A Bayesian non-parametric model for small population mortality

Hong Li, Yang Lu

A Bayesian non-parametric model for small population mortality

Hong Li, Yang Lu

A Bayesian non-parametric model for small population mortality

Hong Li[a] and Yang Lu[b‡]

[a]School of Finance, Nankai University, Tianjin, P.R. China; [b]Aix-Marseille School of Economics, Aix-Marseille University, France

ABSTRACT

This paper proposes a Bayesian non-parametric mortality model for a small population, when a benchmark mortality table is also available and serves as part of the prior information. In particular, we extend the Poisson-gamma model of Hardy and Panjer to incorporate correlated and age-specific mortality coefficients. These coefficients, which measure the difference in mortality levels between the small and the benchmark population, follow an age-indexed autoregressive gamma process, and can be stochastically extrapolated to ages where the small population has no historical exposure. Our model substantially improves the computation efficiency of existing two-population Bayesian mortality models by allowing for closed form posterior mean and variance of the future number of deaths, and an efficient sampling algorithm for the entire posterior distribution. We illustrate the proposed model with a life insurance portfolio from a French insurance company.

ARTICLE HISTORY

Received 1 June 2017
Accepted 13 December 2017

KEYWORDS

Autoregressive gamma; credibility; parameter uncertainty; two-population mortality model

1. Introduction

Projecting mortality is of great concern for many disciplines, such as demographic analysis, social planning, policy-making, and actuarial practices. In recent decades, various methodologies for mortality modeling and projections have been proposed in the literature, such as the Lee–Carter model (1992), the Cairns–Blake–Dowd (2006) model, and their various extensions (see e.g. Booth and Tickle 2008, Cairns *et al.* 2011a). Moreover, government agencies and societies of actuaries publish national or industry-level regulatory mortality tables on a regular basis. However, in many situations, the population of interest, such as the portfolios of pension funds or insurance companies, have limited size and/or short observation window. For such portfolios, the observed mortality experience is typically erratic, with hard-to-identify, or even unreliable longevity trend.

However, such small populations are often sub-populations of a larger population, such as a national population or a larger collective portfolio. The mortality pattern of the latter is usually much more regular, and easier to forecast. In this case, mortality table of the larger population may serve as a benchmark when modeling and forecasting mortality experience of the small population.

To this end, different solutions have been proposed in the literature. The first branch of literature is based on deterministic approaches, such as relational models (see e.g. Himes *et al.* 1994, Brouhns *et al.* 2002, Tomas and Planchet 2015), as well as methods based on the distortion of probability distributions (see e.g. Denuit *et al.* 2007, Bienvenüe and Rullière 2012). These models specify a deterministic, linear, or non-linear functional relationship between the mortality curves of the two populations. Their drawbacks are that, first, the functional relationship typically involves a

CONTACT Yang Lu yang.lu@univ-paris13.fr

‡Department of Economics (CEPN), University of Paris 13, Villetaneuse, France

small number of parameters,[1] which limits the flexibility of the model; second, and more importantly, the deterministic relationship does not capture the impact of *parameter uncertainty* on pricing/reserving/capital requirement calculation, which could be substantial for small populations.

The second class of models are Bayesian (see e.g. Dowd *et al.* 2011, Cairns *et al.* 2011b, Antonio *et al.* 2015, van Berkum *et al.* 2017). While they are flexible, and account for parameter uncertainty in a natural way, these models are currently rather computationally intensive. While credibility theory has been proposed to simplify such models (see e.g. Hardy and Panjer 1998, Olivieri and Pitacco 2012, Salhi *et al.* 2015), the computational gain usually relies on strong parametric restrictions.

This paper proposes a Bayesian non-parametric model for the mortality of a small population, when mortality table of a larger population is exogenously given as the benchmark. The proposed model has a flexible structure, while at the same time being less computationally intensive than the existing Bayesian multi-population models. In the proposed model, the small population is assumed to follow the same longevity trend as the benchmark population, but has different, age-specific mortality levels. These levels are captured by a set of correlated and age-specific coefficients. *A priori*, the mortality level of the small population is the same as the benchmark population for all ages, and its forecasts depend solely on that of the latter. As mortality data accumulate, the estimation of the mortality coefficients are updated by the Bayesian rule, and the small population has increasing weights when forecasting its own mortality.

From the computational side, the model gives *closed form* expressions of the posterior moments, such as mean and variance, of the next period's total death count of the small population. Moreover, in an insurance context, where the observations are claims with heterogeneous payments, closed form expressions are available for moments of the total claim amount as well. These closed form expressions significantly reduce the computational requirement of our model compared to existing Bayesian models. Therefore, we are able to investigate, with limited computational cost, the impact of the prior distribution on the expected future death counts and claim amount and the corresponding uncertainty. These parameters include, among others, the size of the population of interest, the length of the observation window, and the choice of the prior distribution. From the flexibility side, our approach is non-parametric, in the sense that the mortality coefficients, which measure the difference of mortality level, are age-specific. Moreover, the mortality coefficients are correlated across different ages, with stronger correlation when two ages are closer. Such correlation leads to smoother projected mortality curves. Compared to conventional non-parametric models such as Olivieri and Pitacco (2012), the Bayesian aspect of the model allows us to conduct stochastic extrapolation of the mortality coefficients to ages with no historical risk exposure.

The proposed model has also implications for regulation and risk management. Firstly, the closed form expressions allow us to conveniently decompose the total variance into two parts: the diversifiable part due to sampling variation, and the non-diversifiable part due to parameter uncertainty. In the empirical analysis, it is shown that, for a mildly large life insurance portfolio with around 15,000 policyholders and a fairly long observation period (11 years), the contribution of the parameter uncertainty to the total variance can still be rather substantial (about 12%). This result illustrates the danger of under-estimating the risk if a deterministic relational model is used. Secondly, our model may be suitable for building an internal model under the new insurance regulations, such as Solvency II. These regulations encourage the use of proprietary mortality tables that better reflect company-specific risks. Nevertheless, their implementation has to be monitored carefully, in order to prevent insurers from regulatory arbitrage. When a regulatory life table (often built with some risk margin on top of the 'best estimate') serves as the benchmark, our model leads to projections that are closer to the insurer's own mortality experience when its mortality experience is more reliable (e.g. a larger portfolio size with a longer observation window).

In a numerical illustration, we apply the non-parametric Bayesian model to a French life insurance portfolio. In particular, we compute the one-year-ahead predictive expectation and variance of the

[1] See however, Tomas and Planchet (2013), Cadena and Denuit (2016) for recent advances in flexible relational models.

total death counts, and illustrate the impact of the portfolio size and the choice of prior distributions on the risk management of the portfolio. We illustrate, for a given prior distribution of the mortality coefficients, how a larger portfolio gains more weight on its own mortality experience when projecting future mortality, whereas a smaller portfolio relies more heavily on the benchmark mortality table. Besides, we also discuss the joint impact of the portfolio size and the prior distribution, which measures the degree of belief on the benchmark mortality table.

The paper is organized as follows. Section 2 reviews the Poisson-gamma model of Hardy and Panjer (1998). Section 3 discusses the drawbacks of this benchmark approach, and introduces the Bayesian non-parametric model. Section 4 provides the Bayesian updating formula for the mortality forecasts, in particular the closed form expression of the conditional expectation and variance of the total death count (claim amount). Section 5 illustrates the model with a real life insurance portfolio. Section 6 concludes. Technical proofs are gathered in Appendices.

2. The benchmark Poisson-gamma model

The Poisson-gamma credibility model of Hardy and Panjer (1998) is based on the conjugacy property of the gamma distribution with respect to the Poisson likelihood function, inspired by studies of non-life insurance portolios (see e.g. Dionne and Vanasse 1989). Assume that an insurance portfolio (the small population) contains mortality experience of I ages and T years. Ages are indexed by 1, 2, …, I, with indices '1' and 'I' referring to the smallest and the largest ages in the portfolio, rather than the biological age 1 and I. The number of death at each age x and year t is conditionally Poisson distributed (see e.g. Czado *et al.* 2005):

$$d_{x,t} \sim \mathcal{P}(\theta_x \mu_{0,x,t} e_{x,t}), \qquad \forall x = 1, \ldots, I, \quad t = 1, \ldots, T, \tag{2.1}$$

where

- $e_{x,t}$ is the exposure of the portfolio at age x and year t.
- $\mu_{0,x,t}$ is the force of mortality of a benchmark population at the same age and year. This can be, for instance, mortality of the corresponding national population, a collective portfolio, or a regulatory mortality table. In general, mortality of this population can be stochastic with a *longevity trend*, i.e. mortality rates at each age are non-stationary, and decrease over time. $\mu_{0,x,t}$ can be estimated from mortality data of the benchmark population, or exogenously given, e.g. by regulators.
- the age-specific mortality coefficient, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_I)'$, are time-invariant. They capture the relative risk of the portfolio with respect to the benchmark population. For instance, if θ_x is smaller than 1, then the insurance portfolio has lower expected mortality at age x than the benchmark. The mortality coefficients are unobserved and stochastic, inducing uncertainty on the portfolio's mortality.

The assumption that $\boldsymbol{\theta}$ is constant means that the mortality ratios between the two populations are stable. In other words, the small population shares the same longevity trend with the benchmark population. This assumption is also used in the subsequent literature (see e.g. Olivieri and Pitacco 2012, Salhi *et al.* 2015, van Berkum *et al.* 2017). This assumption is motivated by the following arguments:

- It allows the mortality improvements in these two populations to be coherent in the long-run, which is an important criterion for modeling mortality of closely related populations.[2]
- When the small population has a reduced observation window, which is often the case in insurance application, it is extremely difficult to reliably derive the mortality trend of the small

[2] Coherence assumption is typically imposed when the modeled populations have similar socioeconomic characteristics (Li and Lee 2005), or when one population is a sub-population of another (Dowd *et al.* 2011).

population based on its own data. Therefore, it is more reasonable to assume that the two populations share the same mortality trends.

Let us now specify the prior distribution of $\boldsymbol{\theta}$. Model (2.1) is first proposed by Hardy and Panjer (1998), who assume that θ_x is common for all ages. Specifically, $\theta_x = \theta$ for all x's, where θ has the prior distribution $\gamma(\nu, \nu)$. In this way, the prior mean of θ equals to 1, and thus $\mathbb{E}[d_{x,t}] = \mu_{0,x,t}e_{x,t}$. In other words, the portfolio is expected to have *a priori* the same mortality experience as the benchmark population. As is shown by Salhi *et al.* (2015), this condition is often not satisfied in practice. In particular, the common θ assumption may lead to over-estimation of $d_{x,t}$ at certain ages, and under-estimation at some others.

As an alternative to the common θ assumption, Olivieri and Pitacco (2012) and van Berkum *et al.* (2017) assume that $\boldsymbol{\theta}$ is a vector of i.i.d. random variables, each with the prior distribution $\gamma(\nu, \nu)$. This assumption is *Bayesian non-parametric*, in the sense that the number of parameters vary with the number of ages included in the analysis. In the sequel, we refer to the model with i.i.d. $\boldsymbol{\theta}$ assumption as the benchmark Poisson-gamma model.

As observations accumulate over time, the conditional distribution of each θ_x is regularly updated using Bayes' rule. For each T, denote by $D_x = (d_{x,1}, \ldots, d_{x,T}, e_{x,1}, \ldots, e_{x,T}, \mu_{0,x,1}, \ldots, \mu_{0,x,T})$ the set of death counts and exposure of the portfolio and the force of mortality of the benchmark population at age x, the conditional density of D_x given θ_x is:

$$\begin{aligned} l(D_x|\theta_x) &= \prod_{t=1}^{T} l(d_{x,t}|\theta_x) = \prod_{t=1}^{T} \frac{(\theta_x e_{x,t}\mu_{0,x,t})^{d_{x,t}} \exp\left(-\theta_x e_{x,t}\mu_{0,x,t}\right)}{\Gamma(d_{x,t}+1)} \\ &\propto \theta_x^{\sum_{t=1}^{T} d_{x,t}} \exp\left(-\theta_x \sum_{t=1}^{T} e_{x,t}\mu_{0,x,t}\right). \end{aligned} \tag{2.2}$$

The following closed form expression of the posterior distribution is the key contribution to the popularity of the benchmark Poisson-gamma model. Using Equation (2.2), the conditional distribution of θ_x given the death counts is:

$$l(\theta_x \mid D_x) \propto \theta_x^{\nu-1+\sum_{t=1}^{T} d_{x,t}} \exp\left(-\theta_x\left(\nu + \sum_{t=1}^{T} e_{x,t}\mu_{0,x,t}\right)\right). \tag{2.3}$$

That is, $l(\theta_x \mid D_x)$ follows the gamma distribution $\gamma(\nu + \sum_{t=1}^{T} d_{x,t}, \nu + \sum_{t=1}^{T} \mu_{0,x,t}e_{x,t})$. Thus, the predicted mean of next year's death count is:

$$\begin{aligned} \mathbb{E}[d_{x,T+1}|D_x] &= \mu_{0,x,T+1}e_{x,T+1}\frac{\nu + \sum_{t=1}^{T} d_{x,t}}{\nu + \sum_{t=1}^{T} \mu_{0,x,t}e_{x,t}} \\ &= e_{x,T+1}\mu_{0,x,T+1}\left(\frac{\nu}{\nu + \sum_{t=1}^{T} \mu_{0,x,t}e_{x,t}} + \frac{\sum_{t=1}^{T} \mu_{0,x,t}e_{x,t}}{\nu + \sum_{t=1}^{T} \mu_{0,x,t}e_{x,t}} \frac{\sum_{t=1}^{T} d_{x,t}}{\sum_{t=1}^{T} \mu_{0,x,t}e_{x,t}}\right). \end{aligned} \tag{2.4}$$

Therefore, the posterior expected number of deaths can be interpreted as the prior expected number, $e_{x,T+1}\mu_{0,x,T+1}$, multiplied by an average of 1 and the ratio between the observed and the expected number of deaths, $\frac{\sum_{t=1}^{T} d_{x,t}}{\sum_{t=1}^{T} \mu_{0,x,t}e_{x,t}}$.

In this model, parameter ν captures the belief of the benchmark mortality table: if ν is much larger than the cumulative sum, $\sum_{t=1}^{T} \mu_{0,x,t}e_{x,t}$, then we have:

$$\mathbb{E}[d_{x,T+1}|D_x] \approx e_{x,T+1}\mu_{0,x,T+1}.$$

In other words, we have strong belief that mortality of the small population would be close to that of the benchmark population. Contrarily, if ν is small, we have:

$$\begin{aligned}\mathbb{E}[d_{x,T+1}|D_x] &\approx e_{x,T+1}\frac{\sum_{t=1}^{T} d_{x,t}}{\sum_{t=1}^{T}\frac{\mu_{0,x,t}}{\mu_{0,x,T+1}}e_{x,t}}\\ &= e_{x,T+1}\mu_{0,x,T+1}\frac{\sum_{t=1}^{T} d_{x,t}}{\sum_{t=1}^{T}\mu_{0,x,t}e_{x,t}}.\end{aligned}$$

In this case, the posterior expected death count depends on the prior expectation, $e_{x,T+1}\mu_{0,x,T+1}$, multiplied by an adjustment factor, $\frac{\sum_{t=1}^{T} d_{x,t}}{\sum_{t=1}^{T}\mu_{0,x,t}e_{x,t}}$, which reflects the difference between the observed and the expected number of deaths. In other words, predictive death counts in the small population are heavily driven by its own historical pattern.

Let us now compare the variance of the posterior distribution with that of the prior distribution. We have:

$$\mathbb{V}[\theta_x|D_x] = \frac{\nu + \sum_{t=1}^{T} d_{x,t}}{(\nu + \sum_{t=1}^{T}\mu_{0,x,t}e_{x,t})^2}, \qquad \mathbb{V}[\theta_x] = \frac{1}{\nu}. \tag{2.5}$$

Therefore, we have $\mathbb{V}[\theta_x] > \mathbb{V}[\theta_x|D_x]$ if and only if:

$$\nu\left(\sum_{t=1}^{T} d_{x,t} - 2\sum_{t=1}^{T}\mu_{0,x,t}e_{x,t}\right) < \left(\sum_{t=1}^{T}\mu_{0,x,t}e_{x,t}\right)^2.$$

A sufficient condition of this latter inequality is:

$$\sum_{t=1}^{T} d_{x,t} < 2\sum_{t=1}^{T}\mu_{0,x,t}e_{x,t}. \tag{2.6}$$

When the mortality of the small population is lower, or close to that of the benchmark population, Equation (2.6) holds with a large probability. Hence, the posterior variance $\mathbb{V}[\theta_x|D_x]$ is generically smaller than the prior variance $\mathbb{V}[\theta_x]$. In other words, information updating lowers the uncertainty of the conditional distribution of $l(\theta_x|D_x)$.

Let us now discuss the difference of the Poisson-gamma model with the relational models, such as the Brass model (see e.g. Brouhns *et al.* 2002). The latter model assumes that the force of mortality of a population of interest, μ_1, is related to that of a benchmark population, μ_2, via an affine relationship:

$$f(\mu_1(x,t)) = \theta_1 f(\mu_2(x,t)) + \theta_2,$$

where f is some link function, and θ_1 and θ_2 are typically deterministic and constant across different ages. The Poisson-gamma model can be viewed as having a similar spirit, in the sense that $f(x) = x$, $\theta_2 = 0$. However, θ_x is age-specific and stochastic, rather than deterministic and common across ages.

3. The autoregressive gamma-Poisson model

The benchmark Poisson-gamma model assumes independence between the mortality coefficients $\boldsymbol{\theta}$. As a consequence, the posterior distribution of θ_x depends only on the small population's mortality experience at age x, and the resulting curve of $\mathbb{E}[\theta_x|D_x]$ can be rather erratic. In particular, at ages

with small (or no) exposure, $\mathbb{E}[\theta_x|D_x]$ is close to the prior mean, 1; whereas at ages with a large exposure, $\mathbb{E}[\theta_x|D_x]$ could be rather different than 1.

In this paper, we relax the independence assumption on the mortality coefficients by introducing the autoregressive gamma (ARG) process. This allows the posterior distribution of each θ_x to depend not only on mortality experience at age x, but the whole sample. Moreover, the ARG process leads to stochastic and smooth extrapolation of the portfolio mortality table to ages with no historical experience, rather than a point estimate given by the deterministic extrapolation methods (see e.g. Salhi *et al.* 2015).

The ARG process is the exact time discretization of the continuous time Cox–Ingersoll–Ross process (see Nieto-Barajas and Walker 2002, Gouriéroux and Jasiak 2006). It has gamma marginal distribution, and nests the benchmark Poisson-gamma credibility model. To define the ARG process, we can start from either end of the age group, i.e. θ_1 or θ_I, and construct the process iteratively with an auxiliary counting process $\boldsymbol{N} = (N_1, \ldots, N_I)'$. For example, the construction procedure starting from θ_1 is as follows:

- θ_1 follows a gamma distribution, $\gamma(\nu, \nu)$.
- Given θ_1, introduce a latent count variable, N_1, which follows a Poisson conditional distribution with parameter $\beta\theta_1$. Given (N_1, θ_1), θ_2 then follows the conditional gamma distribution $\gamma(\nu + N_1, \beta + \nu)$.
- The next latent count variable, N_2, follows the conditional Poisson distribution $\mathcal{P}(\beta\theta_2)$, and θ_3 follows $\gamma(\nu + N_2, \nu + \beta)$, etc.

To summarize, we have the following causal chain:

$$\theta_1 \to N_1 \to \theta_2 \to N_2 \to \theta_3 \cdots \tag{3.1}$$

This is a Markov chain, in the sense that the conditional distribution of each variable given variables on its left hand side depends only on its nearest left neighbour. For instance,

$$l(N_2|\theta_2, N_1, \theta_1) = l(N_2|\theta_2), \qquad l(\theta_2|N_1, \theta_1) = l(\theta_2|N_1).$$

Moreover, $\boldsymbol{N}$ and $\boldsymbol{\theta}$ are both Markov processes with respect to their own history. For instance, the conditional distribution of N_3 given N_1, N_2 only depends on N_2.

Alternatively, we can define the joint distribution of $\boldsymbol{\theta}$ from the oldest age $x = I$, and construct the process in the reverse direction:

$$\cdots, \theta_{I-2} \leftarrow N_{I-2} \leftarrow \theta_{I-1} \leftarrow N_{I-1} \leftarrow \theta_I,$$

where N_{I-1} follows $\mathcal{P}(\beta\theta_I)$, and θ_{I-1} follows $\gamma(\nu + N_{I-1}, \nu + \beta)$, etc. It will become clear in Proposition 1 that process $\boldsymbol{\theta}$ has the time reversibility property. Intuitively, this property means that these two methods of construction lead to the same joint distribution of $\boldsymbol{\theta}$, and hence the definition of its dynamics does not depend on the choice of the direction.

This specification has a similar spirit to the lognormal autoregressive specification proposed by van Berkum *et al.* (2017), in the sense that both models allow correlation between different θ_x's. However, under the log-normal model, moments of the posterior distribution need to be computed using simulations.

It is also worth mentioning that the counting variables N_x's are latent and have no direct link with the observed mortality counts $d_{x,t}$'s. They can be interpreted as (age-dependent) regimes, since the larger N_x, the larger θ_x in expectation. Although the ARG process can also be introduced without these count variables (see e.g. Lu Forthcoming) for an introduction via the conditional Laplace transform], they are crucial to the closed form posterior moment expressions. Now we summarize some properties of the ARG process that are useful for our application.

Proposition 1 (see Gouriéroux and Jasiak 2006):

(1) *(Marginal distribution) For each age x, variable θ_x is marginally $\gamma(\nu,\nu)$ distributed.*
(2) *(Serial correlation) For each h, we have:*

$$\mathbb{E}[\theta_{x+h}|\theta_x] = 1-\rho^h+\rho^h\theta_x, \tag{3.2}$$

$$\mathbb{V}[\theta_{x+h}|\theta_x] = \frac{(1-\rho^h)^2}{\nu} + 2\frac{\rho^h(1-\rho^h)}{\nu}\theta_x. \tag{3.3}$$

As a consequence, we have

$$corr[\theta_x,\theta_{x+h}] = \rho^h,$$

where $\rho := \frac{\beta}{\beta+\nu}$. When β is equal to zero, $\theta_1,\theta_2,\ldots,\theta_x$ are independent, and we get the benchmark Poisson-gamma model in Section 2 as a special case.

(3) *(Reversibility) The dynamics of the process $\boldsymbol{\theta}$ is reversible, that is, in the reverse direction, the process θ_I, θ_{I-1}, …, θ_1 has the same Markov, ARG dynamics. As a consequence, we have also:*

$$\mathbb{E}[\theta_{x-h}|\theta_x] = 1-\rho^h+\rho^h\theta_x, \qquad \forall h\in\mathbb{N}. \tag{3.4}$$

The proof of these properties can be found in Gouriéroux and Jasiak (2006), and is thus omitted. The first property on marginal distribution is due to the fact that the gamma distribution is a conjugate prior to the Poisson likelihood function. This conjugacy property greatly simplifies the updating formula and explains why gamma distribution is the invariant distribution of the process. It also implies that the proposed model is a direct generalization of the benchmark model. Indeed, if we condition only on the death counts at age x, we still have the linear prediction formula:

$$\mathbb{E}[d_{x,T+1}|D_x] = e_{x,T+1}\mu_{0,x,T+1}\frac{\nu+\sum_{t=1}^T d_{x,t}}{\nu+\sum_{t=1}^T \mu_{0,x,t}e_{x,t}}.$$

The second property shows that the auto-correlation of the process $\boldsymbol{\theta}$ is the same as an AR(1) process. This explains the terminology ARG. The larger ρ, the smoother the (prior) trajectory of $\boldsymbol{\theta}$, and as a consequence, the smoother its posterior trajectory. Moreover, the dependence structure between the different risk parameters θ_x's is *spatial*, in the sense that the closer the two ages x_1 and x_2, the stronger the correlation between θ_{x_1} and θ_{x_2}.

The conditional expectation formula (3.2) allows us to extrapolate the mortality table to ages older than I. Indeed, the expected value of θ_{I+h} given observations at *all* ages is:

$$\mathbb{E}[\theta_{I+h}|\underline{D_I}] = \mathbb{E}\Big[\mathbb{E}[\theta_{I+h}|\theta_I,\underline{D_I}] \mid \underline{D_I}\Big] = \mathbb{E}\Big[\mathbb{E}[\theta_{I+h}|\theta_I] \mid \underline{D_I}\Big] = 1-\rho^h+\rho^h\mathbb{E}[\theta_I|\underline{D_I}], \tag{3.5}$$

where $\underline{D_x} = (D_1, D_2, \ldots, D_x)$. In particular, $\underline{D_I}$ is the whole dataset of death counts and exposure of the portfolio, and the corresponding death probabilities of the benchmark population.

The extrapolation method introduced above is smooth, while being able to take into account uncertainties in θ_x. First, for a fixed h, if ρ is close to 0, then the RHS of Equation (3.5) is close to 1, i.e. the projected mortality at the advanced ages are nearly not affected by observed mortality of the portfolio. On the other hand, if ρ is close to 1, then this term is close to $\mathbb{E}[\theta_I|\underline{D_I}]$. Therefore, with appropriate choice of ρ, the correlation structure of $\boldsymbol{\theta}$ ensures that the projected mortality in each year vary smoothly across ages, rather than having a jump at age $I+1$. Second, compared to existing deterministic extrapolation methods, this method takes into account the uncertainty of the extrapolated mortality coefficients. For instance, by Equation (3.3), the variance of the extrapolation at age $I+h$ is:

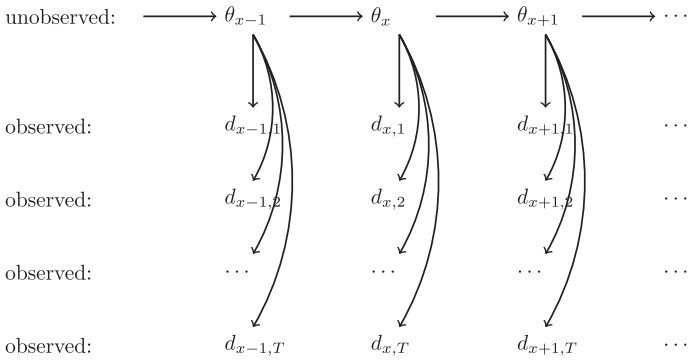


Figure 1. The age-indexed state-space model.

$$\begin{aligned}\mathbb{V}[\theta_{I+h}|\underline{D_I}] &= \mathbb{V}\Big[\mathbb{E}[\theta_{I+h}|\theta_I]|\underline{D_I}\Big] + \mathbb{E}\Big[\mathbb{V}[\theta_{I+h}|\theta_I]|\underline{D_I}\Big] \\ &= \rho^{2h}\mathbb{V}[\theta_I|\underline{D_I}] + \frac{(1-\rho^h)^2}{\nu} + 2\frac{\rho^h(1-\rho^h)}{\nu}\mathbb{E}[\theta_I|\underline{D_I}]. \end{aligned} \tag{3.6}$$

Hence, when $h = 0$, the RHS in Equation (3.6) is equal to $\mathbb{V}[\theta_I|\underline{D_I}]$; whereas when the extrapolation horizon h is large, the RHS converges to $\frac{1}{\nu} = \mathbb{V}[\theta_I]$. Finally, the reversibility property of the ARG process allows us to extrapolate the mortality of the portfolio to lower ages without historical data.

4. The Bayesian updating algorithm

After specifying the prior distribution of $\boldsymbol{\theta}$, it remains to compute their joint posterior distribution given all historical data, $l(\boldsymbol{\theta}|\underline{D_I})$. The introduction of the serial correlation between θ_x's makes this task more difficult than the benchmark case. In particular, the posterior distribution is no longer gamma for each x, as opposed to Equation (2.3). This problem is usually tackled using computationally intensive MCMC techniques (see e.g. Pitt and Shephard 1999, Czado *et al.* 2005, Pedroza 2006, Antonio *et al.* 2015, Koopman *et al.* 2015). In the rest of this section, we formulate our model in a state-space form, and derive closed form expressions of the posterior predictive mean and variance using the chain structure of the ARG process (Equation (3.1)). A simple sampling algorithm for the joint distribution $l(\boldsymbol{\theta}|\underline{D_I})$ is also proposed.

4.1. The state-space representation

A state-space model is a time series model in which there is a latent process in parallel to the observed processes. In our application, this 'time series' refers to the age dimension, and the latent state variables are the mortality coefficients, $\boldsymbol{\theta}$, whereas the observed variables are the observed death counts $d_{x,1}, \ldots, d_{x,T}$. The system has the following state-space representation[3]:

In Figure 1, the vectors $(d_{x,1}, \ldots, d_{x,T})$'s are i.i.d. across x conditional on the benchmark force of mortality, $(\mu_{0,x,t})_{x,t}$, the portfolio exposure $(e_{x,t})_{x,t}$, and $\boldsymbol{\theta}$. The arrow between θ_x and $d_{x,t}$, $t = 1, \ldots, T$ indicates that each $d_{x,t}$ depends on θ_x via Equation (2.1).

[3]The state space based on the reverse order, i.e. from θ_I to θ_1, can be constructed in the similar manner.

In this state-space representation, transition density of each θ_x can be derived by conditioning on the counting variable:

$$\begin{aligned} l(\theta_x|\theta_{x-1}) &= \sum_{n=0}^{\infty} \mathbb{P}[N_{x-1}=n] l(\theta_x|n,\theta_{x-1}) \\ &= \sum_{n=0}^{\infty} \frac{e^{-\beta\theta_{x-1}}(\beta\theta_{x-1})^n}{\Gamma(n+1)} \frac{(\beta+\nu)^{\nu+n}}{\Gamma(\nu+n)} e^{-(\beta+\nu)\theta_x} \theta_x^{\nu+n-1}, \end{aligned}$$

for $x = 2,\ldots,I$. This expansion has infinitely many terms, but can be approximated by the sum of the $K+1$ first terms. In other words, we neglect the possibility of the count variable N_{x-1} taking a value larger than K. Thus we get:

$$\begin{aligned} l(\theta_x|\theta_{x-1}) &= \sum_{n=0}^{\infty} \mathbb{P}[N_{x-1}=n] l(\theta_x|n,\theta_{x-1}) \\ &= \sum_{n=0}^{\infty} \frac{e^{-\beta\theta_{x-1}}(\beta\theta_{x-1})^n}{\Gamma(n+1)} \frac{(\beta+\nu)^{\nu+n}}{\Gamma(\nu+n)} e^{-(\beta+\nu)\theta_x} \theta_x^{\nu+n-1} \\ &\approx \sum_{n=0}^{K} \frac{e^{-\beta\theta_{x-1}}(\beta\theta_{x-1})^n}{\Gamma(n+1)} \frac{(\beta+\nu)^{\nu+n}}{\Gamma(\nu+n)} e^{-(\beta+\nu)\theta_x} \theta_x^{\nu+n-1} \\ &\equiv F'(\theta_{x-1}) G(\theta_x), \end{aligned} \tag{4.1}$$

where F and G are both $(K+1)$ dimensional vectors given by

$$\begin{aligned} F(x) &= \left(\frac{e^{-\beta\theta_{x-1}}(\beta\theta_{x-1})^0}{\Gamma(0+1)}, \ldots, \frac{e^{-\beta\theta_{x-1}}(\beta\theta_{x-1})^K}{\Gamma(K+1)} \right)', \\ G(x) &= \left(\frac{(\beta+\nu)^{\nu+0}}{\Gamma(\nu+0)} e^{-(\beta+\nu)\theta_x} \theta_x^{\nu+0-1}, \ldots, \frac{(\beta+\nu)^{\nu+K}}{\Gamma(\nu+K)} e^{-(\beta+\nu)\theta_x} \theta_x^{\nu+K-1} \right)'. \end{aligned}$$

Clearly, when K goes to infinity, the approximation formula becomes exact. In practice, the value of K is fixed to be mildly large to ensure a good quality of approximation. The choice of K will be discussed in Appendix 3. Given a fixed K, the conditional distribution, $l(\theta_x|\underline{D_I})$, can be computed using an algorithm similar to the standard Hamilton filter for Markov switching model. The posterior densities are summarized in the following proposition.

Proposition 2:

$$l(\theta_x|\underline{D_I}) = \frac{P'(I-1) l(D_I|\theta_x) G(\theta_x)}{P'(I-1) Q(I)}, \qquad \text{if} \quad x = I, \tag{4.2}$$

and

$$l(\theta_x|\underline{D_I}) = \frac{P'(x-1) G(\theta_x) F'(\theta_x) l(D_x|\theta_x) Q(x+1)}{P'(I-1) Q(I)}, \qquad \text{if} \quad x < I, \tag{4.3}$$

where:

- $P'(0)$ *is the row vector of the marginal distribution of* N_x*, i.e. a negative binomial distribution truncated at* K*.*

$$P_j(0) \propto C_j^{j+\nu-1} p^{\nu} (1-p)^j, \quad j = 0,\ldots,K, \tag{4.4}$$

  where $p = 1-\rho$*.*
- *The subsequent value of* $P'(x)$ *is obtained recursively and forward by:*

$$P'(x) = P'(x-1) M_x,$$

- *The sequence $Q(x)$ is obtained recursively and backward by:*

$$Q(I) = \int_0^\infty l(D_I|\theta_I)G(\theta_I)\mathrm{d}\theta_I$$
$$Q(x-1) = M_x Q(x).$$

- *the age-indexed matrix M_x is given by:*

$$M_x = \int_0^\infty G(\theta_x)F'(\theta_x)l(D_x|\theta_x)\mathrm{d}\theta_x. \tag{4.5}$$

This approach has recently been suggested by Creal (2017). For the sake of completeness, we provide the proof of this proposition in Appendix 1. The interpretation of this result is that given $\underline{D_I}$, the process $\boldsymbol{N}$ is still a (time-inhomogeneous) Markov chain, with transition matrix from N_{x-1} to N_x given by, up to a multiple constant, the matrix M_x. Moreover, the vectors $P(x)$ and $Q(x)$ are, up to appropriate normalization, the vector of probabilities of N_x given $\underline{D_x} = D_1, \ldots, D_x$ and $\overline{D_x} = D_x, D_{x+1}, \ldots, D_I$, respectively.

Thus, based on Proposition 2, the computation of the smoothing density necessitates only the computation of the matrices M_x, whose (i,j)th entry has the following explicit expression:

$$\begin{aligned} M_x(i,j) &= \int_0^\infty \frac{(\beta+\nu)^{\nu+i}}{\Gamma(\nu+i)} e^{-(\beta+\nu)\theta_x}\theta_x^{\nu+i-1}\frac{e^{-\beta\theta_x}(\beta\theta_x)^j}{\Gamma(j+1)}\theta_x^{\sum_{t=1}^T d_{x,t}}\exp\left(-\theta_x\sum_{t=1}^T e_{x,t}\mu_{0,x,t}\right)\mathrm{d}\theta_x \\ &= \frac{(\beta+\nu)^{\nu+i}\beta^j}{\Gamma(\nu+i)\Gamma(j+1)}\frac{\Gamma(\nu+i+j+\sum_{t=1}^T d_{x,t})}{\left(\beta+\nu+\beta+\sum_{t=1}^T e_{x,t}\mu_{0,x,t}\right)^{\nu+i+j+\sum_{t=1}^T d_{x,t}}}, \qquad \forall i,j=0,\ldots,K. \end{aligned} \tag{4.6}$$

We can also remark that the smoothing density $l(\theta_x|\underline{D_I})$ is a linear combination of gamma densities $\gamma(\delta+k)$, where k varies between 0 and $2K$. As a consequence, the lth moment of this smoothing density $\mathbb{E}[\theta_x|\underline{D_I}]$ is:

$$\mathbb{E}[\theta_x^l|\underline{D_I}] = \frac{P'(x-1)M_{l,x}Q(x+1)}{P'(I-1)Q(I)}, \qquad \text{with} \qquad M_{l,x} = \int_0^\infty \theta_x^l G(\theta_x)F'(\theta_x)l(D_x|\theta_x)\mathrm{d}\theta_x. \tag{4.7}$$

4.2. Predictive mean and variance of the one-year-ahead death count

From Equation (4.7), we obtain the posterior prediction of the one-year-ahead *total* death count:

$$\mathbb{E}\Big[\sum_x d_{x,T+1}|\underline{D_I}\Big] = \sum_x \mu_{0,x,T+1}e_{x,T+1}\mathbb{E}\Big[\theta_x|\underline{D_I}\Big]. \tag{4.8}$$

For the conditional variance, we first decompose it into the sum of the conditional variance of individual $d_{x,T+1}$'s and the covariance of $d_{x,T+1}$ for different x:

$$\mathbb{V}\left[\sum_x d_{x,T+1}|\underline{D_I}\right] = \sum_x \mathbb{V}[d_{x,T+1}|\underline{D_I}] + 2\sum_{x<x'}\mathrm{CoV}[d_{x,T+1}, d_{x',T+1}|\underline{D_I}]. \tag{4.9}$$

By the law of total variance, the variance term $\mathbb{V}[d_{x,T+1}|\underline{D_I}]$ can be further decomposed into:

$$\begin{aligned}\mathbb{V}\Big[d_{x,T+1}|\underline{D_I}\Big] &= \mathbb{E}\Big[\mathbb{V}(d_{x,T+1}|\theta_x,\underline{D_I})|\underline{D_I}\Big] + \mathbb{V}\Big[\mathbb{E}(d_{x,T+1}|\theta_x,\underline{D_I})|\underline{D_I}\Big] \\ &= \mathbb{E}\Big[\mu_{0,x,T+1}e_{x,T+1}\theta_x|\underline{D_I}\Big] + \mathbb{V}[\mu_{0,x,T+1}e_{x,T+1}\theta_x|\underline{D_I}]. \end{aligned} \tag{4.10}$$

The first term on the RHS is the contribution of the idiosyncratic variation, and can be derived from (4.7), whereas the second term is the contribution of the uncertainty around the unobserved mortality coefficient θ_x, and is equal to:

$$\mathbb{V}[\mu_{0,x,T+1}e_{x,T+1}\theta_x|\underline{D_I}] = {\mu_{0,x,T+1}}^2 e_{x,T+1}^2\Big(\mathbb{E}[\theta_x|\underline{D_I}]^2 - \mathbb{E}[\theta_x^2|\underline{D_I}]\Big).$$

Note that this formula has been derived under the implicit assumption that the future value of $\mu_{0,x,T+1}$ is deterministically given. This can be motivated, for instance, when $\mu_{0,x,T+1}$ is given by the regulator, which usually does not provide uncertainties around the mortality table. In many situations, however, uncertainties in longevity improvements of the benchmark population should be taken into account when forecasting death counts in the insurance portfolio. Accounting for uncertainties of the benchmark mortality rates is straightforward, under the assumption that $\mu_{0,x,T+1}$ is independent of θ_x for each x. In this case, we have:

$$\mathbb{V}[\mu_{0,x,T+1}e_{x,T+1}\theta_x|\underline{D_I},\mathcal{F}_T] = e_{x,T+1}^2\Big(\mathbb{E}[\mu_{0,x,T+1}|\mathcal{F}_T]^2\mathbb{E}[\theta_x|\underline{d_I}]^2 - \mathbb{E}[{\mu_{0,x,T+1}}^2|\mathcal{F}_T]\mathbb{E}[\theta_x^2|\underline{D_x}]\Big), \tag{4.11}$$

where $\mathcal{F}_T$ denotes the mortality history of the benchmark population, i.e. $(\mu_{0,x,t})_{x=1,\dots,I,t=1,\dots,T}$.The independence assumption between each $\mu_{0,x,T+1}$ and θ_x indicates that differences in mortality level between the two populations are stable, regardless the mortality improvement in the benchmark population. In other words, the insurance portfolio has the same mortality trend as the benchmark population, i.e. mortality trends in the two populations are coherent (Li and Lee 2005). Under the independence assumption, computing the variance of the total death counts in the portfolio amounts to computing the first two conditional moments of $\mu_{0,x,T+1}$ for all x. This can be conducted by simulation when a stochastic mortality model, such as the Lee and Carter (1992) model, is specified for the benchmark population.

Similarly, the covariance term in (4.9) is given by:

$$\begin{aligned}\mathrm{Cov}[d_{x,T+1},d_{x',T+1}|\underline{d_I}] &= \mu_{0,x,T+1}e_{x,T+1}\mu_{0,x',T+1}e_{x',T+1}\mathrm{Cov}[\theta_x,\theta_{x'}|\underline{D_I}] \\ &= \mu_{0,x,T+1}e_{x,T+1}\mu_{0,x',T+1}e_{x',T+1}\Big(\mathbb{E}[\theta_x\theta_{x'}|\underline{D_I}] - \mathbb{E}[\theta_x|\underline{d_I}]\mathbb{E}[\theta_{x'}|\underline{D_I}]\Big). \end{aligned} \tag{4.12}$$

Thus, the computation of the covariance term involves solving for $\mathbb{E}[\theta_x\theta_{x'}|\underline{D_I}]$, which is given in the following proposition.

Proposition 3: *For two ages $x < x'$, $\mathbb{E}[\theta_x\theta_{x'}|\underline{D_I}]$ is given by:*

$$\mathbb{E}[\theta_x\theta_{x'}|\underline{D_I}] = \frac{P'(0)\Big[M_1\cdots M_{x-1}\Big]M_{1,x}\Big[M_{x+1}\cdots M_{x'-1}\Big]M_{1,x'}\Big[M_{x'+1}\cdots M_{I-1}\Big]\Big[\int_0^\infty l(d_I|\theta_I)G(\theta_I)\mathrm{d}\theta_I\Big]}{P'(0)\Big[M_1\cdots M_{I-1}\Big]\Big[\int_0^\infty l(d_I|\theta_I)G(\theta_I)\mathrm{d}\theta_I\Big]},$$

with M_x and $M_{1,x}$ given in Equations (4.5) and (4.7), respectively.

Proof: See Appendix 2. □

Finally, when uncertainties of $\mu_{0,x,T+1}$ are incorporated, the covariance term becomes:

$$\begin{aligned}\mathrm{Cov}[d_{x,T+1}, d_{x',T+1}|\underline{d_I}] = {} & e_{x,T+1}e_{x',T+1}\mathbb{E}[\mu_{0,x,T+1}\mu_{0,x',T+1}|\mathcal{F}_T]\mathbb{E}[\theta_x\theta_{x'}|\underline{D_I}] \\ & - e_{x,T+1}e_{x',T+1}\mathbb{E}[\mu_{0,x,T+1}|\mathcal{F}_T]\mathbb{E}[\mu_{0,x',T+1}|\mathcal{F}_T]\mathbb{E}[\theta_x|\underline{D_I}]\mathbb{E}[\theta_{x'}|\underline{D_I}].\end{aligned} \tag{4.13}$$

4.3. Accounting for heterogeneous policy face values

Besides the number of death, we can also calculate the posterior moments of the total claim amount. For illustration, we consider a portfolio of policies only with death benefits. In the simplest scenario where all policies have the same face value, multiplying the forecast of the total death count by this face value (resp. the squared face value) yields the next year's expected total claim amount (resp. the variance of the claim).

In practice, however, insurance policies have different face values. Suppose that there are $e_{x,t}$ policies with age x at year t. Their face values are chosen independently by their policyholders, and we denote by $\mathcal{V}_{x,t}$ the set of these $e_{x,t}$ observed face values. Then the sum of the face values of the $d_{x,t}$ death events at age x and time t, $A_{x,t}$, has the compound Poisson representation:

$$A_{x,t} = \sum_{i=1}^{d_{x,t}} z_{i,x,t}, \tag{4.14}$$

where $z_{i,x,t}$ are mutually exclusive draws from the set of face values $\mathcal{V}_{x,t}$. When $e_{x,t}$ is large, these draws can be approximately regarded as i.i.d., and the mean (resp. variance) of each $z_{i,x,t}$ is well approximated by the empirical mean $m_{x,t}$ (resp. empirical variance $\sigma^2_{x,t}$) of these observed face values, which are easily computable. Thus, we have:

$$\mathbb{E}[A_{x,t+1}|\underline{D_I}] = m_{x,t}\mathbb{E}[d_{x,t+1}|\underline{D_I}]$$

and

$$\begin{aligned}\mathbb{V}[A_{x,t+1}|\underline{D_I}] &= \mathbb{V}\Big[\mathbb{E}[A_{x,t+1}|d_{x,t+1}, \underline{D_I}]|\underline{D_I}\Big] + \mathbb{E}\Big[\mathbb{V}[A_{x,t+1}|d_{x,t+1}, \underline{D_I}]|\underline{D_I}\Big] \\ &= m^2_{x,t}\mathbb{V}[d_{x,t+1}|\underline{D_I}] + \sigma^2_{x,t}\mathbb{E}[d_{x,t+1}|\underline{D_I}].\end{aligned}$$

Similarly, the covariance is given by:

$$\mathrm{Cov}[A_{x,t+1}, A_{x',t+1}|\underline{D_I}] = m_{x,t}m_{x',t}\mathrm{Cov}[d_{x,t+1}, d_{x',t+1}|\underline{D_I}].$$

In practice, $m_{x,t}$ and $\sigma^2_{x,t}$ are unknown, but can be estimated from historical claim experience.

Finally, the third and fourth conditional moments of $\sum_{x=1}^{I} A_{x,t}$ given $\underline{D_I}$ can also be obtained in a similar manner, using the law of total cumulance. These computations involve the third and fourth order (co-)moments of d_x's (and hence those of θ_x's), which have closed form expressions as well (see Section 4.2). They can be used to approximate the conditional distribution of the total claim amount using the moment matching technique (see e.g. Ghysels and Wang 2014). Besides, one can also obtain the posterior distribution of the total claim amount via simulation, which is discussed in the next subsection.

4.4. Simulating trajectories of θ

Our previous closed form predictive formulae concern the mean and variance of the next year's death count. However, other risk metrics, such as Value-at-Risk (VaR) and Conditional-Tail-Expectation (CTE), may also be interesting in solvency calculation. These risk metrics can be calculated using simulation. This can be done by first simulating trajectories from the posterior distributions of $\boldsymbol{\theta}$, then, for each simulated $\boldsymbol{\theta}$, simulating the number of death for each age. Furthermore, the simulation exercise is useful when the forecast horizon is longer than one year, which is particularly relevant for annuity or pension businesses.

Simulation of the posterior trajectory in standard state space models is based on computationally intensive particle filter, or MCMC (see e.g. Pedroza 2006). In our model, the auxiliary counting variable can be used to simplify the sampling procedure (see Nieto-Barajas and Walker 2002, Creal 2017). The basic idea is that, in order to simulate trajectories of $\boldsymbol{\theta}$, we first simulate a trajectory of $\boldsymbol{N}$ from its posterior distribution. Simulating trajectories of $\boldsymbol{N}$ is simpler, since each N_x is discrete and has an (approximated) finite state space $\{0, \ldots, K\}$. Then, in the second step, we simulate each θ_x conditional on d_x, N_x, and N_{x+1}. The detailed sampling algorithm is provided in Appendix 4.

5. Application

In this section, we fit the ARG model to a portfolio of endowment policies (that is, policies with death benefits) from an anonymous French life insurance company. The data contain about 15,000 male policyholders over an observation window of 11 years (1997–2007). The portfolio is not closed, that is, new policyholders can enter the portfolio after the inception date, while existing policyholders can choose to lapse their contract and leave the portfolio at any time during the observation period. In this paper, the French national male mortality data during the same period is used as the benchmark mortality table. The benchmark mortality data are downloaded from the Human Mortality Database.[4] For illustration purpose, mortality development of the benchmark population is assumed to be deterministic.

5.1. Data

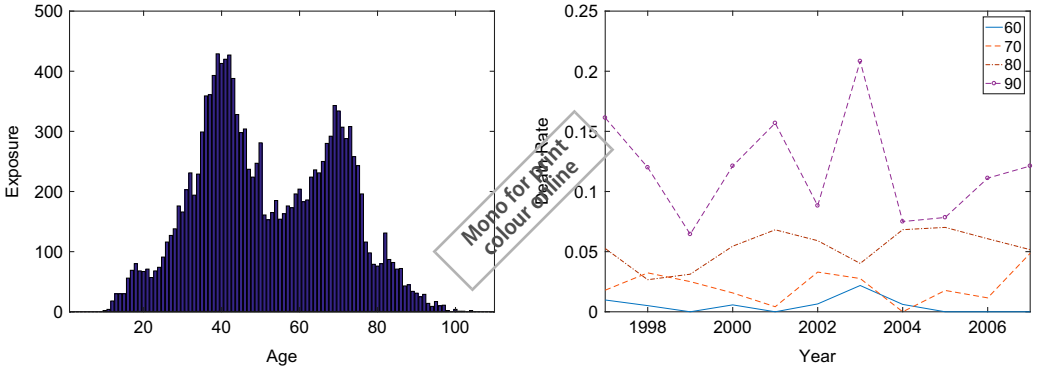
Figure 2(a) displays the age-specific exposure of the portfolio in year 1997. The exposure is adjusted for lapsation.[5] Policyholders' ages range between 10 and 102, and has a two-modal distribution. We see that the majority of policyholders are active lives. However, the second peak at around 70 indicates that recently retired people tend to buy endowment policies as well, possibly because of heritage motive. In particular, the proportion of policyholders above age 65 in the portfolio is around 27% in year 1997, while the analogous number is 11% in the French male population. Therefore, uncertainties regarding the total number of deaths in the portfolio is likely to be substantial, due to the higher elderly proportion. Figure 2(b) displays the age-specific mortality rates of the portfolio from year 1997 to 2007. Due to higher longevity risks, the mortality rates are more erratic at older ages. Moreover, the small exposures at these ages could also contribute to the uncertainties. Finally, no clear mortality improvement is observed in the insurance portfolio, probably due to the short observation window. This motivates the use of a benchmark population with reliable mortality trend.

5.2. Impact of the age dependence

Since $\boldsymbol{\theta}$ are correlated, the posterior distribution of each θ_x depends not only on D_x, but also on other ages. To illustrate the impact of the auto-correlation coefficient ρ, we plot in Figure 3 the posterior mean of the process $(\theta_x)_{x=1,\ldots,110}$ for $\rho = 0.1, 0.9$, and $\nu = 10$. For comparison, the posterior means

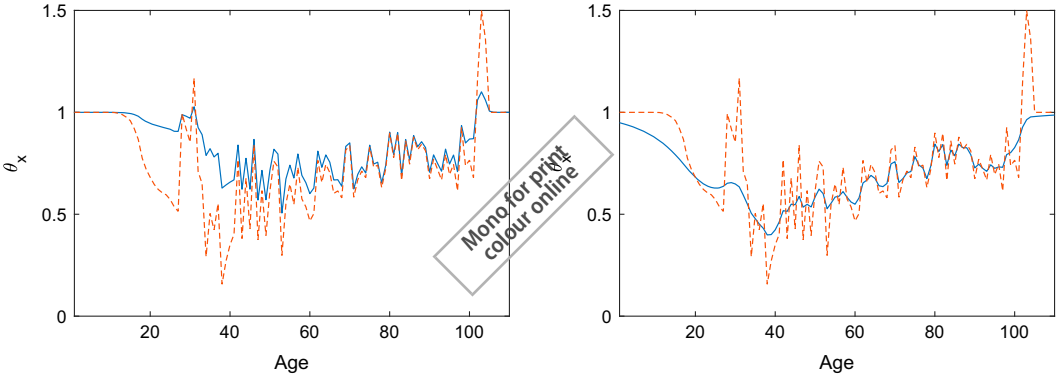
[4] See: http://www.mortality.org/. The application of the ARG model is not restricted to the national mortality table. Regulatory or industrial aggregate mortality tables can also be used as the benchmark mortality table.

[5] For example, if a policyholder lapses his/her policy at sometime between 1 January and 31 December in year t, the year t exposure only counts the time elapsed between 1 January and the time of lapse, but not one entire year.



(a) Age-specific exposure in year 1997. (b) Age-specific mortality rates from year 1997 to 2007.

Figure 2. The exposure and mortality rates of the portfolio at different ages and different years.



(a) $\rho = 0.1$ (b) $\rho = 0.9$

Figure 3. Posterior mean of θ_x with age dependence (full line, corresponding to $\rho \neq 0$) and without age dependence (dashed line, that is, $\rho = 0$) with $\nu = 10$.

of θ_x under the standard Poisson-gamma model ($\rho = 0$) are shown in each panel as well. We let $K = 200$ for the truncation of the Poisson distribution, as we find that in both models, the probability of N_x taking values larger than 200 is negligible (see Appendix 3).

Several observations can be made from Figure 3. First, the posterior means of θ_x are smaller than 1 for most ages and all three model specifications. This is reasonable, since mortality rates of life insurance policyholders are in general lower than those of the national average. Second, the posterior mean of θ_x converges to 1 when x reaches either the minimal age, 1, or the maximal age, 110. This is due to the fact that exposure in the portfolio decreases drastically when approaching these limiting ages, and thus the posterior means of θ_x's are converging to their prior means (see Equation (3.5)). Third, the convergence speed is the fastest in the independent case, and the slowest with $\rho = 0.9$. The reason is that the correlation structure of $\boldsymbol{\theta}$ leads to a smoothing effect on their posterior means. Moreover, this effect is more substantial at ages with a small exposure and when ρ is larger. In particular, for $\rho = 0.9$, the spikes on the independent trajectory (dashed line) near age 30 and 105 have almost been eliminated.

Table 1. The posterior predictive mean, variance, and 99.5% value-at-risk of the total number of deaths next year in the portfolio using different values of ρ and ν. The predictive mean (row 1) and variance (row 2) are obtained from closed-form expressions derived in Section 4, while the VaR (row 3) is based on simulation.

ν	1			5			10		
ρ	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
$\mathbb{E}\left[\sum_x d_{x,T+1}\vert\underline{D_I}\right]$	237.69	237.60	237.01	243.52	240.31	237.61	249.13	243.32	237.95
$\mathbb{V}\left[\sum_x d_{x,T+1}\vert\underline{D_I}\right]$	266.27	266.57	266.95	270.84	268.35	265.99	275.18	270.54	265.91
VaR	285	283	282	288	287	286	296	295	293

5.3. Projected total number of deaths

After forecasting $\boldsymbol{\theta}$, we forecast the one-year-ahead total death count. In Table 1, we show, for different values of ρ and ν, the posterior mean, variance, and 95% VaR of next year's total number of deaths. The first two are calculated using the closed form formulae in Sections 4.1 and 4.2. The 95% VaR is obtained by 100,000 simulated trajectories of $\boldsymbol{\theta}$ based on the sampling algorithm in Section 4.4.

Generically speaking, when ρ is fixed, the predictive mean and variance are increasing in ν. The reason is that ν is a belief parameter and measures the weight assigned to the benchmark mortality, which has in general higher age-specific mortality rates (see Equation (2.4)). Therefore, the larger ν, the larger the weight of the benchmark mortality, and so are the forecasted $\boldsymbol{\theta}$. Moreover, with fixed ν, we see that the predictive mean and variance are decreasing in ρ. Indeed, larger ρ implies stronger correlation between θ_x's, and thus lower posterior means for the θ_x's with small exposure. As a result, the projected mortality rates in the insurance portfolio become smaller when ρ increases.

5.3.1. Variance decomposition of the total death count

Besides projecting the mean and variance of the insurance portfolio's death count in the next period, the ARG model allows us to evaluate the extent to which parameter uncertainty on $\boldsymbol{\theta}$ affects the projection. Specifically, the conditional variance of the total number of deaths is given by:

$$\mathbb{V}\left[\sum_x d_{x,T+1}\vert\underline{D_I}\right] = \mathbb{E}\left[\sum_x d_{x,T+1}\vert\underline{D_I}\right] + \left(\mathbb{V}\left[\sum_x d_{x,T+1}\vert\underline{D_I}\right] - \mathbb{E}\left[\sum_x d_{x,T+1}\vert\underline{D_I}\right]\right). \tag{5.1}$$

The first part on the right hand side of Equation (5.1) corresponds to the (diversifiable) contribution from the sampling variation, whereas the second part is the (non-diversifiable) contribution from the uncertainty on θ_x, i.e. the *parameter uncertainty*. By law of large numbers, the diversifiable part is decreasing (relative to the conditional mean) in the size of the portfolio. As for the second, non-diversifiable part, when the portfolio size and/or the observation window increases, this uncertainty of the underlying mortality rates becomes smaller as well.[6]

Table 1 shows that, although the sampling variation is larger, the contribution of the uncertainty on θ_x accounts for, roughly speaking, 12% of the estimated number of death, $\mathbb{E}\left[\sum_x d_{x,T+1}\vert\underline{D_I}\right]$. This number is far from negligible, even when the data covers an observation window of 11 years, which is relatively long for one insurance product. Therefore, by the arguments of Section 2 (see e.g. Equation (2.3)), we can expect even larger contribution from the uncertainty of θ_x when the observation window is shorter.

To further illustrate the realized economy due to the decrease of the parameter uncertainty, we estimate the model using different observation windows, and compare the results. In particular,

[6] A similar conclusion has been reached in Hardy and Panjer (1998), who report that a large company (say, with 1 million policies) can release up to 50 % of the regulatory capital compared to a small company (say, with 50,000 policies). See also equation (2.5) for a discussion.

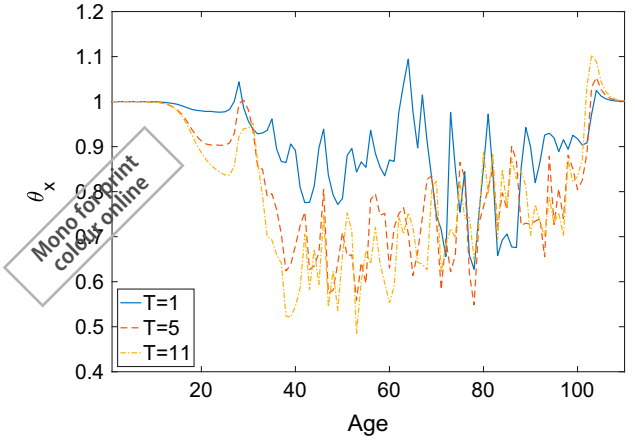


Figure 4. Posterior mean of θ_x for different lengths of observation window: $T = 1$ in full line, $T = 5$ in dashed line, and $T = 11$ in dot-dash line.

for different T's smaller than the maximal possible value $T_{\max} = 11$, we use the first T years of observations to estimate $\mathbb{E}[\theta_x|\underline{D_I}]$, then compute the predictive mean and variance of the total death count at year $T + 1$.

Figure 4 displays the posterior mean of $\boldsymbol{\theta}$ with $T = 1$, 5, and 11. We see that the posterior mean of $\boldsymbol{\theta}$ becomes smaller when T increases. This result is intuitive, since as the observation accumulates, portfolio's own mortality rate, which is lower than the benchmark population, has larger impact on the estimation of $\boldsymbol{\theta}$.

Let us now examine how the predictive mean and variance of the total number of deaths evolves when the observation window increases from $T = 1$ to $T = 11$. Figure 5 shows the predictive mean and variance of next year's death count as a function of T. First, we observe a hump shape for both the mean and the variance. This is the result of two opposite effects. On the one hand, total exposure in the portfolio decreased by around 15% over the 11 years, which would reduce the projected death count. On the other hand, the average age of policyholders had been increasing over time, which lead to higher average mortality rates. More importantly, we see that increasing the observation window clearly reduces the predictive variance relative to the predictive mean, especially when T is small. In other words, the contribution of parameter uncertainty to the variance of total death count becomes less substantial as observation accumulates.

Finally, we illustrate the simulation of the trajectories of $\boldsymbol{\theta}$. Figure 6 displays 100 simulated trajectories of $\boldsymbol{\theta}$ drawn from the posterior distribution $l(\boldsymbol{\theta}|\underline{D_I})$, with $\nu = 10$ and $\rho = 0.5$. We observe that the average of these trajectories correspond roughly to Figure 3(b), i.e. $\mathbb{E}[\theta_x|\underline{D_I}]$. Moreover, as expected, the trajectories are more erratic at lower and higher ages, due to the lack of exposure.

5.4. Implications on regulation and reinsurance pricing

Let us now discuss the implications of the ARG model on regulations and pricing, as well as how to determine the values of ν and ρ. In this paper, we assume that ν and ρ are exogenously given. In a **frequentist** framework, however, it is possible to estimate ν and ρ from the insurance portfolio. Roughly speaking, at age x, θ_x is approximately the ratio between the historical death counts of the insurance portfolio and the benchmark population. Therefore, a reasonable proxy of θ_x would be $\frac{\sum_t d_{x,t}}{\sum_t \mu_{x,t} e_{x,t}}$. Using this approximation, we can calibrate the values of ν and ρ.

Nevertheless, from the regulatory point of view, this frequentist approach is not recommended. Indeed, the benchmark mortality table for life insurance policies (resp. life annuities) is often a 'conservative' table which overestimates (resp. underestimates) the mortality rates of the portfolio. Thus if the previous estimation approach is directly applied, we would get raw estimate of θ_x which

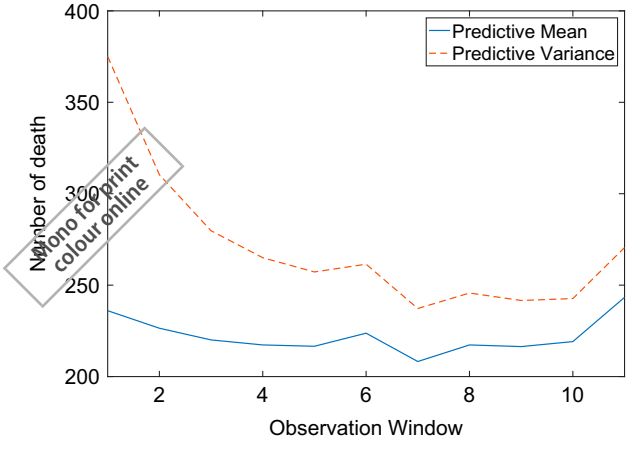


Figure 5. Predictive mean (in full line) and variance (in dashed line) of total death count in the next year, for different lengths of observation window.

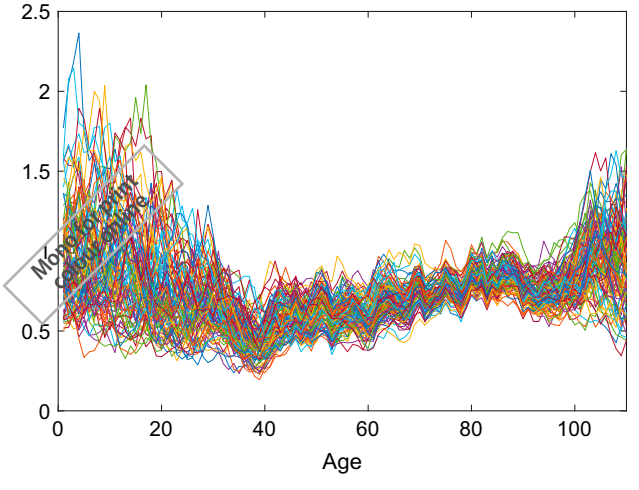


Figure 6. 100 simulated trajectories of $\boldsymbol{\theta}$ with $\nu = 10$ and $\rho = 0.9$.

are usually smaller than 1 for a life insurance portfolio. This is contrary to our prior assumption that $\mathbb{E}[\theta_x] = 1$.

So why do we make, at the beginning, this unitary mean assumption? Indeed the possibility for the insurer to set its own proprietary life table has to be monitored very carefully by the regulator, in order to prevent them from regulatory arbitrage. Without the prior assumption $\mathbb{E}[\theta_x] = 1$, it would be very tempting for insurers to apply the ARG model, or other similar models, to obtain a proprietary life table, so long as the portfolio mortality level seems to be lower than that implied by the regulatory table. regardless of the reliability of such a conclusion. Using such proprietary life tables is very risky if the observed mortality difference does not hold in the future. In fact, the use of proprietary table should only be allowed when there is strong evidence that the risks of the insurer's portfolio is significantly lower than implied by the benchmark life table. This strong evidence might be the fact that the portfolio is sufficiently large, or that a lower mortality level has been observed over a sufficiently long period.

The **Bayesian** approach proposed in this paper is more adapted, as it allows to automatically evaluate the reliability of a portfolio's own mortality experience. In our approach, the prior mean,

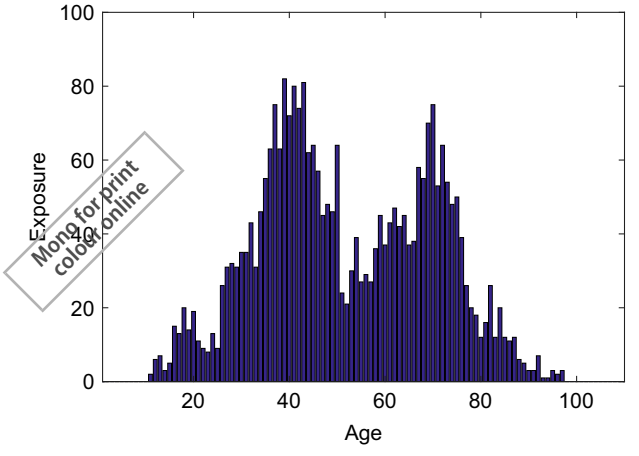


Figure 7. The age-specific exposure in the small portfolio in year 1997.

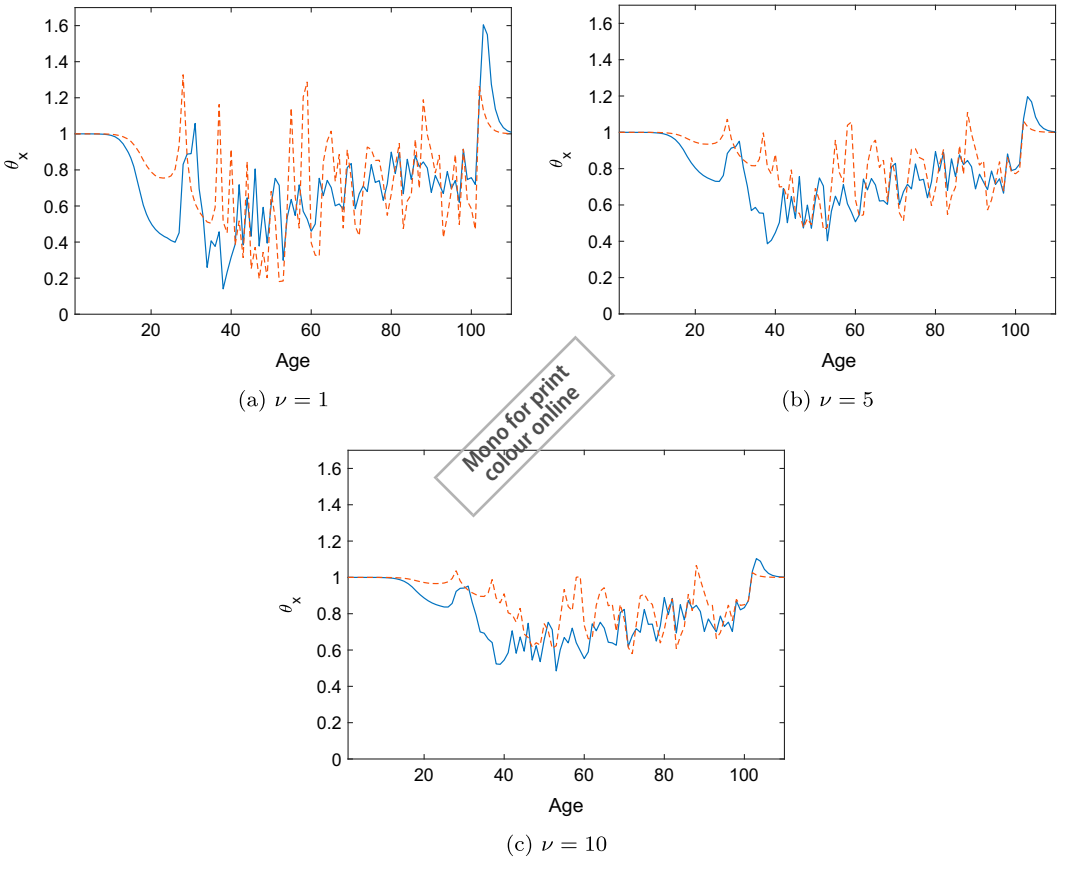


Figure 8. Posterior mean of $\boldsymbol{\theta}$ from the whole portfolio (full line) and the small portfolio (dashed line) for different values of ν with $\rho = 0.5$.

$\mathbb{E}[\theta_x]$, is fixed rather than estimated. As a result, even when the portfolio's mortality rates seem to be lower, we would still get the posterior means of θ_x's close to 1 if *i*) the portfolio size is not sufficiently large; *ii*) or the observation window is not long enough. Contrarily, the predicted mortality would be close to the insurer's own mortality experience when the portfolio size is large, or when the observation window is long. In other words, under the Bayesian approach, the more reliable the proprietary mortality experience, to a larger extent the insurer can take it into account when building the internal model.

Therefore, to ensure coherence, the values of ρ and ν, which characterize the prior distribution of $\boldsymbol{\theta}$, should be fixed by the regulator at the industry level, rather than be estimated for each portfolio, and their values will reflect the regulator's risk aversion (via parameter ν), as well as its emphasis on smooth mortality forecasts (via parameter ρ). This approach is also in line with the existing literature (see e.g. Olivieri and Pitacco 2012, van Berkum *et al.* 2017).

Besides solvency calculation, our model is also useful for reinsurance pricing, in which case the role of the regulator is replaced by the reinsurer, and the regulatory mortality table is replaced by the reference pricing table of the reinsurer. In this case, the values of ρ and ν could be estimated by the reinsurer, as it has access to a wide range of insurance portfolios. This is the approach adopted by Hardy and Panjer (1998) and Salhi *et al.* (2015). As a result, the Bayesian approach allows to derive reinsurance quotes that take into account both the portfolio's own mortality risk and the industry experience.

In the context of a reinsurance quote, the willingness of the reinsurer to price the contract using the insurer's own mortality experience depends on its knowledge of the latter. If the insurer's portfolio is small, then the reinsurer will put more credibility on a conservative reference mortality table. When the size of the insurer's portfolio becomes larger, the degree of the asymmetric information becomes smaller, and the reinsurer becomes more willing to take into account the insurer's own mortality experience. Note, also, that when an industry-level data is used, the conditions $\mathbb{E}[\theta_x] = 1$ for each x ensures that this pricing approach is fair for the reinsurer. That is to say, when the reinsurer underwrites lots of portfolios, it can expect that the average cost across all the portfolios is equal to that predicted by the benchmark mortality table.

To conclude this section, we provide an example to illustrate the impact of parameter ν and the portfolio size on the credibility of its own mortality experience. Specifically, we create a small portfolio by randomly drawing policyholders from the whole portfolio. The small portfolio consists of around 2,800 policyholders in year 1997, thus its size is about 20% of the whole portfolio. The age composition of the small portfolio is similar to the whole portfolio, and is given in Figure 7.

Figure 8 displays the posterior means of $\boldsymbol{\theta}$ for the whole portfolio and the small portfolio with different values of ν and $\rho = 0.5$.[7] First, we see that the posterior means from both portfolios are more volatile with a smaller ν. The reason is that, when the prior belief is weaker (with a smaller ν), the portfolio's own mortality experience has heavier influence on the posterior distribution of $\boldsymbol{\theta}$. Second, with $\nu = 5$ and 10, i.e. when the posterior means are less volatile, the whole portfolio has in general lower posterior means of θ_x. Therefore, given a fixed set of ρ and ν, a larger portfolio could gain more credibility on its own mortality experience in mortality forecasting.

6. Conclusion

This paper proposes a Bayesian non-parametric mortality model for a small population, where mortality of a larger population is exogenously given as the benchmark. The proposed model captures differences in mortality level between the small and the benchmark population by an autoregressive-gamma (ARG) process. Compared with existing deterministic models, the proposed model is more flexible whereas at the same time ensures smoothness of the mortality projection across different ages. Moreover, the model is associated with closed form expressions of the (posterior) moments, and a

[7] The comparison is qualitatively similar for different values of ρ.

computationally efficient sampling algorithm of the whole posterior predictive distribution, of future death counts of the small population. This makes the ARG model more accessible to practitioners.

In a numerical application, we apply the ARG model on a French life insurance portfolio. In particular, we examine how the portfolio size and the length of observation window would affect the weight assigned to the portfolio's own mortality experience in mortality forecasting. Further, we discuss the impact of the prior distribution, which measures the degree of belief on the benchmark mortality table, on the risk management of the insurance portfolio.

Finally, while the primary application in this paper relates to mortality forecasting of a small population, the proposed model can also be used as building block of a two-population Bayesian mortality model (see e.g. Antonio *et al.* 2015). Indeed, the estimation of Bayesian mortality models typically require substantial computational efforts (see also Czado *et al.* 2005), and the computational costs are even higher in the two-population case. The ARG model provides a tractable distribution for the conditional mortality rate of one population, when the mortality of the other population is assumed to be exogenously given. This could substantially reduce computational costs resulting from simultaneously modeling of two populations.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Yang Lu http://orcid.org/0000-0002-7141-1048

References

Antonio, K., Bardoutsos, A. & Ouburg, W. (2015). Bayesian Poisson log-bilinear models for mortality projections with multiple populations. *European Actuarial Journal* **5**(2), 245–281.

Bienvenüe, A. & Rullière, D. (2012). Iterative adjustment of survival functions by composed probability distortions. *The Geneva Risk and Insurance Review* **37**(2), 156–179.

Booth, H. & Tickle, L. (2008). Mortality modelling and forecasting: a review of methods. *Annals of Actuarial Science* **3**(1–2), 3–43.

Brouhns, N., Denuit, M. & Vermunt, J. (2002). A Poisson log-bilinear regression approach to the construction of projected life tables. *Insurance: Mathematics and Economics* **31**(3), 373–393.

Cadena, M. & Denuit, M. (2016). Semi-parametric accelerated hazard relational models with applications to mortality projections. *Insurance: Mathematics and Economics* **68**, 1–16.

Cairns, A. J. G., Blake, D. & 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. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D. & Khalaf-Allah, M. (2011a). Mortality density forecasts: an analysis of six stochastic mortality models. *Insurance: Mathematics and Economics* **48**(3), 355–367.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. & Khalaf-Allah, M. (2011b). Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin* **41**(1), 29–59.

Creal, D. D. (2017). A class of non-Gaussian state space models with exact likelihood inference. *Journal of Business & Economic Statistics* **35**(4).

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

Denuit, M., Devolder, P. & Goderniaux, A.-C. (2007). Securitization of longevity risk: pricing survivor bonds with Wang transform in the Lee-Carter framework. *Journal of Risk and Insurance* **74**(1), 87–113.

Dionne, G. & Vanasse, C. (1989). A generalization of automobile insurance rating models: the negative binomial distribution with a regression component. *Astin Bulletin* **19**(2), 199–212.

Dowd, K., Cairns, A. J. G., Blake, D., Coughlan, G. & Khalaf-Allah, M. (2011). A gravity model of mortality rates for two related populations. *North American Actuarial Journal* **15**(2), 334–356.

Ghysels, E. & Wang, F. (2014). Moment-implied densities: properties and applications. *Journal of Business & Economic Statistics* **32**(1), 88–111.

Gouriéroux, C. & Jasiak, J. (2001). State-space models with finite dimensional dependence. *Journal of Time Series Analysis* **22**(6), 665–678.

Gouriéroux, C. & Jasiak, J. (2006). Autoregressive gamma processes. *Journal of Forecasting* **25**(2), 129–152.

Hardy, M. R. & Panjer, H. H. (1998). A credibility approach to mortality risk. *Astin Bulletin* **28**(02), 269–283.
Himes, C. L., Preston, S. H. & Condran, G. A. (1994). A relational model of mortality at older ages in low mortality countries. *Population Studies* **48**(2), 269–291.
Koopman, S. J., Lucas, A. & Scharth, M. (2015). Numerically accelerated importance sampling for nonlinear non-Gaussian state-space models. *Journal of Business & Economic Statistics* **33**(1), 114–127.
Lee, R. & Carter, L. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association* **87**(419), 659–671.
Li, N. & Lee, R. (2005). Coherent mortality forecasts for a group of populations: an extension of the Lee-Carter method. *Demography* **42**(3), 575–594.
Lu, Y. (Forthcoming). Dynamic frailty count process in insurance: a unified framework for estimation, pricing, and forecasting. *Journal of Risk and Insurance*. DOI: 10.1111/jori.12190
Nieto-Barajas, L. E. & Walker, S. G. (2002). Markov beta and gamma processes for modelling hazard rates. *Scandinavian Journal of Statistics* **29**(3), 413–424.
Olivieri, A. & Pitacco, E. (2012). Life tables in actuarial models: from the deterministic setting to a Bayesian approach. *AStA Advances in Statistical Analysis* **96**(2), 127–153.
Pedroza, C. (2006). A Bayesian forecasting model: predicting US male mortality. *Biostatistics* **7**(4), 530–550.
Pitt, M. K. & Shephard, N. (1999). Filtering via simulation: auxiliary particle filters. *Journal of the American Statistical Association* **94**(446), 590–599.
Salhi, Y., Thérond, P.-E. & Tomas, J. (2015). A credibility approach of the Makeham mortality law. *European Actuarial Journal* **6**(1), 61–96.
Tomas, J. & 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.
Tomas, J. & Planchet, F. (2015). Prospective mortality tables: taking heterogeneity into account. *Insurance: Mathematics and Economics* **63**, 169–190.
van Berkum, F., Antonio, K. & Vellekoop, M. (2017). A Bayesian joint model for population and portfolio-specific mortality. *ASTIN Bulletin* **47**(3), 681–713.

Appendix 1. Proof of Proposition 2

The derivation of the smooth distribution is provided in Section 3.3 in Gouriéroux and Jasiak (2001) for a general class of models with *finite dimensional dependence*. For the sake of completeness, we provide in this appendix a simplified proof using the latent switching interpretation of the ARG process. The proof is decomposed into two parts. We first derive the filtering formula of $\boldsymbol{\theta}$ in Appendix A.1, then derive the smoothing formula in Appendix A.2.

A.1. The filtering formula

Let us compute recursively the filtering distribution $l(\theta_x|\underline{D_x})$, i.e. the conditional distribution of θ_x given deaths occurred at ages non larger than x. As discussed in the paper, it is easier to work with the discrete state variable N_x in the derivation. Therefore, we introduce the filtering distribution of N_x given death count observations up to age x:

$$P(D_x) := (\mathbb{P}[N_x = 0|\underline{D_x}], \ldots, \mathbb{P}[N_x = K|\underline{D_x}]). \tag{A1}$$

Then we can remark that the link between the filtering distribution of N_x and that of θ_x is the following:

$$l(\theta_x|\underline{D_x}) = \frac{P'(\underline{D_{x-1}})G(\theta_x)l(D_x|\theta_x)}{P'(\underline{D_{x-1}})\left[\int_0^\infty G(\theta_x)l(D_x|\theta_x)\mathrm{d}\theta_x\right]}. \tag{A2}$$

This formula is a direct consequence of the Bayes formula and thus its proof is omitted.

Hence, in order to obtain the filtering density, it suffices to obtain the vector of probabilities $P(\underline{D_x})$. Let us start by providing its initial value $P(0)$, when the conditioning set is degenerated. That is to say, $P(0)$ is the stationary distribution of N_x (truncated by upper bound K). Since θ_x has a gamma $\gamma(\nu, \nu)$ marginal distribution and N_x is conditionally Poisson $\mathcal{P}(\beta\theta_x)$, the marginal distribution of N_x is Negative Binomial with size parameter ν and probability of success $p = 1 - \rho$. Hence we have formula (4.4).

Let us now provide a simple recursive updating formula to update $P(\underline{D_x})$. We have the following property:

Lemma 1: *The recursive updating formula of the sequence* $(P'(\underline{D_x}))$ *is given by:*

$$P'(\underline{D_x}) \propto P'(\underline{D_{x-1}}) \underbrace{\int_0^\infty G(\theta_x)F'(\theta_x)l(D_x|\theta_x)\mathrm{d}\theta_x}_{:=M_x}, \tag{A3}$$

where the normalization constant to divide is to ensure that the components of $P'(\underline{D_x})$ sum up to unity.

In other words, $P'(\underline{D_x})$ is equal to, up to a multiplication constant, the vector $\overline{P'(x)}$ introduced in Proposition 2:

$$P'(x) \propto P'(\underline{D_x}), \qquad \forall x.$$

Equation (A3) is just a matrix formulation of the following elementary identity, obtained via the Bayes' rule:

$$\begin{aligned}\mathbb{P}[N_x = j|\underline{D_x}] &= \sum_{i=0}^{K} \mathbb{P}[N_x = j|N_{x-1} = i, D_x]\mathbb{P}[N_{x-1} = i|\underline{D_{x-1}}] \\ &\propto \sum_{i=0}^{K} \left[\int G_i(\theta_x)F_j(\theta_x)l(D_x|\theta_x)\mathrm{d}x\right]\mathbb{P}\left[N_{x-1} = i|\underline{D_{x-1}}\right].\end{aligned}$$

It is then easily checked that this formula leads to the filtering formula (4.2).

A.2. *The smoothing formula* (4.2)

Let us consider the joint distribution:

$$\begin{aligned}&l(D_I, \theta_I, D_{I-1}, \theta_{I-1}, \dots D_{x+1}, \theta_{x+1}, \theta_x|\underline{D_x}) \\ &\quad = l(\theta_x|\underline{D_x})l(\theta_{x+1}|\theta_x)l(D_{x+1}|\theta_{x+1})\cdots l(D_{I-1}|\theta_{I-1})l(\theta_I|\theta_{I-1})l(D_I|\theta_I) \\ &\quad = l(\theta_x|\underline{D_x})F'(\theta_x)G(\theta_{x+1})l(D_{x+1}|\theta_{x+1})F'(\theta_{x+1})G(\theta_{x+2})l(D_{x+2}|\theta_{x+2})\cdots F'(\theta_{I-1})G(\theta_I)l(D_I|\theta_I)\end{aligned} \tag{A4}$$

By integrating out $\theta_{x+1}, \dots, \theta_I$, we obtain:

$$\begin{aligned}&l(D_I, D_{I-1}, \dots, D_{x+1}, \theta_x|\underline{D_x}) \\ &\quad = \left[l(\theta_x|\underline{D_x})F'(\theta_x)\right]\left[\int_0^\infty G(\theta_{x+1})l(d_{x+1}|\theta_{x+1})F'(\theta_{x+1})\mathrm{d}\theta_{x+1}\right]\times\cdots \\ &\quad \left[\int_0^\infty G(\theta_{I-1})F'(\theta_{I-1})l(d_{I-1}|\theta_{I-1})\mathrm{d}\theta_{I-1}\right]\left[\int_0^\infty l(D_I|\theta_I)G(\theta_I)\mathrm{d}\theta_I\right],\end{aligned} \tag{A5}$$

where the term in the first pair of brackets is a row matrix, the term in the last pair of brackets is a column matrix, and all the intermediate terms are square matrices that have already been computed during the filtering stage.

Finally, by Bayes' formula, the smoothing density is equal to:

$$l(\theta_x|\underline{D_I}) = \frac{l(D_I, D_{I-1}, \dots, D_{x+1}, \theta_x|\underline{D_x})}{l(D_I, D_{I-1}, \dots, D_{x+1}|\underline{D_x})},$$

that is the ratio between the RHS of Equation (A5) and its integral with respect to θ_x.

Appendix 2. Proof of Proposition 3

It suffices to show that the joint smoothing distribution $l(\theta_x, \theta_{x'}|\underline{d_I})$ is equal to:

$$\begin{aligned}l(\theta_x, \theta_{x'}|\underline{D_I}) \propto P'(0)&\left[M_1\cdots M_{x-1}\right]G(\theta_x)F'(\theta_x)l(d_x|\theta_x)\left[M_{x+1}\cdots M_{x'-1}\right] \\ &\times G(\theta_{x'})F'(\theta_{x'})l(D_{x'}|\theta_{x'})\left[M_{x'+1}\cdots M_{I-1}\right]\left[\int_0^\infty l(D_I|\theta_I)G(\theta_I)\mathrm{d}\theta_I\right],\end{aligned} \tag{B1}$$

where the normalization constant is equal to $P'(0)\left[M_1\cdots M_{I-1}\right]\left[\int_0^\infty l(D_I|\theta_I)G(\theta_I)\mathrm{d}\theta_I\right]$.

The proof is similar as Proposition 2. Indeed, from Equation (A4), we integrate out $\theta_{x+1}, \cdots, \theta_{x'-1}, \theta_{x'+1}, \cdots, \theta_I$, to obtain the counterpart of Equation (A5). Then by Bayes' formula we have:

$$l(\theta_x, \theta_{x'}|\underline{D_I}) = \frac{l(D_I, D_{I-1}, \dots, D_{x+1}, \theta_x, \theta_{x'}|\underline{D_x})}{l(D_I, D_{I-1}, \dots, D_{x+1}|\underline{D_x})}.$$

Appendix 3. The choice of K

In this paper, we let K, the truncated upper bound of the stationary distribution of N_x, to be 200 for each x. Figure C1 displays the stationary distribution of N_x with $\nu = 1, 5, 10$, and $\rho = 0.9$. We see that $K = 200$ seems sufficient, as probability of N_x exceeding 200 is negligible for all ν's. From the relation $\beta = \frac{\nu\rho}{1-\rho}$ and Equation (4.4), we have that the prior distribution of N_x has a smaller mean and variance when either ρ or ν decreases. Therefore, $K = 200$ would be sufficient for the prior distribution of N_x with any pair of $(\rho, \nu) \leq (0.9, 10)$.

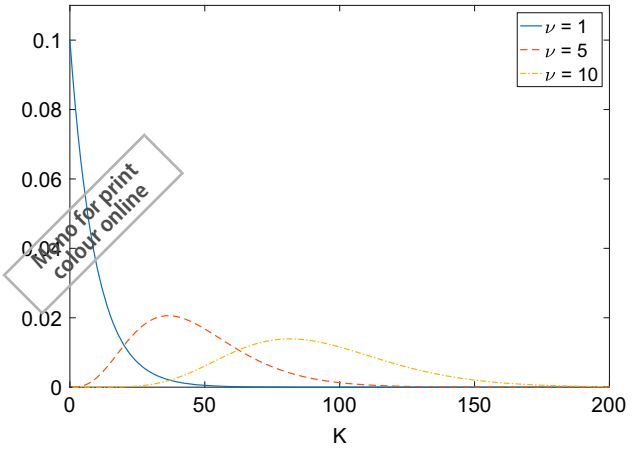
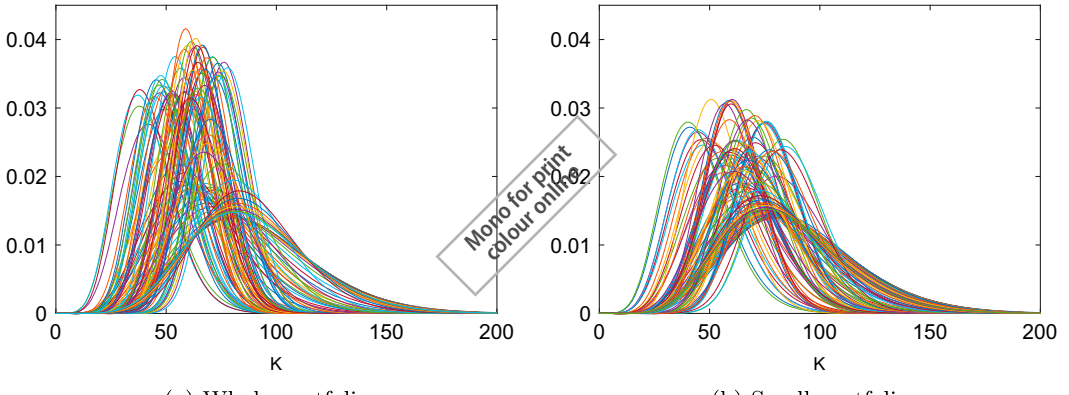


Figure C1. The stationary distribution of N_x with $\nu = 1, 5, 10$ and $\rho = 0.9$.

Figure C2 displays the posterior distribution of N_x for both the whole portfolio and the small portfolio with $\rho = 0.9$ and $\nu = 10$. We see that the posterior distribution of N_x becomes more concentrated for ages with nonzero observed number of deaths. For the rest of the choice of ρ and ν in the paper, the posterior distributions have smaller means and variances. Therefore $k = 200$ is also sufficient for the posterior distribution of N_x in our analysis.



(a) Whole portfolio (b) Small portfolio

Figure C2. Posterior distribution of N_x for the whole portfolio and the small portfolio with $\rho = 0.9$ and $\nu = 10$.

Appendix 4. Sampling of θ

In order to simulate trajectories of $\boldsymbol{\theta}$ from the posterior distribution $\boldsymbol{\theta}|\underline{D_I}$, let us first simulate trajectories of $\boldsymbol{N}$. This is conducted backwardly as follows.

Algorithm 1 (see Creal 2017, Section 3.3):

- *Draw N_{I-1} using the vector of elementary probabilities*

$$p_j \propto P_j(I-1)Q_j(I), \qquad \forall j = 0, \dots, K,$$

where vectors P and Q are introduced in Section 4.2. In other words, N_{I-1} is drawn from the distribution of $N_{I-1}|\underline{D_I}$.
- *For each $x \in [1, I-2]$ and given $N_{x+1}, \dots, N_I$, draw N_x using the vector of elementary probabilities:*

$$p_j \propto P_j(x) \int l(D_{x+1}|\underline{\theta_{x+1}})G_j(\theta_{x+1})\mathrm{d}\theta_{x+1} \times \int l(D_{x+1}|\theta_{x+1})G_j(\theta_{x+1})F_{N_{x+1}}(\theta_{x+1})\mathrm{d}\theta_{x+1}, \qquad j = 0, \dots, K.$$

In this equation, the term $P_j(x)\int l(D_{x+1}|\underline{\theta_{x+1}})G_j(\theta_{x+1})\mathrm{d}\theta_{x+1}$ is, up to normalization constant, the probability $\mathbb{P}[N_x = j|\underline{D_{x+1}}]$, whereas the second integral

$$\int l(D_{x+1}|\theta_{x+1})G_j(\theta_{x+1})F_{N_{x+1}}(\theta_{x+1})\mathrm{d}\theta_{x+1}$$

is the (N_{x+1}, j)th element of matrix M_{x+1}, that is (up to normalization constant), the transition probability of N_{x+1} given N_x and D_{x+1}.

Once we have a trajectory of $\boldsymbol{N}$, we can simulate a trajectories of $\boldsymbol{\theta}$ from the distribution $\boldsymbol{\theta}|\underline{N_I}, \underline{D_I}$. Using the Markov chain structure, we can remark that θ_x depends on $\underline{N_I}$ and $\underline{D_I}$ only via its left and right neighbours, N_x and N_{x+1} (except for the initial and the final value), as well as the death counts D_x. In particular, the sampling algorithm of $\boldsymbol{\theta}$ is summarized as follows.

Algorithm 2:

- *θ_I is sampled from the distribution:*

$$l(\theta_I|\underline{N_I}, \underline{D_I}) \propto G_{N_{I-1}}(\theta_I)l(D_I|\theta_I) \propto \exp\left(-(\beta+\nu)\theta_I - \theta_I \sum_{t=1}^{T} e_{I,t}\mu_{0,I,t}\right)\theta_I^{\nu+N_{I-1}+\sum_{t=1}^{T} d_{I,t}},$$

that is θ_I depends only on D_I and N_{I-1}, and has the gamma conditional distribution with shape parameter $\nu + N_{I-1} + \sum_{t=1}^{T} d_{I,t}$ and rate parameter $\beta + \nu + \sum_{t=1}^{T} e_{I,t}\mu_{0,I,t}$.
- *for $x = 2, \dots, I-1$, θ_x is sampled from the distribution:*

$$l(\theta_x|\underline{N_I}, \underline{D_I}) \propto F_{N_{x-1}}(\theta_x)G_{N_x}(\theta_x)l(D_x|\theta_x) \propto \exp\left(-(2\beta+\nu)\theta_x - \theta_x \sum_{t=1}^{T} e_{x,t}\mu_{0,x,t}\right)\theta_x^{N_x+N_{x-1}+\nu+\sum_{t=1}^{T} d_{x,t}},$$

that is the gamma distribution with shape parameter $N_x + N_{x-1} + \nu + \sum_{t=1}^{T} d_{x,t}$ and rate parameter $2\beta + \nu + \sum_{t=1}^{T} e_{x,t}\mu_{0,x,t}$.
- *θ_1 is sampled from the distribution:*

$$l(\theta_1|\underline{N_I}, \underline{D_I}) \propto F_{N_1}(\theta_1)l(D_1|\theta_1) \propto \exp\left(-\beta\theta_1 - \theta_1 \sum_{t=1}^{T} e_{1,t}\mu_{0,1,t}\right)\theta_1^{N_1+\sum_{t=1}^{T} d_{1,t}},$$

that is the gamma distribution with shape parameter $N_1 + \sum_{t=1}^{T} d_{1,t}$ and rate parameter $\beta + \sum_{t=1}^{T} e_{1,t}\mu_{0,1,t}$.

In terms of complexity, the sampling of a trajectory of $\boldsymbol{\theta}$ involves only the simulation of $2I$ variables, all of which have closed form (approximated) density functions. This means that the complexity is significantly lower than standard MCMC methods, which rely on the (often slow) convergence of some auxiliary Markov chains to sample variables with untractable distributions.