Journal of Econometrics and Statistics
Vol. 2, Issue 2, 2022, pp. 149-185
https://doi.org/10.47509/JES.2022.v02i02.02

Bayesian Poisson Log-normal Model with Regularized Time Structure for Multi-population Mortality Projection

Zhen Liu*, Xiaoqian Sun†, Leping Liu‡, Yu-Bo Wang§

Article History
Received : 30 May 2022; Revised : 29 June 2022; Accepted : 12 July 2022; Published : 15 December 2022

Abstract

Mortality projection is a pivotal topic in the diverse branches related to insurance, demography, and public policy. Motivated by the thread of Lee-Carter related models, we propose a Bayesian model to estimate and predict mortality rates for multi-population. This new model features in information borrowing among populations and properly reflecting variations of data. It also provides a solution to a long-time overlooked problem: model selection for dependence structures of population-specific time parameters. By introducing a novel dirac spike function that hierarchically follows the conditional autoregressive model via the probit link, simultaneous model selection and estimation for population-specific time effects can be achieved without much extra computational cost. Additionally, this selection procedure can leverage spatial information to inform about the geographic proximity in adjacent areas. Via the Brook's lemma and data augmentation steps, a computationally efficient MCMC sampling algorithm is also developed. We use the Japanese mortality data sets from Human Mortality Database to illustrate the desirable properties of our model.

keywords: Bayesian Poisson LC model, Dirac spike, Mortality projection, Conditional autoregressive model, Probit regression model.

1 INTRODUCTION

Mortality projection has become an important topic in demographics since it is greatly involved in many policy makings including but not limited to public health, pension, retirement system and labor resources. Especially for those developed and developing countries that experience population aging due to rapid growth of life expectancy and decline of birth rate after 1950s (Tuljapurkar et al., 2000), a

*School of Mathematical and Statistical Sciences, Clemson University.
†School of Mathematical and Statistical Sciences, Clemson University.
‡Department of Statistics, Tianjin University of Finance and Economics.
§School of Mathematical and Statistical Sciences, Clemson University. Corresponding author: yubow@clemson.edu

To cite this paper
Zhen Liu, Xiaoqian Sun, Leping Liu & Yu-Bo Wang (2022). Bayesian Poisson Log-normal Model with Regularized Time Structure for Multi-population Mortality Projection. *Journal of Econometrics and Statistics*. 2(2), 149-185.

thorough and well established policy relies on an accurate prediction of mortality trajectory.

Over last decades, stochastic models have been widely applied to mortality projection because the produced forecasts along with intervals can properly capture uncertainties over time and inform decision makings. The Lee-Carter (LC) model, a leading model proposed by Lee and Carter (1992), decomposes the centered mortality force in log scale as the product of age and time effects, and considers a random walk with drift model on time effect profile for the prediction purpose. This log-bilinear model was first developed for the U.S. mortality data from 1933 to 1987, and now becomes a benchmark widely implemented in all-cause or cause-specific mortality data. Following this structure, Brouhns et al. (2002) proposed the Poisson model for the number of deaths instead of directly modeling the observed mortality rate. Although it may encounter overdispersion due to the limitation of a Possion distribution, the Poisson LC model distinguishes the cases with the same observed rate but different exposures at risk, and hence, leverages more information from the data. On this basis, Czado et al. (2005) extended to a Bayesian framework to bypass the two-stage estimation procedure while preserving uncertainty from the model in the posterior predictive distributions of mortality rates. Wong et al. (2018) further introduced a random effect to accommodate overdispersion. Other related works can be referred to Girosi and King (2003), Renshaw and Haberman (2003), Cairns et al. (2006) and Plat (2009).

Motivated by the benefits of borrowing information among populations, many works, such as Li and Lee (2005), Cairns et al. (2011), Li and Hardy (2011), and Antonio et al. (2015), have focused on simultaneously projecting the mortality rates of multiple groups by encapsulating the common and population-specific age and time components in the models. In this paper, we revisit the works of Wong et al. (2018) and Antonio et al. (2015), and develop a new multi-population model that can address overdispersion in the count data and leverage spatial information in determining the dependence structures of population-specific time parameters. Specifically, we consider the autoregressive model of order one (AR(1)) with a drift for each population-specific time effect profile followed by a dirac spike setting (George and McCulloch, 1993; Ishwaran et al., 2005; Malsiner-Walli and Wagner, 2018) on the drift term and the slope associated with time. To improve this simultaneous model selection and estimation procedure with spatial information, the inclusion probability that controls binary switches between the spike and slab components is further fitted with a probit regression model, where the latent variable (Albert and Chib, 1993) for each population profile jointly follows a conditional autoregressive (CAR) model (Cressie and Chan, 1989; LeSage, 2000).

The remainder of the paper is organized as follows. In Section 2, we review the Lee-Carter model and its recent developments. Section 3 introduces the proposed model along with the prior settings and detailed steps of an Markov chain Monte Carlo (MCMC) sampling. In Section 4, we apply the proposed method to two Japanese mortality data sets from Human Mortality Database (HMD, 2021). Finally, we conclude with a discussion in Section 5.

2 LEE-CARTER MODEL AND ITS EXTENSIONS

Lee and Carter (1992) introduced a stochastic model for modeling the US mor-

tality data from 1933 to 1987 in an attempt to forecast the future mortality rate during 1988-2065. Suppose $\Theta_{\text{age}} = \{x_1, x_1+1, \ldots, x_1+M-1\} \equiv \{x_1, x_2, \ldots, x_M\}$ and $\Theta_{\text{time}} = \{t_1, t_1+1, \ldots, t_1+N-1\} \equiv \{t_1, t_2, \ldots, t_N\}$ denote the sets of age and time considered in the training dataset, respectively, the Lee-Carter model is then given by

$$\log m_{x,t} = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}, \tag{1}$$

where $m_{x,t}$ is the observed mortality rate for the group aged x at time t, $\epsilon_{x,t}$ is the error term, and $x \in \Theta_{\text{age}}$ and $t \in \Theta_{\text{time}}$. Essentially, this model is a special case of log-linear model in a cross table because $\log m_{x,t}$ is decomposed as the product of age (β_x) and time (κ_t) effects plus an age-specific intercept (α_x), where β_x is a constant over time while an additional time series model is placed on κ_t for the prediction purpose. To make α_x, β_x, and κ_t in (1) estimable, two constraints are imposed in Lee and Carter (1992): $\sum_{x\in\Theta_{\text{age}}} \beta_x = 1$ and $\sum_{t\in\Theta_{\text{time}}} \kappa_t = 0$. With such constraints, the age-specific intercept α_x is first estimated as the mean of log rates at age x observed across time, and then the singular value decomposition (SVD) is applied to the matrix of centered log rates, $\log m_{x,t} - \hat{\alpha}_x$, to estimate β_x and κ_t. Based on $\{\hat{\kappa}_t, \text{for } t \in \Theta_{\text{time}}\}$, the autoregressive integrated moving average (ARIMA) model is separately fitted to forecast the future time components κ_t and thus the mortality projection for any future year can be obtained.

Considering additional information contained in the exposure at risk ($E_{x,t}$), Brouhns et al. (2002) modified the LC model into the following Poisson framework

$$D_{x,t} \mid \mu_{x,t} \sim \text{Poisson}(E_{x,t}\mu_{x,t}) \quad \text{with} \quad \log \mu_{x,t} = \alpha_x + \beta_x \kappa_t, \tag{2}$$

where $D_{x,t}$ is the death toll for the group aged x at time t, and $\mu_{x,t}$ is the corresponding theoretic mortality rate. Note that $\mu_{x,t}$ differs from $m_{x,t} = D_{x,t}/E_{x,t}$ in (1), and that the cases with the same observed rate will have different likelihood values if their $E_{x,t}$s' differ. With the same constraints on β_x and κ_t, Brouhns et al. (2002) adopted the maximum likelihood estimation for α_x, β_x and κ_t in (2), and similarly, fitted $\{\hat{\kappa}_t, \text{for } t \in \Theta_{\text{time}}\}$ with the ARIMA model afterwards.

It is clear that both the LC and Poisson LC models are two-stage analyses, where the main model (that is, (1) or (2)) and the ARIMA model are fitted for estimation and prediction, respectively. Consequently, it may underestimate the uncertainty of the mortality projection. To properly reflect the uncertainty from the estimation process in the main model into forecasting, Czado et al. (2005) considered the Poisson LC model in Bayesian framework, where an MCMC sample is drawn from the posterior distribution of the joint model and used to construct the posterior predictive distribution of mortality rates in the future. Another efforts on improving the Poisson LC model can be found in Wong et al. (2018), where the proposed method tackles with overdispersion potentially encountered in the Poisson model. Letting $\nu_{x,t}$ denote a random effect following $N(0, \sigma^2)$, the normal distribution with mean 0 and variance σ^2, they proposed the Poisson log-normal Lee-Carter (PLNLC) model as

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

With this additional diffusion $\nu_{x,t}$, the PLNLC model relaxes the equality assump-

tion of mean and variance as follows

$$\begin{aligned}
\mathrm{E}[D_{x,t}] &= \mathrm{E}[\mathrm{E}(D_{x,t}|\nu_{x,t})] = E_{x,t}\exp(\alpha_x + \beta_x\kappa_t + \frac{1}{2}\sigma^2),\\
\mathrm{Var}[D_{x,t}] &= \mathrm{E}[\mathrm{Var}(D_{x,t}|\nu_{x,t})] + \mathrm{Var}[\mathrm{E}(D_{x,t}|\nu_{x,t})]\\
&= \mathrm{E}[D_{x,t}] \times \{1 + \mathrm{E}[D_{x,t}] \times [\exp(\sigma^2) - 1]\} \geq \mathrm{E}[D_{x,t}],
\end{aligned}$$

and hence, has a wider application in mortality data.

Besides, inspired from Li and Lee (2005) and Renshaw and Haberman (2003), the works considering two bilinear terms, Antonio et al. (2015) extended (2) to the following Poisson log-bilinear model for a n-population data set

$$D_{x,t}^{(i)} \mid \mu_{x,t}^{(i)} \sim \text{Poisson}(E_{x,t}^{(i)}\mu_{x,t}^{(i)}) \quad \text{with} \quad \log\mu_{x,t}^{(i)} = \alpha_x^{(i)} + \beta_x\kappa_t + \beta_x^{(i)}\kappa_t^{(i)}, \tag{4}$$

where the first bilinear term $\beta_x\kappa_t$ now denotes the overall effect shared by all populations aged x at time t, and the superscript (i) marks the i^{th} population-specific term so that $\alpha_x^{(i)}$ and $\beta_x^{(i)}\kappa_t^{(i)}$, for $i = 1, 2, , \ldots, n$, are a population-specific intercept and effect, respectively. To identify (4), additional constraints on population-specific age and time effects are required: $\left\|\beta_x^{(i)}\right\|_2 = 1$ and $\sum_{t\in\Theta_{\text{time}}}\kappa_t^{(i)} = 0$, where $\|.\|_2$ represents the L_2 norm of a vector. Through jointly investigating related populations, (4) tends to be more efficient than a separate modeling using PLC on each population.

In this paper, we consider pros and cons of the works mentioned above, and propose the Poisson Log-normal model for mortality projection of multi-population in the Bayesian framework. This new model not merely combines the PLNLC model with (4), but can also incorporate spatial information in model selection of time structures of $\kappa_t^{(i)}$ in a one-stage analysis. As a result, it can serve for more varieties of mortality data. We introduce the formulation of our model in Section 3.

3 THE PROPOSED MODEL

3.1 Bayesian Poisson Log-normal Lee-Carter Model with Regularized Time Structure for Multi-population

Let $\nu_{x,t}^{(i)}$ denote the i^{th} population-specific random effect following $N(0, \sigma_i^2)$ for $i = 1, 2, \ldots, n$. We propose the Bayesian Poisson log-normal Lee-Carter regularized model for n-population (BPLNLCrm) as follows

$$\begin{aligned}
&D_{x,t}^{(i)} \mid \mu_{x,t}^{(i)} \sim \text{Poisson}(E_{x,t}^{(i)}\mu_{x,t}^{(i)}) \quad \text{with} \quad \log\mu_{x,t}^{(i)} = \alpha_x^{(i)} + \beta_x\kappa_t + \beta_x^{(i)}\kappa_t^{(i)} + \nu_{x,t}^{(i)},\\
&\kappa_t = \varphi_1 + \varphi_2 t + \rho[\kappa_{t-1} - \varphi_1 - \varphi_2(t-1)] + \epsilon_t,\\
&\kappa_t^{(i)} = \varphi_1^{(i)} + \varphi_2^{(i)} t + \rho^{(i)}[\kappa_{t-1}^{(i)} - \varphi_1^{(i)} - \varphi_2^{(i)}(t-1)] + \epsilon_t^{(i)},
\end{aligned} \tag{5}$$

where $\epsilon_t \overset{i.i.d.}{\sim} N(0, \sigma_\kappa^2)$, $\epsilon_t^{(i)} \overset{i.i.d.}{\sim} N(0, \sigma_{\kappa^{(i)}}^2)$, $i = 1, 2, \ldots, n$, $x \in \Theta_{\text{age}}$, and $t \in \Theta_{\text{time}}$. A few comments are warranted. First, the first line of (5) can be viewed as a generalization of (3) to a multi-population problem while the last two equations describe the dependence structures of κ_t and $\kappa_t^{(i)}$.

Let

$$\boldsymbol{U}_{N\times N}=\begin{bmatrix}1 & 0 & \cdots & \cdots & 0\\ -\rho & 1 & & & \vdots\\ 0 & -\rho & \ddots & & \vdots\\ \vdots & \ddots & \ddots & \ddots & \\ 0 & \cdots & & -\rho & 1\end{bmatrix},\boldsymbol{U}^{(i)}_{N\times N}=\begin{bmatrix}1 & 0 & \cdots & \cdots & 0\\ -\rho^{(i)} & 1 & & & \vdots\\ 0 & -\rho^{(i)} & \ddots & & \vdots\\ \vdots & \ddots & \ddots & \ddots & \\ 0 & \cdots & & -\rho^{(i)} & 1\end{bmatrix},\boldsymbol{W}=\begin{bmatrix}1 & t_1\\ \vdots & \vdots\\ 1 & t_N\end{bmatrix},$$

and also define $\boldsymbol{\varphi}=(\varphi_1,\varphi_2)'$, $\boldsymbol{\varphi}^{(i)}=(\varphi_1^{(i)},\varphi_2^{(i)})'$, $\boldsymbol{Q}=\boldsymbol{U}'\boldsymbol{U}$, and $\boldsymbol{Q}^{(i)}=\boldsymbol{U}^{(i)'}\boldsymbol{U}^{(i)}$. These $n+1$ dependence structures of time effects can also be written as

$$\boldsymbol{\kappa}\sim N(\boldsymbol{W}\boldsymbol{\varphi},\sigma^2_{\kappa}\boldsymbol{Q}^{-1}),$$

and

$$\boldsymbol{\kappa}^{(i)}\sim N(\boldsymbol{W}\boldsymbol{\varphi}^{(i)},\sigma^2_{\kappa^{(i)}}(\boldsymbol{Q}^{(i)})^{-1}), \tag{6}$$

where $\boldsymbol{\kappa}=(\kappa_1,\kappa_2,\ldots,\kappa_N)'$ and $\boldsymbol{\kappa}^{(i)}=(\kappa_1^{(i)},\kappa_2^{(i)},\ldots,\kappa_N^{(i)})'$. Secondly, when all $\boldsymbol{\varphi}^{(i)}$'s are zeros, the dependence structures reduce back to the AR(1) model used in Antonio et al. (2015). Although Antonio et al. (2015) justified this special setting in certain applications, we prefer to consider a more general structure and let data speak out the truth of each $\varphi_1^{(i)}$ and $\varphi_2^{(i)}$. With such a motivation and considering the geographic proximity in adjacent areas, we develop a novel dirac spike function to each $\varphi_1^{(i)}$ and $\varphi_2^{(i)}$ as follows

$$\varphi_l^{(i)}\sim w_l^{(i)}N(0,\sigma^2_{\kappa^{(i)}})+(1-w_l^{(i)})\delta_l^{(i)}, \tag{7}$$

$$w_l^{(i)}\mid p_l^{(i)}\sim \text{Bernoulli}(p_l^{(i)}\equiv P(\lambda_l^{(i)}>0)), \tag{8}$$

$$\boldsymbol{\lambda}_l\mid\boldsymbol{\eta}_l,\tau_l\sim N(\boldsymbol{\eta}_l,(\boldsymbol{I}_n-\tau_l\boldsymbol{S})^{-1}), \tag{9}$$

where $w_l^{(i)}$ is a binary variable with success probability $p_l^{(i)}$ that is determined by the latent variable $\lambda_l^{(i)}$, $\delta_l^{(i)}$ is a point mass at zero, $\boldsymbol{\lambda}_l=(\lambda_l^{(1)},\lambda_l^{(2)},\ldots,\lambda_l^{(n)})'$, $\boldsymbol{\eta}_l=(\eta_l^{(1)},\eta_l^{(2)},\ldots,\eta_l^{(n)})'$, $\boldsymbol{I}_n$ is an identity matrix of size n, $\boldsymbol{S}$ is a $n\times n$ adjacency matrix with $s_{i,i'}=1$ if the i^{th} and i'^{th} regions are neighboring, $s_{i,i'}=0$ otherwise, and $l=1,2$.

A few comments are warranted. First, via the dirac spike setting in (7), our model can explore the model space of 2^{2n} possible dependence structures of $\boldsymbol{\kappa}^{(i)}$ in a single analysis; i.e., simultaneous model selection and parameter estimation and prediction. When n is big, it can greatly ease computation in model selection compared to the traditional criteria-based approaches, such as the marginal likelihood criterion and the Akaike information criterion. Besides, the introduced CAR model in (9) can incorporate spatial information into the selection procedure so that retain a more efficient MCMC sample. Also, note that when τ_l in (9) is set as zero, the proposed hierarchical model reduces back to the traditional beta prior for $p_l^{(i)}$ via the probit link. Therefore, our method can be viewed as a generalization of the standard dirac spike, and is also suitable for a multi-population problem when extra spatial information is unavailable. Lastly, unlike the original CAR model, where the covariance matrix is set as $\sigma_l^2(I_n-\tau_l\boldsymbol{S})^{-1}$, we specify $\sigma_l^2=1$ to avoid the

identification problem. As a result, the full conditional of $\lambda_l^{(i)}$ happens to be the same data augmentation as in Albert and Chib (1993) and is analytically tractable.

We also want to point out that with the same constraints as used in Antonio et al. (2015),

$$\sum_{x\in\Theta_{\text{age}}} \beta_x = 1,$$
$$\sum_{t\in\Theta_{\text{time}}} \kappa_t = 0,$$
$$\left\|\beta_x^{(i)}\right\|_2 = 1,$$
$$\sum_{t\in\Theta_{\text{time}}} \kappa_t^{(i)} = 0.$$

The interpretation of each parameter in (5) is similar to the one in Antonio et al. (2015). However, due to the existence of $\nu_{x,t}^{(i)}$, $\alpha_x^{(i)}$ can only be interpreted as the approximation of mean of log rates at age x across time in the i^{th} population. See Antonio et al. (2015) in details for the advantages of such a constraint setting.

3.2 Prior Specifications

3.2.1 Prior Distributions for Age Parameters

To assure the tractable full conditional distribution of $\alpha_x^{(i)}$, we conduct the same variable transformation $e_x^{(i)} = \exp(\alpha_x^{(i)})$ as Czado et al. (2005) and Antonio et al. (2015) and propose

$$e_x^{(i)} \sim \text{Gamma}(a_x^{(i)}, b_x^{(i)}),$$

with the corresponding density

$$\pi(e_x^{(i)}) = \frac{(b_x^{(i)})^{a_x^{(i)}}}{\Gamma(a_x^{(i)})}(e_x^{(i)})^{a_x^{(i)}-1}\exp(-e_x^{(i)}b_x^{(i)}),$$

where $a_x^{(i)}$ and $b_x^{(i)}$ are pre-specified constants. As for $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_M)'$ and $\boldsymbol{\beta}^{(i)} = (\beta_1^{(i)}, \beta_2^{(i)}, \ldots, \beta_M^{(i)})'$, we consider the following non-informative priors

$$\boldsymbol{\beta} \mid \sigma_\beta^2 \sim N\left(\frac{1}{M}\boldsymbol{J}_M, \sigma_\beta^2 \boldsymbol{I}_M\right),$$
$$\boldsymbol{\beta}^{(i)} \mid \sigma_{\beta^{(i)}}^2 \sim N\left(\frac{1}{M}\boldsymbol{J}_M, \sigma_{\beta^{(i)}}^2 \boldsymbol{I}_M\right),$$
$$\sigma_\beta^2 \sim \text{InvGamma}(a_\beta, b_\beta),$$
$$\sigma_{\beta^{(i)}}^2 \sim \text{InvGamma}(a_\beta^{(i)}, b_\beta^{(i)}),$$

where $\boldsymbol{J}_M$ is a $M \times 1$ vector with all elements equal to 1, and a_β, b_β, $a_\beta^{(i)}$, and $b_\beta^{(i)}$ are pre-specified constants. Note that the proposed priors for $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^{(i)}$ are non-informative in the sense that they are centered at $1/M$; i.e., the constraint (=1) equally shared by M age groups. Also, note that the density of an InvGamma(a_β, b_β) random variable σ_β^2 is given by

$$\pi(\sigma_\beta^2) = \frac{b_\beta^{a_\beta}}{\Gamma(a_\beta)}(\sigma_\beta^2)^{-a_\beta-1}\exp(-b_\beta/\sigma_\beta^2).$$

3.2.2 Prior Distributions for Time Parameters

We consider the following priors for the parameters associated with κ

$$\begin{aligned}
&\boldsymbol{\varphi} \sim N_2(\boldsymbol{\varphi}_0, \boldsymbol{\Sigma}_0),\\
&\rho \sim N(0, \sigma_\rho^2)\mathbf{1}\left\{\rho \in (-1,1)\right\},\\
&\sigma_\kappa^2 \sim \text{InvGamma}(a_\kappa, b_\kappa),
\end{aligned}$$

where $\boldsymbol{\varphi}_0$, $\boldsymbol{\Sigma}_0$, σ_ρ^2, a_κ, and b_κ are pre-specified hyperparameters, and $\mathbf{1}\left\{\rho \in (-1,1)\right\}$ is an indicator function equal to 1 when ρ is between -1 and 1.

As for $\boldsymbol{\kappa}^{(i)}$ in (6), we follow the proposed model in (7)-(9) and have

$$\begin{aligned}
&\boldsymbol{\eta}_l \sim N(\boldsymbol{\eta}_0, \boldsymbol{I}_n),\\
&\tau_l \sim Unif(\tau_{min}^{-1}, \tau_{max}^{-1}),\\
&\rho^{(i)} \sim N(0, \sigma_{\rho^{(i)}}^2)\mathbf{1}\left\{\rho^{(i)} \in (-1,1)\right\},\\
&\sigma_{\kappa^{(i)}}^2 \sim \text{InvGamma}(a_\kappa^{(i)}, b_\kappa^{(i)}),
\end{aligned}$$

where τ_{min} and τ_{max} are minimum and maximum eigenvalues of the adjacent matrix $\boldsymbol{S}$, respectively, and $\boldsymbol{\eta}_0$, $\sigma_{\rho^{(i)}}^2$, $a_\kappa^{(i)}$, and $b_\kappa^{(i)}$ are pre-specified hyperparameters.

3.2.3 Prior Distributions for Overdispersion Parameters

Last, following the practical purpose as mentioned in Gelman et al. (2006), we assign an Inverse Gamma distribution for σ_i^2

$$\sigma_i^2 \sim \text{InvGamma}(a_\mu^{(i)}, b_\mu^{(i)}),$$

where $a_\mu^{(i)}$ and $b_\mu^{(i)}$ are pre-specified.

3.3 Posterior Computation

3.3.1 Posterior Distributions for Age Parameters

Let $\boldsymbol{\theta} = (e^{(1)}, e^{(2)}, \ldots, e^{(n)}, \boldsymbol{\beta}', (\boldsymbol{\beta}^{(1)})', (\boldsymbol{\beta}^{(2)})', \ldots, (\boldsymbol{\beta}^{(n)})', \sigma_\beta^2, \sigma_{\beta^{(1)}}^2, \sigma_{\beta^{(2)}}^2, \ldots, \sigma_{\beta^{(n)}}^2,$
$\boldsymbol{\kappa}',$
$(\boldsymbol{\kappa}^{(1)})', (\boldsymbol{\kappa}^{(2)})', \ldots, (\boldsymbol{\kappa}^{(n)})', \boldsymbol{\varphi}', (\boldsymbol{\varphi}^{(1)})', (\boldsymbol{\varphi}^{(2)})', \ldots, (\boldsymbol{\varphi}^{(n)})', \rho, \rho^{(1)}, \rho^{(2)}, \ldots, \rho^{(n)}, \sigma_\kappa^2, \sigma_{\kappa^{(1)}}^2,$
$\sigma_{\kappa^{(2)}}^2, \ldots, \sigma_{\kappa^{(n)}}^2, w_1^{(1)}, w_1^{(2)}, \ldots, w_1^{(n)}, w_2^{(1)}, w_2^{(2)}, \ldots, w_2^{(n)}, \lambda_1^{(1)}, \lambda_1^{(2)}, \ldots, \lambda_1^{(n)}, \lambda_2^{(1)}, \lambda_2^{(2)}, \ldots, \lambda_2^{(n)},$
$\eta_1^{(1)}, \eta_1^{(2)}, \ldots, \eta_1^{(n)}, \eta_2^{(1)}, \eta_2^{(2)}, \ldots, \eta_2^{(n)}, \tau_1, \tau_2, \sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2, \nu_{x,t}^{(1)}, \nu_{x,t}^{(2)}, \ldots, \nu_{x,t}^{(n)})'$, and let the notation " | ." represent "conditional on all other parameters and the data G". The

full conditional distributions of age parameters are given by

$$\pi(e_x^{(i)} \mid .) \propto \exp(-c_x^{(i)} e_x^{(i)})(e_x^{(i)})^{D_{x,\cdot}^{(i)}} \left| \frac{d}{de_x^{(i)}} g^{-1}(\alpha_x^{(i)}) \right| \pi(e_x^{(i)})$$
$$\propto \exp\left[-(b_x^{(i)} + c_x^{(i)}) e_x^{(i)}\right] (e_x^{(i)})^{a_x^{(i)} + D_{x,\cdot}^{(i)} - 1}, \tag{10}$$

$$\pi(\beta_x \mid .) \propto \prod_{i=1}^{n} \prod_{t \in \Theta_{\text{time}}} \exp\left[\beta_x \kappa_t D_{x,t}^{(i)} - E_{x,t}^{(i)} \exp(\alpha_x^{(i)} + \beta_x \kappa_t + \beta_x^{(i)} \kappa_t^{(i)} + \nu_{x,t}^{(i)})\right]$$
$$\times \exp\left[-\frac{(\beta_x - \frac{1}{M})^2}{2\sigma_\beta^2}\right], \tag{11}$$

$$\pi(\beta_x^{(i)} \mid .) \propto \prod_{t \in \Theta_{\text{time}}} \exp\left[\beta_x^{(i)} \kappa_t^{(i)} D_{x,t}^{(i)} - E_{x,t}^{(i)} \exp(\alpha_x^{(i)} + \beta_x \kappa_t + \beta_x^{(i)} \kappa_t^{(i)} + \nu_{x,t}^{(i)})\right]$$
$$\times \exp\left[-\frac{(\beta_x^{(i)} - \frac{1}{M})^2}{2\sigma_{\beta^{(i)}}^2}\right], \tag{12}$$

and

$$\pi(\sigma_\beta^2 \mid .) \propto (\sigma_\beta^2)^{-\tilde{a}_\beta - 1} \exp(-\tilde{b}_\beta / \sigma_\beta^2), \tag{13}$$
$$\pi(\sigma_{\beta^{(i)}}^2 \mid .) \propto (\sigma_{\beta^{(i)}}^2)^{-\tilde{a}_\beta^{(i)} - 1} \exp(-\tilde{b}_\beta^{(i)} / \sigma_{\beta^{(i)}}^2), \tag{14}$$

where $c_x^{(i)} = \sum_{t \in \Theta_{\text{time}}} E_{x,t}^{(i)} \exp(\beta_x \kappa_t + \beta_x^{(i)} \kappa_t^{(i)} + \nu_{x,t}^{(i)})$, $D_{x,\cdot}^{(i)} = \sum_{t \in \Theta_{\text{time}}} D_{x,t}^{(i)} - 1$, $\tilde{a}_\beta = a_\beta + \frac{M}{2}$, $\tilde{b}_\beta = b_\beta + \frac{1}{2}\left(\boldsymbol{\beta} - \frac{1}{M}\boldsymbol{J}_M\right)'\left(\boldsymbol{\beta} - \frac{1}{M}\boldsymbol{J}_M\right)$, $\tilde{a}_\beta^{(i)} = a_\beta^{(i)} + \frac{M}{2}$, and $\tilde{b}_\beta^{(i)} = b_\beta^{(i)} + \frac{1}{2}\left(\boldsymbol{\beta}^{(i)} - \frac{1}{M}\boldsymbol{J}_M\right)'\left(\boldsymbol{\beta}^{(i)} - \frac{1}{M}\boldsymbol{J}_M\right)$. From (10), (13), and (14), we have

$$e_x^{(i)} \mid . \sim \text{Gamma}(a_x^{(i)} + D_{x,\cdot}^{(i)}, b_x^{(i)} + c_x^{(i)}),$$
$$\sigma_\beta^2 \mid . \sim \text{InvGamma}(\tilde{a}_\beta, \tilde{b}_\beta),$$
$$\sigma_{\beta^{(i)}}^2 \mid . \sim \text{InvGamma}(\tilde{a}_\beta^{(i)}, \tilde{b}_\beta^{(i)}),$$

and thus they can be easily sampled in each iteration.

Due to unidentifiable kernels in (11) and (12), the Metropolis-Hastings (MH) sampling is applied to update β_x and $\beta_x^{(i)}$. Specifically, let $\beta_x^{[j]}$ denote the j^{th} iteration of β_x, and let $\boldsymbol{A}\backslash\{\}$ denote all parameters in $\boldsymbol{A}$ except the ones in $\{\}$. Assuming that $\beta_y^{[j-1]}$ for $y \geq x$ and $\boldsymbol{\theta}^{[j]}\backslash\{\beta_y^{[j]} \text{ for } y \geq x\}$ are ready, we consider $\beta_x^* \sim N(\beta_x^{[j-1]}, \sigma_x^2)$ as the proposal distribution, where σ_x^2 is chosen to ensure the acceptance probability between 20% and 40%. With this symmetric proposal, the acceptance probability

$$\Phi(\beta_x^{[j-1]}, \beta_x^*) = \min\left\{1, \frac{\pi(\beta_x^* | \{\beta_y^{[j-1]} \text{for } y > x\}, \boldsymbol{\theta}^{[j]}\backslash\{\beta_y^{[j]} \text{ for } y \geq x\}, G)}{\pi(\beta_x^{[j-1]} | \{\beta_y^{[j-1]} \text{for } y > x\}, \boldsymbol{\theta}^{[j]}\backslash\{\beta_y^{[j]} \text{ for } y \geq x\}, G)}\right\}$$

is compared with a random value u from the Uniform(0,1) and

$$\beta_x^{[j]} = \begin{cases} \beta_x^* & \text{if } u \leq \Phi(\beta_x^{[j-1]}, \beta_x^*) \\ \beta_x^{[j-1]} & \text{o.w.} \end{cases}.$$

To satisfy the constraint $\sum_{x\in\Theta_{\text{age}}}\beta_x = 1$, we further calculate $\tilde{B} = \sum_{y\le x}\beta_y^{[j]} + \sum_{y>x}B_y^{[j-1]}$, and update

$$(\beta_{x_1}^{[j]},\ldots,\beta_x^{[j]},\beta_{x+1}^{[j-1]},\ldots,\beta_{x_M}^{[j-1]}) \leftarrow \frac{(\beta_{x_1}^{[j]},\ldots,\beta_x^{[j]},\beta_{x+1}^{[j-1]},\ldots,\beta_{x_M}^{[j-1]})}{\tilde{B}}$$

and $\boldsymbol{\kappa}^{[j]} \leftarrow \boldsymbol{\kappa}^{[j]}\tilde{B}$. We repeat all steps above till $x = x_M$ to retain the j^{th} iteration of $\boldsymbol{\beta}$. For updating $\beta_x^{(i)}$ (and also $\boldsymbol{\kappa}_t^{(i)}$), the similar steps are implemented except that $\tilde{B}$ is now calculated as the L_2 norm.

3.3.2 Posterior Distributions for Time Parameters

In this section, we separately discuss the sampling algorithms for common and population-specific time parameters since the dependence structure of the latter is further regularized by the dirac spike setting and available spatial information. Let $\boldsymbol{\kappa}_{-t} = \boldsymbol{\kappa}\backslash\{\kappa_t\}$; i.e., $(\kappa_1,\ldots,\kappa_{t-1},\kappa_{t+1},\ldots,\kappa_{t_N})'$, and $\eta_t = \varphi_1 + \varphi_2 t$, the full conditional distributions of $\boldsymbol{\kappa},\boldsymbol{\varphi},\rho$, and σ_κ^2 are proportional to

$$\pi(\kappa_t \mid .) \propto \prod_{i=1}^{n}\prod_{x\in\Theta_{\text{age}}} \exp\left[\beta_x\kappa_t D_{x,t}^{(i)} - E_{x,t}^{(i)}\exp(\alpha_x^{(i)} + \beta_x\kappa_t + \beta_x^{(i)}\kappa_t^{(i)} + \nu_{x,t}^{(i)})\right] \times f(\kappa_t \mid \boldsymbol{\kappa}_{-t}), \tag{15}$$

$$\pi(\boldsymbol{\varphi} \mid .) \propto \exp\left[-\frac{1}{2\sigma_\kappa^2}(\boldsymbol{\varphi}'(\boldsymbol{\Sigma}^*)^{-1}\boldsymbol{\varphi} - 2(\boldsymbol{\kappa}'\boldsymbol{Q}\boldsymbol{W} + \sigma_\kappa^2\boldsymbol{\varphi}_0'\boldsymbol{\Sigma}_0^{-1})\boldsymbol{\varphi})\right], \tag{16}$$

$$\pi(\rho \mid .) \propto \exp\left[-\frac{1}{2\sigma_\kappa^2}\left(a_\rho\rho^2 + \frac{\sigma_\kappa^2}{\sigma_\rho^2}\rho^2 - 2b_\rho\rho\right)\right]\mathbf{1}\{\rho\in(-1,1)\}, \tag{17}$$

$$\pi(\sigma_\kappa^2 \mid .) \propto (\sigma_\kappa^2)^{-(a_\kappa+N/2)-1}\exp\left[-\frac{1}{\sigma_\kappa^2}\left(b_\kappa + \frac{1}{2}(\boldsymbol{\kappa}-\boldsymbol{W}\boldsymbol{\varphi})'\boldsymbol{Q}(\boldsymbol{\kappa}-\boldsymbol{W}\boldsymbol{\varphi})\right)\right], \tag{18}$$

where $f(\kappa_t \mid \boldsymbol{\kappa}_{-t})$ are the conditional distribution of κ_t based on AR(1) with a drift in (5), $\boldsymbol{\Sigma}^* = (\boldsymbol{W}'\boldsymbol{Q}\boldsymbol{W} + \sigma_\kappa^2\boldsymbol{\Sigma}_0^{-1})^{-1}$, $a_\rho = \sum_{t=t_2}^{t_N}(\kappa_{t-1}-\eta_{t-1})^2$, and $b_\rho = \sum_{t=t_2}^{t_N}(\kappa_t - \eta_t)(\kappa_{t-1}-\eta_{t-1})$. Note that when $t = t_1$,

$$\begin{aligned} f(\kappa_t \mid \boldsymbol{\kappa}_{-t}) &\propto f(\kappa_t)f(\kappa_{t+1} \mid \kappa_t) \\ &\propto \exp\left[-\frac{1}{2\sigma_\kappa^2}[(\kappa_t-\eta_t)^2 + (\kappa_{t+1}-\eta_{t+1}-\rho(\kappa_t-\eta_t))^2]\right]; \end{aligned}$$

when $t_1 < t < t_N$,

$$\begin{aligned} f(\kappa_t \mid \boldsymbol{\kappa}_{-t}) &\propto f(\kappa_{t+1} \mid \kappa_t)f(\kappa_t \mid \kappa_{t-1}) \\ &\propto \exp\left[-\frac{1}{2\sigma_\kappa^2}[(\kappa_t-\eta_t-\rho(\kappa_{t-1}-\eta_{t-1}))^2 + (\kappa_{t+1}-\eta_{t+1}-\rho(\kappa_t-\eta_t))^2]\right]; \end{aligned}$$

and when $t = t_N$,

$$f(\kappa_t \mid \boldsymbol{\kappa}_{-t}) \propto f(\kappa_t \mid \kappa_{t-1}) \propto \exp\left[-\frac{1}{2\sigma_\kappa^2}(\kappa_t-\eta_t-\rho(\kappa_{t-1}-\eta_{t-1}))^2\right].$$

From (16), (17), and (18), $\boldsymbol{\varphi}, \rho$ and σ_κ^2 are updated by

$$\begin{aligned}
&\boldsymbol{\varphi} \mid . \sim N(\boldsymbol{\Sigma}^*(\boldsymbol{W}'\boldsymbol{Q}\boldsymbol{\kappa} + \sigma_\kappa^2\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\varphi}_0), \sigma_\kappa^2\boldsymbol{\Sigma}^*),\\
&\rho \mid . \sim N\left(\frac{b_\rho}{a_\rho + \frac{\sigma_\kappa^2}{\sigma_\rho^2}}, \frac{\sigma_\kappa^2}{a_\rho + \frac{\sigma_\kappa^2}{\sigma_\rho^2}}\right)\mathbf{1}\left\{\rho \in (-1,1)\right\},\\
&\sigma_\kappa^2 \mid . \sim \text{InvGamma}\left(a_\kappa + \frac{N}{2}, b_\kappa + \frac{1}{2}(\boldsymbol{\kappa} - \boldsymbol{W}\boldsymbol{\varphi})'\boldsymbol{Q}(\boldsymbol{\kappa} - \boldsymbol{W}\boldsymbol{\varphi})\right).
\end{aligned}$$

To update κ_t in (15), we let $\kappa_t^{[j]}$ denote the j^{th} iteration of κ_t and assume that $\kappa_z^{[j-1]}$ for $z \geq t$ and $\boldsymbol{\theta}^{[j]}\backslash\{\kappa_z^{[j]} \text{ for } z \geq t\}$ are available. Considering $\kappa_t^* \sim N(\kappa_t^{[j-1]}, \sigma_t^2)$ as the proposal for the MH sampling (similarly, σ_t^2 is selected to have the acceptance probability around $20\% \sim 40\%$), we then have

$$\kappa_t^{[j]} = \begin{cases} \kappa_t^* & \text{if } u \leq \Phi(\kappa_t^{[j-1]}, \kappa_t^*) \\ \kappa_t^{[j-1]} & \text{o.w.} \end{cases},$$

where

$$\Phi(\kappa_t^{[j-1]}, \kappa_t^*) = \min\left\{1, \frac{\pi(\kappa_t^*|\{\kappa_z^{[j-1]}\text{for } z > t\}, \boldsymbol{\theta}^{[j]}\backslash\{\kappa_z^{[j]} \text{ for } z \geq t\}, G)}{\pi(\kappa_t^{[j-1]}|\{\kappa_z^{[j-1]}\text{for } z > t\}, \boldsymbol{\theta}^{[j]}\backslash\{\kappa_z^{[j]} \text{ for } z \geq t\}, G)}\right\}$$

and $u \sim \text{Uniform}(0,1)$. With the constraint $\sum_{t\in\Theta_{\text{time}}}\kappa_t = 0$, we next have

$$(\kappa_{t_1}^{[j]}, \ldots, \kappa_t^{[j]}, \kappa_{t+1}^{[j-1]}, \ldots, \kappa_{t_N}^{[j-1]}) \leftarrow (\kappa_{t_1}^{[j]}, \ldots, \kappa_t^{[j]}, \kappa_{t+1}^{[j-1]}, \ldots, \kappa_{t_N}^{[j-1]}) - \tilde{K}$$

and $(\alpha_x^{(i)})^{[j]} \leftarrow (\alpha_x^{(i)})^{[j]} + \beta_x^{[j]}\tilde{K}$, where $\tilde{K} = (\sum_{z\leq t}\kappa_z^{[j]} + \sum_{z>t}\kappa_z^{[j-1]})/N$. Repeat all procedures till $t = t_N$ to complete the updates of $\boldsymbol{\kappa}^{[j]}$.

To update the parameters associated with population-specific time effect profiles, we first follow the Brook's lemma: $\lambda_l^{(i)} \mid \boldsymbol{\lambda}_l^{(-i)}, \eta_l^{(i)}, \tau_l \sim N(\mu_i^*, 1)$ to obtain

$$\begin{aligned}
&\pi(\lambda_l^{(i)} \mid .) \propto \exp\left[-\frac{1}{2}(\lambda_l^{(i)} - \mu_i^*)^2\right]\left(\mathbf{1}\{w_l^{(i)} = 1\}\mathbf{1}\{\lambda_l^{(i)} > 0\} + \mathbf{1}\{w_l^{(i)} = 0\}\mathbf{1}\{\lambda_l^{(i)} \leq 0\}\right),\\
&\pi(w_l^{(i)} = 1 \mid \boldsymbol{\lambda}_l, \eta_l^{(i)}, \tau_l) = p(\lambda_l^{(i)} > 0 \mid \boldsymbol{\lambda}_l^{(-i)}, \eta_l^{(i)}, \tau_l) = \Phi(\mu_i^*),
\end{aligned}$$

where $\boldsymbol{\lambda}_l^{(-i)} = \boldsymbol{\lambda}_l\backslash\lambda_l^{(i)}$, and $\mu_i^* = \eta_l^{(i)} + \tau_l\sum_{j\neq i} w_{i,j}(\lambda_l^{(j)} - \eta_l^{(j)})$. Accordingly, we have

$$\lambda_l^{(i)} \mid . \sim \begin{cases} N(\mu_i^*, 1)\mathbf{1}\{\lambda_l^{(i)} > 0\} & \text{if } w_l^{(i)} = 1, \\ N(\mu_i^*, 1)\mathbf{1}\{\lambda_l^{(i)} \leq 0\} & \text{if } w_l^{(i)} = 0, \end{cases}$$

and the status of $w_l^{(i)}$ is determined via

$$w_l^{(i)} \mid . \sim \text{Bernoulli}\left(\tilde{p}_l^{(i)} = \frac{m(w_l^{(i)} = 1)\Phi(\mu_i^*)}{m(w_l^{(i)} = 1)\Phi(\mu_i^*) + m(w_l^{(i)} = 0)[1 - \Phi(\mu_i^*)]}\right),$$

where $m(\cdot) = \exp\left[-(\boldsymbol{\kappa}^{(i)} - \boldsymbol{W}\boldsymbol{\varphi})^T\boldsymbol{Q}^{(i)}(\boldsymbol{\kappa}^{(i)} - \boldsymbol{W}\boldsymbol{\varphi})/(2\sigma_{\kappa^{(i)}}^2)\right]$ is the conditional marginal likelihood measuring the overall model fitting to the data when specifying the value

of $w_l^{(i)}$. Secondly, with the identifiable full conditional distributions, $\boldsymbol{\eta}_l$, $\boldsymbol{\varphi}^{(i)}$, $\rho^{(i)}$, and $\sigma^2_{\kappa^{(i)}}$ are updated by

$$\begin{aligned}
&\boldsymbol{\eta}_l \mid . \sim N\left((\boldsymbol{V}_l + \boldsymbol{I}_n)^{-1}(\boldsymbol{\eta}_0 + \boldsymbol{V}_l\boldsymbol{\lambda}_l), (\boldsymbol{V}_l + \boldsymbol{I}_n)^{-1}\right), \\
&\boldsymbol{\varphi}^{(i)} \mid . \sim N(\boldsymbol{a}^*, \boldsymbol{A}^*\sigma^2_{\kappa^{(i)}}) \quad \text{if} \quad w_1^{(i)} = w_2^{(i)} = 1, \\
&\rho^{(i)} \mid . \sim N\left(\frac{b_\rho^{(i)}}{a_\rho^{(i)} + \frac{\sigma^2_{\kappa^{(i)}}}{\sigma^2_{\rho^{(i)}}}}, \frac{\sigma^2_{\kappa^{(i)}}}{a_\rho^{(i)} + \frac{\sigma^2_{\kappa^{(i)}}}{\sigma^2_{\rho^{(i)}}}}\right) \mathbf{1}\left\{\rho^{(i)} \in (-1,1)\right\}, \\
&\sigma^2_{\kappa^{(i)}} \mid . \sim \text{InvGamma}\left(a_k^{(i)} + \frac{N}{2}, b_\kappa^{(i)} + \frac{1}{2}(\boldsymbol{\kappa}^{(i)} - \boldsymbol{W}\boldsymbol{\varphi}^{(i)})'\boldsymbol{Q}^{(i)}(\boldsymbol{\kappa}^{(i)} - \boldsymbol{W}\boldsymbol{\varphi}^{(i)})\right),
\end{aligned}$$

where $\boldsymbol{V}_l = I_n - \tau_l\boldsymbol{S}$, $\boldsymbol{a}^* = \boldsymbol{A}^*\boldsymbol{W}'\boldsymbol{Q}^{(i)}\boldsymbol{\kappa}^{(i)}$, $\boldsymbol{A}^* = (\boldsymbol{W}'\boldsymbol{Q}^{(i)}\boldsymbol{W} + \boldsymbol{I}_2)^{-1}$, $\eta_t^{(i)} = \varphi_1^{(i)} + \varphi_2^{(i)}t$, $a_\rho^{(i)} = \sum_{t=t_2}^{t_N}(\kappa_{t-1}^{(i)} - \eta_{t-1}^{(i)})^2$, $b_\rho^{(i)} = \sum_{t=t_2}^{t_N}(\kappa_t^{(i)} - \eta_t^{(i)})(\kappa_{t-1}^{(i)} - \eta_{t-1}^{(i)})$. Note that when $w_1^{(i)} = w_2^{(i)} = 0$, $\boldsymbol{\varphi}^{(i)}$ is simply updated as $(0,0)'$, and that when $w_{-l}^{(i)} = 0, w_l^{(i)} = 1$, $\varphi_{-l}^{(i)}$ is 0 while $\varphi_l^{(i)}$ is updated from the marginal normal distribution with mean and variance equal to the l^{th} and (l,l) elements in $\boldsymbol{a}^*$ and $\boldsymbol{A}^*\sigma^2_{\kappa^{(i)}}$, respectively. Next, due to a complex structure

$$\pi(\tau_l \mid .) \propto |\boldsymbol{V}_l|^{\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{\lambda}_l - \boldsymbol{\eta}_l)'\boldsymbol{V}_l(\boldsymbol{\lambda}_l - \boldsymbol{\eta}_l)\right] \mathbf{1}\{\tau_{min}^{-1}, \tau_{max}^{-1}\},$$

we conduct the MH sampling and consider $\tau_l^* \sim N(\tau_l^{[j-1]}, \sigma_l^2)\mathbf{1}\{\tau_{min}^{-1}, \tau_{max}^{-1}\}$ as the proposed density for τ_l. Finally, due to the same structure of $\pi(\kappa_t^{(i)} \mid .)$ as $\pi(\kappa_t \mid .)$ and the same constraint of $\kappa_t^{(i)}$ as κ_t, the procedure for $\kappa_t^{(i)}$ is similar to the one for κ_t.

3.3.3 Posterior Distributions for Overdispersion Parameters

Since the full conditional distributions of σ_i^2 and $\nu_{x,t}^{(i)}$ are proportional to

$$\begin{aligned}
&\pi(\sigma_i^2 \mid .) \propto (\sigma_i^2)^{-(a_\mu^{(i)} + MN/2)-1} \exp\left[-\frac{1}{\sigma_i^2}\left(b_\mu^{(i)} + \frac{1}{2}\sum_{x\in\Theta_{\text{age}}}\sum_{t\in\Theta_{\text{time}}}(\nu_{x,t}^{(i)})^2\right)\right], \\
&\pi(\nu_{x,t}^{(i)} \mid .) \propto \exp\left[-E_{x,t}^{(i)}\exp(\alpha_x^{(i)} + \beta_x\kappa_t + \beta_x^{(i)}\kappa_t^{(i)} + \nu_{x,t}^{(i)}) + \nu_{x,t}^{(i)}D_{x,t}^{(i)} - \frac{(\nu_{x,t}^{(i)})^2}{2\sigma_i^2}\right],
\end{aligned}$$

we have

$$\sigma_i^2 \mid . \sim \text{InvGamma}\left(a_\mu^{(i)} + \frac{MN}{2}, b_\mu^{(i)} + \frac{1}{2}\sum_{x\in\Theta_{\text{age}}}\sum_{t\in\Theta_{\text{time}}}(\nu_{x,t}^{(i)})^2\right),$$

and update $\nu_{x,t}^{(i)}$ via the MH sampling. Specifically, we propose $(\nu_{x,t}^{(i)})^* \sim N((\nu_{x,t}^{(i)})^{[j-1]}, \sigma_q^2)$ with σ_q^2 chosen to have the acceptance probability around 20%∼40%.

4 NUMERICAL ANALYSIS

4.1 Data Description

Two Japanese mortality data sets from the Japanese Mortality Database are used to illustrate our proposed method. In the first application, we apply to the national gender-specific deaths and exposures from 1951 to 2000 to retain the mortality projections from 2001 to 2016. In this two-population problem; i.e., each gender is considered as a single population, we pre-specify $\tau_1 = \tau_2 = 0$ and evaluate the performance of BPLNLCrm under a situation lacking spatial information. Then, our focus switches to the regional mortality rates, where the death tolls from 1975 to 2016 of six Northeast regions (Akita, Aomori, Fukushima, Iwate, Miyagi, and Yamagata) along with two Kanto regions (Tokyo and Kanagawa) are collectively investigated in the second analysis. By utilizing the well-defined adjacency matrix seized in our hierarchical structure, we empirically demonstrate the advantages of BPLNLCrm in both model selection procedure and prediction. Both analyses were run in R-4.2.1 using the Palmetto Cluster, which is Clemson University's primary high-performance computing resource.

4.2 Initial Settings of Prior Distributions

Following the prior specifications in Sections 3.2.1-3.2.3, we consider $a_x^{(i)} = b_x^{(i)} = 1$, $a_\beta = b_\beta = 0.01$, and $a_\beta^{(i)} = b_\beta^{(i)} = 0.001$ as age-related hyperparameters, and set $a_\mu^{(i)} = b_\mu^{(i)} = 2.5$ for overdispersion parameters. For those hyperparameters related to the time factors, they are set as $a = b = 1$, $a_\kappa = b_\kappa = 0.001$, $a_\kappa^{(i)} = b_\kappa^{(i)} = 0.001$, $\sigma_\rho^2 = 1$, $\sigma_{\rho^{(i)}}^2 = 0.1$, $\boldsymbol{\varphi}_0 = (0,0)'$, $\Sigma_0 = \text{diag}(10,10)$, and $\boldsymbol{\eta}_0 = (0.1, 0.1, \ldots, 0.1)'$, respectively. It has to be mentioned that the pre-specified values here are similar to the ones in Czado et al. (2005) and Antonio et al. (2015), and non-informative relative to the size of our analyzing data sets.

4.3 National Gender-specific Data

To evaluate the performance of BPLNLCrm, an MCMC sample of 20,000 iterations is generated with the first 10,000 as burn-ins. We then select the most dominant combination of $(w_1^{(F)}, w_2,^{(F)}, w_1^{(M)}, w_2^{(M)})$ in the dirac spike setting. Table 1 lists out the top four dependence structures of gender-specific time effect profiles in the MCMC chain. Accordingly, our final chosen models for $\boldsymbol{\kappa}^{(F)}$ and $\boldsymbol{\kappa}^{(M)}$ are given by

$$\kappa_t^{(F)} = \varphi_1^{(F)} + \varphi_2^{(F)} t + \rho^{(F)} \left(\kappa_{t-1}^{(F)} - \varphi_1^{(F)} - \varphi_2^{(F)} (t-1) \right) + \epsilon_t^{(F)},$$

and

$$\kappa_t^{(M)} = \varphi_1^{(M)} + \varphi_2^{(M)} t + \rho^{(M)} \left(\kappa_{t-1}^{(M)} - \varphi_1^{(M)} - \varphi_2^{(M)} (t-1) \right) + \epsilon_t^{(M)},$$

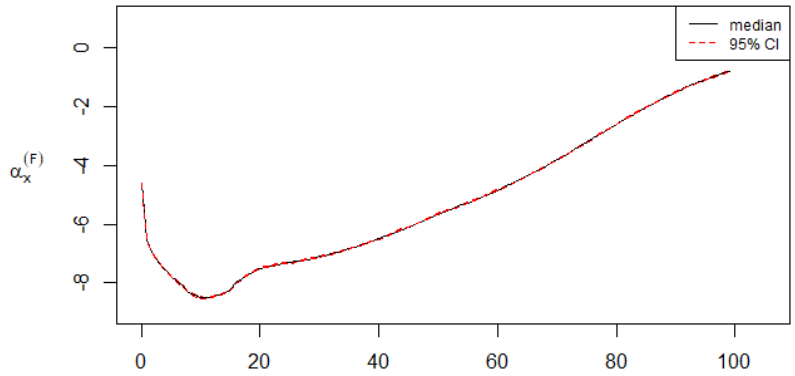
respectively. Based on this 47% of MCMC sample, we further present the posterior medians of $\boldsymbol{\alpha}^{(i)}$, $\boldsymbol{\beta}$, $\boldsymbol{\beta}^{(i)}$, $\boldsymbol{\kappa}$ and $\boldsymbol{\kappa}^{(i)}$ along with the 95% highest posterior density (HPD) intervals in Figures 1-3. Last, we assess the overall model fitting and prediction by comparing with the method by Antonio et al. (2015). For MCMC

convergence diagnostics, trace plots of selected parameters are also provided in the supplementary materials.

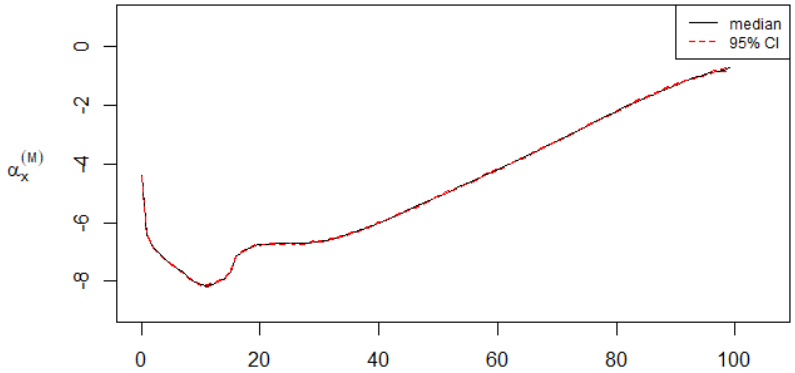
Table 1: Top four dependence structures of $\boldsymbol{\kappa}^{(F)}$ and $\boldsymbol{\kappa}^{(M)}$.

Dependence Structure	Proportion
$w_1^{(F)} = 1, w_1^{(M)} = 1, w_2^{(F)} = 1, w_2^{(M)} = 1$	0.47
$w_1^{(F)} = 1, w_1^{(M)} = 1, w_2^{(F)} = 1, w_2^{(M)} = 0$	0.08
$w_1^{(F)} = 1, w_1^{(M)} = 1, w_2^{(F)} = 0, w_2^{(M)} = 1$	0.06
$w_1^{(F)} = 1, w_1^{(M)} = 0, w_2^{(F)} = 0, w_2^{(M)} = 0$	0.05

Figures 1(a) and 1(b) present the results of $\alpha_x^{(F)}$ and $\alpha_x^{(M)}$ under the BPLNLCrm model, where the 95% HPD intervals are obtained by the method suggested in Hoff (2009). From Figure 1, we notice that the posterior distributions of $\alpha_x^{(F)}$ and $\alpha_x^{(M)}$ have small variances, and it is worth pointing out some features in the estimated curves: First, male tends to have a higher mortality rate than woman. This justifies our multi-population modeling in this case. Secondly, the decline from the infant stage to teenager is likely related to the immune system strengthened with growing age. Then, when ages are around 16-21, the health condition may not be the only decisive factor for the hump. It might be blamed on unnatural deaths caused by immature behaviors in this rebellious stage such as alcohol and drug uses, crimes, and careless drivings etc. For the adult-and-elder stage, the curves consistently go up since deaths happening in this stage are more relevant to natural causes.



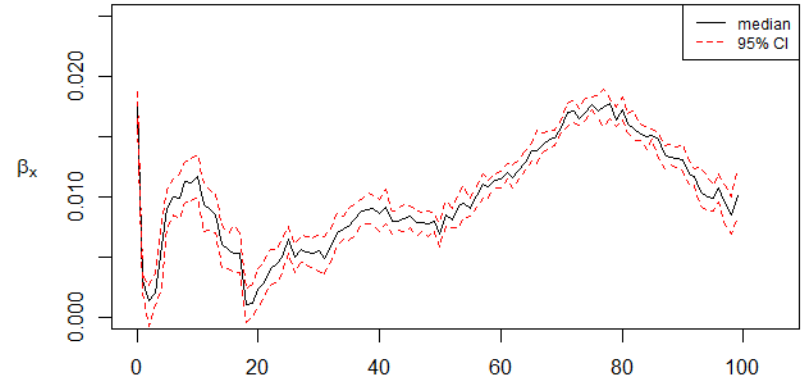
(a)



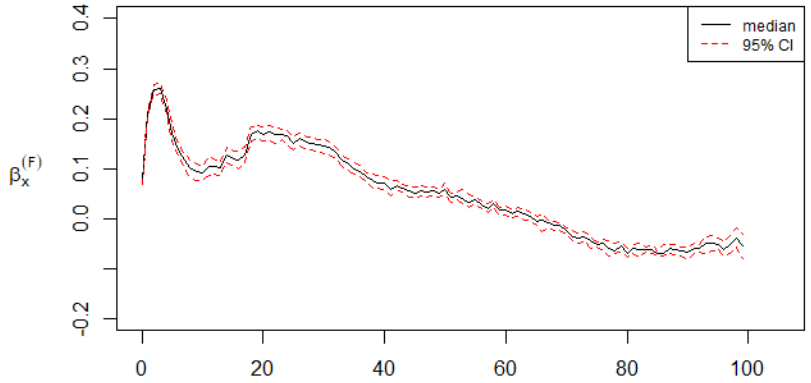
(b)

Figure 1: Plots of the posterior medians of $\alpha_x^{(F)}$ and $\alpha_x^{(M)}$ with their 95% HDP intervals.

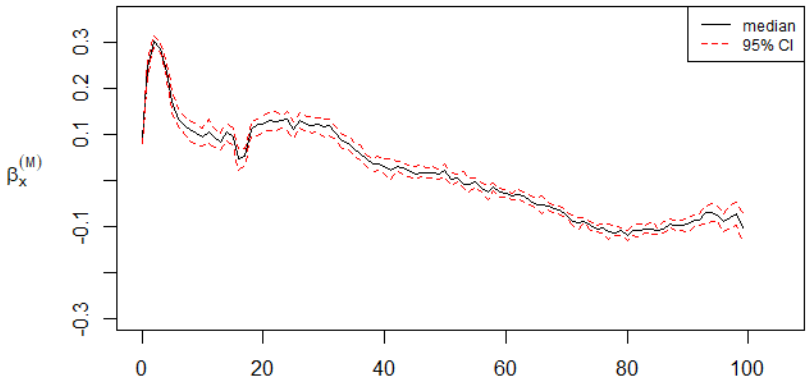
Similarly, Figure 2(a) shows the posterior median and 95% HPD interval of common factor β_x while Figures 2(b) and 2(c) are for the population-specific parameters $\beta_x^{(F)}$ and $\beta_x^{(M)}$, respectively. It can be seen that the corresponding posterior distributions are concentrated, indicating that the effect sizes of β_x and $\beta_x^{(i)}$ are not sensitive to all time change at any ages.



(a)



(b)



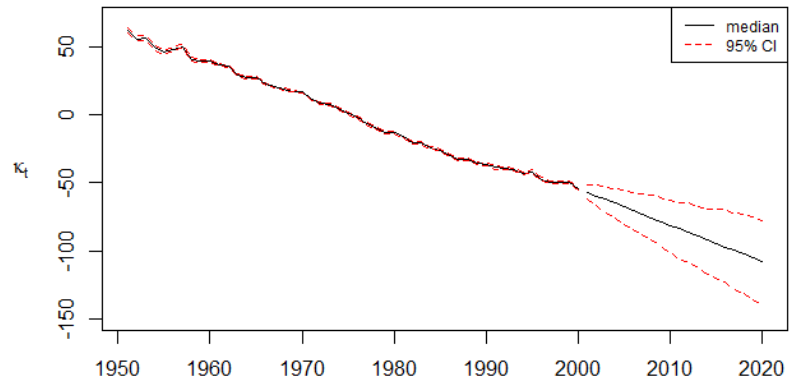
(c)

Figure 2: Plots of the posterior medians of β_x, $\beta_x^{(F)}$ and $\beta_x^{(M)}$ with their 95% HDP intervals.

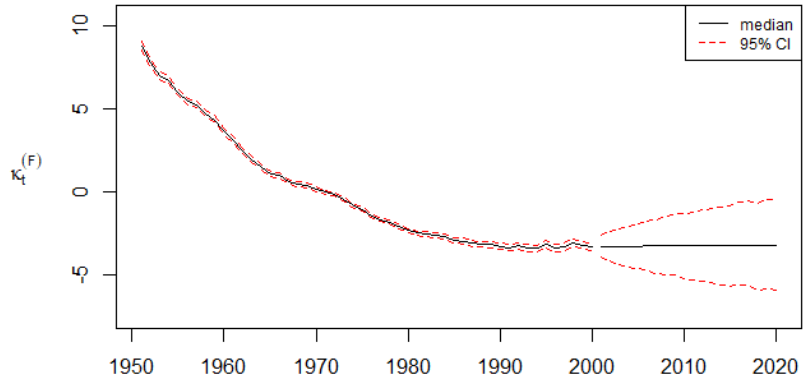
In Figures 3(a)-3(c), we present the posterior medians and 95% HPD intervals of κ_t, $\kappa_t^{(F)}$, and $\kappa_t^{(M)}$, respectively. In addition to the years 1951-2000, the follow-up 20-year ahead projections of time effects are also provided via the posterior predictive distributions. To obtain a sample from the posterior predictive distribution of κ_t, we exploit the second equation in (5) iteratively. Specifically, we have

$$\kappa_{N+t'}^{[j]} \sim N\left(\varphi_1^{[j]} + \varphi_2^{[j]}(N+t') + \rho^{[j]}[\kappa_{N+t'-1}^{[j]} - \varphi_1^{[j]} - \varphi_2^{[j]}(N+t'-1)], (\sigma_\kappa^2)^{[j]}\right)$$

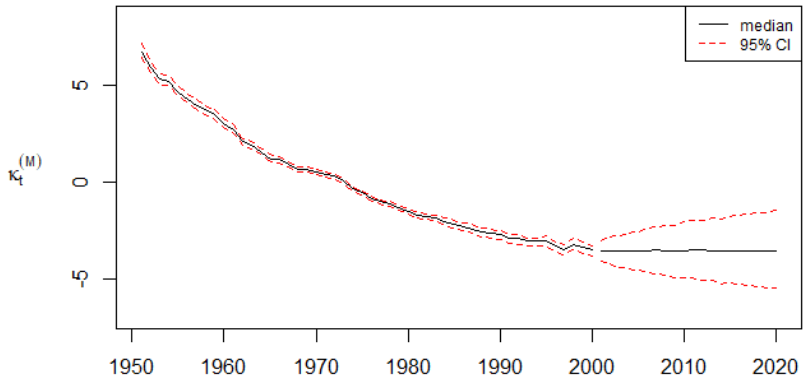
for $t' = 1, 2, \ldots, 20$. A similar procedure on the third equation in (5) is implemented for $\kappa_t^{(i)}$. From Figure 3, we observe decreasing trends in most of time windows, which might be attributed to the advances in medical technology and social welfare. We also observe that $\kappa_t^{(F)}$ and $\kappa_t^{(M)}$ have similar estimated curves and converge to the same size when time passes, meaning that the gender-specific time effects will reach a stable status in the long run as mentioned in Li and Lee (2005).



(a)



(b)



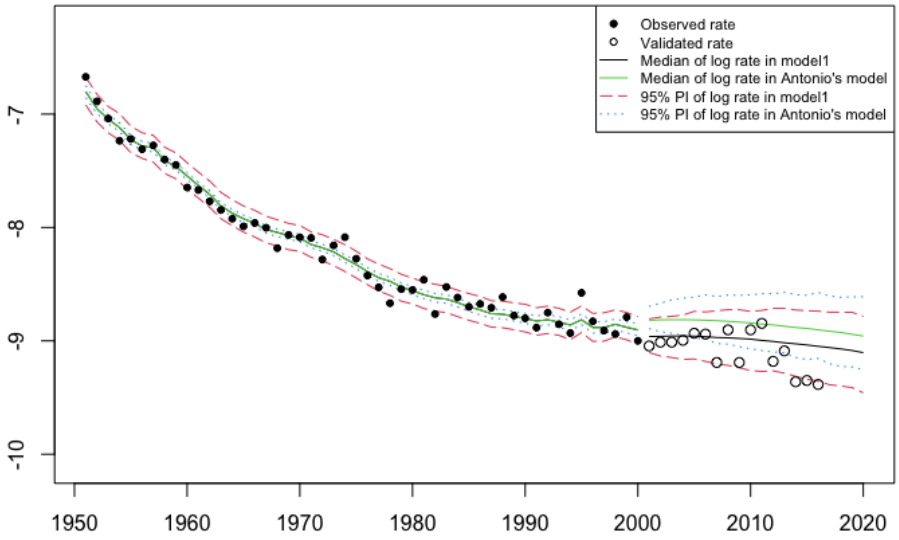
(c)

Figure 3: Plots of the posterior medians of κ_t, $\kappa_t^{(F)}$ and $\kappa_t^{(M)}$ with their 95% HDP intervals and the corresponding 20-years ahead projections.

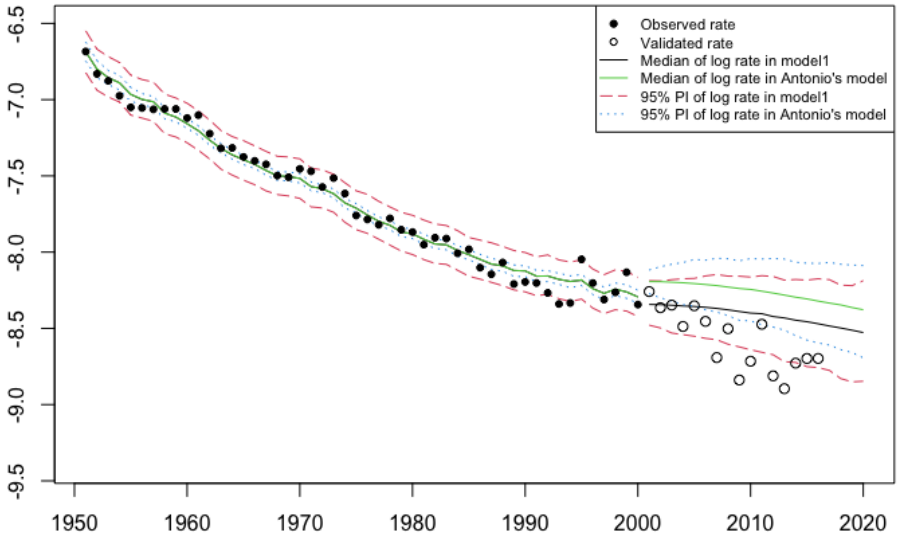
The way to empirically peek the posterior distributions of future time effects (that is, $\kappa_{2001}, \kappa_{2002}, \ldots, \kappa_{2020}$) above can also be used to assess the overall model fitting and its prediction ability. In particular, we compare mortality rates $\mu_{x,t}^{(i)}$ with the ones simulated from the fitted model, where $(\mu_{x,t}^{(i)})^{[j]} = \exp\Big[(\alpha_x^{(i)})^{[j]} + \beta_x^{[j]}\kappa_t^{[j]} + (\beta_x^{(i)})^{[j]}(\kappa_t^{(i)})^{[j]} + (\nu_{x,t}^{(i)})^{[j]}\Big]$.

Figures 4-6 present the medians and 95% HPD intervals of simulated log mortality rates at three selected ages 15, 55, and 70, respectively. In addition to the training time window (years 1950-2000), 20-years ahead projections are provided to assess the prediction ability of BPLNLCrm (marked as "model 1"). We also include the simulated results of the method by Antonio et al. (2015) (marked as "Antonio") as a comparison.

From Figures 4-6, we observe that the estimated curves (black and green) are close to each other within the training time window, but become bifurcating in the validation. Overall, the BPLNLCrm model provides better 20-years ahead projections because the true log rates are closer to the black predicted curves. We also notice that although the model by Antonio et al. (2015) tends to produce shorter credible intervals (blue dashed curves), those intervals fail to contain many of observed and validated log rates, implying the potential underestimation of variability inherited in model 2. In contrast, the wider credible intervals based on BPLNLCrm, which contain reasonable number of points, may properly present the variability of data by introducing additional overdispersion term, and are preferred.

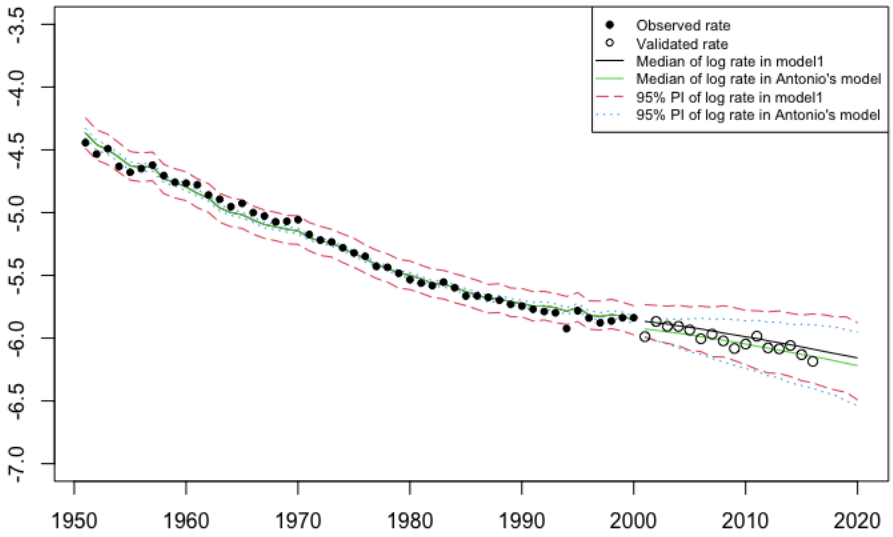


(a)

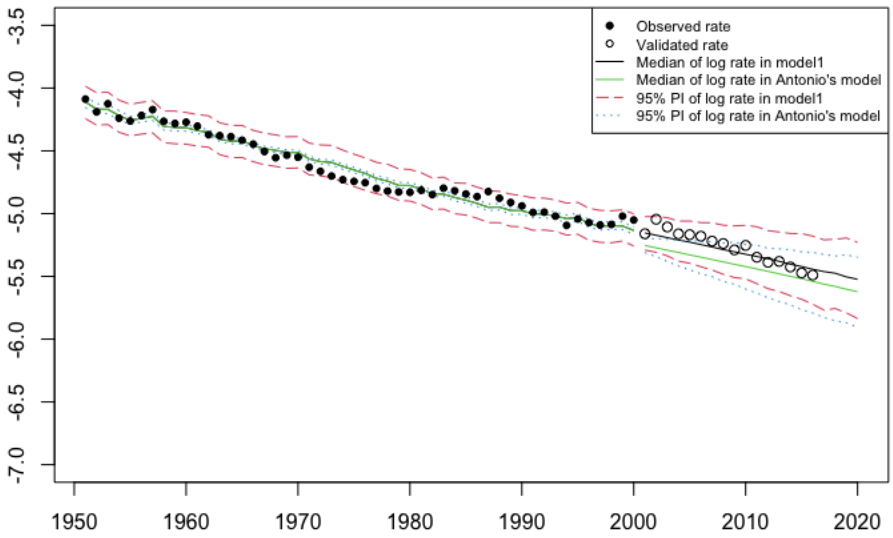


(b)

Figure 4: Plots of the observed and simulated log death rates at age 15 along with 20-years ahead projections and 95% HDP intervals for (a) female and (b) male.

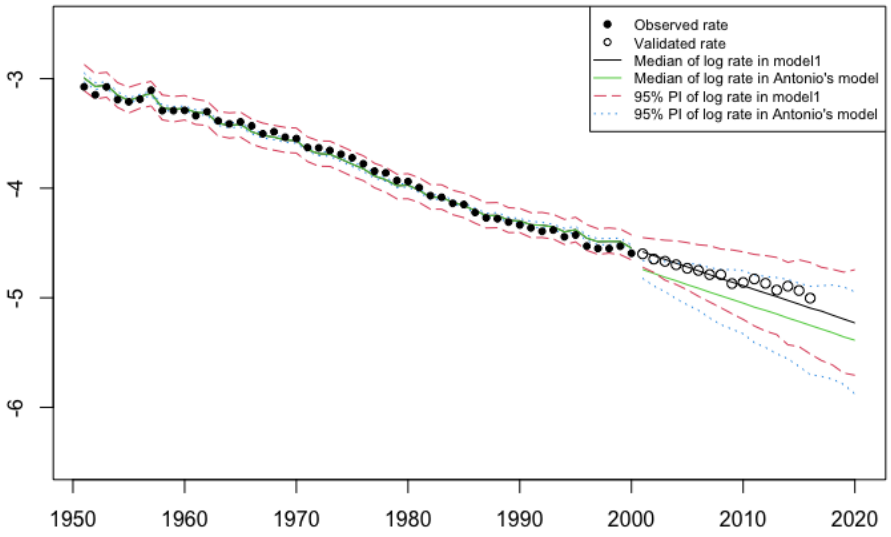


(a)

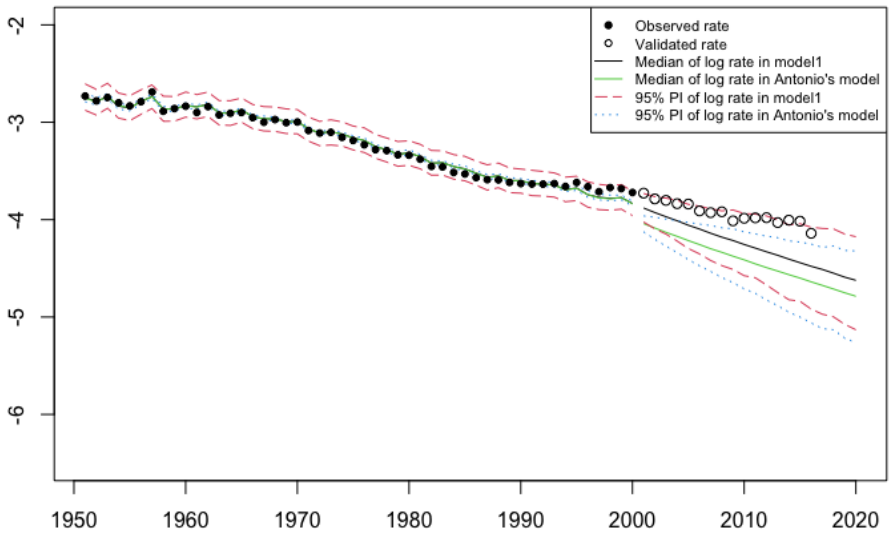


(b)

Figure 5: Plots of the observed and simulated log death rates at age 55 along with 20-years ahead projections and 95% HDP intervals for (a) female and (b) male.



(a)



(b)

Figure 6: Plots of the observed and simulated log death rates at age 70 along with 20-years ahead projections and 95% HDP intervals for (a) female and (b) male.

4.4 Northeast Region-specific Data

In this section, we apply BPLNLCrm to the mortality data with spatial information. Specifically, we consider six Northeast regions (Akita, Aomori, Fukushima, Iwate, Miyagi, and Yamagata) and two Kanto regions (Tokyo and Kanagawa) in the analysis. This choice reflects two different lifestyles and disconnected geospatial topologies, which may have an impact on the mortality. To evaluate the performance of our proposed model, which can leverage spatial information into the model selection and fitting, we compare with the results by BPLNLCrm with pre-specified $\tau_1 = \tau_2 = 0$. For convenience, we follow Section 4.3, and let "model 1" and "model 2" stand for BPLNLCrm with or without pre-specified τ_ls', respectively.

One challenge of the dirac spike setting in prediction-oriented models is that it usually requires a separate MCMC sample from the final chosen model to make predictions. This is because only a small portion of the original MCMC sample corresponds to the final choice and could be used to construct the posterior predictive distributions of future mortality rates. For example, as shown in Table 1, only 47% of MCMC sample is from the dominant dependence structure. This situation could become deteriorated when more populations are involved in the study. To improve the selection procedure and achieve a true one step analysis for both model selection and prediction, we use the hierarchical spatial model in (9) and facilitate the MCMC visits to the best model. Similar to Table 1, Tables 2 and 3 summarize the top four models and their relative frequencies of visits by an MCMC sample (10,000 iterations after 10,000 as burn-ins). As expected, the spatial information that contains geographic similarities and dissimilarities can inform the selection procedure and yield around 50% ($\approx (0.38 - 0.25)/0.25$) more visits to the best model. It is also worth mentioning that the final model indeed presents different dependence structures of time effects between Northeast and Kanto regions (only $w_2^{(7)} = w_2^{(8)} = 0$).

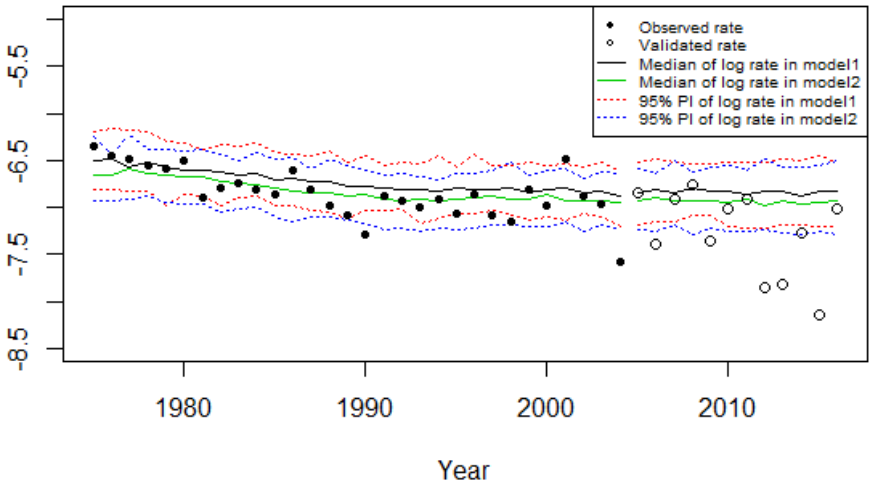
Table 2: Top four dependence structures of $\boldsymbol{\kappa}^{(i)}$ when τ_1 and τ_2 are set as 0.

Variables	Proportion
$\boldsymbol{w_1}; w_2^{(1)}, w_2^{(2)}, w_2^{(3)}, w_2^{(4)}, w_2^{(5)}, w_2^{(6)}$	0.25
$\boldsymbol{w_1}; w_2^{(1)}, w_2^{(2)}, w_2^{(3)}, w_2^{(4)}, w_2^{(5)}, w_2^{(6)}, w_2^{(7)}$	0.12
$\boldsymbol{w_1}; w_2^{(1)}, w_2^{(2)}, w_2^{(3)}, w_2^{(4)}, w_2^{(5)}, w_2^{(6)}, w_2^{(8)}$	0.10
$\boldsymbol{w_1}; w_2^{(1)}, w_2^{(2)}, w_2^{(3)}, w_2^{(4)}, w_2^{(5)}, w_2^{(6)}, w_2^{(7)}, w_2^{(8)}$	0.05

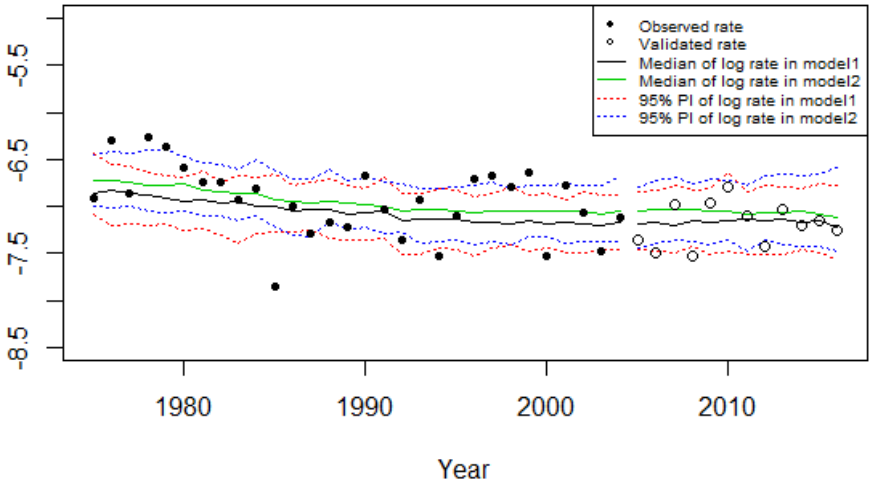
Table 3: Top four dependence structures of $\boldsymbol{\kappa}^{(i)}$.

Variables	Proportion
$\boldsymbol{w_1}; w_2^{(1)}, w_2^{(2)}, w_2^{(3)}, w_2^{(4)}, w_2^{(5)}, w_2^{(6)}$	0.38
$\boldsymbol{w_1}; w_2^{(1)}, w_2^{(2)}, w_2^{(3)}, w_2^{(4)}, w_2^{(5)}, w_2^{(6)}, w_2^{(7)}$	0.07
$\boldsymbol{w_1}; w_2^{(1)}, w_2^{(2)}, w_2^{(3)}, w_2^{(4)}, w_2^{(5)}, w_2^{(6)}, w_2^{(8)}$	0.05
$\boldsymbol{w_1}; w_2^{(1)}, w_2^{(2)}, w_2^{(3)}, w_2^{(4)}, w_2^{(5)}, w_2^{(6)}, w_2^{(7)}, w_2^{(8)}$	0.02

Following the chosen model and 38% of an MCMC sample (25% for "model 1"), we further evaluate the model fitting and prediction ability of BPLNLCrm by comparing its mortality projections with the observed and validated rates (training data set: $1975 \sim 2004$ & validation data set: $2005 \sim 2016$). Figures 7-9 present the simulated log mortality rates along with 95% HPD intervals for the regions Akita and Aomori at the three selected ages 35, 55, and 75, respectively. For other six regions (Fukushima, Iwate, Miyagi, Yamagata, Tokyo and Kanagawa), the reader may refer to Figures 10-18 in the Appendix. It is clear that two models produce quite different mortality projections (black and green curves) because the observed differences are in log scale. Besides, we observe that 95% HPD intervals (blue dot curves) by "model 2" have overall better coverage of true mortality rates than the red dot ones by "model 1", suggesting that the additional hierarchical structure may yield better projections.

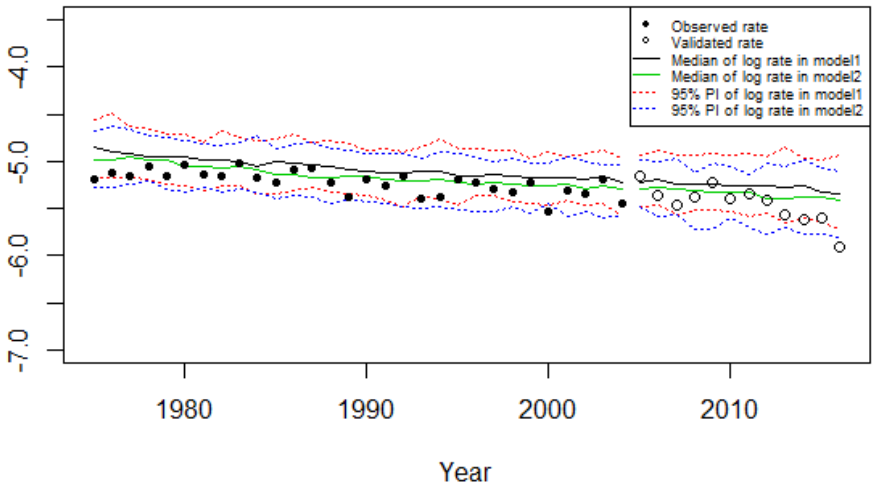


(a) Akita

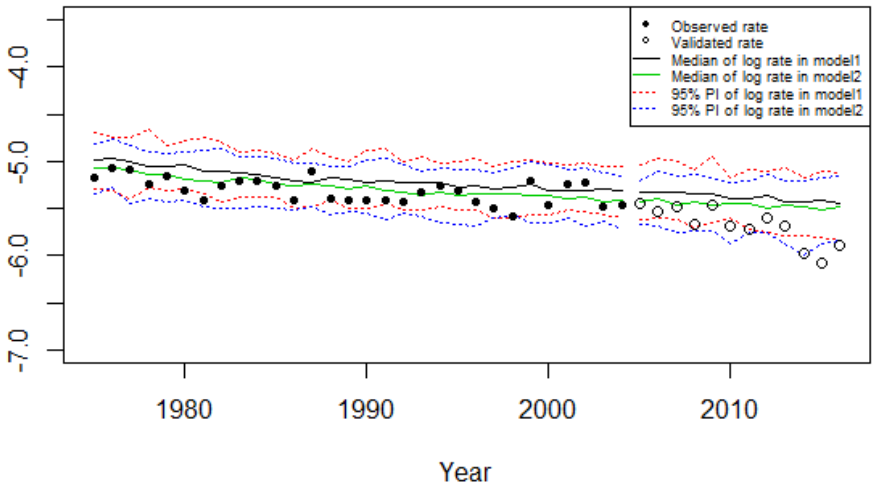


(b) Aomori

Figure 7: Plots of the observed and simulated log death rates at age 35 along with 12-year ahead projections and 95% HDP intervals for (a) Akita and (b) Aomori.

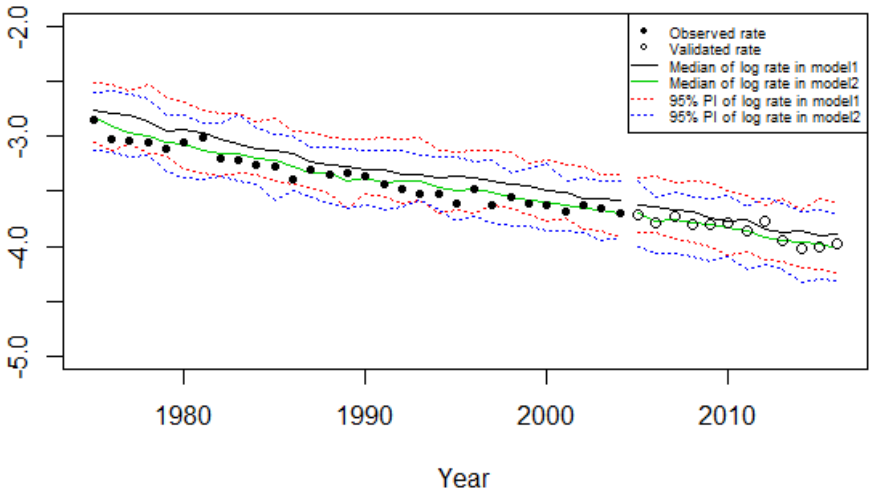


(a) Akita

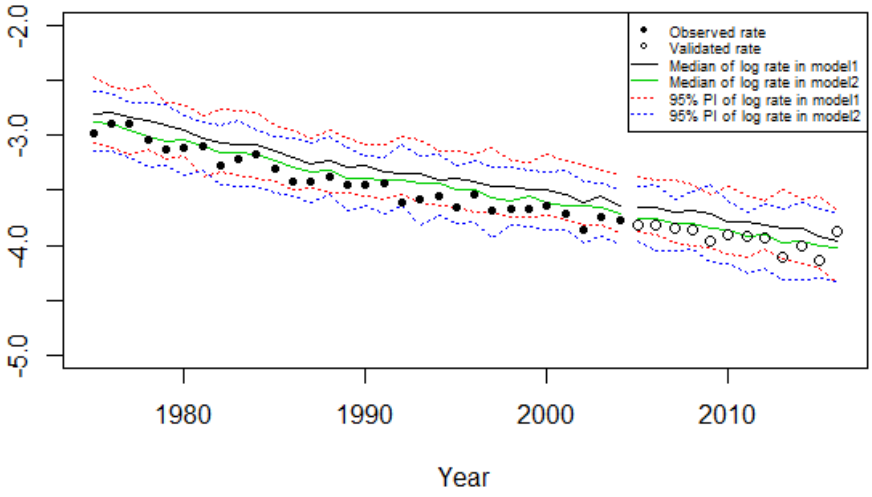


(b) Aomori

Figure 8: Plots of the observed and simulated log death rates at age 55 along with 12-year ahead projections and 95% HDP intervals for (a) Akita and (b) Aomori.



(a) Akita



(b) Aomori

Figure 9: Plots of the observed and simulated log death rates at age 75 along with 12-year ahead projections and 95% HDP intervals for (a) Akita and (b) Aomori.

5 DISCUSSION

This paper presents a Bayesian approach to estimate and predict mortality for multiple populations under Poisson log-normal assumption. It combines the model by Antonio et al. (2015) with PLNLC, granting the new model to properly reflect the variations of mortality in a multi-population problem. Additionally, by introducing a novel dirac spike function, the new model can leverage additional spatial information to improve one step analysis; i.e., simultaneously model selection and estimation of population-specific time effects. As a result, with this affordable computation even when n is big, it can avoid unnecessary assumptions on dependence structures of $\kappa_t^{(i)}$. We note that the current model considers the adjacency matrix as the only resource of spatial information. When more region-specific factors such as cultures, dietary habits, levels of economic development, and regional climate changes are available, BPLNLCrm could include them as covariates in the mean function of λ_l. We also note that the similar modeling strategies could be applied to a negative binomial setting, which is a different over-dispersed extension of the Poisson and can incorporate a random effect into the model. The comparisons and generalizations of these two versions of LC model require further investigations.

Conflict of Interest

The authors confirm that this article content has no conflict of interest.

Acknowledgement

Part of Dr. Leping Liu's research has been supported by National Science Foundation of China (NSFC) under Grant 71771163.

References

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, 88(422):669–679.

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

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

Cairns, A. J., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *Journal of Risk and Insurance*, 73(4):687–718.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., and Khalaf-Allah, M. (2011). Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin: The Journal of the IAA*, 41(1):29–59.

Cressie, N. and Chan, N. H. (1989). Spatial modeling of regional variables. *Journal of the American Statistical Association*, 84(406):393–401.

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

Gelman, A. et al. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian analysis*, 1(3):515–534.

George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.

Girosi, F. and King, G. (2003). *Demographic forecasting*. PhD thesis, Harvard University.

HMD (2021). Human Mortality Database. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). http://www.mortality.org/.

Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*. Springer Publishing Company, Incorporated, 1st edition.

Ishwaran, H., Rao, J. S., et al. (2005). Spike and slab variable selection: frequentist and Bayesian strategies. *The Annals of Statistics*, 33(2):730–773.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American statistical association*, 87(419):659–671.

LeSage, J. P. (2000). Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis*, 32(1):19–35.

Li, J. S.-H. and Hardy, M. R. (2011). Measuring basis risk in longevity hedges. *North American Actuarial Journal*, 15(2):177–200.

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

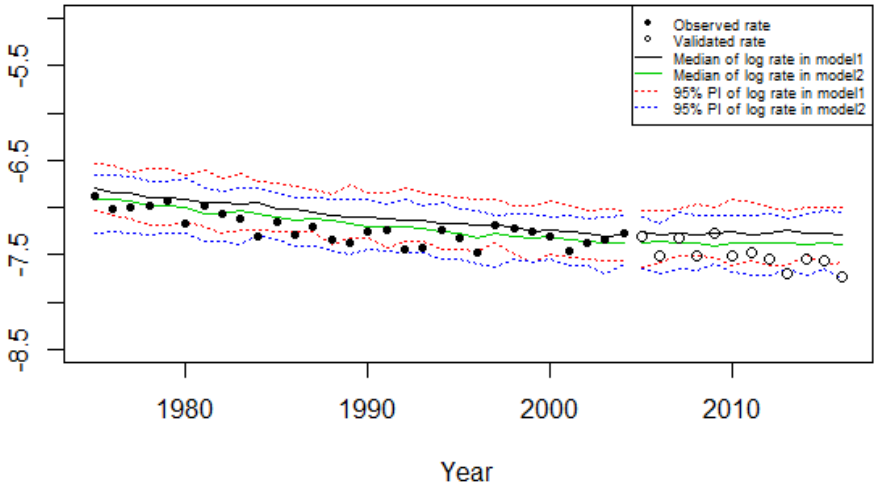
Malsiner-Walli, G. and Wagner, H. (2018). Comparing spike and slab priors for Bayesian variable selection. *arXiv preprint arXiv:1812.07259*.

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

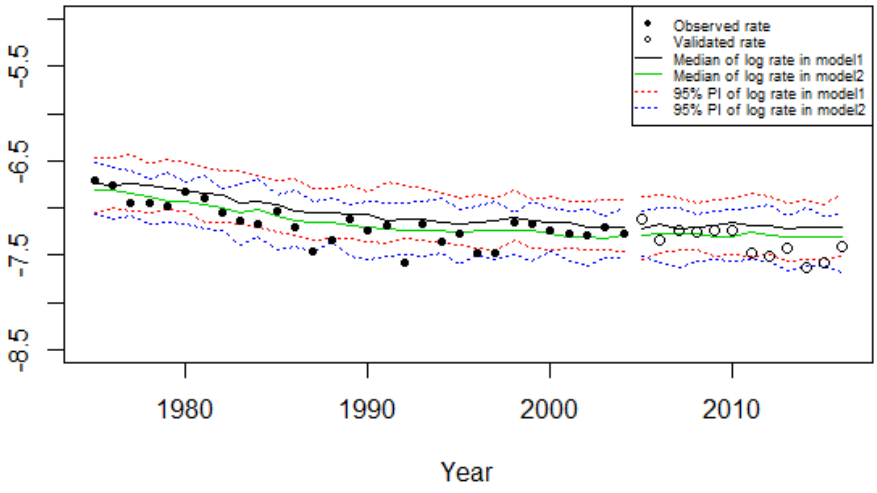
Renshaw, A. E. and Haberman, S. (2003). Lee-carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics*, 33(2):255–272.

Tuljapurkar, S., Li, N., and Boe, C. (2000). A universal pattern of mortality decline in the G7 countries. *Nature*, 405(6788):789.

Wong, J. S., Forster, J. J., and Smith, P. W. (2018). Bayesian mortality forecasting with overdispersion. *Insurance: Mathematics and Economics*, 83:206–221.

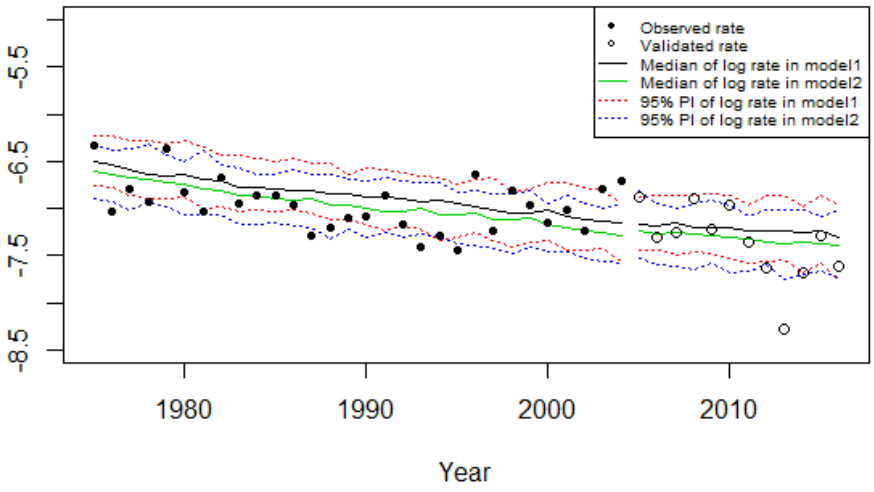


(a) Fukushima

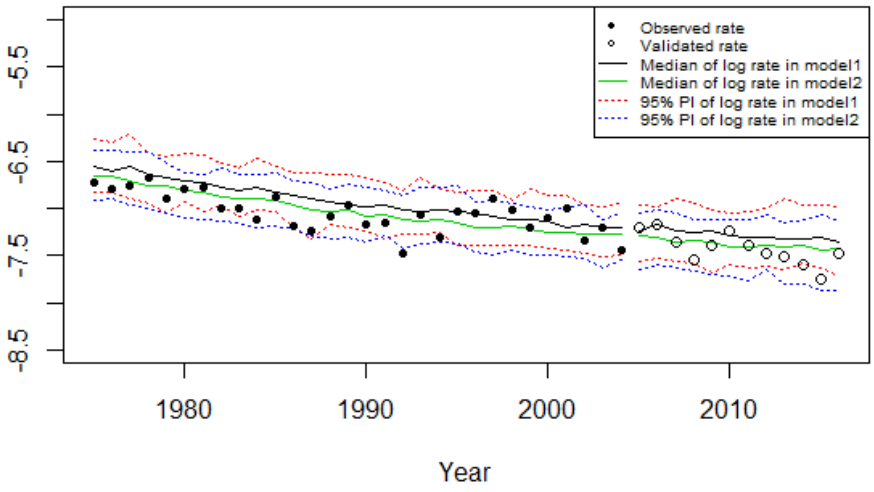


(b) Iwate

Figure 10: Plots of the observed and simulated log death rates at age 35 along with 12-year ahead projections and 95% HDP intervals for (a) Fukushima and (b) Iwate.

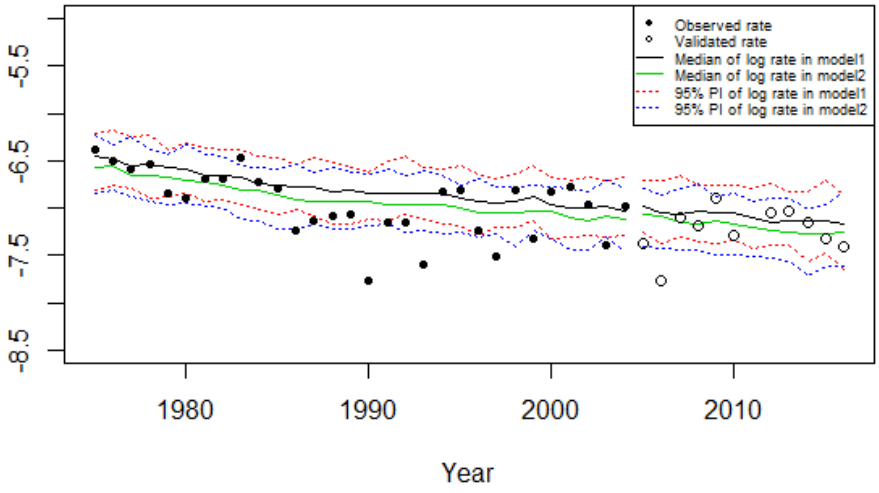


(a) Miyagi

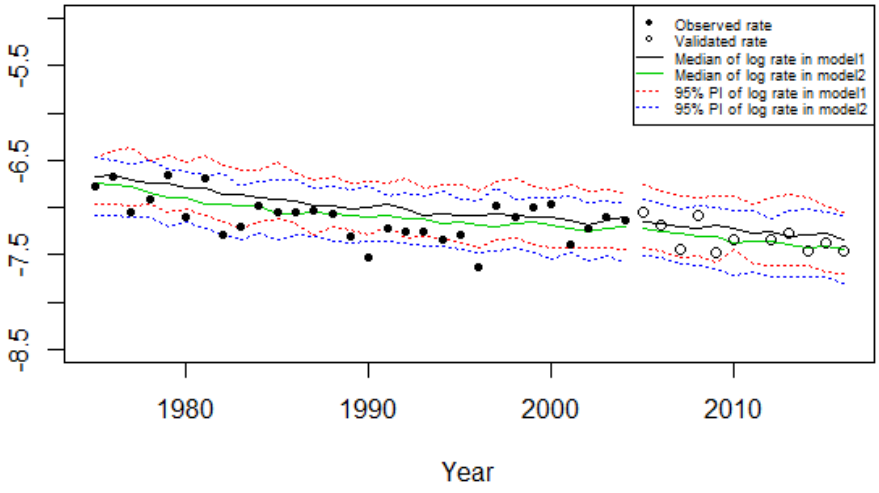


(b) Yamagata

Figure 11: Plots of the observed and simulated log death rates at age 35 along with 12-year ahead projections and 95% HDP intervals for (a) Miyagi and (b) Yamagata.

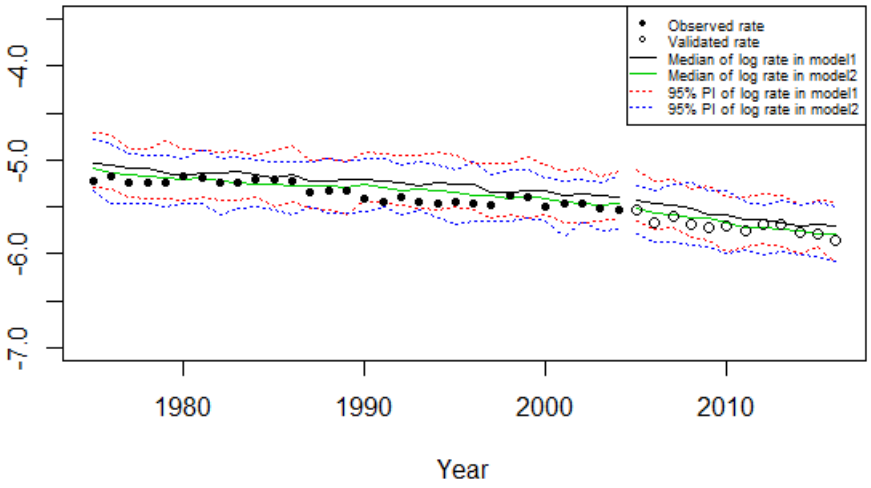


(a) Tokyo

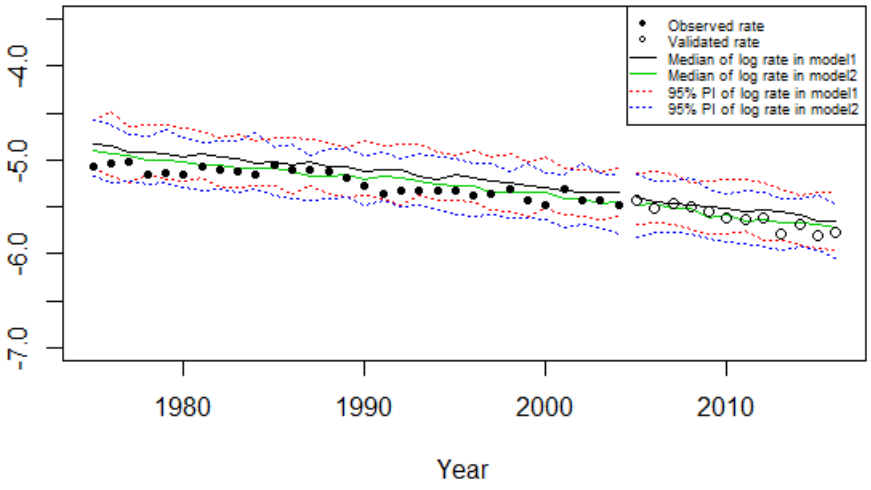


(b) Kanagawa

Figure 12: Plots of the observed and simulated log death rates at age 35 along with 12-year ahead projections and 95% HDP intervals for (a) Tokyo and (b) Kanagawa.

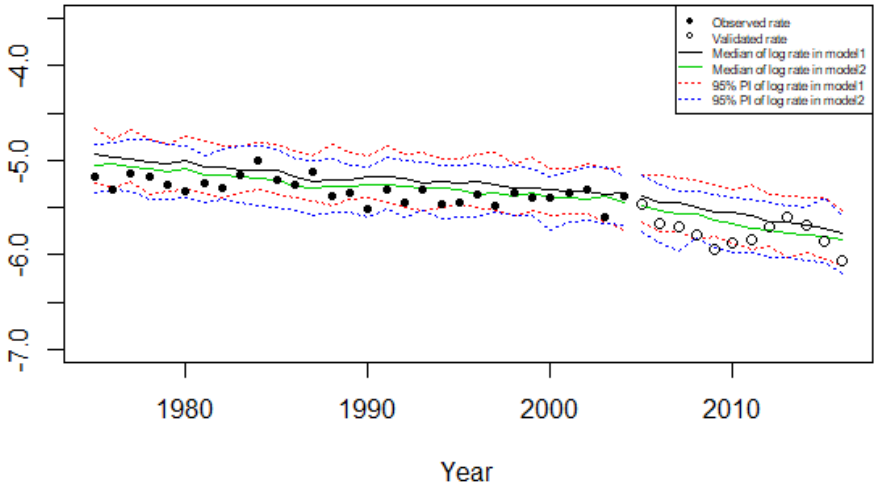


(a) Fukushima

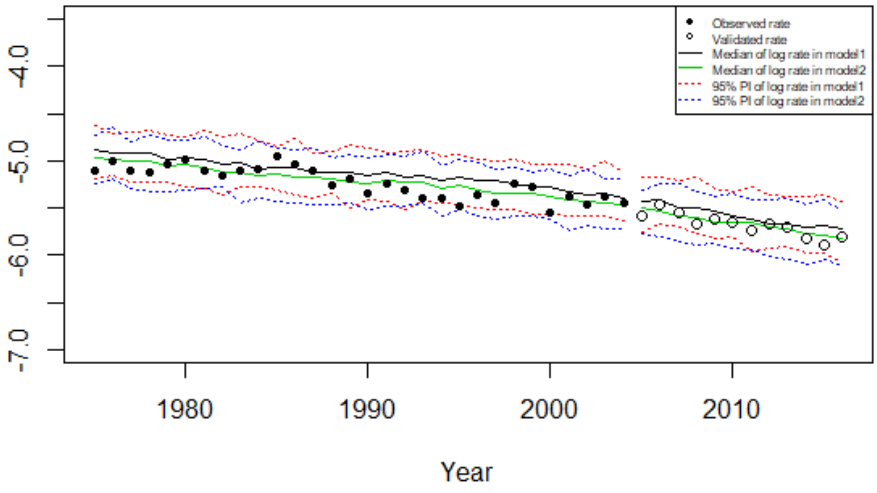


(b) Iwate

Figure 13: Plots of the observed and simulated log death rates at age 55 along with 12-year ahead projections and 95% HDP intervals for (a) Fukushima and (b) Iwate.

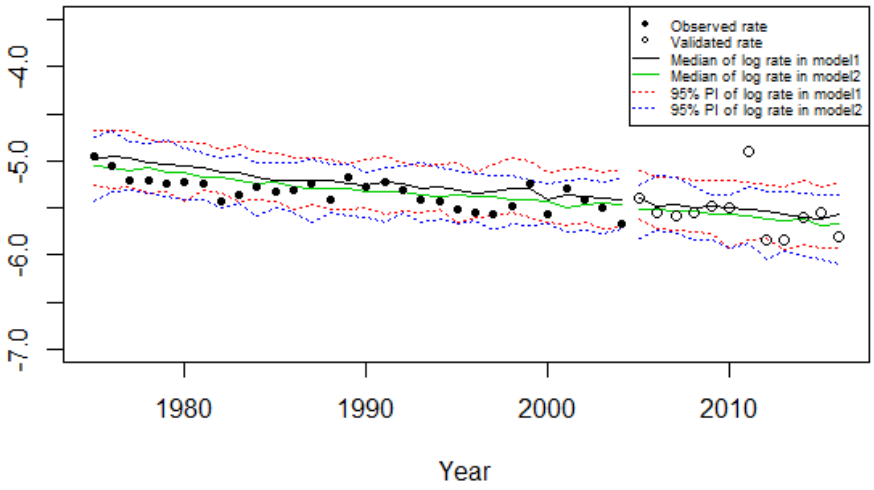


(a) Miyagi

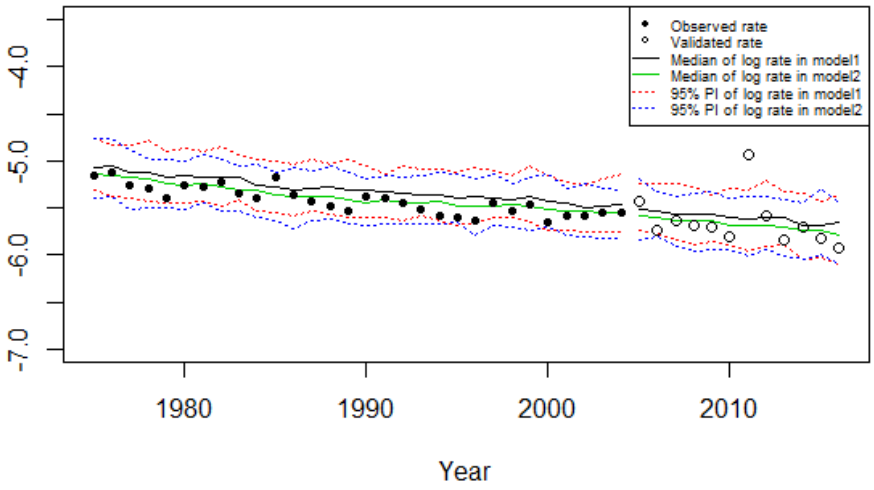


(b) Yamagata

Figure 14: Plots of the observed and simulated log death rates at age 55 along with 12-year ahead projections and 95% HDP intervals for (a) Miyagi and (b) Yamagata.

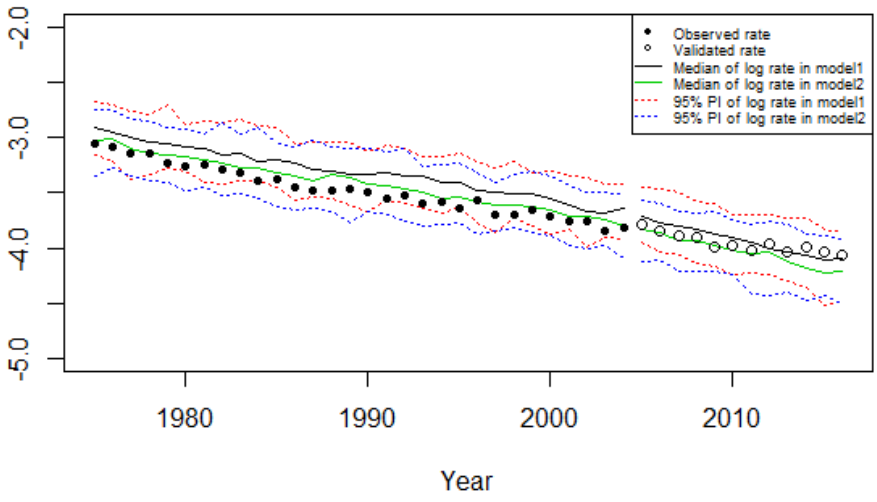


(a) Tokyo

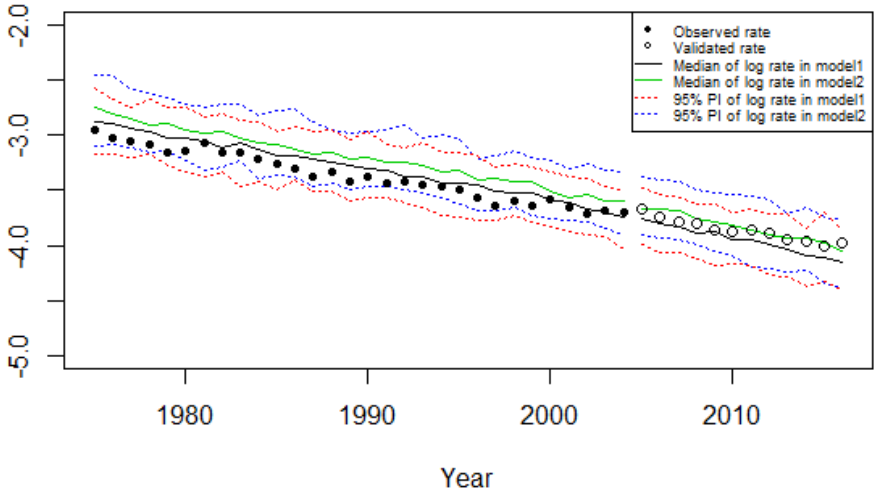


(b) Kanagawa

Figure 15: Plots of the observed and simulated log death rates at age 55 along with 12-year ahead projections and 95% HDP intervals for (a) Tokyo and (b) Kanagawa.

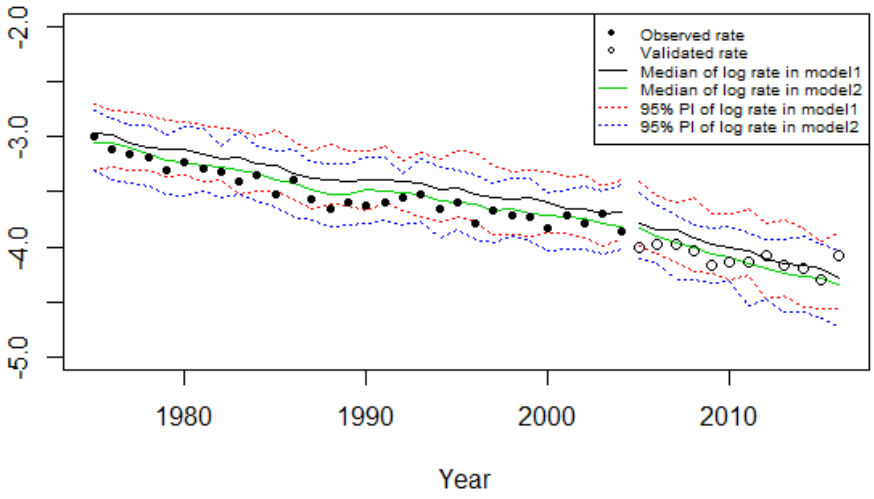


(a) Fukushima

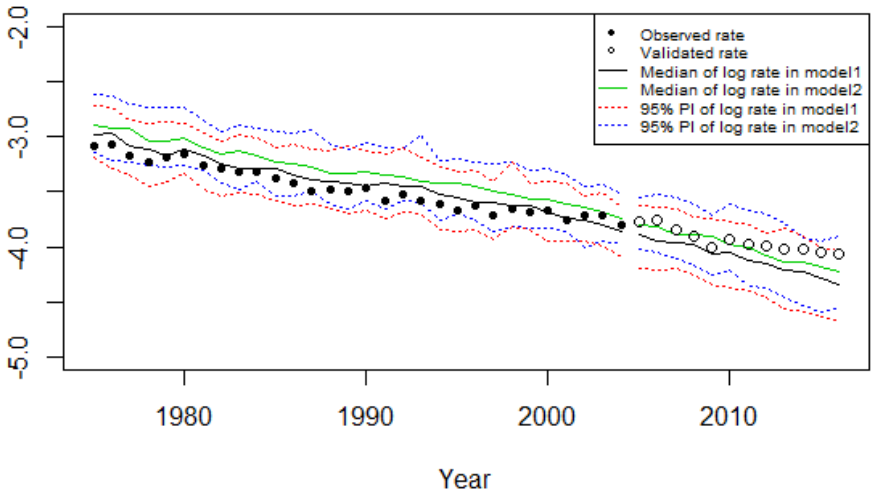


(b) Iwate

Figure 16: Plots of the observed and simulated log death rates at age 75 along with 12-year ahead projections and 95% HDP intervals for (a) Fukushima and (b) Iwate.

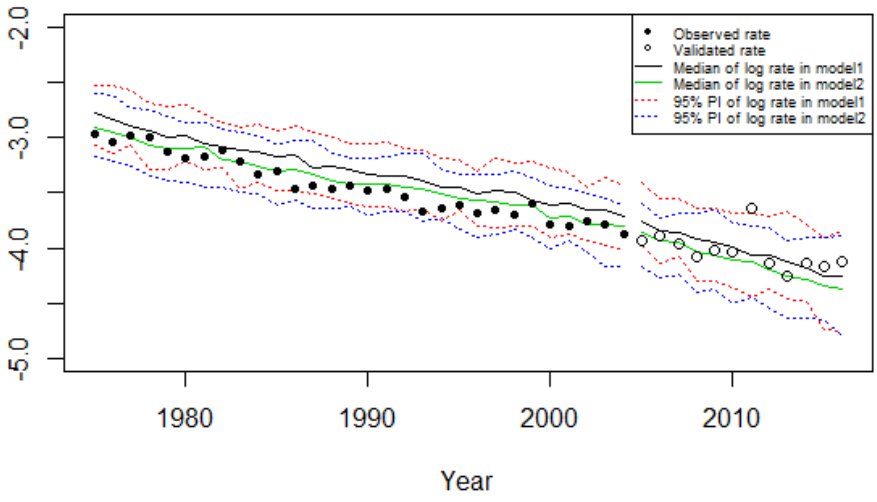


(a) Miyagi

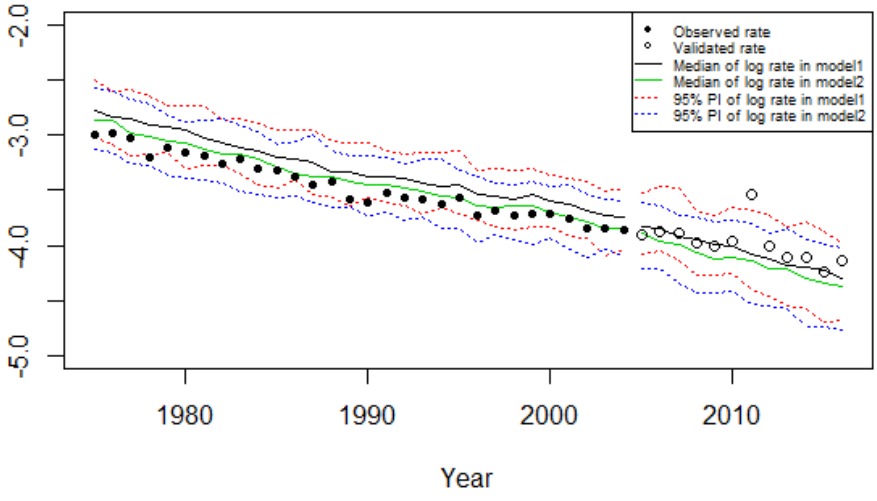


(b) Yamagata

Figure 17: Plots of the observed and simulated log death rates at age 75 along with 12-year ahead projections and 95% HDP intervals for (a) Miyagi and (b) Yamagata.



(a) Tokyo



(b) Kanagawa

Figure 18: Plots of the observed and simulated log death rates at age 75 along with 12-year ahead projections and 95% HDP intervals for (a) Tokyo and (b) Kanagawa.