



ARC Centre of Excellence in Population Ageing Research

Working Paper 2022/12

Affine Mortality Models with Jumps: Parameter Estimation and Forecasting

Len Patrick Dominic M. Garces, Jovana Kolar, Michael Sherris, and Francesco Ungolo

This paper can be downloaded without charge from the ARC Centre of Excellence in Population Ageing Research Working Paper Series available at www.cepar.edu.au

Affine Mortality Models with Jumps: Parameter Estimation and Forecasting

Len Patrick Dominic M. Garces ^{*a}, Jovana Kolar^a, Michael Sherris^{a, b}, and Francesco Ungolo^{b, c}

^aARC Centre of Excellence in Population Ageing Research (CEPAR), University of New South Wales, Kensington NSW 2052, Australia

^bSchool of Risk and Actuarial Studies, UNSW Business School, University of New South Wales, Kensington NSW 2052, Australia

^cLehrstuhl für Finanzmathematik, Technische Universität München, 85748 Garching bei München, Germany

September 16, 2022

Abstract

In this paper, we investigate the dynamics of age-cohort survival curves under the assumption that the instantaneous mortality intensity is driven by an affine jump-diffusion (AJD) process. Advantages of an AJD specification of mortality dynamics include the availability of closed-form expressions for survival probabilities afforded by an affine mortality specification and the ease with which we can incorporate sudden positive and negative shocks in mortality dynamics, reflecting events such as wars, pandemics, and medical advancements. As we are interested in modelling the evolution of mortality rates, we propose a state-space approach to calibrate the parameters of the affine mortality process. This ensures consistent survival curves in the sense that forecasts of survival probabilities have the same parametric form as the fitted survival curves. As the resulting state-space model is non-Gaussian due to the presence of jumps, we apply and assess a particle filter-based Markov chain Monte Carlo approach to estimate the model parameters. We illustrate our methodology by fitting one-factor Cox-Ingersoll-Ross and Blackburn-Sherris mortality models with asymmetric double exponential jumps to historical age-cohort mortality data from USA. We find that these one-factor models with jumps have good in-sample fit, but their forecasting performance suggests the need for additional latent factors to improve the accuracy of forecasts.

JEL CLASSIFICATION CODES: C12, C32, G22, J11

KEY WORDS. Affine mortality models, affine jump-diffusion, age-cohort mortality rates, particle filter, particle Markov chain Monte Carlo

*E-mail addresses: LPDM Garces (corresponding author), 1.garces@unsw.edu.au; J Kolar, j.kolar@unsw.edu.au; M Sherris, m.sherris@unsw.edu.au; F Ungolo, f.ungolo@unsw.edu.au

1 Introduction

In addition to financial risk, life insurance companies also face longevity risk, the risk that a population survives longer than expected. While improvements in life expectancy is overall a positive outcome, from an insurer’s perspective, understanding systematic longevity risk is crucial to maintaining long-run solvency while offering guaranteed lifetime income products.¹ Insurers can manage longevity risk in various ways: product re-design, risk-sharing arrangements, hedging, and reinsurance.² Regardless of which risk management strategy an insurer chooses, there is a need to understand the evolution of mortality rates to quantify longevity risk.

It is widely accepted that mortality rates evolve over time in a stochastic manner. While several trends in mortality rates, including (1) an increase in life expectancy and reduction in mortality at younger ages, (2) the rectangularization phenomenon (the concentration of the modal age of death at older ages), and (3) the expansion phenomenon (shift of the Lexis point to older ages), have been observed at the aggregate level across countries (Ebeling et al. 2018; Macdonald et al. 1998; Wilmoth and Horiuchi 1999), the rates of improvement in mortality still varied substantially over time and across age groups (Cairns et al. 2006). Thus, stochastic mortality models have been proposed, ranging from time-series models to models based on short-rate models in finance; see for example Booth and Tickle (2008) and Cairns et al. (2009) for a survey and comparison of these models.

Furthermore, while historical mortality data suggest a steady improvement in human life expectancy worldwide over the years since World War II, the COVID-19 pandemic has shown mortality rates can rapidly increase as a result of unexpected phenomena. In an analysis of short-term mortality fluctuations data for the year 2020, Regis and Jevtić (2022) show that there has been a sharp increase in the weekly measured total death rate compared to the average over the period 2015-2019. As the long-term effects of COVID-19 are still under investigation, it is currently unclear how mortality rates will evolve over time and how the pandemic will affect future trends in mortality. Thus, there is a greater need to better understand the evolution of mortality rates given that catastrophic shocks may occur, albeit rarely.

In this work, we study the dynamics of age-cohort survival curves under the assumption that the instantaneous mortality intensity is driven by an affine jump-diffusion process. Advantages of an affine jump diffusion specification of mortality dynamics include the availability of closed-form expressions for survival probabilities afforded by an affine mortality specification (see e.g. Duffie et al. 2000) and the ease with which we can incorporate sudden positive and negative shocks in mortality dynamics, reflecting events such as wars, pandemics, and medical advancements. Furthermore, we are interested in age-cohort mortality as age-cohort data is more well-suited to pricing longevity-linked financial and insurance products. In addition, the use of continuous-time models, such as the affine jump-diffusion model, affords greater synergy

¹We distinguish between systematic and unsystematic longevity risk, since the latter, which represents the randomness of deaths within a given portfolio, is diversifiable.

²See, for example, Bravo and Nunes (2021), Cairns et al. (2006), D’Amato et al. (2018), Fung et al. (2019), and the references therein for other longevity risk (and mortality risk) management strategies.

with existing insurance or financial pricing models which are typically set in continuous-time (see e.g. [Biffis 2005](#); [Jevtić et al. 2013](#); [Xu et al. 2020b](#); [Zhou et al. 2021](#)). This paper follows the stream of research on modelling mortality rates, particularly the intensity-based approach using affine continuous-time stochastic processes as the mortality intensity.³

We contribute to the literature in the following aspects. First, we model the evolution of age-cohort mortality rates by assuming the instantaneous mortality intensity is an affine jump diffusion process. In this regard, we extend prior work that focuses only on affine diffusion models for the mortality intensity and analyze the evolution of mortality rates across cohorts, in contrast to earlier work that focus on fitting the survival curve of a single cohort. Second, to estimate the model parameters, we formulate a non-Gaussian state-space representation of the average force of mortality and develop a particle filter-based Markov chain Monte Carlo (PMCMC) parameter estimation method. The PMCMC method proposed in this paper can also be used to estimate the parameters of affine mortality models driven by square-root processes, which are typically estimated using quasi-linear Kalman filtering.⁴ Finally, we assess our mortality models using USA age-cohort mortality rates and show that one-factor affine mortality models with jumps have an in-sample fit that is comparable to three-factor affine diffusion models. We also find that although forecasts are robust to changes in model parameter values, our models' forecasting performance suggest the need for additional latent factors to improve the accuracy of forecasts.

[Bravo and Nunes \(2021\)](#), [Luciano and Vigna \(2008\)](#), and [Russo et al. \(2011\)](#) specify the instantaneous mortality intensity as an affine jump-diffusion process.⁵ However, they focus on a single cohort or a single fixed age when calibrating their models. In particular, [Bravo and Nunes \(2021\)](#) and [Luciano and Vigna \(2008\)](#) calibrate their models by minimizing the sum-of-squared errors between the survival curve implied by the mortality model and the observed survival curve. Similarly, [Russo et al. \(2011\)](#) calibrate their models against mortality rates implied by term assurance premiums for a given fixed age of the policyholder. As a result, the model parameters change depending on the reference cohort considered. In contrast, we assume an affine jump diffusion process to model the age-cohort mortality rates across several cohorts within a pre-specified age range. Thus, our approach allows us to characterize the evolution of mortality rates and to forecast mortality rates for future cohorts.

Furthermore, we extend the affine diffusion model specification of [Blackburn and Sherris \(2013\)](#); [Jevtić et al. \(2013\)](#); [Huang et al. \(2022\)](#); [Ungolo et al. \(2021\)](#) and incorporate jump dynamics in the instantaneous mortality intensity. In [Section 2](#), we specify a general affine jump-diffusion model for a latent process driving the instantaneous mortality intensity, derive the exponential affine representation of survival probabilities, and discuss change-of-measure results

³A non-exhaustive list of earlier work on continuous-time stochastic mortality models is as follows: [Biffis \(2005\)](#); [Blackburn and Sherris \(2013\)](#); [Bravo and Nunes \(2021\)](#); [Dahl \(2004\)](#); [Huang et al. \(2022\)](#); [Jevtić et al. \(2013\)](#); [Jevtić and Regis \(2019, 2021\)](#); [Luciano and Vigna \(2008\)](#); [Milevsky and Promislow \(2001\)](#); [Russo et al. \(2011\)](#); [Schrager \(2006\)](#); [Ungolo et al. \(2021\)](#); [Xu et al. \(2020a,b\)](#); [Zhu and Bauer \(2022\)](#).

⁴See for example [Huang et al. \(2022\)](#); [Jevtić and Regis \(2021\)](#); [Ungolo et al. \(2021\)](#).

⁵[Regis and Jevtić \(2022\)](#) proposed an extension of their multi-population mortality model ([Jevtić and Regis 2019](#)) to account for jumps via an affine jump diffusion process, but do not provide further details on its implementation.

required to transition from the risk-neutral probability setting to the real-world probability space.

To estimate the parameters of our model, we formulate a state-space model where the measurement equation is given by the affine representation of the age-cohort average force of mortality and the state transition equation is given by a discretization of the continuous-time affine jump diffusion model for the latent factors. In contrast to the single-cohort approach of [Bravo and Nunes \(2021\)](#) and [Luciano and Vigna \(2008\)](#), our approach produces consistent survival curves in the sense that forecasts of survival probabilities have the same parametric form as the fitted survival curves ([Blackburn and Sherris 2013](#)). This also allows us to explicitly include a time-varying measurement error term, whose variance we specify to have a parametric form that captures the exponentially increasing variation in mortality rates at older ages.⁶

Parameter estimation in this setting is more complicated since the presence of jumps in the mortality intensity process implies that the state-transition equation is non-Gaussian. This rules out the use of the Kalman filter which can only be applied to linear Gaussian state-space models.⁷ We thus propose a particle Markov chain Monte Carlo (MCMC) approach which allows us to estimate the posterior distribution of the model parameters when the state-space model is neither linear nor Gaussian ([Andrieu et al. 2010](#); [Dahlin and Schön 2019](#)); this includes the case when the state transition model is a jump-diffusion process (see e.g. [Johannes et al. 2009](#)). In Section 4, we formulate the state-space model and provide a contextualized overview of the particle filter and the particle Metropolis-Hastings algorithm. The specific configuration of these general algorithms to suit our analysis is discussed in Section 4.4.

In this paper, we focus on the one-factor Cox-Ingersoll-Ross and Blackburn-Sherris mortality models with jumps. Given the relatively simple model specification, we elaborate on the parameter estimation and forecasting methodology, which can then be extended to the case of multiple factors in a straightforward manner. Through an analysis of in-sample fit and forecasting performance, we also assess whether a one-factor model with jumps is sufficient or if additional factors are required. We illustrate our methodology using yearly age-cohort mortality rates for males aged 50 to 99 born in the USA in the period from 1883 to 1915 (i.e. cohorts with complete data). We also discuss forecasts of future age-cohort survival curves using our mortality models and conduct a sensitivity analysis to numerically determine the effect of individual parameters on the shape of the survival curve forecasts. The results of our parameter estimation, forecasting, and sensitivity analysis results are discussed in Section 5. We summarize and conclude our investigation in Section 6.

We emphasize that, under our model formulation, jumps manifest as cohort effects rather than period effects. That is, mortality shocks affect individuals through their birth-year. Furthermore, our formulation assumes that mortality shocks are not transitory, as the mortality intensity of future cohorts is affected by the mortality intensity of previous cohorts which may

⁶See [Blackburn and Sherris \(2013\)](#) for similar comments in the age-period model setting.

⁷See for example [Blackburn and Sherris \(2013\)](#); [Huang et al. \(2022\)](#); [Jevtić et al. \(2013\)](#); [Jevtić and Regis \(2019\)](#); [Ungolo et al. \(2021\)](#); [Xu et al. \(2020a\)](#) for the implementation of the Kalman filter to estimate the model parameters under an affine diffusion specification of the latent factors.

have had a jump event. Such an approach makes sense in the event of positive mortality improvements (e.g. advances in medical treatments) which has a non-transitory effect on mortality rates or large-scale geo-political conflicts which may have long-lasting adverse effects on life expectancy.

2 Model

In this section, we specify our one-factor affine mortality models with jumps. Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{Q})$ be a filtered probability space where $\mathbb{F} = \{\mathcal{F}_t\}_{t \geq 0}$ is a \mathbb{Q} -complete, right-continuous filtration under \mathbb{Q} . As we are interested in pricing applications, we assume that $(\Omega, \mathcal{F}, \mathbb{Q})$ corresponds to an arbitrage-free market and that the probability measure \mathbb{Q} can be used for pricing basic financial and insurance products. This approach ensures a consistent pricing property for the fitted survival probabilities; see, for example, [Biffis \(2005, Section 5\)](#) or [Blackburn and Sherris \(2013, Definition 1\)](#). We consider a finite time horizon $T > 0$ and we set $\mathcal{F} = \mathcal{F}_T$.

We assume there exist \mathbb{Q} -complete and right-continuous sub-filtrations $\mathbb{G} = \{\mathcal{G}_t\}_{t \geq 0}$ and $\mathbb{H} = \{\mathcal{H}_t\}_{t \geq 0}$ of \mathbb{F} such that $\mathcal{F}_t := \mathcal{G}_t \vee \mathcal{H}_t$, the smallest σ -algebra containing \mathcal{G}_t and \mathcal{H}_t .

Without loss of generality, we consider a newborn whose random lifetime is denoted by τ . Let $H := \{H_t\}_{t \geq 0}$ be the process given by $H_t := \mathbf{1}_{\{\tau \leq t\}}$, $t \geq 0$ ($\mathbf{1}_{\{\cdot\}}$ is the indicator function) and consider the natural filtration \mathbb{H} generated by H , i.e. $\mathbb{H} := \{\mathcal{H}_t\}_{t \geq 0}$, where $\mathcal{H}_t := \sigma(H_s : s \leq t)$. Hence, \mathbb{H} is the smallest filtration such that τ is an \mathbb{H} -stopping time. We refer to [Biffis \(2005\)](#) and [Blackburn and Sherris \(2013\)](#) for further details on the stochastic mortality intensity modelling framework.

2.1 Affine Jump-Diffusion Mortality Intensity Process

Let $X := \{X_t\}_{t \geq 0}$ denote an affine time-homogeneous jump-diffusion process satisfying the stochastic differential equation

$$dX_t = \xi(\eta - X_{t-}) dt + \sigma \sqrt{\gamma + \delta X_{t-}} dW_t + dJ_t, \quad X_0 = x_0 \quad (1)$$

where $\xi, \eta \in \mathbb{R}$, $x_0, \sigma, \gamma, \delta \geq 0$, W is a standard Brownian motion under \mathbb{Q} , and J_t is a compound Poisson process, given by

$$J_t := \sum_{i=1}^{N_t} Z_i,$$

where N is a Poisson counting process with intensity $\lambda \geq 0$, and $\{Z_i\}_{i=1}^{\infty}$ is a collection of i.i.d. random variables with common probability law $m(dz)$. We assume that W , N , and $\{Z_i\}$ are pairwise independent. We express the dynamics of X in terms of the compensated compound Poisson process \tilde{J}_t ,

$$\tilde{J}_t := J_t - \lambda \mathbb{E}_{\mathbb{Q}}[Z]t,$$

giving us,

$$dX_t = \xi \left(\eta + \frac{\lambda}{\xi} \mathbb{E}_{\mathbb{Q}}[Z] - X_{t-} \right) dt + \sqrt{\gamma + \delta X_{t-}} dW_t + d\tilde{J}_t, \quad X_0 = x_0. \quad (2)$$

We thus define \mathbb{G} by $\mathcal{G}_t := \sigma(W_s, J_s : 0 \leq s \leq t)$. We assume that X is a càdlàg, \mathbb{G} -adapted process (hence, X is \mathbb{G} -progressively measurable (Karatzas and Shreve 1988, Proposition 1.13)).

A standard result for affine jump-diffusion process is the exponential-affine representation

$$\mathbb{E}_{\mathbb{Q}} \left[\exp \left\{ - \int_t^T (\mu_0 + \mu_1 X_s) ds \right\} e^{u X_T} \middle| \mathcal{G}_t \right] = \exp \{ A(t, T) + B(t, T) X_t \}, \quad 0 \leq t < T, \quad (3)$$

where $\mu_0, \mu_1 \in \mathbb{R}$, $u \in \mathbb{C}$, and A and B are solutions to a system of complex-valued ordinary differential equations, which we state below. See Duffie et al. (2000) for a proof and a more general discussion.

Theorem 2.1. *Suppose A and B are unique solutions of the complex-valued ODEs*

$$\frac{dB(t, T)}{dt} = \mu_1 + \xi B(t, T) - \frac{1}{2} \sigma^2 \delta B^2(t, T) \quad (4)$$

$$\frac{dA(t, T)}{dt} = \mu_0 - \xi \eta B(t, T) - \frac{1}{2} \sigma^2 \gamma B^2(t, T) - \lambda (\mathbb{E}_{\mathbb{Q}}^Z [\exp \{ B(t, T) Z \}] - 1), \quad (5)$$

with terminal conditions $B(T, T) = u \in \mathbb{C}$ and $A(T, T) = 0$ (the superscript Z in $\mathbb{E}_{\mathbb{Q}}^Z[\cdot]$ indicates taking the mean with respect to Z). Suppose further that the following integrability conditions hold:

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}} \left[\int_0^T \left| B(t, T) \sigma \sqrt{\gamma + \delta X_t} \varphi_t \right|^2 dt \right] &< \infty \\ \mathbb{E}_{\mathbb{Q}} \left[\int_0^T \left| \lambda (\mathbb{E}_{\mathbb{Q}}^Z [\exp \{ B(t, T) Z \}] - 1) \varphi_t \right|^2 dt \right] &< \infty, \end{aligned}$$

where $\varphi = \{\varphi_t\}$ is the process defined by

$$\varphi_t := \exp \left\{ - \int_0^t (\mu_0 + \mu_1 X_s) ds \right\} e^{A(t, T) + B(t, T) X_t}, \quad t \in [0, T].$$

Then the representation (3) holds.

We now make additional assumptions on the probability distribution of τ . We assume there exists a nonnegative \mathbb{G} -progressive process $\mu := \{\mu_t\}_{t \geq 0}$ such that

$$\mathbb{Q}[\tau > t | \mathcal{G}_t] = \exp \left\{ - \int_0^t \mu_s ds \right\}. \quad (6)$$

That is, the time of death τ is said to admit an \mathbb{F} -intensity μ . Given (6), the risk-neutral survival probability $S(t, T)$ of an individual, conditional on being alive at time t , surviving until

time $T > t$ is given by

$$S(t, T) := \mathbf{1}_{\{\tau > t\}} \mathbb{Q}[\tau > T | \mathcal{F}_t] = \mathbf{1}_{\{\tau > t\}} \mathbb{E}_{\mathbb{Q}} \left[\exp \left\{ - \int_t^T \mu_s ds \right\} \middle| \mathcal{G}_t \right] \quad (7)$$

(see for example [Filipović 2009](#), Lemma 12.3). In (7), we note that $\mathbf{1}_{\{\tau > t\}} = 1$ since we assume that the individual is alive at time t , so $S(t, T)$ is simply equal to the conditional expectation. We also assume that τ satisfies $\mathbb{Q}[\tau > t | \mathcal{G}_{\infty}] = \mathbb{Q}[\tau > t | \mathcal{G}_t]$, for all $t \geq 0$. Thus, we say that the stopping time τ is doubly stochastic.

Suppose the intensity process μ is an affine function of X_t , i.e.

$$\mu_t = \mu_0 + \mu_1 X_{t-}, \quad t \in [0, T]$$

with $\mu_0, \mu_1 \geq 0$. This specification is permitted since X is \mathbb{G} -progressively measurable and is strictly positive with positive probability, provided $x_0 > 0$. Thus, (3) is precisely the risk-neutral survival probability $S(t, T)$ defined in (7); that is,

$$S(t, T) = e^{A(t, T) + B(t, T)X_t}, \quad 0 \leq t \leq T, \quad (8)$$

where A and B are solutions to (4) and (5). Consequently, the average force of mortality $\bar{\mu}(t, T)$, defined as

$$\bar{\mu}(t, T) := -\frac{\log S(t, T)}{T - t},$$

is an affine function of X_t ,

$$\bar{\mu}(t, T) = -\frac{A(t, T)}{(T - t)} - \frac{B(t, T)}{(T - t)} X_t, \quad 0 \leq t \leq T. \quad (9)$$

Hereafter, we shall refer to A and B as *factor loadings* and X as the *latent factor*.

2.2 Specific Models

We consider the case where the mortality intensity is X , i.e. $\mu = X$ or where $\mu_0 = 0$ and $\mu_1 = 1$. We concentrate on two models nested within (1):

1. **Blackburn-Sherris model with jumps (BSj):** This model is given by

$$dX_t = -\xi X_{t-} dt + \sigma dW_t + dJ_t. \quad (10)$$

This model is an Ornstein-Uhlenbeck process with jumps. A multi-factor version of this model with no jumps was used by [Blackburn and Sherris \(2013\)](#) and [Jevtić et al. \(2013\)](#).

2. **Cox-Ingersoll-Ross model with jumps (CIRj):** This model is given by

$$dX_t = \xi(\eta - X_{t-}) dt + \sigma \sqrt{X_{t-}} dW_t + dJ_t. \quad (11)$$

Given evidence that mortality rates are mean-reverting over time for specific ages (see for example [Njenga and Sherris 2011](#)), we may assume that $\xi > 0$ for both the BSj and CIRj models. However, since the BSj model specified in (10) has a long-run mean of zero, imposing the condition $\xi > 0$ may lead to a negative mortality intensity, which violates the assumption that $\mu = X$ must be a nonnegative process. Thus, in our parameter estimation, we allow ξ to be any nonzero real number for the BSj model and rely on the estimation methodology to ascertain any mean-reverting behavior in the mortality rates under this model.⁸ For the CIRj model, we impose the restrictions $\xi > 0$ and $\eta > 0$.

Under the BSj and CIRj model, the factor loadings A and B of the risk-neutral survival probability (8) are available in closed form. The expressions for the factor loadings are presented in the following proposition.

Proposition 2.2. *Under the BSj and CIRj models, the risk-neutral survival probability $S(t, T; a)$ is given by (8), where A and B are:*

1. **BSj:**

$$B(t, T) = -\frac{1 - e^{-\xi(T-t)}}{\xi} \quad (12)$$

$$A(t, T) = -\lambda(T-t) + \frac{1}{2} \frac{\sigma^2}{\xi^3} \left[\xi(T-t) - 2(1 - e^{-\xi(T-t)}) + \frac{1}{2}(1 - e^{-2\xi(T-t)}) \right] + \lambda \int_t^T M_Z(B(s, T)) ds \quad (13)$$

2. **CIRj:**

$$B(t, T) = -\frac{2(e^{\vartheta(T-t)} - 1)}{(\xi + \vartheta)(e^{\vartheta(T-t)} - 1) + 2\vartheta} \quad (14)$$

$$A(t, T) = -\lambda(T-t) + \frac{2\xi\eta}{\sigma^2} \ln \left(\frac{2\vartheta e^{\frac{1}{2}(\xi+\vartheta)(T-t)}}{(\xi + \vartheta)(e^{\vartheta(T-t)} - 1) + 2\vartheta} \right) + \lambda \int_t^T M_Z(B(s, T)) ds, \quad (15)$$

where $\vartheta := \sqrt{\xi^2 + 2\sigma^2}$.

In both cases, $M_Z(\cdot)$ is the moment generating function (mgf) of Z under \mathbb{Q} .

Proof. The derivation $B(t, T)$ for both models is similar to that in the no-jump case, since the jump intensity is constant (see for example [Blackburn and Sherris 2013](#); [Huang et al. 2022](#)). Given the explicit form of $B(t, T)$, the integration of (5) also proceeds as in the no-jump case. In our case, we leave the integral involving the mgf of the jump size distribution as is, although under alternative model parameterizations, this integral can be computed explicitly (for instance, see [Bravo and Nunes 2021](#)). ■

⁸In their single-cohort analysis, [Bravo and Nunes \(2021\)](#) show that non-mean-reverting processes fit observed survival curves better than mean-reverting processes. We emphasize, however, that they look at the dynamics of mortality rates across ages for a fixed cohort, and indeed they find that mortality rates are exponentially increasing at older ages.

Models (10) and (11) for the instantaneous mortality intensity imply that we consider jumps in mortality rates to be persistent mortality shocks, in the sense of [Chen and Cox \(2009\)](#). That is, a mortality shock experienced by one cohort will affect subsequent cohort mortality rates.⁹ The assumption of persistence in mortality shocks makes more sense in our age-cohort context since historical and natural events (e.g. wars, pandemics, climate changes and catastrophes) and medical advancements often have persistent effects on mortality rates.

2.3 Jump Size Distribution

Following [Bravo and Nunes \(2021\)](#), we assume that the jump size random variable Z has an asymmetric double exponential (ADE) distribution. This jump size distribution, used by [Kou \(2002\)](#) and [Ramezani and Zeng \(1998\)](#) as an extension of the Gaussian jump size assumption of [Merton \(1976\)](#) for asset returns, allows for both positive and negative jumps. This is important since we are also interested in modelling sudden upward or downward mortality shocks.

Under the ADE distribution, the jump size random variable Z has the probability density function

$$f_Z(z) = \rho\phi_1 e^{-\phi_1 z} \mathbf{1}_{\{z \geq 0\}} + (1 - \rho)\phi_2 e^{\phi_2 z} \mathbf{1}_{\{z < 0\}}, \quad z \in \mathbb{R},$$

with parameters $\rho \geq 0$, and $\phi_1, \phi_2 > 0$. As such, we write $Z \sim \text{ADE}(\rho, \phi_1, \phi_2)$. The parameter ρ represents the probability of a positive jump, with mean $\frac{1}{\phi_1} > 0$; similarly, $1 - \rho$ represents the probability of a negative jump, with mean $\frac{1}{\phi_2} > 0$. The corresponding moment generating function is

$$M_Z(u) = \frac{\rho\phi_1}{\phi_1 - u} + \frac{(1 - \rho)\phi_2}{\phi_2 + u}, \quad -\phi_2 < u < \phi_1. \quad (16)$$

The change-of-measure result of the jump component of the model is discussed in terms of the Poisson random measure $N(dt, dz)$ associated to the counting process N_t and the jump size distribution $m_{\mathbb{Q}}(dz) := f_Z(z) dz$. We establish the required notation here. For any $t \geq 0$ and $A \in \mathcal{B}(\mathbb{R})$ (the Borel σ -algebra on \mathbb{R}), the random measure $N((0, t], A) := N_t(A)$ counts the number of jumps with size in A occurring on the interval $(0, t]$. For $A = \mathbb{R}$ we simply write $N_t := N_t(\mathbb{R})$. Thus, the compound Poisson process may be written as

$$J_t = \sum_{i=1}^{N_t} Z_i = \int_0^t \int_{\mathbb{R}} z N(dt, dz)$$

(see for example [Runggaldier 2003](#), equation 6). The corresponding compensated random measure $\tilde{N}(dt, dz)$ is given by $\tilde{N}(dt, dz) := N(dt, dz) - \lambda m_{\mathbb{Q}}(dz) dt$, which is related to the compensated compound Poisson process \tilde{J} via $\tilde{J}_t = \int_0^t \int_{\mathbb{R}} z \tilde{N}(dt, dz)$.

⁹In contrast, non-persistent mortality shocks can be modelled by subtracting the jumps experienced at time $t - 1$ from the latent state value at time t ; see [Chen and Cox \(2009\)](#) for a more detailed discussion in the context of the Lee-Carter model with jumps.

2.4 Historical Dynamics of the Mortality Intensity Process

To forecast future survival curves based on historical data, we need to specify the evolution of the latent process X under the historical probability measure \mathbb{P} also defined on the measurable space (Ω, \mathcal{F}) . The combined financial and insurance markets are generally considered to be incomplete, hence the risk-neutral measure \mathbb{Q} is not unique. Thus, we are free to specify \mathbb{Q} , specifically its density process relative to \mathbb{P} , such that X has affine dynamics in both \mathbb{P} and \mathbb{Q} .

We start with the \mathbb{P} dynamics of X , which we also assume to be an affine jump-diffusion process

$$dX_t = \xi^P(\eta^P - X_{t-}) dt + \sigma\sqrt{\gamma + \delta X_{t-}} dW_t^P + dJ_t, \quad (17)$$

where $\xi^P, \eta^P \in \mathbb{R}$ and $\sigma, \gamma, \delta \geq 0$ are constants, W^P is a standard Brownian motion under \mathbb{P} , and $J_t := \sum_{i=1}^{N_t} Z_i$ is a compound Poisson process whose corresponding counting process N_t has intensity λ^P and jumps are i.i.d. with common probability density function f_Z^P . The corresponding \mathbb{P} -probability law of the jumps is denoted by $m_{\mathbb{P}}(dz)$. We denote by $\tilde{N}^P := \{\tilde{N}_t^P\}$ the compensated Poisson counting process $N_t^P := N_t - \lambda^P t$ and by $\tilde{J}^P := \{\tilde{J}_t^P\}$ the compensated compound Poisson process $J_t^P := J_t - \lambda \mathbb{E}_{\mathbb{P}}[Z]t$. The \mathbb{P} -compensated random counting measure associated to J_t is denoted by $\tilde{N}^P(dt, dz)$. As before, W , N , and $\{Z_i\}$ are pairwise independent. In particular, the BSj and CIRj models under the \mathbb{P} measure are given by

1. **BSj:**

$$dX_t = -\xi^P X_{t-} dt + \sigma dW_t^P + dJ_t \quad (18)$$

2. **CIRj:**

$$dX_t = \xi^P(\eta^P - X_{t-}) dt + \sigma\sqrt{X_{t-}} dW_t^P + dJ_t. \quad (19)$$

The transition from \mathbb{P} to \mathbb{Q} is facilitated by the following Girsanov Theorem for Itô-Lévy processes. The statement of the theorem below is a one-dimensional version of the statement in [Øksendal and Sulem \(2019, Theorem 1.33\)](#) (see also [Runggaldier \(2003, Theorem 2.5\)](#)). The change-of-measure result for the jump part is expressed in terms of the compensated random counting measure.

Theorem 2.3. *Let $\psi^D : \Omega \times [0, T] \rightarrow \mathbb{R}$ and $\psi^J : \Omega \times [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ be \mathbb{F} -predictable processes with $\psi_t^J(z) \leq 1$ \mathbb{P} -almost surely such that the process $L := \{L_t\}_{t \in [0, T]}$ defined by*

$$L_t := \exp \left\{ - \int_0^t \psi_s^D dW_s^P - \frac{1}{2} \int_0^t (\psi_s^D)^2 ds + \int_0^t \int_{\mathbb{R}} \ln(1 - \psi_s^J(z)) \tilde{N}^P(dt, dz) + \int_0^T \int_{\mathbb{R}} \left[\ln(1 - \psi_s^J(z)) + \psi_s^J(z) \right] \lambda^{\mathbb{P}} m_{\mathbb{P}}(dz) ds \right\}$$

is well-defined, is strictly positive for $0 \leq t \leq T$, and satisfies $\mathbb{E}_{\mathbb{P}}[L_T] = 1$.

Define the probability measure \mathbb{Q} equivalent to \mathbb{P} on (Ω, \mathcal{F}_T) by

$$d\mathbb{Q}(\omega)|_{\mathcal{F}_T} = L_T d\mathbb{P}(\omega)|_{\mathcal{F}_T}.$$

Furthermore, we define the processes W and the random measure $\tilde{N}(dt, dz)$ by

$$\begin{aligned} dW_t &= \psi_t^D dt + dW_t^P \\ \tilde{N}(dt, dz) &= \psi_t^J(z) \lambda^P m_{\mathbb{P}}(dz) dt + \tilde{N}^P(dt, dz). \end{aligned}$$

Then W is a standard Brownian motion under \mathbb{Q} and $\tilde{N}(dt, dz)$ is the compensated compound Poisson counting process under \mathbb{Q} corresponding to J .

To ensure that the process X , satisfying (17), is affine under both \mathbb{P} and \mathbb{Q} , we set

$$\psi_t^D := \zeta_1 \sqrt{\gamma + \delta X_{t-}} + \zeta_2 \bar{D}(X_{t-}) X_{t-}, \quad (20)$$

where $\zeta_1, \zeta_2 \in \mathbb{R}$ and

$$\bar{D}(x) = \begin{cases} (\gamma + \delta x)^{-1/2} & \text{if } \inf_{x \geq 0} \{\gamma + \delta x\} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

This is the *essentially affine* specification of Duffee (2002) for the market price of risk process.¹⁰ This specification yields the following \mathbb{Q} -dynamics of the analyzed models:

1. **BSj:** Since $\gamma = 1$ and $\delta = 0$, we have $\inf_{x \geq 0} \{\gamma + \delta x\} > 0$, so $D(X_{t-}) = (\gamma + \delta X_{t-})^{-1/2}$. This implies that

$$\psi_t^D = \zeta_1 + \zeta_2 X_{t-}.$$

As we are free to choose ζ_1 and ζ_2 , we set $\zeta_1 = 0$ (this effectively means that the long-run mean under \mathbb{Q} is also zero), and so the dynamics of X under \mathbb{Q} are

$$\begin{aligned} dX_t &= -\xi^P X_{t-} dt + \sigma(dW_t - \zeta_2 X_{t-} dt) + dJ_t \\ &= -(\xi^P + \sigma \zeta_2) X_{t-} dt + \sigma dW_t + dJ_t. \end{aligned}$$

That is, we can relate the risk-neutral and real-world mean-reversion parameters ξ and ξ^P via

$$\xi = \xi^P + \sigma \zeta, \quad (21)$$

where, for convenience, we set $\zeta_2 = \zeta$.

2. **CIRj:** Since $\gamma = 0$ and $\delta = 1$, $\gamma + \delta x$ is not bounded away from zero, hence $D(X_{t-}) = 0$. This results to

$$\psi_t^D = \zeta \sqrt{X_{t-}},$$

where for convenience we set $\zeta_1 = \zeta$. We can thus write (17) as,

$$dX_t = \xi^P (\eta^P - X_{t-}) dt + \sigma \sqrt{X_{t-}} (dW_t - \zeta \sqrt{X_t} dt) + dJ_t$$

¹⁰See also Cheredito et al. (2007).

$$= (\xi^P + \sigma\zeta) \left(\frac{\xi^P \eta^P}{\xi^P + \sigma\zeta} - X_{t-} \right) dt + \sigma \sqrt{X_{t-}} dW_t + dJ_t.$$

That is, the risk-neutral and real-world parameters are related via the equations

$$\xi = \xi^P + \sigma\zeta, \quad \eta = \frac{\xi^P \eta^P}{\xi^P + \sigma\zeta}. \quad (22)$$

Although the process $\psi_t^J(z)$ does not appear explicitly in the risk-neutral dynamics of the latent factor, it affects the intensity of the counting process N_t and the jump size distribution under \mathbb{Q} . We parameterize ψ^J as

$$\psi_t^J(z) = \psi^J(z) := 1 - \exp\{\psi_0 + \psi_1 z\}, \quad z \in \mathbb{R} \quad (23)$$

where $\psi_0, \psi_1 \in \mathbb{R}$. With this specification, we can write

$$\tilde{N}(dt, dz) = N(dt, dz) - \lambda m_{\mathbb{Q}}(dz) dt,$$

where

$$\lambda := e^{\psi_0} \mathbb{E}_{\mathbb{P}}[e^{\psi_1 Z}] \lambda^P, \quad m_{\mathbb{Q}}(dz) := \frac{e^{\psi_1 z}}{\mathbb{E}_{\mathbb{P}}[e^{\psi_1 Z}]} m_{\mathbb{P}}(dz). \quad (24)$$

From equation (24) we deduce that, under \mathbb{Q} , the intensity of the counting process N_t is λ and the law of jump sizes is $m_{\mathbb{Q}}(dz)$.

The transformation of the \mathbb{P} -probability law of the jumps in (24) is an Esscher-type transform of the jump size distribution with parameter ψ_1 (Gerber and Shiu 1994). Unfortunately, the Esscher transform fails to preserve the asymmetric double exponential jump size distribution when changing from \mathbb{P} to \mathbb{Q} , unlike the case of distributions belonging to the exponential family of distributions (e.g. Gaussian, exponential). However, for the purposes of calculating the factor loading $A(t, T)$, it is sufficient to obtain an expression for the moment generating function $M_Z(z)$ of Z under \mathbb{Q} in terms of the real-world and change of measure parameters. Specifically, we assume that under \mathbb{P} , $Z \sim \text{ADE}(\rho^P, \phi_1^P, \phi_2^P)$. From equation (24), it follows that $M_Z(z)$ is given by

$$M_Z(c) = \frac{M_Z^{\mathbb{P}}(c + \psi_1)}{M_Z^{\mathbb{P}}(\psi_1)}, \quad \text{where} \quad M_Z^{\mathbb{P}}(u) = \frac{\rho^P \phi_1^P}{\phi_1^P - u} + \frac{(1 - \rho^P) \phi_2^P}{\phi_2^P + u}. \quad (25)$$

We note that $M_Z(c)$ is defined whenever $-\phi_2^P < c < \phi_1^P$ and $-\phi_2^P < c + \psi_1 < \phi_1^P$.

3 Data

The models introduced in Section 2 are used for the analysis of the age-cohort mortality rates, as sourced from the Human Mortality Database for the cohorts of USA males born between 1883 to 1915, aged 50 to 99. We chose to analyze USA mortality data to represent the mortality experience of a large developed country. In addition, the cohorts born between 1883 and 1915 will have experienced key historical events, namely the World Wars and the influenza pandemic,

which had a significant effect on mortality rates. More precisely, we can obtain the number of central exposure-at-risk $E_{x,t}^c$ and the number of deaths $D_{x,t}$ for each age x and cohort t . We then estimate the central rate of mortality by

$$m_{x,t} = \frac{D_{x,t}}{E_{x,t}^c},$$

which are equal to the force of mortality assuming that the force of mortality is constant between integer ages.

In this analysis, we set the base age to 50. The probability that an individual, aged 50 in calendar year y , will survive until age $50 + T - y$ (i.e. until calendar year T) can be calculated as

$$S(y, T) = \prod_{k=1}^{T-y} e^{-m_{50+k-1, y-50}}.$$

Letting $t := y - 50$ denote this individual's birth year (cohort), the corresponding average force of mortality is given by

$$\bar{\mu}_{t,k} = \frac{1}{k} \sum_{j=1}^k m_{50+j-1, t}, \quad (26)$$

for $k = 1, 2, \dots, 50$, indicating our interest in survival probabilities up to age 99 for individuals born in cohort t currently aged 50. We note that $\bar{\mu}_{t,k}$ in (26) is $\bar{\mu}(t, t+k)$ in (9).

4 Parameter Estimation

In what follows we describe the proposed parameter estimation methodology. We first formulate the discrete-time state-space models corresponding to the affine jump diffusion mortality models in Section 2. For completeness, we provide a summary of the particle filtering algorithm for our state-space model and discuss the implementation of the particle filter that is specific to our state-space models in Algorithm 1. Then we discuss the particle MCMC algorithm to estimate the parameter posterior distribution of the state-space models. Lastly, we provide the specifics of our the implementation of the methodology.

4.1 State-Space Representation

We characterize the parameter estimation problem of the affine jump diffusion models presented in this work in terms of a state-space representation, similar to Blackburn and Sherris (2013), Huang et al. (2022), Ungolo et al. (2021), and Xu et al. (2020a). This approach differs from the earlier works of Luciano and Vigna (2008), Jevtić et al. (2013), Bravo and Nunes (2021), and Bravo (2021), who use cohort-specific parameters estimated by minimizing the square difference between observed and theoretical survival probabilities.

Therefore, the approach of this work yields consistent survival curves, which depend on the same parameters across all cohorts under analysis. In this way, it is possible to estimate the

dynamics of the latent variable X_t which is of great importance when projecting survival curves for future cohorts.

Despite the continuous-time nature of the modelling framework, we need to work with a one-year discretized version of the state-space modelling approach given the nature of available data.

To this end, note that the evolution of X_t under the CIRj model (19) over the time interval $[t - \Delta t, t]$ (with $0 \leq \Delta t < t$) can be approximated as

$$X_t = X_{t-\Delta t} + \xi^P(\eta^P - X_{t-\Delta t})\Delta t + \sigma\sqrt{X_{t-\Delta t}}\Delta W_t^P + \sum_{i=N_{t-\Delta t}+1}^{N_t} Z_i,$$

where an Euler-Maruyama scheme is employed, along the lines of [Golightly \(2009\)](#) and [Johannes et al. \(2009\)](#). Here, $\Delta W_t^P := W_t^P - W_{t-\Delta t}^P$ is the increment of W^P which has a normal distribution $N(0, \Delta t)$. For simplicity, we assume that there is at most one jump occurring on $[t - \Delta t, t]$. In this context the jump is interpreted as a positive or negative mortality shock experienced by the cohort born in year t . Thus, $\sum_{i=N_{t-\Delta t}+1}^{N_t} Z_i$ can be written as $Z_t\Delta N_t$, where $Z_t \sim \text{ADE}(\rho^P, \phi_1^P, \phi_2^P)$ and $\Delta N_t := N_t - N_{t-\Delta t}$ is equal to 1 if a jump occurs during the period $[t - \Delta t, t]$ and 0 otherwise. This is consistent with the assumptions of the Poisson model for jump-diffusion processes (see e.g. [Merton 1976](#); [Runggaldier 2003](#)), in which it is assumed that $\mathbb{P}[\Delta N_t = 1] = \lambda^P \Delta t + O(\Delta t)$, where $O(\cdot)$ is the big-O notation. Hence, when estimating the model, we approximate the distribution of ΔN_t by means of a Bernoulli($\lambda^P \Delta t$) distribution (see also [Golightly \(2009\)](#) and [Johannes et al. \(2009\)](#)). Finally, setting $\Delta t = 1$ corresponding to the annual frequency of mortality data, we can simplify the notation and obtain

$$X_t = X_{t-1} + \xi^P(\eta^P - X_{t-1}) + \sigma\sqrt{X_{t-1}}\varpi_t + Z_t\Delta N_t, \quad X_0 = x_0, \quad (27)$$

where $\varpi_t \sim N(0, 1)$, $Z_t \sim \text{ADE}(\rho^P, \phi_1^P, \phi_2^P)$, and $\Delta N_t \sim \text{Bernoulli}(\lambda^P)$.

Similarly, for the BSj model, we obtain the following state-transition equation:

$$X_t = X_{t-1} - \xi^P X_{t-1} + \sigma\varpi_t + Z_t\Delta N_t, \quad X_0 = x_0, \quad (28)$$

where $\varpi_t \sim N(0, 1)$, $Z_t \sim \text{ADE}(\rho^P, \phi_1^P, \phi_2^P)$, and $\Delta N_t \sim \text{Bernoulli}(\lambda^P)$.

The state-space characterization of the parameter estimation problem is completed by the measurement equation, where the average force of mortality is an affine function of the latent state variable X_t :

$$\bar{\mu}_t = \bar{A} + \bar{B}X_t + \epsilon_t, \quad \epsilon_t \sim N(0, H), \quad (29)$$

where $\bar{\mu}_t$, \bar{A} , and \bar{B} are 50×1 vectors with entries

$$\bar{\mu}_t = [\bar{\mu}_{t,k}]_{k=1}^{50}, \quad \bar{A} = \left[-\frac{\hat{A}(k)}{k} \right]_{k=1}^{50}, \quad \bar{B} = \left[-\frac{\hat{B}(k)}{k} \right]_{k=1}^{50},$$

with $\bar{\mu}_{t,k}$ given by equation (26), $\hat{A}(k) = A(0, k)$, and $\hat{B}(k) = B(0, k)$. From Proposition 2.2, we note that $A(t, t+k)$ and $B(t, t+k)$ only depend on $(t+k) - t = k$ and not on t , so for convenience we set $t = 0$ when computing the factor loading vectors \bar{A} and \bar{B} . The measurement error covariance matrix H is assumed to be a diagonal matrix, reflecting the assumption that measurement noise is independent between ages. Furthermore, to account for the increasing variation in mortality rates at older ages, we assume that the measurement error variances $H_{k,k}$ have the parametric form

$$H_{k,k} = e^{r_c} + \frac{e^{r_1}}{k} \sum_{i=1}^k e^{e^{r_2} i}, \quad k = 1, \dots, 50, \quad (30)$$

where $r_1, r_2, r_c \in \mathbb{R}$ are additional parameters to be estimated from the data¹¹.

Let Θ denote the set of parameters which characterize the affine models of this work:

1. **BSj:** $\Theta = (x_0, \xi^P, \sigma, \lambda^P, \rho^P, \phi_1^P, \phi_2^P, \zeta, \psi_0, \psi_1, r_1, r_2, r_c)$
2. **CIRj:** $\Theta = (x_0, \xi^P, \eta^P, \sigma, \lambda^P, \rho^P, \phi_1^P, \phi_2^P, \zeta, \psi_0, \psi_1, r_1, r_2, r_c)$.

Key to the estimation of Θ , is the likelihood function of $\bar{\mu}_{1:T}$, denoted by $p(\bar{\mu}_{1:T}|\Theta)$, which can be conveniently factorized as:

$$p(\bar{\mu}_{1:T}|\Theta) = p(\bar{\mu}_1|\Theta) \prod_{t=2}^T p(\bar{\mu}_t|\bar{\mu}_{1:t-1}, \Theta). \quad (31)$$

Let $L_t = (X_t, Z_t, \Delta N_t)$ denote the vector of latent variables.¹² Furthermore, we denote by $g_\theta(\bar{\mu}_t|L_t) = g(\bar{\mu}_t|L_t, \Theta = \theta)$ the conditional probability density function of $\bar{\mu}_t$ given L_t and $\Theta = \theta$ and by $f_\theta(L_t|L_{t-1}) = f(L_t|L_{t-1}, \Theta = \theta)$ the conditional probability density function of L_t given L_{t-1} and Θ . The measurement model and the state transition model are characterized by the following two densities $g_\theta(\bar{\mu}_t|L_t)$ and $f_\theta(L_t|L_{t-1})$, respectively. Using the Bayesian filtering recursions (see e.g. Särkkä 2013, Theorem 12.1) we can write $p_\theta(\bar{\mu}_t|\bar{\mu}_{1:t-1})$ as

$$\begin{aligned} p_\theta(\bar{\mu}_t|\bar{\mu}_{1:t-1}) &= \int g_\theta(\bar{\mu}_t|L_t) p_\theta(L_t|\bar{\mu}_{1:t-1}) dL_t \\ p_\theta(L_t|\bar{\mu}_{1:t-1}) &= \int f_\theta(L_t|L_{t-1}) p_\theta(L_{t-1}|\bar{\mu}_{1:t-1}) dL_{t-1} \\ p_\theta(L_t|\bar{\mu}_{1:t}) &= \frac{g_\theta(\bar{\mu}_t|L_t) p_\theta(L_t|\bar{\mu}_{1:t-1})}{p_\theta(\bar{\mu}_t|\bar{\mu}_{1:t-1})}, \end{aligned} \quad (32)$$

for $t = 2, 3, \dots, T$, where

$$g_\theta(\bar{\mu}_t|L_t) = \mathbf{N}_\theta(\bar{\mu}_t; \bar{A} + \bar{B}X_t, H) \quad (33)$$

¹¹Huang et al. (2022) and Xu et al. (2020a) refer to this as the ‘‘Poisson’’ variation in historical mortality rates reflecting the size of the population at each age.

¹²We treat the jump size and indicator random variables Z_t and ΔN_t that appear in the dynamics of X_t as additional latent variables of interest. As a result, conditional on Z_t and ΔN_t , X_t has a normal distribution driven by the white noise error term ϖ_t .

$$f_\theta(L_t|L_{t-1}) = p_\theta(X_t|X_{t-1}, Z_t, \Delta N_t)p_\theta(Z_t|\Delta N_t)p_\theta(\Delta N_t), \quad (34)$$

with

$$p_\theta(X_t|X_{t-1}, Z_t, \Delta N_t) = \begin{cases} \mathbf{N}_\theta(X_t; X_{t-1} - \xi^P X_{t-1} + Z_t \Delta N_t, \sigma^2) & \text{for BSj} \\ \mathbf{N}_\theta(X_t; X_{t-1} + \xi^P (\eta^P - X_{t-1}) + Z_t \Delta N_t, \sigma^2 X_{t-1}) & \text{for CIRj} \end{cases}$$

and $p_\theta(Z_t|\Delta N_t)p_\theta(\Delta N_t)$ given by

$$p_\theta(z|\Delta n)p_\theta(\Delta n) = (\lambda^P)^{\Delta n} \left[\rho^P \phi_1^P e^{-\phi_1^P z} \mathbf{1}_{\{z \geq 0\}} + (1 - \rho^P) \phi_2^P e^{\phi_2^P z} \mathbf{1}_{\{z < 0\}} \right] + (1 - \lambda^P)^{1 - \Delta n},$$

for $z \in \mathbb{R}$ and $\Delta n = 0, 1$.

The presence of non Gaussian latent components following the inclusion of jumps, yields a non-Gaussian likelihood function. This means that there is not a closed form solution for the integrals in equation (32). For this reason, in this work we propose the use of particle filtering for approximating such integrals and obtain a likelihood function of the data, for a given value of the parameter Θ .

4.2 Particle Filtering Algorithm

Introduced by [Gordon et al. \(1993\)](#), the particle filter is a sequential Monte Carlo method to estimate the distribution of the latent state variables by using a set of random samples, or *particles*. Due to its flexibility, the particle filter has been applied to perform inferences on stochastic volatility and jump-diffusion models for stock prices (see for example [Johannes et al. 2009](#); [Golightly 2009](#)). For a more general discussion, see for example [Andrieu et al. \(2010\)](#); [Doucet and Johansen \(2011\)](#); [Särkkä \(2013\)](#).

The particle filter approximates (sequentially for $t = 1, \dots, T$) the filtering distribution $p_\theta(L_t|\bar{\boldsymbol{\mu}}_{1:t})$ (see equation (32)) using a system $\{w_t^{(i)}, L_t^{(i)}\}_{i=1}^{N_p}$ of N_p particles $L_t^{(i)}$ and associated weights $w_t^{(i)}$.

The particles are sampled from an *importance density* $q_\theta(L_{1:t}|\bar{\boldsymbol{\mu}}_{1:t})$ instead of the full posterior density $p_\theta(L_{1:t}|\bar{\boldsymbol{\mu}}_{1:t})$ as the latter is unknown. The importance density is chosen such that $p_\theta(L_{1:t}|\bar{\boldsymbol{\mu}}_{1:t}) > 0$ implies $q_\theta(L_{1:t}|\bar{\boldsymbol{\mu}}_{1:t}) > 0$. Owing to the sequential nature of the algorithm, the importance density satisfies the following recursion

$$q_\theta(L_{1:t}|\bar{\boldsymbol{\mu}}_{1:t}) = q_\theta(L_t|L_{1:t-1}, \bar{\boldsymbol{\mu}}_{1:t})q_\theta(L_{1:t-1}|\bar{\boldsymbol{\mu}}_{1:t-1}), \quad (35)$$

for some function $q_\theta(L_t|L_{1:t-1}, \bar{\boldsymbol{\mu}}_{1:t})$.

In our implementation, we choose the so called *bootstrap filter* ([Gordon et al. 1993](#)), which implies $q_\theta(L_t|L_{1:t-1}, \bar{\boldsymbol{\mu}}_{1:t}) = f_\theta(L_t|L_{t-1})$. This choice is motivated by the ease with which we can sample from $f_\theta(L_t|L_{t-1})$ via the state transition equation (28) or (27). An alternative option is the use of the auxiliary particle filter (APF) ([Pitt and Shephard 1999](#); [Johansen and Doucet 2008](#)) which incorporates future data $\bar{\boldsymbol{\mu}}_{t+1}$ in the importance distribution. However, the APF

requires the evaluation or approximation of

$$p_\theta(\bar{\boldsymbol{\mu}}_{t+1}|L_t) = \int g_\theta(\bar{\boldsymbol{\mu}}_{t+1}|L_{t+1})f_\theta(L_{t+1}|L_t) dL_{t+1},$$

which is not tractable in our setting since L_t is multi-dimensional.

Hence at time t , we have the particles $\{L_{1:t-1}^{(i)}\}_{i=1}^{N_p}$ with corresponding weights $\{w_{t-1}^{(i)}\}_{i=1}^{N_p}$. Before we generate new particles at time t , a further *resampling step* is performed to avoid generating too many particles with zero, or very close to zero weight, the so called *degeneracy* problem. Following [Andrieu et al. \(2010\)](#) and [Dahlin and Schön \(2019\)](#) we perform the multinomial resampling step of the particles at time $t - 1$. To keep track of the resampling history, we let $a_t^{(i)}$ denote the parent index of particle i at time step t .

Let $v_t^{(i)}$ and $w_t^{(i)}$ denote the unnormalized and the normalized weights respectively, calculated as

$$v_t^{(i)} = \frac{g_\theta(\bar{\boldsymbol{\mu}}_t|L_t^{(i)})f_\theta(L_t^{(i)}|L_{t-1}^{a_t^{(i)}})}{q_\theta(L_t^{(i)}|L_{t-1}^{a_t^{(i)}}, \bar{\boldsymbol{\mu}}_{1:t})}, \quad w_t^{(i)} = \frac{v_t^{(i)}}{\sum_{j=1}^{N_p} v_t^{(j)}}. \quad (36)$$

An estimate of the factors in the likelihood function of equation (31) is given by

$$\hat{p}_\theta(\bar{\boldsymbol{\mu}}_t|\bar{\boldsymbol{\mu}}_{1:t-1}) \approx \frac{1}{N_p} \sum_{i=1}^{N_p} v_t^{(i)}, \quad (37)$$

resulting to the following estimate for the marginal likelihood,

$$\hat{p}_\theta(\bar{\boldsymbol{\mu}}_{1:T}) = \hat{p}_\theta(\bar{\boldsymbol{\mu}}_1) \prod_{t=2}^T \left(\frac{1}{N_p} \sum_{i=1}^{N_p} v_t^{(i)} \right).$$

The use of particle filtering for estimating the likelihood function for the two mortality models of this work is summarized in [Algorithm 1](#). [Algorithm 1](#) shows in further detail our implementation of the bootstrap particle filter to approximate the marginal likelihood $p_\theta(\bar{\boldsymbol{\mu}}_{1:T})$ for both the BSj and CIRj models. Note that we use the log-sum-exp trick, similar to [Dahlin and Schön \(2019\)](#), to avoid overflow or underflow problems in the numerical computations for the likelihood. The procedure formulated here closely follows the particle filter discussed by [Andrieu et al. \(2010\)](#); [Dahlin and Schön \(2019\)](#), among others.

4.3 Particle Markov Chain Monte Carlo Algorithm

The bootstrap particle filter discussed in the previous section is then integrated into a Markov chain Monte Carlo (MCMC) algorithm to obtain a sample of N_m draws $\{\theta^{[k]}\}_{k=1}^{N_m}$, from the posterior distribution of the parameters

$$p(\theta|\bar{\boldsymbol{\mu}}_{1:T}) = \frac{p(\theta)p(\bar{\boldsymbol{\mu}}_{1:T}|\Theta = \theta)}{\int p(\theta')p(\bar{\boldsymbol{\mu}}_{1:T}|\Theta = \theta') d\theta'},$$

where $p(\theta)$ is the parameter prior density.

Algorithm 1: Bootstrap filter to approximate $p(\bar{\boldsymbol{\mu}}_{1:T}|\Theta = \theta) = p_\theta(\bar{\boldsymbol{\mu}}_{1:T})$.

Result: Likelihood estimate $\hat{p}_\theta(\bar{\boldsymbol{\mu}}_{1:T})$
Input: Model parameters $\Theta = \theta$, average force of mortality data $\bar{\boldsymbol{\mu}}_{1:T}$, number of particles N_p
/ Construct the measurement model, and the state-transition model */*
 Compute the risk-neutral parameters using (21) or (22);
 Compute the coefficients \bar{A} and \bar{B} and the covariance matrix \mathbf{H} of the measurement error of the measurement equation (33);
 Define $g_\theta(\bar{\boldsymbol{\mu}}_t|L_t)$ and $f_\theta(L_t|L_{t-1})$ as functions of $\bar{\boldsymbol{\mu}}_t$, $L_t = (X_t, Z_t, \Delta N_t)$, and X_{t-1} following (33) and (34);
 Initialize the log-likelihood $\log \hat{p}_\theta(\bar{\boldsymbol{\mu}}_{1:0}) = 0$;
 Set the weights $w_0^{(i)} := \frac{1}{N_p}$, $i = 1, \dots, N_p$;
/ Likelihood estimation at time $t = 1, \dots, T$ */*
for $t = 1$ **to** T **do**
 Sample $a_t^{(1:N_p)}$ from a multinomial distribution with probabilities $(w_{t-1}^{(1)}, \dots, w_{t-1}^{(N_p)})$;
for $i = 1$ **to** N_p **do**
 Sample $\Delta N_t^{(i)} \sim \text{Bernoulli}(\lambda^P)$ and $\varpi_t^{(i)} \sim \mathbf{N}(0, 1)$;
 Sample $u^{(i)} \sim \text{Unif}(0, 1)$;
if $u^{(i)} \leq \rho^P$ **then**
 | Sample $Z_t^{(i)} \sim \text{Exponential}(\phi_1^P)$;
else
 | Sample $Z_t^{(i)} \sim \text{Exponential}(\phi_2^P)$ and set $Z_t^{(i)} \leftarrow -Z_t^{(i)}$;
end
 Compute

$$X_t^{(i)} = \begin{cases} X_{t-1}^{a_t^{(i)}} - \xi^P X_{t-1}^{a_t^{(i)}} + \sigma \varpi_t^{(i)} + Z_t^{(i)} \Delta N_t^{(i)} & \text{for BSj} \\ X_{t-1}^{a_t^{(i)}} + \xi^P (\eta^P - X_{t-1}^{a_t^{(i)}}) + \sigma \sqrt{X_{t-1}^{a_t^{(i)}} \varpi_t^{(i)} + Z_t^{(i)} \Delta N_t^{(i)}} & \text{for CIRj} \end{cases}$$

 (for CIRj, set $X_t^{(i)} \leftarrow |X_t^{(i)}|$);
 Form $L_t^{(i)} := (X_t^{(i)}, Z_t^{(i)}, \Delta N_t^{(i)})$;
end
 Compute the unnormalized weights $v_t^{(i)}$ using equation (36);
end
 Compute the normalized weights

$$w_t^{(i)} = \frac{v_t^{(i)}}{\sum_{j=1}^{N_p} v_t^{(j)}} = \frac{\exp\{\tilde{v}_t^{(i)}\}}{\sum_{j=1}^{N_p} \exp\{\tilde{v}_t^{(j)}\}}$$

where $\tilde{v}_t^{(i)} := \log(v_t^{(i)}) - \max\{\log v_t^{(1:N_p)}\}$;
 Compute the log predictive likelihood estimate

$$\log \hat{p}_\theta(\bar{\boldsymbol{\mu}}_t | \bar{\boldsymbol{\mu}}_{1:t-1}) = \max\{\log v_t^{(1:N_p)}\} + \log \left(\sum_{i=1}^{N_p} \exp\{\tilde{v}_t^{(i)}\} \right) - \log(N_p)$$

and update the log-likelihood estimate $\log \hat{p}_\theta(\bar{\boldsymbol{\mu}}_{1:t}) = \log \hat{p}_\theta(\bar{\boldsymbol{\mu}}_{1:t-1}) + \log \hat{p}_\theta(\bar{\boldsymbol{\mu}}_t | \bar{\boldsymbol{\mu}}_{1:t-1})$;
 Compute the filtered state estimate $\hat{L}_t = \sum_{i=1}^{N_p} w_t^{(i)} L_t^{(i)}$;
 Store particles $L_t^{(1:N_p)}$, ancestor indices $a_t^{(1:N_p)}$, normalized weights $w_t^{(1:N_p)}$, and filtered state estimate \hat{L}_t ;

Before we discuss the MCMC algorithm, we recall the restrictions on our model parameters. In the CIRj model we require

$$x_0, \xi^P, \eta^P, \sigma, \phi_1^P, \phi_2^P \in (0, \infty), \quad \lambda^P, \rho^P \in [0, 1].$$

For convenience, we remove the parameter constraints by sampling from the posterior distribution of their unconstrained transformations. Therefore, for the nonnegative parameters, we use the log-transformation, while for the parameters constrained between 0 and 1, we use the logit transformation. Therefore, let $\tilde{\Theta}$ denote the set of unconstrained parameters

$$\tilde{\Theta} := \{x_0^l, \xi^{P,l}, \eta^{P,l}, \sigma^l, \lambda^{P,\ell}, \rho^{P,\ell}, \phi_1^{P,l}, \phi_2^{P,l}, \zeta, \psi_0, \psi_1, r_1, r_2, r_c\}, \quad (38)$$

where the superscript l denotes the log transformation and the superscript ℓ denotes the logit transformation. For the BSj model, we perform the same transformations, except for the mean reversion rate ξ^P . Since all transformations applied are one-to-one functions, the marginal likelihood $p_\theta(\bar{\boldsymbol{\mu}}_{1:T})$ can still be calculated by first recovering the original parameters Θ from the given transformed parameter values $\tilde{\Theta}$ and then proceeding as discussed in Section 4.2.

To generate the Markov chain $\{\tilde{\theta}^{[k]}\}_{k=1}^{N_m}$ sampled from the parameter posterior distribution $p(d\tilde{\theta}|\bar{\boldsymbol{\mu}}_{1:T})$, we first set a parameter prior $p(\tilde{\theta})$ and a proposal density $q(\tilde{\theta}'|\tilde{\theta}^{[k-1]})$ from which we sample new candidate parameter values conditional on the previous state $\tilde{\theta}^{[k-1]}$ of the Markov chain. The proposal density must be selected such that the support of the target distribution is covered. Given an initial value $\tilde{\theta}^{[0]}$, the rest of the Markov chain is generated via a *particle Metropolis-Hastings (PMH)* algorithm.

Pseudocode Outline 4.1 (Particle Metropolis-Hastings). *Given the parameter prior $p(\tilde{\theta})$, the proposal density $q(\tilde{\theta}'|\tilde{\theta}^{[k-1]})$, an initial value $\tilde{\theta}^{[0]}$, and the inputs for the particle filter in Algorithm 1, do for $k = 1, \dots, N_m$*

1. *Sample a candidate $\tilde{\theta}^* \sim q(\tilde{\theta}^*|\tilde{\theta}^{[k-1]})$*
2. *Reverse the parameter transformations to obtain θ^* and $\theta^{[k-1]}$*
3. *Using the particle filter approximations for $p_{\theta^*}(\bar{\boldsymbol{\mu}}_{1:T})$ and $p_{\theta^{[k-1]}}(\bar{\boldsymbol{\mu}}_{1:T})$ calculate the acceptance probability*

$$\alpha(\tilde{\theta}^*, \tilde{\theta}^{[k-1]}) = \min \left\{ 1, \frac{p(\tilde{\theta}^*)p_{\theta^*}(\bar{\boldsymbol{\mu}}_{1:T})}{p(\tilde{\theta}^{[k-1]})p_{\theta^{[k-1]}}(\bar{\boldsymbol{\mu}}_{1:T})} \frac{q(\tilde{\theta}^{[k-1]}|\tilde{\theta}^*)}{q(\tilde{\theta}^*|\tilde{\theta}^{[k-1]})} \right\}.$$

4. *Generate a sample $u \sim \text{Unif}(0, 1)$ and set $\tilde{\theta}^{[k]} = \tilde{\theta}^*$ if $u \leq \alpha(\tilde{\theta}^*, \tilde{\theta}^{[k-1]})$; otherwise, reject $\tilde{\theta}^*$ and set $\tilde{\theta}^{[k]} = \tilde{\theta}^{[k-1]}$.*

The parameters are sampled by using Gibbs sampling for blocks of parameters (Chib and Greenberg 1995; Hastings 1970), defined as follows:

1. initial state: $\tilde{\Theta}_1 = x_0^l$

2. diffusion parameters: $\tilde{\Theta}_2 = (\xi^{P,l}, \eta^{P,l}, \sigma^l)$ (for the CIRj model), $\tilde{\Theta}_2 = (\xi^P, \sigma^l)$ (for the BSj model)
3. jump parameters: $\tilde{\Theta}_3 = (\lambda^{P,\ell}, \rho^{P,\ell}, \phi_1^{P,l}, \phi_2^{P,l})$
4. change of measure parameters: $\tilde{\Theta}_4 = (\zeta, \psi_0, \psi_1)$
5. measurement error parameters: $\tilde{\Theta}_5 = (r_1, r_2, r_c)$

We assume that the parameter prior $p(\tilde{\theta})$ can be decomposed as $p(\tilde{\theta}) = \prod_{g=1}^5 p(\tilde{\theta}_g)$, i.e. the parameter groups are independently distributed *a priori*.

To generate candidate parameter values, we use a multivariate Gaussian distribution as the proposal distribution for each of the parameter groups. However, instead of specifying a constant covariance matrix of the Gaussian proposal distribution, we use the robust adaptive Metropolis (RAM) algorithm, proposed by Vihola (2012), in which the covariance matrix is updated after each MCMC iteration. The RAM algorithm is a modification of standard adaptive Metropolis algorithms (see for example Andrieu and Thoms 2008; Haario et al. 2001, for a summary) in which one can simultaneously “learn” about the shape of the target distribution and coerce a particular acceptance rate. The RAM algorithm also has the advantage that it does not rely on the empirical covariance of previously generated samples, and thus it can be used to approximate target distributions with heavy tails. In our implementation, the RAM algorithm is implemented as follows.

Pseudocode Outline 4.2 (Robust Adaptive Metropolis (RAM)). *Given the parameter prior $p(\tilde{\theta}) = \prod_{g=1}^5 p(\tilde{\theta}_g)$, an initial parameter value $\tilde{\theta}^{[0]} = (\tilde{\theta}_1^{[0]}, \dots, \tilde{\theta}_5^{[0]})$, initial covariance matrices $\mathbf{S}_1^{[0]}, \dots, \mathbf{S}_5^{[0]}$, a number $\varepsilon \in (\frac{1}{2}, 1]$, the desired mean acceptance rate $\alpha^* \in (0, 1)$, and the inputs for the particle filter in Algorithm 1, do*

For $k = 1, \dots, N_m$

For $g = 1, \dots, 5$

1. *Generate a candidate $\tilde{\theta}_g^* = \tilde{\theta}_g^{[k-1]} + \mathbf{C}_g^{[k-1]} \mathbf{r}_k$, where $\mathbf{C}_g^{[k-1]}$ is the lower Cholesky factor of $\mathbf{S}_g^{[k-1]}$ and $\mathbf{r}_k \sim \mathbf{N}(0, \mathbf{I})$ of the appropriate size.*
2. *Using the particle filter approximations, calculate the acceptance probability*

$$\alpha_g^{[k]} = \min \left\{ 1, \frac{p(\tilde{\theta}_g^*)}{p(\tilde{\theta}_g^{[k-1]})} \frac{p_g^*(\bar{\boldsymbol{\mu}}_{1:T})}{p_g^{[k-1]}(\bar{\boldsymbol{\mu}}_{1:T})} \right\}.$$

Observe that the proposal density does not appear here since the Gaussian proposal is symmetric, i.e. $q(\tilde{\theta}^ | \tilde{\theta}^{[k-1]}) = q(\tilde{\theta}^{[k-1]} | \tilde{\theta}^*)$.*

See Remark 4.3 for further details on the likelihood values $p_g^(\bar{\boldsymbol{\mu}}_{1:T})$ and $p_g^{[k-1]}(\bar{\boldsymbol{\mu}}_{1:T})$ introduced here in view of the Gibbs sampling procedure.*

3. *Sample $u \sim \text{Unif}(0, 1)$ and set $\tilde{\theta}_g^{[k]} = \tilde{\theta}_g^*$ if $u \leq \alpha_g^{[k]}$; otherwise, reject $\tilde{\theta}_g^*$ and set $\tilde{\theta}_g^{[k]} = \tilde{\theta}_g^{[k-1]}$.*

4. Calculate the updated covariance matrix $\mathbf{S}_g^{[k]}$ using the equation

$$\mathbf{S}_g^{[k]} = \mathbf{C}_g^{[k-1]} \left(\mathbf{I} + k^{-\varepsilon} (\alpha_g^{[k]} - \alpha^*) \frac{\mathbf{r}_m \mathbf{r}_m^\top}{|\mathbf{r}_m|^2} \right) (\mathbf{C}_g^{[k-1]})^\top,$$

where $|\mathbf{r}_m|$ is the Euclidean norm of \mathbf{r}_m .

Remark 4.3. For the g th group, with $g \geq 2$, the calculation of the particle filter-approximated likelihood values $p_g^*(\bar{\boldsymbol{\mu}}_{1:T})$ and $p_g^{[k-1]}(\bar{\boldsymbol{\mu}}_{1:T})$ use the k th iteration values $\tilde{\theta}_1^{[k]}, \dots, \tilde{\theta}_{g-1}^{[k]}$. For example, if $g = 3$, then

$$\begin{aligned} p_3^*(\bar{\boldsymbol{\mu}}_{1:T}) &:= p(\bar{\boldsymbol{\mu}}_{1:T} | \tilde{\Theta}_1 = \tilde{\theta}_1^{[k]}, \tilde{\Theta}_2 = \tilde{\theta}_2^{[k]}, \tilde{\Theta}_3 = \tilde{\theta}_3^*, \tilde{\Theta}_4 = \tilde{\theta}_4^{[k-1]}, \tilde{\Theta}_5 = \tilde{\theta}_5^{[k-1]}) \\ p_3^{[k-1]}(\bar{\boldsymbol{\mu}}_{1:T}) &:= p(\bar{\boldsymbol{\mu}}_{1:T} | \tilde{\Theta}_1 = \tilde{\theta}_1^{[k]}, \tilde{\Theta}_2 = \tilde{\theta}_2^{[k]}, \tilde{\Theta}_3 = \tilde{\theta}_3^{[k-1]}, \tilde{\Theta}_4 = \tilde{\theta}_4^{[k-1]}, \tilde{\Theta}_5 = \tilde{\theta}_5^{[k-1]}) \end{aligned}$$

Furthermore, since the parameter groups are assumed to be independent in the prior, only the prior of the current group g is needed in the calculation of $\alpha_g^{[k]}$.

4.4 Implementation

Configuration of the Prior Distribution

The prior distribution of the parameters must respect the constraint that survival curves are monotonically decreasing functions of residual life time, eventually decaying to zero at very old ages. Furthermore, candidate parameters must also result in survival curves that capture the observed rectangularization and the expansion phenomenon in recent mortality trends (see for example [Ebeling et al. 2018](#); [Macdonald et al. 1998](#); [van Raalte 2021](#); [Wilmoth and Horiuchi 1999](#)).

To this end, for each model, we generated parameter values from independent uniform distributions with widely spaced bounds and produced the corresponding survival curve using the exponential affine form (8) until we obtained a survival curve that has the “correct” shape. The corresponding parameter values are then taken to be the mean of the prior distribution of the unknown parameters.

To ensure that all candidate parameters generated through the PMH algorithm also produce reasonably shaped survival curves, we use a Gaussian parameter prior for each group of parameters in the PMH-RAM algorithm. For each parameter group, we assume a diagonal prior covariance matrix. The parameters of the Gaussian prior are provided in [Table 1](#).

Implementation of the PMH Algorithm

In the implementation of the robust adaptive Metropolis algorithm, we set the initial covariance matrix equal to the covariance matrix of the prior distribution and used the values $\varepsilon = 0.6$ and $\alpha^* = 0.234$, following [Vihola \(2012\)](#). In our numerical experiments, we found no considerable improvement in the algorithm’s performance for other values of ε and α^* .

Initial parameter values were chosen randomly from the collection of parameter values from which the prior distribution means were extracted. Several initial values were used to assess whether the convergence of the RAM/PMH algorithm is robust with respect to the starting values.

We use $N_p = 100$ particles in each run of the particle filter to estimate the log-likelihood at each iteration of the PMH/RAM algorithm. This choice provides a satisfactory level of precision in the log-likelihood estimator, while allowing us to minimize the computational time required to generate sufficiently many Markov chain samples.¹³

The construction of Markov chains with varying starting points was implemented using parallel computing. To this end, we used an RStudio Server (R version 4.0.2) on the Katana computational cluster housed in UNSW Sydney (<https://doi.org/10.26190/669x-a286>).¹⁴ For each model, the PMH algorithm was run for 72 hours. This resulted to around 12,000 Markov chain values for each parameter set. We discarded the first 4,000 iterations as the burn-in and used the remaining 8,000 samples for parameter inference. To reduce the autocorrelation in the samples used for parameter inference, we only retain the 5th draw of the sample after burn-in.

Parameters of the no-jump version of our mortality models were also estimated using the PMH algorithm. In this case, we skip the proposal step for the jump parameters and set $\lambda_0^P = 0$ in the particle filter function in the proposal step for all other parameter groups.¹⁵ For brevity, the no-jump versions of our models shall be referred to as the 1FBS and 1FCIR models.

5 Results

5.1 Parameter Estimates

The parameter estimates were derived from the Markov chains as follows. After removing the burn-in samples and thinning the remainder with a lag of 5, we calculate the mean of the thinned sample and reverse the transformation applied (see equation (38)) to obtain an estimate for the model parameters. For example, we set $\hat{x}_0 = \exp\left\{\frac{1}{N_{\text{thin}}} \sum_{k=1}^{N_{\text{thin}}} x_0^{l[k]}\right\}$, where N_{thin} is the number of observations in the thinned sample.¹⁶ Estimates of the jump parameters λ^P and ρ^P are

¹³For general guidance see [Doucet et al. \(2015\)](#).

¹⁴We note that the computations performed in this paper can also be implemented on a standard multi-core personal computer at no additional computational expense. The use of parallel computing is only useful in generating Markov chains from different initial points in parallel.

¹⁵Alternatively, a Kalman filter-based algorithm can be designed to estimate the parameters of the one-factor Blackburn-Sherris model with no jumps, as it has a linear Gaussian state-space representation (see e.g. [Blackburn and Sherris 2013](#)). However, the same cannot be done for the CIR model with no jumps since its state transition distribution is a non-central χ^2 distribution; thus, a Kalman filter approach to the CIR model is only an approximation. Thus, our approach using the particle filter is a unifying methodology as it is able to handle nonlinear and/or non-Gaussian state-space models.

¹⁶Estimating x_0 by first undoing the transformation on x_0^l and taking the mean results to a larger estimate since, by Jensen's inequality, we have

$$\tilde{x}_0 := \frac{1}{N_{\text{thin}}} \sum_{k=1}^{N_{\text{thin}}} e^{x_0^{l[k]}} \geq \exp\left\{\frac{1}{N_{\text{thin}}} \sum_{k=1}^{N_{\text{thin}}} x_0^{l[k]}\right\} = \hat{x}_0.$$

Table 1: Model parameter estimates based on the mean of the parameter posterior and goodness-of-fit measures.

	BSj	CIRj	BS	CIR
x_0	0.010020	0.010094	0.010174	0.010159
ξ^P	-0.074686	0.001992	-0.072817	0.001993
η^P	-	0.005996	-	0.005987
σ	0.000598	0.008394	0.000549	0.008385
λ^P	0.030039	0.004991	-	-
ρ^P	0.849973	0.899953	-	-
ϕ_1^P	1.490847	7.000935	-	-
ϕ_2^P	1700.668653	700.092918	-	-
ζ	0.019804	-9.013548	0.047416	-9.013413
ψ_0	-3.498794	-2.501417	-3.428912	-2.499303
ψ_1	0.605229	-1.199590	0.670053	-1.200624
r_1	-19.010625	-23.006194	-19.205511	-23.006434
r_2	-1.551787	-1.063222	-1.554578	-1.065118
r_c	-15.003413	-15.000239	-14.808785	-14.999570
Neg. Log-lik.	-9061.315453	-8767.583533	-9053.121720	-8781.302991
RMSE $_{\bar{\mu}}$	0.001922	0.004516	0.001846	0.004440
RMSE $_S$	0.004740	0.005812	0.004968	0.005742
AIC	-18096.630906	-17507.167065	-18080.243440	-17534.605982
BIC	24777.369094	28664.832935	11593.756560	15437.394018

treated similarly; we first take the mean of the thinned samples for $\lambda^{P,\ell}$ and $\rho^{P,\ell}$ then undo the logit transformation. Parameter estimates are shown in Table 1.

The latent process $\{X_t\}_{t=1}^T$ is then estimated by the filtered state estimates $\{\hat{X}_t\}_{t=1}^T$ derived from the particle filter using the final parameter estimates shown in Table 1 and 10,000 particles. The filtered state estimates and the estimated factor loadings for each model can be found in Figures 1 and 2, respectively. Figure 1, in particular, provides evidence towards the improvement of life expectancy of more recent cohorts, as implied by the decreasing behavior of the filtered estimates of the instantaneous mortality intensity, regardless of the stochastic model. In addition, all models exhibit the same fluctuation pattern (see e.g. the filtered state estimates around the cohort born in 1900), regardless of whether there are jumps in the underlying latent state dynamics. The presence of jumps, however, tends to induce a change in the overall level of the latent state estimates. In particular, the jumps affect the Blackburn-Sherrie models more than the CIR models, as seen by the larger gap between the filtered state estimates in Figure 1 and the difference in the minimum value of $B(t, T)$ in Figure 2. Furthermore, from Figure 2,

We use the estimator \hat{x}_0 since the PMH algorithm generates the posterior distribution for x_0^l , not x_0 (see also Ungolo et al. 2020, Section 7.2). Alternatively, one can design the PMH algorithm so that the prior is set on the original parameters, but proposals are generated for the transformed parameters (see for example Dahlin and Schön 2019, Section 6.3). This approach accounts for the parameter transformation by including a Jacobian term in the calculation of the acceptance probability.

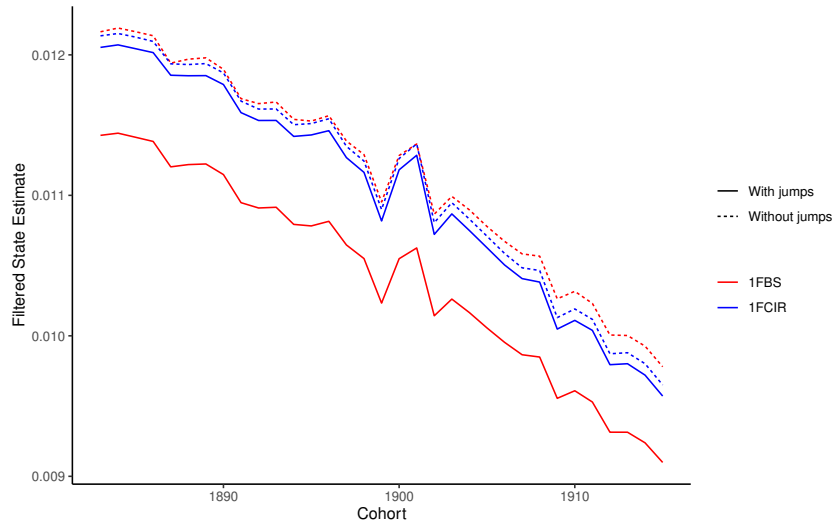


Figure 1: Filtered state estimates \hat{X}_t by model

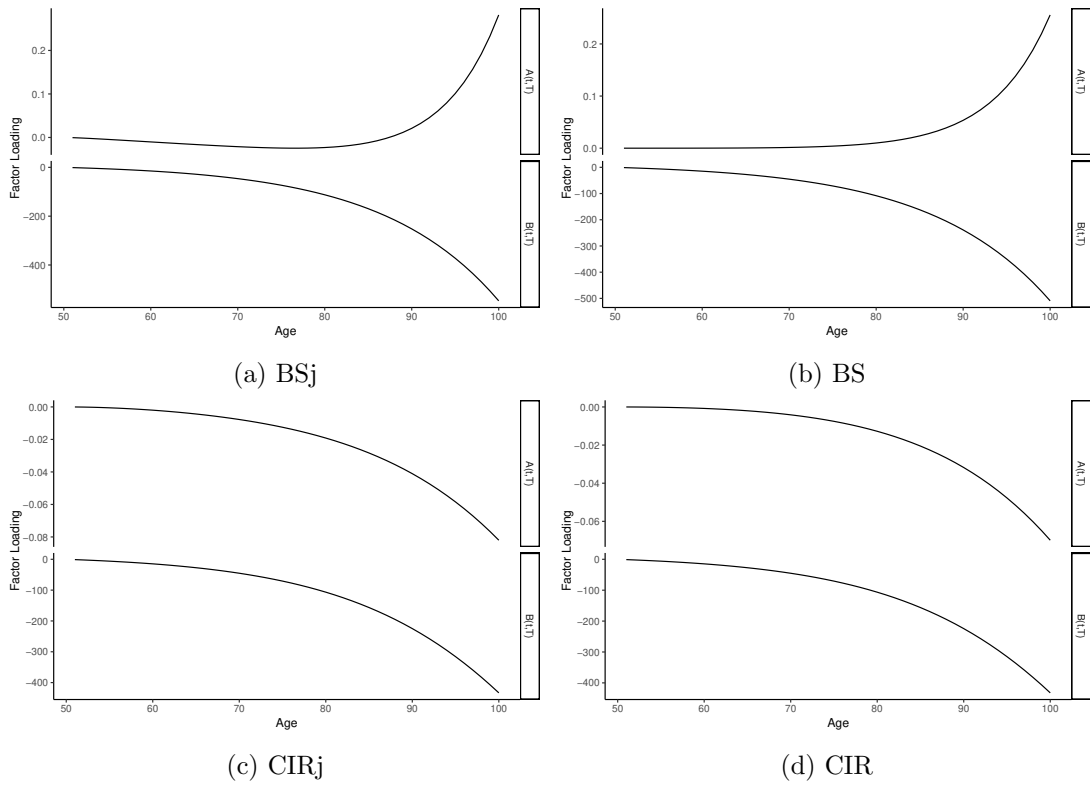


Figure 2: Factor loadings $A(t, T)$ and $B(t, T)$ by model.

we see that the magnitude of $B(t, T)$ is larger at older ages, so the latent state X_t thus tends to have a greater effect at older ages.

At this juncture, we emphasize that the current approach is unable to determine which cohorts experienced mortality shocks, since the inclusion of jumps in the latent factor dynamics induces a parallel shift in latent state estimates, rather than pronounced peaks or troughs. This is because, due to the relatively small number of cohorts in the USA data set, it is not possible to filter out the unobserved jump parts separately. Nonetheless, the jumps in our models can capture the variability in mortality rates beyond what can be captured by a pure diffusion process.

For the 1FBSj model and its no-jump counterpart, applying a nonnegativity constraint on ξ^P (by considering the transformed variable $\xi^{P,l}$) actually produces worse results. Under this model, the latent factor process has a long-run mean of $\eta^P = 0$. Thus, if we force the latent factor process to be mean-reverting, then the particle filter estimate of $\{X_t\}$ tends to contain negative values. This is not appropriate since, in the one-factor setting, $\{X_t\}$ represents the instantaneous mortality intensity which needs to be a nonnegative process. Thus, it is better not to impose a nonnegativity constraint on ξ^P for the 1FBSj model.

In addition to model parameter estimates, Table 1 also contains estimates of the negative log-likelihood, the root mean squared error (RMSE) for both the average force of mortality and the survival curves, and the AIC and BIC for each model. Here, the RMSE for the survival curves $S(t, T)$ (RMSE_S) is computed as

$$\text{RMSE}_S = \sqrt{\frac{1}{50 \times 33} \sum_{k=1}^{50} \sum_{t=1}^{33} [S(t, t+k) - \hat{S}(t, t+k)]^2},$$

as we have 33 cohorts and 50 ages in the data set; a similar formula was used for the RMSE for $\bar{\mu}$ ($\text{RMSE}_{\bar{\mu}}$). The AIC and BIC were computed as

$$\begin{aligned} \text{AIC} &= 2 \times \text{Neg. Log-Lik.} + 2 \times \sharp(\Theta) \\ \text{BIC} &= 2 \times \text{Neg. Log-Lik.} + 2 \times \sharp(\Theta) \times 33 \times 50, \end{aligned}$$

where $\sharp(\Theta)$ is the number of parameters.

Additional summaries and visualizations of the Markov chains generated by our implementation of the PMH algorithm can be found in Appendix S2.2. In particular, Table 2 provides the mean and standard deviation of the (marginal) transformed parameter posterior distribution based on the thinned sample after excluding the burn-in period. Figures S3 to S10 show the trace plots and the marginal posterior densities of the transformed parameters. The trace plots show several Markov chains with different initial points and indicate that convergence was attained after a few thousand iterations. In the trace plots, we changed the y -axis scales after the burn-in period of 4000 iterations. The density plots combines all chains for each parameter and excludes the burn-in period. For all parameters and for all models, we observe that the marginal posterior distributions are almost symmetric. A few marginal posteriors appear to be

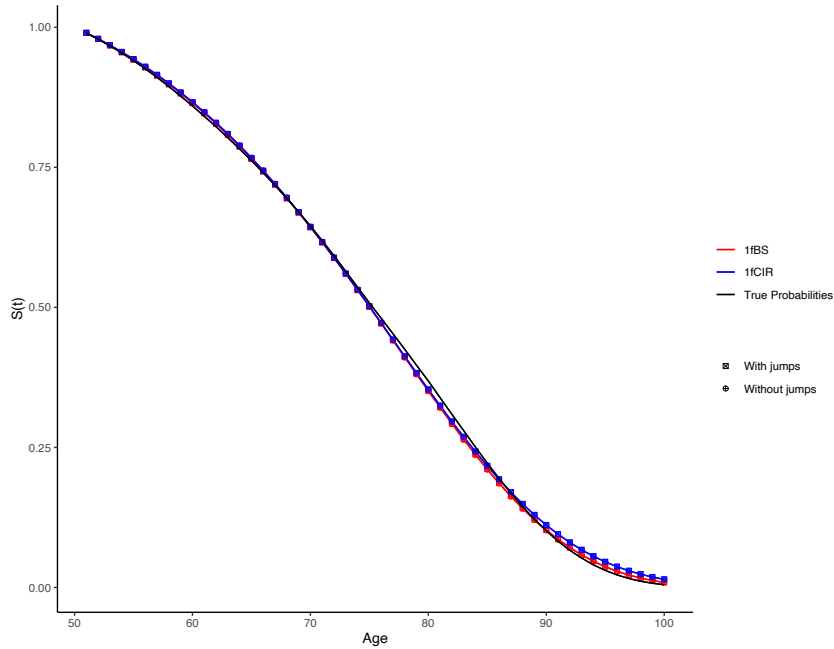


Figure 3: Estimated 1915 cohort survival probabilities

bimodal, but the two modes are very close to each other.

5.2 In-Sample Fit Assessment

We now discuss how well the one-factor affine mortality models (with and without jumps) estimate the survival probabilities and average force of mortality relative to the data used in the parameter estimation.

Based on the error metrics, AIC, and BIC shown in Table 1, the BSj and BS models provide a better fit to both the average force of mortality and survival probabilities across all cohorts, compared to their CIR counterparts. In addition, the presence of jumps in the latent factor dynamics only provides a marginal improvement in model fit. This is consistent with the very small magnitude of the estimated λ^P , especially in the CIRj model. Interestingly, the BSj and BS models return a lower negative log-likelihood despite having fewer parameters than their CIR counterparts.

Figure 3 shows the estimated survival curve for the cohort of USA males currently aged 50 born in the year 1915 alongside the observed survival curve. Overall, all models are able to capture the general shape of the survival curve. However, we observe that all models tend to underestimate the survival probability at the middle ages (around 75 to 85) and overestimate the survival probability at the older ages (around 85 to 100).¹⁷ While the underestimation in the middle ages is similar among the four models, the Blackburn-Sherris models are more accurate than the CIR models when estimating the survival probability at older ages. We also note that, for both the Blackburn-Sherris and the CIR models, survival probability estimates from the

¹⁷Bravo and Nunes (2021) observed a similar phenomenon when modelling survival curves for the USA individuals born in 1885.

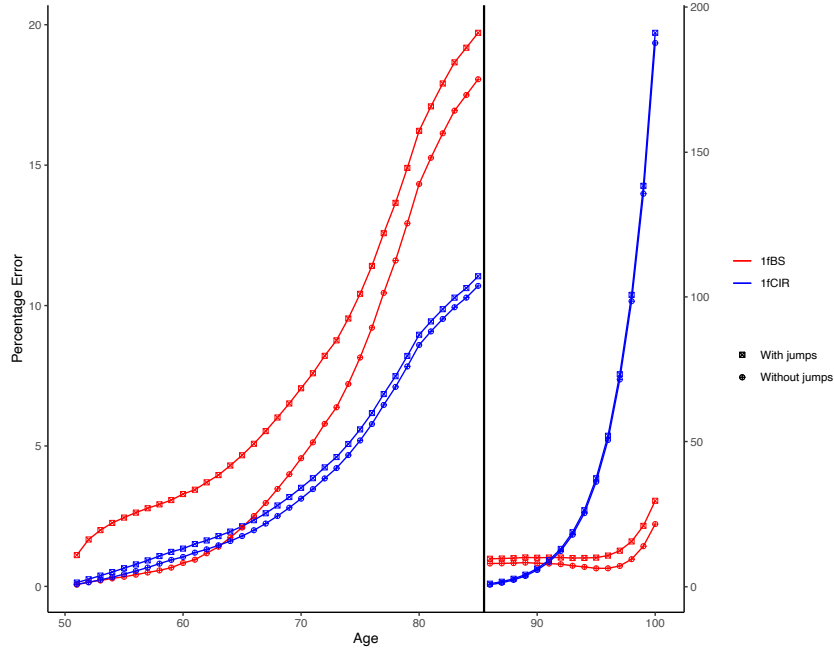


Figure 4: In-sample mean absolute percentage errors (MAPE) for estimated survival probabilities

model with and without jumps are very close. Separate comparisons between the actual and estimated survival curves for each model can be seen in Figures S1 and S2.

We assess the accuracy of the models with respect to the entire estimation data set by plotting the mean absolute percentage errors (MAPE) by age. The MAPE for the survival probabilities is computed for each age k by

$$\text{MAPE}_k = \frac{1}{33} \sum_{t=1}^{33} \left| \frac{S(t, t+k) - \hat{S}(t, t+k)}{S(t, t+k)} \right| \times 100\%, \quad k = 1, \dots, 50$$

where $S(t, t+k)$ is the observed k -year survival probability for an individual aged 50 born in the year t and $\hat{S}(t, t+k)$ is the estimated survival probability computed using the final parameter estimates and the exponential-affine formula, $\hat{S}(t, t+k) := \exp\{\hat{A}(k) + \hat{B}(k)\hat{X}_t\}$ under the relevant model (see Section 4.1 for the relevant notation; recall that we considered 33 cohorts born from 1883 to 1915). The MAPE for the average force of mortality is calculated similarly by replacing $S(t, t+k)$ by $\bar{\mu}_{t,k}$.

We plot the MAPE for the average force of mortality and the survival probabilities in Figures 5 and 4, respectively. In Figure 4, we change the y -axis scaling at age 85 since the MAPE for the CIR models increases sharply towards the older ages. Both figures show that while the CIR model tends to more accurately estimate mortality rates at the middle ranges, these models perform poorly at very old ages compared to the Blackburn-Sherris models. Models with no jumps also tend to perform marginally better than their counterparts with jumps.

Figure 6 shows the heat map of raw residuals (observed less estimated value) for survival probabilities. Notably, the Blackburn-Sherris models more accurately estimate the survival

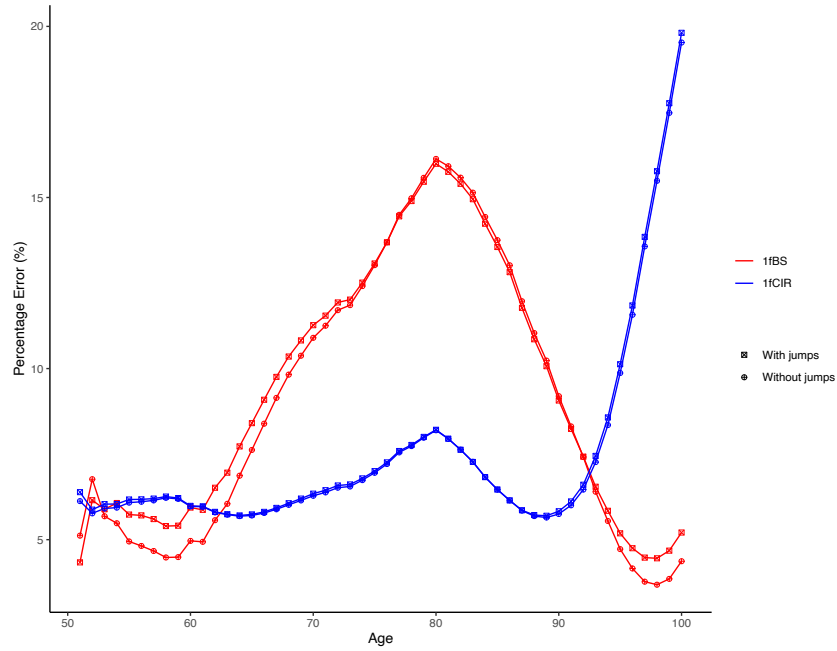


Figure 5: In-sample mean absolute percentage errors (MAPE) for estimated average force of mortality

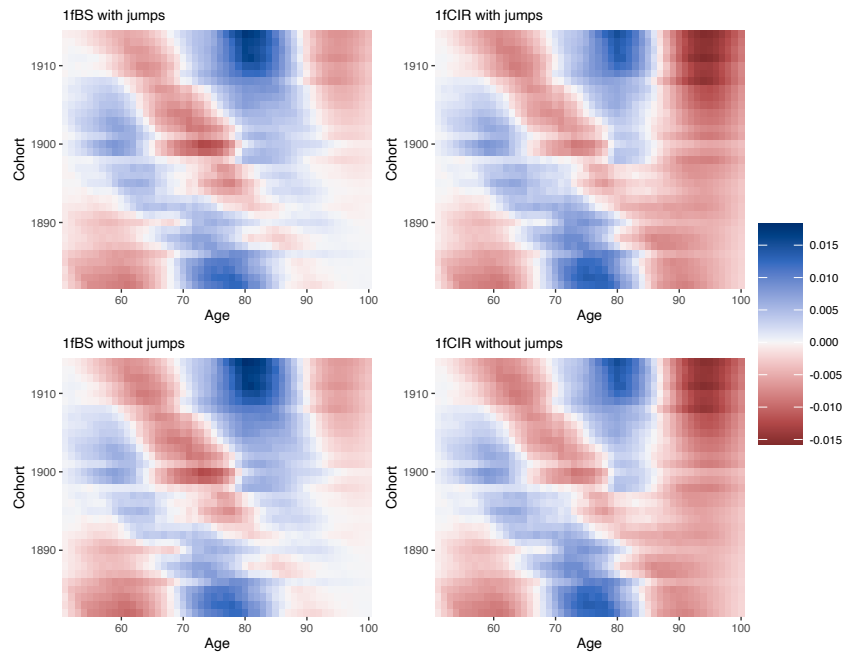


Figure 6: Heat map of residuals for estimated survival probabilities

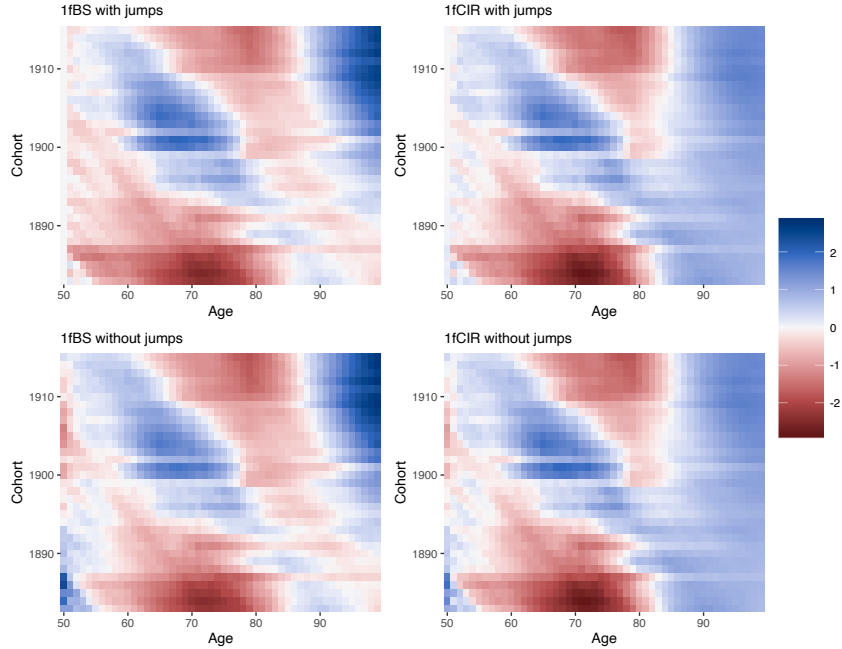


Figure 7: Heat map of standardized residuals for estimated average force of mortality

probabilities across all ages, compared to the CIR models, with errors only occurring in the more recent cohorts in the data set. In contrast, the CIR models tend to accurately capture the survival probabilities before age 90, and thereafter consistently overestimate the survival probability at older ages. We note however that there is no considerable difference in the residuals between the models with and without jumps. Furthermore, we note that the in-sample fit achieved by our one-factor models is inferior to that achieved by the three-factor models considered by [Huang et al. \(2022\)](#) and [Ungolo et al. \(2021\)](#), especially at older ages.

Additional insight on the model fit can be gleaned from the heat map of standardized residuals for the average force of mortality in [Figure 7](#). For each cohort t , the standardized residual ν_t is computed as $\nu_t = \mathbf{L}^{-1}(\bar{\boldsymbol{\mu}}_t - \hat{\boldsymbol{\mu}}_t)$, where \mathbf{L} is a lower-triangular 50×50 matrix such that $\widehat{\text{Cov}}(\bar{\boldsymbol{\mu}}_t) = \mathbf{L}\mathbf{L}^\top$ and $\bar{\boldsymbol{\mu}}_t$ and $\hat{\boldsymbol{\mu}}_t$ are the observed and estimated average force of mortality, respectively. From [\(29\)](#), we have $\widehat{\text{Cov}}(\bar{\boldsymbol{\mu}}_t) = H + \bar{B}\text{Var}(X_t)\bar{B}^\top$; we approximate $\text{Var}(X_t)$ via the sample variance of the particles $\{X_t^{(i)}\}_{i=1, \dots, N_p}$. The standardized residuals show that the CIR and CIRj models consistently underestimate the average force of mortality at older ages, although we observe larger standardized residuals with the BS and BSj models, especially in the more recent cohorts. However, the smaller magnitude of standardized residuals in the CIR and CIRj models may be due to the larger values of $\widehat{\text{Cov}}(\bar{\boldsymbol{\mu}}_t)$ under these models.

In both residual plots, we observe a period effect which manifests in the diagonals, given our age-cohort setting. In particular, we note that in the 1950s and 1970s, the observed $\bar{\boldsymbol{\mu}}_t$ is lower than model estimates. These may refer to a period of increased improvements in mortality outcomes past World War II. [Njenga and Sherris \(2011\)](#) also identified a similar trend in USA mortality rates for males around this period: there is a slight increase in mortality rates in the 1950s and in the 1970s, with mortality rates trending downward thereafter (see [Figure 1](#) of [Njenga and Sherris \(2011\)](#)). This analysis also shows that the jumps in our models are unable

to capture period effects, since the jumps are indexed by the cohort year.

In terms of the overall ability to capture the dynamics of the average force of mortality and the survival curves across cohorts, the above analysis shows that the Blackburn-Sherris model outperforms the CIR model.

5.3 Forecasting Performance

In this analysis, we forecast the survival curve for the cohort born in 1916 and compare the forecast to the observed survival curve. Forecasts can be generated by calculating the conditional mean of the latent factor X_{1916} given $X_{1915} = \hat{X}_{1915}$. In general, if $h \geq 0$ denotes our forecast horizon (e.g. the one-cohort-ahead forecast corresponds to $h = 1$), then forecasts of the average force of mortality and the survival probability can be calculated as

$$\begin{aligned}\hat{\mu}_{t+h,k} &= -\frac{\hat{A}(k)}{k} - \frac{\hat{B}(k)}{k} \mathbb{E}_{\mathbb{P}}[X_{t+h}|X_t] \\ \hat{S}(t+h, t+h+k) &= \exp\{\hat{A}(k) + \hat{B}(k) \mathbb{E}_{\mathbb{P}}[X_{t+h}|X_t]\},\end{aligned}$$

for each $k = 1, \dots, 50$. These equations generate forecasts that are optimal under a quadratic loss function (Christensen et al. 2011, Section 5.1). For the one-factor Blackburn-Sherris model with jumps, the conditional mean $\mathbb{E}_{\mathbb{P}}[X_{t+h}|X_t]$ is available in closed form and is given by

$$\mathbb{E}_{\mathbb{P}}[X_{t+h}|X_t] = e^{-\xi^P h} X_t + \frac{\lambda^P \mathbb{E}_{\mathbb{P}}[Z]}{\xi^P} (1 - e^{-\xi^P h})$$

(see Appendix S1 for the derivation). The conditional mean for the no-jump Blackburn-Sherris model can be obtained by setting $\lambda^P = 0$. In contrast, the conditional mean for the one-factor CIR model with jumps is not available in closed form, so we resort to Monte Carlo simulations using the state-transition equation (27) to generate forecasts under the CIR model.

In Figure 8, we compare forecasts of the 1916 cohort survival curve with the observed survival curve. Immediately, we observe that forecasts from the CIR models outperform those from the Blackburn-Sherris models. The forecast from the CIR model also mirrors the features observed from the in-sample estimates in that there is some degree of underestimation in the middle ages and overestimation in the older ages. In contrast, the forecasts from the Blackburn-Sherris models consistently underestimate the survival probability at all ages. This inaccuracy arises from the non-mean-reverting feature of the latent factor under the Blackburn-Sherris models. Given our parameter estimates, forecasts for X_t under the Blackburn-Sherris models tend to increase with t , implying a worsening in mortality rates for future cohorts. In addition, since $\hat{B}(k) > 0$ given our parameter estimates, increasing values in X_t result to smaller survival probabilities. From this figure, we also see that there is no noticeable difference in the forecasting performance of the CIR model with and without jumps. On the contrary, the presence of jumps in the Blackburn-Sherris model leads to a bigger underestimation in survival probabilities at younger ages, compared to the no-jump case. The discrepancy, however, tapers off towards the middle and older ages.

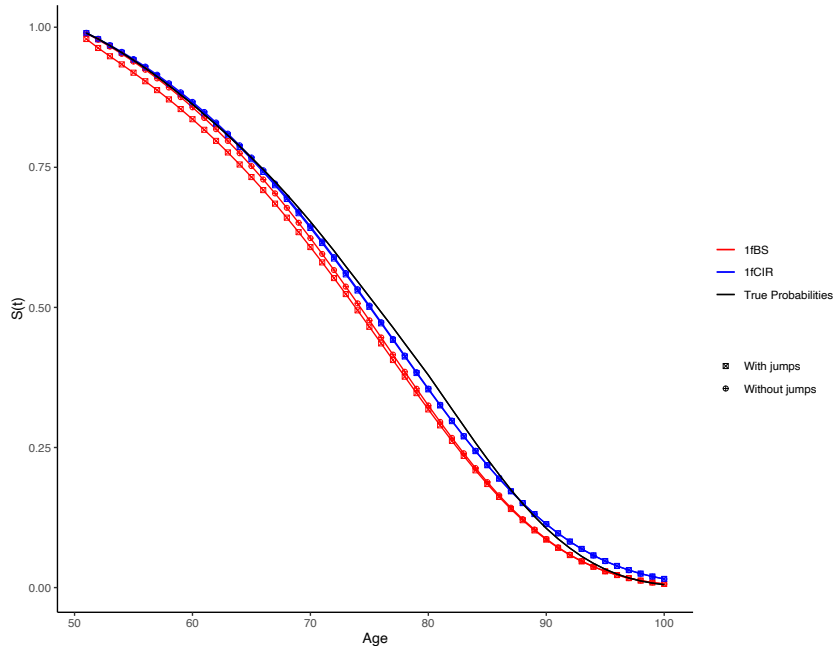
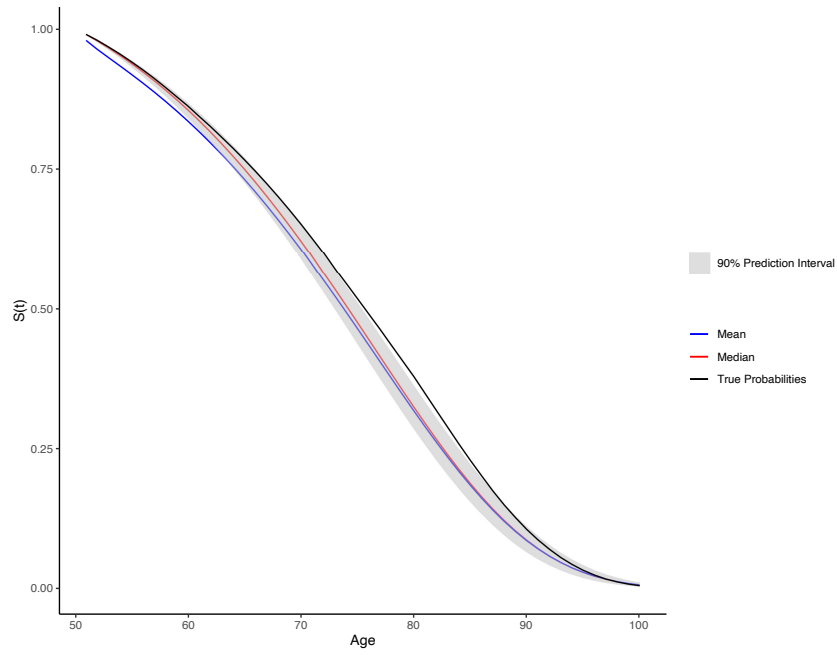


Figure 8: 1916 survival curve forecasts based on the estimated conditional mean of the latent factor

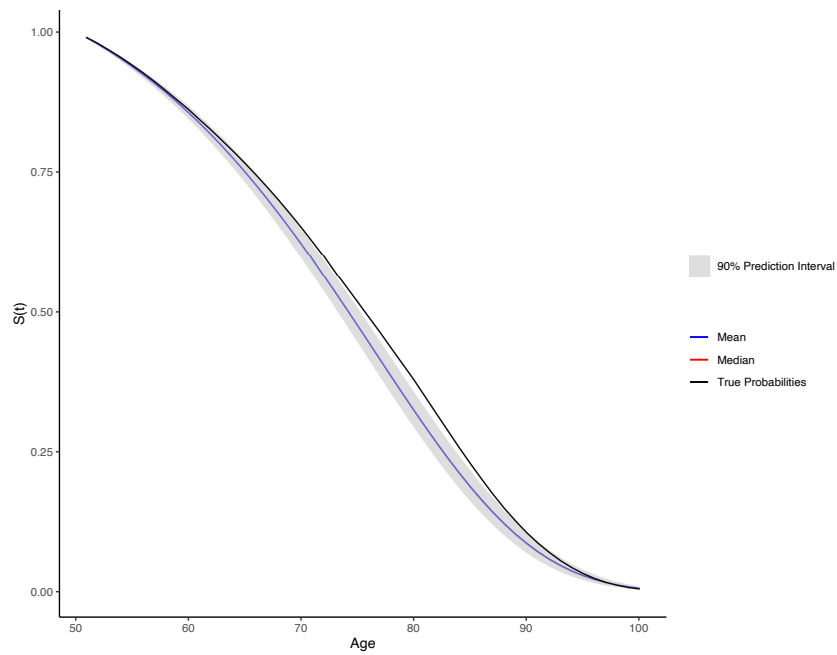
In the one factor setting, prediction bands for the survival curve can be generated by replacing $\mathbb{E}_{\mathbb{P}}[X_{1916}|X_{1915}]$ in the forecasting equations with the appropriate quantile of the simulated values of X_{1916} . Given that extreme values may occur in the simulations due to the presence of jumps, we also calculate the “median” survival curve which corresponds to the median of the simulated values of X_{1916} .

Figures 9 and 10 show forecasts of the 1916 cohort survival curve using the mean and median value of X_{1916} and a 90% prediction band based on the simulated quantiles. From these figures, it is clear that CIR has a superior forecasting performance compared to the Blackburn-Sherris models, since the true survival curve is contained within the CIR models’ prediction band, even at older ages. In addition, it is only with the Blackburn-Sherris model with jumps that we observe a discrepancy between the survival curve forecasts generated using the mean and the median of X_{1916} , indicating that the evolution of the survival curve may be more sensitive to jumps if the latent factor indeed follows the 1FBSj dynamics.

The jumps in our mortality models allow us to forecast scenarios where there are mortality jumps; that is, our mortality rate forecasts incorporate possible stress or shock mortality events. This may be one factor which contributes to the quality of the forecasts from the CIRj models. On the other hand, we may attribute the poor forecasting performance of the Blackburn-Sherris models to the use of just one latent factor in the model. We note that the forecasts we generate from the CIR and CIRj models are comparable to those of [Huang et al. \(2022\)](#) and [Ungolo et al. \(2021\)](#) who used three-factor models (with no jump). This may be due to our use of the particle filter, which more accurately captures the non-Gaussian distribution of the latent factor under the CIR and CIRj models. In contrast, [Huang et al. \(2022\)](#) and [Ungolo et al. \(2021\)](#) employ a quasi-linear Kalman filter for the three-factor CIR model, which approximates the latent state

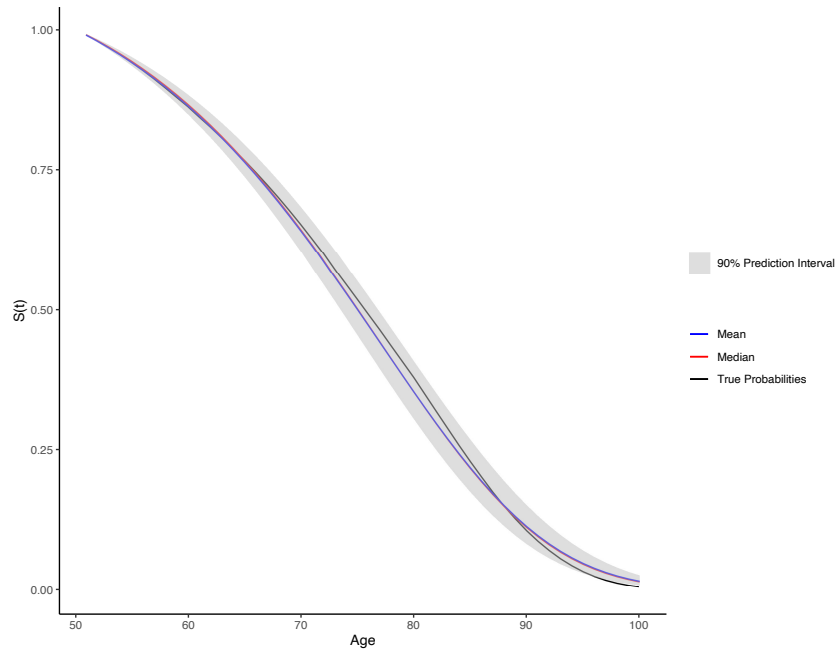


(a) with jumps

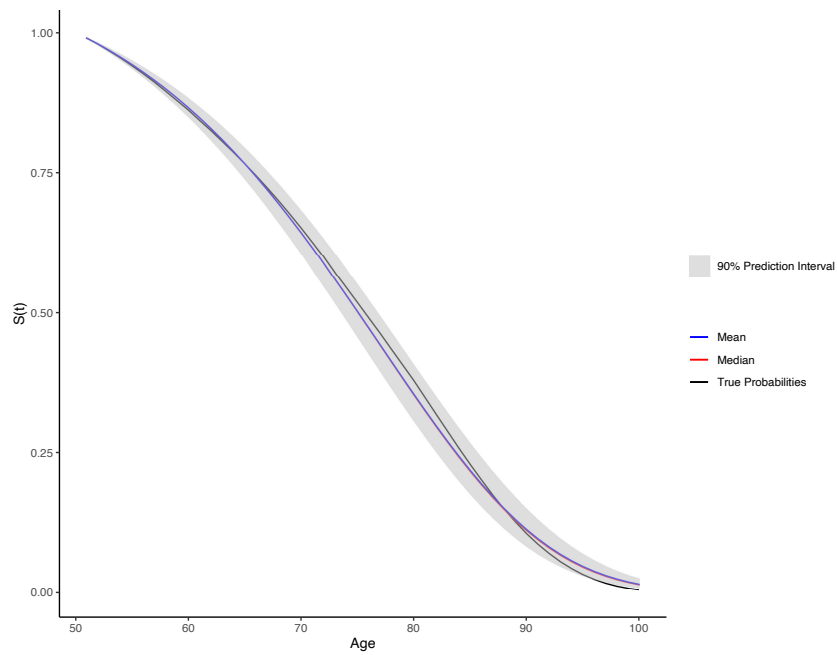


(b) without jumps

Figure 9: 1916 survival curve forecasts with the one-factor Blackburn-Sherris model with a 90% prediction interval



(a) with jumps



(b) without jumps

Figure 10: 1916 survival curve forecasts with the one-factor Cox-Ingersoll-Ross model with a 90% prediction interval

distribution with a Gaussian distribution. Nevertheless, a multi-factor extension of our affine mortality models with jumps may be necessary to improve the general forecasting performance.

5.4 Sensitivity Analysis

In this section, we investigate how sensitive forecasts of future age-cohort survival curves are to changes in instantaneous mortality intensity model parameters. Specifically, assuming the mortality model for cohorts born from 1883 to 1915 is correct (see Table 1), we analyze how forecasts of the survival curve for the cohort born in 1916 will be affected by changes in the dynamics of the mortality intensity process. We do this by introducing shocks to each model parameter based on the standard deviation of the parameter posterior distribution (see Table 2). Shocks in the value of X_{1915} used in projections of X_{1916} are based on the standard deviation computed using the filtering distribution estimated via the particle filter given the parameter values in Table 1. Figures S11 to S14 display the forecasts for the 1916 cohort survival curve resulting from the shocks in the model parameters alongside the observed survival curve.

We find that survival curve forecasts for all models are robust to shocks in the model parameters and the latent mortality intensity for the 1915 cohort, when the shocks are consistent with the degree of parameter uncertainty implied by the parameter posterior distribution. That is, survival curve forecasts obtained by shocking the transformed parameter by units of its respective posterior standard deviation (i.e. $\pm i$ standard deviations from the posterior mean, $i = 1, 2, 3, 4$) are very close to the forecasts obtained in the previous section.

However, since the forecasts are similar despite introducing parameter shocks, there is no substantial improvement in the accuracy of the forecasts. In Figures S11 and S12, forecasts obtained with the Blackburn-Sherris model with and without jumps still exhibit noticeable errors at the middle ages relative to the true survival curve. Similarly, Figures S13 and S14 show that there is still a slight underestimation and a slight overestimation of survival probabilities at middle and older ages, respectively, under the CIR model with and without jumps. This suggests that, to improve the accuracy of forecasts, we must revisit the mortality model structure itself, i.e. consider multi-factor mortality models with jumps.

6 Conclusion

In this paper, we studied a stochastic mortality model where the mortality intensity process is an affine jump-diffusion process. This specification allows one to model sudden positive or negative mortality shocks (in the form of jumps in the mortality intensity) which may not be captured with pure diffusion mortality models (see e.g. Blackburn and Sherris 2013; Jevtić et al. 2013; Xu et al. 2020b). This specification is also mathematically advantageous in that survival probabilities are available in closed form up to the solution of a system of ODEs, which in our setting can also be solved in closed form. Thus, affine jump-diffusion mortality models lend themselves easily towards pricing life insurance and other longevity-linked products within continuous-time valuation frameworks. In our setting, mortality shocks manifest as cohort

effects, rather than period effects, and are persistent shocks whose affect permeates across cohorts.

Using one-factor Blackburn-Sherris and Cox-Ingersoll-Ross mortality intensity models with jumps and age-cohort mortality rates for USA males born from 1883 to 1915, we analyze the evolution of the unobserved mortality intensity across cohorts. This approach eventually allows us to forecast survival probabilities for future cohorts. As such, our work extends that of [Bravo and Nunes \(2021\)](#) and [Luciano and Vigna \(2008\)](#) who adopt a single-cohort approach in their use of affine jump-diffusion models to analyze survival probabilities. Since the corresponding state-space representation of our model is nonlinear and non-Gaussian, we propose a particle filter-based Metropolis-Hastings algorithm to estimate our model parameters and to facilitate statistical inference. This Bayesian approach is informed by historical mortality trends and parameter values that lead to the usual shape of survival curves. This approach more accurately captures the non-Gaussian distribution of the latent factor, in contrast to quasi-Kalman filtering approaches used by [Huang et al. \(2022\)](#), [Jevtić and Regis \(2021\)](#), and [Ungolo et al. \(2021\)](#) for CIR-type models.

Our analysis shows that the Blackburn-Sherris models, with and without jumps, have better in-sample fit compared to their CIR counterparts and that the presence of jumps in the latent factor dynamics introduce only a marginal improvement in fit. In particular, we found that the CIR models tend to overestimate the survival probabilities at older ages. A simulation-based forecasting analysis, however, shows that the CIR models yield more accurate forecasts of the survival curves. Nevertheless, the inclusion of jumps in our model allows us to incorporate stress or shock events when making mortality rate forecasts for future cohorts. A sensitivity analysis also shows that survival curve forecasts are robust relative to the parameter uncertainty implied by the posterior parameter distributions.

This investigation can be extended in several directions. First, our age-cohort mortality model can be modified so that the jumps manifest as period effects, since mortality shocks tend to be identified through historical events (e.g. wars, pandemics, medical advancements) rather than through cohorts. Second, forecasting performance may be improved by modelling the mortality intensity as the sum of multiple latent factors, each modelled as an affine jump-diffusion process. The parameter estimation framework set forth in this paper can be extended to the multi-factor case, but may result to added computational complexity due to the increase in the dimension of the parameter space.¹⁸ Third, since this paper uses complete cohort data, the parameter estimation methodology must be modified if incomplete cohort data are used for forecasting mortality rates for yet-to-be-born cohorts. Lastly, the parameter estimation framework proposed here can accommodate an extension of the multi-population mortality models of [Jevtić and Regis \(2019\)](#) to include jumps, consistent with what was proposed by [Regis and Jevtić \(2022\)](#). Work along these lines are currently in progress and shall be reported in future papers.

¹⁸In the multi-factor setting, we can then also consider the arbitrage-free Nelson-Siegel model with jumps.

Author Contributions

LPDM Garces: Conceptualization, methodology, software, formal analysis, investigation, writing–original draft preparation, writing–reviewing and editing, visualization; **J Kolar:** Software, investigation, writing–review and editing, visualization; **M Sherris:** Conceptualization, formal analysis, writing–review and editing, supervision, funding acquisition; **F Ungolo:** Methodology, formal analysis, writing–review and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This research was supported by the Australian Research Council Centre of Excellence in Population Ageing Research (Project Number CE170100005).

Data Availability

This analysis uses publicly accessible mortality data from the [Human Mortality Database](#) (downloaded 5 June 2022).

Acknowledgments

This research includes computations using the computational cluster Katana supported by Research Technology Services at UNSW Sydney. We also thank Yuxin Zhou for providing valuable feedback on an earlier version of this manuscript. F. Ungolo also acknowledges financial support from ERGO Center of Excellence in Insurance.

Declaration of Interests

The authors declare no conflict of interest.

References

- Andrieu, C., Doucet, A., and Holenstein, R. 2010. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society B* 72, 3 (2010), 269–342. <https://doi.org/10.1111/j.1467-9868.2009.00736.x>
- Andrieu, C. and Thoms, J. 2008. A tutorial on adaptive MCMC. *Statistics and Computing* 18 (2008), 343–374. <https://doi.org/10.1007/s11222-008-9110-y>
- Biffis, E. 2005. Affine processes for dynamic mortality and actuarial valuations. *Insurance: Mathematics and Economics* 37 (2005), 443–468. <https://doi.org/10.1016/j.insmatheco.2005.05.003>

- Blackburn, C. and Sherris, M. 2013. Consistent dynamic affine mortality models for longevity risk applications. *Insurance: Mathematics and Economics* 43 (2013), 64–73. <https://doi.org/10.1016/j.insmatheco.2013.04.007>
- Booth, H. and Tickle, L. 2008. Mortality modelling and forecasting: a review of methods. *Annals of Actuarial Science* 3, 1-2 (2008), 3–43. <https://doi.org/10.1017/S1748499500000440>
- Bravo, J. M. 2021. Pricing survivor bonds with affine-jump diffusion stochastic mortality models. In *The 5th International Conference on E-Commerce, E-Business and E-Government (ICEEG '21), April 28-30, 2021*. ACM, New York, 91–96. <https://doi.org/10.1145/3466029.3466037>
- Bravo, J. M. and Nunes, J. P. V. 2021. Pricing longevity derivatives via Fourier transforms. *Insurance: Mathematics and Economics* 96 (2021), 81–97. <https://doi.org/10.1016/j.insmatheco.2020.10.008>
- Cairns, A. J. G., Blake, D., and Dowd, K. 2006. Pricing death: frameworks for the valuation and securitization of mortality risk. *ASTIN Bulletin* 36, 1 (2006), 79–120. <https://doi.org/10.2143/AST.36.1.2014145>
- Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., and Balevich, I. 2009. A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal* 13, 1 (2009), 1–35. <https://doi.org/10.1080/10920277.2009.10597538>
- Chen, H. and Cox, S. H. 2009. Modeling mortality with jumps: applications to mortality securitization. *The Journal of Risk and Insurance* 76, 3 (2009), 727–751. <https://doi.org/10.1111/j.1539-6975.2009.01313.x>
- Cheredito, P., Filipović, D., and Kimmel, R. L. 2007. Market price of risk specifications for affine models: Theory and evidence. *Journal of Financial Economics* 83 (2007), 123–170. <https://doi.org/10.1016/j.jfineco.2005.09.008>
- Chib, S. and Greenberg, E. 1995. Understanding the Metropolis-Hastings algorithm. *The American Statistician* 49, 4 (1995), 327–335. <https://doi.org/10.2307/2684568>
- Christensen, J. H. E., Diebold, F. X., and Rudebusch, G. D. 2011. The affine arbitrage-free class of Nelson-Siegel term structure models. *Journal of Econometrics* 164 (2011), 4–20. <https://doi.org/10.1016/j.jeconom.2011.02.011>
- Dahl, M. 2004. Stochastic mortality in life insurance: market reserves and mortality-linked insurance contracts. *Insurance: Mathematics and Economics* 35 (2004), 113–136. <https://doi.org/10.1016/j.insmatheco.2004.05.003>
- Dahlin, J. and Schön, T. B. 2019. Getting started with particle Metropolis-Hastings for inference in nonlinear dynamical systems. *Journal of Statistical Software* 88 (2019), 1–41. <https://doi.org/10.18637/jss.v088.c02>
- D’Amato, V., Di Lorenzo, E., Haberman, S., Sagoo, P., and Sibillo, M. 2018. De-risking strategy: longevity spread buy-in. *Insurance: Mathematics and Economics* 79 (2018), 124–136. <https://doi.org/10.1016/j.insmatheco.2018.01.004>

- Doucet, A. and Johansen, A. M. 2011. A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later. In *The Oxford Handbook of Nonlinear Filtering*, Dan Crisan and Boris Rozovskiĭ (Eds.). Oxford University Press, New York, Chapter 24, 656–704.
- Doucet, A., Pitt, M. K., Deligiannidis, G., and Kohn, R. 2015. Efficient implementation of the Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika* 102, 2 (2015), 295–313. <https://doi.org/10.1093/biomet/asu075>
- Duffee, G. R. 2002. Term premia and interest rate forecasts in affine models. *The Journal of Finance* 57, 1 (2002), 405–443. <https://doi.org/10.1111/1540-6261.00426>
- Duffie, D., Pan, J., and Singleton, K. 2000. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68, 6 (2000), 1343–1376. <https://doi.org/10.1111/1468-0262.00164>
- Ebeling, M., Rau, R., and Baudisch, A. 2018. Rectangularization of the survival curve reconsidered: the maximum inner rectangle approach. *Population Studies: A Journal of Demography* 72, 3 (2018), 369–379. <https://doi.org/10.1080/00324728.2017.1414299>
- Filipović, D. 2009. *Term-Structure Models: A Graduate Course*. Springer, Germany. <https://doi.org/10.1007/978-3-540-68015-4>
- Fung, M. C., Ignatieva, K., and Sherris, M. 2019. Managing systematic mortality risk in life annuities: an application of longevity derivatives. *Risks* 7, 2 (2019), 1–25. <https://doi.org/10.3390/risks7010002>
- Gerber, H. U. and Shiu, E. S. W. 1994. Option pricing by Esscher transforms. *Transactions of the Society of Actuaries* 46 (1994), 99–140.
- Golightly, A. 2009. Bayesian filtering for jump-diffusions with application to stochastic volatility. *Journal of Computational and Graphical Statistics* 18, 2 (2009), 384–400. <https://doi.org/10.1198/jcgs.2009.07137>
- Gordon, N. J., Salmond, D. J., and Smith, A. F. M. 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proceedings F (Radar and Signal Processing)* 140, 2 (1993), 107–113. <https://doi.org/10.1049/ip-f-2.1993.0015>
- Haario, H., Saksman, E., and Tamminen, J. 2001. An adaptive Metropolis algorithm. *Bernoulli* 7, 2 (2001), 223–242. <https://doi.org/10.2307/3318737>
- Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 1 (1970), 97–109. <https://doi.org/10.2307/2334940>
- Huang, Z., Sherris, M., Villegas, A. M., and Ziveyi, J. 2022. Modelling USA age-cohort mortality: a comparison of multi-factor affine mortality models. *Risks* 10, 183 (2022), 1–28. <https://doi.org/10.3390/risks10090183>
- Human Mortality Database. 2002. 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 05 June 2022.
- Jevtić, P., Luciano, E., and Vigna, E. 2013. Mortality surface by means of continuous time cohort models. *Insurance: Mathematics and Economics* 53 (2013), 122–133. <https://doi.org/10.1016/j.insmatheco.2013.04.005>

- Jevtić, P. and Regis, L. 2019. A continuous-time stochastic model for the mortality surface of multiple populations. *Insurance: Mathematics and Economics* 88 (2019), 181–195. <https://doi.org/10.1016/j.insmatheco.2019.07.001>
- Jevtić, P. and Regis, L. 2021. A square-root factor-based multi-population extension of the mortality laws. *Mathematics* 9 (2021), 1–17. <https://doi.org/10.3390/math9192402>
- Johannes, M. S., Polson, N. G., and Stroud, J. R. 2009. Optimal filtering of jump diffusions: Extracting latent states from sset prices. *The Review of Financial Studies* 22, 7 (2009), 2759–2799. <https://doi.org/10.1093/rfs/hhn110>
- Johansen, A. M. and Doucet, A. 2008. A note on auxiliary particle filters. *Statistics and Probability Letters* 78 (2008), 1498–1504. <https://doi.org/10.1016/j.spl.2008.01.032>
- Karatzas, I. and Shreve, S. E. 1988. *Brownian Motion and Stochastic Calculus* (2nd ed.). Springer-Verlag New York, Inc., New York, USA.
- Kou, S. G. 2002. A jump-diffusion model for option pricing. *Management Science* 48, 8 (2002), 1086–1101.
- Luciano, E. and Vigna, E. 2008. Mortality risk via affine stochastic intensities: calibration and empirical relevance. *Belgian Actuarial Bulletin* 8, 1 (2008), 5–16.
- Macdonald, A. S., Cairns, A. J. G., Gwilt, P. L., and Miller, K. A. 1998. An international comparison of recent trends in population mortality. *British Actuarial Journal* 4, 1 (1998), 3–141.
- Merton, R. C. 1976. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Econometrics* 3 (1976), 125–144.
- Milevsky, M. A. and Promislow, S. D. 2001. Mortality derivatives and the option to annuitise. *Insurance: Mathematics and Economics* 29 (2001), 299–318. [https://doi.org/10.1016/S0167-6687\(01\)00093-2](https://doi.org/10.1016/S0167-6687(01)00093-2)
- Njenga, C. N. and Sherris, M. 2011. Longevity Risk and the Econometric Analysis of Mortality Trends and Volatility. *Asia-Pacific Journal of Risk and Insurance* 5, 2 (2011), 1–52. <https://doi.org/10.2202/2153-3792.1115>
- Øksendal, B. and Sulem, A. 2019. *Applied Stochastic Control of Jump Diffusions* (3rd ed.). Springer Nature Switzerland AG, Cham, Switzerland. <https://doi.org/10.1007/978-3-540-69826-5>
- Pitt, M. K. and Shephard, N. 1999. Filtering via simulation: Auxiliary particle filters. *J. Amer. Statist. Assoc.* 94, 446 (1999), 590–599. <https://doi.org/10.2307/2670179>
- Ramezani, C. A. and Zeng, Y. 1998. Maximum likelihood estimation of asymmetric jump-diffusion processes: Application to security prices. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=606361
- Regis, L. and Jevtić, P. 2022. Stochastic Mortality Models and Pandemic Shocks. In *Pandemics: Insurance and Social Protection*, María del Carmen Boado-Penas, Julia Eisenberg, and Şule Şahin (Eds.). Springer, Cham, Switzerland, Chapter 4, 61–74. https://doi.org/10.1007/978-3-030-78334-1_4

- Runggaldier, W. J. 2003. Jump-Diffusion Models. In *Handbook of Heavy-Tailed Distributions in Finance*, S. T. Rachev (Ed.). Elsevier Science B. V., Amsterdam, The Netherlands, Chapter 5, 170–209.
- Russo, V., Giacometti, R., Ortobelli, S., Rachev, S., and Fabozzi, F. J. 2011. Calibrating affine stochastic mortality models using term insurance premiums. *Insurance: Mathematics and Economics* 49 (2011), 53–60. <https://doi.org/10.1016/j.insmatheco.2011.01.015>
- Särkkä, S. 2013. *Bayesian Filtering and Smoothing*. Cambridge University Press, Cambridge, UK. <https://doi.org/10.1017/CB09781139344203>
- Schrager, D. F. 2006. Affine stochastic mortality. *Insurance: Mathematics and Economics* 38 (2006), 81–97. <https://doi.org/10.1016/j.insmatheco.2005.06.013>
- Ungolo, F., Kleinow, T., and Macdonald, A. S. 2020. A hierarchical model for the joint mortality analysis of pension scheme data with missing covariates. *Insurance: Mathematics and Economics* 91 (2020), 68–84.
- Ungolo, F., Sherris, M., and Zhou, Y. 2021. *Multi-factor, age-cohort, affine mortality models: a multi-country comparison*. Working Paper 2021/26. ARC Centre of Excellence in Population Ageing Research, Sydney, Australia. <https://www.cepar.edu.au/publications/working-papers/multi-factor-age-cohort-affine-mortality-models-multi-country-comparison>
- van Raalte, A. A. 2021. What have we learned about mortality patterns over the past 25 years? *Population Studies: A Journal of Demography* 75, sup1 (2021), 105–132. <https://doi.org/10.1080/00324728.2021.1967430>
- Vihola, M. 2012. Robust adaptive Metropolis algorithm with coerced acceptance rate. *Statistics and Computing* 22 (2012), 997–1008. <https://doi.org/10.1007/s11222-011-9269-5>
- Wilmoth, J. R. and Horiuchi, S. 1999. Rectangularization revisited: variability of age at death within human populations. *Demography* 36, 4 (1999), 475–495.
- Xu, Y., Sherris, M., and Ziveyi, J. 2020a. Continuous-time multi-cohort mortality modelling with affine processes. *Scandinavian Actuarial Journal* 2020, 6 (2020), 526–552. <https://doi.org/10.1080/03461238.2019.1696223>
- Xu, Y., Sherris, M., and Ziveyi, J. 2020b. Market price of longevity risk for a multi-cohort mortality model with application to longevity bond option pricing. *The Journal of Risk and Insurance* 87, 3 (2020), 571–595. <https://doi.org/10.1111/jori.12273>
- Zhou, Y., Sherris, M., Ziveyi, J., and Xu, M. 2021. An innovative design of flexible, bequest-enhanced life annuity with natural hedging. *Scandinavian Actuarial Journal* 0, 0 (2021), 1–22. <https://doi.org/10.1080/03461238.2021.1997795>
- Zhu, N. and Bauer, D. 2022. Modeling the risk in mortality projections. *Operations Research* 0, 0 (2022), 1–16. <https://doi.org/10.1287/opre.2021.2255>

Supplemental Materials

Affine Mortality Models with Jumps: Parameter Estimation and Forecasting

S1 Forecasting Equations for Blackburn-Sherris Model with Jumps

This appendix shows the derivation of a closed-form expression for the conditional mean $\mathbb{E}_{\mathbb{P}}[X_t|X_s]$ for any $0 \leq s \leq t$, where X is given by the one-factor Blackburn-Sherris model with jumps (18).

First, we note that equation (18) can be written as

$$dX_t = -(\xi^P X_{t-} - \lambda^P \mathbb{E}_{\mathbb{P}}[Z]) dt + \sigma dW_t^P + d\tilde{J}_t^P,$$

where $\tilde{J}_t^P := J_t - \lambda^P \mathbb{E}_{\mathbb{P}}[Z]$ is the compensated compound Poisson process under \mathbb{P} .

Next, we consider the process $Y_t = f(t, X_t)$, where $f(t, x) = -\xi^P e^{\xi^P t} X_t$. Using the Itô formula for jump-diffusion processes, we obtain

$$\begin{aligned} dY_t = & \left[-\xi^P e^{\xi^P t} X_{t-} - (\xi^P X_{t-} - \lambda^P \mathbb{E}_{\mathbb{P}}[Z])(-\xi^P e^{\xi^P t}) \right] dt + \sigma(-\xi^P e^{\xi^P t}) dW_t^P \\ & + \mathbb{E}_{\mathbb{P}}^Z \left[-\xi^P e^{\xi^P t} (X_{t-} + Z_{N_t}) + \xi^P e^{\xi^P t} X_{t-} - Z_{N_t}(-\xi^P e^{\xi^P t}) \right] \lambda^P dt \\ & + \left[-\xi^P e^{\xi^P t} (X_{t-} + Z_{N_t}) + \xi^P e^{\xi^P t} X_{t-} \right] d\tilde{N}_t^P, \end{aligned}$$

where $\tilde{N}_t^P := N_t - \lambda^P t$ is the compensated Poisson counting process under \mathbb{P} . Simplifying and integrating over the interval $[s, t]$ yields

$$Y_t = Y_s - \lambda^P \mathbb{E}_{\mathbb{P}}[Z] \xi^P \int_s^t e^{\xi^P u} du - \int_s^t \sigma \xi^P e^{\xi^P u} dW_u^P - \int_s^t \xi^P e^{\xi^P u} Z_{N_u} d\tilde{N}_u^P.$$

Putting back $Y_t = -\xi^P e^{\xi^P t} X_t$ and simplifying yields the expression

$$X_t = e^{-\xi^P(t-s)} X_s + \frac{\lambda^P \mathbb{E}_{\mathbb{P}}[Z]}{\xi^P} (1 - e^{-\xi^P(t-s)}) + \int_s^t \sigma e^{-\xi^P(t-u)} dW_u^P + \int_s^t e^{-\xi^P(t-u)} Z_{N_u} d\tilde{N}_u^P.$$

We can then take the conditional expectation of X_t given X_s , giving us

$$\mathbb{E}_{\mathbb{P}}[X_t|X_s] = e^{-\xi^P(t-s)} X_s + \frac{\lambda^P \mathbb{E}_{\mathbb{P}}[Z]}{\xi^P} (1 - e^{-\xi^P(t-s)}). \quad (\text{S1})$$

The stochastic integrals vanish upon taking the conditional expectation as they are independent of X_s and are increments of \mathbb{P} -martingale processes. For purposes of forecasting the annual mortality intensity given the previous year's mortality intensity, we can specialize (S1) by taking $s = t - 1$, giving us

$$\mathbb{E}_{\mathbb{P}}[X_t|X_{t-1}] = e^{-\xi^P} X_{t-1} + \frac{\lambda^P \mathbb{E}_{\mathbb{P}}[Z]}{\xi^P} (1 - e^{-\xi^P}). \quad (\text{S2})$$

In the no-jump case, conditional on X_{t-1} , X_t has a normal distribution whose mean and variance are given by

$$\mathbb{E}[X_t|X_{t-1}] = e^{-\xi^P} X_{t-1}, \quad \text{Var}[X_t|X_{t-1}] = \frac{\sigma^2}{2\xi^P}(1 - e^{2\xi^P}).$$

S2 Supplementary Figures

S2.1 Comparison with 1915 Survival Curve by Model

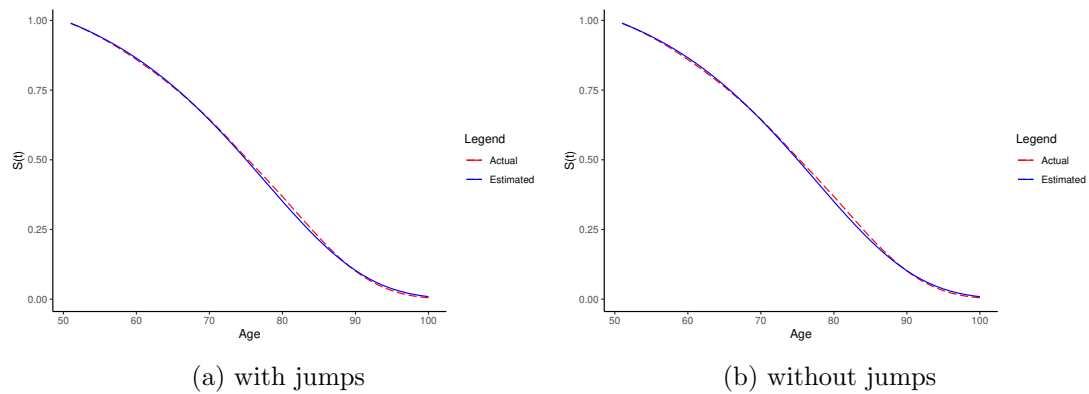


Figure S1: Comparison between actual and estimated 1915 survival curves under the one-factor Blackburn-Sherris model

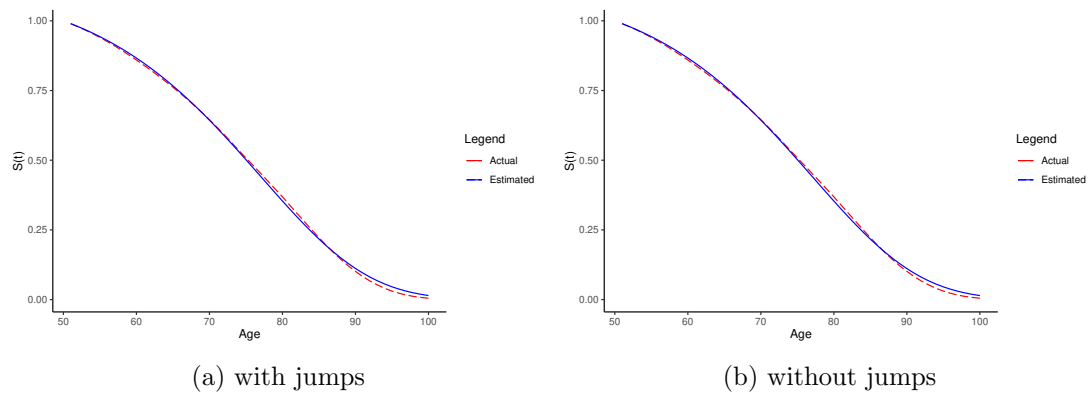


Figure S2: Comparison between actual and estimated 1915 survival curves under the one-factor CIR model

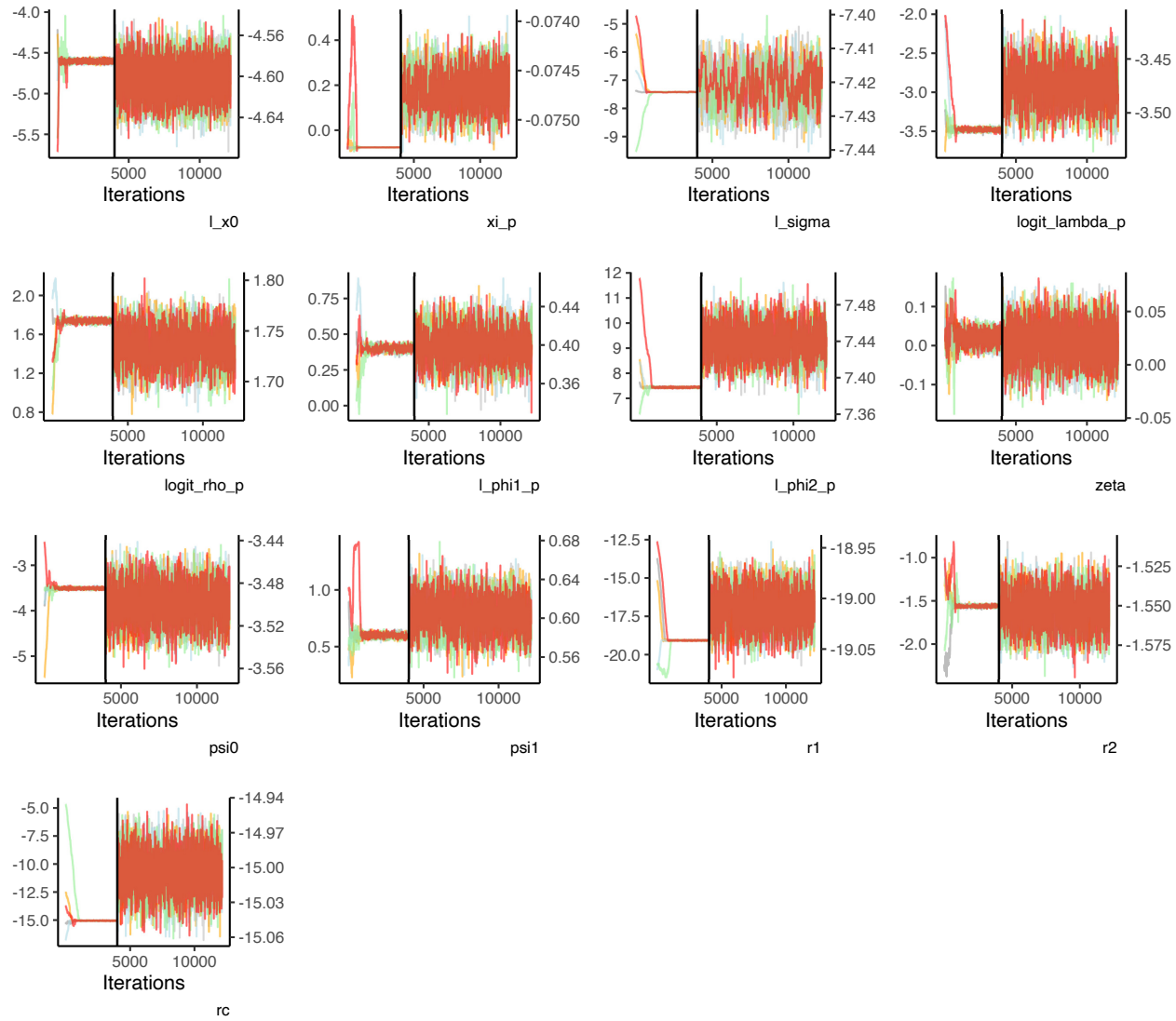
S2.2 MCMC Prior Distribution Parameters, Results, and Plots

Table 1: Means and standard deviations of the Gaussian parameter prior.

	BSj		CIRj		BS		CIR	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
x_0^l	-4.6052	0.0100	-4.6052	0.0100	-4.6052	0.0100	-4.6052	0.0100
ξ^P	-0.0750	0.0032	-	0.0100	-0.0730	0.0032	-	-
$\xi^{P,l}$	-	-	-6.2146	0.0100	-	-	-6.2146	0.0100
$\eta^{P,l}$	-	-	-5.1160	0.0100	-	-	-5.1160	0.0100
σ^l	-7.4186	0.0032	-4.8283	0.0100	-7.5056	0.0032	-4.8283	0.0100
$\lambda^{P,\ell}$	-3.4761	0.0100	-5.2933	0.0100	-	-	-	-
$\rho^{P,\ell}$	1.7346	0.0100	2.1972	0.0100	-	-	-	-
$\phi_1^{P,l}$	0.4055	0.0100	1.9459	0.0100	-	-	-	-
$\phi_2^{P,l}$	7.4384	0.0100	6.5511	0.0100	-	-	-	-
ζ	0.0200	0.0100	-9.0000	0.0100	0.0480	0.0100	-9.0000	0.0100
ψ_0	-3.5000	0.0100	-2.5000	0.0100	-3.4300	0.0100	-2.5000	0.0100
ψ_1	0.6000	0.0100	-1.2000	0.0100	0.6700	0.0100	-1.2000	0.0100
r_1	-19.0000	0.0100	-23.0000	0.0100	-19.2000	0.0100	-23.0000	0.0100
r_2	-1.5000	0.0100	-1.0000	0.0100	-1.5400	0.0100	-1.0000	0.0100
r_c	-15.0000	0.0100	-15.0000	0.0100	-14.8000	0.0100	-15.0000	0.0100

Table 2: Parameter posterior means and standard deviations after excluding the burn-in period and applying a thinning interval of 5.

	BSj		CIRj		BS		CIR	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
x_0^l	-4.603139	0.016629	-4.595778	0.020312	-4.587940	0.022389	-4.589426	0.024703
ξ^P	-0.074686	0.000167	-	-	-0.072817	0.000242	-	-
$\xi^{P,l}$	-	-	-6.218819	0.022660	-	-	-6.218233	0.028930
$\eta^{P,l}$	-	-	-5.116663	0.022118	-	-	-5.118181	0.028092
σ^l	-7.421121	0.005224	-4.780276	0.004773	-7.507142	0.008877	-4.781338	0.006017
$\lambda^{P,\ell}$	-3.474760	0.016696	-5.295118	0.023011	-	-	-	-
$\rho^{P,\ell}$	1.734386	0.016697	2.196699	0.023319	-	-	-	-
$\phi_1^{P,l}$	0.399344	0.016552	1.946044	0.022951	-	-	-	-
$\phi_2^{P,l}$	7.438777	0.016601	6.551213	0.023686	-	-	-	-
ζ	0.019804	0.016651	-9.013548	0.022387	0.047416	0.027966	-9.013413	0.029177
ψ_0	-3.498794	0.016532	-2.501417	0.023326	-3.428912	0.028169	-2.499303	0.028983
ψ_1	0.605229	0.016890	-1.199590	0.023450	0.670053	0.028313	-1.200624	0.030052
r_1	-19.010625	0.016769	-23.006194	0.023335	-19.205511	0.028272	-23.006434	0.029147
r_2	-1.551787	0.010409	-1.063222	0.010783	-1.554578	0.017838	-1.065118	0.013409
r_c	-15.003413	0.016515	-15.000239	0.022651	-14.808785	0.027855	-14.999570	0.029408



5

Figure S3: Trace plots for the parameter Markov chains generated by the PMH algorithm for the BSj model.

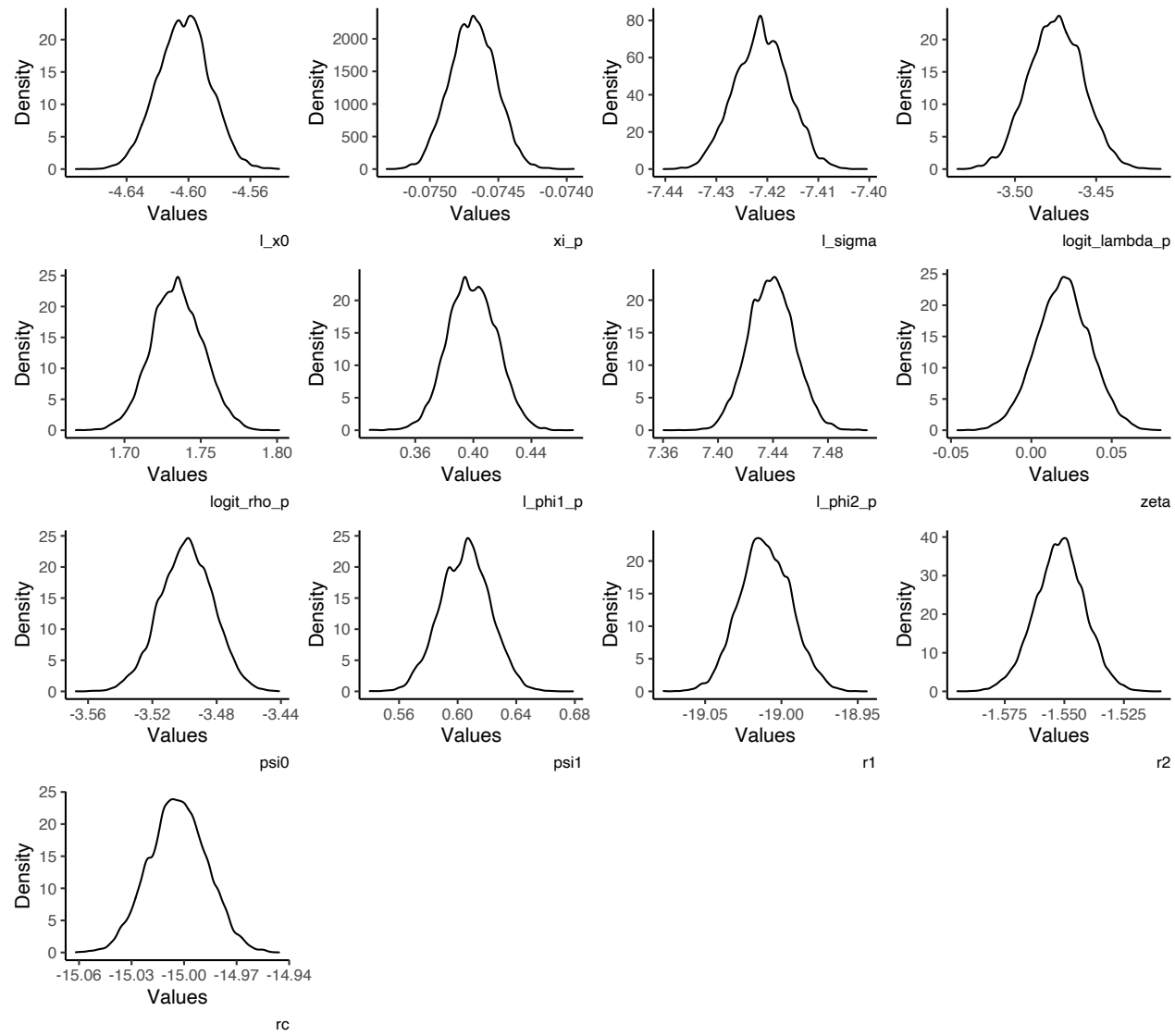


Figure S4: Density plots for the parameter Markov chains generated by the PMH algorithm for the BSj model.

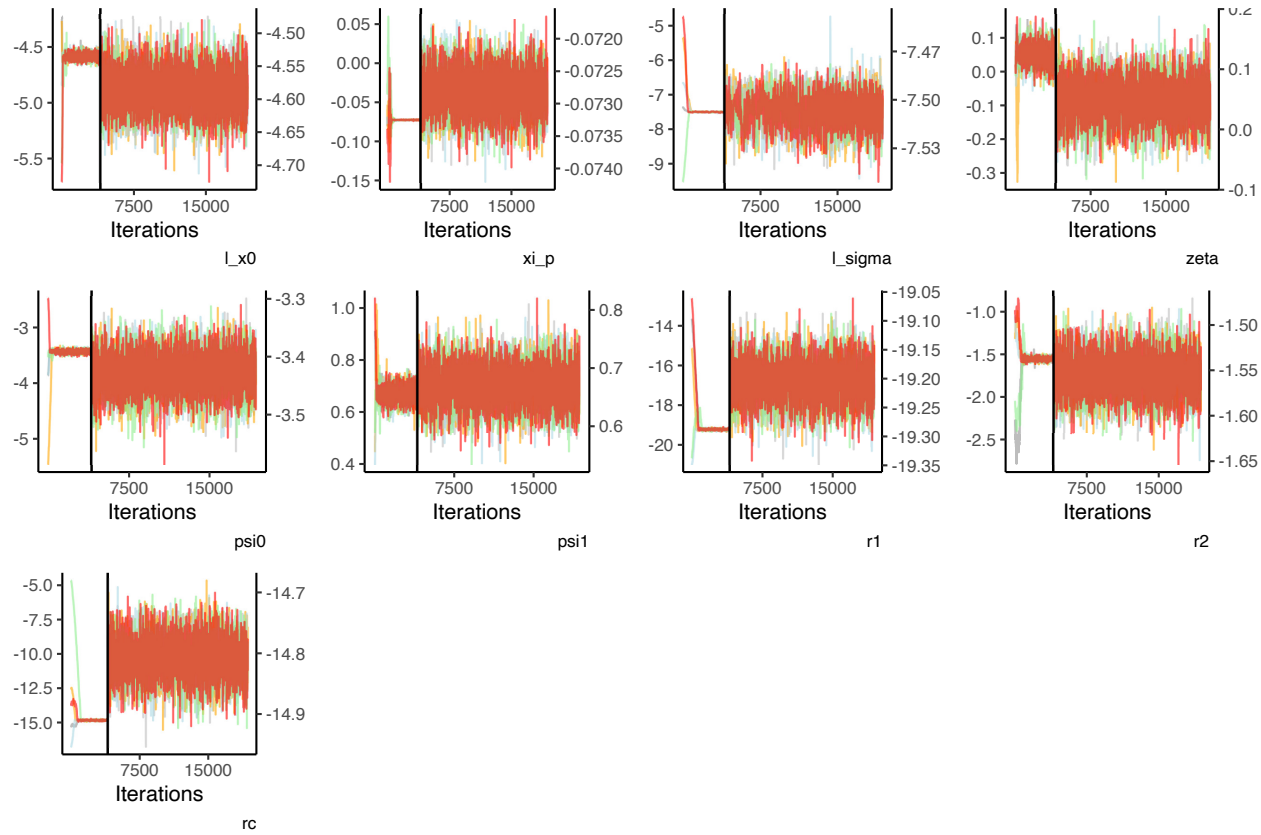


Figure S5: Trace plots for the parameter Markov chains generated by the PMH algorithm for the BS model.

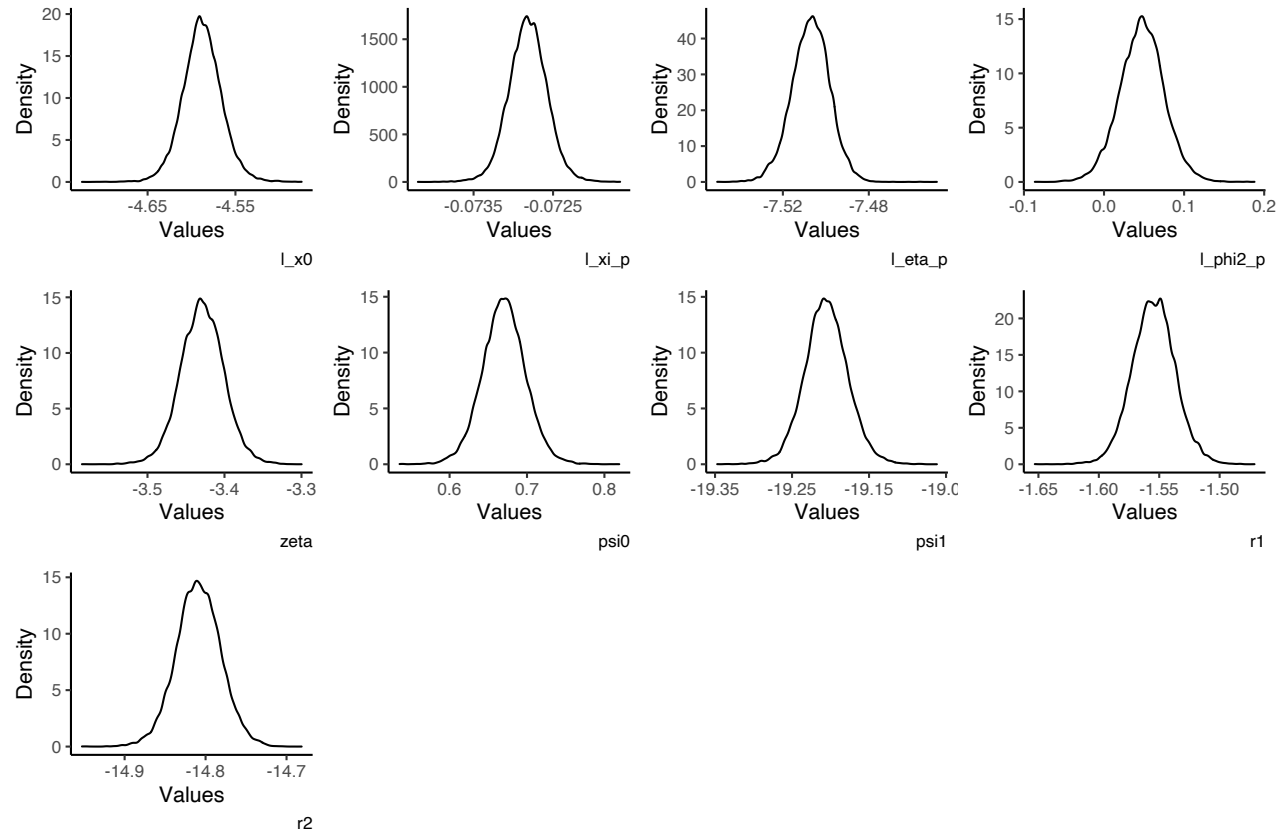


Figure S6: Density plots for the parameter Markov chains generated by the PMH algorithm for the BS model.

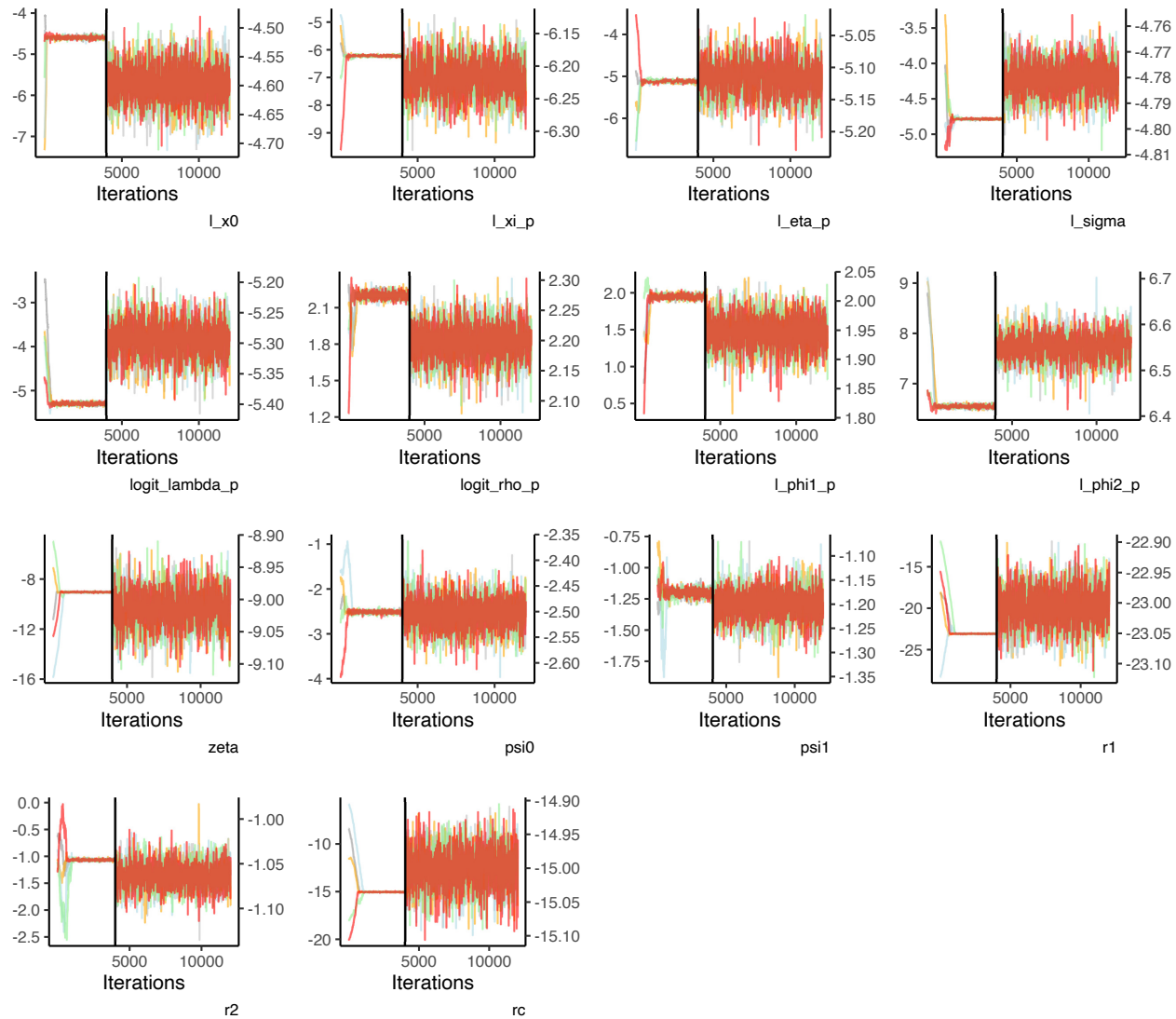


Figure S7: Trace plots for the parameter Markov chains generated by the PMH algorithm for the CIRj model.

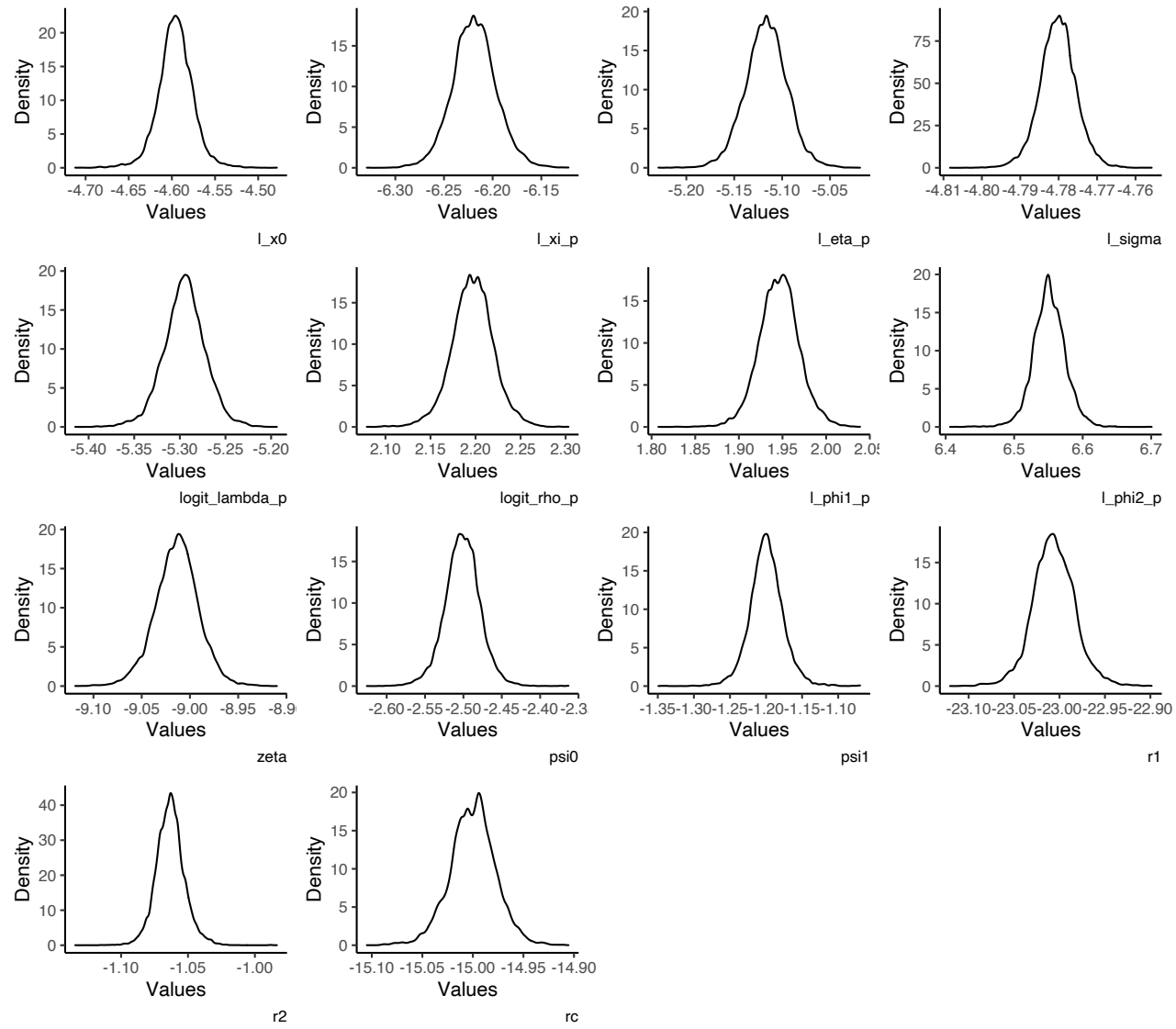


Figure S8: Density plots for the parameter Markov chains generated by the PMH algorithm for the CIRj model.

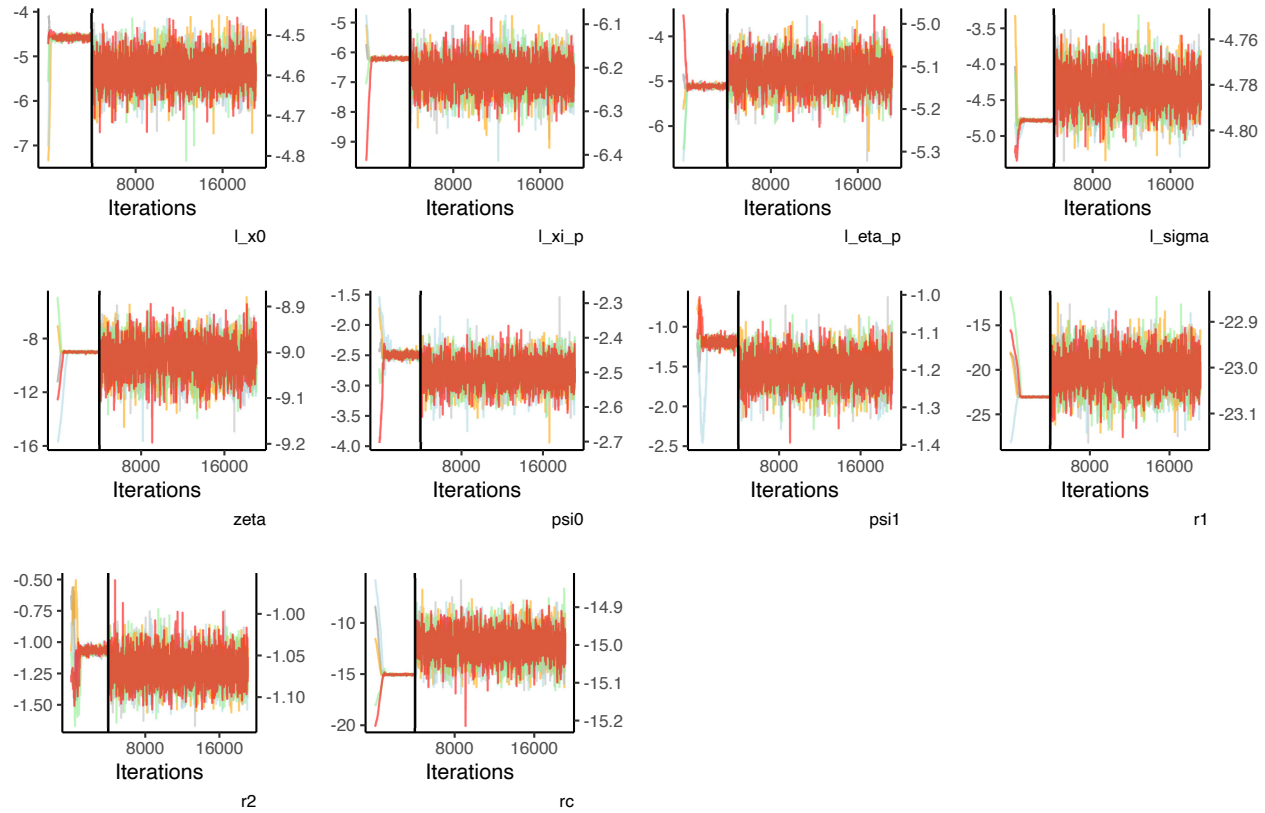


Figure S9: Trace plots for the parameter Markov chains generated by the PMH algorithm for the CIR model.

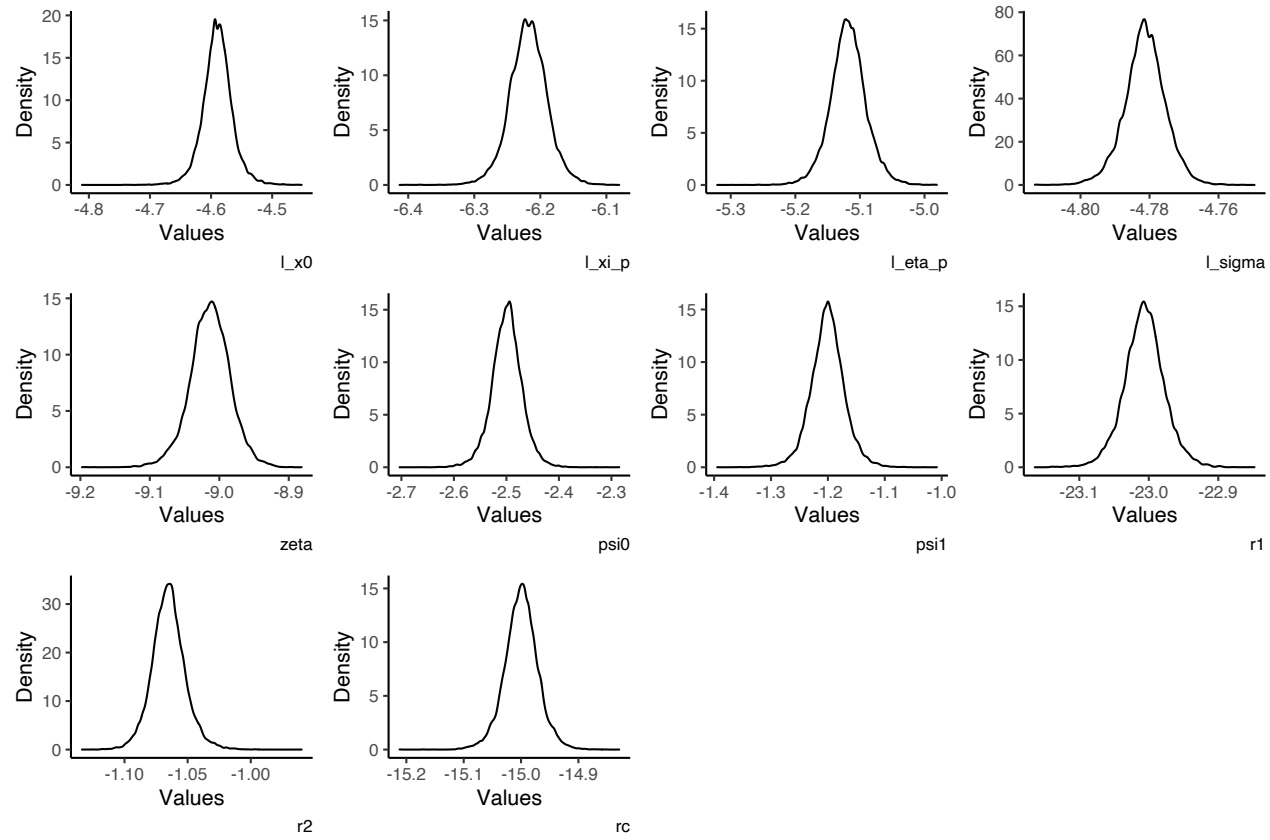


Figure S10: Density plots for the parameter Markov chains generated by the PMH algorithm for the CIR model.

S2.3 Sensitivity Analysis

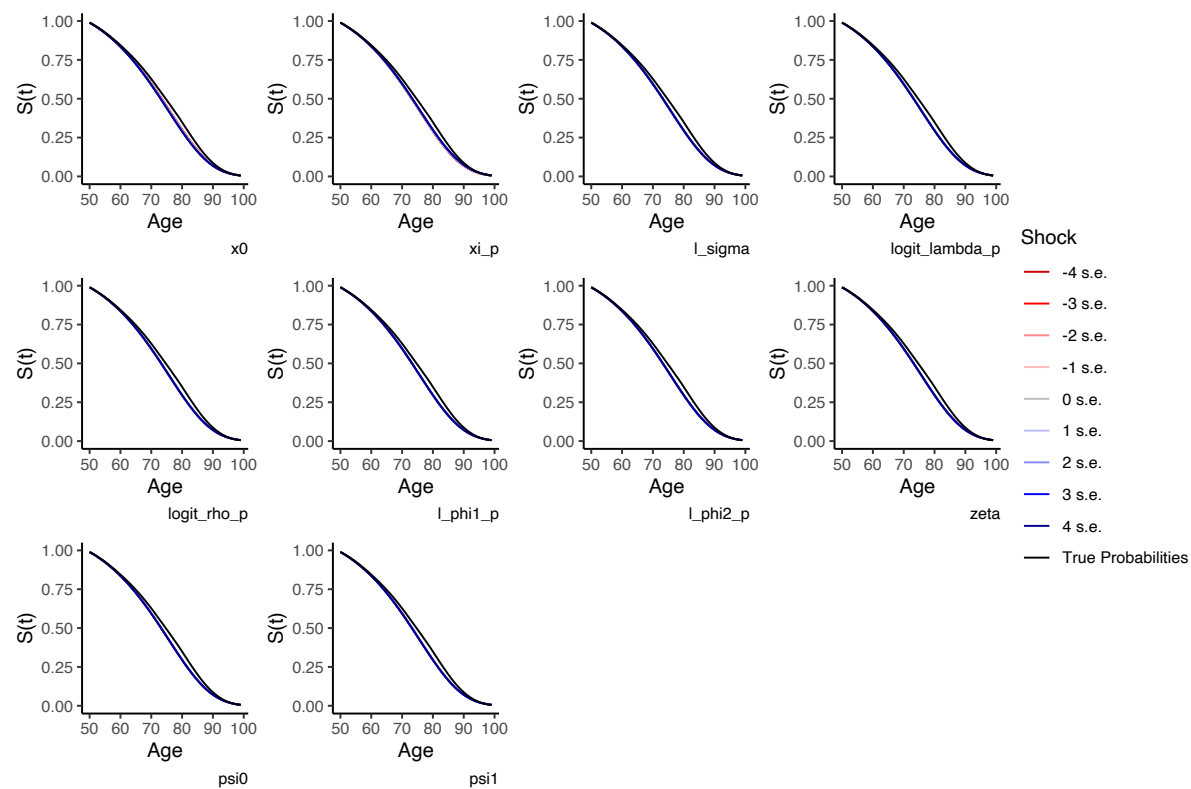


Figure S11: Sensitivity of 1916 cohort survival curve forecast with respect to model parameters under the BSj model. Note: x_0 above represents the initial value used in the simulation of X_{1916} , which in this case is X_{1915} .

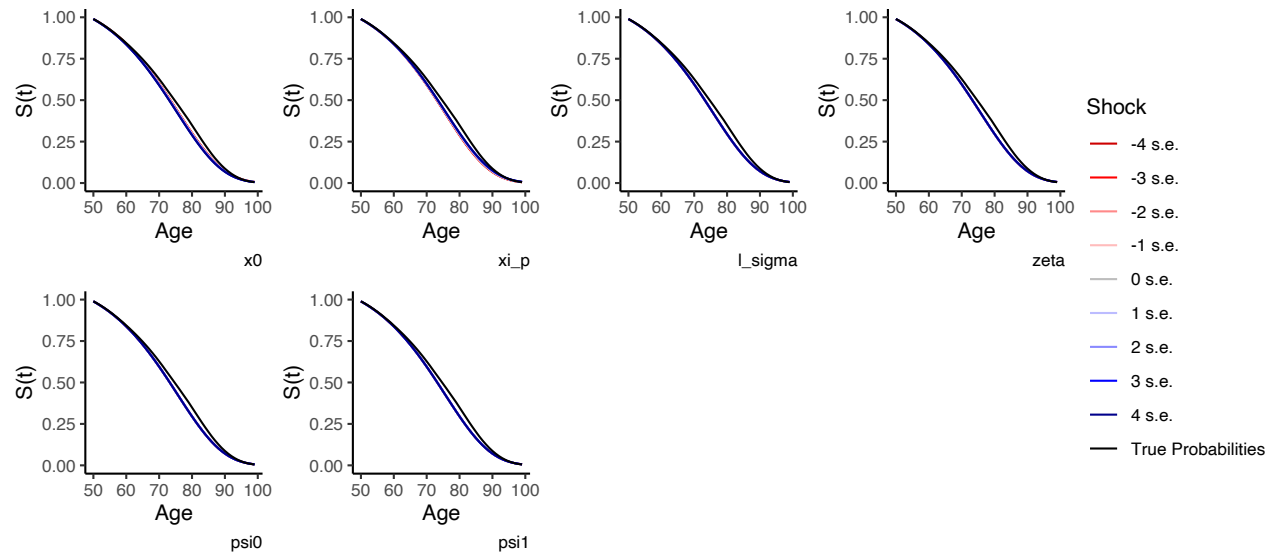


Figure S12: Sensitivity of 1916 cohort survival curve forecast with respect to model parameters under the BS model. Note: x_0 above represents the initial value used in the simulation of X_{1916} , which in this case is X_{1915} .

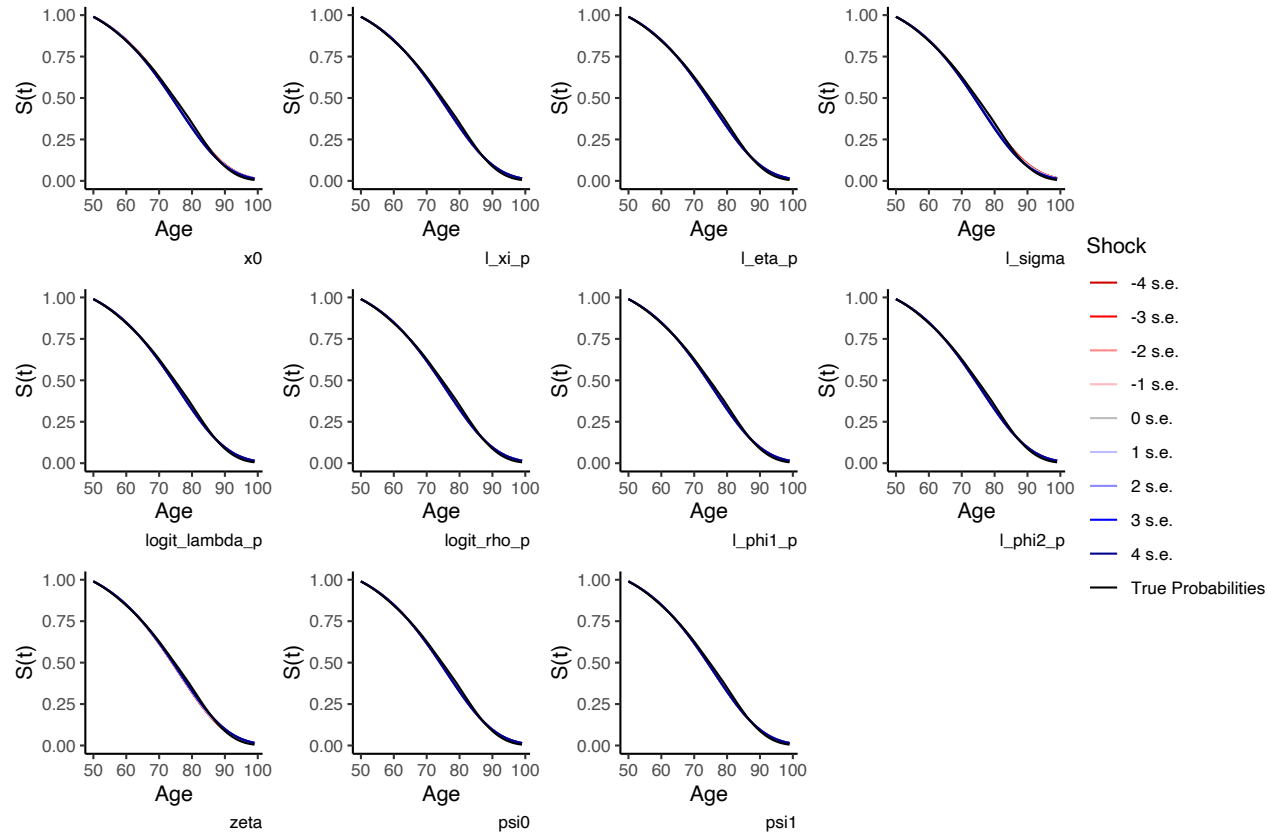


Figure S13: Sensitivity of 1916 cohort survival curve forecast with respect to model parameters under the CIRj model. Note: x_0 above represents the initial value used in the simulation of X_{1916} , which in this case is X_{1915} .

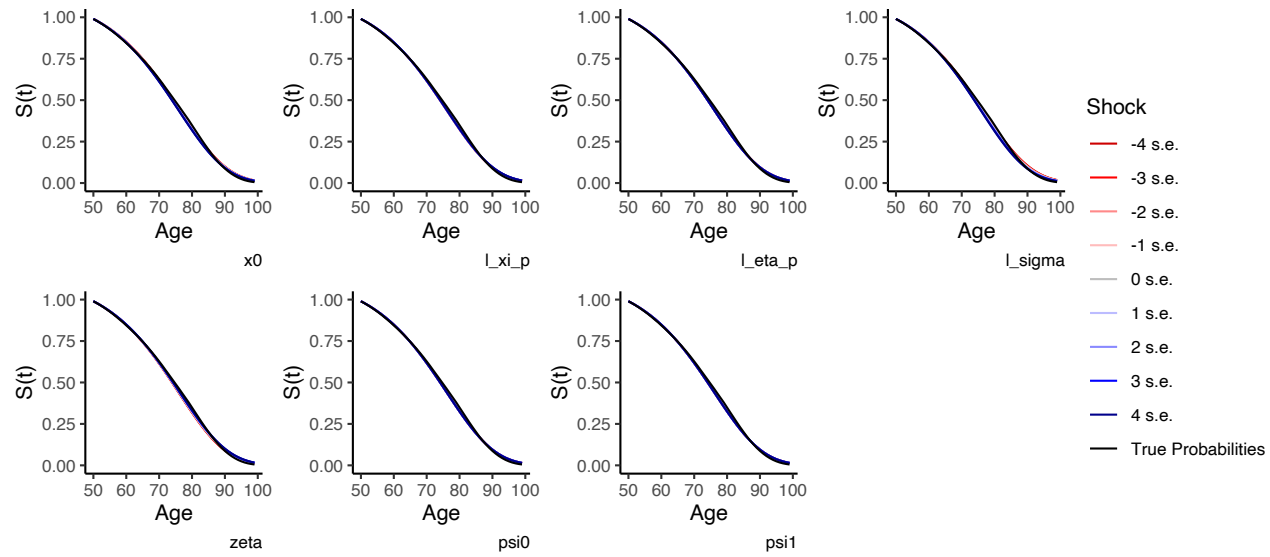


Figure S14: Sensitivity of 1916 cohort survival curve forecast with respect to model parameters under the CIR model. Note: x_0 above represents the initial value used in the simulation of X_{1916} , which in this case is X_{1915} .