A BAYESIAN JOINT MODEL FOR POPULATION AND PORTFOLIO-SPECIFIC MORTALITY

BY

FRANK VAN BERKUM, KATRIEN ANTONIO AND MICHEL VELLEKOOP

ABSTRACT

Insurance companies and pension funds must value liabilities using mortality rates that are appropriate for their portfolio. These can only be estimated in a reliable way from a sufficiently large historical dataset for such portfolios, which is often not available. We overcome this problem by introducing a model to estimate portfolio-specific mortality simultaneously with population mortality. By using a Bayesian framework, we automatically generate the appropriate weighting for the limited statistical information in a given portfolio and the more extensive information that is available for the whole population. This allows us to separate parameter uncertainty from uncertainty due to the randomness in individual deaths for a given realization of mortality rates. When we apply our method to a dataset of assured lives in England and Wales, we find that different prior specifications for the portfolio-specific factors lead to significantly different posterior distributions for hazard rates. However, in short-term predictive distributions for future numbers of deaths, individual mortality risk turns out to be more important than parameter uncertainty in the portfolio-specific factors, both for large and for small portfolios.

KEYWORDS

Bayesian analysis, portfolio-specific mortality, mortality projection, von Mises–Fisher prior, smoothing prior.

1. INTRODUCTION

Life insurance companies and pension funds need to value their liabilities using mortality rates appropriate for their portfolio. For many countries, projections of mortality rates are available for the entire population, but substantial heterogeneity in mortality rates exists between individuals within a population, which is caused among others by differences in socioeconomic classes, see Villegas and

Haberman (2014). Lantz *et al.* (1998) argue that individuals with a higher education tend to live more healthily, which may help to explain these differences in mortality.

Heterogeneity in mortality also exists between individuals since they may have different motivations to buy insurance. Finkelstein and Poterba (2002) show that differences in mortality even exist between individuals with voluntary annuities, compulsory annuities or without annuities. Pitacco *et al.* (2009) discuss the presence of select mortality when individuals are subject to medical tests when starting a life insurance policy. Policyholders with a longer duration since the test may experience higher mortality than policyholders that have been accepted more recently.

Therefore, an insurance company or pension fund cannot use mortality projections for the whole population without making any adjustments. The difference between mortality in a population and a portfolio is often called basis risk, see, for example, Barrieu *et al.* (2012).

In current practice, portfolio-specific mortality rates are often constructed by multiplying projections of country-wide mortality rates with portfolio-specific factors. These portfolio-specific factors, also called experience factors, thus represent the relative difference between the mortality rates of the population and the portfolio under consideration. In Solvency II, insurance companies are obliged to derive portfolio-specific mortality rate projections and analyze the uncertainty in these projections.

We propose a model to estimate population and portfolio-specific mortality simultaneously. To account for yearly fluctuations in small portfolios, we use a Poisson distribution to model individual deaths for a given realization of hazard rates, as in Brouhns *et al.* (2002). We view the portfolio as part of the population and use a baseline mortality trend for the population. The larger dataset for the population allows us to generate reliable estimates for the dynamics of mortality in the wider population. The relative difference between the population and the portfolio is modeled using a portfolio-specific and age-dependent random effect. Such random effects reflect the remaining heterogeneity among policyholders which is not captured by the observable risk factors. See Denuit *et al.* (2007) and Antonio and Zhang (2014) for similar examples in pricing models for non-life insurance, where policy(holder)-specific behavior is captured by such a random effect.

We use the Lee–Carter model for population mortality. In our Bayesian setting, we consider two prior distributions for the portfolio-specific factors. The first prior distribution assumes independent factors for different ages and independence between groups (portfolio and rest). The second prior distribution is an autoregressive smoothing prior which implies dependence between ages but independence between the factor for our own portfolio and the factor for the rest of the population. We describe population mortality and portfolio-specific mortality simultaneously, in contrast to the multi-step method that is required in a frequentist approach. This helps to distinguish volatility in the time series for the population, parameter uncertainty in the
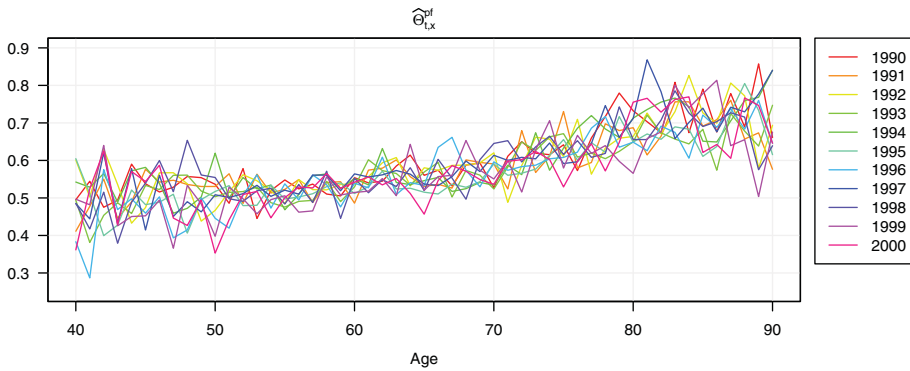
$$\widehat{\Theta}_{t,x}^{pf}$$

FIGURE 1: Observed portfolio-specific factors (the ratio of death rates in a portfolio and death rates in the whole country) for the CMI portfolio of assured male lives in England and Wales. (Color online)

model for the population, and parameter uncertainty in the portfolio-specific factors.

To illustrate this point, Figure 1 shows observed portfolio-specific factors for the CMI dataset on assured lives in England and Wales.[1] These factors are the ratio of death rates in the CMI portfolio and death rates in the whole of England and Wales, for different years and ages. The observations are very volatile when considered as a function of age and they can fluctuate wildly over consecutive years. These fluctuations are mainly due to the randomness in the number of individual deaths for a given fixed mortality rate. In order to take this into account, we will explicitly model the noise in the outcomes that we can actually observe (the number of deaths), by specifying that these follow a Poisson distribution when conditioned on the unobserved hazard rates that contain the unknown portfolio-specific factors that we are ultimately interested in. In a case study based on this dataset for England and Wales, we will show that parameter uncertainty in portfolio-specific factors can be substantial but that its impact on mortality projections is relatively small compared to the impact of the Poisson noise in individual deaths.

In Section 2, we give an overview of existing approaches to model portfolio-specific mortality. In Section 3, we introduce our own method and describe the prior distributions that are used in our Bayesian setting. Section 4 contains the illustration of our approach using the dataset on assured male lives from England and Wales, and Section 5 concludes.

2. LITERATURE OVERVIEW

General population mortality. Let $d_{t,x}$ be the observed number of deaths in a population at age x in calendar year t, and let $E_{t,x}$ be the corresponding exposure. The observed death rate is defined as $m_{t,x} = \frac{d_{t,x}}{E_{t,x}}$. Under the assumption of a constant force of mortality or hazard rate $\mu_{t,x}$ on the interval

$[t, t+1) \times [x, x+1)$, and in the absence of any further model structure, the maximum likelihood estimate $\hat{\mu}_{t,x}$ of the force of mortality $\mu_{t,x}$ equals the observed death rate $m_{t,x}$. The mortality rate $q_{t,x}$ is the probability that a person aged exactly x at the beginning of calendar year t dies within the next year, and under the assumption of a constant force of mortality it equals $q_{t,x} = 1 - \exp[-\mu_{t,x}]$.

Lee and Carter (1992) introduce the seminal mortality model to explain observed death rates:

$$\ln m_{t,x} = \alpha_x + \beta_x \kappa_t + \eta_{t,x}, \qquad (1)$$

with the $\eta_{t,x}$ iid stochastic variables with mean zero. Lee and Carter apply least squares estimation to this model, which is equivalent to maximum likelihood estimation under the assumption of normality and homoscedasticity for the $\eta_{t,x}$'s. Since this model is a single-factor model, mortality improvements for all ages are assumed to be perfectly correlated. First, they estimate the parameters α_x, β_x and κ_t using a Singular Value Decomposition, and then they model the period effect κ_t as a random walk with drift to generate mortality projections:

$$\kappa_t = \kappa_{t-1} + \delta + \varepsilon_t, \qquad \varepsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_\varepsilon^2), \qquad (2)$$

for independent $\eta_{t,x}$ and ε_t. Brouhns et al. (2002) model the force of mortality instead of the death rate $m_{t,x}$ and specify that the observed $d_{t,x}$ are realizations of stochastic variables $D_{t,x}$ with the following structure:

$$D_{t,x}|\mu_{t,x} \sim \mathrm{Poisson}(E_{t,x}\mu_{t,x}), \qquad \text{with } \ln \mu_{t,x} = \alpha_x + \beta_x \kappa_t. \qquad (3)$$

Whereas the error terms in (1) have the same variance for all ages and year, this assumption is relaxed through the specification in (3). Pitacco et al. (2009) show that a likelihood based on yearly individual survival observations is proportional to the likelihood of a Poisson distribution as specified above. Therefore, the model specification using a Poisson distribution is appropriate for small and large portfolios.

For an overview of extensions to the Lee–Carter model in a single population setting, we refer to Cairns et al. (2009), Haberman and Renshaw (2011) and van Berkum et al. (2016).

Multiple population mortality models. Mortality developments in a population can be strongly time varying. Periods of small mortality improvements may be followed by periods of larger ones, and a rapidly changing mortality trend is difficult to project. Therefore, extensions to the Lee–Carter model have been proposed to incorporate information from different but comparable populations in the estimation process. This can lead to a more stable, global mortality trend, which also provides insight in population-specific deviations from the general pattern. A disadvantage is that a sufficiently large historical data is needed to analyze such population-specific deviations. If there is only limited historical data available for a portfolio, application of the multiple population

approach to portfolio data must contain a careful analysis of the uncertainty in the estimates.

Li and Lee (2005) propose the augmented common factor model for multiple populations (indexed by i)

$$\ln m_{t,x}^i = \alpha_x^i + B_x K_t + \beta_x^i \kappa_t^i + \epsilon_{t,x,i}, \qquad \epsilon_{t,x,i} \overset{\text{iid}}{\sim} \text{N}(0, \sigma_i^2). \qquad (4)$$

The term $B_x K_t$ represents the common factor for the different populations, α_x^i is the average mortality for age x in population i over time, and the term $\beta_x^i \kappa_t^i$ is a population-specific, age-dependent mortality development. Li and Lee (2005) estimate this model using Singular Value Decomposition, whereas Antonio *et al.* (2015) use a Bayesian framework. For a related alternative where different groups have a common age effect for mortality improvements, see Kleinow (2015).

Dowd *et al.* (2011) investigate mortality in two populations using a gravity model

$$\ln m_{t,x}^i = \alpha_x^i + \kappa_t^i + \gamma_{t-x}^i, \qquad i = 1, 2, \qquad (5)$$

with γ_{t-x}^i a term representing a cohort effect. For forecasting purposes, the cohort effect must be projected in a similar way as the variables κ_t^i, but for γ_{t-x}^i this often proves to be more difficult because of its volatile behavior, see Haberman and Renshaw (2011). The first population is defined as the dominant population and the second population is of smaller size and is therefore considered to be the subordinate population. In a related model of Dowd *et al.* (2011), the time series of the subordinate population depends on the difference in mortality between the two populations, which is called the "spread". Cairns *et al.* (2011) estimate parameters for this model using a Bayesian approach. By defining the dependence between the two populations slightly differently, they arrive at a specification that can be used for a combination of a dominant and a subordinate population, but also for a combination of two equal-sized populations. This makes it suitable to model mortality in different countries but also for mortality in a country and in a large pension fund.

Villegas and Haberman (2014) consider mortality of five different socioeconomic classes in England. Mortality for the reference population is described using an extension of the Lee–Carter model, and mortality for different socioeconomic classes is modeled relative to the population. Haberman *et al.* (2014) use a similar approach to model mortality in an insurance book, but a wider collection of mortality models is considered. For books with large exposures and sufficient historical observations, a variant of the model introduced in Cairns *et al.* (2006) is suggested.

Portfolio-specific mortality models. Apart from the multiple population approach, other methods have been suggested to characterize portfolio-specific mortality. Often, population mortality is assumed given, or a (smooth) baseline mortality rate is estimated beforehand. Even when these models are able to

explain historical observations well, they may be less appropriate for projection purposes when population and portfolio-specific mortality are not estimated simultaneously. Below we discuss several approaches to modeling portfolio-specific factors. In Section 3.1, we combine these ideas with the multiple population approach and introduce a new method to simultaneously estimate population and portfolio-specific mortality.

Plat (2009) considers realized portfolio-specific factors defined by

$$P_{t,x} = \frac{m_{t,x}^A}{m_{t,x}^{\mathrm{pop}}}, \tag{6}$$

where $m_{t,x}^A$ is the observed death rate in the portfolio based on insured amounts, and $m_{t,x}^{\mathrm{pop}}$ is the observed death rate in the population. As an example, a linear age effect is assumed, using

$$P_{t,x} = a_t + b_t x + \varepsilon_{t,x}, \qquad \varepsilon_{t,x} \overset{\mathrm{iid}}{\sim} \mathrm{N}(0, \sigma_\varepsilon^2). \tag{7}$$

The parameters a_t and b_t are estimated using regression techniques, and portfolio-specific factors for future years are obtained by projecting a_t and b_t using time series models.

Using only 5 years of historical data, Gschlössl et al. (2011) do not include time dynamics in their model for portfolio-specific mortality. First, they estimate a baseline force of mortality on portfolio data which is a smooth function of age. Remaining heterogeneity is then captured by observable risk factors in a Poisson GLM framework. Richards et al. (2013) model the force of mortality using a time-varying version of the Makeham–Beard law and estimate the parameters on 5 years of historical portfolio data for individual lives. Their approach can, therefore, not be used when only aggregated portfolio data are available.

Olivieri (2011) considers a Bayesian setting of the form

$$D_{t,x} \sim \mathrm{Poisson}(E_{t,x} q_{t,x}^* Z_{t,x}), \tag{8}$$

where $q_{t,x}^*$ is a best estimate mortality rate published by an independent institution, and $Z_{t,x} \sim \mathrm{Gamma}(\alpha_{t,x}, \beta_{t,x})$ is a random adjustment. Starting with values for $\alpha_{0,x}$ and $\beta_{0,x}$, subsequent values of $\alpha_{t,x}$ and $\beta_{t,x}$ can be computed in closed form when new mortality observations become available, since the Gamma distribution is the conjugate of the Poisson distribution. Kan (2012) considers a similar framework, but a different method is used to estimate the population mortality rate.

3. BAYESIAN PORTFOLIO-SPECIFIC MORTALITY

As Section 2 illustrates, different approaches to modeling portfolio-specific factors exist, which are suitable for different types of datasets. We consider the
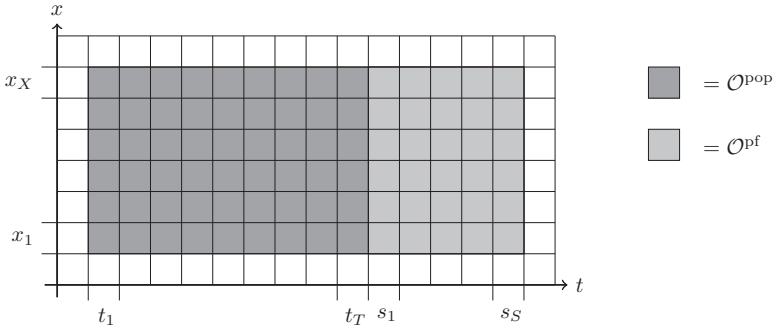
FIGURE 2: Illustration of $\mathcal{O}^{\mathrm{pf}}$ and $\mathcal{O}^{\mathrm{pop}}$.

situation where only limited historical portfolio data is available, which hinders reliable estimation of mortality developments if only portfolio data would be used. We, therefore, simultaneously estimate mortality in the population and the portfolio-specific factors.

Let the observed number of deaths for group i during calendar year t for ages in $[x, x+1)$ be $d_{t,x}^{i}$, and denote the exposure in this group for that period by $E_{t,x}^{i}$. The groups we consider are the entire population of a country (pop), the portfolio under investigation (pf), and the part of the population which is not included in the portfolio under consideration (hereafter referred to as the "rest"), so $i \in \{\mathrm{pop}, \mathrm{pf}, \mathrm{rest}\}$. The observed portfolio and the rest thus form the total population and we have that $d_{t,x}^{\mathrm{pf}} + d_{t,x}^{\mathrm{rest}} = d_{t,x}^{\mathrm{pop}}$ and $E_{t,x}^{\mathrm{pf}} + E_{t,x}^{\mathrm{rest}} = E_{t,x}^{\mathrm{pop}}$. We need to define the rest group explicitly, to ensure that we always consider all information available in the population.

To estimate parameters, we extend the portfolio dataset with observations of the total population. In the dataset of the portfolio, we consider X ages and S years, and in the population X ages and T years. We define the set of cells (t, x) for which we have observations from both our portfolio and the rest (the light gray cells in Figure 2) as $\mathcal{O}^{\mathrm{pf}} = \mathcal{S} \times \mathcal{X}$ with $\mathcal{S} = \{s_1, s_1 + 1, \ldots, s_S\}$ and $\mathcal{X} = \{x_1, x_1 + 1 \ldots, x_X\}$. We can only measure the heterogeneity between the portfolio and the rest on *this* set of observations. The set for which we have observations from the population but not separately for our portfolio and the rest (the dark gray cells in Figure 2) is defined by $\mathcal{O}^{\mathrm{pop}} = \mathcal{T} \times \mathcal{X}$, with $\mathcal{T} = \{t_1, t_1 + 1 \ldots, t_T\}$ and $t_1 \leq t_T < t_T + 1 = s_1 \leq s_S$, and by construction $\mathcal{O}^{\mathrm{pop}} \cap \mathcal{O}^{\mathrm{pf}} = \emptyset$.[2] In total, $T^* = T + S$ years are included in the dataset.

We introduce indicator variables that will turn out to be useful when working with likelihoods:

$$I_{t,x}^{\mathrm{pf}} = I_{t,x}^{\mathrm{rest}} = \begin{cases} 1 & \text{if } (t, x) \in \mathcal{O}^{\mathrm{pf}} \\ 0 & \text{otherwise,} \end{cases} \qquad I_{t,x}^{\mathrm{pop}} = \begin{cases} 1 & \text{if } (t, x) \in \mathcal{O}^{\mathrm{pop}} \\ 0 & \text{otherwise.} \end{cases} \qquad (9)$$

3.1. Model formulation and implementation

We assume that people in our own portfolio and the rest of the population share a baseline force of mortality which is denoted by $\mu_{t,x}$. Heterogeneity between groups is captured by a random effect Θ_x^i which depends on age. This leads to the following specification:

$$D_{t,x}^{\text{pop}}|\mu_{t,x} \sim \text{Poisson}(E_{t,x}^{\text{pop}}\mu_{t,x}), \qquad \text{for } (t, x) \in \mathcal{O}^{\text{pop}}, \qquad (10)$$

and

$$D_{t,x}^{\text{pf}}|(\mu_{t,x}, \Theta_x^{\text{pf}}) \sim \text{Poisson}(E_{t,x}^{\text{pf}}\mu_{t,x}\Theta_x^{\text{pf}}), \qquad \text{for } (t, x) \in \mathcal{O}^{\text{pf}} \qquad (11)$$

$$D_{t,x}^{\text{rest}}|(\mu_{t,x}, \Theta_x^{\text{rest}}) \sim \text{Poisson}(E_{t,x}^{\text{rest}}\mu_{t,x}\Theta_x^{\text{rest}}), \qquad (12)$$

with the Lee–Carter model for the baseline

$$\ln \mu_{t,x} = \alpha_x + \beta_x \kappa_t. \qquad (13)$$

This implies that we consider all deaths in the population for every cell (t, x), by either using $D_{t,x}^{\text{pop}}$ or both $D_{t,x}^{\text{pf}}$ and $D_{t,x}^{\text{rest}}$.[3]

The random effects Θ_x^i are independent between groups i, but there may be dependence for different ages x.[4] In Section 3.2, we will consider two prior specifications for Θ_x^i, a Gamma prior and a lognormal prior. In the first one, we assume independence between ages x and between groups i, but in the second one, we assume dependence between ages and independence between groups. Given the baseline force of mortality $\mu_{t,x}$ and the portfolio-specific factors Θ_x^i, the Poisson distributed numbers of deaths are independent between ages, calendar years and groups.

To project mortality into the future, we need to impose a time series model on the period effect κ_t. Two time series specifications that are often used for projecting the period effect in the Lee–Carter model are a trend stationary and a difference stationary model (also known as a random walk with possibly a drift). As discussed in van Berkum *et al.* (2016), we believe a difference stationary model to be more appropriate to model the period effect for a single country so that is what we will use in this paper.

In order to generate samples of posterior distributions, we use the Markov chain Monte Carlo method (MCMC) with a burn-in period which allows the chain to move toward the desired distribution before we start taking samples. Since the MCMC algorithm requires Metropolis(-Hastings) sampling for our model, the burn-in period is also used to calibrate scale parameters for the distribution to propose new samples. We calibrate the scale parameters such that the acceptance probabilities are within the interval [20%, 30%]. Only samples that are found after the burn-in period are used for inference on parameters and for prediction purposes.

In a frequentist setting, parameter constraints are needed to uniquely identify the Lee–Carter model, since linear transformations can be applied which change the value of the parameters α_x, β_x and κ_t without changing the forces of mortality. In a Bayesian framework, parameters are random variables so there is no identifiability problem due to the introduction of priors. However, the presence of identification problems in a frequentist setting suggests that we may also encounter convergence problems for the MCMC algorithm in a Bayesian implementation of the same model. We, therefore, apply two parameter constraints:

$$\kappa_{t_1} = 0 \qquad \text{and} \qquad \|\boldsymbol{\beta}\|^2 = \sum_{x \in \mathcal{X}} \beta_x^2 = 1, \tag{14}$$

through the specification of the prior distributions. Further, for $t \geq s_1$, we have more information: two observations per cell (t, x). As a result, in a frequentist setting, the forces of mortality for the portfolio and the rest group are invariant to the following transformation:

$$\Theta_x^{\text{pf}} \rightarrow \Theta_x^{\text{pf}} \cdot \exp(c\beta_x), \quad \Theta_x^{\text{rest}} \rightarrow \Theta_x^{\text{rest}} \cdot \exp(c\beta_x) \quad \text{and} \quad \kappa_t \rightarrow \kappa_t - c I_{[t \in \mathcal{S}]}.$$

We observed that we may encounter convergence problems in κ_t and Θ_x^i if no additional constraint is imposed, and therefore we impose

$$\kappa_{s_1} = \kappa_{t_T}. \tag{15}$$

With this constraint, we further ensure that the parameters α_x and β_x can be used for both $\mathcal{O}^{\text{pop}}$ and $\mathcal{O}^{\text{pf}}$. The parameter constraint in (15) is appropriately taken into account in the prior specification and in the derivation of the posterior distributions, and this has no impact on the random walk specification.

3.2. Prior distributions

We will now describe the prior distributions for parameters and hyperparameters, to complete the Bayesian specification of the model.

3.2.1. *Population mortality parameters.*

Prior distribution for α_x. Following Czado *et al.* (2005) and Antonio *et al.* (2015), we use the following prior for α_x with $x = x_1, \ldots, x_X$:

$$e_x = \exp(\alpha_x) \overset{\text{iid}}{\sim} \text{Gamma}(a_x, b_x). \tag{16}$$

Prior distribution for β_x. For the vector of parameters $\boldsymbol{\beta} = \{\beta_{x_1}, \ldots, \beta_{x_X}\}$, we choose a prior distribution that automatically satisfies the constraint in (14): the Von Mises–Fisher distribution, which has its origins in directional statistics (von Mises, 1918; Fisher, 1953). To the best of our knowledge, Antoniadis *et al.*

(2004) were the first to use this distribution as a prior in Bayesian analysis. The prior distribution for β is denoted by

$$\beta \sim \text{vMF}(\mu_\beta, c_\beta), \tag{17}$$

for constants μ_β (the mean direction vector) and c_β (the concentration parameter) with $\|\mu_\beta\| = 1$ and $c_\beta > 0$. The probability density function is given by

$$f_X(\beta | \mu_\beta, c_\beta) = C_X(c_\beta) \exp\left(c_\beta \mu_\beta^T \beta\right), \tag{18}$$

where the normalization constant $C_X(c)$ equals

$$C_X(c) = \frac{c^{X/2-1}}{(2\pi)^{X/2} I_{X/2-1}(c)}, \tag{19}$$

with I_v the modified Bessel function of the first kind of order v. See Hoff (2009) for details on how to sample from this distribution.

Note that our approach differs from what is usually done in the actuarial literature (see, for example, Czado *et al.* (2005), Li (2014), Antonio *et al.* (2015)), in the sense that often transformations are applied in a Metropolis–Hastings step after a sample has been accepted. In our approach, every proposed sample already satisfies the necessary constraints because of our choice of the priors.

Prior specification for κ_t.   In line with van Berkum *et al.* (2016), we assume a random walk with drift for the period effect κ_t. The prior distribution is specified by

$$\delta \sim \text{N}(\mu_\delta, \sigma_\delta^2), \tag{20}$$

$$\sigma_\varepsilon \sim \text{Uniform}(0, A_\varepsilon), \tag{21}$$

$$\kappa_t = \kappa_{t-1} + \delta + \varepsilon_t, \quad \text{with} \quad \varepsilon_t \overset{\text{iid}}{\sim} \text{N}(0, \sigma_\varepsilon^2) \quad \text{for} \quad t > t_1 \quad \text{and} \quad t \neq s_1$$
$$\text{and} \quad \kappa_{t_1} = 0, \quad \kappa_{s_1} = \kappa_{t_T}. \tag{22}$$

For variance hyperparameters, Gelman (2006) suggests using a Uniform(0, A) prior on σ instead of an Inverse-Gamma(ϵ, ϵ) prior on σ^2, because if the estimate of σ is close to zero, the posterior density will be sensitive to the choice of ϵ. Therefore, we use a uniform prior on σ.

3.2.2. *Portfolio-specific factors.*   The portfolio-specific factor Θ_x^i represents the ratio of the hazard rate for group i at age x and the hazard rate for the whole population at age x, where $i \in \{\text{pf, rest}\}$. We do not want to make a priori assumptions on whether mortality in a group is higher or lower than the baseline mortality, and therefore we impose $\mathbb{E}(\Theta_x^i) = 1$ ($\forall x, \forall i$). We consider two prior distributions for Θ_x^i, an independent (Gamma) prior and a lognormal prior.

Gamma prior. The Gamma prior on the age-dependent factors for group i is given by

$$\Theta^i_x \sim \text{Gamma}(c^i_x, c^i_x), \quad \text{for } x_1 \leq x \leq x_X. \tag{23}$$

The factors are independent over ages x and between groups i. By choosing equal values for the two parameters in the Gamma distribution, we ensure that $\mathbb{E}(\Theta^i_x) = 1$ for all x and i, and the variance of the prior can be controlled by the choice of c^i_x.

Lognormal prior. In this specification, we assume a mean reverting process (AR(1)) for the logarithm of the age-dependent factors. This ensures that the factors are non-negative. The lognormal prior on the age-dependent factors for group i is given by

$$\text{logit}\,(\rho_i) \sim \text{N}(\mu_{\rho_i}, \sigma^2_{\rho_i}), \tag{24}$$

$$\sigma_i \sim \text{Uniform}(0, A_i), \tag{25}$$

$$\ln \Theta^i_x = \mu_i + \rho_i \ln \Theta^i_{x-1} + \eta^i_x, \quad \text{with } \eta^i_x \stackrel{\text{iid}}{\sim} \text{N}(0, \sigma^2_i(1 - \rho^2_i)) \text{ for } x_1 < x \leq x_X,$$
$$\text{and } \ln \Theta^i_{x_1} \stackrel{\text{iid}}{\sim} \text{N}(-\tfrac{1}{2}\sigma^2_i, \sigma^2_i), \tag{26}$$

and all the $\ln \Theta^i_{x_1}$ and η^i_x are independent. Note that this implies that there may be dependence between group-specific mortality factors for different ages x, while factors for different groups i are independent. Due to the autoregressive structure in (26), this prior is also often referred to as an autoregressive smoothing prior.

The prior for ρ_i is chosen such that it is restricted to the interval $(-1, 1)$, and the prior for σ_i is chosen in line with the other variance prior specifications. The mean parameter is chosen to be $\mu_i = -\frac{1}{2}(1 - \rho_i)\sigma^2_i$ and the prior for $\Theta^i_{x_1}$ is chosen such that $\mathbb{E}(\Theta^i_x) = 1$ for all x and i, see Denuit *et al.* (2005).

In the next section, we specify the constants that are needed to finalize the specification of the prior distributions for the parameters and hyperparameters. With the definition of the prior distributions, our model is completely specified, and posterior distributions can thus be calculated. They can be found in Appendix A.

4. EMPIRICAL STUDY

In this section, we apply our model to data from the Continuous Mortality Investigation (CMI), which contains mortality statistics of assured male lives in England and Wales. We use the years $s \in \mathcal{S} = \{s_1 = 1990, \ldots, s_S = 2000\}$ and the ages $x \in \mathcal{X} = \{x_1 = 40, \ldots, x_X = 90\}$. Dowd *et al.* (2011) also use the CMI
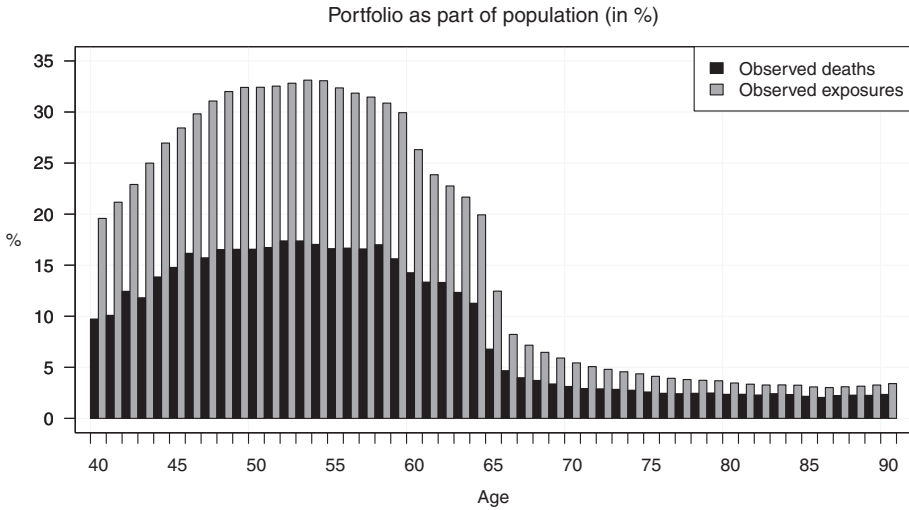
FIGURE 3: The relative size of the portfolio in terms of observed deaths and observed exposures. For each age, the relative size is computed as $\sum_t d_{t,x}^{\text{pf}} / \sum_t d_{t,x}^{\text{pop}}$ and $\sum_t e_{t,x}^{\text{pf}} / \sum_t e_{t,x}^{\text{pop}}$, where each summation is over $t \in \mathcal{S}$.

dataset in estimating a two-population mortality model, but they use the years 1961–2005 and the ages 60–84.

We extend the dataset with mortality data on the England and Wales population for the years $t \in \mathcal{T} = \{t_1 = 1950, \ldots, t_T = 1989\}$ and the same ages $x \in \mathcal{X}$ to ensure we obtain mortality forecasts consistent with population mortality forecasts.[5] We use population mortality data for $t \in \mathcal{S}$ to construct the rest group by subtracting portfolio deaths and exposures from the population deaths and exposures in those cells (t, x) for which portfolio data are available.

The size of the portfolio as a portion of the population, measured in observed deaths and observed exposures, is shown in Figure 3. In total, there were around 28.5 million years of exposure and 159,029 observed deaths. If mortality in the portfolio were similar to that in the population, we would expect the observed deaths and observed exposures to be of similar relative size. However, the observed deaths and observed exposures clearly differ, and we see that mortality in the portfolio is lower than in the population as whole.

We estimate four different models:[6]

1. The Lee–Carter model is used for population mortality for the England and Wales population for $t \in \{\mathcal{T}, \mathcal{S}\}$ and $x \in \mathcal{X}$, and parameters are estimated using maximum likelihood. This method is referred to as POP(f).
2. The Lee–Carter model is used for population mortality for the England and Wales population for $t \in \{\mathcal{T}, \mathcal{S}\}$ and $x \in \mathcal{X}$, and parameters are estimated in a Bayesian framework. This method is referred to as POP(B).

3. The model described in Section 3.1 is used, with a Gamma prior for Θ_x^i. Population and group-specific mortality are estimated simultaneously in a Bayesian framework. This method is referred to as PF(B-G).
4. The model described in Section 3.1 is used, with a logNormal prior for Θ_x^i. Population and group-specific mortality are estimated simultaneously in a Bayesian framework. This method is referred to as PF(B-logN).

4.1. CMI assured lives — original dataset

In this section, we consider the original CMI dataset, and we use the ages and years as described above. In the next section, we reduce the size of the CMI dataset to investigate the effect of portfolio size on the posterior distribution of the parameters.

Prior distributions. To complete the specifications of the prior distributions, we have to choose the constants used in these specifications. We do this in such a way that the priors contain little information about our prior beliefs, i.e. such that the prior variance is large.

We run four MCMC chains in parallel. For the population mortality parameters α_x, β_x and κ_t, we use frequentist estimates $\hat{\alpha}_x$, $\hat{\beta}_x$ and $\hat{\kappa}_t$ as starting points, but in each chain, we add some random Gaussian noise to obtain different starting values. Using the starting values for β_x and κ_t, we obtain maximum likelihood estimates in each chain for σ_β^2, δ and σ_ε^2, and we use these as initial values for the hyperparameters. For the portfolio-specific factors Θ_x^i, we take the initial draw of the MCMC simulations equal to 1. For the hyperparameters of Θ_x^i, we start with $\rho_i = 0.8$ and $\sigma_i^2 = 1$. The constants that complete the specification of the prior distributions and the sampling variances used in the Gibbs and Metropolis(–Hastings) sampling algorithms are chosen as follows:

- To ensure the prior does not contain much information, we use $a_x = b_x \cdot \exp(\hat{\alpha}_x)$ and $b_x = 0.01$, see Antonio *et al.* (2015). This way, $\mathbb{E}[\exp(\alpha_x)] = \exp(\hat{\alpha}_x)$ with large variance.
- For β, we use $\boldsymbol{\mu}_\beta = \frac{1}{\sqrt{X}} \cdot \mathbf{1}_X$ with $\mathbf{1}_X$ a vector with ones of length X, and $c_\beta = 0.01$.
- We use $\mu_\delta = \hat{\delta}$ (the Maximum Likelihood estimate of the drift, as obtained from the frequentist estimates of the κ_t) and $\sigma_\delta^2 = 0.5^2$. For the variance hyperparameter, we use $A_\varepsilon = 10$.
- For the Gamma prior on the portfolio-specific factors, we use $c_x^i = 1$ for all x and for $i \in \{pf, rest\}$. As a result, the prior 95% confidence interval for Θ_x^i is approximately $(0, 4)$.
- For the logNormal prior on the portfolio-specific factors, we use $\mu_{\rho_i} = 0$ and $\sigma_{\rho_i}^2 = 1$, and for the variance hyperparameter we use $A_i = 10$ for $i \in \{pf, rest\}$.
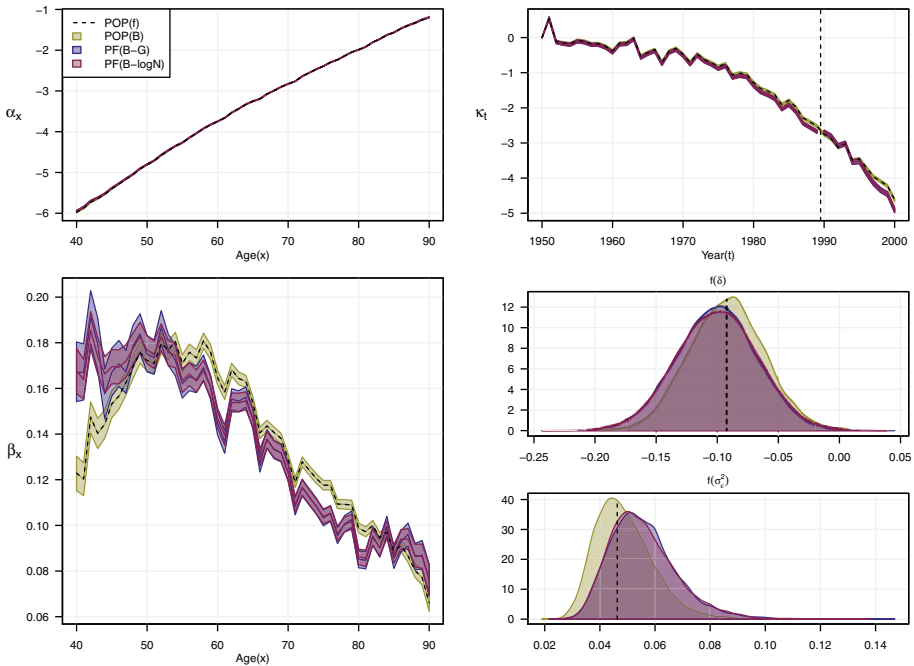
FIGURE 4: Parameter estimates for the CMI and England and Wales datasets using portfolio data for 1990–2000 and ages 40–90. For the frequentist method (POP(f)), we show the Maximum Likelihood estimates, and for the Bayesian methods (POP(B), PF(B–G) and PF(B–logN)) we show the 95% credible interval (equal tailed) of the posterior distributions. (Color online)

- For the scale parameters used in the proposal densities, we start with $d_\beta = 10^5$, $s_{\kappa_t}^2 = 0.05^2$, $s_{\Theta_x^i}^2 = 2^2$, $s_{\rho_i}^2 = 0.05^2$ and $s_{\sigma_i^2}^2 = 0.5^2$. For the definition of scale parameters, we refer to Appendix A.

Convergence diagnostics.   We run 1,100,000 iterations in each chain of the MCMC algorithm. We save every 500th iteration, and during the first 100,000 iterations, we calibrate the scale parameters of the proposal distributions every 100th iteration.[7] Our trace plots show good mixing properties, the calculated Gelman and Rubin statistics converge rapidly toward 1, and density plots of the parameters in different chains overlap almost perfectly.[8] Our final sample size is 8,000.

Estimation results.   Figure 4 shows frequentist and Bayesian estimation results for the population mortality parameters. The parameter estimates for POP(f) are represented by dashed black lines, and the median and the 95% equal-tailed credible intervals derived from the posterior distributions for POP(B), PF(B-G) and PF(B-logN) by, respectively, green, blue and red lines and areas.

The estimates for POP(f) and POP(B) overlap which means that estimating the Lee–Carter model and the time series model simultaneously gives practically
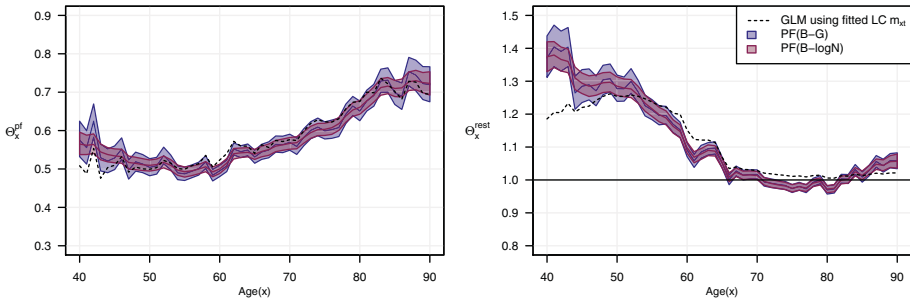
FIGURE 5: Parameter estimates for Θ_x^{pf} and Θ_x^{rest} using the original CMI portfolio. (Color online)
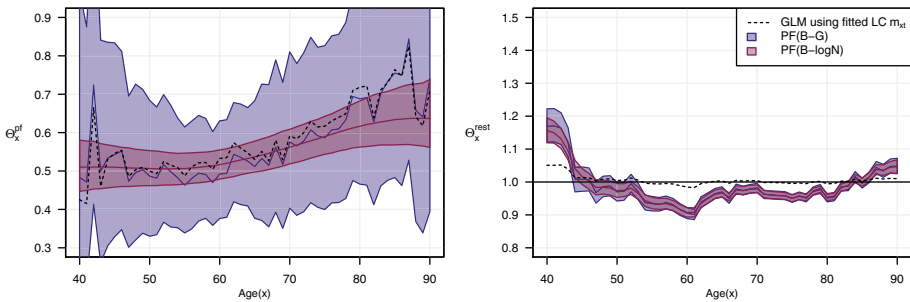
FIGURE 6: Parameter estimates for Θ_x^{pf} and Θ_x^{rest} when the CMI portfolio is reduced by a factor of 100. (Color online)

the same best estimates as a two-step frequentist approach. The specification of the hyperparameters therefore seems to have a limited effect on the posterior distributions of the parameters. In models $\mathsf{PF}(\mathsf{B}\text{-}\mathsf{G})$ and $\mathsf{PF}(\mathsf{B}\text{-}\mathsf{logN})$, we also include portfolio data. The credible intervals for α_x and κ_t are similar to the ones found for $\mathsf{POP}(\mathsf{B})$. For β_x, we observe differences for all ages. The prior specification for Θ_x^i (Gamma versus lognormal) does not have a large effect on the credible interval for β_x. The posterior distributions for the hyperparameters δ and σ_ε^2 are also similar for all model specifications.

Figure 5 shows estimates for the portfolio-specific factors using the different methods. The black line represents a frequentist method that corresponds to methods used in practice. First, the Lee–Carter model is estimated on population mortality. A Poisson GLM with age-dependent factors is then estimated in which the deaths in the portfolio are explained using the portfolio exposure and the fitted population mortality rate as offset:

$$D_{t,x}^i \sim \text{Poisson}(E_{t,x}^i \hat{\mu}_{t,x}^{\text{LC}} \cdot \Theta_x^i). \tag{27}$$

The blue and red areas again correspond to the 95% equal-tailed credible intervals for $\mathsf{PF}(\mathsf{B}\text{-}\mathsf{G})$ and $\mathsf{PF}(\mathsf{B}\text{-}\mathsf{logN})$.

The factors for the portfolio are all below 1, implying that mortality in the portfolio is lower than the baseline mortality rate, and in the rest group, the

factors are generally above 1. The baseline mortality rate $\mu_{t,x}$ in our model is estimated using $\mathcal{O}^{pop} \cup \mathcal{O}^{pf}$. Therefore, $\Theta_x^{pf} < 1$ does not automatically imply $\Theta_x^{rest} > 1$ or vice versa, since both Θ_x^{pf} and Θ_x^{rest} apply only to $\mathcal{O}^{pf}$. Estimated portfolio factors below 1 are in line with the results in Dowd *et al.* (2011) where mortality for the CMI dataset is shown to be significantly lower than for the England and Wales population.

The estimated factors from PF(B-G) show equally irregular behavior as the frequentist estimates for the factors. We find different estimated Lee–Carter parameters for POP(f) on the one hand and PF(B-G) or PF(B-logN) on the other hand. This leads to different baseline hazard rates $\mu_{t,x}$ which explains why the frequentist portfolio-specific factor estimates differ slightly from their Bayesian counterparts. The estimated factors for PF(B-logN), which incorporate dependence between ages within a group, are much smoother than the ones for PF(B-G), where we assume independence.

The posterior means of the mean reversion coefficients for the logNormal prior specification of Θ_x^i are $\rho_{\theta^{pf}} = 0.997$ and $\rho_{\theta^{rest}} = 0.999$.[9] We see in Figure 5 that the posterior distributions of Θ_x^{pf} have smaller credible intervals than the posterior distributions of Θ_x^{rest} for most ages. This can be explained by the fact that the portfolio is apparently more homogeneous than the remainder of the population for those ages.

We have investigated different constants for the prior distributions, but the estimated effects are hardly affected. Therefore, we conclude that any differences in parameter estimates are caused by differences in model and prior specification rather than prior constants.

Forecasting mortality.   Figure 7 shows projections of mortality rates from (1) a combination of POP(f) and frequentist estimates of portfolio-specific factors (hereafter indicated by PF(f)), (2) PF(B-G) and (3) PF(B-logN).[10] In these graphs, we only show fitted mortality rates for observations that are included in the likelihood, which means we consider the population for $t < 1990$ and the portfolio and the rest group for $t \geq 1990$. Projected mortality rates for the portfolio are less uncertain than the ones for the rest group in absolute terms, but not when the uncertainty is expressed as a percentage of the best estimate.

Projections of mortality rates in a Bayesian setting using the two different prior distributions for Θ_x^i show little difference; both the medians and standard deviations of the projections are similar.

We further observe that the prediction intervals from PF(B-G) and PF(B-logN) are similar to those from PF(f), though only the first two include parameter uncertainty. Our projections include uncertainty in the variance parameter σ_ϵ^2 in the time series model, and a higher variance leads to wider prediction intervals whereas a lower variance leads to narrower prediction intervals. Including the uncertainty in the variance parameter, therefore, does not necessarily lead to wider prediction intervals. The slightly wider prediction intervals further in the future are mainly caused by uncertainty in the drift parameter δ.
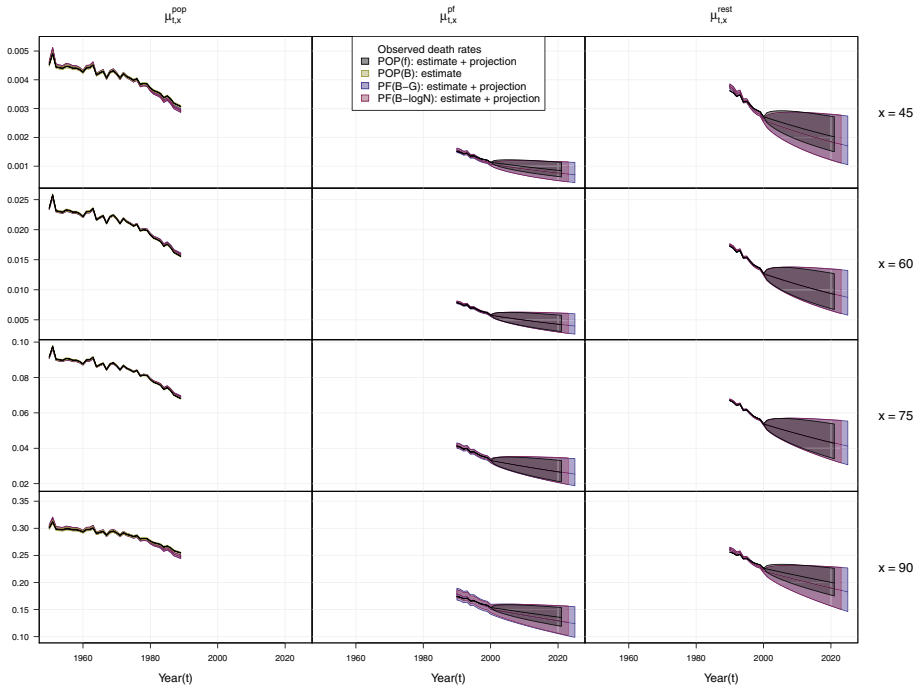
FIGURE 7: Estimated and projected mortality rates from POP(f) in combination with frequentist estimates of group-specific factors (red lines and areas), and from PF(B-G) and PF(B-logN) (blue and gray areas, respectively) using the original CMI portfolio. (Color online)

4.2. CMI assured lives — reduced portfolio size

The CMI dataset is much larger than any portfolio for a single insurance company. Haberman *et al.* (2014) consider a minimum annual exposure of 25,000 life years and a minimum of 8 years of observations sufficient to estimate a mortality model on the portfolio book itself. To assess how well our model performs on smaller datasets, we artificially reduce the size of the CMI portfolio. We divide observed deaths and exposures by a factor of 100, and the resulting deaths are subsequently rounded to the nearest integer. This ensures that the crude portfolio-specific factors remain largely the same as in the original dataset, which facilitates a comparison of the outcomes. The resulting dataset has on average 25,000 life years annually. We have again defined the rest group in such a way that the population is the disjoint union of the portfolio and the rest group for $(t, x) \in \mathcal{O}^{\text{pf}}$.

We use the same constants to define the prior distributions and the same initial values and settings in the MCMC algorithm as in the previous subsection. Convergence diagnostics again show good behavior; they have been made available in an online appendix.
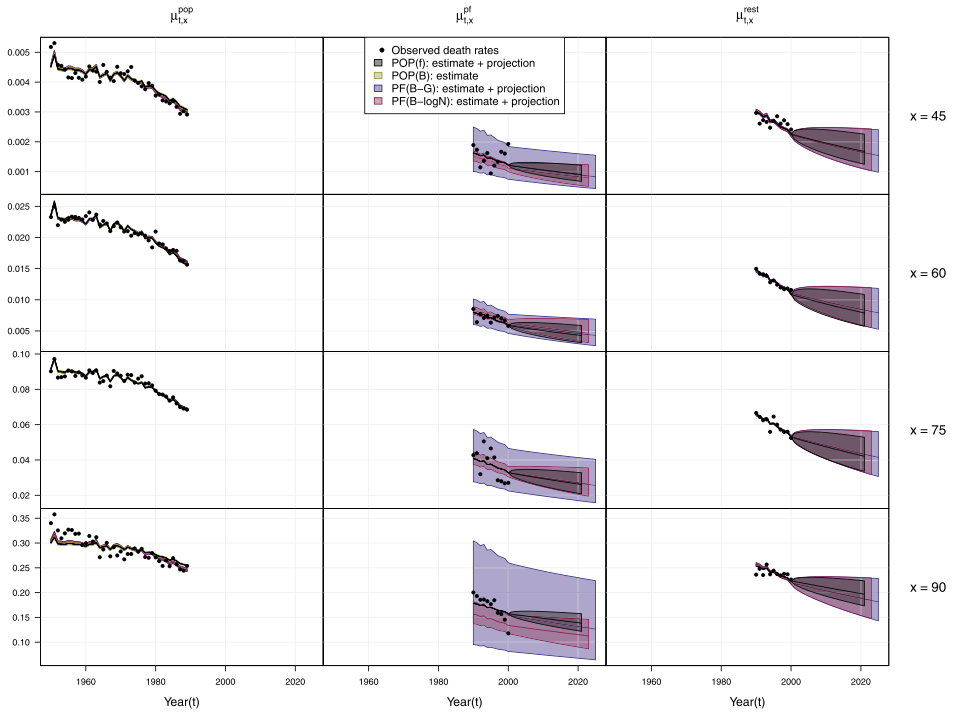
FIGURE 8: Estimated and projected mortality rates from POP(f) in combination with frequentist estimates of group-specific factors (black lines and gray areas), and from PF(B-G) and PF(B-logN) (blue and red areas, respectively) using the reduced CMI portfolio. (Color online)

Estimation results.   The posterior distributions for the Lee–Carter parameters are similar to those in Figure 4, so we do not show these. Figure 6 shows the portfolio-specific factors when estimated for the reduced portfolio. Since there are now fewer lives in our dataset, the observed portfolio death rates show more volatile behavior over the years.

The right graph of Figure 6 shows the estimated factors for the rest group, Θ_x^{rest}. Now that the portfolio has become smaller, this group constitutes a larger part of the general population. As a result, the posterior means are, in general, closer to 1 and the posterior credible intervals are slightly smaller. In the left graph, we observe how this results in frequentist estimates for Θ_x^{pf} which fluctuate more over the years (see the black line). The estimates for Θ_x^{pf} are also more volatile in the PF(B-G) model (in blue) when compared with the original portfolio, and the corresponding posterior credible interval is much wider. In the PF(B-logN) model (in red), the estimates are smoother than in Figure 5 while the posterior credible interval is again wider than before, but much less so than for PF(B-G). This is due to the smoothing characteristic of PF(B-logN): information from ages near x influences the estimates for Θ_x^i. The PF(B-logN) prior is more parsimonious than the one for PF(B-G) (it is a "shrinkage prior") and the effect of the prior specification on the posterior distribution is stronger

if less data is available. We believe it is reasonable to assume that portfolio-specific factors for ages close to each other are related, which makes the posterior credible intervals from PF(B-logN) more plausible than those from PF(B-G).

Forecasting mortality. Figure 8 shows projections of mortality rates from PF(f), PF(B-G) and PF(B-logN) using the reduced CMI portfolio, which can be compared to the mortality rate projections in Figure 7 where the whole CMI portfolio is used.

The credible intervals and prediction intervals from PF(B-G) for mortality rates in the reduced CMI portfolio are much wider which is caused by the wider credible intervals for Θ_x^{pf}, see Figure 6. The intervals for mortality rates for PF(B-logN) are a bit wider when the smaller CMI portfolio is used, but much less so than the ones for PF(B-G). Based on Figures 6 and 8, we thus conclude that parameter uncertainty in portfolio-specific factors can be substantial for small portfolios, and that the amount of uncertainty may strongly depend on the prior specification for the portfolio-specific factors.

Hoem (1973) already identified different sources of uncertainty in mortality predictions but no quantification was given for the relative impact of the difference sources of uncertainty. In Table 1, we investigate the relative impacts by showing the mean and standard deviation of the predicted numbers of deaths for future times $t \in \{2001, 2010, 2025\}$ based on observations until $s_S = 2000$, for $x = \{45, 65, 85\}$ and for all ages combined. To get the appropriate comparison, we use the exposures at time s_S for later times as well, so we take $E_{t,x}^{\text{pf}} = E_{s_S,x}^{\text{pf}}$ for $t \geq s_S$. The mortality scenarios correspond to the ones used in Figures 7 and 8, and the figures presented in Table 1 are constructed using four different methods as follows:

- Using PF(f), we predict mortality rates $\mu_{t,x}^{\text{pf}}$, taking into account uncertainty in the projection of the time series κ_t (but not parameter uncertainty). These mortality rates are then multiplied by the exposures $E_{t,x}^{\text{pf}} = E_{T,x}^{\text{pf}}$ to generate the expected number of deaths, given the scenario for mortality rates $\mathbb{E}[D_{t,x}^{\text{pf}} \mid \mu_{t,x}^{\text{pf}}]$ (first column with intervals).
- Using PF(f), we predict mortality rates $\mu_{t,x}^{\text{pf}}$, taking into account uncertainty in the projection of κ_t. We still do not include parameter uncertainty but for each generated mortality scenario, we include Poisson randomness by drawing random numbers of deaths $D_{t,x}^{\text{pf}} \sim \text{Poisson}(E_{t,x}^{\text{pf}} \mu_{t,x}^{\text{pf}})$ (second column).
- Using PF(B-G), we predict mortality rates $\mu_{t,x}^{\text{pf}}$, taking into account uncertainty in the projection of κ_t. Parameter uncertainty is now included, since we use the MCMC samples. For each generated mortality scenario, we draw random numbers of deaths $D_{t,x}^{\text{pf}} \sim \text{Poisson}(E_{t,x}^{\text{pf}} \mu_{t,x}^{\text{pf}})$ (third column).
- For PF(B-logN), our approach is similar to that for PF(B-G) (fourth column).

PREDICTIVE MEAN AND STANDARD DEVIATION FOR FUTURE NUMBERS OF DEATHS FOR DIFFERENT PROJECTION HORIZONS, SELECTED AGES AND FOR THE SUM OVER ALL AGES, FOR DIFFERENT MODELS THAT INCLUDE DIFFERENT SOURCES OF UNCERTAINTY AND FOR THE ORIGINAL AND REDUCED PORTFOLIO SIZE.

Original Portfolio Size

	Year	PF(f) TS	PF(f) TS + Pois	PF(B-G) TS + Pois + PU	PF(B-logN) TS + Pois + PU
$x = 45$	2001	(58.1 ; 1.9)	(58.1 ; 7.9)	(54.4 ; 7.8)	(54.6 ; 7.8)
	2010	(51.4 ; 5.4)	(51.4 ; 9.0)	(47.2 ; 9.5)	(47.3 ; 9.6)
	2025	(41.9 ; 7.0)	(41.9 ; 9.5)	(37.4 ; 11.1)	(37.4 ; 11.2)
$x = 65$	2001	(277.5 ; 9.2)	(277.5 ; 19.0)	(270.7 ; 19.5)	(274.1 ; 19.4)
	2010	(245.4 ; 25.8)	(245.4 ; 30.2)	(238.5 ; 32.8)	(241.7 ; 33.2)
	2025	(199.9 ; 33.4)	(199.9 ; 36.2)	(194.1 ; 44.0)	(196.6 ; 44.8)
$x = 85$	2001	(190.4 ; 3.6)	(190.4 ; 14.3)	(187.7 ; 14.8)	(189.6 ; 14.6)
	2010	(177.1 ; 10.7)	(177.1 ; 17.1)	(174.0 ; 18.4)	(175.6 ; 18.5)
	2025	(157.0 ; 15.1)	(157.0 ; 19.6)	(153.7 ; 23.1)	(154.8 ; 23.5)
$40 - 90$	2001	(10302.8 ; 307.8)	(10302.9 ; 324.1)	(10077.4 ; 334.0)	(10077.1 ; 332.7)
	2010	(9225.2 ; 868.2)	(9225.2 ; 873.5)	(8995.6 ; 970.7)	(9002.3 ; 972.8)
	2025	(7682.1 ; 1134.5)	(7682.1 ; 1137.9)	(7484.7 ; 1416.9)	(7493.6 ; 1425.5)

Reduced Portfolio Size

	Year	PF(f) TS	PF(f) TS + Pois	PF(B-G) TS + Pois + PU	PF(B-logN) TS + Pois + PU
$x = 45$	2001	(0.6 ; 0.0)	(0.6 ; 0.8)	(0.6 ; 0.8)	(0.6 ; 0.8)
	2010	(0.6 ; 0.1)	(0.6 ; 0.7)	(0.6 ; 0.8)	(0.5 ; 0.7)
	2025	(0.4 ; 0.1)	(0.4 ; 0.7)	(0.5 ; 0.7)	(0.4 ; 0.6)
$x = 65$	2001	(2.7 ; 0.1)	(2.7 ; 1.7)	(2.7 ; 1.7)	(2.9 ; 1.7)
	2010	(2.4 ; 0.3)	(2.4 ; 1.6)	(2.4 ; 1.6)	(2.5 ; 1.6)
	2025	(2.0 ; 0.3)	(2.0 ; 1.4)	(2.0 ; 1.5)	(2.1 ; 1.5)
$x = 85$	2001	(2.1 ; 0.0)	(2.1 ; 1.4)	(2.1 ; 1.5)	(1.7 ; 1.3)
	2010	(1.9 ; 0.1)	(1.9 ; 1.4)	(1.9 ; 1.5)	(1.6 ; 1.3)
	2025	(1.7 ; 0.2)	(1.7 ; 1.3)	(1.7 ; 1.4)	(1.4 ; 1.2)
$40–90$	2001	(103.2 ; 3.1)	(103.2 ; 10.6)	(104.7 ; 11.1)	(102.3 ; 10.9)
	2010	(92.4 ; 8.7)	(92.4 ; 12.9)	(94.3 ; 14.2)	(92.0 ; 14.0)
	2025	(77.0 ; 11.3)	(77.0 ; 14.3)	(79.3 ; 17.5)	(77.5 ; 17.4)

If only uncertainty in the evolution of the time series κ_t is taken into account, the uncertainty in the conditional expectation of $D_{t,x}^{\text{pf}}$ given $\mu_{t,x}^{\text{pf}}$ can be very small for small portfolios, see the first column in the bottom panel of Table 1. For larger portfolios, the uncertainty is much larger due to the higher exposures, as shown in the top panel of that table. The uncertainty also becomes larger if uncertainty

in the individual number of deaths (Poisson noise) is added. This is shown in the second column, and the effect is of course stronger for the smaller portfolio.

If we compare the results with parameter uncertainty (third and fourth column) and without parameter uncertainty (second column), we conclude that the impact on the predicted numbers of deaths is negligible compared to the impact of the Poisson noise due to individual deaths. The mortality prediction intervals for the smaller portfolio size with the Gamma prior are very wide, as shown in Figure 8. However, the uncertainty in the future numbers of deaths is similar to the cases where we only include time series and Poisson uncertainty. Therefore, we conclude that for the time horizons considered here, the Poisson noise due to individual deaths is more important than parameter uncertainty in the portfolio-specific factors, and this turns out to be true for the smaller but also for the larger portfolio.[11]

Given fixed mortality rates, death numbers at different ages are independent (but of course not identically distributed). Therefore, one might expect that the relevance of Poisson randomness disappears if we consider a whole insurance portfolio and thus look at the sum of random death numbers over all ages. Predicted means and standard deviations for this sum over all ages are shown in the bottom rows in both panels in Table 1. We observe that for the small portfolio including Poisson uncertainty leads to larger uncertainty even at the portfolio level, but this is not true for large portfolios. From comparing the second to the third and fourth column, we further see that parameter uncertainty has little impact on the short horizon, but the effect of parameter uncertainty increases with the projection horizon, regardless of portfolio size. This is explained by the uncertainty in the mortality trend parameter δ which has an effect on mortality rates that increases over time. We conclude the following:

- For large portfolios, individual mortality risk (modeled through Poisson noise) is important for individual ages, but not at the portfolio level. For small portfolios, individual mortality risk is important both for individual ages and at the portfolio level.
- For both large and small portfolios, parameter uncertainty in portfolio-specific factors is not relevant since it is overshadowed by Poisson noise.
- For both large and small portfolios, parameter uncertainty in the mortality trend is not relevant in the short term, but of increasing importance if the projection horizon increases.

The results regarding trend uncertainty are in line with the results for small portfolios in Haberman *et al.* (2014). However, we show that even at the portfolio level, Poisson and trend uncertainty cannot be ignored.

5. CONCLUSION

Proper risk management for portfolios in life insurance companies or pension funds requires a reliable method to estimate the distribution of future

deaths in such portfolios. This involves the modeling of population-wide mortality trends, a specification of portfolio-specific deviations from this trend, and the conditional distribution for the individual deaths in a portfolio, given its mortality rates. In this paper, we use Bayesian inference to analyze these three sources of uncertainty in life insurance portfolio data. This may help to generate scenarios for survival in a portfolio in which these three different components in the predictions can be explicitly distinguished.

The law of large numbers implies that the last component will be relatively small for very large portfolios. But when the portfolio under consideration is small or when observations have only been available for a limited number of years, it may be difficult to know a priori what part of the fluctuations in the observations over age and time should be assigned to genuine changes in mortality over time, to noise in the observations and to parameter uncertainty. For those cases, we believe that our method may turn out to be a useful alternative to what has been proposed in the actuarial literature so far.

By using both the CMI dataset of assured male lives and a scaled version of that dataset, we show that estimates of the difference between country-wide and portfolio-specific hazard rates strongly depend on a priori assumptions about the age dependence of that difference. Assuming that there is no dependence for different ages can give unrealistically large posterior credible intervals for portfolio-specific factors in small portfolios, while an alternative based on a autoregressive smoothing prior gives much more satisfactory results.

However, the impact of uncertainty in the portfolio-specific factors on the predictive distributions of future number of deaths in the portfolio is negligible compared to the Poisson noise that is added by individual deaths, regardless of the projection horizon. As the projection horizon increases, the effect of parameter uncertainty on predictive distributions becomes increasingly relevant, which is solely caused by uncertainty in the mortality trend. This reinforces our conclusion that a full analysis for small portfolios must always be based on an explicit description of the different sources for uncertainty in the predictive distributions of future deaths.

Supplementary Materials

For supplementary material for this article, please visit https://doi.org/10.1017/asb.2017.17

Notes

1. See Section 4 for a description of the CMI dataset on assured lives. Colored versions of all figures can be found online.

2. We could use a wider age range for the population, but we choose to use the same set of ages in the population as we have available for the portfolio. This way, we ensure the parameters α_x, β_x and κ_t are most appropriate for projection of mortality for the portfolio.

3. We use the Lee–Carter model to specify the baseline mortality, but our model can easily be extended to include e.g. a cohort effect as in Renshaw and Haberman (2006). Also, note that our model differs from the augmented common-factor model as in (4) since we do not include an extra dynamic factor when modeling subpopulations. The absence of such a term makes our model similar to the common factor model also discussed in Li and Lee (2005).

4. We use time-independent portfolio factors, because estimating time dynamics on a few historical years can lead to spurious forecasting results. As a result, mortality improvements are perfectly correlated between groups, and our model is less appropriate for assessing basis risk in longevity hedging as in Haberman *et al.* (2014).

5. Population mortality data is obtained from the Human Mortality Database. The Human Mortality Database is a joint project of the University of California, Berkeley (USA) and the Max Planck Institute for Demographic Research (Germany). Data are available at http://www.mortality.org.

6. For POP(f) and POP(B) , we only apply the parameter restrictions in (14), and for PF(B-G) and PF(B-logN) we apply the parameter restrictions in (14) and (15).

7. The large number of required iterations may be due to the high dimension of our model. However, since our Metropolis–Hastings algorithm for β consists of only one step, instead of the usual loop over all ages (see e.g. Czado *et al.* (2005) and Antonio *et al.* (2015)), using the Von Mises–Fisher distribution as proposal density speeds up the algorithm considerably.

8. Convergence diagnostics are available in an online appendix.

9. Parameters ρ_i close to 1 imply that a random walk (with drift) model might be more appropriate for log Θ_x^i. For this alternative approach, see Congdon (2009). However, since the estimates of the parameters ρ_i are already close to 1, we expect that the posterior distributions for other parameters will not differ significantly for a random walk specification.

10. These mortality projections are constructed as follows. For each MCMC sample, we generate 100 scenarios for future κ_t's using κ_T, δ and σ_ε^2. The mortality rates are then constructed using the other parameters α_x, β_x and Θ_x^i from that sample. Hence, a total of 800,000 scenarios are used to construct the prediction intervals in Figure 7.

11. When sampling the individual deaths in practice (e.g. for portfolio valuation purposes), one may prefer to use the Bernoulli distribution. Here we use the Poisson distribution to remain consistent with the approach used for estimation.

References

ANTONIADIS, A., GRÉGOIRE, G. and MCKEAGUE, I. (2004) Bayesian estimation in single-index models. *Statistica Sinica*, **14**, 1147–1164.

ANTONIO, K., BARDOUTSOS, A. and OUBURG, W. (2015) A Bayesian Poisson log-bilinear model for mortality projections with multiple populations. *European Actuarial Journal*, **5**(2), 245–281.

ANTONIO, K. AND ZHANG, Y. (2014) Nonlinear mixed models. In *Predictive Modeling Applications in Actuarial Science* (eds. E. Frees, R. Derrig and G. Meyers), volume 1, pp. 398–426. New York: Cambridge University Press.

BARRIEU, P., BENSUSAN, H., KAROUI, N.E., HILLAIRET, C., LOISEL, S., RAVANELLI, C. and SALHI, Y. (2012) Understanding, modelling and managing longevity risk: Key issues and main challenges. *Scandinavian Actuarial Journal*, **3**, 203–231.

BROUHNS, N., DENUIT, M. and VERMUNT, J. (2002) A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, **31**(3), 373–393.

CAIRNS, A., BLAKE, D. and DOWD, K. (2006) A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance*, **73**(4), 687–718.

CAIRNS, A., BLAKE, D., DOWD, K., COUGHLAN, G., 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), 1–35.

CAIRNS, A., BLAKE, D., DOWD, K., COUGHLAN, G. and KHALAF-ALLAH, M. (2011) Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin*, **41**(1), 25–59.

CONGDON, P. (2009) Life expectancies for small areas: A Bayesian random effects methodology. *International Statistical Review*, **77**(2), 222–240.

CZADO, C., DELWARDE, A. and DENUIT, M. (2005) Bayesian Poisson log-bilinear mortality projections. *Insurance: Mathematics and Economics*, **36**(3), 260–284.

DENUIT, M., DHAENE, J., GOOVAERTS, M. and KAAS, R. (2005) *Actuarial Theory for Dependent Risks*. Chichester, UK: John Wiley & Sons.

DENUIT, M., MARÉCHAL, X., PITREBOIS, S. and WALHIN, J.-F. (2007) *Actuarial Modelling of Claim Counts*. John Wiley & Sons, Ltd.

DOWD, K., CAIRNS, A., BLAKE, D., COUGHLAN, G. and KHALAF-ALLAH, M. (2011) A gravity model of mortality rates for two related populations. *North American Actuarial Journal*, **15**(2), 334–356.

FINKELSTEIN, A. and POTERBA, J. (2002) Selection effects in the United Kingdom individual annuities market. *The Economic Journal*, **112**(476), 28–50.

FISHER, R. (1953) Dispersion on a sphere. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, **217**(1130), 295–305.

GELMAN, A. (2006) Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, **1**(3), 515–534.

GSCHLÖSSL, S., SCHOENMAEKERS, P. and DENUIT, M. (2011) Risk classification in life insurance: Methodology and case study. *European Actuarial Journal*, **1**, 23–41.

HABERMAN, S., KAISHEV, V., MILLOSSOVICH, P., VILLEGAS, A., BAXTER, S., GACHES, A., GUNNLAUGSSON, S. and SISON, M. (2014) Longevity basis risk: A methodology for assessing basis risk. Technical report, Institute and Faculty of Actuaries. Available online at: www.actuaries.org.uk/events/pages/sessional-research-programme.

HABERMAN, S. and RENSHAW, A. (2011) A comparative study of parametric mortality projection models. *Insurance: Mathematics and Economics*, **48**(1), 35–55.

HOEM, J. (1973) Levels of error in population forecasts. *Artikler fra Statistisk Sentralbyrå*, **61**, 1–46.

HOFF, P. (2009) Simulation of the matrix Bingham-von-Mises-Fisher distribution, with applications to multivariate and relational data. *Journal of Computation and Graphical Statistics*, **18**(3), 438–456.

KAN, H. (2012) A Bayesian mortality forecasting framework for population and portfolio mortality. MSc thesis, University of Amsterdam, The Netherlands.

KLEINOW, T. (2015) A common age effect model for the mortality of multiple populations. *Insurance: Mathematics and Economics*, **63**, 147–152.

LANTZ, P., HOUSE, J., LEPKOWSKI, J., WILLIAMS, D., MERO, R. and CHEN, J. (1998) Socioeconomic factors, health behaviors, and mortality: Results from a nationally representative prospective study of US adults. *Journal of the American Medical Association*, **279**(21), 1703–1708.

LEE, R. and CARTER, L. (1992) Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, **87**(419), 659–671.

LI, J. (2014) An application of MCMC simulation in mortality projection for populations with limited data. *Demographic Research*, **30**(1), 1–48.

LI, N. AND LEE, R. (2005) Coherent mortality forecasts for a group of populations: An extension of the Lee–Carter method. *Demography*, **42**(3), 575–594.

OLIVIERI, A. (2011) Stochastic mortality: Experience-based modeling and application issues consistent with Solvency 2. *European Actuarial Journal*, **1**, S101–S125.

PITACCO, E., DENUIT, M., HABERMAN, S. and OLIVIERI, A. (2009) *Modelling Longevity Dynamics for Pensions and Annuity Business*. New York: Oxford University Press.

PLAT, R. (2009) Stochastic portfolio specific mortality and the quantification of mortality basis risk. *Insurance: Mathematics and Economics*, **45**, 123–132.

PURCARU, O., GUILLÉN, M. and DENUIT, M. (2004) Linear credibility models based on time series for claim counts. *Belgian Actuarial Bulletin*, **4**(1), 62–74.

RENSHAW, A. and HABERMAN, S. (2006) A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, **38**(3), 556–570.

RICHARDS, S., KAUFHOLD, K. and ROSENBUSCH, S. (2013) Creating portfolio-specific mortality tables: A case study. *European Actuarial Journal*, **3**, 295–319.

VAN BERKUM, F., ANTONIO, K. and VELLEKOOP, M. (2016) The impact of multiple structural changes on mortality predictions. *Scandinavian Actuarial Journal*, **7**, 581–603.

VILLEGAS, A. and HABERMAN, S. (2014) On the modeling and forecasting of socioeconomic mortality differentials: An application to deprivation and mortality in England. *North American Actuarial Journal*, **18**(1), 168–193.

VON MISES, R. (1918) Über die 'Ganzzahligkeit' der Atomgewicht und verwandte Fragen. *Physikalische Zeitschrift*, **19**, 490–500.

FRANK VAN BERKUM (Corresponding author)
Faculty of Economics and Business
University of Amsterdam
Amsterdam, The Netherlands
E-Mail: f.vanberkum@uva.nl

KATRIEN ANTONIO
Faculty of Economics and Business
University of Amsterdam
Amsterdam, The Netherlands
Faculty of Economics and Business
KU Leuven, Belgium

MICHEL VELLEKOOP
Faculty of Economics and Business
University of Amsterdam
Amsterdam, The Netherlands

APPENDIX

A. POSTERIOR DISTRIBUTIONS

We derive the posterior distribution for all parameters in the model. For convenience, we define the following variables:

$$D = \{D^{\mathrm{pop}}, D^{\mathrm{pf}}, D^{\mathrm{rest}}\}, \qquad E = \{E^{\mathrm{pop}}, E^{\mathrm{pf}}, E^{\mathrm{rest}}\},$$

$$\alpha = \{\alpha_{x_1}, \ldots, \alpha_{x_X}\}, \qquad \beta = \{\beta_{x_1}, \ldots, \beta_{x_X}\}, \qquad \kappa = \{\kappa_{t_1}, \ldots, \kappa_{t_T}\},$$

$$\Theta = \{\Theta^{\mathrm{pf}}, \Theta^{\mathrm{rest}}\}, \qquad \rho_\theta = \{\rho_{\mathrm{pf}}, \rho_{\mathrm{rest}}\}, \qquad \sigma_\theta^2 = \{\sigma_{\mathrm{pf}}^2, \sigma_{\mathrm{rest}}^2\},$$

with $(D_{t,x}^{\mathrm{pop}}, E_{t,x}^{\mathrm{pop}})$ defined on $(t, x) \in \mathcal{O}^{\mathrm{pop}}$ and $(D_{t,x}^i, E_{t,x}^i)$ defined on $(t, x) \in \mathcal{O}^{\mathrm{pf}}$ for $i \in \{\mathrm{pf}, \mathrm{rest}\}$. See Section 3 for the definition of $\mathcal{O}^{\mathrm{pop}}$ and $\mathcal{O}^{\mathrm{pf}}$.

We further define the set Λ that contains both data and parameters:

$$\Lambda = \{D, E, \alpha, \beta, \kappa, \delta, \sigma_\varepsilon^2, \Theta, \rho_\theta, \sigma_\theta^2\},$$

and remark that ρ_θ and σ_θ^2 are not needed when we use a Gamma prior for the portfolio-specific factors.

A.1. Age parameters for population mortality

A.1.1. *Gibbs sampling for α_x.* The individual α_x's are independent. Therefore, the posterior distribution for a single $e_x = \exp(\alpha_x)$ with $x_1 \leq x \leq x_X$ is given by

$$f(e_x|\Lambda \setminus \{e_x\}) \propto f(D|E, e, \beta, \kappa, \Theta^{\mathrm{pf}}, \Theta^{\mathrm{rest}}) f(e_x) \tag{A1}$$

$$\propto \prod_{t \in \mathcal{T}} \left(e^{-E_{t,x}^{\mathrm{pf}} e_x \exp[\beta_x \kappa_t] \Theta_x^{\mathrm{pf}}} \frac{\left(E_{t,x}^{\mathrm{pf}} e_x \exp[\beta_x \kappa_t] \Theta_x^{\mathrm{pf}}\right)^{D_{t,x}^{\mathrm{pf}}}}{D_{t,x}^{\mathrm{pf}}!} \right)^{I_{t,x}^{\mathrm{pf}}}$$

$$\times \prod_{t \in \mathcal{T}} \left(e^{-E_{t,x}^{\mathrm{rest}} e_x \exp[\beta_x \kappa_t] \Theta_x^{\mathrm{rest}}} \frac{\left(E_{t,x}^{\mathrm{rest}} e_x \exp[\beta_x \kappa_t] \Theta_x^{\mathrm{rest}}\right)^{D_{t,x}^{\mathrm{rest}}}}{D_{t,x}^{\mathrm{rest}}!} \right)^{I_{t,x}^{\mathrm{rest}}}$$

$$\times \prod_{t \in \mathcal{T}} \left(e^{-E_{t,x}^{\mathrm{pop}} e_x \exp[\beta_x \kappa_t]} \frac{\left(E_{t,x}^{\mathrm{pop}} e_x \exp[\beta_x \kappa_t]\right)^{D_{t,x}^{\mathrm{pop}}}}{D_{t,x}^{\mathrm{pop}}!} \right)^{I_{t,x}^{\mathrm{pop}}}$$

$$\times \frac{b_x^{a_x}}{\Gamma(a_x)} e_x^{a_x - 1} \exp[-b_x e_x]$$

$$\propto \exp[-(b_x + d_x)e_x] \cdot e_x^{a_x + D_{\bullet x} - 1},$$

with

$$d_x = \sum_{t \in \mathcal{T}} \left\{ I_{t,x}^{\mathrm{pf}} \left(E_{t,x}^{\mathrm{pf}} \exp[\beta_x \kappa_t] \Theta_x^{\mathrm{pf}}\right) + I_{t,x}^{\mathrm{rest}} \left(E_{t,x}^{\mathrm{rest}} \exp[\beta_x \kappa_t] \Theta_x^{\mathrm{rest}}\right) + I_{t,x}^{\mathrm{pop}} \left(E_{t,x}^{\mathrm{pop}} \exp[\beta_x \kappa_t]\right) \right\}$$

and

$$D_{\bullet x} = \sum_{t \in \mathcal{T}} \left\{ I_{t,x}^{\mathrm{pf}} \cdot D_{t,x}^{\mathrm{pf}} + I_{t,x}^{\mathrm{rest}} \cdot D_{t,x}^{\mathrm{rest}} + I_{t,x}^{\mathrm{pop}} \cdot D_{t,x}^{\mathrm{pop}} \right\} = \sum_{t \in \mathcal{T}} D_{t,x}^{\mathrm{pop}}.$$

The last line in (A1) is proportional to a Gamma$(a_x + D_{\bullet x}, b_x + d_x)$ distribution. Therefore, we can use Gibbs sampling to draw a new value of e_x, which can subsequently be transformed into a new value of α_x.

A.1.2. *Metropolis sampling for β_x.* The posterior distribution for $\boldsymbol{\beta}$ is given by

$$f(\boldsymbol{\beta}|\boldsymbol{\Lambda}\setminus\{\boldsymbol{\beta}\}) \propto f(\boldsymbol{D}|\boldsymbol{E}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{\Theta}^{\mathrm{pf}}, \boldsymbol{\Theta}^{\mathrm{rest}}) f(\boldsymbol{\beta}) \tag{A2}$$

$$\propto \prod_{x\in\mathcal{X}} \prod_{t\in\mathcal{T}} \left(e^{-E_{t,x}^{\mathrm{pf}} e_x \exp[\beta_x \kappa_t] \Theta_x^{\mathrm{pf}}} \frac{\left(E_{t,x}^{\mathrm{pf}} e_x \exp[\beta_x \kappa_t] \Theta_x^{\mathrm{pf}}\right)^{D_{t,x}^{\mathrm{pf}}}}{D_{t,x}^{\mathrm{pf}}!} \right)^{I_{t,x}^{\mathrm{pf}}}$$

$$\times \prod_{x\in\mathcal{X}} \prod_{t\in\mathcal{T}} \left(e^{-E_{t,x}^{\mathrm{rest}} e_x \exp[\beta_x \kappa_t] \Theta_x^{\mathrm{rest}}} \frac{\left(E_{t,x}^{\mathrm{rest}} e_x \exp[\beta_x \kappa_t] \Theta_x^{\mathrm{rest}}\right)^{D_{t,x}^{\mathrm{rest}}}}{D_{t,x}^{\mathrm{rest}}!} \right)^{I_{t,x}^{\mathrm{rest}}}$$

$$\times \prod_{x\in\mathcal{X}} \prod_{t\in\mathcal{T}} \left(e^{-E_{t,x}^{\mathrm{pop}} e_x \exp[\beta_x \kappa_t]} \frac{\left(E_{t,x}^{\mathrm{pop}} e_x \exp[\beta_x \kappa_t]\right)^{D_{t,x}^{\mathrm{pop}}}}{D_{t,x}^{\mathrm{pop}}!} \right)^{I_{t,x}^{\mathrm{pop}}}$$

$$\times \exp(c_\beta \boldsymbol{\mu}_\beta^T \boldsymbol{\beta}).$$

Given a current value $\tilde{\boldsymbol{\beta}}$ and scaling parameter d_β, we sample a proposal $\hat{\boldsymbol{\beta}}$ from the distribution $\mathrm{vMF}(\tilde{\boldsymbol{\beta}}, d_\beta)$. The proposal distribution is symmetric, and the acceptance probability is thus given by

$$\phi = \min\left\{ \frac{f(\hat{\boldsymbol{\beta}}; |\boldsymbol{\Lambda}\setminus\{\hat{\boldsymbol{\beta}}\})}{f(\tilde{\boldsymbol{\beta}}|\boldsymbol{\Lambda}\setminus\{\tilde{\boldsymbol{\beta}}\})}; 1 \right\}.$$

A.2. Period parameters for population mortality

A.2.1. *Metropolis sampling for κ_t.* Define $\boldsymbol{\kappa}_{-t} = \{\kappa_{t_1}, \ldots, \kappa_{t-1}, \kappa_{t+1}, \ldots, \kappa_{s_S}\}$. The posterior distribution of κ_t for $t_1 < t < t_T$ and $s_1 < t \leq s_S$ is given by

$$f(\kappa_t|\boldsymbol{\Lambda}\setminus\{\kappa_t\}) \propto f(\boldsymbol{D}|\boldsymbol{E}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{\Theta}^{\mathrm{pf}}, \boldsymbol{\Theta}^{\mathrm{rest}}) f(\boldsymbol{\kappa}|\kappa_{t_1}, \delta, \sigma_\varepsilon^2) \tag{A3}$$

$$\propto \prod_{x\in\mathcal{X}} \left[\exp\left(-E_{t,x}^{\mathrm{pf}} \exp[\alpha_x + \beta_x \kappa_t] \Theta_x^{\mathrm{pf}}\right) \exp\left(D_{t,x}^{\mathrm{pf}} \beta_x \kappa_t\right) \right]^{I_{t,x}^{\mathrm{pf}}}$$

$$\times \prod_{x\in\mathcal{X}} \left[\exp\left(-E_{t,x}^{\mathrm{rest}} \exp[\alpha_x + \beta_x \kappa_t] \Theta_x^{\mathrm{rest}}\right) \exp\left(D_{t,x}^{\mathrm{rest}} \beta_x \kappa_t\right) \right]^{I_{t,x}^{\mathrm{rest}}}$$

$$\times \prod_{x\in\mathcal{X}} \left[\exp\left(-E_{t,x}^{\mathrm{pop}} \exp[\alpha_x + \beta_x \kappa_t]\right) \exp\left(D_{t,x}^{\mathrm{pop}} \beta_x \kappa_t\right) \right]^{I_{t,x}^{\mathrm{pop}}}$$

$$\times f(\kappa_t|\boldsymbol{\kappa}_{-t}, \delta, \sigma_\varepsilon^2),$$

in which the expression in the last line can be simplified:

- for $t_1 < t < t_T$ and $s_1 < t < s_S$:

$$f(\kappa_t|\boldsymbol{\kappa}_{-t}, \delta, \sigma_\varepsilon^2) \propto f(\kappa_t|\kappa_{t-1}, \delta, \sigma_\varepsilon^2) f(\kappa_{t+1}|\kappa_t, \delta, \sigma_\varepsilon^2)$$

$$\sim \mathrm{N}\left(\frac{1}{2}(\kappa_{t-1} + \kappa_{t+1}), \frac{1}{2}\sigma_\varepsilon^2 \right),$$

- for $t = s_S$:

$$f(\kappa_t | \boldsymbol{\kappa}_{-t}, \delta, \sigma_\varepsilon^2) \propto f(\kappa_t | \kappa_{t-1}, \delta, \sigma_\varepsilon^2)$$
$$\sim N\left(\kappa_{t-1} + \delta, \sigma_\varepsilon^2\right).$$

For κ_{t_T} and κ_{s_1}, we derive the *joint* posterior distribution, since we have applied the restriction $\kappa_{t_T} = \kappa_{s_1}$. It is given by

$$f(\kappa_{t_T}, \kappa_{s_1} | \boldsymbol{\Lambda} \backslash \{\kappa_{t_T}, \kappa_{s_1}\}) \propto f(\boldsymbol{D} | \boldsymbol{E}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{\Theta}^{\text{pf}}, \boldsymbol{\Theta}^{\text{rest}}) f(\boldsymbol{\kappa} | \kappa_{t_1}, \delta, \sigma_\varepsilon^2) \tag{A4}$$

$$\propto \prod_{t=t_T}^{s_1} \prod_{x \in \mathcal{X}} \left[\exp\left(-E_{t,x}^{\text{pf}} \exp[\alpha_x + \beta_x \kappa_t] \Theta_x^{\text{pf}}\right) \exp\left(D_{t,x}^{\text{pf}} \beta_x \kappa_t\right)\right]^{I_{t,x}^{\text{pf}}}$$

$$\times \prod_{t=t_T}^{s_1} \prod_{x \in \mathcal{X}} \left[\exp\left(-E_{t,x}^{\text{rest}} \exp[\alpha_x + \beta_x \kappa_t] \Theta_x^{\text{rest}}\right) \exp\left(D_{t,x}^{\text{rest}} \beta_x \kappa_t\right)\right]^{I_{t,x}^{\text{rest}}}$$

$$\times \prod_{t=t_T}^{s_1} \prod_{x \in \mathcal{X}} \left[\exp\left(-E_{t,x}^{\text{pop}} \exp[\alpha_x + \beta_x \kappa_t]\right) \exp\left(D_{t,x}^{\text{pop}} \beta_x \kappa_t\right)\right]^{I_{t,x}^{\text{pop}}}$$

$$\times f(\kappa_{t_T}, \kappa_{s_1} | \boldsymbol{\kappa}_{-\{t_T, s_1\}}, \delta, \sigma_\varepsilon^2),$$

in which the expression in the last line can be simplified:

$$f(\kappa_{t_T}, \kappa_{s_1} | \boldsymbol{\kappa}_{-\{t_T, s_1\}}, \delta, \sigma_\varepsilon^2) = f(\kappa_{t_T} | \boldsymbol{\kappa}_{-\{t_T, s_1\}}, \delta, \sigma_\varepsilon^2)$$
$$\propto f(\kappa_{t_T} | \kappa_{t_T-1}, \delta, \sigma_\varepsilon^2) f(\kappa_{s_1+1} | \kappa_{s_1}, \delta, \sigma_\varepsilon^2)$$
$$\sim N\left(\frac{1}{2}(\kappa_{t_T-1} + \kappa_{s_1+1}), \frac{1}{2}\sigma_\varepsilon^2\right).$$

Given a current value $\tilde{\kappa}_t$ and Metropolis sampling variance $s_{\kappa_t}^2$, we sample a proposal $\hat{\kappa}_t$ from the distribution $N(\tilde{\kappa}_t, s_{\kappa_t}^2)$. This proposal distribution is symmetric, and the acceptance probability is thus given by

$$\phi = \min\left\{\frac{f(\hat{\kappa}_t | \boldsymbol{\Lambda} \backslash \{\hat{\kappa}_t\})}{f(\tilde{\kappa}_t | \boldsymbol{\Lambda} \backslash \{\tilde{\kappa}_t\})}; 1\right\}.$$

A.2.2. *Gibbs sampling for* δ. Define $\Delta\kappa_t = \kappa_t - \kappa_{t-1}$. Note that we have applied the restriction $\kappa_{s_1} := \kappa_{s_1-1} = \kappa_{t_T}$, and that summations are therefore not simply over all t. The posterior distribution of δ is given by

$$f(\delta | \boldsymbol{\Lambda} \backslash \{\delta\}) \propto f(\boldsymbol{\kappa} | \kappa_1, \delta, \sigma_\varepsilon^2) f(\delta) \tag{A5}$$

$$\propto \exp\left[-\sum_{t=t_1+1}^{t_T} \frac{[\Delta\kappa_t - \delta]^2}{2\sigma_\varepsilon^2} - \sum_{t=s_1+1}^{s_S} \frac{[\Delta\kappa_t - \delta]^2}{2\sigma_\varepsilon^2}\right] \cdot \exp\left[-\frac{[\delta - \mu_\delta]^2}{2\sigma_\delta^2}\right]$$

$$\propto \exp\left[-\frac{1}{2a_\delta}\left(\delta^2 - 2\delta b_\delta\right)\right]$$

$$\sim N\left(b_\delta, a_\delta\right),$$

with

$$a_\delta = \frac{\sigma_\varepsilon^2}{(T^* - 2) + \sigma_\varepsilon^2/\sigma_\delta^2},$$

and

$$b_\delta = \frac{(T^* - 2)}{(T^* - 2) + \sigma_\varepsilon^2/\sigma_\delta^2} \cdot \left(\frac{1}{(T^* - 2)} \left\{ \sum_{t=t_1+1}^{t_T} \Delta\kappa_t + \sum_{t=s_1+1}^{s_S} \Delta\kappa_t \right\} \right) + \frac{\sigma_\varepsilon^2/\sigma_\delta^2}{(T^* - 2) + \sigma_\varepsilon^2/\sigma_\delta^2} \cdot \mu_\delta.$$

We can use Gibbs sampling to draw a new value for δ.

A.2.3. *Gibbs sampling for σ_ε^2.* The posterior distribution of σ_ε^2 is given by

$$f(\sigma_\varepsilon^2|\mathbf{\Lambda}\backslash\{\sigma_\varepsilon^2\}) \propto f(\kappa|\kappa_1, \delta, \sigma_\varepsilon^2) f(\sigma_\varepsilon^2) \tag{A6}$$

$$= \prod_{t=t_1+1}^{t_T} \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left[-\frac{[\Delta\kappa_t - \delta]^2}{2\sigma_\varepsilon^2} \right] \cdot \prod_{t=s_1+1}^{s_S} \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left[-\frac{[\Delta\kappa_t - \delta]^2}{2\sigma_\varepsilon^2} \right]$$

$$\times \sigma_\varepsilon^{-1} \cdot 1_{[0 \leq \sigma_\varepsilon \leq A_\varepsilon]}$$

$$\propto (\sigma_\varepsilon^{-2})^{\frac{T^*-1}{2}} \exp\left[-(\sigma_\varepsilon^{-2}) \cdot \frac{1}{2} \left(\sum_{t=t_1+1}^{t_T} (\Delta\kappa_t - \delta)^2 + \sum_{t=s_1+1}^{s_S} (\Delta\kappa_t - \delta)^2 \right) \right].$$

Therefore, we know that the posterior distribution of σ_ε^{-2} is

$$f(\sigma_\varepsilon^{-2}|\mathbf{\Lambda}\backslash\{\sigma_\varepsilon^2\}) \propto (\sigma_\varepsilon^{-2})^{\frac{T^*-1}{2}-1-1} \cdot \exp\left[-(\sigma_\varepsilon^{-2}) \cdot \frac{1}{2} \left(\sum_{t=t_1+1}^{t_T} (\Delta\kappa_t - \delta)^2 + \sum_{t=s_1+1}^{s_S} (\Delta\kappa_t - \delta)^2 \right) \right]$$

$$\sim \text{Gamma}\left(\frac{T^*-3}{2}, \frac{1}{2} \left\{ \sum_{t=t_1+1}^{t_T} (\Delta\kappa_t - \delta)^2 + \sum_{t=s_1+1}^{s_S} (\Delta\kappa_t - \delta)^2 \right\} \right).$$

We can use Gibbs sampling to draw new values of σ_ε^{-2} which can be transformed to σ_ε^2.

A.3. Portfolio-specific mortality — Gamma prior

The posterior distribution of Θ_x^i for $i \in \{\text{pf}, \text{rest}\}$ and $y_1 \leq x \leq y_Y$ is given by

$$f(\Theta_x^i|\mathbf{\Lambda}\backslash\{\Theta_x^i\}) \propto f(\mathbf{D}|\mathbf{E}, \mathbf{\Theta}^{\text{pf}}, \mathbf{\Theta}^{\text{rest}}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}) f(\Theta_x^i) \tag{A7}$$

$$\propto \prod_{t \in \mathcal{S}} \left(e^{-E_{t,x}^i \exp[\alpha_x + \beta_x\kappa_t]\Theta_x^i} \frac{(E_{t,x}^i \exp[\alpha_x + \beta_x\kappa_t]\Theta_x^i)^{D_{t,x}^i}}{D_{t,x}^i!} \right)^{I_{t,x}^i}$$

$$\times \frac{(c_x^i)^{c_x^i}}{\Gamma(c_x^i)} (\Theta_x^i)^{c_x^i-1} \exp[-c_x^i\Theta_x^i]$$

$$\propto \exp[-(c_x^i + f_x^i)\Theta_x^i] \cdot (\Theta_x^i)^{c_x^i + D_{\bullet,x}^i - 1},$$

with

$$f_x^i = \sum_{t\in\mathcal{S}} I_{t,x}^i \cdot E_{t,x}^i \exp[\alpha_x + \beta_x\kappa_t] \quad \text{and} \quad D_{\bullet,x}^i = \sum_{t\in\mathcal{S}} I_{t,x}^i \cdot D_{t,x}^i.$$

The last line in (A7) is proportional to a Gamma($c_x^i + D_{\bullet,x}^i$, $c_x^i + f_x^i$) distribution and we can, therefore, use Gibbs sampling to obtain new values for Θ_x^i. Note that the posterior mean can be written as

$$\frac{c_x^i}{c_x^i + \sum_{t\in\mathcal{S}} I_{t,x}^i \cdot E_{t,x}^i \mu_{t,x}} \cdot 1 + \frac{\sum_{t\in\mathcal{S}} I_{t,x}^i \cdot E_{t,x}^i \mu_{t,x}}{c_x^i + \sum_{t\in\mathcal{S}} I_{t,x}^i \cdot E_{t,x}^i \mu_{t,x}} \cdot \frac{\sum_{t\in\mathcal{S}} I_{t,x}^i \cdot D_{t,x}^i}{\sum_{t\in\mathcal{S}} I_{t,x}^i \cdot E_{t,x}^i \mu_{t,x}}.$$

If c_x^i is chosen small relative to $\sum_{t\in\mathcal{S}} I_{t,x}^i \cdot E_{t,x}^i \mu_{t,x}$, the posterior mean is close to $\frac{\sum_{t\in\mathcal{S}} I_{t,x}^i \cdot D_{t,x}^i}{\sum_{t\in\mathcal{S}} I_{t,x}^i \cdot E_{t,x}^i \mu_{t,x}}$ which is often used in practice to determine portfolio-specific factors.

A.4. Portfolio-specific mortality — lognormal prior

Define $\Theta^i = \{\Theta_{y_1}^i, \dots, \Theta_{y_Y}^i\}$. The mean reverting process on the log of the portfolio-specific factors in (26) can also be written as a multivariate lognormal distribution (Purcaru *et al.*, 2004, Section 3.3.2):

$$\Theta^i \sim \text{lognormal}(\tilde{\boldsymbol{\mu}}_i, \boldsymbol{\Sigma}_i), \tag{A8}$$

with $\tilde{\boldsymbol{\mu}}_i = -\frac{1}{2}\sigma_i^2 \mathbf{1}_Y$ where $\mathbf{1}_Y$ is a column vector of ones of length Y and $(\boldsymbol{\Sigma}_i)_{xy} = \rho_i^{|x-y|}\sigma_i^2$. Before we derive the posterior distribution for Θ_x^i and the hyperparameters, we define the following variables and relations:

$$\boldsymbol{\Sigma}_i = \sigma_i^2 \cdot \boldsymbol{\Gamma}(\rho_i)$$

$$\boldsymbol{\Sigma}_i^{-1} = \frac{1}{\sigma_i^2} \cdot \boldsymbol{\Gamma}^{-1}(\rho_i) = \frac{1}{\sigma_i^2}\frac{1}{1-\rho_i^2} \cdot \tilde{\boldsymbol{\Gamma}}^{-1}(\rho_i)$$

$$|\boldsymbol{\Sigma}_i| = \left|\sigma_i^2 \cdot \boldsymbol{\Gamma}(\rho_i)\right| = (\sigma_i^2)^Y \cdot (1-\rho_i^2)^{Y-1}$$

$$\boldsymbol{\Psi}^i = \ln\boldsymbol{\Theta}^i - \tilde{\boldsymbol{\mu}}_i = \ln\boldsymbol{\Theta}^i + \frac{1}{2}\sigma_i^2\mathbf{1}_Y,$$

with

$$\boldsymbol{\Gamma}(\rho) = \begin{pmatrix} 1 & \rho & \cdots & \rho^{Y-2} & \rho^{Y-1} \\ \rho & 1 & \cdots & \rho^{Y-3} & \rho^{Y-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho^{Y-2} & \rho^{Y-3} & \cdots & 1 & \rho \\ \rho^{Y-1} & \rho^{Y-2} & \cdots & \rho & 1 \end{pmatrix}, \quad \tilde{\boldsymbol{\Gamma}}^{-1}(\rho) = \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & \ddots & 0 \\ 0 & -\rho & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & 1+\rho^2 & -\rho \\ 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}.$$

A.4.1. *Metropolis–Hastings sampling for* Θ_x^i. Define $\Theta_{-j}^i = \{\Theta_{y_1}^i, \dots, \Theta_{j-1}^i, \Theta_{j+1}^i, \dots, \Theta_{y_Y}^i\}$. The posterior distribution of Θ_x^i for $i \in \{\text{pf, rest}\}$ and

$y_1 \leq x \leq y_Y$ is given by

$$f(\Theta^i_x | \Lambda \setminus \{\Theta^i_x\}) \propto f(\boldsymbol{D} | \boldsymbol{E}, \boldsymbol{\Theta}^{\mathrm{pf}}, \boldsymbol{\Theta}^{\mathrm{rest}}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}) f(\boldsymbol{\Theta}^i | \rho_i, \sigma_i^2) \qquad \text{(A9)}$$

$$\propto \prod_{t \in \mathcal{S}} \left(e^{-E^i_{t,x} \exp[\alpha_x + \beta_x \kappa_t] \Theta^i_x} (\Theta^i_x)^{D^i_{t,x}} \right)^{I^i_{t,x}}$$

$$\times f(\Theta^i_x | \boldsymbol{\Theta}^i_{-x}, \rho_i, \sigma_i^2).$$

In this last equation, we can simplify $f(\Theta^i_x | \boldsymbol{\Theta}^i_{-x}, \rho_i, \sigma_i^2)$ for different x:

- for $x = y_1$:

$$f(\Theta^i_x | \boldsymbol{\Theta}^i_{-x}, \rho_i, \sigma_i^2) \propto f(\Theta^i_x | \rho_i, \sigma_i^2) \cdot f(\Theta^i_{x+1} | \Theta^i_x, \rho_i, \sigma_i^2)$$

$$= \frac{1}{\Theta^i_x \sqrt{2\pi\sigma_i^2}} \cdot \exp\left[-\frac{(\ln \Theta^i_x + \frac{1}{2}\sigma_i^2)^2}{2\sigma_i^2} \right]$$

$$\times \frac{1}{\Theta^i_{x+1}\sqrt{2\pi\sigma_i^2(1-\rho_i^2)}} \cdot \exp\left[-\frac{(\ln \Theta^i_{x+1} + \frac{1}{2}\sigma_i^2(1-\rho_i) - \rho_i \ln \Theta^i_x)^2}{2\sigma_i^2(1-\rho_i^2)} \right]$$

$$\propto \frac{1}{\Theta^i_x \sqrt{2\pi\sigma_i^2(1-\rho_i^2)}} \cdot \exp\left[-\frac{1}{2\sigma_i^2(1-\rho_i^2)} \left(\ln \Theta^i_x + \frac{1}{2}\sigma_i^2 - \rho_i(\ln \Theta^i_{x+1} + \frac{1}{2}\sigma_i^2) \right)^2 \right]$$

$$\sim \mathrm{logN}\left(-\frac{1}{2}\sigma_i^2 + \rho_i(\ln \Theta^i_{x+1} + \frac{1}{2}\sigma_i^2), \sigma_i^2(1-\rho_i^2) \right),$$

- for $y_1 < x < y_Y$:

$$f(\Theta^i_x | \boldsymbol{\Theta}^i_{-x}, \rho_i, \sigma_i^2) \propto f(\Theta^i_x | \Theta^i_{x-1}, \rho_i, \sigma_i^2) \cdot f(\Theta^i_{x+1} | \Theta^i_x, \rho_i, \sigma_i^2)$$

$$\sim \mathrm{logN}\left(-\frac{1}{2}\sigma_i^2 + \frac{\rho_i}{1+\rho_i^2} \left(\ln \Theta^i_{x-1} + \ln \Theta^i_{x+1} + \sigma_i^2 \right), \sigma_i^2 \frac{(1-\rho_i^2)}{(1+\rho_i^2)} \right),$$

- for $x = y_Y$:

$$f(\Theta^i_x | \boldsymbol{\Theta}^i_{-x}, \rho_i, \sigma_i^2) \propto f(\Theta^i_x | \Theta^i_{x-1}, \rho_i, \sigma_i^2)$$

$$\sim \mathrm{logN}\left(-\frac{1}{2}\sigma_i^2 + \rho_i(\ln \Theta^i_{x-1} + \frac{1}{2}\sigma_i^2), \sigma_i^2(1-\rho_i^2) \right).$$

Given a current $\widetilde{\Theta}^i_x$ and Metropolis–Hastings sampling variance $s^2_{\Theta^i_x}$, we draw a proposal $\widehat{\Theta}^i_x$ from the distribution $\ln \widehat{\Theta}^i_x \sim N(\ln \widetilde{\Theta}^i_x - \frac{1}{2}s^2_{\Theta^i_x}, s^2_{\Theta^i_x})$. Using this proposal distribution ensures that $\mathbb{E}[\widehat{\Theta}^i_x] = \exp[\ln \widetilde{\Theta}^i_x - \frac{1}{2}s^2_{\Theta^i_x} + \frac{1}{2}s^2_{\Theta^i_x}] = \widetilde{\Theta}^i_x$. The proposal distribution is not symmetric and the acceptance probability is thus given by

$$\phi = \min\left\{ \frac{f(\widehat{\Theta}^i_x | \Lambda \setminus \{\widehat{\Theta}^i_x\})}{f(\widetilde{\Theta}^i_x | \Lambda \setminus \{\widetilde{\Theta}^i_x\})} \cdot \frac{g(\widetilde{\Theta}^i_x | \widehat{\Theta}^i_x)}{g(\widehat{\Theta}^i_x | \widetilde{\Theta}^i_x)}; 1 \right\}.$$

Here, $g(.|\Theta_x)$ is the lognormal density which gives the logarithm of the stochastic variable mean $\ln \Theta_x - \frac{1}{2}s^2_{\Theta_x}$ and variance $s^2_{\Theta_x}$.

A.4.2. *Metropolis–Hastings sampling for ρ_i.* The posterior distribution of ρ_i for $i \in$ {pf, rest} is given by

$$f(\rho_i|\mathbf{\Lambda}\backslash\{\rho_i\}) \propto f(\mathbf{\Theta}^i|\sigma_i^2, \rho_i) \cdot f(\rho_i) \tag{A10}$$

$$= \frac{1}{(2\pi)^{Y/2}\Theta_{y_1}^i \cdots \Theta_{y_Y}^i \cdot |\mathbf{\Sigma}_i|^{1/2}}$$

$$\times \exp\left[-\frac{1}{2}(\ln\mathbf{\Theta}^i - \boldsymbol{\mu}_i)\mathbf{\Sigma}_i^{-1}(\ln\mathbf{\Theta}^i - \boldsymbol{\mu}_i)'\right]$$

$$\times \frac{1}{\sqrt{2\pi\sigma_{\rho_i}^2}} \cdot \exp\left[-\frac{(\text{logit}\,(\rho_i) - \mu_{\rho_i})^2}{2\sigma_{\rho_i}^2}\right] \cdot \frac{1}{\rho_i(1-\rho_i)}$$

$$\propto \frac{1}{\rho_i(1-\rho_i)(1-\rho_i^2)^{\frac{Y-1}{2}}}$$

$$\times \exp\left[-\frac{a_\rho^i}{2\sigma_i^2(1-\rho_i^2)}\left(\rho_i - \frac{b_\rho^i}{a_\rho^i}\right)^2\right] \cdot \exp\left[-\frac{(\text{logit}\,(\rho_i) - \mu_{\rho_i})^2}{2\sigma_{\rho_i}^2}\right],$$

with $a_\rho^i = \sum_{x=y_1+1}^{y_Y-1}(\Psi_x^i)^2$ and $b_\rho^i = \sum_{x=y_1+1}^{y_Y}\Psi_{x-1}^i\Psi_x^i$. This final expression will be used in the Metropolis–Hastings sampling algorithm. For a given current value $\tilde{\rho}_i$ and Metropolis–Hastings sampling variance $s_{\rho_i}^2$, we draw a proposal $\hat{\rho}_i$ from the distribution $\hat{\rho}_i \sim \text{TN}(\tilde{\rho}_i, s_{\rho_i}^2|0, 1)$, with $\text{TN}(a, b|c, d)$ a truncated normal distribution with mean a, variance b, lower and upper bound c and d, respectively. We use the truncated normal distribution to ensure the proposal is between 0 and 1. The proposal distribution is not symmetric and the acceptance probability is thus given by

$$\phi = \min\left\{\frac{f(\hat{\rho}_i^2|\mathbf{\Lambda}\backslash\{\hat{\rho}_i^2\})}{f(\tilde{\rho}_i^2|\mathbf{\Lambda}\backslash\{\tilde{\rho}_i^2\})} \cdot \frac{g(\tilde{\rho}_i^2|\hat{\rho}_i^2)}{g(\hat{\rho}_i^2|\tilde{\rho}_i^2)}; 1\right\},$$

where g is the density for the truncated normal distribution as described above.

A.4.3. *Metropolis–Hastings sampling for σ_i^2.* The posterior distribution of σ_i^2 is given by

$$f(\sigma_i^2|\mathbf{\Lambda}\backslash\{\sigma_i^2\}) \propto f(\mathbf{\Theta}^i|\sigma_i^2, \rho_i) \cdot f(\sigma_i^2) \tag{A11}$$

$$= \frac{1}{(2\pi)^{Y/2}\Theta_{y_1}^i \cdots \Theta_{y_Y}^i|\mathbf{\Sigma}_i|^{1/2}} \cdot \exp\left[-\frac{1}{2}(\ln\mathbf{\Theta}^i - \boldsymbol{\mu}_i)'\mathbf{\Sigma}_i^{-1}(\ln\mathbf{\Theta}^i - \boldsymbol{\mu}_i)\right]$$

$$\times \sigma_i^{-1}$$

$$\propto \frac{1}{\sigma_i^{Y+1}} \cdot \exp\left[-\sigma_i^{-2}\frac{1}{2}(\ln\mathbf{\Theta}^i - \boldsymbol{\mu}_i)'\mathbf{\Gamma}^{-1}(\rho_i)(\ln\mathbf{\Theta}^i - \boldsymbol{\mu}_i)\right].$$

We use the final expression for the Metropolis–Hastings sampling algorithm. Given a current value $\tilde{\sigma}_i^2$ and Metropolis–Hastings sampling variance $s_{\sigma_i^2}^2$, we draw a new candidate $\hat{\sigma}_i^2$ from the proposal distribution $\ln\hat{\sigma}_i^2 \sim \text{N}(\ln\tilde{\sigma}_i^2 - \frac{1}{2}s_{\sigma_i^2}^2, s_{\sigma_i^2}^2)$. The proposal distribution is not

symmetric and the acceptance probability is thus given by

$$\phi = \min \left\{ \frac{f(\hat{\sigma}_i^2 | \mathbf{\Lambda} \backslash \{\hat{\sigma}_i^2\})}{f(\tilde{\sigma}_i^2 | \mathbf{\Lambda} \backslash \{\tilde{\sigma}_i^2\})} \cdot \frac{g(\tilde{\sigma}_i^2 | \hat{\sigma}_i^2)}{g(\hat{\sigma}_i^2 | \tilde{\sigma}_i^2)}; 1 \right\}.$$

Here, $g(.|\sigma_i^2)$ is the lognormal density which gives the logarithm of the stochastic variable mean $\ln \sigma_i^2 - \frac{1}{2} s_{\sigma_i^2}^2$ and variance $s_{\sigma_i^2}^2$.