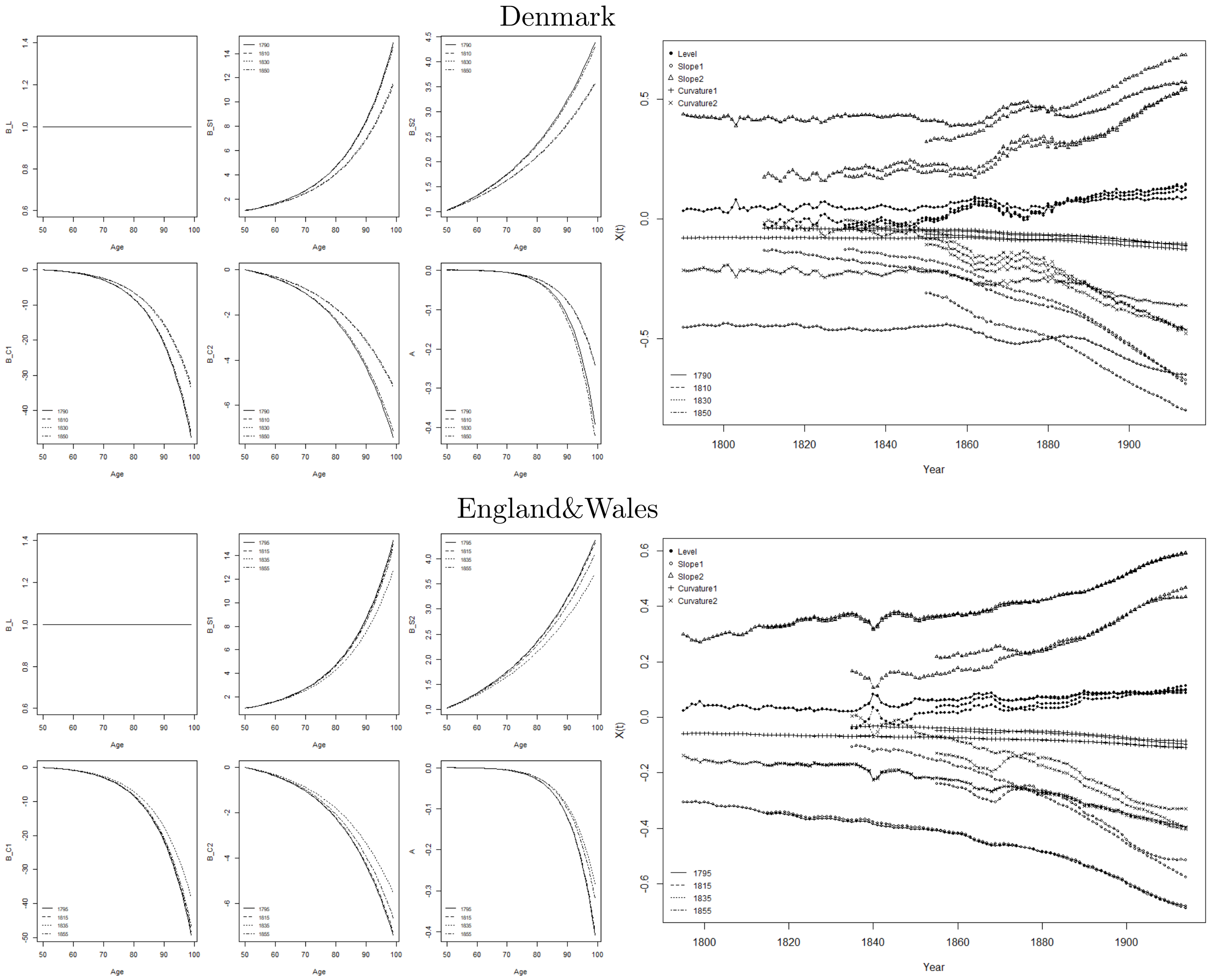



Figure 4.7: Factor loadings (left) and factor values $X(t)$ (right) for the AFGNS model with dependent factors for the Denmark (top) and the E&W datasets (bottom).

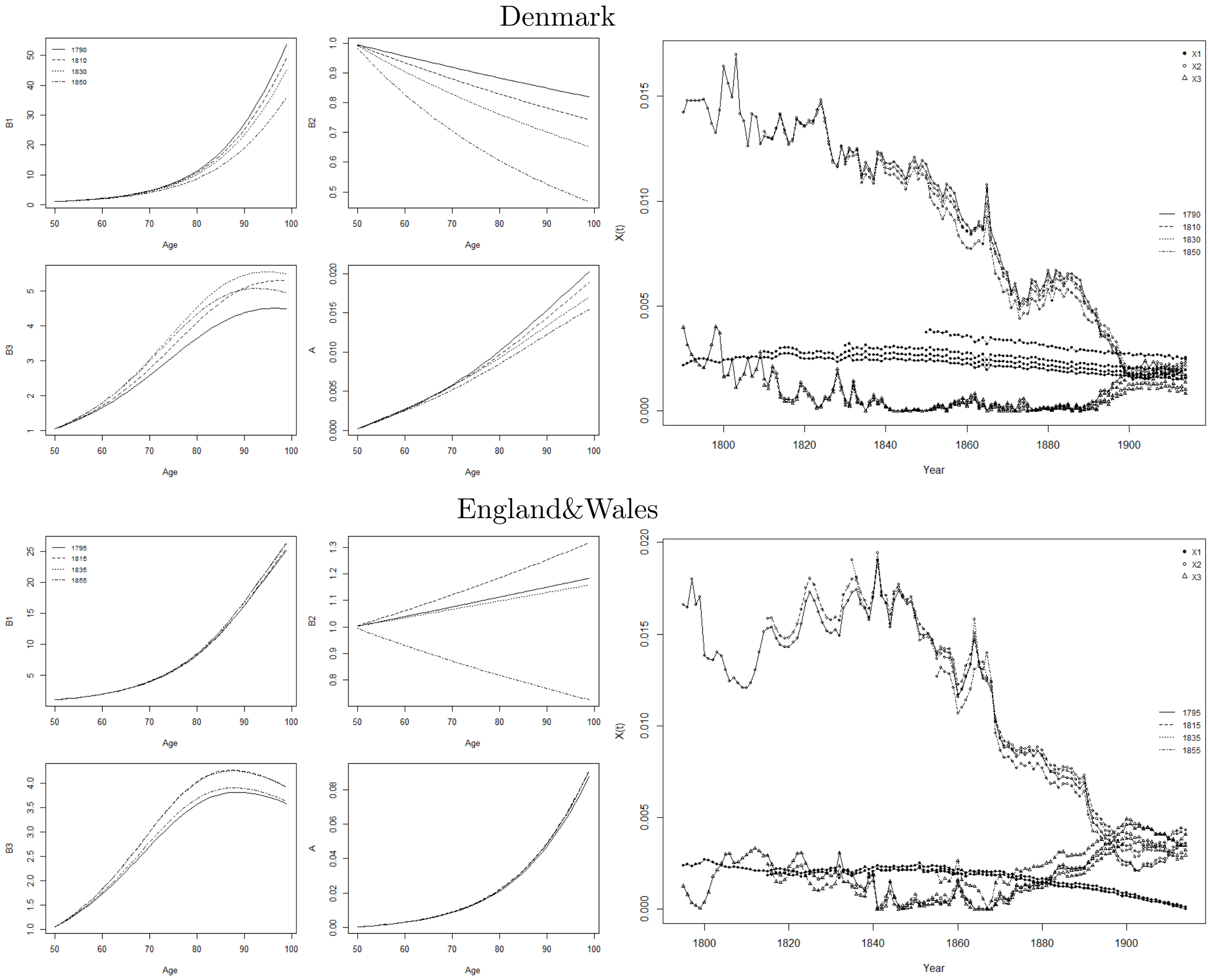



Figure 4.8: Factor loadings (left) and factor values $X(t)$ (right) for the CIR model with three factors for the Denmark (top) and the E&W datasets (bottom).

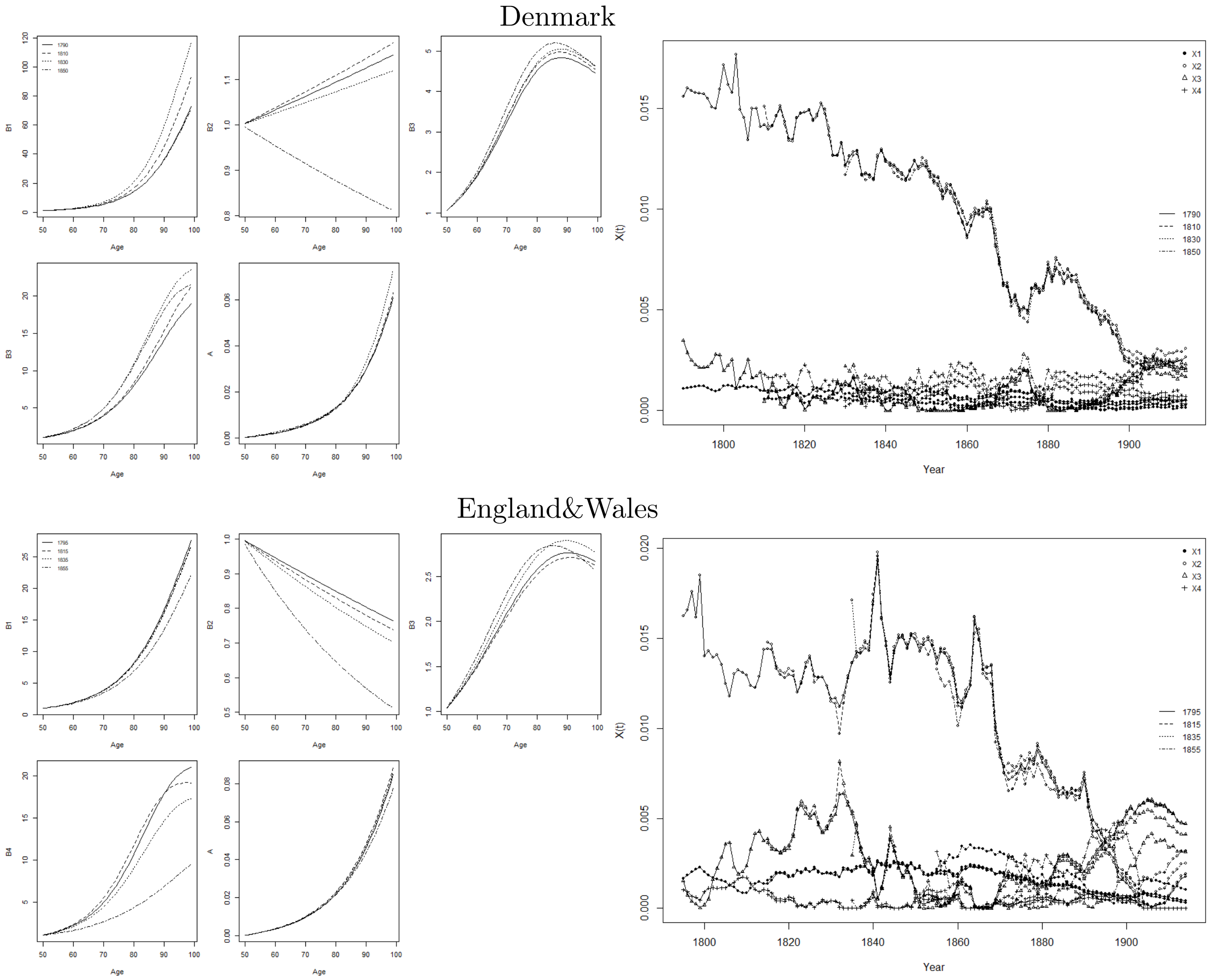



Figure 4.9: Factor loadings (left) and factor values $X(t)$ (right) for the CIR model with four factors for the Denmark (top) and the E&W datasets (bottom).

# 5 Plot of projected survival curves and their Mean Absolute Percentage Error (MAPE)

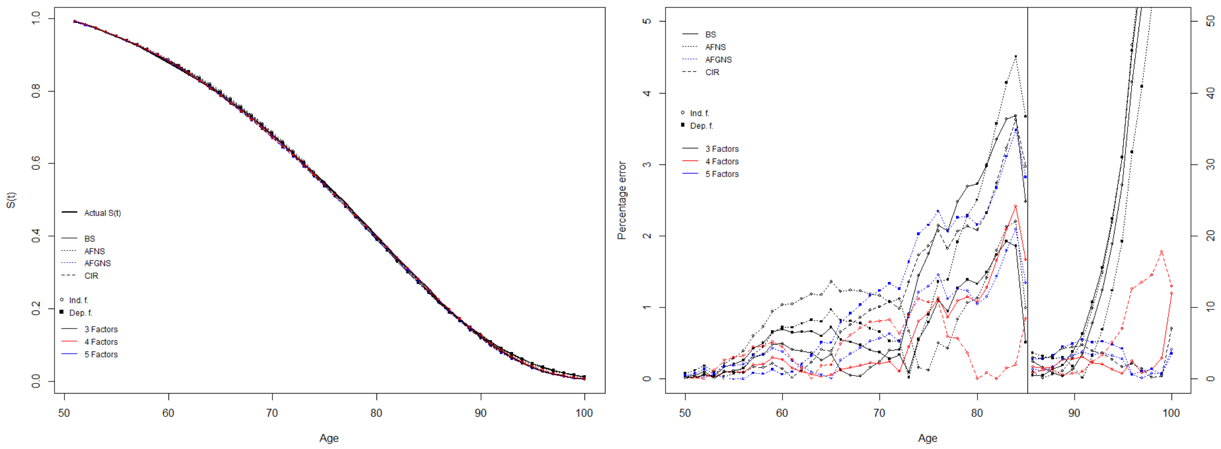



Figure 5.1: Projection of the survival curves for the Australian male cohort born in 1917 under each model (left) and their MAPE (right).

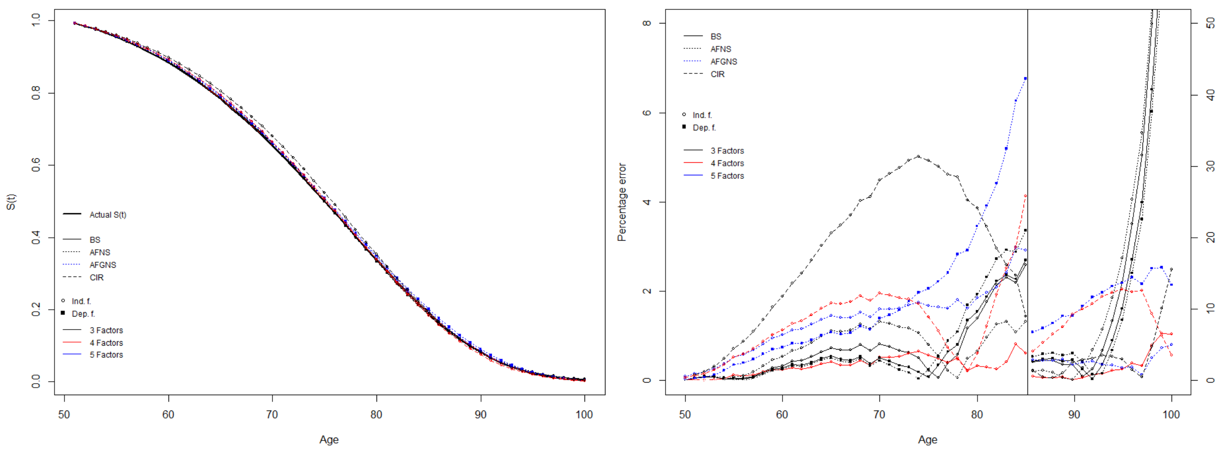



Figure 5.2: Projection of the survival curves for the England & Wales male cohort born in 1915 under each model (left) and their MAPE (right).

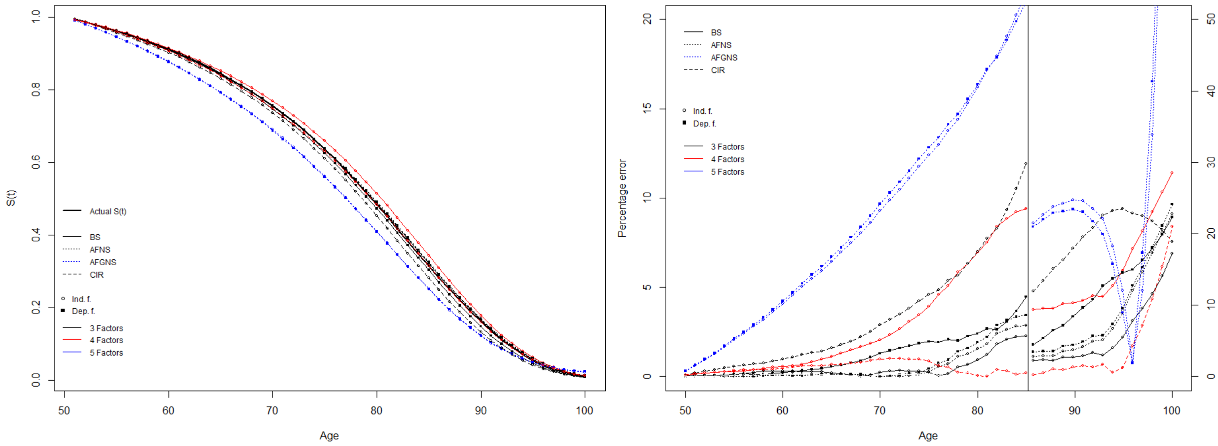



Figure 5.3: Projection of the survival curves for the Japanese male cohort born in 1917 under each model (left) and their MAPE (right).

## 6 Parameter estimation: illustration with R

The estimation of continuous-time affine mortality models follows along the code skeleton illustrated below. We show two examples: the Blackburn-Sherris model with independent factors, and the AFNS model with dependent factors. Further details about the use of the code and of the functions therein are available in Ungolo et al. (2021).

```
<readFunctions>
<defineVariables>
<loadData>
<prepareData>
<fitModelCoordAscent>
<improveEstLocalSearch>
```

<readFunctions> We load the the package which is needed for reading the data (we use HMDHFDplus by Riffe (2015)) and read the functions which need to be optimized ('Est_fun.R') by means of Coordinate Ascent ('Coordinate_Ascent.R') and its subsequent improvement by local search ('Full_opt.R'):

```
<readFunctions>:
source('Est_fun.R')
source('Coordinate_Ascent.R')
source('Full_opt.R')
```

<defineVariables> and <loadData> wrap the code for defining the variables and load the dataset from the IMD. <prepareData> wraps the code which is needed to produce the working dataset of average mortality rates used for fitting the models (not shown here).

We now illustrate the remaining wrappers for the Blackburn-Sherris model with independent factors and for the AFNS model with dependent factors.

## 6.1 Blackburn Sherris with independent factors

The parameter estimation process for the Blackburn-Sherris model with independent factors requires the specification of the number of factors, which is given by the length of the parameter vector $\kappa = (\kappa_1, \kappa_2, \ldots)$.

Hence, `<fitModelCoordAscent>` wraps the code which uses the function `co_asc_BSi`, which performs the optimization process by iteratively optimizing the negative log-likelihood function by group of parameters. This process is repeated for a maximum number of iterations `max_iter` (set by the researcher, default set to 200) by using as starting values the parameter estimates obtained at the previous iteration, until the log-likelihood function increase is below a tolerance level (`tol_lik`, set by the researcher, default value of 0.1). The researcher should also specify the starting value for each parameter vector (default values are also provided).

```
<fitModelCoordAscent>:
pe_CA_BSi <- co_asc_BSi(mu_bar=dataset, max_iter=200, tol_lik=0.1)
```

If we use the default starting values used in our code, by running `pe_CA_BSi <- co_asc_BSi()` this yields the following output when fitted using the US dataset of this work:

```
> pe_CA_BSi
$par_est
$par_est$x0
       x0_1        x0_2        x0_3
0.001551705 0.005618262 0.007011683

$par_est$delta
    delta_1     delta_2     delta_3
 0.04268782 -0.03122758 -0.08573677

$par_est$kappa
      kappa_1       kappa_2       kappa_3
 1.475362e-02 -4.096367e-05  1.156081e-02

$par_est$sigma
     sigma_1      sigma_2      sigma_3
7.941997e-04 6.671747e-04 9.359528e-05

$par_est$r1
          r1
2.156668e-15

$par_est$r2
       r2
0.5546705

$par_est$rc
          rc
9.494266e-08
```

```
$log_lik
log_lik.
10638.34
```

In addition, the function `co_asc_BSi` provides also a table (not shown here) containing the value of the parameters and of the log-likelihood function at each iteration.

### 6.2 AFNS with dependent factors

For this model, the number of factors is set to three by default. `<fitModelFullOptim>` is as follows:

```
<fitModelCoordAscent>:
fit_AFNSi_CA <- co_asc_AFNSd(mu_bar=dataset, x0=c(0,0,0),
delta=-0.05, kappa=c(0.01,0.01,0.01), sigma_dg=c(0.01,0.01,0.01),
Sigma_cov=c(0,0,0), r=c(1e-5,0.3,1e-5), max_iter=200, tol_lik=0.1)
```

The starting values of $\Sigma$ are supplied separately, based on whether they are the terms on the diagonals (`sigma_dg`, corresponding to $\sigma_L$, $\sigma_S$ and $\sigma_C$, i.e. the standard deviations, not the variances), or its off-diagonal elements, the covariances, provided in the following order: $\sigma_{LS}$, $\sigma_{LC}$ and $\sigma_{SC}$.

The output of this function has the same structure described for the Blackburn-Sherris independent factor model implementation. An analogous reasoning applies for the other R functions.

## References


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*

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*

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*

Ungolo, F., Garces, L. P. D. M., Sherris, M. & Zhou, Y. (2021), '`affine_mortality`: A github repository for estimation and analysis of affine mortality models', *Working paper* .