Network for Studies on Pensions, Aging and Retirement

NETSPAR ACADEMIC SERIES

A Bayesian Joint Model for Population and Portfolio-Specific Mortality

Frank van Berkum, Katrien Antonio and Michel Vellekoop

DP 11/2015-034

A Bayesian joint model for population and portfolio-specific mortality

Frank van Berkum* [1], Katrien Antonio[1,2], and Michel Vellekoop[1]

[1]Faculty of Economics and Business, University of Amsterdam, The Netherlands.
[2]Faculty of Economics and Business, KU Leuven, Belgium.

This version† 17 November 2015

Abstract

Insurers and pension funds must value liabilities using mortality rates that are appropriate for their portfolio. Current practice is to multiply available projections of population mortality with portfolio-specific factors, which are often determined using Generalised Linear Models. Alternatively, one of the well-known stochastic mortality models can directly be applied to portfolio data to construct portfolio-specific projections without the use of population data. However, this requires a sufficiently large historical dataset for the portfolio, which is often not available.

We overcome this problem by introducing a model to estimate portfolio-specific mortality simultaneously with population mortality. We use a Bayesian framework, which automatically generates the appropriate weighting of the limited statistical information for a given portfolio and the more extensive information that is available for the whole population. It also allows us to incorporate parameter uncertainty when projecting portfolio-specific mortality rates.

We apply our method to a dataset of assured lives in England & Wales. We find that uncertainty in portfolio-specific factors can be significant, and that confidence intervals for portfolio-specific mortality projections are slightly wider than those resulting from frequentist projections.

Key words: Bayesian analysis, portfolio-specific mortality, mortality experience, random effects, smoothing prior, simultaneous modelling

*Corresponding author. Email: f.vanberkum@uva.nl
†This version is similar to the version submitted to ASTIN Bulletin.

1 Introduction

Life insurers 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 amongst others by differences in socioeconomic classes, see Villegas and Haberman (2014). Lantz et al. (1998) argue that individuals with higher education tend to live more healthily which might explain these differences in mortality.

Heterogeneity in mortality also exists between individuals with different motivation for buying 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) discusses the presence of select mortality. If individuals are subject to medical selection when starting a life insurance policy, policyholders with a longer duration since policy issue may experience higher mortality than policyholders recently accepted. Therefore, an insurance company or pension fund generally can not use population mortality rates without making any adjustments to the country-specific mortality projections. Barrieu et al. (2012) define this as basis risk.

In current practice, portfolio-specific mortality rates are often constructed by multiplying projections of country-specific mortality rates with portfolio-specific factors. These portfolio-specific factors, also called experience factors, represent the relative difference between the mortality rates of the population and the portfolio under consideration. For Solvency II insurance companies must derive portfolio-specific mortality rate projections and analyse the uncertainty associated within these projections, and The Dutch Bank has published a guideline on how pension funds in the Netherlands can derive their portfolio-specific factors which are to be applied relative to population mortality projections, see The Dutch Bank (2012).

Lee and Carter (1992) introduce the seminal stochastic mortality model known as the Lee-Carter model. It is a single-factor model, meaning that there is one time series involved. Many generalisations and extensions of the Lee-Carter model have been proposed since then. Renshaw and Haberman (2006) introduce a cohort effect, Cairns et al. (2006) introduce a two-factor model specifically designed for higher ages, and Plat (2009*a*) introduces age-specific improvements. Extensions using information from other populations are made by Li and Lee (2005), Dowd et al. (2011), Jarner and Kryger (2011) and Antonio et al. (2015). Advantages of these multiple population models is that the inclusion of mortality from other populations leads to a more stable mortality trend which results in projections that are more robust with respect to the calibration period. In some countries these multiple population models are now used for forecasting of population mortality, see for example Koninklijk Actuarieel Genootschap (2014) and Institute of Actuaries in Belgium (2015).

Multiple population models are also used for modelling portfolio-specific mortality. For example, Cairns et al. (2011) use an age-period-cohort model in a Bayesian framework with two populations of different size (e.g. a population and a large pension portfolio), Villegas and Haberman (2014) model five different socioeconomic classes in England & Wales relative to the total population of England & Wales, and Danesi et al. (2015) use a multiple population approach to model mortality in different regions in Italy. Such models can be used to obtain stochastic projections for the portfolio, but sufficient historical

portfolio data is needed for plausible time series projections which is often not the case for portfolios of pension funds and insurers.

Another stream of research puts focus on age-dependent relative factors with respect to population mortality, which is assumed given or a deterministic baseline mortality rate is estimated. Plat (2009*b*) determines the realised portfolio-specific factors and models these directly with a linear regression model, whereas Gschlössl et al. (2011) and Olivieri (2011) use a Poisson framework to model the number of deaths in an insured portfolio. Tomas and Planchet (2013) also use a Poisson likelihood to model mortality of long-term care claimants using duration-dependent factors. These models can be used to explain historical portfolio mortality, but projections of portfolio-specific mortality rates that include uncertainty both in the mortality trend and in the level cannot be obtained in a natural way. Further, frailty models that use parametric mortality laws like Gompertz, Makeham and variations thereof are investigated by Butt and Haberman (2004) and Richards et al. (2013), but mortality projections can not be obtained using such frailty models.

We propose a model to estimate population and portfolio-specific mortality simultaneously. To account for yearly fluctuations in small portfolios we use a Poissonian likelihood assumption, as introduced by Brouhns et al. (2002). We view the portfolio as part of the population and model a baseline mortality trend in the population. By using a larger dataset for the population than we have available for the portfolio, we are able to adequately estimate the dynamics of the mortality trend. The relative difference between the population and the portfolio is modelled using a random effect. Such random effects can be used to reflect remaining heterogeneity among policyholders which is not captured by the observable risk factors. See Denuit et al. (2005) and Antonio and Zhang (2014) for examples in pricing models for non-life insurance. We define the difference between the population and the portfolio as the 'rest', and use this to model the total number of deaths in the population for each calendar year and each age. We use the Lee-Carter model to estimate population mortality and approach this in a Bayesian way. Age-dependent portfolio-specific factors are estimated relative to population mortality. We consider two prior distributions for the portfolio-specific factors: a Gamma prior which implies independence between different ages and portfolios (our portfolio and the rest), and a logNormal prior which implies dependence between ages but independence between portfolios. Through the Bayesian approach we can model population and portfolio-specific mortality simultaneously in contrast to a multistep approach that is often used in a frequentist approach. Further, the Bayesian approach provides insight concerning the parameter uncertainty in the stochastic population mortality model, in the time series model used in the stochastic population mortality model, and in the portfolio-specific factors. As a result, we are able to project portfolio-specific mortality and assess the uncertainty in these projections in a natural way.

Section 2 gives an overview of existing approaches to modelling portfolio-specific mortality and introduces our model. In Section 3 we describe the prior distributions used in our Bayesian setting, and in Section 4 we derive the posterior distributions of all parameters. Section 5 contains an illustration of our proposed model using a dataset on assured male lives from England & Wales, and Section 6 concludes.

2 Bayesian portfolio-specific mortality

2.1 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 $\mu_{t,x}$ on the interval $[t, t+1) \times [x, x+1)$, the maximum likelihood estimate $\hat{\mu}_{t,x}$ of the force of mortality $\mu_{t,x}$ equals the observed death rate $m_{t,x}$ (Pitacco et al. (2009)). 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.

Lee and Carter (1992) introduce the seminal mortality model to explain observed death rates. Their model contains a single period effect and is therefore called a single-factor model:

$$\ln m_{t,x} = \alpha_x + \beta_x \kappa_t + \epsilon_{t,x}, \qquad \epsilon_{t,x} \overset{\text{iid}}{\sim} (0, \sigma^2). \tag{1}$$

They estimate this model using a Singular Value Decomposition (SVD), and they model the period effect κ_t using a random walk with drift to obtain mortality projections:

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

Brouhns et al. (2002) investigate this model in a Poisson framework, thereby accounting for sampling randomness:

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

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. (2014).

Multiple population mortality models. Mortality developments in a single country can be strongly time-varying. Periods of low mortality improvements may be followed by periods of high mortality improvements. A rapidly changing mortality trend is complicated to project. Therefore, extensions to the Lee-Carter model have been proposed to incorporate information from different but comparable countries to estimate a more stable, global mortality trend, which allows insight in country-specific deviations from this trend. The approach of multiple population modelling has clear advantages, but a disadvantage is that sufficient historical data is necessary to model the country-specific deviations. If there is only limited historical data available for a portfolio, applying the multiple population approach to portfolio data is less appropriate. Below we discuss some approaches for multiple population modelling to illustrate ways to model a mortality trend in a reference population.

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

$$\ln m^i_{t,x} = \alpha^i_x + B_x K_t + \beta^i_x \kappa^i_t + \epsilon_{t,x,i}, \qquad \epsilon_{t,x,i} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_i^2). \tag{4}$$

The term $B_x K_t$ represents the common factor for the different countries considered, α_x^i is defined as the average mortality in country i over time, and the term $\beta_x^i \kappa_t^i$ is a country-specific, age-dependent mortality development. Li and Lee (2005) estimate this model using SVD, whereas Antonio et al. (2015) investigate this model and variations thereof in a Bayesian setting with a Poissonian likelihood. These models are designed mainly for simultaneous estimation of mortality in different countries, see Koninklijk Actuarieel Genootschap (2014) and Institute of Actuaries in Belgium (2015).

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

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

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. Dowd et al. (2011) model the time series of the subordinate population as a spread relative to the time series of the dominant population. Cairns et al. (2011) consider this model in a Bayesian setting. By specifying the dependence between the two populations slightly differently, their specification can be used for a combination of a dominant and subordinate population and for a combination of two equal-sized populations. These models can for example be used for mortality in different countries or 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 modelled using an extension of the Lee-Carter model:

$$\ln \mu_{t,x} = \alpha_x + \beta_x \kappa_t + \gamma_{t-x}. \tag{6}$$

Mortality for socioeconomic class g is modelled relative to the population

$$\ln {}_n\mu_{t,x,g} = \ln {}_n\mu_{t,x} + \alpha_{x,g} + \beta_x \kappa_{t,g}, \tag{7}$$

where

$${}_n\mu_{t,x} = \left(\prod_{i=0}^{n-1} \mu_{t,x+i}\right)^{\frac{1}{n}} \tag{8}$$

is the geometric average of the forces of mortality in the reference population between age x and age $x + n - 1$. This model is specifically designed for subpopulations within a larger population, where the larger population is for example a country.

Portfolio-specific mortality models. Alongside the multiple population approach, other models have been suggested, which estimate portfolio-specific factors relative to a population. In these models population mortality is often assumed given or a (smooth) baseline mortality rate is estimated beforehand. These models can be used to explain historical observations, but are less appropriate for projection purposes because population and portfolio-specific mortality are not estimated simultaneously. Below we discuss several approaches to modelling portfolio-specific factors. In Section 2.2 we combine these ideas with the multiple population approach and introduce a new model to simultaneously estimate population and portfolio-specific mortality.

Plat (2009*b*) focusses on realised portfolio-specific factors defined by

$$P_{t,x} = \frac{q_{t,x}^A}{q_{t,x}^{\text{pop}}}, \tag{9}$$

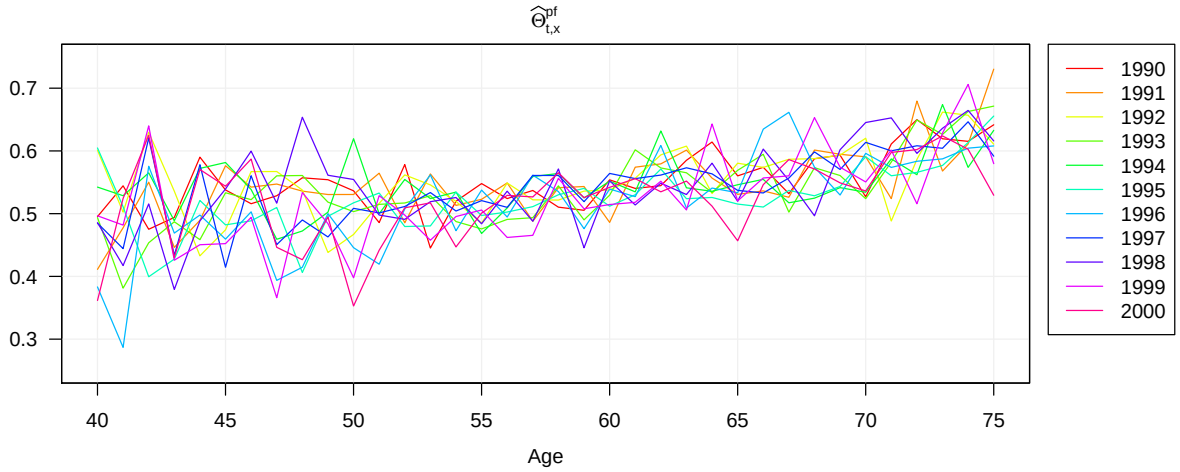

Figure 1: Observed portfolio-specific factors computed as $\widehat{\Theta}^{\text{pf}}_{t,x} = m^{\text{pf}}_{t,x}/m^{\text{pop}}_{t,x}$ for the CMI dataset on assured male lives in England & Wales (see Section 5). (Coloured versions of the figures can be found online.)

where $q^{A}_{t,x}$ is the realised mortality rate in the portfolio based on insured amounts, and $q^{\text{pop}}_{t,x}$ is the realised mortality rate in the population. Several models are suggested for modelling the portfolio-specific factors $P_{t,x}$. As an example, Plat (2009*b*) estimates a linear trend through ages on fourteen years of data

$$P_{t,x} = a_t + b_t x + \varepsilon_{t,x}, \qquad \varepsilon_{t,x} \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2_\varepsilon). \tag{10}$$

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.

Figure 1 shows observed portfolio-specific factors $\widehat{\Theta}^{\text{pf}}_{t,x}$ (computed as $m^{\text{pf}}_{t,x}/m^{\text{pop}}_{t,x}$) for the CMI dataset on assured lives in England & Wales (see Section 5). Within a calendar year the observations are very volatile; these factors jump up and down from one age to the other. This is not only uncertainty in the portfolio-specific factors, but also Poisson or Binomial volatility from random deaths given a fixed mortality rate. In order to appropriately model the volatility due to random deaths, we will directly model observed deaths instead of observed portfolio-specific factors.

Using only five years of historical data, Gschlössl et al. (2011) do not include time dynamics. First they estimate a baseline force of mortality on portfolio data for each age using 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 five years of historical portfolio data on individual lives. Their model is designed for data on individual lives, and this approach therefore can not be used when only aggregated portfolio data is available.

Olivieri (2011) considers a Bayesian setting of the form

$$D_{t,x} \sim \text{Poisson}(E_{t,x} q^{*}_{t,x} Z_{t,x}), \tag{11}$$

where $q^{*}_{t,x}$ is a best estimate mortality rate published by an independent institution, and $Z_{t,x} \sim \text{Gamma}(\alpha_{t,x}, \beta_{t,x})$ is a random adjustment to the best estimate mortality rate. Starting with values for $\alpha_{0,x}$ and $\beta_{0,x}$, subsequent values of $\alpha_{t,x}$ and $\beta_{t,x}$ can analytically be

computed as new mortality observations becomes available since the Gamma distribution is the conjugate of the Poisson distribution. Kan (2012) considers a similar framework, but estimates a population mortality using the Lee-Carter model on population mortality data.

2.2 Model formulation, likelihood specification and parameter constraints

As Section 2.1 illustrates, different approaches to modelling portfolio-specific factors exist, each approach suitable for different types of datasets. We consider the situation where only limited historical portfolio data is available which hinders reliable estimation of a mortality trend in the portfolio. To obtain projections of mortality rates specific to a portfolio, we model portfolio-specific factors relative to population mortality, and we use the Lee-Carter model to project population mortality. This way, we ensure the portfolio-specific factors are consistent with population mortality projections.

Mortality observations. Let the number of deaths for group i during calendar year t aged x at death be $d^i_{t,x}$, and the exposure for group i aged x during calendar year t be $E^i_{t,x}$. We consider the entire population of a country ('pop'), the portfolio under investigation ('pf'), and the difference between the entire population and the portfolio under consideration (hereafter referred to as the 'rest'), thus $i \in \{\text{pop}, \text{pf}, \text{rest}\}$. We assume that the observed portfolio and the rest sum to the total population: $d^{\text{pf}}_{t,x} + d^{\text{rest}}_{t,x} = d^{\text{pop}}_{t,x}$ and $E^{\text{pf}}_{t,x} + E^{\text{rest}}_{t,x} = E^{\text{pop}}_{t,x}$. We need to define the rest to ensure we consider all information available in the population.

To estimate parameters in our model, we extend the dataset with observations of the total population. As such, we obtain a larger dataset which enables the simultaneous estimation of population mortality and portfolio-specific mortality. We define the set of cells (t, x) for which we have observations about both our portfolio and the rest as $\mathcal{O}^{\text{pf}}$ (the red cells in Figure 2), and as a result we can only measure the heterogeneity between the portfolio and the rest on this set. The set for which we have observations about the population but not separately about our portfolio and the rest is defined by $\mathcal{O}^{\text{pop}}$ (the green cells in Figure 2).

For the portfolio we have observations for $(t, x) \in \mathcal{S} \times \mathcal{Y}$ with $\mathcal{S} = \{s_1, s_1 + 1, \ldots, s_S\}$ and $\mathcal{Y} = \{y_1, y_1 + 1 \ldots, y_Y\}$. For the population we have observations for cells $(t, x) \in \mathcal{T} \times \mathcal{X}$ with $\mathcal{T} = \{t_1, t_1+1 \ldots, t_T\}$ and $\mathcal{X} = \{x_1, x_1+1 \ldots, x_X\}$ with $t_1 \leq s_1 \leq s_S \leq t_T$ and $x_1 \leq y_1 \leq y_Y \leq x_X$ (see Figure 2), but we include only observations for the population such that $\mathcal{O}^{\text{pop}} \cap \mathcal{O}^{\text{pf}} = \emptyset$. In the dataset of the portfolio we have Y ages and S years, whereas in the population we have X ages and T years. We introduce indicator variables that are useful when working with likelihoods:

$$I^{\text{pf}}_{t,x} = I^{\text{rest}}_{t,x} = \begin{cases} 1 & \text{if } (t,x) \in \mathcal{O}^{\text{pf}} \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

and

$$I^{\text{pop}}_{t,x} = \begin{cases} 1 & \text{if } (t,x) \in \mathcal{O}^{\text{pop}} \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

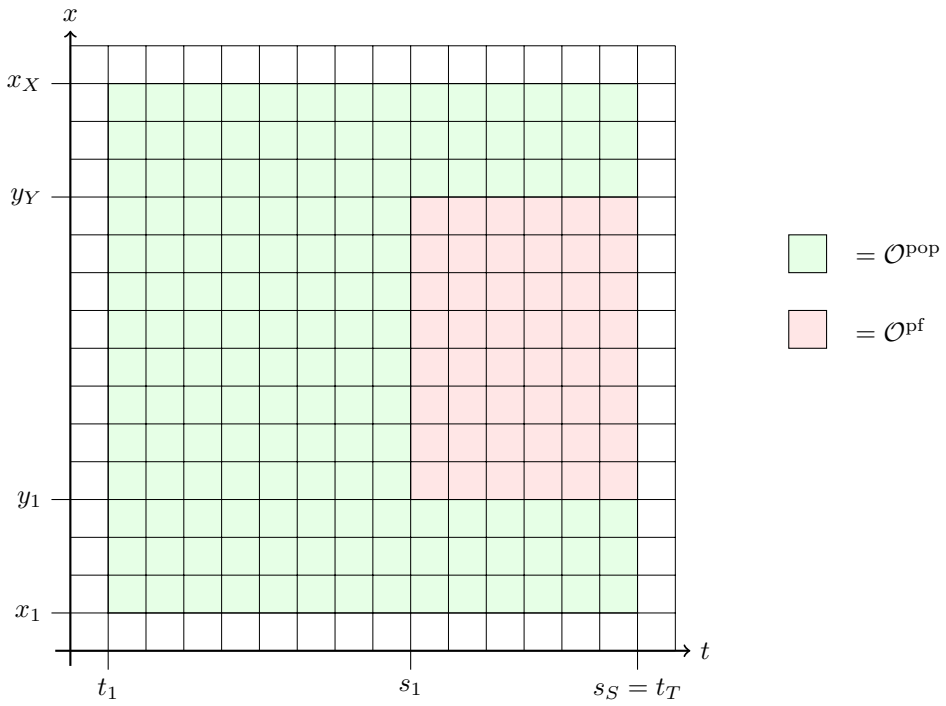

Figure 2: Illustration of overlap between portfolio and population observations. (Coloured versions of the figures can be found online.)

Model formulation and likelihood specification. We assume that all groups in the population (our own portfolio and the rest) share a baseline force of mortality defined by $\mu_{t,x}$; we model this force of mortality using the Lee-Carter model as in Equation (3). However, since there is heterogeneity between groups, mortality in a group will deviate from the baseline force of mortality. This heterogeneity is captured by a random effect that depends on age, and the effects are different for our own portfolio and for the rest. The model is specified as:

$$D^{\text{pop}}_{t,x} | \mu_{t,x} \sim \text{Poisson}(E^{\text{pop}}_{t,x} \mu_{t,x}), \quad \text{for } (t,x) \in \mathcal{O}^{\text{pop}} \tag{14}$$

and

$$D^{\text{pf}}_{t,x} | (\mu_{t,x}, \Theta^{\text{pf}}_x) \sim \text{Poisson}(E^{\text{pf}}_{t,x} \mu_{t,x} \Theta^{\text{pf}}_x), \quad \text{for } (t,x) \in \mathcal{O}^{\text{pf}} \tag{15}$$

$$D^{\text{rest}}_{t,x} | (\mu_{t,x}, \Theta^{\text{rest}}_x) \sim \text{Poisson}(E^{\text{rest}}_{t,x} \mu_{t,x} \Theta^{\text{rest}}_x), \tag{16}$$

with

$$\mu_{t,x} = \exp[\alpha_x + \beta_x \kappa_t]. \tag{17}$$

By specifying the model this way, we include all deaths in the population for every cell (t,x), either by directly modelling $D^{\text{pop}}_{t,x}$ or by modelling both $D^{\text{pf}}_{t,x}$ and $D^{\text{rest}}_{t,x}$.

The (in)dependence assumptions are summarised as follows. All groups in a country share a baseline force of mortality that is modelled by $\mu_{t,x}$. The random effects Θ^i_x are independent between groups, but there might be dependence between ages. In Section 3 we consider two prior specifications for Θ^i_x, a Gamma prior and a logNormal prior. The first

implies independence between ages and between groups, the second implies dependence between ages but independence between groups. Given the baseline force of mortality $\mu_{t,x}$ and the portfolio-specific factors Θ_x^i the 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[1]. As discussed in van Berkum et al. (2014) we believe a difference stationary model to be more appropriate for modelling the period effect for a single country and that is what we use in this paper. In the next section we describe the prior distributions that we use for the parameters and hyperparameters to complete the Bayesian specification of the model.

Parameter constraints. In a frequentist setting, parameter constraints are needed to uniquely identify the Lee-Carter model. Linear transformations can be applied to the parameters α_x, β_x and κ_t without changing the fitted mortality rates and thus without changing the likelihood of the model.

We consider the model in a Bayesian framework which means that we approach all parameters as random variables and specify a prior for them. Applying linear transformations to the parameters α_x, β_x and κ_t leads to a different joint posterior distribution because the likelihood of the parameters given their prior distribution changes after a transformation.

However, the prior distribution of the parameters has little effect on the posterior distribution compared with the impact of the likelihood of the observations, and as a result the MCMC procedure may not converge. To facilitate convergence in the Bayesian setting, we apply two parameter constraints:

$$\kappa_{t_1} = 0 \qquad \text{and} \qquad |\boldsymbol{\beta}|^2 = \sum_{x \in \mathcal{X}} \beta_x^2 = 1. \tag{18}$$

The first constraint is applied through the model specification of the time series by fixing the starting point of the time series. Applying the second constraint can not be performed without changing the likelihood of the data, but Bafumi et al. (2005) suggest normalising parameters after estimation is completed. Therefore, we normalise the β_x's after all simulations have been performed and we apply corresponding transformations to the κ_t's and the hyperparameters δ and σ_ε^2 to ensure the posterior distribution of κ_t does not change.

3 Prior distributions

3.1 Age parameters for population mortality

Following Czado et al. (2005) and Antonio et al. (2015) we use the following prior for α_x:

$$e_x = \exp(\alpha_x) \overset{\text{iid}}{\sim} \text{Gamma}(a_x, b_x). \tag{19}$$

[1] The difference stationary model is also known as a random walk (possibly with a drift).

Our prior distribution for β_x is inspired by Antonio et al. (2015):

$$\beta_x \overset{\text{iid}}{\sim} \text{N}(\mu_\beta, \sigma_\beta^2). \tag{20}$$

However, we assume the mean of the β_x's is a hyperparameter with the following prior

$$\mu_\beta \sim \text{N}(\nu_\beta, c_\beta^2). \tag{21}$$

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 the following prior for the variance hyperparameter:

$$\sigma_\beta \sim \text{Uniform}(0, A_\beta), \tag{22}$$

which implies

$$\begin{aligned} f_{\sigma_\beta^2}(\sigma^2) &\propto \sigma^{-1} \quad \text{for} \quad 0 \leq \sigma^2 \leq A_\beta^2 \\ &= \sigma^{-1} \cdot 1_{[0 \leq \sigma \leq A_\beta]}. \end{aligned} \tag{23}$$

3.2 Period parameters for population mortality

In line with van Berkum et al. (2014) we consider a random walk with drift for the period effect κ_t. The prior distribution is specified by

$$\kappa_t = \kappa_{t-1} + \delta + \varepsilon_t, \quad \text{with} \quad \kappa_{t_1} = 0 \quad \text{and} \quad \varepsilon_t \overset{\text{iid}}{\sim} \text{N}(0, \sigma_\varepsilon^2) \quad \text{for} \quad t > t_1. \tag{24}$$

For the hyperparameter δ we use the prior

$$\delta \sim \text{N}(\mu_\delta, \sigma_\delta^2), \tag{25}$$

and for the hyperparameter σ_ε^2 we use the prior

$$\sigma_\varepsilon \sim \text{Uniform}(0, A_\varepsilon). \tag{26}$$

3.3 Portfolio-specific mortality

The portfolio-specific factors Θ_x^i represent the relative difference between mortality in the population and in group i with $i \in \{\text{pf}, \text{rest}\}$. We do not want to make prior assumptions on whether mortality in a group is higher or lower than in the population. Therefore, we choose our prior distribution such that $\mathbb{E}(\Theta_x^i) = 1$ $(\forall x, \forall i)$. We consider two prior distributions for Θ_x^i, a Gamma prior and a lognormal prior.

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

$$\Theta_x^i \sim \text{Gamma}(c_x^i, c_x^i), \quad \text{for } y_1 \leq x \leq y_Y. \tag{27}$$

The resulting age-dependent and group-specific factors are independent over ages x and between groups i. By choosing equal parameters for the Gamma distribution we ensure $\mathbb{E}(\Theta_x^i) = 1$ $(\forall x, \forall i)$, and the variance of the prior distribution can be set by choosing c_x^i accordingly.

Lognormal prior The lognormal prior on the age-dependent factors for group i is given by

$$\ln \Theta^i_x = \mu_i + \rho_i \ln \Theta^i_{x-1} + \eta^i_x, \quad \text{with } \eta^i_x \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_i^2(1-\rho_i^2)) \quad \text{for } y_1 < x \leq y_Y, \qquad (28)$$
$$\text{and } \ln \Theta^i_{y_1} \overset{\text{iid}}{\sim} \mathrm{N}(-\tfrac{1}{2}\sigma_i^2, \sigma_i^2),$$

where $\ln \Theta^i_{y_1}$ and η^i_x are independent. This process is a mean reverting process over the log of the group-specific factors, through which we achieve dependence between group-specific mortality factors over ages x, but not between groups i. The mean parameter to which the factors revert is defined as $\mu_i = -\frac{1}{2}(1-\rho_i)\sigma_i^2$; choosing it as such ensures $\mathbb{E}(\Theta^i_x) = 1$ $(\forall x, \forall i)$, see Purcaru et al. (2004).

Define $\boldsymbol{\Theta}^i = \{\Theta^i_{y_1}, \ldots, \Theta^i_{y_Y}\}$. The mean reverting process on the log of the portfolio-specific factors can also be written as a multivariate logNormal distribution (Purcaru et al., 2004, Section 3.3.2)

$$\boldsymbol{\Theta}^i \sim \text{logNormal}(\tilde{\boldsymbol{\mu}}_i, \boldsymbol{\Sigma}_i), \qquad (29)$$

with $\tilde{\boldsymbol{\mu}}_i = -\frac{1}{2}\sigma_i^2 \mathbf{1}_Y$ and $(\boldsymbol{\Sigma}_i)_{xy} = \rho_i^{|x-y|}\sigma_i^2$. For the variance hyperparameter we use the prior

$$\sigma_i \sim \text{Uniform}(0, A_i), \qquad (30)$$

and for the mean reversion parameter we use the prior

$$\text{logit}\,(\rho_i) \sim \mathrm{N}(\mu_{\rho_i}, \sigma^2_{\rho_i}). \qquad (31)$$

Using this logit prior assumption ensures a mean reversion parameter in the interval $(0, 1)$.

4 Posterior distributions

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

$$\boldsymbol{D} = \{\boldsymbol{D}^{\text{pop}}, \boldsymbol{D}^{\text{pf}}, \boldsymbol{D}^{\text{rest}}\}, \qquad \boldsymbol{E} = \{\boldsymbol{E}^{\text{pop}}, \boldsymbol{E}^{\text{pf}}, \boldsymbol{E}^{\text{rest}}\},$$
$$\boldsymbol{\alpha} = \{\alpha_{x_1}, \ldots, \alpha_{x_X}\}, \qquad \boldsymbol{\beta} = \{\beta_{x_1}, \ldots, \beta_{x_X}\}, \qquad \boldsymbol{\kappa} = \{\kappa_{t_1}, \ldots, \kappa_{t_T}\},$$
$$\boldsymbol{\Theta} = \{\boldsymbol{\Theta}^{\text{pf}}, \boldsymbol{\Theta}^{\text{rest}}\}, \qquad \boldsymbol{\rho}_\theta = \{\rho_{\text{pf}}, \rho_{\text{rest}}\}, \qquad \boldsymbol{\sigma}^2_\theta = \{\sigma^2_{\text{pf}}, \sigma^2_{\text{rest}}\}.$$

We further define the set $\boldsymbol{\Lambda}$ that contains both data and parameters:

$$\boldsymbol{\Lambda} = \{\boldsymbol{D}, \boldsymbol{E}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mu_\beta, \sigma^2_\beta, \boldsymbol{\kappa}, \delta, \sigma^2_\varepsilon, \boldsymbol{\Theta}, (\ldots)\}, \qquad (32)$$

where the term between brackets is empty in case of a Gamma prior on the portfolio-specific factors, and it is $\{\boldsymbol{\rho}_\theta, \boldsymbol{\sigma}^2_\theta\}$ for the logNormal prior.

4.1 Age parameters for population mortality

4.1.1 Gibbs sampling for α_x

Since the individual α_x's are independent, we derive the posterior for a single $e_x(= \exp(\alpha_x))$ and $x_1 \leq x \leq x_X$ as follows:

$$
\begin{aligned}
f(e_x|\boldsymbol{\Lambda}\backslash\{e_x\}) &\propto f(\boldsymbol{D}|\boldsymbol{E},\boldsymbol{e},\boldsymbol{\beta},\boldsymbol{\kappa},\boldsymbol{\Theta}^{\text{pf}},\boldsymbol{\Theta}^{\text{rest}})f(e_x) \qquad (33)\\
&\propto \prod_{t\in\mathcal{T}}\left(e^{-E^{\text{pf}}_{t,x}e_x\exp[\beta_x\kappa_t]\Theta^{\text{pf}}_x}\frac{(E^{\text{pf}}_{t,x}e_x\exp[\beta_x\kappa_t]\Theta^{\text{pf}}_x)^{D^{\text{pf}}_{t,x}}}{D^{\text{pf}}_{t,x}!}\right)^{I^{\text{pf}}_{t,x}}\\
&\quad\times\prod_{t\in\mathcal{T}}\left(e^{-E^{\text{rest}}_{t,x}e_x\exp[\beta_x\kappa_t]\Theta^{\text{rest}}_x}\frac{(E^{\text{rest}}_{t,x}e_x\exp[\beta_x\kappa_t]\Theta^{\text{rest}}_x)^{D^{\text{rest}}_{t,x}}}{D^{\text{rest}}_{t,x}!}\right)^{I^{\text{rest}}_{t,x}}\\
&\quad\times\prod_{t\in\mathcal{T}}\left(e^{-E^{\text{pop}}_{t,x}e_x\exp[\beta_x\kappa_t]}\frac{(E^{\text{pop}}_{t,x}e_x\exp[\beta_x\kappa_t])^{D^{\text{pop}}_{t,x}}}{D^{\text{pop}}_{t,x}!}\right)^{I^{\text{pop}}_{t,x}}\\
&\quad\times\frac{b_x^{a_x}}{\Gamma(a_x)}e_x^{a_x-1}\exp[-b_xe_x]\\
&\propto \exp[-(b_x+d_x)e_x]\cdot e_x^{a_x+D_{\bullet x}-1},
\end{aligned}
$$

with

$$
d_x=\sum_{t\in\mathcal{T}}\left\{I^{\text{pf}}_{t,x}\left(E^{\text{pf}}_{t,x}\exp[\beta_x\kappa_t]\Theta^{\text{pf}}_x\right)+I^{\text{rest}}_{t,x}\left(E^{\text{rest}}_{t,x}\exp[\beta_x\kappa_t]\Theta^{\text{rest}}_x\right)+I^{\text{pop}}_{t,x}\left(E^{\text{pop}}_{t,x}\exp[\beta_x\kappa_t]\right)\right\}
$$

and

$$
D_{\bullet x}=\sum_{t\in\mathcal{T}}\left\{I^{\text{pf}}_{t,x}\cdot D^{\text{pf}}_{t,x}+I^{\text{rest}}_{t,x}\cdot D^{\text{rest}}_{t,x}+I^{\text{pop}}_{t,x}\cdot D^{\text{pop}}_{t,x}\right\}=\sum_{t\in\mathcal{T}}D^{\text{pop}}_{t,x}.
$$

The last line in (33) is proportional to a Gamma$(a_x+D_{\bullet x}, b_x+d_x)$ distribution. Therefore, we can use direct Gibbs sampling to draw a new value of e_x which can be transformed into a new value of α_x.

4.1.2 Metropolis sampling for β_x

Since the individual β_x's are independent, we derive the posterior for a single β_x and $x_1 \leq x \leq x_X$ as follows

$$
\begin{aligned}
f(\beta_x|\boldsymbol{\Lambda}\backslash\{\beta_x\}) &\propto f(\boldsymbol{D}|\boldsymbol{E},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\kappa},\boldsymbol{\Theta}^{\text{pf}},\boldsymbol{\Theta}^{\text{rest}},\mu_\beta,\sigma^2_\beta)f(\boldsymbol{\beta}|\mu_\beta,\sigma^2_\beta) \qquad (34)\\
&\propto \prod_{t\in\mathcal{T}}\left[\exp\left(-E^{\text{pf}}_{t,x}\exp[\alpha_x+\beta_x\kappa_t]\Theta^{\text{pf}}_x\right)\exp\left(D^{\text{pf}}_{t,x}\beta_x\kappa_t\right)\right]^{I^{\text{pf}}_{t,x}}\\
&\quad\times\prod_{t\in\mathcal{T}}\left[\exp\left(-E^{\text{rest}}_{t,x}\exp[\alpha_x+\beta_x\kappa_t]\Theta^{\text{rest}}_x\right)\exp\left(D^{\text{rest}}_{t,x}\beta_x\kappa_t\right)\right]^{I^{\text{rest}}_{t,x}}\\
&\quad\times\prod_{t\in\mathcal{T}}\left[\exp\left(-E^{\text{pop}}_{t,x}\exp[\alpha_x+\beta_x\kappa_t]\right)\exp\left(D^{\text{pop}}_{t,x}\beta_x\kappa_t\right)\right]^{I^{\text{pop}}_{t,x}}\\
&\quad\times f(\beta_x|\mu_\beta,\sigma^2_\beta).
\end{aligned}
$$

We cannot use Gibbs sampling here, thus we resort to Metropolis sampling. Given a current value $\tilde{\beta}_x$ and Metropolis sampling variance $s^2_{\beta_x}$, we draw a candidate $\hat{\beta}_x$ from the candidate distribution $\mathrm{N}(\tilde{\beta}_x, s^2_{\beta_x})$. The candidate distribution is symmetric and the acceptance probability is thus given by

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

4.1.3 Gibbs sampling for μ_β

The posterior density of μ_β is given by

$$\begin{aligned} f(\mu_\beta|\boldsymbol{\Lambda}\backslash\{\mu_\beta\}) &\propto f(\boldsymbol{\beta}|\mu_\beta, \sigma^2_\beta) f(\mu_\beta) \\ &\propto \exp\left[-\sum_{x\in\mathcal{X}} \frac{(\beta_x - \mu_\beta)^2}{2\sigma^2_\beta}\right] \cdot \exp\left[-\frac{(\mu_\beta - \nu_\beta)^2}{2c^2_\beta}\right] \\ &\propto \exp\left[-\frac{1}{2d_\beta}\left(\mu^2_\beta - 2\mu_\beta e_\beta\right)\right], \end{aligned} \tag{35}$$

with

$$d_\beta = \frac{\sigma^2_\beta}{X + \sigma^2_\beta/c^2_\beta} \quad \text{and} \quad e_\beta = \frac{X}{X + \sigma^2_\beta/c^2_\beta} \cdot \frac{1}{X}\sum_{x\in\mathcal{X}} \beta_x + \frac{\sigma^2_\beta/c^2_\beta}{X + \sigma^2_\beta/c^2_\beta} \cdot \nu_\beta.$$

The last line in (35) is proportional to a Normal distribution, so we can use Gibbs sampling to obtain a new value for μ_β from the distribution $\mathrm{N}(e_\beta, d_\beta)$.

4.1.4 Gibbs sampling for σ^2_β

The posterior of σ^2_β is given by

$$\begin{aligned} f(\sigma^2_\beta|\boldsymbol{\Lambda}\backslash\{\sigma^2_\beta\}) &\propto f(\boldsymbol{\beta}|\mu_\beta, \sigma^2_\beta) f(\sigma^2_\beta) \\ &= \frac{1}{(2\pi)^{X/2}(\sigma^2_\beta)^{X/2}} \cdot \exp\left[-\sum_{x\in\mathcal{X}} \frac{(\beta_x - \mu_\beta)^2}{2\sigma^2_\beta}\right] \\ &\quad \times \sigma_\beta^{-1} \cdot 1_{[0\le\sigma_\beta\le A_\beta]} \\ &\propto (\sigma_\beta^{-2})^{\frac{X+1}{2}} \cdot \exp\left[-(\sigma_\beta^{-2}) \cdot \frac{1}{2}\sum_{x\in\mathcal{X}} (\beta_x - \mu_\beta)^2\right]. \end{aligned} \tag{36}$$

Therefore, we know that the posterior of its reciprocal value, σ_β^{-2}, is

$$f(\sigma_\beta^{-2}|\boldsymbol{\Lambda}\backslash\{\sigma^2_\beta\}) \sim \text{Gamma}\left(\frac{X-1}{2}, \frac{1}{2}\sum_{x\in\mathcal{X}} (\beta_x - \mu_\beta)^2\right),$$

and we can use Gibbs sampling to draw a new value of σ_β^{-2} which can be transformed to σ^2_β.

4.2 Period parameters for population mortality

4.2.1 Metropolis sampling for κ_t

Define $\boldsymbol{\kappa}_{-t} = \{\kappa_{t_1}, \ldots, \kappa_{t-1}, \kappa_{t+1}, \ldots, \kappa_{t_T}\}$. The posterior distribution of κ_t for $t_1 < t \leq t_T$ is given by

$$\begin{aligned} f(\kappa_t|\boldsymbol{\Lambda}\backslash\{\kappa_t\}) &\propto f(\boldsymbol{D}|\boldsymbol{E}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{\Theta}^{\text{pf}}, \boldsymbol{\Theta}^{\text{rest}}) f(\boldsymbol{\kappa}|\kappa_{t_1}, \delta, \sigma_\varepsilon^2) \\ &\propto \prod_{x\in\mathcal{X}} \left[\exp\left(-E_{t,x}^{\text{pf}} \exp[\alpha_x + \beta_x\kappa_t]\Theta_x^{\text{pf}}\right) \exp\left(D_{t,x}^{\text{pf}}\beta_x\kappa_t\right)\right]^{I_{t,x}^{\text{pf}}} \\ &\quad \times \prod_{x\in\mathcal{X}} \left[\exp\left(-E_{t,x}^{\text{rest}} \exp[\alpha_x + \beta_x\kappa_t]\Theta_x^{\text{rest}}\right) \exp\left(D_{t,x}^{\text{rest}}\beta_x\kappa_t\right)\right]^{I_{t,x}^{\text{rest}}} \\ &\quad \times \prod_{x\in\mathcal{X}} \left[\exp\left(-E_{t,x}^{\text{pop}} \exp[\alpha_x + \beta_x\kappa_t]\right) \exp\left(D_{t,x}^{\text{pop}}\beta_x\kappa_t\right)\right]^{I_{t,x}^{\text{pop}}} \\ &\quad \times f(\kappa_t|\boldsymbol{\kappa}_{-t}, \delta, \sigma_\varepsilon^2). \end{aligned} \tag{37}$$

The expression in the last line can be simplified:

- for $t_1 < t < t_T$:

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

- for $t = t_T$:

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

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

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

4.2.2 Gibbs sampling for δ

Define $\Delta\kappa_t = \kappa_t - \kappa_{t-1}$. The posterior distribution of δ is given by

$$\begin{aligned} f(\delta|\boldsymbol{\Lambda}\backslash\{\delta\}) &\propto f(\boldsymbol{\kappa}|\kappa_1, \delta, \sigma_\varepsilon^2) f(\delta) \\ &\propto \exp\left[-\sum_{t=t_1+1}^{t_T} \frac{[\Delta\kappa_t - \delta]^2}{2\sigma_\varepsilon^2}\right] \cdot \exp\left[-\frac{[\delta - \mu_\delta]^2}{2\sigma_\delta^2}\right] \\ &\propto \exp\left[-\frac{1}{2a_\delta}\left(\delta^2 - 2\delta b_\delta\right)\right] \\ &\sim \text{N}\left(b_\delta, a_\delta\right), \end{aligned} \tag{38}$$

with

$$a_\delta = \frac{\sigma_\varepsilon^2}{(T-1) + \sigma_\varepsilon^2/\sigma_\delta^2}$$

and

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

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

4.2.3 Gibbs sampling for σ_ε^2

The posterior of σ_ε^2 is given by

$$\begin{aligned} f(\sigma_\varepsilon^2|\boldsymbol{\Lambda}\backslash\{\sigma_\varepsilon^2\}) &\propto f(\boldsymbol{\kappa}|\kappa_1, \delta, \sigma_\varepsilon^2) f(\sigma_\varepsilon^2) \\ &= \prod_{t=t_1+1}^{t_T} \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left[-\frac{[\Delta\kappa_t - \delta]^2}{2\sigma_\varepsilon^2} \right] \\ &\quad \times \sigma_\varepsilon^{-1} \cdot 1_{[0 \leq \sigma_\varepsilon \leq A_\varepsilon]} \\ &\propto (\sigma_\varepsilon^{-2})^{\frac{T}{2}} \exp\left[-(\sigma_\varepsilon^{-2}) \cdot \left(\frac{1}{2} \sum_{t=t_1+1}^{t_T} (\Delta\kappa_t - \delta)^2 \right) \right]. \end{aligned} \tag{39}$$

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

$$\begin{aligned} f(\sigma_\varepsilon^{-2}|\boldsymbol{\Lambda}\backslash\{\sigma_\varepsilon^2\}) &\propto (\sigma_\varepsilon^{-2})^{\frac{T}{2}-1-1} \cdot \exp\left[-(\sigma_\varepsilon^{-2}) \cdot \left(\frac{1}{2} \sum_{t=t_1+1}^{t_T} (\Delta\kappa_t - \delta)^2 \right) \right] \\ &\sim \text{Gamma}\left(\frac{T-2}{2}, \frac{1}{2} \sum_{t=t_1+1}^{t_T} (\Delta\kappa_t - \delta)^2 \right). \end{aligned}$$

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

4.3 Portfolio-specific mortality - Gamma prior

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

$$\begin{aligned} f(\Theta_x^i|\boldsymbol{\Lambda}\backslash\{\Theta_x^i\}) &\propto f(\boldsymbol{D}|\boldsymbol{E}, \boldsymbol{\Theta}^{\text{pf}}, \boldsymbol{\Theta}^{\text{rest}}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}) f(\Theta_x^i) \\ &\propto \prod_{t \in \mathcal{S}} \left(e^{-E_{t,x}^i \exp[\alpha_x + \beta_x\kappa_t]\Theta_x^i} \frac{(E_{t,x}^i \exp[\alpha_x + \beta_x\kappa_t]\Theta_x^i)^{D_{t,x}^i}}{D_{t,x}^i!} \right)^{I_{t,x}^i} \\ &\quad \times \frac{(c_x^i)^{c_x^i}}{\Gamma(c_x^i)} (\Theta_x^i)^{c_x^i - 1} \exp[-c_x^i \Theta_x^i] \\ &\propto \exp[-(c_x^i + f_x^i)\Theta_x^i] \cdot (\Theta_x^i)^{c_x^i + D_{\bullet x}^i - 1} \end{aligned} \tag{40}$$

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 (40) is proportional to a Gamma$(c_x^i + D_{\bullet x}^i, c_x^i + f_x^i)$ distribution and thus we can 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.

4.4 Portfolio-specific mortality - logNormal prior

Before we derive the posterior distribution for Θ_x^i and the hyperparameters, we define the following variables and relations:

$$\begin{aligned}
\boldsymbol{\Sigma}_i &= \sigma_i^2 \cdot \boldsymbol{\Gamma}(\rho_i) \\
\boldsymbol{\Sigma}_i^{-1} &= \tfrac{1}{\sigma_i^2} \cdot \boldsymbol{\Gamma}^{-1}(\rho_i) = \tfrac{1}{\sigma_i^2}\tfrac{1}{1-\rho_i^2} \cdot \widetilde{\boldsymbol{\Gamma}}^{-1}(\rho_i) \\
|\boldsymbol{\Sigma}_i| &= \left|\sigma_i^2 \cdot \boldsymbol{\Gamma}(\rho_i)\right| = (\sigma_i^2)^Y \cdot (1-\rho_i^2)^{Y-1} \\
\boldsymbol{\Psi}^i &= \ln \boldsymbol{\Theta}^i - \tilde{\boldsymbol{\mu}}_i = \ln \boldsymbol{\Theta}^i + \tfrac{1}{2}\sigma_i^2 \mathbf{1}_Y,
\end{aligned}$$

with

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

and $\mathbf{1}_Y$ is a column vector of ones of length Y.

4.4.1 Metropolis-Hastings sampling for Θ_x^i

Define $\boldsymbol{\Theta}_{-j}^i = \{\Theta_{y_1}^i, \ldots, \Theta_{j-1}^i, \Theta_{j+1}^i, \ldots, \Theta_{y_Y}^i\}$. The posterior density of Θ_x^i for $i \in \{\text{pf}, \text{rest}\}$ and $y_1 \le x \le y_Y$ is given by

$$\begin{aligned}
f(\Theta_x^i | \boldsymbol{\Lambda} \backslash \{\Theta_x^i\}) &\propto f(\boldsymbol{D} | \boldsymbol{E}, \boldsymbol{\Theta}^{\text{pf}}, \boldsymbol{\Theta}^{\text{rest}}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}) f(\boldsymbol{\Theta}^i | \rho_i, \sigma_i^2) \qquad (41) \\
&\propto \prod_{t\in\mathcal{S}} \left(e^{-E_{t,x}^i \exp[\alpha_x + \beta_x \kappa_t] \Theta_x^i} (\Theta_x^i)^{D_{t,x}^i} \right)^{I_{t,x}^i} \\
&\quad \times f(\Theta_x^i | \boldsymbol{\Theta}_{-x}^i, \rho_i, \sigma_i^2).
\end{aligned}$$

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

- for $x = y_1$:

$$\begin{aligned} f(\Theta_x^i|\mathbf{\Theta}_{-x}^i, \rho_i, \sigma_i^2) &\propto f(\Theta_x^i|\rho_i, \sigma_i^2) \cdot f(\Theta_{x+1}^i|\Theta_x^i, \rho_i, \sigma_i^2) \\ &= \frac{1}{\Theta_x^i\sqrt{2\pi\sigma_i^2}} \cdot \exp\left[-\frac{(\ln\Theta_x^i + \frac{1}{2}\sigma_i^2)^2}{2\sigma_i^2}\right] \\ &\quad \times \frac{1}{\Theta_{x+1}^i\sqrt{2\pi\sigma_i^2(1-\rho_i^2)}} \cdot \exp\left[-\frac{(\ln\Theta_{x+1}^i + \frac{1}{2}\sigma_i^2(1-\rho_i) - \rho_i\ln\Theta_x^i)^2}{2\sigma_i^2(1-\rho_i^2)}\right] \\ &\propto \frac{1}{\Theta_x^i\sqrt{2\pi\sigma_i^2(1-\rho_i^2)}} \cdot \exp\left[-\frac{1}{2\sigma_i^2(1-\rho_i^2)}\left(\ln\Theta_x^i + \frac{1}{2}\sigma_i^2 - \rho_i(\ln\Theta_{x+1}^i + \frac{1}{2}\sigma_i^2)\right)^2\right] \\ &\sim \text{logN}\left(-\tfrac{1}{2}\sigma_i^2 + \rho_i(\ln\Theta_{x+1}^i + \tfrac{1}{2}\sigma_i^2), \sigma_i^2(1-\rho_i^2)\right), \end{aligned}$$

- for $y_1 < x < y_Y$:

$$\begin{aligned} f(\Theta_x^i|\mathbf{\Theta}_{-x}^i, \rho_i, \sigma_i^2) &\propto f(\Theta_x^i|\Theta_{x-1}^i, \rho_i, \sigma_i^2) \cdot f(\Theta_{x+1}^i|\Theta_x^i, \rho_i, \sigma_i^2) \\ &\sim \text{logN}\left(-\frac{1}{2}\sigma_i^2 + \frac{\rho_i}{1+\rho_i^2}\left(\ln\Theta_{x-1}^i + \ln\Theta_{x+1}^i + \sigma_i^2\right), \sigma_i^2\frac{(1-\rho_i^2)}{(1+\rho_i^2)}\right), \end{aligned}$$

- for $x = y_Y$:

$$\begin{aligned} f(\Theta_x^i|\mathbf{\Theta}_{-x}^i, \rho_i, \sigma_i^2) &\propto f(\Theta_x^i|\Theta_{x-1}^i, \rho_i, \sigma_i^2) \\ &\sim \text{logN}\left(-\tfrac{1}{2}\sigma_i^2 + \rho_i(\ln\Theta_{x-1}^i + \tfrac{1}{2}\sigma_i^2), \sigma_i^2(1-\rho_i^2)\right). \end{aligned}$$

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

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

Here, $g(.|\Theta_x)$ is the logNormal density with mean $\ln\Theta_x - \frac{1}{2}s^2_{\Theta_x}$ and variance $s^2_{\Theta_x}$.

4.4.2 Metropolis-Hastings sampling for ρ_i

The posterior distribution of ρ_i for $i \in \{\text{pf}, \text{rest}\}$ is given by

$$\begin{aligned} f(\rho_i|\mathbf{\Lambda}\backslash\{\rho_i\}) &\propto f(\mathbf{\Theta}^i|\sigma_i^2, \rho_i) \cdot f(\rho_i) \qquad (42) \\ &= \frac{1}{(2\pi)^{Y/2}\Theta_{y_1}^i\cdots\Theta_{y_Y}^i \cdot |\mathbf{\Sigma}_i|^{1/2}} \\ &\quad \times \exp\left[-\frac{1}{2}(\ln\mathbf{\Theta}^i - \boldsymbol{\mu}_i)\mathbf{\Sigma}_i^{-1}(\ln\mathbf{\Theta}^i - \boldsymbol{\mu}_i)'\right] \\ &\quad \times \frac{1}{\sqrt{2\pi\sigma_{\rho_i}^2}} \cdot \exp\left[-\frac{(\text{logit}\,(\rho_i) - \mu_{\rho_i})^2}{2\sigma_{\rho_i}^2}\right] \cdot \frac{1}{\rho_i(1-\rho_i)} \end{aligned}$$

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

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. Given current value $\tilde{\rho}_i$ and Metropolis-Hastings sampling variance $s_{\rho_i}^2$, we draw a candidate $\hat{\rho}_i$ 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 candidate is between 0 and 1. The candidate distribution is not symmetric and the acceptance probability is thus given by

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

4.4.3 Metropolis-Hastings sampling for σ_i^2

The posterior distribution of σ_i^2 is given by

$$\begin{aligned} f(\sigma_i^2|\mathbf{\Lambda}\backslash\{\sigma_i^2\}) &\propto f(\mathbf{\Theta}^i|\sigma_i^2,\rho_i)\cdot f(\sigma_i^2) \\ &= \frac{1}{(2\pi)^{Y/2}\Theta_{y_1}^i\cdots\Theta_{y_Y}^i|\mathbf{\Sigma}_i|^{1/2}} \cdot \exp\left[-\tfrac{1}{2}(\ln\mathbf{\Theta}^i - \boldsymbol{\mu}_i)'\mathbf{\Sigma}_i^{-1}(\ln\mathbf{\Theta}^i - \boldsymbol{\mu}_i)\right] \\ &\quad \times \sigma_i^{-1} \\ &\propto \frac{1}{\sigma_i^{Y+1}} \cdot \exp\left[-\sigma_i^{-2}\tfrac{1}{2}(\ln\mathbf{\Theta}^i - \boldsymbol{\mu}_i)'\mathbf{\Gamma}^{-1}(\rho_i)(\ln\mathbf{\Theta}^i - \boldsymbol{\mu}_i)\right] \end{aligned} \tag{43}$$

We use the final expression in a 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 candidate 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 candidate 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 with mean $\ln\sigma_i^2 - \frac{1}{2}s_{\sigma_i^2}^2$ and variance $s_{\sigma_i^2}^2$.

5 Empirical study

In this section we illustrate our model using a real-world dataset. We estimate four different models[2]:

1. The Lee-Carter model is used to estimate population mortality for the England & Wales population for $t \in \mathcal{T}$ and $x \in \mathcal{X}$. Parameters are estimated in a frequentist framework, and this model is referred to as POP(f);

[2] We use the same parameter constrains in all models to uniquely identify the population mortality parameters.

2. The Lee-Carter model is used to estimate population mortality for the England & Wales population for $t \in \mathcal{T}$ and $x \in \mathcal{X}$. Parameters are estimated in a Bayesian framework, and this model is referred to as POP(B);

3. The model described in Section 2.2 is estimated using a Gamma prior for Θ_x^i. In this model population and group-specific mortality are estimated simultaneously in a Bayesian framework, and this model is referred to as PF(B-G);

4. The model described in Section 2.2 is estimated using a Gamma prior for Θ_x^i. In this model population and group-specific mortality are estimated simultaneously in a Bayesian framework, and this model is referred to as PF(B-logN).

These four different estimations allow us to illustrate:

- The effect on parameter estimates when moving from a frequentist to a Bayesian framework[3];
- The effect on population mortality parameters when estimating only population mortality in a Bayesian framework versus jointly estimating population mortality and group-specific mortality in a Bayesian framework;
- The effect of the different prior distributions for Θ_x^i on posterior credible intervals.

First we discuss the datasets used and the initialisation of the MCMC algorithm, then we present estimation results and mortality predictions.

Data For the portfolio under consideration we use data from the Continuous Mortality Investigation which contains mortality data of assured male lives in England & Wales. We use the years $s \in \{s_1 = 1990, \ldots, s_S = 2000\}$ and the ages $y \in \{y_1 = 45, \ldots, y_Y = 75\}$. The size of the portfolio relative to the population measured in observed deaths and observed exposures is shown in Figure 3. In total there were 28.0 million life-years and 125,390 observed deaths. If mortality in the portfolio was 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 differ clearly, and we expect mortality in the portfolio to be lower than in the population.

Eleven years of data is too little to estimate a mortality trend, and therefore we extend the dataset with population mortality data from England & Wales for the years $t \in \{t_1 = 1950, \ldots, t_T = 2000\}$ and we use the ages $x \in \{x_1 = 20, \ldots, x_X = 90\}$ to ensure we obtain mortality forecasts consistent with population mortality forecasts.[4] The rest is constructed by subtracting portfolio deaths and exposures from the population deaths and exposures for those cells (t, x) for which $E_{t,x}^{\text{pf}} > 0$.

[3] To the best of our knowledge, this is the first paper to show the parameter estimates from a frequentist Lee-Carter model and the corresponding credible intervals from a Bayesian Lee-Carter model, both with Poissonian likelihood and assuming a random walk with drift for the period effect.

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

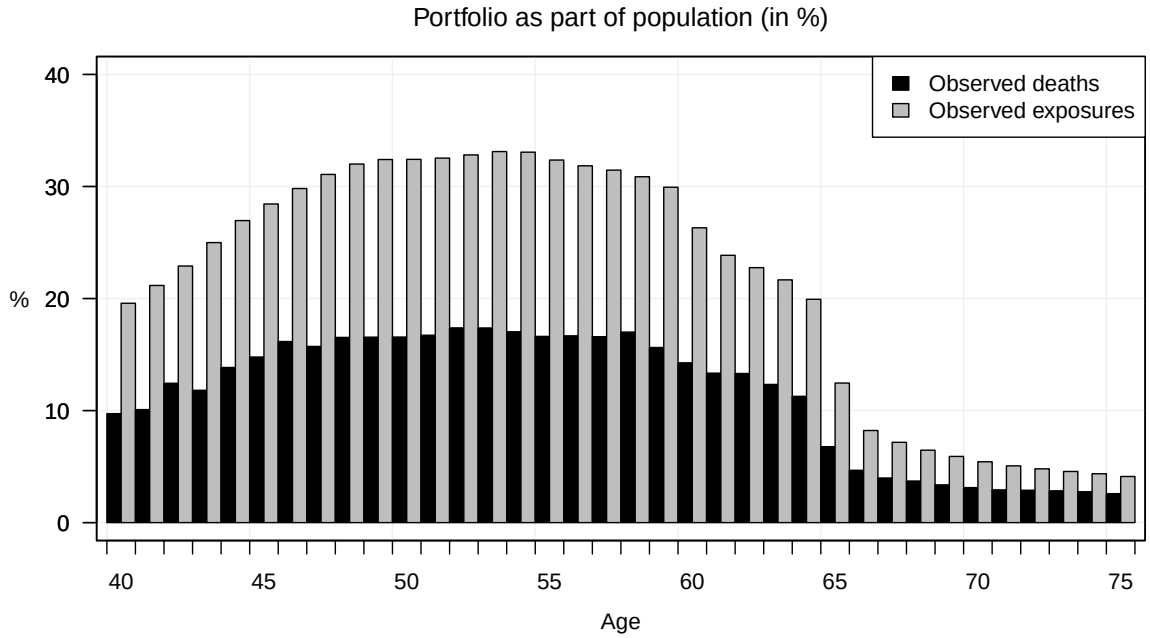

Figure 3: The relative size of the portfolio as a portion of the population measured in observed deaths and observed exposures. For each age the relative size is computed as $\sum_t d^{\text{pf}}_{t,x} / \sum_t d^{\text{pop}}_{t,x}$ and $\sum_t e^{\text{pf}}_{t,x} / \sum_t e^{\text{pop}}_{t,x}$ where each summation is over $t \in \mathcal{S}$.

Initialising prior distributions To initialise the prior distributions proposed in Section 3 we have to set the constants used in these specifications. We do this in such a way that the priors contain little information about our prior belief, i.e. such that the prior variance is large and the impact of the prior distribution on the posterior distribution is limited.

Regarding the initial draw of the MCMC simulations, we take the following approach. Let $\hat{\alpha}_x$, $\hat{\beta}_x$ and $\hat{\kappa}_t$ denote the frequentist Poisson parameter estimates of the Lee-Carter model calibrated to the England & Wales population mortality data. We use these estimates as initial values for the population parameters and set the initial portfolio-specific factors equal to one. Using the frequentist parameter estimates for β_x and κ_t, we obtain maximum likelihood estimates for μ_β, σ^2_β, δ and σ^2_ε, and use these as initial values for the hyperparameters. For the hyperparameters of Θ^i_x we start with $\rho_i = 0.8$ and $\sigma^2_i = 1$.

- 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.

- As described in Section 3, we apply a parameter constraint to $\boldsymbol{\beta}$. Therefore, we use $\nu_\beta = X^{-\frac{1}{2}}$ and $c^2_\beta = (0.5 \cdot X^{-\frac{1}{2}})^2$, which implies that the prior 95% confidence interval contains only positive values for μ_β. For the variance hyperparameter we use $A_\beta = 1$.

- Given that we expect the κ_t's to exhibit a downward trend, we use $\mu_\delta = -2$ and $\sigma^2_\delta = 0.5^2$. For the variance hyperparameter we use $A_\varepsilon = 5$.

- For the Gamma prior on the portfolio-specific factors we use $c^i_x = 1$ for all x and for

$i \in \{\text{pf}, \text{rest}\}$. 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 $A_i = 1$ for $i \in \{\text{pf}, \text{rest}\}$.
- For all Metropolis(-Hastings) sampling variances in the proposal distributions, we use $s^2 = 0.5^2$.

Convergence diagnostics We run 450,000 simulations of the MCMC algorithm for both prior distributions of the group-specific factors. We remove an initial 50,000 simulations (the burn-in period), and then collect information from every 200th iteration[5], resulting in a final sample size of 2,000. We perform the usual convergence checks[6].

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

The simultaneous estimation of the Lee-Carter and the time series model leads to slightly different parameter estimates. The largest differences occur at the estimates for β_x: for low and high ages the estimates in POP(B) are lower than in POP(f), and for the mid-ages it is vice versa. As a result, fitted mortality rates and projections of mortality rates for POP(f) and POP(B) will differ slightly. In models PF(B-G) and PF(B-logN) we also include portfolio data. The credible intervals for the parameters in the baseline force of mortality $\mu_{t,x}$ are similar for α_x and κ_t. For β_x we observe differences compared to POP(B), but the prior distribution on Θ_x^i does not have a significant effect on the credible interval for β_x. The posterior distributions for the hyperparameters δ and σ_ε^2 are similar for all model specifications.

Figure 5 shows estimated portfolio-specific factors using different methods. The factors represented by the green lines are estimated using a frequentist method that corresponds to methods used in practice. First, the observed population mortality rate is computed. Then, a Poisson GLM is estimated in which the deaths in the portfolio are explained using the portfolio exposure and the observed population mortality rate as offset, and age-dependent factors are used as explanatory variables:

$$D_{t,x}^i \sim \text{Poisson}(E_{t,x}^i \mu_{t,x}^{\text{obs}} \cdot \Theta_x^i)$$

The factors represented by the red lines are obtained in a similar way, but in this case fitted population mortality rates from the Lee-Carter model are used in the offset of the Poisson estimation instead of observed mortality rates. The blue and grey areas correspond to the 95% credible intervals (equal-tailed) for PF(B-G) and PF(B-logN).

The estimated factors for the portfolio are all below one, implying that mortality in the portfolio is lower than in the population, and in the rest the factors are generally above one.

[5] This high lag is needed to ensure convergence when using the logNormal prior on the group-specific factors.

[6] An appendix with selected convergence diagnostics is available upon request from the corresponding author.

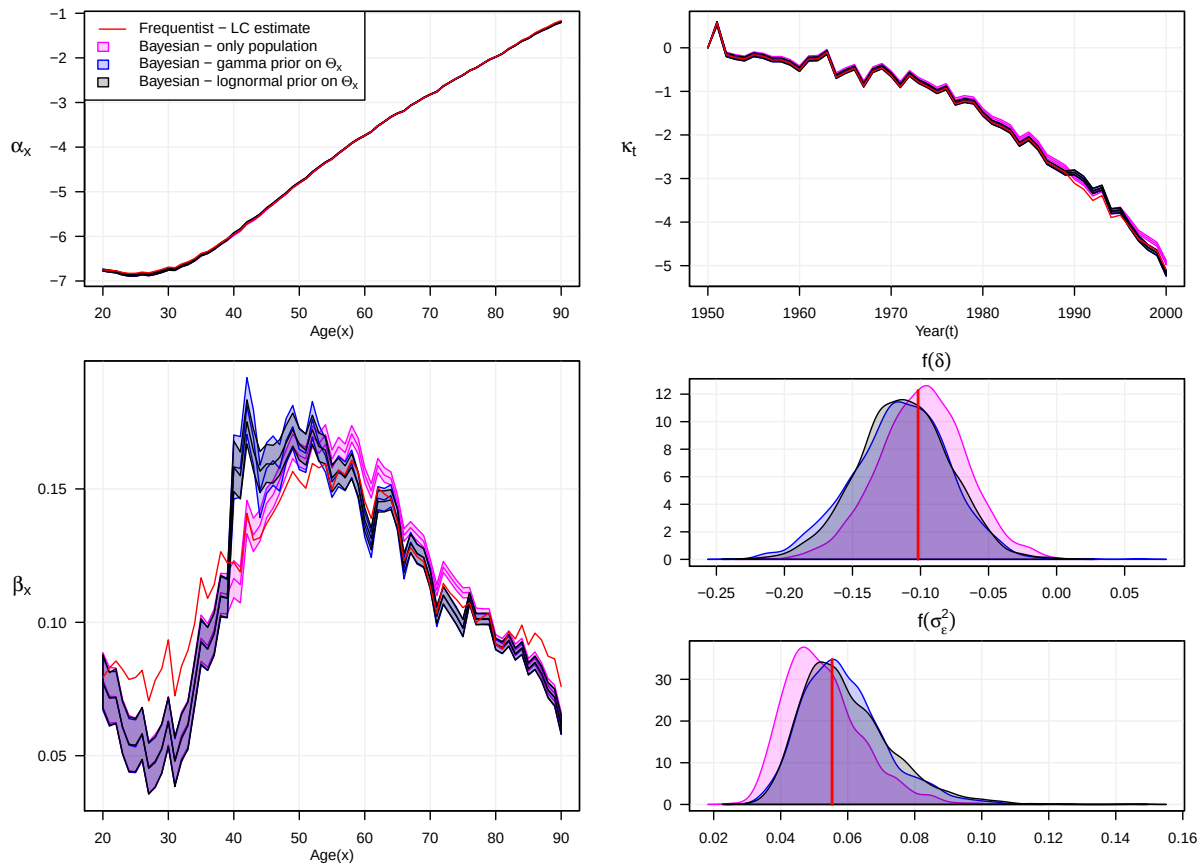

Figure 4: Parameter estimates for the model as specified in Section 2 applied to the CMI and England & Wales datasets using portfolio data from 1990-2000. In the five top panels the red lines correspond with parameter estimates for POP(f), and the shaded areas and corresponding lines show the 95% credible interval (equal-tailed) and the median of the posterior distribution for the different parameters. The magenta areas relates to the posterior distributions for POP(B), the blue areas to PF(B-G), and the grey areas to PF(B-logN). (Coloured versions of the figures can be found online.)

Further, the estimated factors from the frequentist methods show similar patterns, and the credible intervals for PF(B-G) show similar spiky behaviour as the frequentist factors, but the level differs. This can be explained by the difference in parameter estimates for β_x in the frequentist and Bayesian case. As a result of this difference, the baseline force of mortality has changed which justifies portfolio-specific factors at a different level. The credible intervals for PF(B-logN) in which dependence between factors within a group is assumed are much smoother than the factors estimated assuming independence between factors within a group. The posterior means of the mean reversion coefficients are $\rho_{\theta^{\text{pf}}} = 0.992$ and $\rho_{\theta^{\text{rest}}} = 0.999$. Finally, the posterior distributions of Θ_x^{pf} have smaller credible intervals than the posterior distributions of Θ_x^{rest}. This can be explained by the fact that the portfolio is apparently more homogeneous than the remainder of the population.

Forecasting of mortality Figure 6 shows projections of mortality rates from 1) a combination of POP(f) and frequentist estimates of portfolio-specific factors (hereafter:

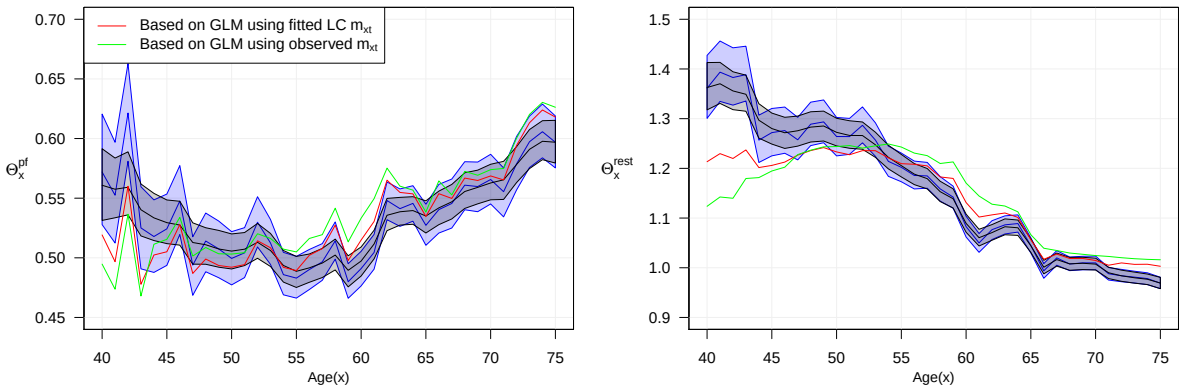

Figure 5: Parameter estimates for the model as specified in Section 2 applied to the CMI and England & Wales datasets using portfolio data from 1990-2000. The green and red lines show the estimated factors that are estimated using a Poisson GLM in which the portfolio deaths are explained using age-dependent factors with the portfolio exposure and a force of mortality as offset. For the green lines the observed population force of mortality is used, and for the red lines we used the fitted Lee-Carter force of mortality on population data. The shaded areas and corresponding lines show the 95% credible interval (equal-tailed) and the median of the posterior distribution for the different parameters. The magenta areas relates to the posterior distributions for POP(B), the blue areas to PF(B-G), and the grey areas to PF(B-logN). (Coloured versions of the figures can be found online.)

PF(f))[7], 2) PF(B-G), and 3) PF(B-logN). In these graphs we only show fitted mortality rates for the observations that are included in the likelihood, thus for the population for $t < 1990$ and for the portfolio and the rest for $t \geq 1990$. We notice again that mortality in the portfolio is not only lower, but also less uncertain than in the rest of the population.

Projections of mortality rates in a Bayesian setting using the two different prior distributions for Θ_x^i show little differences; both direction and uncertainty in the projections are similar[8]. However, there are differences between the Bayesian and frequentist projections. Especially for the lower ages, see for example $m_{t,40}^{pf}$ in Figure 6, the in-sample estimates of the mortality rates are closer to the observations for the Bayesian setting including portfolio-specific mortality than for the frequentist two-step approach. As a result, the projections for portfolio-specific mortality rates connect better with the latest observations.

We further observe that the confidence intervals for the projections from PF(B-G) and PF(B-logN) are similar to those from PF(f), though only the first two include parameter uncertainty. When we include in our projections the uncertainty in the variance parameter in the time series model, a higher variance leads to larger projection intervals whereas a

[7] We used the estimated portfolio-specific factors that are estimated using a Poisson GLM to explain the portfolio deaths with the portfolio exposure and fitted POP(f) force of mortality as offset. These portfolio-specific factors are displayed in Figure 5 by the red lines.

[8] These differences are small for the ages shown. For some ages, the projections start at a different level, because their group-specific factor is estimated at a different level, see e.g. Θ_{43}^{pf} in Figure 5. However, the frequentist projections also start at a more distant level. Relevant plots are included in an appendix that is available upon request from the corresponding author.

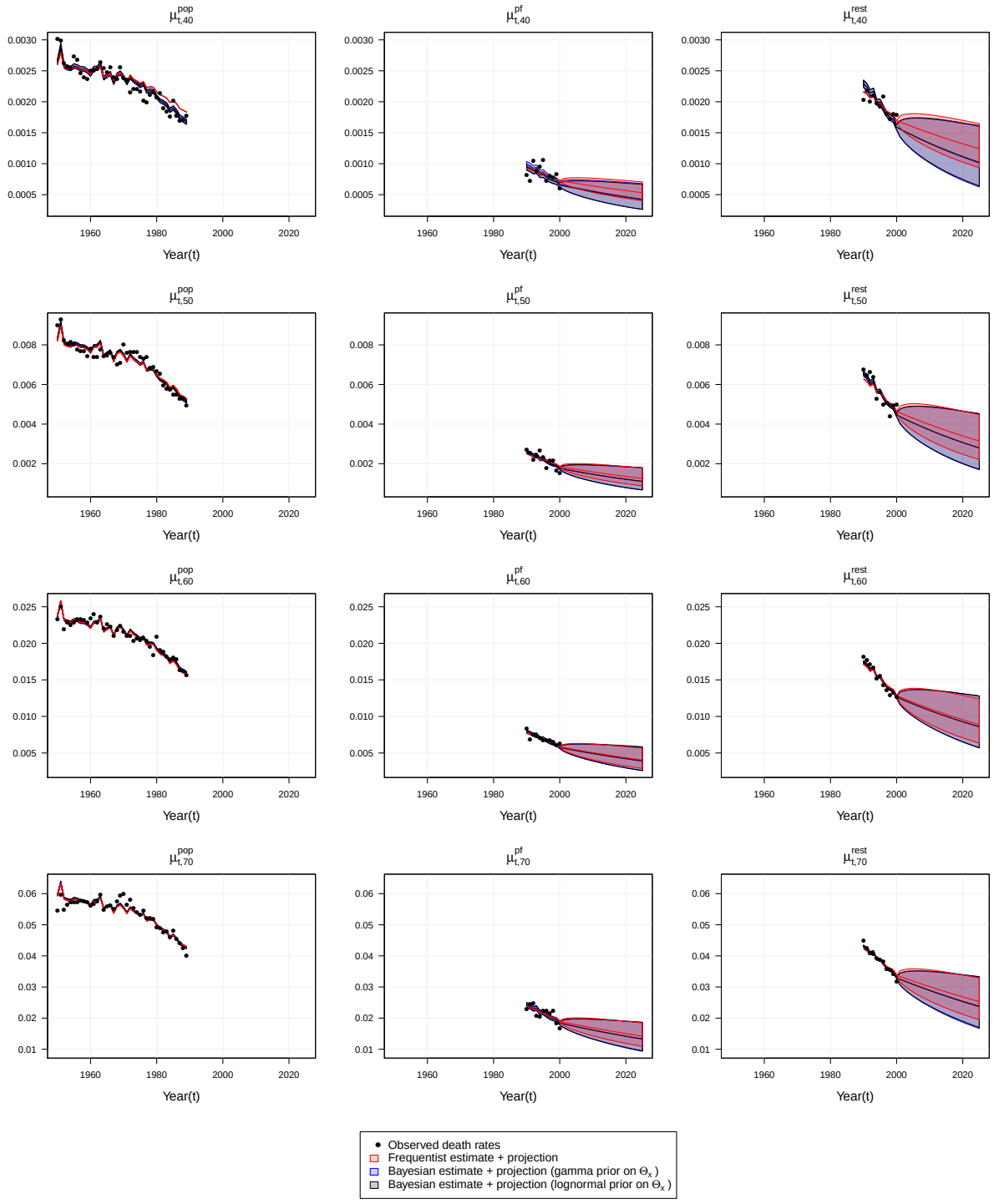

Figure 6: Estimated and projected mortality rates from POP(f) in combination with frequentist estimates of group-specific factors (red lines and areas), and from PF(B-G) and PF(B-logN) (blue and grey areas respectively). (Coloured versions of the figures can be found online.)

lower variance leads to smaller projection intervals. Including parameter uncertainty in projections of mortality rates therefore does not automatically lead to larger projection intervals.

6 Conclusion

In this paper we present a Bayesian model to simultaneously estimate population mortality and portfolio-specific mortality. This model is especially useful when there is only a limited number of years with observations available for the portfolio, since it allows for corrections on the population mortality rate projections.

We illustrate our model using the dataset on CMI assured male lives in combination with data on the England & Wales population. Parameter estimates for the population mortality model are similar in the Bayesian and frequentist setting, but the portfolio-specific factors differ from frequentist estimates. Also, the posterior distributions for the portfolio-specific factors are smoother when assuming dependence in the prior distribution. Our approach to model population mortality and portfolio-specific mortality simultaneously may help to better explain the data. Fitted values are closer to observed central death rates, which subsequently leads to projections of portfolio-specific mortality rates that connect better with recent observations. Further, this approach provides an objective way to determine to what extent mortality in the portfolio is different from mortality in the population.

We show projections of forces of mortality in a frequentist framework and from our proposed models in a Bayesian framework. For the projections in a frequentist framework we do not include parameter uncertainty, but for our models in the Bayesian framework we can incorporate parameter uncertainty in the projections in a natural way. The confidence intervals are similar, indicating that parameter uncertainty in the parameters of the time series model does not necessarily increase uncertainty in mortality rate projections.

The datasets considered in this paper only contain calendar year and age as explanatory variables. The proposed model can be extended to include other explanatory variables, see for example Gschlössl et al. (2011) in a frequentist setting. Further, in a frequentist setting the estimates of β_x can be rather volatile over ages. In a Bayesian setting often independence is assumed in the prior of β_x, and the posterior distribution of β_x is as volatile as the frequentist estimate. A smoother posterior distribution for β_x may lead to biologically more plausible projections of mortality rates. It is therefore worth exploring how mortality projections change when assuming dependence between β_x's.

Acknowledgements

The authors gratefully thank Anastasios Bardoutsos for helpful discussions and useful comments on an earlier draft of this paper.

References

Antonio, K., Bardoutsos, A. and Ouburg, W. (2015), 'A Bayesian Poisson log-bilinear model for mortality projections with multiple populations', *European Actuarial Journal*

p. (to appear).
URL: *http://ssrn.com/abstract=257287*

Antonio, K. and Zhang, Y. (2014), Nonlinear mixed models, *in* E. Frees, R. Derrig and G. Meyers, eds, 'Predictive Modeling Applications in Actuarial Science', Vol. 1, Cambridge University Press, pp. 398 – 426.

Bafumi, J., Gelman, A., Park, D. and Kaplan, N. (2005), 'Practical issues in implementing and understanding Bayesian ideal point estimation', *Political Analysis* **13**, 171 – 187.

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.

Butt, Z. and Haberman, S. (2004), 'Application of frailty-based mortality models using generalized linear models', *Astin Bulletin* **34**(1), 175 – 197.

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.

Czado, C., Delwarde, A. and Denuit, M. (2005), 'Bayesian Poisson log-bilinear mortality projections', *Insurance: Mathematics and Economics* **36**(3), 260 – 284.

Danesi, I., Haberman, S. and Millossovich, P. (2015), 'Forecasting mortality in subpopulations using Lee–Carter type models: A comparison', *Insurance: Mathematics and Economics* **62**, 151 – 161.

Denuit, M., Dhaene, J., Goovaerts, M. and Kaas, R. (2005), *Actuarial theory for dependent risks*, John Wiley & Sons.

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.

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

Institute of Actuaries in Belgium (2015), 'The IA|BE 2015 mortality projection for the Belgian population'.
URL: *http://www.iabe.be/nl/iabe-mortality-tables*

Jarner, S. and Kryger, E. (2011), 'Modelling adult mortality in small populations: the SAINT model', *Astin Bulletin* **41**(2), 377 – 418.

Kan, H. (2012), 'A Bayesian mortality forecasting framework for population and portfolio mortality'.

Koninklijk Actuarieel Genootschap (2014), 'Prognosetafel AG2014'.
URL: *http://www.ag-ai.nl/view.php?action=view&Pagina_Id=480*

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, 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*, Oxford University Press, New York.

Plat, R. (2009*a*), 'On stochastic mortality modeling', *Insurance: Mathematics and Economics* **45**(3), 393 – 404.

Plat, R. (2009*b*), '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.

The Dutch Bank (2012), 'Good Practice Gebruik Fondsspecifieke Ervaringssterfte'.
URL: *http://www.toezicht.dnb.nl/binaries/50-226664.pdf*

Tomas, J. and Planchet, F. (2013), 'Multidimensional smoothing by adaptive local kernel-weighted log-likelihood: Application to long-term care insurance', *Insurance: Mathematics and Economics* **52**(3), 573 – 589.

van Berkum, F., Antonio, K. and Vellekoop, M. (2014), 'The impact of multiple structural changes on mortality predictions', *Scandinavian Actuarial Journal* p. (to appear).
URL: *http://tandfonline.com/doi/full/10.1080/03461238.2014.987807*

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.