Stochastic Mortality Modelling: Key Drivers and Dependent Residuals

George Mavros*, Andrew J.G. Cairns*, Torsten Kleinow, George Streftaris
Maxwell Institute for Mathematical Sciences, Edinburgh, and
Department of Actuarial Mathematics & Statistics
Heriot-Watt University, Edinburgh, EH14 4AS, UK.

Abstract

This article proposes an alternative framework for modelling the stochastic dynamics of mortality rates. We employ a common mortality basis of age and period effects and compare the implications if the remaining structure is modelled, first, via a univariate cohort process and, second, via a multivariate autoregressive model for the residuals surface. The residuals model introduces dependencies between adjacent age-period cells of the mortality matrix that, amongst other things, can be structured to capture cohort effects in a transparent manner. The framework is also capable of incorporating across ages correlations in a natural way, and it turns out that such an assumption impacts the smoothness of the forecasted rates in the short-term. The age and period related latent states of the mortality basis are more robust when the residuals surface is modelled via the multivariate time-series model, implying that the process acts indeed independently of the assumed mortality basis. Under the Bayesian paradigm, the posterior distribution of the models is considered in order to explore coherently the extent of parameter uncertainty. Samples from the posterior predictive distribution are used to project mortality, and an in-depth sensitivity analysis is conducted. The methodology is easily extendable in multiple ways which give a different form and degree of significance to the different components of mortality dynamics.

Keywords: Stochastic Mortality Dynamics, Robustness, Parameter uncertainty, Correlation modelling.

* Corresponding authors: grgmavros@gmail.com, A.J.G.Cairns@hw.ac.uk

1 Introduction

Stochastic mortality models allow us to identify the impact of systematic risk on uncertainty in numbers of deaths, and a vast variety of such models has been developed during the last two decades. Some of the most influential ideas underlying the time-series nature of the field have been set up by Lee and Carter (1992), Brouhns et al. (2002), Booth et al. (2002), Renshaw and Haberman (2003, 2006), Cairns et al. (2006) and Plat (2009). In-depth comparative studies of various competitors have been conducted by Cairns et al. (2009) and Haberman and Renshaw (2011), and certain models from therein are today of the most elaborate academic and industrial approaches in modelling mortality rates. All the above, and indeed many more, models are based on the assumption that human mortality dynamics may be broken down into a mixture of eventually independent deterministic and stochastic components which capture age, period and, in some cases, cohort effects. The number and form of these types of effects is usually what distinguishes one model from another. Often, some of these effects are unobservable, and their latent states are estimated through optimisation of an objective function under some assumption about the distribution of deaths. The Poisson and the Binomial models are two standard choices. Extrapolation is then performed by independent modelling and projection of the relevant stochastic factors.

In this article we suggest an alternative framework, where the existence of dominant deterministic and stochastic factors responsible for the main shape of the mortality surface is also admitted, but at the same time we investigate the possibility of short-term and contemporaneous interdependencies within the data. Such dependence structures are modelled, first, by estimating the residuals of a simple mortality structure jointly with all other latent effects, and second, by structuring an appropriate multivariate stochastic model for projecting the newly introduced residuals process. In contrast to the mainstream approach where an additional univariate cohort process is used to model the stationary part of the mortality dynamics, our framework is able to accommodate both serial and cross-age dependence structures. This turns out to be sufficient for capturing the underlying dynamics, although in a qualitatively distinct manner. The residuals augmentation builds up a hierarchical structure of conditionally independent processes, which we then embed within the Bayesian paradigm. By constructing a Markov chain which has as its stationary distribution the target posterior of the model, we sample from it through a hybrid Markov chain Monte Carlo (MCMC) algorithm. The usage of the full posterior distribution of the model imposes the correct degree of parameter uncertainty in estimates and projections, hence removing the requirement for any additional bootstrap exercise in the modelling procedure. Our approach shares common elements with that of D'Amato et al. (2012) in the sense that time-series models are employed in both cases for the residuals matrix. However, D'Amato et al. (2012) work with the Lee-Carter model, whereas we use an enhanced version of the CBD model. More importantly, we use a parsimonious multivariate Markov autoregression instead of the single univariate autoregressive model for each age in the data in the D'Amato et al. (2012) approach. Mortality residuals have also been directly modelled by Debón et al. (2008, 2010) under a spatial dependence approach specified by a parametric covariance structure. In the present article, the main driver of the residuals model is the autoregression, and the covariance of the model enters as an additional differentiating step.

The structure of the article is as follows. Section 2 introduces the notation and modelling framework, and develops the proposed model. Section 3 briefly illustrates the fitting procedure and outlines the related algorithm for the employed MCMC scheme. In Section 4 we present our results and examine the sensitivity of the posterior distributions of the latent states and parameters of the models to different estimation time-frames. Section 5 includes comparative projections and numerical applications of the models, while Section 6 concludes.

2 Model Description

In this section we develop the proposed framework and specify the particular modelling assumptions. We assign a particular parametric residuals model to illustrate the idea which will be examined in the following sections, and briefly present and comment on the models which will be used for comparative purposes.

2.1 Notation and Assumptions

Time is assumed to be measured in years, so that calendar year t has the meaning of the time interval $[t, t+1)$. $D(x,t)$ denotes the number of deaths in year t among individuals aged x last birthday on the date of death, and $E(x,t)$ denotes the central exposure-to-risk for age x during year t, which refers to the population exposed to the risk of death.

Usually mortality models describe the central death rate, $m(x,t)$, which is presumed constant within each cell of the data, through some distributional assumption about deaths, $D(x,t)$. The conditional Poisson assumption with mean $E(x,t)m(x,t)$, independently in each age-year cell given $m(x,t)$ is employed here, where the exposures $E(x,t)$ are considered constants provided in the data. Hence, the distributional assumption that accompanies all the models in this article is stated as follows:

$$D(x,t)|m(x,t) \sim \text{Poisson}\Big(E(x,t)m(x,t)\Big). \tag{1}$$

Under this choice we ensure that all the models will be consistently implemented and compared under the same source of idiosyncratic risk. We denote by n_x and n_t the number of ages and years, respectively. The data also include $n_c = n_x + n_t - 1$ number of cohorts across the diagonals of the mortality table, given as $c = t - x$.

2.2 Modelling Framework

We assume that the principal mortality dynamics are given by the following modelling form:

$$\text{M1:} \quad \log\big(m(x,t)\big) = \alpha_x + \kappa_t^{(1)} + \kappa_t^{(2)}\big(x - \bar{x}\big), \tag{2}$$

where α_x are age-related mortality level parameters, $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ are time varying unobservable states of the two stochastic period factors and $\bar{x}$ is the mean of the ages in the data. The mortality surface is developed while the semi-parametric deterministic age basis consisting of the three vectors α_x, $\mathbf{1}$ (unit vector of length

n_x) and $\left(x-\bar{x}\right)$ varies in accordance to the evolution of the vector of stochastic factors $\boldsymbol{\kappa}_t = \left(\kappa_t^{(1)}, \kappa_t^{(2)}\right)$. Equation (2) is termed M1 in this article and is in parts similar to popular models in the literature. The right hand side of the model excluding α_x is similar to the CBD model (Cairns et al., 2006), except that it models the logarithmic $m(x,t)$'s, while including α_x gives a very basic Plat (2009) model.

Model M1 describes the principal mortality dynamics in the sense that it includes the age related basis component and all of the non-stationary stochastic part of the mortality surface. Although that basis can be extended by a quadratic age term, and the existence of a further stochastic factor, $\kappa_t^{(3)}$, can be statistically justified, we limit ourselves to the above model for simplicity, and so that we have a more informative residual structure. Even if a third term is added in model M1, the age-period interactions alone are known to be inadequate, at least for some available data where prominent stationary variation appears to exist. Such stationary dynamics come up in the form of diagonal structures within the residual surface, and are modelled by the inclusion of year-of-birth depended parameters representing the cohort, c, each cell of the dynamic life-table belongs to. Model M1 is then extended such that:

$$\text{M2:} \qquad \log\left(m(x,t)\right) = \alpha_x + \kappa_t^{(1)} + \kappa_t^{(2)}\left(x-\bar{x}\right) + \gamma_c, \tag{3}$$

where γ_c denote the stochastic cohort effects that are responsible for capturing stationary structures, assuming the same level of persistence for any specific cohort across the whole observation period. That is, for a fixed cohort c, the respective coefficient γ_c will determine the impact of the cohort effect it suffers, or enjoys, for as long as the dynamics of this cohort are being projected. Again, M2 may be viewed as model M6 in Cairns et al. (2009) with the addition of α_x modelling the logarithmic $m(x,t)$'s, or a simplified Plat (2009) model.

The alternative approach suggested by this article encapsulates estimating the residuals matrix of model M1, and the latent residual states are modelled by a multivariate stochastic process directly. Modelling the residuals as additional latent states allows for introducing richer interdependencies within the age-year cells of the population, compared to the restrictive structure of the cohort effects. If the assumption about the constant contribution of the cohort parameters to the rates across the diagonals of the mortality table was really accurate, one would expect under a model that does not include cohort effects to observe residual series on the same diagonal having low variability, *i.e.* the constant cohort effect plus the presumed *i.i.d.* noise. Examination of residual series of the same cohort, under for example M1, reveals persistently excessive standard deviation values, which justifies the adoption of a more flexible modelling framework. By utilising an appropriate multivariate residuals model one may assume various structures across both dimensions of the mortality table, while at the same time capture efficiently the cohort effect. In this case, the augmented modelling equation becomes:

$$\text{M3:} \qquad \log\left(m(x,t)\right) = \alpha_x + \kappa_t^{(1)} + \kappa_t^{(2)}\left(x-\bar{x}\right) + \mathcal{R}(x,t), \tag{4}$$

where an additional residual term $\mathcal{R}(x,t)$ is added for each cell of the data. Similarly to M2, the random vectors across all ages for fixed time t, $\boldsymbol{\mathcal{R}}_t := \mathcal{R}(x,t)$, of M3 are assumed to act independently of all other

stochastic factors of the model. The latent residuals are considered as an additional contribution to the mortality rate explanatory equation according to a multivariate stochastic model of dimension n_x.

2.3 The Residuals Model

We model the residuals matrix by a simple first order vector autoregressive process (VAR). We know a priori that the dynamics within the residuals are principally driven by the diagonal structures. We also allow for interaction between adjacent in age cells of the mortality matrix. Such a relationship is introduced via the following residuals regression equation:

$$\mathcal{R}(x,t) = \alpha\mathcal{R}(x,t-1) + \beta\mathcal{R}(x-1,t-1) + \gamma\mathcal{R}(x+1,t-1) + Z(x,t), \tag{5}$$

where $Z(x,t)$ are cell specific error terms and the coefficient β corresponds to the cohort-type effect. If cohort effects are primarily driven by lifestyle factors, such as smoking, then it is likely that there will be some diffusion between adjacent cohorts, which should then be captured by factors α and γ. Such a parsimonious structure imposes flexibility in the way cohort effects impact mortality rates over time by allowing the introduction of additional serial correlation within the mortality table. The residuals model may be compactly written as:

$$\boldsymbol{\mathcal{R}}_t = \mathcal{A}\boldsymbol{\mathcal{R}}_{t-1} + \boldsymbol{Z}_t, \tag{6}$$

where $\boldsymbol{Z}_t \sim N_{n_x}\big(0, V_Z\big)$ are *i.i.d.* vectors of dimension equal to the number of ages in the data, n_x, and covariance matrix V_Z. Following equation (5), the $n_x \times n_x$ autoregressive matrix $\mathcal{A}$ may be better written as:

$$\mathcal{A} = \begin{pmatrix} \alpha_1 & 0 & 0 & 0 & \dots & \dots & \dots & 0 \\ \alpha_2 & \alpha_3 & 0 & 0 & \dots & \dots & \dots & 0 \\ \alpha_4 & \alpha_5 & \alpha_6 & 0 & \dots & \dots & \dots & 0 \\ 0 & \alpha_4 & \alpha_5 & \alpha_6 & \dots & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \dots & \ddots & \dots \\ 0 & \dots & \dots & 0 & \alpha_4 & \alpha_5 & \alpha_6 & 0 \\ 0 & \dots & \dots & 0 & 0 & \alpha_4 & \alpha_5 & \alpha_6 \end{pmatrix}.$$

Parameters α_4, α_5 and α_6 correspond to α, β and γ in equation (5), respectively. α_5, as well as α_2, carry the cohort information of the model. α_6 and α_3 introduce serial dependence within the residuals model. α_1 is the single regression coefficient for the first age of the data.

A further source of variability within the residuals model comes in the form of the covariance matrix V_Z. Several choices can be made here. A diagonal structure implies independence between the residual states across the age dimension. We use a single common parameter across the diagonal of V_Z for the first version of model M3 which we label M3a. Further assumptions, such as constant or autoregressive correlation, may be imposed by appropriate parameterisation of V_Z. We choose to explore the full underlying structure by allowing a complete parameterisation for V_Z for the the second version of model M3 which we name M3b.

That is, we assume a fully parameterised covariance matrix which will be easily estimated within the MCMC algorithm. If we refer to model M3 means that the assignment of the structure for V_Z is not relevant. Beyond independence between the residual states, version M3a also carries the strong assumption of homoscedastic errors across all ages of the data. In contrast, all diagonal parameters of V_Z for M3b are allowed to vary, so that we can investigate the case of a distinct residuals variance parameter for each age of the data. Matrix V_Z only accounts for the one-step ahead variability of the residuals model. Under stationarity, which is ensured if the eigenvalues of matrix $\mathcal{A}$ belong in the interval $(-1, 1)$, or equivalently if parameters α_1, α_3 and α_6 are in that interval, the long-term covariance matrix, $V_{\mathcal{R}}$, of the VAR model solely depends on $\mathcal{A}$ and V_Z, and it is given as the solution of the system (Lütkepohl, 2005):

$$V_{\mathcal{R}} = \mathcal{A} V_{\mathcal{R}} \mathcal{A}' + V_Z, \tag{7}$$

which is linear in the entries of $V_{\mathcal{R}}$. The solution of the above system is the constant, long-term covariance matrix of the residuals process. So long as the examined population preserves the stationarity requirement, the contribution of the residuals model as the projection horizon extends will be n_x-dimensional vectors with expectation zero and covariance given by $V_{\mathcal{R}}$.

2.4 Further Modelling Considerations

The models presented in the previous sections are driven by the same pair of principle stochastic factors, the vector $\boldsymbol{\kappa}_t$. The components of $\boldsymbol{\kappa}_t$ are usually modelled as correlated random walks with constant drifts, so that they contribute the non-stationary mortality dynamics (Cairns et al., 2006). Hence, for $\boldsymbol{\kappa}_t$ of all models we use the following process:

$$\boldsymbol{\kappa}_t = \boldsymbol{\delta} + \boldsymbol{\kappa}_{t-1} + \boldsymbol{\zeta}_t, \qquad \boldsymbol{\zeta}_t \sim N_2\big(\mathbf{0}, V_\zeta\big), \tag{8}$$

where $\boldsymbol{\delta} = \Big(\delta_1, \delta_2\Big)$ is the constant drift vector and the covariance matrix V_ζ is parameterised as:

$$V_\zeta = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix}.$$

For model M2, in addition to $\boldsymbol{\kappa}_t$ we assign a model for the cohort effects process which is also generally assumed stochastic. A standard choice is usually a first order autoregressive (AR(1)) process with drift (Cairns et al., 2011):

$$\gamma_c = \delta_\gamma + \alpha_\gamma \gamma_{c-1} + \zeta_c, \quad \zeta_c \sim N\big(0, \sigma_\gamma^2\big). \tag{9}$$

Next, we comment on the identifiability constraints for each of the models. A detailed description of these is given in Appendix B. For model M1, equation (2) does not imply a unique set of parameter estimates. The regression equation may become identifiable by imposing the following constraints for the period effects, $\kappa_t^{(i)}, i = 1, 2$:

$$\sum_t \kappa_t^{(i)} = 0. \tag{10}$$

Model M2 is not identifiable if only the constraints in equations (10) are used. Additionally, we impose:

$$\sum_c \gamma_c = 0, \;\; \sum_c (c - \bar{c})\, \gamma_c = 0, \;\; \sum_c (c - \bar{c})^2\, \gamma_c = 0, \tag{11}$$

where $\bar{c}$ is the mean of the cohort years of birth. The constraints in equation (11) ensure γ_c will fluctuate around zero with no identifiable structure, and hence the choice of model (9) is justified.

For models M3, beyond the constraints of equation (10) we also use:

$$\sum_x \mathcal{R}(x,t) = 0, \;\; \sum_x (x - \bar{x})\mathcal{R}(x,t) = 0 \;\text{ and }\; \sum_x \alpha_x \mathcal{R}(x,t) = 0, \;\; \forall t. \tag{12}$$

Note that for this model in addition to the two constraints in equation (10) we have three sets of constraints for each calendar year of the data. This yields $3 \times n_t + 2$ constraints in total.

3 Bayesian Model Fitting

Fundamentally, the vast amount of latent states and parameters of models M3 motivates the adoption of the Bayesian paradigm. Under a usual estimation approach, such as maximum likelihood, each mortality rate would need to serve for the estimation of more than one unknown of models M3. Such an estimate would then be liable to a great variability which we attempt to balance through a more informative estimator, such as those provided using the full posterior distribution. Furthermore, obtaining a point estimate through some optimisation method would also then require a succeeding, and always essential, parameter uncertainty assessment exercise, the most usual being some sort of parametric bootstrap method as in Brouhns et al. (2005). The unified estimation and parameter uncertainty assessment provided by the Bayesian framework, as also described and applied for example by Czado et al. (2005) or Pedroza (2006), is therefore in our view the most coherent and comprehensive solution for the calibration of models M3. Due to the similar structure of all the models, the relevant MCMC algorithms are developed in parallel for all of them. Thus, eventually we compare the framework of residuals modelling, first, against the absence of modelling the stationary component of the mortality data, and second, against modelling with cohort effects as in M2, by keeping all other conditions equal. The Bayesian paradigm consists of combining the likelihood of the model with prior specifications for its parameters, so that the relevant posterior distribution is derived. The global parameter vector for all the presented models is denoted by $\boldsymbol{\theta}$. For convenience and to emphasise the dependence on the parameter vector $\boldsymbol{\theta}$, let also $u(\boldsymbol{\theta}) \equiv u(\boldsymbol{\theta}; x, t)$ generically denote the right-hand-side of equations (2), (3) and (4) for each of M1, M2 and M3, respectively.

3.1 Likelihood Function

The Poisson model of Equation (1) and the bivariate Normal likelihood for $\boldsymbol{\kappa}_t$ are common for all models. The two components of the likelihood are:

$$\begin{aligned}\ell_1(\boldsymbol{\theta}) &= \sum_{x,t}\Big\{D(x,t)\big(u(\boldsymbol{\theta};x,t)\big) - E(x,t)\exp\big(u(\boldsymbol{\theta};x,t)\big)\Big\},\\ \ell_2(\boldsymbol{\theta}) &= -\frac{1}{2}\Bigg\{\big(n_t-1\big)\log\big(|V_\zeta|\big) + \sum_{j=1}^{n_t-1}\Big[\big(\boldsymbol{\kappa}_{j+1}-\boldsymbol{\kappa}_j-\boldsymbol{\delta}\big)^{'}V_\zeta^{-1}\big(\boldsymbol{\kappa}_{j+1}-\boldsymbol{\kappa}_j-\boldsymbol{\delta}\big)\Big]\Bigg\}.\end{aligned}$$

ℓ_1 follows from the conditional Poisson assumption in terms of the parameterisation of the mortality rates as given in Equations (2), (3) and (4) for M1, M2 and M3, respectively. ℓ_2 corresponds to the conditional likelihood of the bivariate random walk model for the period effects, $\boldsymbol{\kappa}_t$, as given by Equation (8).

The likelihood of M2 also includes the model for γ_c as given by equation (9). The first state of the process is assigned the stationary distribution of the model, so that:

$$\gamma_1|\left\{\alpha_\gamma,\delta_\gamma,\sigma_\gamma\right\} \sim N\left(\frac{\delta_\gamma}{1-\alpha_\gamma},\frac{\sigma_\gamma^2}{1-\alpha_\gamma^2}\right). \tag{13}$$

Then, the additional likelihood component for M2 becomes:

$$\begin{aligned}\ell_3(\boldsymbol{\theta}) = \; - \; \frac{1}{2}\Bigg\{ &\log\left(\frac{\sigma_\gamma^2}{1-\alpha_\gamma^2}\right) + \big(n_c-1\big)\log\big(\sigma_\gamma^2\big) + \\ &\frac{1}{\sigma_\gamma^2}\left(\big(1-\alpha_\gamma^2\big)\left(\gamma_1-\frac{\delta_\gamma}{1-\alpha_\gamma}\right)^2 + \sum_{j=2}^{n_c}\big(\gamma_j-\delta_\gamma-\alpha_\gamma\gamma_{j-1}\big)^2\right)\Bigg\}.\end{aligned}$$

For model M3, the likelihood incorporates the term of the residuals model of equation (6). Similarly to above, we assign the stationary distribution implied by the model for the first residuals vector of the data. In this case the relevant stationary distribution is given as:

$$\boldsymbol{\mathcal{R}}_1|\left\{\mathcal{A},V_Z\right\} \sim N_{n_x}(\mathbf{0},V_{\mathcal{R}}), \tag{14}$$

where the covariance matrix $V_{\mathcal{R}}$ is given as the solution of the system in equation (7). The additional component of the likelihood function for M3 is then given as:

$$\begin{aligned}\ell_4(\boldsymbol{\theta}) = \; - \; \frac{1}{2}\Bigg\{ &\big(n_t-1\big)\log\big(|V_Z|\big) + \log\big(|V_{\mathcal{R}}|\big) + \boldsymbol{\mathcal{R}}_1^{'}V_{\mathcal{R}}^{-1}\boldsymbol{\mathcal{R}}_1 + \\ &\sum_{j=1}^{n_t-1}\Big[\big(\boldsymbol{\mathcal{R}}_{j+1}-\mathcal{A}\boldsymbol{\mathcal{R}}_j\big)^{'}V_Z^{-1}\big(\boldsymbol{\mathcal{R}}_{j+1}-\mathcal{A}\boldsymbol{\mathcal{R}}_j\big)\Big]\Bigg\}.\end{aligned}$$

Summarising the above discussion, the relevant likelihood functions, ℓ^{M1},ℓ^{M2} and ℓ^{M3}, for models M1, M2 and M3, respectively, are given as:

$$\begin{aligned}&\text{M1:} && \ell^{M1}(\boldsymbol{\theta}) = \ell_1(\boldsymbol{\theta}) + \ell_2(\boldsymbol{\theta}) + C_1,\\ &\text{M2:} && \ell^{M2}(\boldsymbol{\theta}) = \ell_1(\boldsymbol{\theta}) + \ell_2(\boldsymbol{\theta}) + \ell_3(\boldsymbol{\theta}) + C_2,\\ &\text{M3:} && \ell^{M3}(\boldsymbol{\theta}) = \ell_1(\boldsymbol{\theta}) + \ell_2(\boldsymbol{\theta}) + \ell_4(\boldsymbol{\theta}) + C_3.\end{aligned} \tag{15}$$

where C_1, C_2 and C_3 are constants that do not depend on the parameters of the models.

3.2 Prior Distributions

The second ingredient of our Bayesian models is obtained by the assignment of prior specifications for the lowest level free parameters over all layers of hierarchy.

The prior distribution for model M1, which is also a component of the priors for models M2 and M3, includes two components. First, the drift, $\boldsymbol{\delta}$, of $\boldsymbol{\kappa}_t$ is assigned a bivariate normal prior with mean $\boldsymbol{\delta}_0 = (0,0)^{'}$ and covariance, V_0, the 2×2 identity matrix. The prior is centred close to reasonable values for the parameters; we expect δ_1 to be positive and δ_2 to be negative, but in both cases considerably close to zero. Hence, the chosen prior is quite diffuse and not very restrictive. There is very little change in our results, however, if V_0 was 100 times the 2×2 identity matrix. Matrix V_ζ is given the Jeffreys' non-informative prior (Cairns et al., 2006). This choice conveniently leads in a conjugate Inverse-Wishart (IW) full conditional posterior for V_ζ.

The prior distribution for model M2 includes additional components for the parameters of the model for γ_c, as this is given in Equation (9). For α_γ we use a Uniform$(-1,1)$ distribution, to ensure stationarity of the process and the existence of the distribution of Equation (13). The drift, δ_γ, is given a Uniform$(-\infty,\infty)$ distribution. Finally, the variance parameter, σ^2_γ, is given an Inverse-Gamma (IG), $\sigma^2_\gamma \sim$ IG(a,b), where $a, b = 10^{-4}$. That latter prior is the usual conjugate non-informative choice for the variance of a normally distributed process, such as the one implied by model (9).

The prior distribution for model M3 includes terms for the elements of the autoregression matrix, $\mathcal{A}$, and the covariance structure V_Z. The diagonal elements of $\mathcal{A}$, α_1, α_3 and α_6, are assigned Uniform$(-1,1)$ prior distributions along the same lines of argumentation as for α_γ in M2 and the relevant comments about stationarity in Section 2. For the remaining elements of $\mathcal{A}$ we use Uniform$(-\infty,\infty)$ priors. In the case of a diagonal matrix V_Z, we assume $v \sim$ IG(a,b) a priori, as for σ^2_γ of M2. For the fully parameterised covariance matrix, V_Z, of model M3b, we use again the Jeffreys' non-informative prior, as for V_ζ of M1.

Let now $\boldsymbol{\phi}$ denote the parameter sub-vector of $\boldsymbol{\theta}$ for which prior distributions are supplied. For each of M1, M2 and M3, the respective vectors are $\boldsymbol{\phi}_1 = \left\{\boldsymbol{\delta}, V_\zeta\right\}$, $\boldsymbol{\phi}_2 = \left\{\boldsymbol{\phi}_1, \alpha_\gamma, \delta_\gamma, \sigma^2_\gamma\right\}$ and $\boldsymbol{\phi}_3 = \left\{\boldsymbol{\phi}_1, \mathcal{A}, V_Z\right\}$. We summarise the above discussion in the following mathematical statements, where the prior p_i, $i = 1, 2, 3a, 3b$ is used for each of our four models.

$$
\begin{aligned}
\text{M1:} \quad p_1(\boldsymbol{\phi}_1) &= \sum_{i=1}^{2} \delta_i^2 - \frac{3}{2}\log(|V_\zeta|) + C_1 \\
\text{M2:} \quad p_2(\boldsymbol{\phi}_2) &= p_1(\boldsymbol{\phi}_1) - (a+1)\log(\sigma^2_\gamma) - \frac{b}{\sigma^2_\gamma} + C_2 \\
\text{M3a:} \quad p_{3a}(\boldsymbol{\phi}_3) &= p_1(\boldsymbol{\phi}_1) - (a+1)\log(v) - \frac{b}{v} + C_{3a} \\
\text{M3b:} \quad p_{3b}(\boldsymbol{\phi}_3) &= p_1(\boldsymbol{\phi}_1) - \frac{n_x+1}{2}\log(|V_Z|) + C_{3b}
\end{aligned} \tag{16}
$$

where C_1, C_2, C_{3a} and C_{3b} are constants which do not depend on the parameters of the models.

3.3 Posterior Sampling

The proceeding assumptions for the models lead to their respective posterior distributions. On the log-scale, the posterior distributions of the models, $\pi_j^i(\boldsymbol{\theta})$, $(i,j) = (1,1),(2,2),(3,3a),(3,3b)$, are sums of the respective likelihood and prior terms. Based on the notation of the previous sections the posteriors may be written as:

$$\pi_j^i(\boldsymbol{\theta}) \quad = \quad \ell^i(\boldsymbol{\theta}) + p_j(\boldsymbol{\phi}_i).$$

The above log-posterior densities are complex functions of the parameters and latent states of the models. We employ a general method based on drawing samples of the model parameters from approximate distributions which are improved at each step, known as Markov chain Monte Carlo (MCMC) (Gelman et al., 2003). MCMC sampling converges to the full target posterior distribution of the model. The iterative sampling is based on the full conditional log-posterior distributions of the model. The conditioning in each case is on the global parameter vector except for the parameter, or group of parameters under consideration, denoted by $\boldsymbol{\theta}_{-i}$, where i indicates the parameters that are being updated, while the others are fixed at their latest values. The exact forms of the log-posterior and of the full conditional log-posterior distributions of the models are given in Appendix A. Generally, the parameters of interest may be sampled individually, jointly or appropriately blocked. For models such as those developed herein, blocking usually leads to great simplifications and performance improvements. The hierarchical structure of the models unveils several conjugate relationships, for which the Gibbs sampler is used. In that case, the full conditional log-posterior has a closed form which is used to sample the parameters at each iteration. For instances where no analytical result arises, the Metropolis-Hastings (MH) algorithm, or variants of it, is implemented (Gamerman and Lopes, 2006). For that method an acceptance-rejection calculation is involved, comparing the proposed state against the current. The full conditional log-posterior is again the computational tool required for comparing and choosing between the two states. Next, we sketch out the posterior sampling procedure for all the models. The similar structure of the models helps developing the respective MCMC algorithms jointly.

The age related mortality levels, α_x, are exponentially conjugate for fixed age x, for all models. The full-conditional posteriors, which are used to implement Gibbs steps in the algorithms, are given as follows:

$$\exp(\alpha_x)|\boldsymbol{\theta}_{-i} \sim \text{Gamma}\left(\sum_t D(x,t), \sum_t E(x,t)\exp\left(u\left(\boldsymbol{\theta}\right)\right)\right). \tag{17}$$

The full-conditional log-posteriors of the vector $\boldsymbol{\kappa}_t$ of the two period effects for fixed time t are mixtures of Poisson and bivariate normal densities, commonly for all models. The differentiation from one model to another depends only on the parameterisation of the model, $u\left(\boldsymbol{\theta}\right)$. The functions of the vectors $\boldsymbol{\kappa}_1$ and $\boldsymbol{\kappa}_{n_t}$ include a single term coming from the bivariate normal density. The period effects vectors for intermediate years, $\boldsymbol{\kappa}_j$, include two such terms. Since no analytical result is available for their sampling, we employ MH sampling. Given the current estimate of the vector $\boldsymbol{\kappa}_t$ we propose a new value by sampling from a bivariate normal distribution centred at that current value of the chain and appropriately tuned diagonal covariance matrix. The technique is generally known as random walk MH scheme. Although more elaborate choices for

the proposal covariance structure may be essential for other Bayesian models, the random walk scheme works well for our purposes. Transition, or not, to the proposed state depends on the corresponding acceptance ratio. If the $j-1^{th}$ state of the algorithm is $[\boldsymbol{\kappa}_t]^{(j-1)}$ and the proposed state of the j^{th} iteration is $\widetilde{\boldsymbol{\kappa}_t}$, then the required acceptance ratio is given as:

$$\mathrm{r} = \min\left\{1, \exp\left(\pi\left(\widetilde{\boldsymbol{\kappa}_t}|\boldsymbol{\theta}_{-i}\right) - \pi\Big(\left[\boldsymbol{\kappa}_t\right]^{(j-1)}|\boldsymbol{\theta}_{-i}\Big)\right)\right\}.$$

Model M2 includes the cohort effects, γ_c, as additional latent states. Their full-conditional log-posteriors are again mixtures of Poisson and univariate, in this case, normal terms. There is a single observation for the boundary cohorts of the mortality table, which increase progressively from both sides as we move towards intermediate cohorts. Due to the assumption in Equation (13), the full-conditional log-posterior of the first state of the cohort series includes an additional term from the stationary distribution of the process. Similarly to $\boldsymbol{\kappa}_t$, we employ a univariate random walk MH scheme for sampling the latent states γ_c for fixed cohort c.

Model M3 includes the residual vectors, $\boldsymbol{\mathcal{R}}_t$, as additional latent states. As for $\boldsymbol{\kappa}_t$, the full-conditional log-posteriors for the vectors at the boundary years of the data include a single term from the conditional likelihood of the residuals models, while all intermediate years comprise two such multivariate normal terms. Again, for $\boldsymbol{\mathcal{R}}_1$ we include the stationary term of the multivariate normal which stems from assumption in Equation (14). Sampling the latent residuals vectors is achieved by a multivariate random walk MH scheme centred at that current value of the chain and appropriately tuned diagonal covariance matrix.

Having described the sampling procedure for the latent states for the models, now we detail the method for the parameters of the underlying stochastic processes. The random walk model for $\boldsymbol{\kappa}_t$ is present in all models and its parameters, $\boldsymbol{\delta}$ and V_ζ, are conjugate with bivariate normal and IW full conditional log-posterior distributions, respectively. As such, the Gibbs sampler is implemented to update these parameters.

Model M2 includes the parameters of the cohort process, α_γ, δ_γ and σ^2_γ. The first two have intractable full-conditional log-posterior, and therefore random walk MH sampling is used to approximate their posterior distributions. The full-conditional of σ^2_γ is conjugate such that:

$$\sigma^2_\gamma \sim \mathrm{IG}\left(\frac{n_c}{2} - 1, \left(1-\alpha^2_\gamma\right)\left(\gamma_1 - \frac{\delta_\gamma}{1-\alpha_\gamma}\right) + \sum_{j=2}^{n_c}\left(\gamma_j - \delta_\gamma - \alpha_\gamma\gamma_{j-1}\right)^2\right). \tag{18}$$

Model M3 comprises the parameters of the residuals model, $\mathcal{A}$ and V_Z. Matrix $\mathcal{A}$ is involved in the summation included in the ℓ_4 part of the log-likelihood but moreover, it is also contained in the system whose solution yields the matrix $V_\mathcal{R}$ for the unconditional posterior term of $\boldsymbol{\mathcal{R}}_1$. No compact form arises for the functional specification of the full-conditional log-posteriors of these parameters. Thus, the entries of the autoregressive matrix $\mathcal{A}$, $\alpha_1, ..., \alpha_6$, are updated individually based on normal random walk MH steps centred at the current state of the chain and appropriately tuned proposal variance.

At this point we describe the only differentiation between models M3a and M3b, which occurs for the residuals model covariance. The covariance matrix, V_Z, of M3a is assigned a diagonal structure with a

common parameter v. This form implies independence between the latent residuals across the age dimension for fixed time, t. Parameter v is involved in the stationary covariance $V_{\mathcal{R}}$ so that no precise conjugate relationship holds, and a normal random walk MH step is applied to sample for it. On the other hand, M3b assumes a fully parameterised covariance matrix for V_Z. Again due to the presence of the stationary covariance matrix, $V_{\mathcal{R}}$, the form of the full-conditional log-posterior of V_Z is not conjugate. However, it is close to an IW distribution, which we use as the proposal generator and set up a MH sampling step. This process approximates efficiently the fully parameterised covariance matrix V_Z which captures the correlation structure of the residuals states across all ages.

4 Empirical Results

In this section we present the results of the previous fitting procedures. The Human Mortality Database England and Wales (EW) males exposures and deaths sample is used for 50 years, from 1960 to 2009, and for 30 ages, 60 to 89 inclusive. First, we discuss the fit of the models for the full data-set. Second, we conduct a sensitivity analysis for the parameter estimates, initially, by fitting the models to the final 30 years of the data, from 1980 to 2009, and then, by fitting the models to the first 40 years of data from 1960 to 1999, thus leaving the final 10 years of the complete data-set for validation. The MCMC algorithms are iterated for 1,050,000 cycles, of which only every 50^{th} iteration is kept to yield a posterior sample of size 21,000. The first 1,000 iterations are discarded as the burn-in period for all chains.

4.1 Data Analysis, years 1960-2009

Convergence of all models is fast and the applied burn-in period is possibly longer than required. This is more the case for models M1 and M2, where the algorithms converge within the first 10 to 20 iterations. For models M3, some lower level parameters may require up to 200 to 300 runs. The thinning interval is adequate in the sense that the individual MCMC chains show no significant autocorrelation at any lag, except for some of the elements of matrix $\mathcal{A}$ for models M3. Also, the traces of some elements of matrix $\mathcal{A}$ develop noticeable cross-correlation. The acceptance rates for the MH steps of the algorithms vary between 10-25% across the latent states and parameters of the models, and these values are consistent with what is recommended in the relevant literature. The majority of MCMC traces are as expected, and some of the most indicative are displayed in Figures 1 and 2 below.

First we comment on the convergence trace of the autoregression coefficients of $\mathcal{A}$ for models M3a and M3b. We have excluded the burn-in period for parameters α_4, α_5 and α_6 in order to inspect closely the differences between the two posterior distributions. The distinction between these estimates for the two models is due to the fully populated residuals covariance matrix of M3b, instead of the diagonal structure of M3a. The panels of Figure 1 illustrate that coefficients α_1, α_2 and α_3 converge to different values for the two models. This suggests that the covariance matrix of M3b captures and absorbs significant interaction from the autoregression for the first two ages of the data. A particular point can be made for the behaviour

of α_1. The single variance parameter v of M3a, and the absence of correlation with residuals of other years of the data, leads α_1 to be very close to one, due to the imposed constraint. The implication will be that residuals $\mathcal{R}(60,t)$ will be determined by a very strong autoregressive relationship. On the contrary, α_1 of M3b is centred to significantly lower values and its posterior distribution is more volatile than α_1 of M3a. The posterior spread of all other parameters of matrix $\mathcal{A}$ is greater for model M3a, implying that the fully populated covariance matrix, V_Z, reduces the variability of the autoregressive parameters.

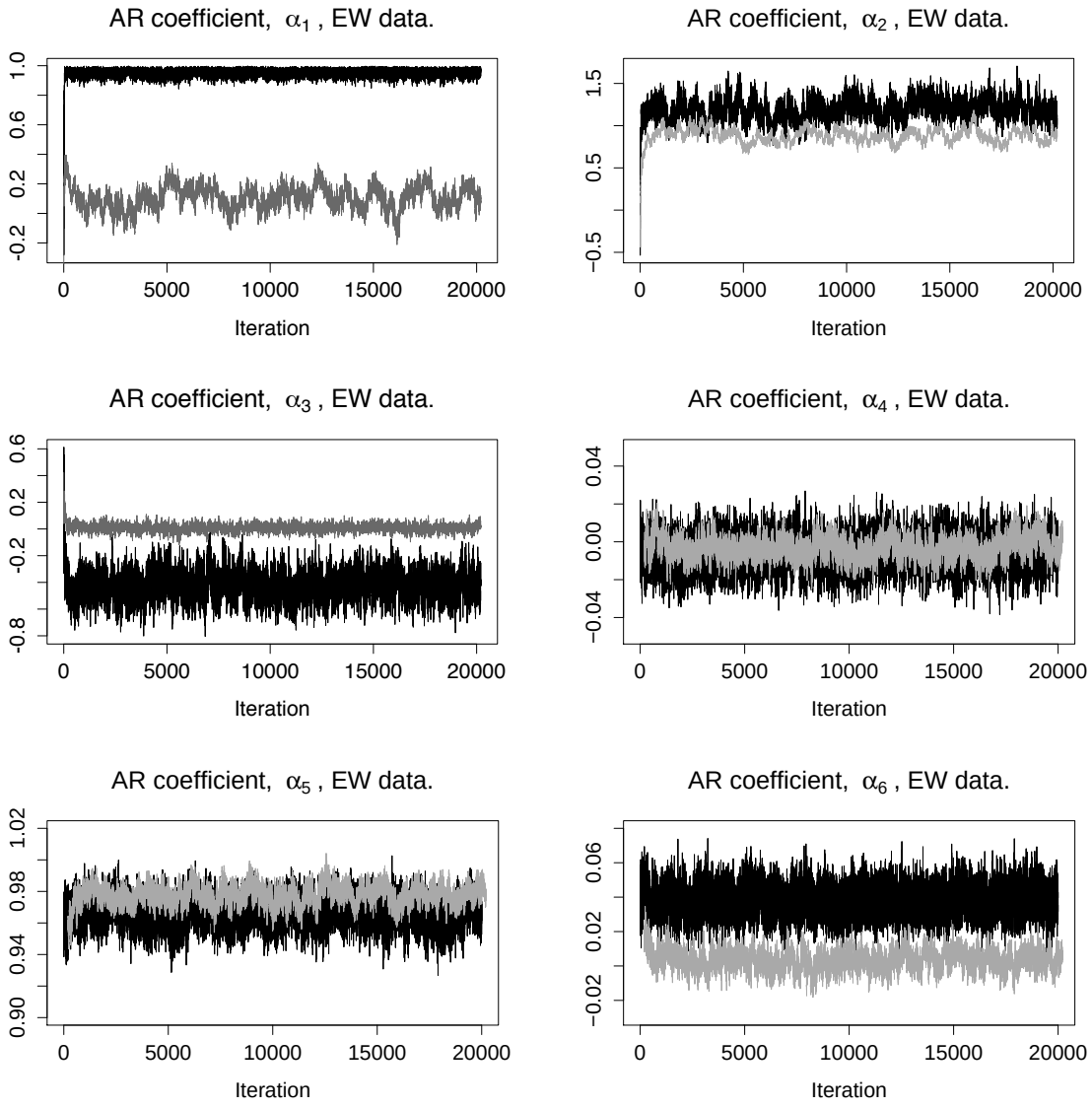

Figure 1: MCMC traces for elements of autoregressive matrix $\mathcal{A}$. Black trace is for model M3a and grey trace is for model M3b.

Next we investigate the behaviour of the MCMC chains for some of the latent residual states and for the residuals covariance matrix elements. Figure 2 compares the MCMC traces of $\mathcal{R}(65,2009), \mathcal{R}(75,2009)$ and $\mathcal{R}(85,2009)$ between models M3a and M3b. These three cases display the distinctive behaviour of the residuals across ages. The latent residual state for age 75 is clearly negative, whereas for ages 65 and 85 it

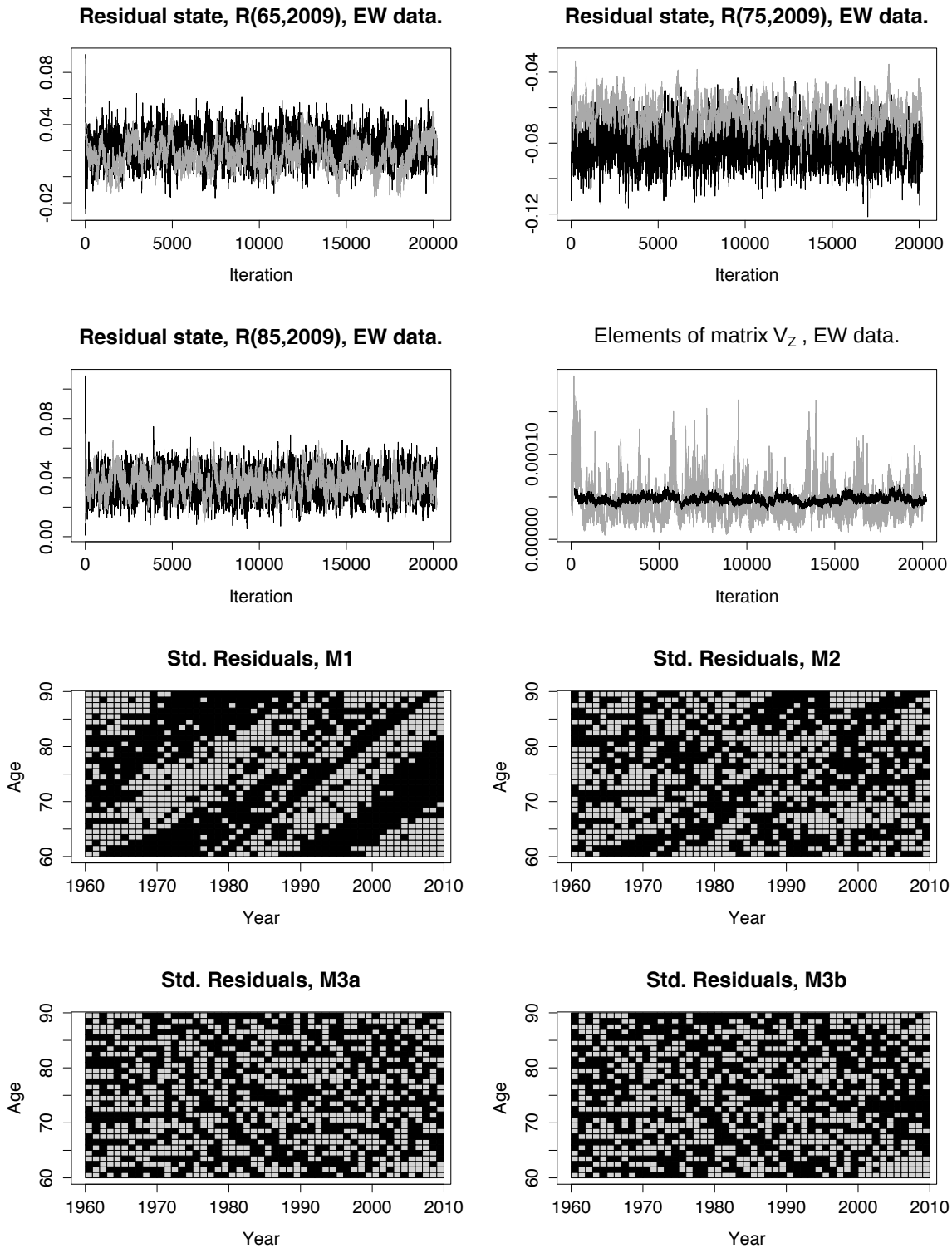

Figure 2: Top and left of 2^{nd} from top: Comparative MCMC traces for models M3a and M3b in black and grey colour, respectively.

2^{nd} from top: Right frame compares the trace of the single variance parameter v of matrix V_Z of M3a against that of the variance parameter for age 65 of matrix V_Z of M3b.

Lower two panels: Standardised residuals plots of all models, indicated in grey and black squares for positive and negative values, respectively.

is positive. In all three cases the residual state compensates for what is not captured by the mortality basis of model M1, resulting in a better fit. The bottom right panel in the second row of Figure 2 includes the plot of the MCMC trace of the variance parameter v, the common variance parameter across all ages, of matrix V_Z for M3a, along with the trace of v_{65} of V_Z, the variance of the residuals series for age 65, for M3b. The posterior distribution of v of M3a has much smaller variance compared to v_{65} of M3b, and this is generally the case across all ages.

We compare the goodness of fit of the models by inspecting the bi-dimensional standardised residuals plots. The lower two panels of Figure 2 display grey-black plots, where grey are positive and black are negative standardised residuals values for all models. The plot of M1 indicates an obvious remaining structure not captured by the model. The plot of M2 is improved but clusters of positive or negative standardised residuals can still be observed. Both models M3 depict fairly random plots across both dimensions. A slight improvement in that respect could result from the addition of a further age-period interaction term in the mortality basis of M1.

Figure 3 plots the posterior medians of the latent age and period effects of all models for the three different time-frame estimations. In this subsection, we focus on the points corresponding to the full data (1960-2009), and comment on the distinctive behaviour of the latent states estimates between the four models. Comments regarding the sensitivity of the latent states to the estimation time-frame are reserved for the next subsection. The age dependent mortality basis, α_x, appears identical for all models. The key drivers of the mortality dynamics as captured by the vector of period effects, $\boldsymbol{\kappa}_t$, signify two main points. First, the estimates of the period effect vectors are not affected by the correlation structure of model M3b, and are practically identical between M3a and M3b. Second, the $\boldsymbol{\kappa}_t$'s of models M3 are much closer to those of M1. This is in contrast to the respective behaviour of M2, where the cohort process distorts the $\boldsymbol{\kappa}_t$ estimates, especially for $\kappa_t^{(2)}$. Small discrepancies between the $\kappa_t^{(2)}$'s of M1 and of models M3 may be noticed for the very final years in the data. The trend of the period effects estimates also determine the parameters, $\boldsymbol{\delta}$ and V_ζ, based on which the random walk model for $\boldsymbol{\kappa}_t$ will be projected. Close inspection of the corresponding densities in Figure 4 shows that although the tilt of the period effects of M2, the only observable difference occurs at the tails of the posterior distributions of parameters σ_2^2 and σ_{12}, while the drifts remain almost unchanged. Nevertheless, the addition of the cohort process, γ_t, appears to affect the $\boldsymbol{\kappa}_t$ series, and models M3 seem to capture the stationary stochastic component of the mortality dynamics in a properly independent manner.

Further, we look into how the estimated residuals model is described by its parameters. The model is primarily determined by the autoregressive matrix, $\mathcal{A}$, and differences in the autoregression coefficients of models M3a and M3b are solely due to the full correlation structure of M3b in contrast to the single-parameter diagonal covariance matrix of M3a. Figure 5 shows the sampled posterior densities of the elements of $\mathcal{A}$ for models M3. As above, we focus in the full data estimated densities in black and reserve comments regarding the sensitivity to the estimation time-frame for the next subsection. As also commented earlier, the most striking differentiation occurs for coefficient α_1, which in the presence of the full covariance matrix of M3b is significantly shifted to lower values. Beyond the qualitatively different behaviour of the $\mathcal{R}(60, t)$ series,

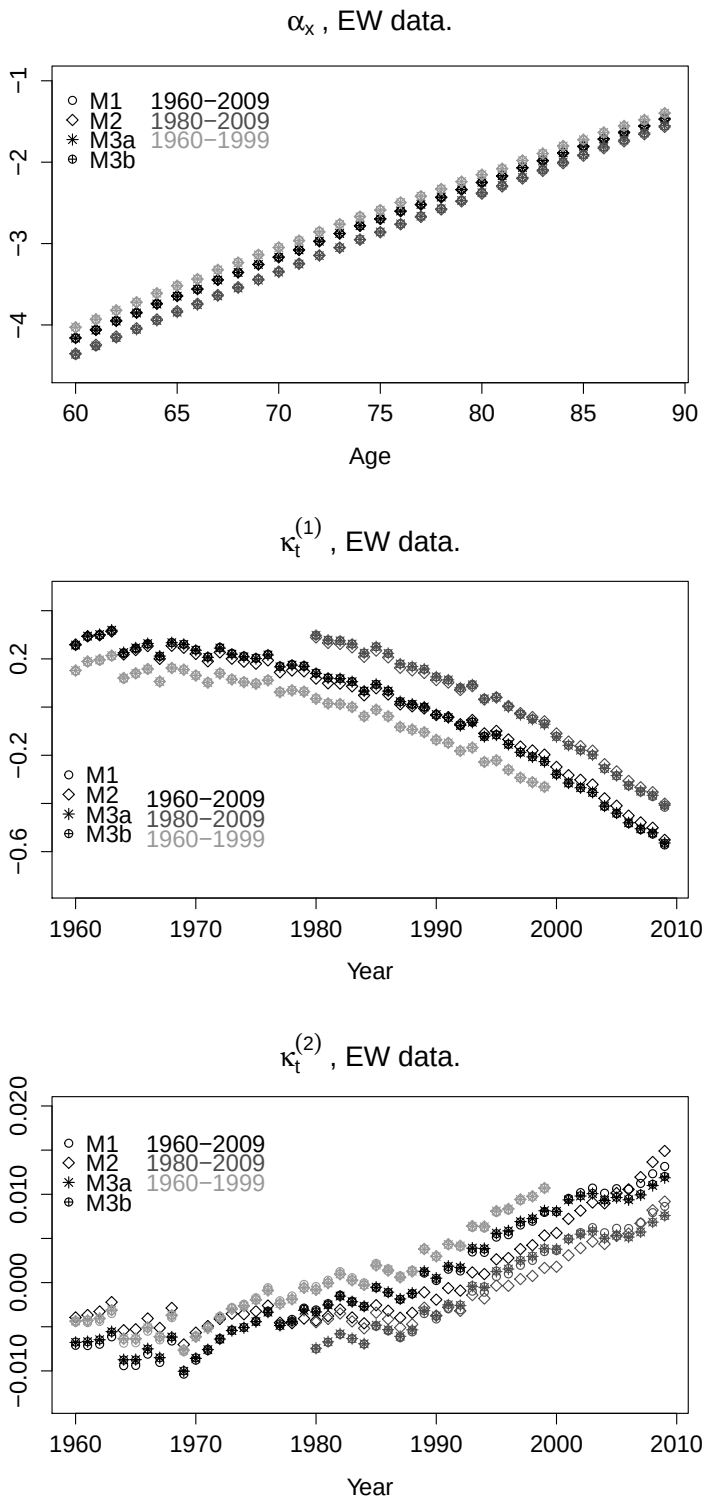

Figure 3: Posterior medians of age dependent mortality basis, α_x, and period effects, $\boldsymbol{\kappa}_t$, for models M1, M2, M3a and M3b.

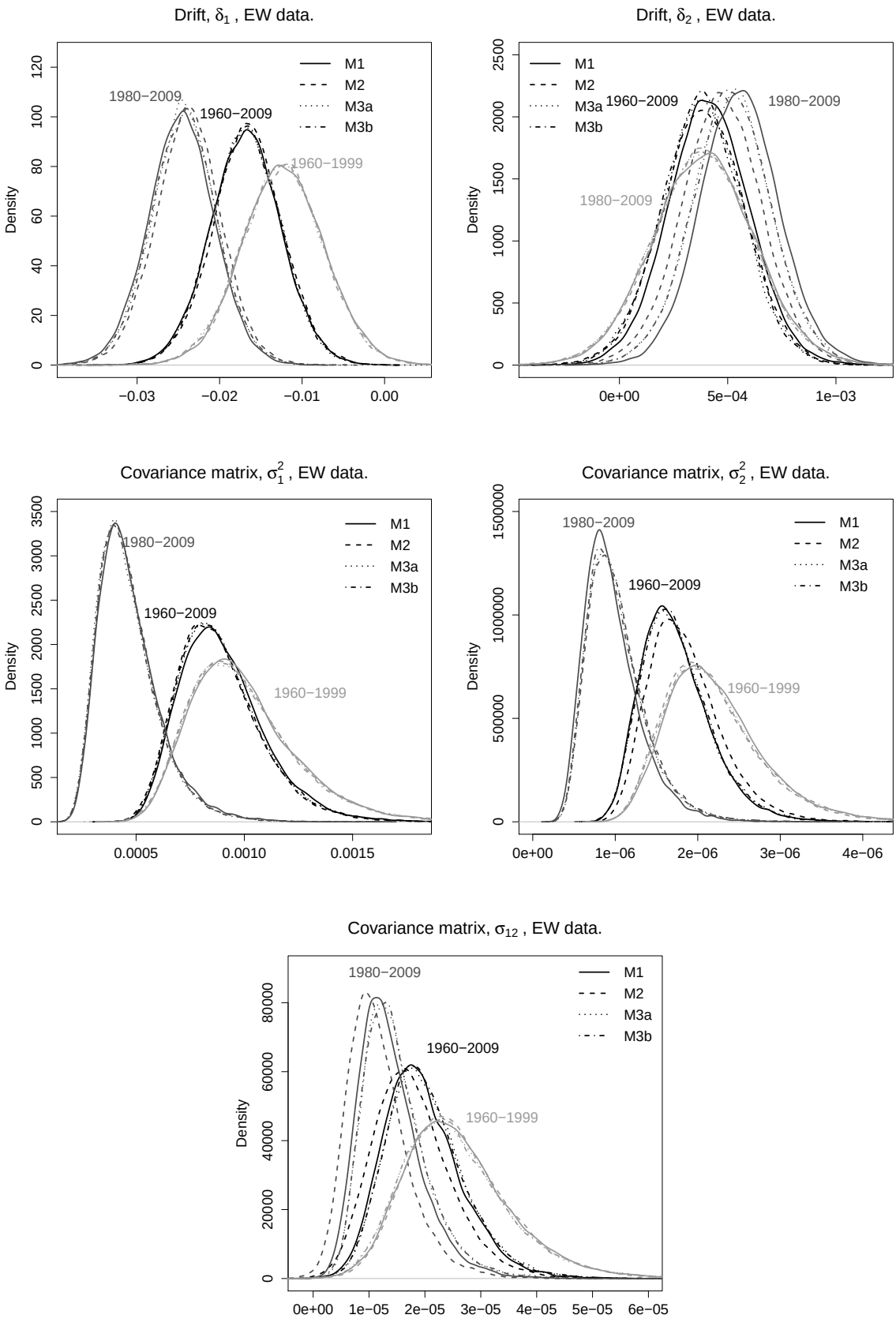

Figure 4: Posterior densities of the bivariate random walk model parameters for the vector $\boldsymbol{\kappa}_t$ for models M1, M2, M3a and M3b.

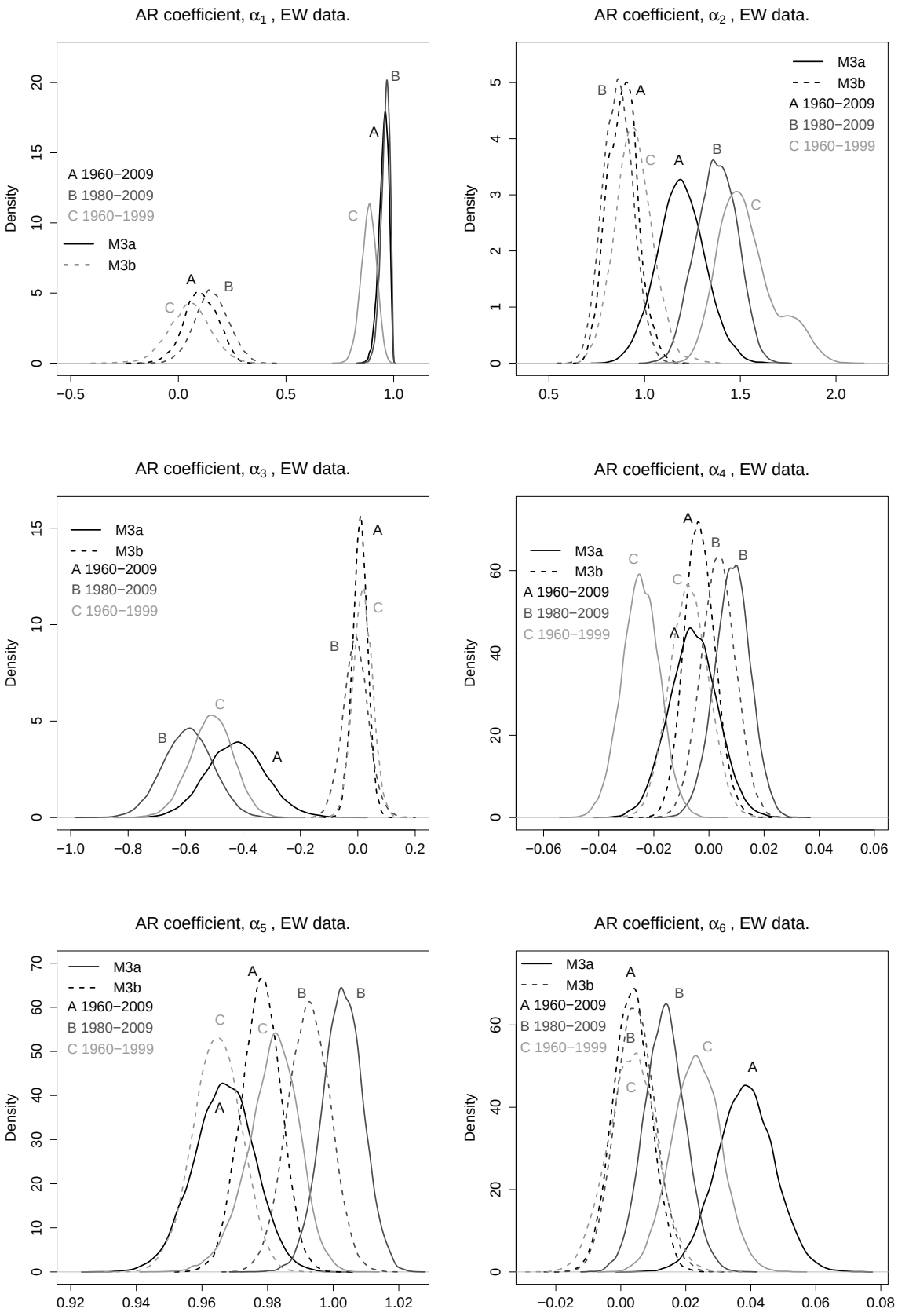

Figure 5: Posterior densities for the elements of the autoregression matrix $\mathcal{A}$ for models M3a and M3b.

a further implication of that shift is that the residuals model of M3b is clearly stationary, whereas M3a strongly relies on the restriction of α_1 being less than 1 within the algorithm. Deviation is also observed for the α_3 coefficient of the second age of the data, which under M3b is clearly centred around a zero.

Coefficients α_4, α_5 and α_6 describe the residuals behaviour for the majority of ages. The main difference between the posterior distributions of these parameters under models M3a and M3b appears for α_6, which in the presence of the full covariance matrix of M3b is centred around zero. Combined with the similar shift of α_3, we conclude that the covariance matrix of M3b absorbs significance from the time dimension of the mortality surface and embodies the dependency in the age dimension.

Figure 6 examines the posterior estimates of the covariance matrix V_Z for models M3. The top panel shows the element-wise posterior median of V_Z of M3 scaled to illustrate the correlation for the full data-set estimation. The graph shows mostly positive residual correlation for adjacent ages, which typically decreases uniformly. This is an additional dependence structure across the ages in the data which will incorporate further variability in the forecasts beyond the cohort effect which is mainly captured by coefficients α_2 and α_5 of matrix $\mathcal{A}$. The variance parameters of M3b across ages are compared the to common variance parameter of M3a in the lower right panel of Figure 6, where the median estimates are shown. The estimate for the first age in the data-set is excluded since it is significantly great and distorts the illustration. We observe that for the majority of ages the residuals variance parameters of M3b are lower than the single parameter of M3a. This was expected due to the effect of the covariance elements of M3b.

4.2 Sensitivity Analysis

Here we refer to the plots of the previous section and examine the robustness of the parameters with respect to the estimation time-frame.

In Figure 3 we examine the impact of the time-frame on the posterior medians of the latent age and period effects. For each different estimation time-frame the age-related basis, α_x, is identical for all models. The period effects $\kappa_t^{(1)}$ perfectly match for the 1960-1999 estimation across all models, whereas with the 1980-2009 data slight deviations occur in some cases under M2, but less intense than those observed for the full data. Similar behaviour is shown for period effects $\kappa_t^{(2)}$, where the estimates for the time-frame 1960-1999 for M2 are closer to those of the other three models. In Figure 4 we assess the impact of the calibration time-frame on the parameter estimates of the random walk model for $\boldsymbol{\kappa}_t$. The greatest differences within the drift vector occur for δ_1, which is shifted significantly lower and higher, compared to the full data medians, for the 1980-2009 and 1960-1999 estimations, respectively. The second component, δ_2, appears to remain centred around a stable region in all three cases, with an obvious shift to higher values under the time-frame 1980-2009. Also, noticeable are the shifts in the densities across the four models for that time-frame, most possibly due to the inclusion of only 30 years. The posterior distributions of the three parameters of the covariance matrix of the random walk model share similar behaviours. First, all three estimated densities are significantly lower for the smaller time-frame, 1980-2009. Second, all three estimated densities for the

time-frame 1960-1999 develop a heavier right tail.

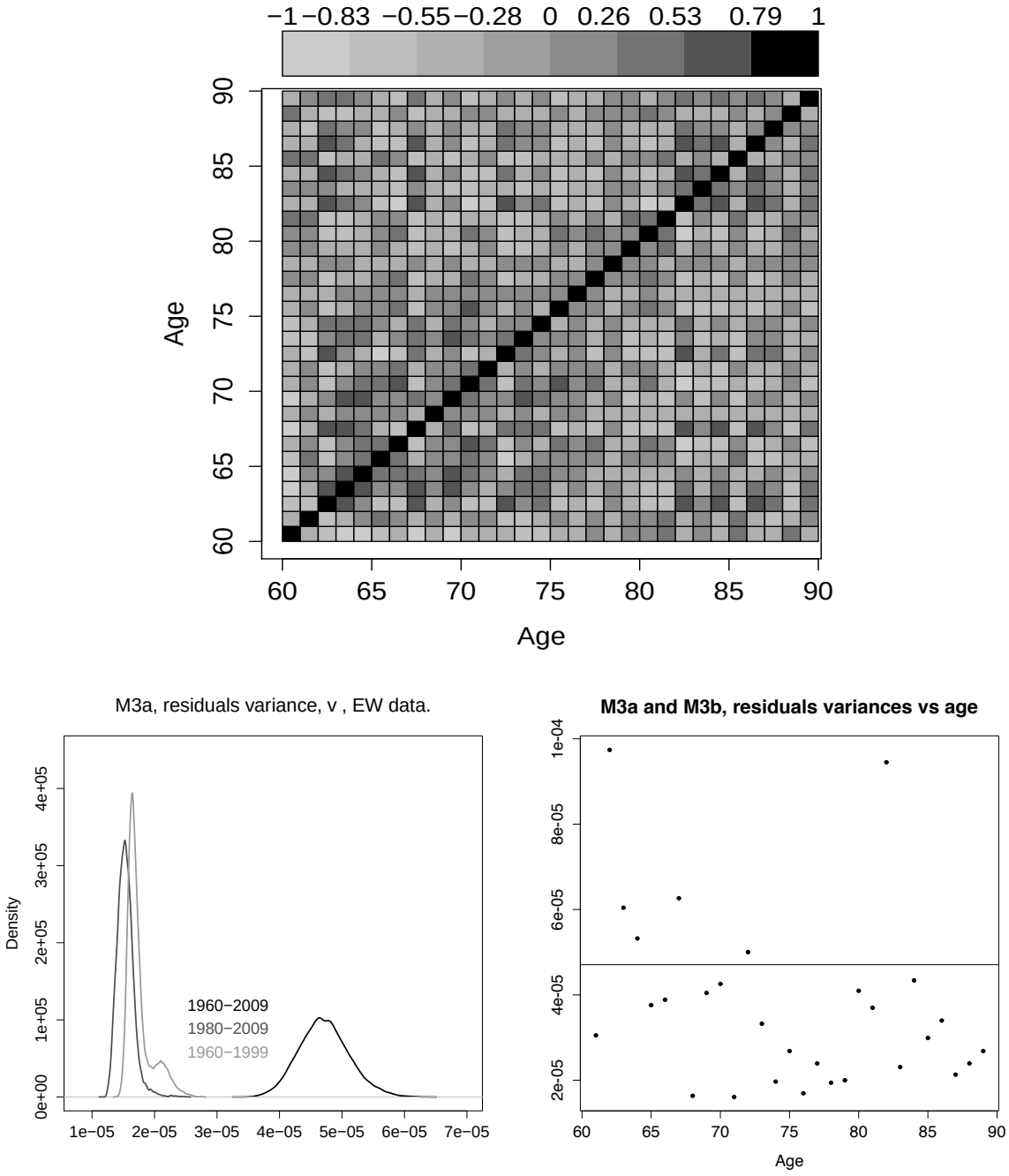

Figure 6: Posterior summaries of covariance structures for models M3a and M3b.

Top: Implied correlation matrix from the posterior medians of the covariance matrix V_Z of model M3b for the full EW data-set.

Bottom left: Posterior density of parameter v of the covariance matrix V_Z of model M3a for all three simulations.

Bottom right: Posterior medians of diagonal elements of V_Z of M3b against the constant parameter v of M3a for the full EW data-set.

A final observation for all three parameters of the covariance matrix concerns the stability of the estimation across models for the 1960-1999 time-frame, where all densities appear to be closer one to another compared to some instances of the other two cases.

Figure 5 focuses on the robustness of the elements of matrix $\mathcal{A}$ of the residuals model. It can be seen that parameters α_2, α_3 and α_6 of M3b do not change substantially under the different time-frames. On the contrary, α_4 and α_5 of M3b appear to shift by changing the estimation time-frame, this being more intense for α_5 which reflects the impact of the cohort effect. Also, the estimates of the elements of $\mathcal{A}$ of M3a seem to be systematically impacted by the change in the estimation time-frame. Finally, in the lower left panel of Figure 6 we compare the change in the estimated posterior density of the single variance parameter v of the diagonal covariance matrix V_Z of model M3a by changing the estimation time-frame. For the smaller data-sets this results in posteriors centred around significantly lower values with reduced posterior variance compared to the density from the estimation of the full time-frame.

5 Applications

In this section we present applications of the mortality models. First we look at the forecasting properties of the models, and we conduct a brief back-testing exercise. Further, we calculate standard mortality risk metrics in order to quantify the differences between the forecasts of the mortality models. Finally, we examine the difference in the forecasts when the stationary mortality dynamics are modelled by a cohort process and when modelled by the residuals.

5.1 Forecasting Properties

First we examine the forecasts of the models under estimation based on the full data, 1960-2009. We have a joint posterior sample of size 20,000 from which we sample without replacement 10,000 times and forecast the models for 120 years ahead.

The full set of historical observations and associated model forecasts for the first 50 years of the log-rates for ages 65, 75 and 85 are shown in Figure 7. The forecasts of M1 contain only non-stationary dynamics. If the stationary mortality dynamics are modelled via the cohort process, γ_c, the graph of M2 shows increasingly volatile forecasts for as long as estimated states are incorporated in the forecasts and until the cohort process converges. This is particularly evident for the higher ages, 75 and 85, of the graphs. Model M3a seems capable of capturing the stationary mortality dynamics in a similar manner to M2 since the same trend is generated by the forecasts for all three ages. However, the observed curvature is created by the Markov multivariate autoregressive relationship of the residuals model rather than the volatile cohort estimates of M2. Hence, our framework captures the actual mortality dynamics instead of just incorporating volatile latent effects that improve the fit of the model to the given data-set. Moreover, the evolution of the projected rates under M3a is much smoother compared to M2. The graph for M3b shows that the smooth projections of M3a are due to the stronger autoregression implied by coefficient α_1 of the residuals model of M3a.

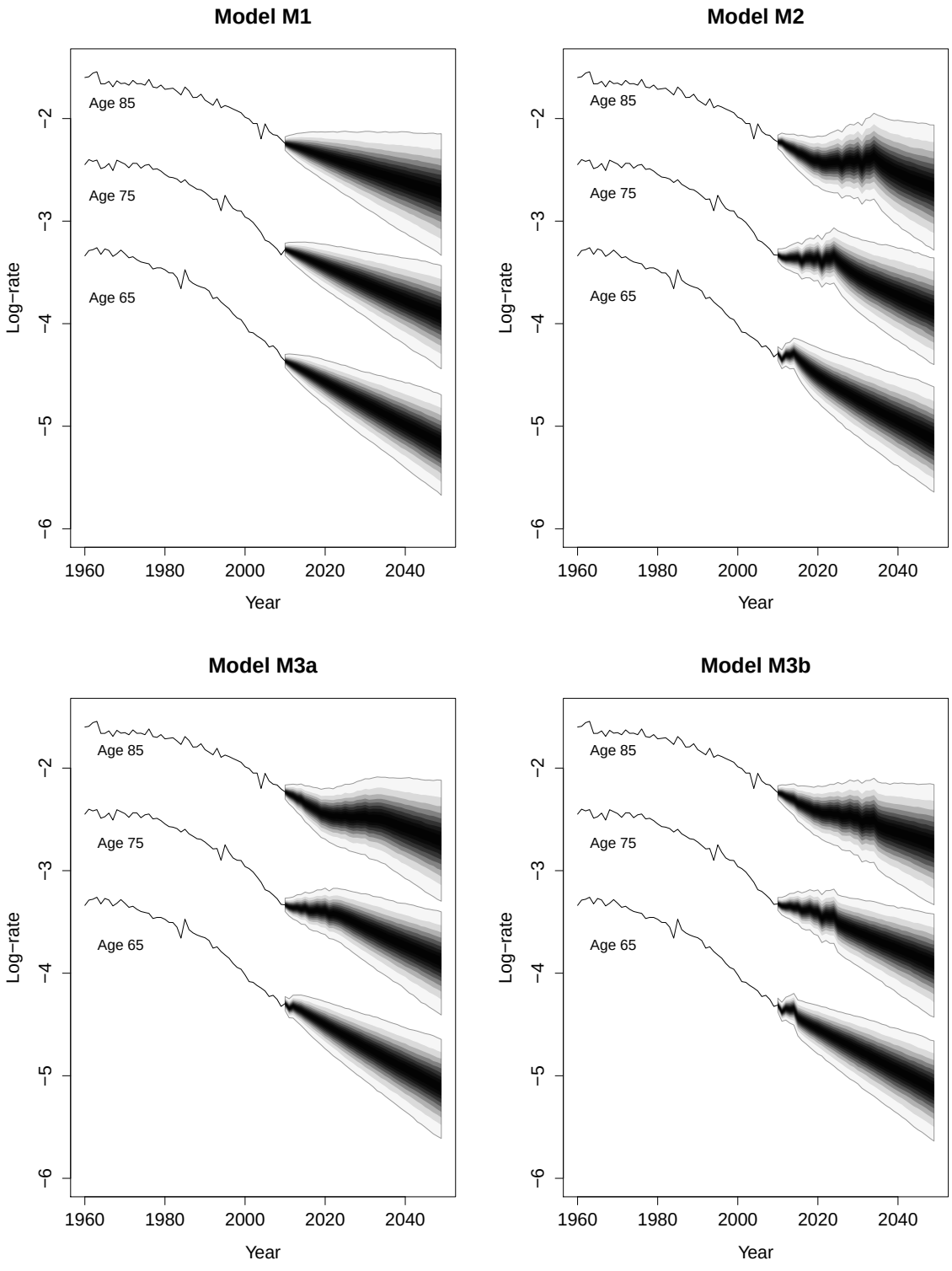

Figure 7: Historical observations and associated forecasts of log-rates for ages 65, 75 and 85 for 50 years ahead for all models. Fan charts illustrate 50% to 95% confidence bands at 5% step.

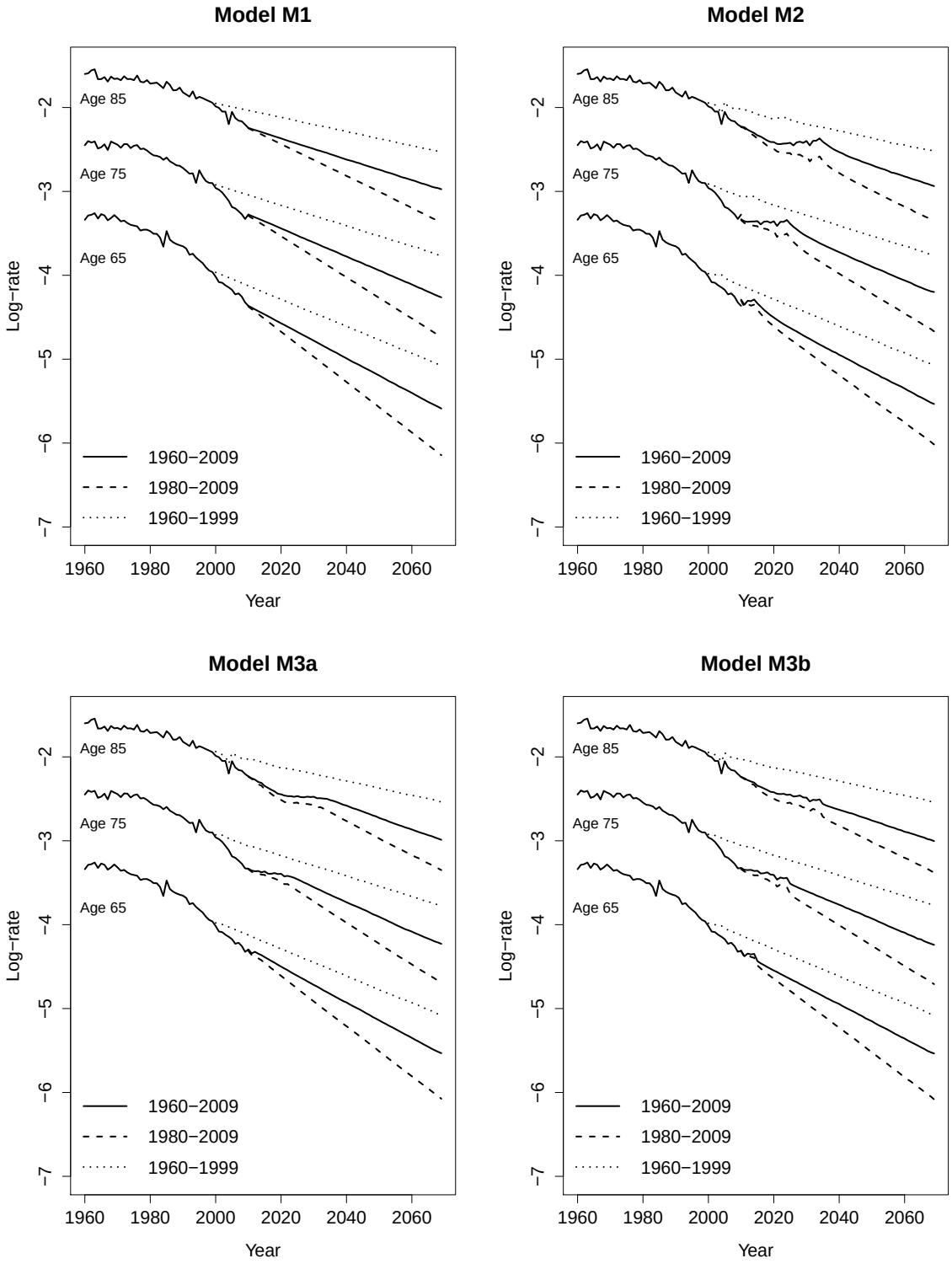

Figure 8: Historical observations and associated forecasts of log-rates for ages 65, 75 and 85 for 60 years ahead for all models. The graphs show the forecasts under all three estimation time-frames.

Figure 8 plots the historical observations and the projected medians of the four models under all three estimation time-frames, this time for half the forecasted period, up until 2070. The graphs allow close examination of the qualitative properties of the forecasts, and in particular the smoothness of model M3a in contrast to the jagged outcomes of models M2 and M3b. A further point to note is the difference between the forecasted medians depending on the estimation time-frame. If the final 30 years of the data-set are only employed to deduce the parameters of the underlying stochastic models, future rates appear significantly lower compared to the forecasts under the full available time-frame. The principal reason for this behaviour is lower estimate of the drift component, δ_1, under the truncated data estimation. Finally, we backtest the mortality models by comparing the projected rates for years 2000-2009 against the observed. In all cases, the models fail to capture the realised rates, and the discrepancies are generally greater as the examined age increases.

5.2 Mortality Risk Metrics

In this section we calculate mortality indices and metrics in order to compare the behaviour of the models and quantify their differences.

A standard mortality related financial measure is the price of an annuity. Based on the forecasts of the previous section, Table 1 shows the mean and standard deviation of the value of a 25-year term immediate unit annuity paid in arrears issued to a life aged 65 in 2010, the first forecasted calendar year, under a constant interest rate of 4%. We have calculated the summary statistics of the annuity based on the fit of the models for two of the estimation time-frames of the previous section. We do expect the annuity distribution to be higher under estimation time-frame 1980-2009, since the forecasted mortality rates are considerably lower in that case. As the results show, this turns to be indeed the case for all our models. Additionally, we see that the standard deviation of the annuity distribution is significantly lower for the 30 years data-set. Focusing on the behaviour across models for fixed estimation window, model M1 yields the highest mean values and model M2 the lowest. Models M3 result in mean values that are close, and the expected annuity value under M3b turns out to be greater for both estimation time-frames. Finally, model M1 returns the lowest annuity standard deviation since it does not take into account the stationary mortality dynamics.

Next we examine the forecasting impact of modelling the stationary mortality dynamics under the three models. In Figure 9 we plot the difference in the median forecasts between M1 and each of the three models which include a stationary dynamics component, for ages 65, 75 and 85. Essentially, we plot the ratio:

$$s(x,t)^{Mj} = \frac{\tilde{m}(x,t)^{Mj} - \tilde{m}(x,t)^{M1}}{\tilde{m}(x,t)^{M1}}, \tag{19}$$

where $\tilde{m}(x,t)^{Mj}$ is the median mortality rate forecast for model Mj = M2, M3a, M3b. $s(x,t)^{Mj}$ can be seen as the median forecasted spread of each model if M1 is the basis.

	1960-2009		1980-2009	
Model	Mean	Std. Dev.	Mean	Std. Dev.
M1	11.84	0.259815	12.06	0.202738
M2	11.59	0.268172	11.88	0.204953
M3a	11.70	0.269849	11.90	0.208190
M3b	11.74	0.265996	11.93	0.206181

Table 1: Summary statistics of a 25-year term immediate annuity paid in arrears issued to a life aged 65 in 2010 under a constant interest rate of 4%.

The graph shows a clear distinction between the behaviour of M2 and models M3. In particular the spread of M2 develops a large jump early in the forecasting period, and then settles down. In contrast, the behaviour of the spreads of models M3 is more stable, especially as age increases. Commonly for all three models the long-term spread for age 65 increasingly converges to some positive level between 5% and 10%. Furthermore, as the examined age increases, the long-term spread of M2 is clearly positive and systematically greater than that of models M3. On the other hand, the long-term spread of models M3 agree, and also vary according to the examined age as they are positive for age 75 but turn negative for age 85. In other words, the residuals modelling framework allows for more flexible mortality improvements across different ages of the data-set. The initial spread paths of all models are close for the first projection years. Those similar paths are followed for more years as age increases. The path of the spread of M3b is systematically different to that of M3a in the medium to long-term due to the residuals autoregression implied by the two models. The strong autoregression for the first age of the data, implied by coefficient α_1 of matrix $\mathcal{A}$, of M3a generates cohorts that produce the slow convergence of the spreads observed across all ages. Overall, we conclude that modelling the stationary mortality dynamics via a residuals model yields more flexible long-term forecasts compared to the cohort process. This is also a matter of the underlying time-series models, but the choice we have made here for the cohort ARIMA model is quite common and compares well with the residuals VAR model.

6 Conclusion

We have described an alternative framework for modelling the stationary structure of mortality data. In contrast to the mainstream method of using a pure univariate cohort process, we approach the problem by assuming a VAR model for the residuals surface. Through our framework it is possible to model efficiently the impact of the cohort effect, while also impose further cross-age and serial dependencies. We have presented the framework by assuming a simple structure for the residuals model which we compared against the absence of explicit modelling of the stationary component of mortality data, as well as against modelling it via a cohort process. We have shown how the correlation structure of the latent residual states across ages can be identified, and that such an assumption controls the smoothness of the forecasted rates.

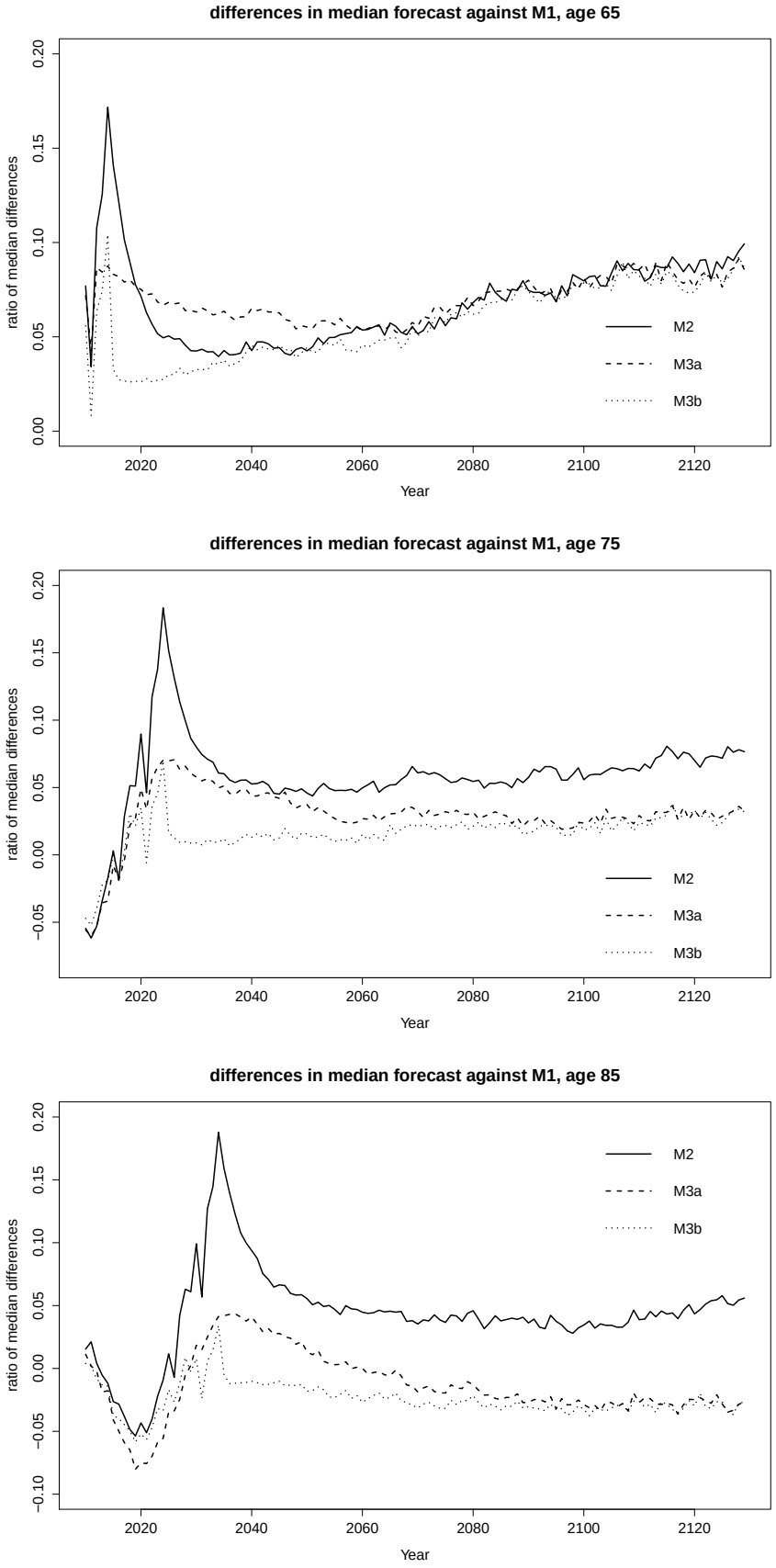

Figure 9: Ratio of Equation (19) for all models for ages of 65, 75 and 85 across the full 120 years forecasting period.

We conducted both sensitivity and backtesting analysis for the parameters and forecasts of all the models involved in the article, and generally conclude that the estimation time-frame is critical for the behaviour of the forecasted rates. The method is easily extendable in several ways. A more complex component may be used for the non-stationary mortality component, which would result in the residuals structure being lighter and the multivariate autoregressive relationship weaker. In such a case, it would be interesting to track the behaviour of the full covariance structure of V_Z of M3b. Conversely, a more complex model, such a higher order VAR, may be considered for the stationary mortality component. The autoregressive parameters of models M3, in particular of α_2 and α_5, are then expected to be mostly affected.

Acknowledgments

We are grateful to an anonymous referee for their helpful comments on an early version of the article.

References

Booth, H., Maindonald, J., and Smith, L. (2002). **"Applying Lee–Carter under conditions of variable mortality decline"**. *Population Studies*, 56: 325–336.

Brouhns, N., Denuit, M., and Van Keilegom, I. (2005). **"Bootstrapping the Poisson log-bilinear model for mortality forecasting"**. *Scandinavian Actuarial Journal*, 2005: 212–224.

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

Cairns, A. J. G., 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: 687–718.

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

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., and Balevich, I. (2009). **"A quantitative comparison of stochastic mortality models using data from England and Wales and the United States"**. *North American Actuarial Journal*, 13: 1–35.

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

D'Amato, V., Haberman, S., Piscopo, G., and Russolillo, M. (2012). **"Modelling dependent data for longevity projections"**. *Insurance: Mathematics and Economics*, 51: 694–701.

Debón, A., Martínez-Ruiz, F., and Montes, F. (2010). **"A geostatistical approach for dynamic life tables: The effect of mortality on remaining lifetime and annuities"**. *Insurance: Mathematics and Economics*, 47: 327–336.

Debón, A., Montes, F., Mateu, J., Porcu, E., and Bevilacqua, M. (2008). **"Modelling residuals dependence in dynamic life tables: A geostatistical approach"**. *Computational Statistics & Data Analysis*, 52: 3128–3147.

Gamerman, D. and Lopes, H. F. (2006). ***"Markov chain Monte Carlo: Stochastic simulation for Bayesian inference"***. CRC Press.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). ***"Bayesian data analysis"***. CRC press.

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

Lee, R. D. and Carter, L. R. (1992). **"Modeling and forecasting US mortality"**. *Journal of the American statistical association*, 87: 659–671.

Lütkepohl, H. (2005). ***"New introduction to multiple time series analysis"***. Cambridge University Press.

Pedroza, C. (2006). **"A Bayesian forecasting model: predicting US male mortality"**. *Biostatistics*, 7: 530–550.

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

Renshaw, A. E. and Haberman, S. (2003). **"Lee–Carter mortality forecasting with age–specific enhancement"**. *Insurance: Mathematics and Economics*, 33: 255–272.

Renshaw, A. E. and Haberman, S. (2006). **"A cohort–based extension to the Lee–Carter model for mortality reduction factors"**. *Insurance: Mathematics and Economics*, 38: 556–570.

Appendix A: Full conditional posterior distributions

Model M1

Model M1 is specified by the following relationships:

$$\begin{aligned} D(x,t) &\sim \text{Poisson}\left(E(x,t)m(x,t)\right), \\ \log\left(m(x,t)\right) &= \alpha_x + \kappa_t^{(1)} + \kappa_t^{(2)}\left(x-\bar{x}\right), \\ \left(\kappa_t^{(1)},\kappa_t^{(2)}\right) = \boldsymbol{\kappa}_t &= \boldsymbol{\delta} + \boldsymbol{\kappa}_{t-1} + \boldsymbol{\zeta}_t, \qquad \boldsymbol{\zeta}_t \sim \text{N}_2\left(0, V_\zeta\right). \end{aligned}$$

The complete parameter vector for model M1 is:

$$\boldsymbol{\theta} = \left\{\alpha_x, \boldsymbol{\kappa}_t, \boldsymbol{\delta}, V_\zeta\right\},$$

and the full log-posterior distribution is:

$$\begin{aligned} \pi\left(\boldsymbol{\theta}\right) = & \sum_{x,t}\left\{D(x,t)\left(\alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)}\right) - E(x,t)\exp\left(\alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)}\right)\right\} - \\ & -\frac{1}{2}\left\{(n_t-1)\log\left(|V_\zeta|\right) + \sum_{j=1}^{n_t-1}\left(\boldsymbol{\kappa}_{j+1}-\boldsymbol{\kappa}_j-\boldsymbol{\delta}\right)' V_\zeta^{-1}\left(\boldsymbol{\kappa}_{j+1}-\boldsymbol{\kappa}_j-\boldsymbol{\delta}\right)\right\} - \\ & -\frac{1}{2}\left\{\left(\boldsymbol{\delta}-\boldsymbol{\delta}_0\right)' V_0^{-1}\left(\boldsymbol{\delta}-\boldsymbol{\delta}_0\right) + 3\log(|V_\zeta|)\right\} + C, \end{aligned}$$

where C is a generic constant which does not depend on $\boldsymbol{\theta}$.

For fixed age-x, the full conditional log-posterior of α_x is:

$$\pi\left(\alpha_x|\boldsymbol{\theta}_{-i}\right) = \alpha_x\sum_t D(x,t) - \exp(\alpha_x)\sum_t E(x,t)\exp\left(\kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)}\right) + C,$$

so that:

$$\exp(\alpha_x)|\boldsymbol{\theta}_{-i} \sim \text{Gamma}\left(\sum_t D(x,t), \sum_t E(x,t)\exp\left(\kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)}\right)\right).$$

For fixed year, t, define the functions:

$$g\left(\boldsymbol{\kappa}_t\right) = \sum_x\left\{D(x,t)\left(\kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)}\right) - E(x,t)\exp\left(\alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)}\right)\right\}.$$

The full conditional log-posteriors for the first, intermediate and final calendar years of the data, 1, j, and n_t, respectively, are as follows:

$$\begin{aligned} \pi\left(\boldsymbol{\kappa}_1|\boldsymbol{\theta}_{-i}\right) &= g\left(\boldsymbol{\kappa}_1\right) - \frac{\left(\boldsymbol{\kappa}_2-\boldsymbol{\kappa}_1-\boldsymbol{\delta}\right)' V_\zeta^{-1}\left(\boldsymbol{\kappa}_2-\boldsymbol{\kappa}_1-\boldsymbol{\delta}\right)}{2} + C, \\ \pi\left(\boldsymbol{\kappa}_j|\boldsymbol{\theta}_{-i}\right) &= g\left(\boldsymbol{\kappa}_j\right) - \frac{\left(\boldsymbol{\kappa}_{j+1}-\boldsymbol{\kappa}_j-\boldsymbol{\delta}\right)' V_\zeta^{-1}\left(\boldsymbol{\kappa}_{j+1}-\boldsymbol{\kappa}_j-\boldsymbol{\delta}\right) + \left(\boldsymbol{\kappa}_j-\boldsymbol{\kappa}_{j-1}-\boldsymbol{\delta}\right)' V_\zeta^{-1}\left(\boldsymbol{\kappa}_j-\boldsymbol{\kappa}_{j-1}-\boldsymbol{\delta}\right)}{2} + C, \\ \pi\left(\boldsymbol{\kappa}_{n_t}|\boldsymbol{\theta}_{-i}\right) &= g\left(\boldsymbol{\kappa}_{n_t}\right) - \frac{\left(\boldsymbol{\kappa}_{n_t}-\boldsymbol{\kappa}_{n_t-1}-\boldsymbol{\delta}\right)' V_\zeta^{-1}\left(\boldsymbol{\kappa}_{n_t}-\boldsymbol{\kappa}_{n_t-1}-\boldsymbol{\delta}\right)}{2} + C. \end{aligned}$$

The full conditional log-posteriors of the drift, $\boldsymbol{\delta}$, and the covariance matrix, V_ζ, of the random walk model model for $\boldsymbol{\kappa}_t$ are as follows:

$$\begin{aligned}
\pi\left(\boldsymbol{\delta}|\boldsymbol{\theta}_{-i}\right) &= -\frac{1}{2}\left\{\boldsymbol{\delta}^{'}\left((n_t-1)V_\zeta^{-1}+V_0^{-1}\right)\boldsymbol{\delta}-2\boldsymbol{\delta}^{'}\left(V_\zeta^{-1}\left(\sum_{j=1}^{n_t-1}\left(\boldsymbol{\kappa}_{j+1}-\boldsymbol{\kappa}_j\right)\right)+V_0^{-1}\boldsymbol{\delta}_0\right)\right\}+C,\\
\pi\left(V_\zeta|\boldsymbol{\theta}_{-i}\right) &= -\frac{1}{2}\left\{(n_t+2)\log\left(|V_\zeta|\right)+\mathrm{tr}\left(\left(\sum_{j=1}^{n_t-1}\left(\boldsymbol{\kappa}_{j+1}-\boldsymbol{\kappa}_j-\boldsymbol{\delta}\right)\left(\boldsymbol{\kappa}_{j+1}-\boldsymbol{\kappa}_j-\boldsymbol{\delta}\right)^{'}\right)V_\zeta^{-1}\right)\right\}+C,
\end{aligned}$$

so that:

$$\begin{aligned}
\boldsymbol{\delta}|\boldsymbol{\theta}_{-i} &\sim \mathrm{N}_2\left(\left((n_t-1)V_\zeta^{-1}+V_0^{-1}\right)^{-1}\left(V_\zeta^{-1}\left(\sum_{j=1}^{n_t-1}\left(\boldsymbol{\kappa}_{j+1}-\boldsymbol{\kappa}_j\right)\right)+V_0^{-1}\boldsymbol{\delta}_0\right),\left((n_t-1)V_\zeta^{-1}+V_0^{-1}\right)^{-1}\right),\\
V_\zeta|\boldsymbol{\theta}_{-i} &\sim \mathrm{IW}\left(n_t-1,\left(\sum_{j=1}^{n_t-1}\left(\boldsymbol{\kappa}_{j+1}-\boldsymbol{\kappa}_j-\boldsymbol{\delta}\right)\left(\boldsymbol{\kappa}_{j+1}-\boldsymbol{\kappa}_j-\boldsymbol{\delta}\right)^{'}\right)^{-1}\right). \qquad (20)
\end{aligned}$$

Model M2

Model M2 is specified by the following relationships:

$$\begin{aligned}
D(x,t) &\sim \mathrm{Poisson}\left(E(x,t)m(x,t)\right),\\
\log\left(m(x,t)\right) &= \alpha_x+\kappa_t^{(1)}+\kappa_t^{(2)}\left(x-\bar{x}\right)+\gamma_c,\\
\left(\kappa_t^{(1)},\kappa_t^{(2)}\right)=\boldsymbol{\kappa}_t &= \boldsymbol{\delta}+\boldsymbol{\kappa}_{t-1}+\boldsymbol{\zeta}_t, \qquad \boldsymbol{\zeta}_t\sim\mathrm{N}_2\left(0,V_\zeta\right),\\
\gamma_c &= \delta_\gamma+\alpha_\gamma\gamma_{c-1}+\zeta_c, \quad \zeta_c\sim\mathrm{N}\left(0,\sigma_\gamma^2\right).
\end{aligned}$$

The complete parameter vector for model M2 is:

$$\boldsymbol{\theta}=\left\{\alpha_x,\boldsymbol{\kappa}_t,\gamma_c,\boldsymbol{\delta},V_\zeta,\alpha_\gamma,\delta_\gamma,\sigma_\gamma^2\right\},$$

and the full log-posterior distribution is:

$$\begin{aligned}
\pi\left(\boldsymbol{\theta}\right) = &\sum_{x,t}\left\{D(x,t)\left(\alpha_x+\kappa_t^{(1)}+(x-\bar{x})\kappa_t^{(2)}+\gamma_c\right)-E(x,t)\exp\left(\alpha_x+\kappa_t^{(1)}+(x-\bar{x})\kappa_t^{(2)}+\gamma_c\right)\right\}-\\
&-\frac{1}{2}\left\{(n_t-1)\log\left(|V_\zeta|\right)+\sum_{j=1}^{n_t-1}\left(\boldsymbol{\kappa}_{j+1}-\boldsymbol{\kappa}_j-\boldsymbol{\delta}\right)^{'}V_\zeta^{-1}\left(\boldsymbol{\kappa}_{j+1}-\boldsymbol{\kappa}_j-\boldsymbol{\delta}\right)\right\}-\\
&-\frac{1}{2}\left\{(n_c-1)\log\left(\sigma_\gamma^2\right)+\frac{1}{\sigma_\gamma^2}\left(\left(1-\alpha_\gamma^2\right)\left(\gamma_1-\frac{\delta_\gamma}{1-\alpha_\gamma}\right)^2+\sum_{j=2}^{n_c}\left(\gamma_j-\delta_\gamma-\alpha_\gamma\gamma_{j-1}\right)^2\right)\right\}-\\
&-\frac{1}{2}\log\left(\frac{\sigma_\gamma^2}{1-\alpha_\gamma^2}\right)-\frac{1}{2}\left\{\left(\boldsymbol{\delta}-\boldsymbol{\delta}_0\right)^{'}V_0^{-1}\left(\boldsymbol{\delta}-\boldsymbol{\delta}_0\right)+3\log(|V_\zeta|)\right\}-(a+1)\log\left(\sigma_\gamma^2\right)-\frac{b}{\sigma_\gamma^2}+C,
\end{aligned}$$

where C is a generic constant which does not depend on $\boldsymbol{\theta}$.

For fixed age-x, the full conditional log-posterior of α_x is:

$$\pi\left(\alpha_x|\boldsymbol{\theta}_{-i}\right)=\alpha_x\sum_t D(x,t)-\exp(\alpha_x)\sum_t E(x,t)\exp\left(\kappa_t^{(1)}+(x-\bar{x})\kappa_t^{(2)}+\gamma_c\right)+C,$$

so that:

$$\exp(\alpha_x)|\boldsymbol{\theta}_{-i} \sim \text{Gamma}\left(\sum_t D(x,t), \sum_t E(x,t)\exp\left(\kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + \gamma_c\right)\right).$$

For fixed year, t, define the functions:

$$g^*\left(\boldsymbol{\kappa}_t\right) = \sum_x \left\{D(x,t)\left(\kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)}\right) - E(x,t)\exp\left(\alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + \gamma_c\right)\right\}.$$

The full conditional log-posteriors of the period effects vectors, $\boldsymbol{\kappa}_t$, for the first, intermediate and final calendar years of the data, 1, j, and n_t, respectively, are as follows:

$$\begin{aligned}
\pi\left(\boldsymbol{\kappa}_1|\boldsymbol{\theta}_{-i}\right) &= g^*\left(\boldsymbol{\kappa}_1\right) - \frac{\left(\boldsymbol{\kappa}_2 - \boldsymbol{\kappa}_1 - \boldsymbol{\delta}\right)' V_\zeta^{-1}\left(\boldsymbol{\kappa}_2 - \boldsymbol{\kappa}_1 - \boldsymbol{\delta}\right)}{2} + C, \\
\pi\left(\boldsymbol{\kappa}_j|\boldsymbol{\theta}_{-i}\right) &= g^*\left(\boldsymbol{\kappa}_j\right) - \frac{\left(\boldsymbol{\kappa}_{j+1} - \boldsymbol{\kappa}_j - \boldsymbol{\delta}\right)' V_\zeta^{-1}\left(\boldsymbol{\kappa}_{j+1} - \boldsymbol{\kappa}_j - \boldsymbol{\delta}\right) + \left(\boldsymbol{\kappa}_j - \boldsymbol{\kappa}_{j-1} - \boldsymbol{\delta}\right)' V_\zeta^{-1}\left(\boldsymbol{\kappa}_j - \boldsymbol{\kappa}_{j-1} - \boldsymbol{\delta}\right)}{2} + C, \\
\pi\left(\boldsymbol{\kappa}_{n_t}|\boldsymbol{\theta}_{-i}\right) &= g^*\left(\boldsymbol{\kappa}_{n_t}\right) - \frac{\left(\boldsymbol{\kappa}_{n_t} - \boldsymbol{\kappa}_{n_t-1} - \boldsymbol{\delta}\right)' V_\zeta^{-1}\left(\boldsymbol{\kappa}_{n_t} - \boldsymbol{\kappa}_{n_t-1} - \boldsymbol{\delta}\right)}{2} + C.
\end{aligned}$$

The full conditional log-posteriors of the drift, $\boldsymbol{\delta}$, and the covariance matrix, V_ζ, of the random walk model model for $\boldsymbol{\kappa}_t$ are not affected by the existence of the cohort process and they are identical to those of model M1.

For fixed year of birth, c, define the functions:

$$h\left(\gamma_c\right) = \sum_{c=t-x} \left\{D(x,t)\gamma_c - E(x,t)\exp\left(\alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + \gamma_c\right)\right\}.$$

The full conditional log-posteriors of the cohort process states, γ_c, for the first, intermediate and final years of birth of the data, 1, j, and n_c, respectively, are as follows:

$$\begin{aligned}
\pi\left(\gamma_1|\boldsymbol{\theta}_{-i}\right) &= h\left(\gamma_1\right) - \frac{1}{2\sigma_\gamma^2}\left(\left(1-\alpha_\gamma^2\right)\left(\gamma_1 - \frac{\delta_\gamma}{1-\alpha_\gamma}\right)^2 + \left(\gamma_2 - \alpha_\gamma\gamma_1 - \delta_\gamma\right)^2\right) + C, \\
\pi\left(\gamma_j|\boldsymbol{\theta}_{-i}\right) &= h\left(\gamma_j\right) - \frac{1}{2\sigma_\gamma^2}\left(\left(\gamma_{j+1} - \alpha_\gamma\gamma_j - \delta_\gamma\right)^2 + \left(\gamma_j - \alpha_\gamma\gamma_{j-1} - \delta_\gamma\right)^2\right) + C, \\
\pi\left(\gamma_{n_c}|\boldsymbol{\theta}_{-i}\right) &= h\left(\gamma_{n_c}\right) - \frac{1}{2\sigma_\gamma^2}\left(\gamma_{n_c} - \alpha_\gamma\gamma_{n_c-1} - \delta_\gamma\right)^2 + C.
\end{aligned}$$

For the drift, δ_γ, and the autoregression parameter, α_γ, we employ the same full conditional log-posterior, namely:

$$\pi\left(\delta_\gamma|\boldsymbol{\theta}_{-i}\right) = \pi\left(\alpha_\gamma|\boldsymbol{\theta}_{-i}\right) = -\frac{1}{2\sigma_\gamma^2}\left(\left(1-\alpha_\gamma^2\right)\left(\gamma_1 - \frac{\delta_\gamma}{1-\alpha_\gamma}\right)^2 + \sum_{j=1}^{n_c-1}\left(\gamma_{j+1} - \alpha_\gamma\gamma_j - \delta_\gamma\right)^2\right) + C.$$

The full conditional log-posterior of σ_γ^2 is given as:

$$\pi\left(\sigma_\gamma^2|\boldsymbol{\theta}_{-i}\right) = -\frac{n_c}{2}\log\left(\sigma_\gamma^2\right) - \frac{1}{2\sigma_\gamma^2}\left(\left(1-\alpha_\gamma^2\right)\left(\gamma_1 - \frac{\delta_\gamma}{1-\alpha_\gamma}\right)^2 + \sum_{j=1}^{n_c-1}\left(\gamma_{j+1} - \alpha_\gamma\gamma_j - \delta_\gamma\right)^2\right) + C,$$

so that:

$$\sigma_\gamma^2|\boldsymbol{\theta}_{-i} \sim \text{IG}\left(\frac{n_c}{2} - 1, \left(1-\alpha_\gamma^2\right)\left(\gamma_1 - \frac{\delta_\gamma}{1-\alpha_\gamma}\right)^2 + \sum_{j=1}^{n_c-1}\left(\gamma_{j+1} - \alpha_\gamma\gamma_j - \delta_\gamma\right)^2\right).$$

Models M3

Models M3 are specified by the following relationships:

$$\begin{aligned} D(x,t) &\sim \text{Poisson}\left(E(x,t)m(x,t)\right), \\ \log\left(m(x,t)\right) &= \alpha_x + \kappa_t^{(1)} + \kappa_t^{(2)}\left(x-\bar{x}\right) + \mathcal{R}(x,t), \\ \left(\kappa_t^{(1)}, \kappa_t^{(2)}\right) = \boldsymbol{\kappa}_t &= \boldsymbol{\delta} + \boldsymbol{\kappa}_{t-1} + \boldsymbol{\zeta}_t, \qquad \boldsymbol{\zeta}_t \sim \text{N}_2\left(0, V_\zeta\right), \\ \boldsymbol{\mathcal{R}}_t &= \mathcal{A}\boldsymbol{\mathcal{R}}_{t-1} + \boldsymbol{Z}_t, \qquad \boldsymbol{Z}_t \sim N\left(0, V_Z\right). \end{aligned}$$

The complete parameter vector for models M3 is:

$$\boldsymbol{\theta} = \left\{\alpha_x, \boldsymbol{\kappa}_t, \boldsymbol{\mathcal{R}}_t, \boldsymbol{\delta}, V_\zeta, \mathcal{A}, V_Z\right\},$$

and the full log-posterior distribution is:

$$\begin{aligned} \pi\left(\boldsymbol{\theta}\right) &= \sum_{x,t}\left\{D(x,t)\left(\alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + \mathcal{R}(x,t)\right) - E(x,t)\exp\left(\alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + \mathcal{R}(x,t)\right)\right\} - \\ &-\frac{1}{2}\left\{(n_t-1)\log\left(|V_\zeta|\right) + \sum_{j=1}^{n_t-1}\left(\boldsymbol{\kappa}_{j+1} - \boldsymbol{\kappa}_j - \boldsymbol{\delta}\right)' V_\zeta^{-1}\left(\boldsymbol{\kappa}_{j+1} - \boldsymbol{\kappa}_j - \boldsymbol{\delta}\right)\right\} - \\ &-\frac{1}{2}\left\{(n_t-1)\log(|V_Z|) + \log(|V_{\mathcal{R}}|) + \boldsymbol{\mathcal{R}}_1' V_{\mathcal{R}}^{-1}\boldsymbol{\mathcal{R}}_1 + \sum_{j=1}^{n_t-1}\left(\boldsymbol{\mathcal{R}}_{j+1} - \mathcal{A}\boldsymbol{\mathcal{R}}_j\right)' V_Z^{-1}\left(\boldsymbol{\mathcal{R}}_{j+1} - \mathcal{A}\boldsymbol{\mathcal{R}}_j\right)\right\} - \\ &-\frac{1}{2}\left\{\left(\boldsymbol{\delta} - \boldsymbol{\delta}_0\right)' V_0^{-1}\left(\boldsymbol{\delta} - \boldsymbol{\delta}_0\right) + 3\log(|V_\zeta|)\right\} + C - \\ &-\left((a+1)\log(v) + \frac{b}{v}\right) \qquad \text{for Model M3a} \\ &-\frac{n_x+1}{2}\log(|V_Z|) \qquad \text{for Model M3b,} \end{aligned}$$

where C is a generic constant which does not depend on $\boldsymbol{\theta}$.

For fixed age-x, the full conditional log-posterior of α_x is:

$$\pi\left(\alpha_x|\boldsymbol{\theta}_{-i}\right) = \alpha_x\sum_t D(x,t) - \exp(\alpha_x)\sum_t E(x,t)\exp\left(\kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + \mathcal{R}(x,t)\right) + C,$$

so that:

$$\exp(\alpha_x)|\boldsymbol{\theta}_{-i} \sim \text{Gamma}\left(\sum_t D(x,t), \sum_t E(x,t)\exp\left(\kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + \mathcal{R}(x,t)\right)\right). \tag{21}$$

For fixed year, t, define the functions:

$$g^{**}\left(\boldsymbol{\kappa}_t\right) = \sum_x\left\{D(x,t)\left(\kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)}\right) - E(x,t)\exp\left(\alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + \mathcal{R}(x,t)\right)\right\}.$$

The full conditional log-posteriors of the period effects vectors, $\boldsymbol{\kappa}_t$, for the first, intermediate and final calendar years of the data, 1, j, and n_t, respectively, are as follows:

$$\begin{aligned} \pi\left(\boldsymbol{\kappa}_1|\boldsymbol{\theta}_{-i}\right) &= g^{**}\left(\boldsymbol{\kappa}_1\right) - \frac{\left(\boldsymbol{\kappa}_2 - \boldsymbol{\kappa}_1 - \boldsymbol{\delta}\right)' V_\zeta^{-1}\left(\boldsymbol{\kappa}_2 - \boldsymbol{\kappa}_1 - \boldsymbol{\delta}\right)}{2} + C, \\ \pi\left(\boldsymbol{\kappa}_j|\boldsymbol{\theta}_{-i}\right) &= g^{**}\left(\boldsymbol{\kappa}_j\right) - \frac{\left(\boldsymbol{\kappa}_{j+1} - \boldsymbol{\kappa}_j - \boldsymbol{\delta}\right)' V_\zeta^{-1}\left(\boldsymbol{\kappa}_{j+1} - \boldsymbol{\kappa}_j - \boldsymbol{\delta}\right) + \left(\boldsymbol{\kappa}_j - \boldsymbol{\kappa}_{j-1} - \boldsymbol{\delta}\right)' V_\zeta^{-1}\left(\boldsymbol{\kappa}_j - \boldsymbol{\kappa}_{j-1} - \boldsymbol{\delta}\right)}{2} + C, \\ \pi\left(\boldsymbol{\kappa}_{n_t}|\boldsymbol{\theta}_{-i}\right) &= g^{**}\left(\boldsymbol{\kappa}_{n_t}\right) - \frac{\left(\boldsymbol{\kappa}_{n_t} - \boldsymbol{\kappa}_{n_t-1} - \boldsymbol{\delta}\right)' V_\zeta^{-1}\left(\boldsymbol{\kappa}_{n_t} - \boldsymbol{\kappa}_{n_t-1} - \boldsymbol{\delta}\right)}{2} + C. \end{aligned} \tag{22}$$

The full conditional log-posteriors of the drift, $\boldsymbol{\delta}$, and the covariance matrix, V_ζ, of the random walk model model for $\boldsymbol{\kappa}_t$ are not affected by the existence of the residuals process and they are identical to those of model M1 and M2.

For fixed year, t, define the functions:

$$h^*\left(\boldsymbol{\mathcal{R}}_t\right) = \sum_x \left\{ D(x,t)\mathcal{R}(x,t) - E(x,t)\exp\left(\alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + \mathcal{R}(x,t)\right)\right\}.$$

The full conditional log-posteriors of the vectors of residuals, $\boldsymbol{\mathcal{R}}_t$, for the first, intermediate and final calendar years of the data, 1, j, and n_t, respectively, are as follows:

$$\begin{aligned}
\pi\left(\boldsymbol{\mathcal{R}}_1|\boldsymbol{\theta}_{-i}\right) &= h^*\left(\boldsymbol{\mathcal{R}}_1\right) - \frac{\boldsymbol{\mathcal{R}}_1^{'} V_{\mathcal{R}}^{-1}\boldsymbol{\mathcal{R}}_1 + \left(\boldsymbol{\mathcal{R}}_2 - \mathcal{A}\boldsymbol{\mathcal{R}}_1\right)^{'} V_Z^{-1}\left(\boldsymbol{\mathcal{R}}_2 - \mathcal{A}\boldsymbol{\mathcal{R}}_1\right)}{2} + C, \\
\pi\left(\boldsymbol{\mathcal{R}}_j|\boldsymbol{\theta}_{-i}\right) &= h^*\left(\boldsymbol{\mathcal{R}}_j\right) - \frac{\left(\boldsymbol{\mathcal{R}}_{j+1} - \mathcal{A}\boldsymbol{\mathcal{R}}_j\right)^{'} V_Z^{-1}\left(\boldsymbol{\mathcal{R}}_{j+1} - \mathcal{A}\boldsymbol{\mathcal{R}}_j\right) + \left(\boldsymbol{\mathcal{R}}_j - \mathcal{A}\boldsymbol{\mathcal{R}}_{j-1}\right)^{'} V_Z^{-1}\left(\boldsymbol{\mathcal{R}}_j - \mathcal{A}\boldsymbol{\mathcal{R}}_{j-1}\right)}{2} + C, \\
\pi\left(\boldsymbol{\mathcal{R}}_{n_t}|\boldsymbol{\theta}_{-i}\right) &= h^*\left(\boldsymbol{\mathcal{R}}_{n_t}\right) - \frac{\left(\boldsymbol{\mathcal{R}}_{n_t} - \mathcal{A}\boldsymbol{\mathcal{R}}_{n_t-1}\right)^{'} V_Z^{-1}\left(\boldsymbol{\mathcal{R}}_{n_t} - \mathcal{A}\boldsymbol{\mathcal{R}}_{n_t-1}\right)}{2} + C. \qquad (23)
\end{aligned}$$

For all parameters of the autoregressive matrix $\mathcal{A}$, α_i, $i = 1, \ldots, 6$, we employ the same full conditional log-posterior, namely:

$$\pi\left(\alpha_i|\boldsymbol{\theta}_{-i}\right) = -\frac{1}{2}\left(\boldsymbol{\mathcal{R}}_1^{'} V_{\mathcal{R}}^{-1}\boldsymbol{\mathcal{R}}_1 + \sum_{j=1}^{n_t-1}\left[\left(\boldsymbol{\mathcal{R}}_{j+1} - \mathcal{A}\boldsymbol{\mathcal{R}}_j\right)^{'} V_Z^{-1}\left(\boldsymbol{\mathcal{R}}_{j+1} - \mathcal{A}\boldsymbol{\mathcal{R}}_j\right)\right]\right) + C. \qquad (24)$$

The full conditional log-posterior distribution of the covariance matrix V_Z is what distinguishes models M3a and M3b.

For model M3a, the full conditional log-posterior for parameter v is:

$$\begin{aligned}
\pi\left(v|\boldsymbol{\theta}_{-i}\right) &= -\frac{1}{2}\left\{(n_t - 1)\log(|V_Z|) + \log(|V_{\mathcal{R}}|) + \boldsymbol{\mathcal{R}}_1^{'} V_{\mathcal{R}}^{-1}\boldsymbol{\mathcal{R}}_1 + \sum_{j=1}^{n_t-1}\left(\boldsymbol{\mathcal{R}}_{j+1} - \mathcal{A}\boldsymbol{\mathcal{R}}_j\right)^{'} V_Z^{-1}\left(\boldsymbol{\mathcal{R}}_{j+1} - \mathcal{A}\boldsymbol{\mathcal{R}}_j\right)\right\} - \\
&\quad - \left((a+1)\log(v) + \frac{b}{v}\right) + C.
\end{aligned}$$

For model M3b, the full conditional log-posterior for the full matrix V_Z is:

$$\begin{aligned}
\pi\left(V_Z|\boldsymbol{\theta}_{-i}\right) &= -\frac{1}{2}\left\{(n_x + n_t)\log\left(|V_Z|\right) + \mathrm{tr}\left(\left(\sum_{j=1}^{n_t-1}\left(\boldsymbol{\mathcal{R}}_{j+1} - \mathcal{A}\boldsymbol{\mathcal{R}}_j\right)\left(\boldsymbol{\mathcal{R}}_{j+1} - \mathcal{A}\boldsymbol{\mathcal{R}}_j\right)^{'}\right) V_Z^{-1}\right)\right\} - \\
&\quad - \frac{1}{2}\left(\log\left(|V_{\mathcal{R}}|\right) + \boldsymbol{\mathcal{R}}_1^{'} V_{\mathcal{R}}^{-1}\boldsymbol{\mathcal{R}}_1\right) + C. \qquad (25)
\end{aligned}$$

The first term of the above density is exactly the following Inverse Wishart distribution:

$$\mathrm{IW}\left(n_t - 1, \left(\sum_{j=1}^{n_t-1}\left(\boldsymbol{\mathcal{R}}_{j+1} - \mathcal{A}\boldsymbol{\mathcal{R}}_j\right)\left(\boldsymbol{\mathcal{R}}_{j+1} - \mathcal{A}\boldsymbol{\mathcal{R}}_j\right)^{'}\right)^{-1}\right), \qquad (26)$$

and will be used as the proposal distribution for V_Z of M3b within the respective MCMC.

Appendix B: Constraints

Model M1

For model M1 we need to satisfy the constraints of equation (10), namely:

$$\sum_t \kappa_t^{(i)} = 0, \quad i = 1, 2. \tag{27}$$

Given any set of period effects $\kappa_t^{(i)}$, we write:

$$\widetilde{\kappa}_t^{(i)} = \kappa_t^{(i)} - \phi_\kappa^{(i)}, \tag{28}$$

and if

$$\widetilde{\phi}_\kappa^{(i)} = \frac{1}{n} \sum_t \kappa_t^{(i)},$$

the parameter sets of equation (28) satisfy constraints (10). Hence, the estimates for M1 are obtained as:

$$\begin{aligned} \widetilde{\kappa}_t^{(i)} &= \kappa_t^{(i)} - \widetilde{\phi}_\kappa^{(i)}, \quad i = 1, 2 \\ \widetilde{\alpha}_x &= \alpha_x + \widetilde{\phi}_\kappa^{(1)} + (x - \bar{x})\, \widetilde{\phi}_\kappa^{(2)} \end{aligned} \tag{29}$$

Model M2

For model M2 we need to additionally satisfy the constraints of equation (11), namely:

$$\sum_c \gamma_c = 0, \quad \sum_c (c - \bar{c})\, \gamma_c = 0, \quad \sum_c (c - \bar{c})^2\, \gamma_c = 0. \tag{30}$$

Given any set of cohort effects, γ_c, we write:

$$\widetilde{\gamma}_c = \gamma_c - \phi_1 - \phi_2 (c - \bar{c}) - \phi_3 (c - \bar{c})^2. \tag{31}$$

The solution

$$\begin{pmatrix} \widetilde{\phi}_1 \\ \widetilde{\phi}_2 \\ \widetilde{\phi}_3 \end{pmatrix} = \begin{pmatrix} n_c & 0 & \sum_c (c - \bar{c})^2 \\ 0 & \sum_c (c - \bar{c})^2 & 0 \\ \sum_c (c - \bar{c})^2 & 0 & \sum_c (c - \bar{c})^4 \end{pmatrix}^{-1} \begin{pmatrix} \sum_c \gamma_c \\ \sum_c (c - \bar{c})\, \gamma_c \\ \sum_c (c - \bar{c})^2\, \gamma_c \end{pmatrix}$$

is used to identify the required cohort effects. The revised estimates are then given as:

$$\begin{aligned} \widetilde{\alpha}_x &= \alpha_x + \widetilde{\phi}_1 + \widetilde{\phi}_3 (x - \bar{x})^2 \\ \widetilde{\kappa}_t^{(1)} &= \kappa_t^{(1)} + \widetilde{\phi}_2 (t - \bar{t}) + \widetilde{\phi}_2 (t - \bar{t})^2 \\ \widetilde{\kappa}_t^{(2)} &= \kappa_t^{(2)} - \widetilde{\phi}_2 (t - \bar{t}) - 2\widetilde{\phi}_3 (t - \bar{t}) \\ \widetilde{\gamma}_c &= \gamma_c - \widetilde{\phi}_1 - \widetilde{\phi}_2 (c - \bar{c}) - \widetilde{\phi}_3 (c - \bar{c})^2. \end{aligned} \tag{32}$$

Given these estimates, equations (29) are then implemented to ensure constraints (10).

Model M3

For models M3 we need to additionally satisfy the constraints of equations (12), namely:

$$\sum_x \mathcal{R}(x,t) = 0, \quad \sum_x (x-\bar{x})\mathcal{R}(x,t) = 0 \text{ and } \sum_x \alpha_x \mathcal{R}(x,t) = 0, \forall t. \tag{33}$$

For each vectors of residuals, $\boldsymbol{\mathcal{R}}_t$, at time t, we write:

$$\widetilde{\boldsymbol{\mathcal{R}}_t} = \boldsymbol{\mathcal{R}}_t - \phi_1 - \phi_2 (x-\bar{x}) - \phi_3 \alpha_x. \tag{34}$$

The solution

$$\begin{pmatrix} \widetilde{\phi}_1 \\ \widetilde{\phi}_2 \\ \widetilde{\phi}_3 \end{pmatrix} = \begin{pmatrix} n_x & 0 & \sum_x \alpha_x \\ 0 & \sum_x (x-\bar{x})^2 & \sum_x \alpha_x (x-\bar{x}) \\ \sum_x \alpha_x & \sum_x \alpha_x (x-\bar{x}) & \sum_x \alpha_x^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_x \boldsymbol{\mathcal{R}}_t \\ \sum_x \boldsymbol{\mathcal{R}}_t (x-\bar{x}) \\ \sum_x \alpha_x \boldsymbol{\mathcal{R}}_t \end{pmatrix}$$

is used to identify the required residuals estimates. The revised parameters after each vector of residuals update are then given as:

$$\begin{aligned} \widetilde{\alpha}_x &= \alpha_x \left(\widetilde{\phi}_3 + 1 \right) \\ \widetilde{\kappa_t}^{(1)} &= \kappa_t^{(1)} + \widetilde{\phi}_1 \\ \widetilde{\kappa_t}^{(2)} &= \kappa_t^{(2)} + \widetilde{\phi}_2 \\ \widetilde{\boldsymbol{\mathcal{R}}_t} &= \boldsymbol{\mathcal{R}}_t - \widetilde{\phi}_2 - \widetilde{\phi}_2 (x-\bar{x}) - \widetilde{\phi}_3 \alpha_x \end{aligned} \tag{35}$$

Given these estimates, equations (29) are then implemented to ensure constraints (10).

Appendix C: Pseudo-algorithm of MCMC for model M3b

Initialise the parameter vector

$$\boldsymbol{\theta}^{(0)} = \left\{ [\alpha_x]^{(0)}, [\boldsymbol{\kappa}_t]^{(0)}, [\boldsymbol{\mathcal{R}}_t]^{(0)}, [\boldsymbol{\delta}]^{(0)}, [V_\zeta]^{(0)}, [\mathcal{A}]^{(0)}, [V_Z]^{(0)} \right\}$$

and set the number of iterations, M.

For $j = 1, 2, \ldots, M$

- For $x = 1, 2, \ldots, n_x$

  Sample and update $[\alpha_x]^{(j)}$ from its full conditional distribution given by equation (21).

- For $t = 1, 2, \ldots, n_t$

  - Sample a proposed vector as $\widetilde{\boldsymbol{\kappa}_t} \sim \mathrm{N}\left([\boldsymbol{\kappa}_t]^{(j-1)}, S_{\boldsymbol{\kappa}} \right)$.

    * Given equations (22), calculate the acceptance ratio:
      $\mathrm{r} = \min\left\{ 1, \exp\left(\pi\left(\widetilde{\boldsymbol{\kappa}_t} | \boldsymbol{\theta}_{-i} \right) - \pi\left([\boldsymbol{\kappa}_t]^{(j-1)} | \boldsymbol{\theta}_{-i} \right) \right) \right\}$

    * If $r > u \sim U(0,1)$, set: $[\boldsymbol{\kappa}_t]^{(j)} = \widetilde{\boldsymbol{\kappa}_t}$,
      otherwise set: $[\boldsymbol{\kappa}_t]^{(j)} = [\boldsymbol{\kappa}_t]^{(j-1)}$.

- For $t = 1, 2, \ldots, n_t$
  - Sample a proposed vector as $\widetilde{\boldsymbol{\mathcal{R}}}_t \sim \mathrm{N}\left([\boldsymbol{\mathcal{R}}_t]^{(j-1)}, S_{\boldsymbol{\mathcal{R}}}\right)$.
    - Given equations (23), calculate the acceptance ratio:
      $\mathrm{r} = \min\left\{1, \exp\left(\pi\left(\widetilde{\boldsymbol{\mathcal{R}}}_t|\boldsymbol{\theta}_{-i}\right) - \pi\left([\boldsymbol{\mathcal{R}}_t]^{(j-1)}|\boldsymbol{\theta}_{-i}\right)\right)\right\}$
    - If $r > u \sim U(0,1)$, set: $[\boldsymbol{\mathcal{R}}_t]^{(j)} = \widetilde{\boldsymbol{\mathcal{R}}}_t$ and apply equations (35),
      otherwise set: $[\boldsymbol{\mathcal{R}}_t]^{(j)} = [\boldsymbol{\mathcal{R}}_t]^{(j-1)}$.
- Apply equations (29).
- Sample for the parameters of the random walk, $[\boldsymbol{\delta}]^{(j)}, [V_\zeta]^{(j)}$, from their full conditional posteriors given in equations (20).
- For $i = 1, 2, \ldots, 6$
  - Sample for each element of the autoregressive matrix $\mathcal{A}$ as $\widetilde{\alpha}_i \sim \mathrm{N}\left([\alpha_i]^{(j-1)}, \sigma^2_{\alpha_i}\right)$.
    - Given equation (24), calculate the acceptance ratio:
      $\mathrm{r} = \min\left\{1, \exp\left(\pi\left(\widetilde{\alpha}_i|\boldsymbol{\theta}_{-i}\right) - \pi\left([\alpha_i]^{(j-1)}|\boldsymbol{\theta}_{-i}\right)\right)\right\}$
    - If $r > u \sim U(0,1)$, set: $[\alpha_i]^{(j)} = \widetilde{\alpha}_i$,
      otherwise set: $[\alpha_i]^{(j)} = [\alpha_i]^{(j-1)}$.
- Sample a proposed matrix, $\widetilde{V_Z}$, from the proposal distribution of equation (26) with density function q_V.
  - Given equation (25), calculate the corresponding acceptance ratio:
    $\mathrm{r} = \min\left\{1, \exp\left(\pi\left(\widetilde{V_Z}|\boldsymbol{\theta}_{-i}\right) + q_V\left([V_Z]^{(j-1)}\right) - \pi\left([V_Z]^{(j-1)}|\boldsymbol{\theta}_{-i}\right) - q_V\left(\widetilde{V_Z}\right)\right)\right\}$
  - If $r > u \sim U(0,1)$, set: $[V_Z]^{(j)} = \widetilde{V_Z}$,
    otherwise set: $[V_Z]^{(j)} = [V_Z]^{(j-1)}$.

End of iteration.