8-2021

Bayesian Poisson Log-normal Model with Regularized Time Structure and Spatial Framework for Mortality Projection of Multi-population

Zhen Liu
Clemson University, zliu2@clemson.edu

Bayesian Poisson Log-normal Model with Regularized Time Structure and Spatial Framework for Mortality Projection of Multi-population

---

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Mathematical Sciences

---

by
Zhen Liu
August 2021

Accepted by:
Dr. Xiaoqian Sun, Committee Chair
Dr. Yu-Bo Wang, Committee Co-Chair
Dr. Andrew Brown
Dr. Whitney Huang

Abstract

The improvement of mortality projection is a pivotal topic in the diverse branches related to insurance, demography, and public health. The dissertation consists of two distinct but related projects about mortality projection. We consider hierarchical bilinear modeling, variable selection methods and spatio-temporal structure under Bayesian frameworks in the estimation of mortality rate for multiple populations.

First, we propose a Bayesian model to estimate and predict mortality rates for multi-population, motivated by the thread of Lee-Carter related models. This new model features 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 Dirac spike function, simultaneous model selection and estimation for population-specific time effects can be achieved without much extra computation cost. Further, this work discusses a Bayesian probit model to provide a dependence structure involving spatial information together with the Dirac spike function. We use the Japanese mortality data from the Human Mortality Database to illustrate the desirable properties of our model.

Second, this work develops Bayesian mortality projection models for multiple populations by considering the stochastic structure and the effect of spatial autocorrelation within the observations. We explain high levels of overdispersion according to adjacent regions based on the conditional autoregressive model. In an empirical study, this dissertation compares different hierarchical projection models for the analysis of geographical diversity in mortality among multiple Japanese counties in consecutive years, grouped by age. We implement Markov chain Monte Carlo (MCMC) computation to conduct parameter estimation. Results have demonstrated the flexibility and predictive performance of our proposed model.

Dedication

This paper is lovingly dedicated to my beloved parents and my lovely wife Hui, for the love and support they have given me and for standing by me during all the tough and exciting moments.

Acknowledgments

I would like to express my deepest gratitude to my advisors Dr. Xiaoqian Sun and Dr. Yu-Bo Wang for their insightful guidance and persistent help in my research journey into spatio-temporal Bayesian modeling. My sincere gratitude also goes out to my committee members, Dr. Andrew Brown and Dr. Whitney Huang for their valuable and helpful advice and enlightening discussions in improving the quality of my research projects. Thanks to Dr. William Bridges for his advise during my early period of graduate study.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Mortality projection has become a topic of great importance in demographics, given its strong association with various social areas, for example, 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 mortality after 1950s (Tuljapurkar et al., 2000), a thorough and well established policy relies on an accurate prediction of mortality trajectory.

Over last several decades, stochastic models have been widely applied to mortality projection because the resulting forecasts along with inferences 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 (RWD) 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, takes advantage of 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 considered 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), Hyndman and Ullah (2007), Pedroza (2006), Delwarde et al. (2007) and Plat (2009).

With the development of people's understanding on these topics, research focused on the mortality information has evolved from single population into multiple populations. Due to strong mutual impacts, it seems improper to forecast mortality rates for some adjacent geographic regions in isolation from one another. In addition, this strong correlation also allows accurate estimates of potential missing data using those observations from neighboring areas. Hence, geographical dependent modeling and stochastic modeling of mortality forecasting for age groups have drawn more attention in decades.

Inspired by spatial statistics for discrete variation (Besag, 1975), Besag et al. (1991) suggested a Bayesian spatial model, which combined an autoregressive component for adjacent spatial effects with independent error terms to explain the unobserved variance using the spatial structure of epidemic data. This model incorporates individual and area effects, and provides additional insights in terms of similarity based on the neighborhood structure of areas in a specific period. However, the temporal effect to the spatial heterogeneity was not discussed to establish the mapping model of risk over time. By further taking the interaction of spatial and temporal factors into consideration, Knorr-Held (2000) introduced a Bayesian spatio-temporal model in Ohio lung cancer prediction to elucidate the variation involved in the geographical diversity from 1968 to 1988. Using a Bayesian hierarchical modeling, expected similarities can be accommodated via a model with positive spatial correlations between the regional parameters. The inclusion of temporal varying factors overcame the drawbacks of the single spatial model for panel data.

In order to tackle the limitations of Bayesian modeling of spatial effects alone without age information, Arató et al. (2006) developed a Bayesian spatial age-dependent mortality model for the research of deaths in 150 geographic regions of Hungary. They proposed a binomial distribution as the underlying model for respective age and region groups in 1997. This framework, however, only focused on the mortality estimation in a single year rather than the projection for following periods. In a similar study for estimating small-area mortality rates at provincial level, Greco and Scalone (2013) established an extended Poisson linear model for mortality risk of every province in Italy from 1992 to 2009, whereas this model did not borrow strength from age groups and ignored the time series structure of forecasting. For cases in more countries, see Waller and Gotway (2004),

Congdon (2007), Rau and Schmertmann (2017), Bergeron-Boucher et al. (2018) and Khana et al. (2018).

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), Antonio et al. (2015) and Liu et al. (2020), 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 the second chapter, we revisit the works of Wong et al. (2018) and Antonio et al. (2015), and develop a new multi-population model that can entertain the potential overdispersion in data. We also relax the model assumption on each population-specific time component by considering an autoregressive model of order one (AR(1)) with a drift, where the drift parameter and the slope associated with time are hierachically regulated by a Dirac spike and slab prior (George and McCulloch, 1993; Kuo and Mallick, 1998; Ishwaran, Rao, et al., 2005; Malsiner-Walli and Wagner, 2018). Considering spatial interaction effects among adjacent regions, an alternative Bayesian approach to probit models allows more information of spatial dependency within model selection for population-specific time parameters (Albert and Chib, 1993; LeSage, 2000; Smith and LeSage, 2004). As a result, model selection between AR(1) with and without a drift on population-specific time components is conducted simultaneously with estimation and prediction.

Most of previous mortality research focused on the modeling of age-time interaction alone or separate spatial effect in a single period. In the third chapter, I intend to take advantage of age-specific factors and stochastic structures in the LC model for multiple nationwide adjacent regions. The proposed model fills the gap of integrated spatial estimation and prediction of age-stratified mortality rates under Bayesian paradigm in a long run. Following this line, an essential feature of our model is to link spatial effects to time effects for each age division within multiple populations. This model features with two settings of spatial factor: first, we can consider it as a measure of heterogeneity in each population; second, as an overdispersion term, the spatial component is extended to a multivariate dependent structure involving more information among areas, such as ethnicity and pandemics. Under the assumption of univariate distributions (Wong et al., 2018), overdispersions were set to be independent of each other. Comparably, our overdispersion terms based on dependent covariance structure consider more interactions among adjacent regions.

Chapter 2

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

2.1 Introduction

The study of variability in mortality rates is a topic of great interest in wide social fields, such as demography, epidemiology and insurance. A thorough and accurate mortality projection will provide essential and important information for policy making in public health, pension and many other systems. Over the past several decades, researchers have applied various types of models to achieve a better projection.

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

$$\log m_{x,t} = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}, \tag{2.1}$$

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 (2.1) 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

$$D_{x,t} | \mu_{x,t} \sim \text{Poisson}(E_{x,t}\mu_{x,t}) \quad \text{with} \quad \log \mu_{x,t} = \alpha_x + \beta_x \kappa_t, \tag{2.2}$$

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 (2.1), 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 (2.2), 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, (2.1) or (2.2)) 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

$$D_{x,t}|\mu_{x,t} \sim \text{Poisson}(E_{x,t}\mu_{x,t}) \quad \text{with} \quad \log \mu_{x,t} = \alpha_x + \beta_x\kappa_t + \nu_{x,t}. \tag{2.3}$$

With this additional diffusion $\nu_{x,t}$, the PLNLC model relaxes the equality constraint on mean and variance as follows

$$\text{E}[D_{x,t}] = \text{E}[\text{E}(D_{x,t}|\nu_{x,t})] = E_{x,t} \exp(\alpha_x + \beta_x\kappa_t + \frac{1}{2}\sigma^2), \tag{2.4}$$

$$\text{Var}[D_{x,t}] = \text{E}[\text{Var}(D_{x,t}|\nu_{x,t})] + \text{Var}[\text{E}(D_{x,t}|\nu_{x,t})] \tag{2.5}$$

$$= \text{E}[D_{x,t}] \times \{1 + \text{E}[D_{x,t}] \times [\exp(\sigma^2) - 1]\} \geq \text{E}[D_{x,t}], \tag{2.6}$$

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 (2.2) to the following Poisson log-bilinear model for a n-population data set

$$D^{(i)}_{x,t}|\mu^{(i)}_{x,t} \sim \text{Poisson}(E^{(i)}_{x,t}\mu^{(i)}_{x,t}) \quad \text{with} \quad \log \mu^{(i)}_{x,t} = \alpha^{(i)}_x + \beta_x\kappa_t + \beta^{(i)}_x\kappa^{(i)}_t, \tag{2.7}$$

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 i^{th} population-specific term so $\alpha^{(i)}_x$ and $\beta^{(i)}_x\kappa^{(i)}_t$, for $i = 1, 2, , \ldots, n$, are a population-specific intercept and effect, respectively. To identify (2.7), additional constraints on population-specific age and time effects are required: $\left\|\beta^{(i)}_x\right\|_2 = 1$ and $\sum_{t\in\Theta_{\text{time}}} \kappa^{(i)}_t = 0$, where $\|.\|_2$ represents the L_2 norm of a vector. Through jointly investigating related populations, (2.7) tends to be more efficient than the separate modeling using PLC on each population.

In this chapter, 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 (2.7), but also simultaneously conducts

model selection of time structures of $\kappa_t^{(i)}$ via a Dirac spike and slab prior and spatial probit model, respectively. As a result, it can serve for more varieties of mortality data. We introduce our model formally in Section 2.2.

The remainder of this chapter is organized as follows. In Section 2.2, we introduce the proposed model along with the prior settings and detailed steps of an Markov chain Monte Carlo (MCMC) sampling. In Section 2.3, the proposed method is applied to Japan gender-specific mortality data between 1951 and 2016 from the Human Mortality Database (HMD). To evaluate its performance, the results based on the model by Antonio et al. (2015) are also included as a comparison. Finally, we conclude with a discussion in Section 2.4.

2.2 The Proposed Model

2.2.1 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 model for n-population (BPLNLCrm) as follows

$$
\begin{aligned}
&D_{x,t}^{(i)} | \mu_{x,t}^{(i)} \sim \text{Poisson}(E_{x,t}^{(i)} \mu_{x,t}^{(i)}) \quad \text{with} \quad \log \mu_{x,t}^{(i)} = \alpha_x^{(i)} + \beta_x \kappa_t + \beta_x^{(i)} \kappa_t^{(i)} + \nu_{x,t}^{(i)}, \\
&\kappa_t = \varphi_1 + \varphi_2 t + \rho[\kappa_{t-1} - \varphi_1 - \varphi_2(t-1)] + \epsilon_t, \\
&\kappa_t^{(i)} = \varphi_1^{(i)} + \varphi_2^{(i)} t + \rho^{(i)}[\kappa_{t-1}^{(i)} - \varphi_1^{(i)} - \varphi_2^{(i)}(t-1)] + \epsilon_t^{(i)},
\end{aligned}
\tag{2.8}
$$

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 $t \in \Theta_{\text{time}}$. Note that the first line can be viewed as a generalization of (2.3) to a multi-population problem, and 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 $\boldsymbol{\varphi} = (\varphi_1, \varphi_2)'$, $\boldsymbol{\varphi}^{(i)} = (\varphi_1^{(i)}, \varphi_2^{(i)})'$, $\boldsymbol{Q} = \boldsymbol{U'U}$, and $\boldsymbol{Q}^{(i)} = \boldsymbol{U}^{(i)'}\boldsymbol{U}^{(i)}$. Then, $n+1$ dependence structures of time effects can be written as

$$\boldsymbol{\kappa} \sim N(\boldsymbol{W\varphi}, \sigma_\kappa^2 \boldsymbol{Q}^{-1}), \tag{2.9}$$

and

$$\boldsymbol{\kappa}^{(i)} \sim N(\boldsymbol{W\varphi}^{(i)}, \sigma^2_{\kappa^{(i)}} (\boldsymbol{Q}^{(i)})^{-1}), \tag{2.10}$$

where $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \ldots, \kappa_N)'$ and $\boldsymbol{\kappa}^{(i)} = (\kappa_1^{(i)}, \kappa_2^{(i)}, \ldots, \kappa_N^{(i)})'$.

We further assign a Dirac spike function to regularize $\varphi_1^{(i)}$ and $\varphi_2^{(i)}$ as follows

$$\varphi_l^{(i)} \sim w_l^{(i)} N(0, \sigma^2_{\kappa^{(i)}}) + (1 - w_l^{(i)}) \delta_l^{(i)} \tag{2.11}$$

where $w_l^{(i)}$ is a binary variable equal to 1 when a more complicated dependence structure of $\boldsymbol{\kappa}^{(i)}$ is needed for fitting the data set, and vice versa, $\delta_l^{(i)}$ is a point mass at zero, and $l = 1, 2$. Note that when all $w_1^{(i)}$s' and $w_2^{(i)}$s' are zero so that $\boldsymbol{\varphi}^{(1)} = \boldsymbol{\varphi}^{(2)} = \cdots = \boldsymbol{\varphi}^{(n)} = (0,0)'$, the dependence structures are the same as those in Antonio et al. (2015). Although Antonio et al. (2015) justified this special setting in some way, we prefer to consider a more general structure and let the data speak out the truth of $\varphi_1^{(i)}$s' and $\varphi_2^{(i)}$s'. With (2.11), our model can explore the model space of 2^{2n} possible dependence structures of $\boldsymbol{\kappa}^{(i)}$ in a single analysis simultaneously obtaining parameter estimation. When n is large, it can ease computation in model selection compared to using the criteria-based approaches, such as the marginal likelihood criterion and the Akaike information criterion.

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 (2.8) 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 approximate the 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.

2.2.2 Prior Specifications

2.2.2.1 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 in Czado et al. (2005) and Antonio et al. (2015) and propose

$$e_x^{(i)} \sim \text{Gamma}(a_x^{(i)}, b_x^{(i)}), \tag{2.12}$$

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 consistent with the previous constraints for $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^{(i)}$.

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

where $\boldsymbol{J}_M$ is a $M \times 1$ vector with all elements equal to 1, and $\boldsymbol{I}_M$ is an identity matrix of size M. The hyperparameters σ^2_β and $\sigma^2_{\beta^{(i)}}$ are assumed to follow the inverse Gamma distributions, that is,

$$\sigma^2_\beta \sim \text{InvGamma}(a_\beta, b_\beta),$$
$$\sigma^2_{\beta^{(i)}} \sim \text{InvGamma}(a^{(i)}_\beta, b^{(i)}_\beta),$$

where a_β, b_β, $a^{(i)}_\beta$, and $b^{(i)}_\beta$ are pre-specified constants such that

$$\pi(\sigma^2_\beta) = \frac{b^{a_\beta}_\beta}{\Gamma(a_\beta)}(\sigma^2_\beta)^{-a_\beta - 1}\exp(-b_\beta/\sigma^2_\beta)$$

and

$$\pi(\sigma^2_{\beta^{(i)}}) = \frac{(b^{(i)}_\beta)^{a^{(i)}_\beta}}{\Gamma(a^{(i)}_\beta)}(\sigma^2_{\beta^{(i)}})^{-a^{(i)}_\beta - 1}\exp(-b^{(i)}_\beta/\sigma^2_{\beta^{(i)}}).$$

Note that the proposed priors are non-informative in the sense that they are all centered at $1/M$, the constraint (=1) equally shared by M age groups.

2.2.2.2 Prior Distributions for Time Parameters

We consider the following priors for the parameters in (2.9) and (2.10)

$$\boldsymbol{\varphi} \sim N_2(\boldsymbol{\varphi}_0, \boldsymbol{\Sigma}_0),$$
$$\rho \sim N(0, \sigma^2_\rho)\mathbf{I}\left\{\rho \in (-1, 1)\right\},$$
$$\sigma^2_\kappa \sim \text{InvGamma}(a_\kappa, b_\kappa),$$

where $\boldsymbol{\varphi}_0$, $\boldsymbol{\Sigma}_0$, σ^2_ρ, a_κ, and b_κ are pre-specified hyperparameters, and $\mathbf{I}\left\{\rho \in (-1, 1)\right\}$ is an indicator function equal to 1 when ρ is between -1 and 1. For the dependence structure of $\boldsymbol{\kappa}^{(i)}$, we propose a conjugate prior for $p^{(i)} \equiv P(w^{(i)}_l = 1)$

$$p^{(i)} \sim \text{Beta}(a, b),$$

where a and b are pre-specified constants such that

$$\pi(p^{(i)}) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}(p^{(i)})^{a-1}(1-p^{(i)})^{b-1},$$

and consider

$$\rho^{(i)} \sim N(0, \sigma^2_{\rho^{(i)}})\mathbf{I}\left\{\rho^{(i)} \in (-1,1)\right\},$$
$$\sigma^2_{\kappa^{(i)}} \sim \text{InvGamma}(a^{(i)}_{\kappa}, b^{(i)}_{\kappa}),$$

where $\sigma^2_{\rho^{(i)}}$, $a^{(i)}_{\kappa}$, and $b^{(i)}_{\kappa}$ are pre-specified hyperparameters.

2.2.2.3 Prior Distributions for Overdispersion Parameters

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

$$\sigma_i^2 \sim \text{InvGamma}(a^{(i)}_{\mu}, b^{(i)}_{\mu}), \tag{2.13}$$

where $a^{(i)}_{\mu}$ and $b^{(i)}_{\mu}$ are pre-specified.

2.2.3 Posterior Computation

2.2.3.1 Posterior Distributions for Age Parameters

Let $\boldsymbol{\theta} = (e^{(1)}, e^{(2)}, \ldots, e^{(n)}, \boldsymbol{\beta}', (\boldsymbol{\beta}^{(1)})', (\boldsymbol{\beta}^{(2)})', \ldots, (\boldsymbol{\beta}^{(n)})', \sigma^2_{\beta}, \sigma^2_{\beta^{(1)}}, \sigma^2_{\beta^{(2)}}, \ldots, \sigma^2_{\beta^{(n)}}, \boldsymbol{\kappa}',$
$(\boldsymbol{\kappa}^{(1)})', (\boldsymbol{\kappa}^{(2)})', \ldots, (\boldsymbol{\kappa}^{(n)})', \boldsymbol{\varphi}', (\boldsymbol{\varphi}^{(1)})', (\boldsymbol{\varphi}^{(2)})', \ldots, (\boldsymbol{\varphi}^{(n)})', \rho, \rho^{(1)}, \rho^{(2)}, \ldots, \rho^{(n)}, \sigma^2_{\kappa}, \sigma^2_{\kappa^{(1)}},$
$\sigma^2_{\kappa^{(2)}}, \ldots, \sigma^2_{\kappa^{(n)}}, w_1^{(1)}, w_1^{(2)}, \ldots, w_1^{(n)}, w_2^{(1)}, w_2^{(2)}, \ldots, w_2^{(n)}, \lambda_1^{(1)}, \lambda_1^{(2)}, \ldots, \lambda_1^{(n)}, \lambda_2^{(1)}, \lambda_2^{(2)}, \ldots, \lambda_2^{(n)},$
$\eta_1^{(1)}, \eta_1^{(2)}, \ldots, \eta_1^{(n)}, \eta_2^{(1)}, \eta_2^{(2)}, \ldots, \eta_2^{(n)}, \tau_1, \tau_2, p^{(1)}, p^{(2)}, \ldots, p^{(n)}, \sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2, \nu_{x,t}^{(1)}, \nu_{x,t}^{(2)}, \ldots, \nu_{x,t}^{(n)})'$. The full conditional distributions of age parameters are given by

$$\begin{aligned}
\pi(e_x^{(i)}|.) &\propto \exp(-c_x^{(i)} e_x^{(i)})(e_x^{(i)})^{D^{(i)}_{x,\cdot}} \left|\frac{d}{de_x^{(i)}} g^{-1}(\alpha_x^{(i)})\right| \pi(e_x^{(i)}) \\
&\propto \exp\left[-(b_x^{(i)} + c_x^{(i)})e_x^{(i)}\right](e_x^{(i)})^{a_x^{(i)} + D^{(i)}_{x,\cdot} - 1}, \qquad (2.14) \\
\pi(\beta_x|.) &\propto \prod_{i=1}^{n} \prod_{t \in \Theta_{\text{time}}} \exp\left[-E^{(i)}_{x,t} \exp(\alpha_x^{(i)} + \beta_x \kappa_t + \beta_x^{(i)} \kappa_t^{(i)} + \nu_{x,t}^{(i)})\right] \times \exp(\beta_x \kappa_t D^{(i)}_{x,t}) \\
&\qquad \times \exp\left[-\frac{(\beta_x - \frac{1}{M})^2}{2\sigma^2_{\beta}}\right], \qquad (2.15) \\
\pi(\beta_x^{(i)}|.) &\propto \prod_{t \in \Theta_{\text{time}}} \exp\left[-E^{(i)}_{x,t} \exp(\alpha_x^{(i)} + \beta_x \kappa_t + \beta_x^{(i)} \kappa_t^{(i)} + \nu_{x,t}^{(i)})\right] \times \exp(\beta_x^{(i)} \kappa_t^{(i)} D^{(i)}_{x,t}) \\
&\qquad \times \exp\left[-\frac{(\beta_x^{(i)} - \frac{1}{M})^2}{2\sigma^2_{\beta^{(i)}}}\right], \qquad (2.16)
\end{aligned}$$

and

$$\pi(\sigma_\beta^2|.) \propto (\sigma_\beta^2)^{-\tilde{a}_\beta - 1} \exp(-\tilde{b}_\beta/\sigma_\beta^2), \tag{2.17}$$

$$\pi(\sigma_{\beta^{(i)}}^2|.) \propto (\sigma_{\beta^{(i)}}^2)^{-\tilde{a}_\beta^{(i)} - 1} \exp(-\tilde{b}_\beta^{(i)}/\sigma_{\beta^{(i)}}^2), \tag{2.18}$$

where the notation "|." represents "conditional on the data G and all other parameters ", $c_x^{(i)} = \sum_{t\in\Theta_{\text{time}}} E_{x,t}^{(i)} \exp(\beta_x\kappa_t + \beta_x^{(i)}\kappa_t^{(i)} + \nu_{x,t}^{(i)})$, $D_{x,\cdot}^{(i)} = \sum_{t\in\Theta_{\text{time}}} D_{x,t}^{(i)} - 1$, $\tilde{a}_\beta = a_\beta + \frac{M}{2}$, $\tilde{b}_\beta = b_\beta + \frac{1}{2}\left(\boldsymbol{\beta} - \frac{1}{M}\boldsymbol{J}_M\right)'\left(\boldsymbol{\beta} - \frac{1}{M}\boldsymbol{J}_M\right)$, $\tilde{a}_\beta^{(i)} = a_\beta^{(i)} + \frac{M}{2}$, 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 (2.14), (2.17), and (2.18), we have

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

and thus they can be easily sampled in each iteration.

Due to unidentifiable kernels in (2.15) and (2.16), the Metropolis-Hastings (MH) sampling is applied to update β_x and $\beta_x^{(i)}$. Let $\beta_x^{[j]}$ denote the j^{th} iteration of β_x, and let $\boldsymbol{\theta}\backslash\{\}$ denote all parameters in $\boldsymbol{\theta}$ 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 let $\tilde{B} = \sum_{y\le x}\beta_y^{[j]} + \sum_{y>x}B_y^{[j-1]}$, 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 obtain 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 calculated by the L_2 norm.

2.2.3.2 Posterior Distributions for Time Parameters

In this section, we separately discuss the sampling algorithms for common and population-specific time parameters because the dependence structure of the latter is further regularized by the Dirac spike. Let $\boldsymbol{\kappa}_{-t} = \boldsymbol{\kappa}\backslash\{\kappa_t\} = (\kappa_1,\ldots,\kappa_{t-1},\kappa_{t+1},\ldots,\kappa_{t_N})'$ 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

$$\begin{aligned}
\pi(\kappa_t|.) \propto \prod_{i=1}^{n}\prod_{x\in\Theta_{\text{age}}} &\exp\left[-E_{x,t}^{(i)}\exp(\alpha_x^{(i)} + \beta_x\kappa_t + \beta_x^{(i)}\kappa_t^{(i)} + \nu_{x,t}^{(i)})\right] \\
&\times \exp(\beta_x\kappa_t D_{x,t}^{(i)}) \times f(\kappa_t|\boldsymbol{\kappa}_{-t}), \qquad (2.19)\\
\pi(\boldsymbol{\varphi}|.) \propto \exp&\left[-\frac{1}{2\sigma_\kappa^2}(\boldsymbol{\varphi}'(\boldsymbol{\Sigma}^*)^{-1}\boldsymbol{\varphi} - 2(\boldsymbol{\kappa}'\boldsymbol{Q}\boldsymbol{W} + \sigma_\kappa^2\boldsymbol{\varphi}_0'\boldsymbol{\Sigma}_0^{-1})\boldsymbol{\varphi})\right], \qquad (2.20)\\
\pi(\rho|.) \propto \exp&\left[-\frac{1}{2\sigma_\kappa^2}\left(a_\rho\rho^2 + \frac{\sigma_\kappa^2}{\sigma_\rho^2}\rho^2 - 2b_\rho\rho\right)\right]\mathbf{I}\{\rho\in(-1,1)\}, \qquad (2.21)\\
\pi(\sigma_\kappa^2|.) \propto (\sigma_\kappa^2)^{-(a_\kappa+N/2)-1}&\exp\left[-\frac{1}{\sigma_\kappa^2}\left(b_\kappa + \frac{1}{2}(\boldsymbol{\kappa}-\boldsymbol{W}\boldsymbol{\varphi})'\boldsymbol{Q}(\boldsymbol{\kappa}-\boldsymbol{W}\boldsymbol{\varphi})\right)\right], \qquad (2.22)
\end{aligned}$$

where $f(\kappa_t|\boldsymbol{\kappa}_{-t})$ are the conditional distribution of κ_t based on AR(1) with a drift in (2.8), $\boldsymbol{\Sigma}^* = (\boldsymbol{W}'\boldsymbol{Q}\boldsymbol{W} + \sigma_\kappa^2\boldsymbol{\Sigma}_0^{-1})^{-1}$, $a_\rho = \sum_{t=t_2}^{t_N}(\kappa_{t-1}-\eta_{t-1})^2$, and $b_\rho = \sum_{t=t_2}^{t_N}(\kappa_t-\eta_t)(\kappa_{t-1}-\eta_{t-1})$. Note that when $t = t_1$,

$$\begin{aligned}
f(\kappa_t|\boldsymbol{\kappa}_{-t}) &\propto f(\kappa_t)f(\kappa_{t+1}|\kappa_t) \\
&\propto \exp\left[-\frac{1}{2\sigma_\kappa^2}[(\kappa_t-\eta_t)^2 + (\kappa_{t+1}-\eta_{t+1}-\rho(\kappa_t-\eta_t))^2]\right];
\end{aligned} \qquad (2.23)$$

when $t_1 < t < t_N$,

$$f(\kappa_t|\boldsymbol{\kappa}_{-t}) \propto f(\kappa_{t+1}|\kappa_t)f(\kappa_t|\kappa_{t-1})$$
$$\propto \exp\left[-\frac{1}{2\sigma_\kappa^2}[(\kappa_t - \eta_t - \rho(\kappa_{t-1} - \eta_{t-1}))^2 + (\kappa_{t+1} - \eta_{t+1} - \rho(\kappa_t - \eta_t))^2]\right]; \quad (2.24)$$

and when $t = t_N$,

$$f(\kappa_t|\boldsymbol{\kappa}_{-t}) \propto f(\kappa_t|\kappa_{t-1}) \propto \exp\left[-\frac{1}{2\sigma_\kappa^2}(\kappa_t - \eta_t - \rho(\kappa_{t-1} - \eta_{t-1}))^2\right]. \quad (2.25)$$

From (2.20), (2.21), and (2.22), $\boldsymbol{\varphi}, \rho$ and σ_κ^2 are updated by

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

To update κ_t, 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 $\sum_{t\in\Theta_{\text{time}}} \kappa_t = 0$,

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

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 obtain the j^{th} iteration of $\boldsymbol{\kappa}$.

To obtain an MCMC sample of $\boldsymbol{\kappa}^{(i)}$, $\boldsymbol{\varphi}^{(i)}$, $\rho^{(i)}$, and $\sigma^2_{\kappa^{(i)}}$, the status of $w_l^{(i)}$ is first determined via $w_l^{(i)} \sim \text{Bernoulli}(\xi_l^{(i)})$.

Then we have

$$\begin{aligned}\xi_l^{(i)} = p(w_l^{(i)} = 1 \mid w_{-l}^{(i)}, G) &= \frac{m(w_{-l}^{(i)}, G \mid w_l^{(i)} = 1) \cdot p^*_{l,i}}{m(w_{-l}^{(i)}, G \mid w_l^{(i)} = 0)(1 - p^*_{l,i}) + m(w_{-l}^{(i)}, G \mid w_l^{(i)} = 1) \cdot p^*_{l,i}} \\ &= \frac{c_l^{(i)}}{c_l^{(i)} + d_l^{(i)}} \end{aligned} \tag{2.26}$$

where $m(w_{-l}^{(i)}, G \mid w_l^{(i)} = 1)$ and $m(w_{-l}^{(i)}, G \mid w_l^{(i)} = 0)$ are conditional marginal likelihoods measuring the overall model fitting to the data when specifying $w_l^{(i)}$; $\boldsymbol{w_l} = \{w_l^{(1)}, \cdots, w_l^{(n)}\}$ and $w_{-l}^{(i)} = \boldsymbol{w_l} \backslash w_l^{(i)}$. Here, $c_l^{(i)} = p^{(i)} \exp\{-\frac{1}{2\sigma^2_{\kappa^{(i)}}}(\boldsymbol{\kappa^{(i)}} - \boldsymbol{W}\boldsymbol{\varphi}^*_{\boldsymbol{w,i}})^T \boldsymbol{Q^{(i)}}(\boldsymbol{\kappa^{(i)}} - \boldsymbol{W}\boldsymbol{\varphi}^*_{\boldsymbol{w,i}})\}$ and
$d_l^{(i)} = (1 - p^{(i)}) \exp\{-\frac{1}{2\sigma^2_{\kappa^{(i)}}}(\boldsymbol{\kappa^{(i)}} - \boldsymbol{W}\boldsymbol{\varphi}^{**}_{\boldsymbol{w,i}})^T \boldsymbol{Q^{(i)}}(\boldsymbol{\kappa^{(i)}} - \boldsymbol{W}\boldsymbol{\varphi}^{**}_{\boldsymbol{w,i}})\}$.

The vector $\boldsymbol{\varphi}^*_{\boldsymbol{w,i}}$ is the column vector of $\boldsymbol{\varphi_{w,i}}$ with the l-th entry replaced by $\varphi_l^{(i)}$, similarly, $\boldsymbol{\varphi}^{**}_{\boldsymbol{w,i}}$ is the column vector of $\boldsymbol{\varphi_{w,i}}$ with the l-th entry replaced by 0.

Then, based on the identifiable kernels of the full conditional distributions of $\boldsymbol{\varphi}^{(i)}$, $\rho^{(i)}$, $\sigma^2_{\kappa^{(i)}}$, and $p^{(i)}$, they are updated by

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

respectively, where $\boldsymbol{A}^* = (\boldsymbol{W}'\boldsymbol{Q}^{(i)}\boldsymbol{W} + \boldsymbol{I}_2)^{-1}$, $\boldsymbol{a}^* = \boldsymbol{A}^*\boldsymbol{W}'\boldsymbol{Q}^{(i)}\boldsymbol{\kappa}^{(i)}$, $\eta_t^{(i)} = \varphi_1^{(i)} + \varphi_2^{(i)} t$, $a_\rho^{(i)} = \sum_{t=t_2}^{t_N}(\kappa_{t-1}^{(i)} - \eta_{t-1}^{(i)})^2$, $b_\rho^{(i)} = \sum_{t=t_2}^{t_N}(\kappa_t^{(i)} - \eta_t^{(i)})(\kappa_{t-1}^{(i)} - \eta_{t-1}^{(i)})$. 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. Finally, due to the similar structure of $\pi(\kappa_t^{(i)}|.)$ as $\pi(\kappa_t|.)$

$$\pi(\kappa_t^{(i)}|.) \propto \prod_{x \in \Theta_{\text{age}}} \exp\left[-E_{x,t}^{(i)} \exp(\alpha_x^{(i)} + \beta_x \kappa_t + \beta_x^{(i)} \kappa_t^{(i)} + \nu_{x,t}^{(i)})\right] \times \exp(\beta_x^{(i)} \kappa_t^{(i)} D_{x,t}^{(i)}) \times f(\kappa_t^{(i)}|\kappa_{-t}^{(i)}),$$

the procedure to update $\kappa_t^{(i)}$ is similar to the one for κ_t.

2.2.3.3 Model Selection for Population-specific Time Parameters with spatial information

Variables located at adjacent spots in space may tend to perform similar behavior. Further, spatial interaction effects can be well exhibited due to the varying geographic location by probit framework with spatial dependencies. Here, we suggest a Bayesian probit model to explain choices of latent variable within the decision of time series structure for $\boldsymbol{\kappa^{(i)}}$.

For non-spatial model, $\lambda^{(i)}$ and $\lambda^{(j)}$ are theoretically uncorrelated without sharing any geographical information each other. Conversely, the sign of latent variable λ in one region affects ones in other adjacent areas based on the proposed spatial probit model. Instead of setting the prior of $p^{(i)}$ as a Beta distribution, we consider there are some spatial information to explore $w_l^{(i)}$ and thus we propose a spatial latent variable $\lambda_l^{(i)} = \eta_l^{(i)} + \epsilon_l^{(i)}$ to construct a following assumption:

$$w_l^{(i)} = \begin{cases} 1 & \text{if} \quad \lambda_l^{(i)} > 0, \\ 0 & \text{otherwise}, \end{cases} \tag{2.27}$$

where $l = 1, 2$ and $i = 1, 2, \cdots, n$.

The hierarchical Bayesian spatial probit model allows for spatial interaction effects as well as heterogeneous individual effects here. The model extends the traditional Bayesian variable selection model by encouraging us to exhibit spatial similarities. In addition to spatial interaction effects, the model also accommodates variation in individual choices presumed to be located at distinct points in space.

Now, in order to bring spatial information, we let $\epsilon_l^{(1)}, \cdots, \epsilon_l^{(n)}$ comply a Conditional Au-

toregressive (CAR) model. Further, we set

$$\boldsymbol{\epsilon_l} \mid \tau_l \sim N(\mathbf{0}, (I_n - \tau_l W_p)^{-1}) \tag{2.28}$$

$$\boldsymbol{\eta_l} \sim N(\boldsymbol{\eta_0}, I_n) \tag{2.29}$$

$$\tau_l \sim Unif(\tau_{min}^{-1}, \tau_{max}^{-1}) \tag{2.30}$$

where $\boldsymbol{\eta_l} = (\eta_l^{(1)}, \eta_l^{(2)}, \cdots, \eta_l^{(n)})$, and $\boldsymbol{\epsilon_l} = (\epsilon_l^{(1)}, \epsilon_l^{(2)}, \cdots, \epsilon_l^{(n)})$; τ_{min},τ_{max} are minimum and maximum eigen-values of the adjacent matrix $W_p = (w_{i.j} : i, j = 1, \ldots, n)$, $w_{i,j} = 1$ if i and j are adjacent regions; $w_{i,j} = 0$, otherwise and where $V_l = I_n - \tau_l W_p$.

The full conditional distribution of $\boldsymbol{\eta_l}$ becomes

$$\begin{aligned} p(\boldsymbol{\eta_l} \mid \boldsymbol{\lambda_l}, G) \propto & \exp\{(-\frac{1}{2}(\boldsymbol{\lambda_l} - \boldsymbol{\eta_l})^T V_l (\boldsymbol{\lambda_l} - \boldsymbol{\eta_l})\} \times \\ & \exp\{-\frac{1}{2}(\boldsymbol{\eta_l} - \boldsymbol{\eta_0})^T (\boldsymbol{\eta_l} - \boldsymbol{\eta_0})\} \end{aligned} \tag{2.31}$$

which is a multinormal distribution,

$$\boldsymbol{\eta_l} \sim N(A_l^{-1} m_l, A_l^{-1}), \tag{2.32}$$

where $A_l = (V_l + I)$, $m_l = (\boldsymbol{\eta_0} + V_l \boldsymbol{\lambda_l})$ and $\boldsymbol{\lambda_l} = (\lambda_l^{(1)}, \lambda_l^{(2)}, \cdots, \lambda_l^{(n)})$.

By proposing a spatial structure, we have

$$\lambda_l^{(i)} \mid \boldsymbol{\lambda_l^{(-i)}} \sim N(\eta_l^{(i)} + \tau_l \sum_{i \neq j} w_{i,j} (\lambda_l^{(j)} - \eta_l^{(j)}), 1) \tag{2.33}$$

$$\sim N(\mu_i^*, 1), \tag{2.34}$$

$$\begin{aligned}\pi(w_l^{(i)} = 1 \mid \boldsymbol{\lambda_l^{(-i)}}) &= p(\lambda_l^{(i)} > 0 \mid \boldsymbol{\lambda_l^{(-i)}}) \\ &= \Phi(\eta_l^{(i)} + \tau_l \sum_{i \neq j} w_{i,j}(\lambda_l^{(j)} - \eta_l^{(j)})) \\ &= \Phi(\mu_i^*) \qquad (2.35) \\ &:= p_{l,i}^* \qquad (2.36)\end{aligned}$$

where

$$\mu_i^* = \eta_l^{(i)} + \tau_l \sum_{i \neq j} w_{i,j}(\lambda_l^{(j)} - \eta_l^{(j)}) \qquad (2.37)$$

and $\boldsymbol{\lambda_l^{(-i)}} = \boldsymbol{\lambda_l} \backslash \lambda_l^{(i)} = (\lambda_l^{(1)}, \cdots, \lambda_l^{(i-1)}, \lambda_l^{(i+1)}, \cdots, \lambda_l^{(n)})$; $p_{l,i}^*$ denotes a probability $p(\lambda_l^{(i)} > 0 \mid \boldsymbol{\lambda_l^{(-i)}})$.

Note that the full conditional distribution of $w_l^{(i)}$ is given by

$$\begin{aligned}p(w_l^{(i)} = 1 \mid w_{-l}^{(i)}, G) &= \frac{m(w_{-l}^{(i)}, G \mid w_l^{(i)} = 1) \cdot p_{l,i}^*}{m(w_{-l}^{(i)}, G \mid w_l^{(i)} = 0)(1 - p_{l,i}^*) + m(w_{-l}^{(i)}, G \mid w_l^{(i)} = 1) \cdot p_{l,i}^*} \\ &= \frac{c_l^{(i)}}{c_l^{(i)} + d_l^{(i)}} \qquad (2.38)\end{aligned}$$

Here, $c_l^{(i)} = p_{l,i}^* \exp\{-\frac{1}{2\sigma^2_{\kappa^{(i)}}}(\boldsymbol{\kappa^{(i)}} - \boldsymbol{W\varphi^*_{w,i}})^T \boldsymbol{Q^{(i)}}(\boldsymbol{\kappa^{(i)}} - \boldsymbol{W\varphi^*_{w,i}})\}$ and $d_l^{(i)} = (1 - p_{l,i}^*) \exp\{-\frac{1}{2\sigma^2_{\kappa^{(i)}}}(\boldsymbol{\kappa^{(i)}} - \boldsymbol{W\varphi^{**}_{w,i}})^T \boldsymbol{Q^{(i)}}(\boldsymbol{\kappa^{(i)}} - \boldsymbol{W\varphi^{**}_{w,i}})\}$.

Also, the conditional posterior distribution of $\boldsymbol{\lambda_l}$ is

$$\begin{aligned}p(\lambda_l^{(i)} \mid \cdot) &\propto p(w_l^{(i)} \mid \lambda_l^{(i)}) \pi(\lambda_l^{(i)} \mid \boldsymbol{\lambda_l^{(-i)}}) \qquad (2.39) \\ &\propto \exp[-\frac{1}{2}(\lambda_l^{(i)} - \mu_i^*)^2] \times \\ &\quad \Big(\mathbf{I}\{w_l^{(i)} = 1\}\mathbf{I}\{\lambda_l^{(i)} > 0\} + \mathbf{I}\{w_l^{(i)} = 0\}\mathbf{I}\{\lambda_l^{(i)} \leq 0\}\Big) \qquad (2.40)\end{aligned}$$

and thus $\lambda_l^{(i)} \mid \cdot \sim N(\mu_i^*, 1)$ left truncated at 0 if $w_l^{(i)} = 1$

$\lambda_l^{(i)} \mid \cdot \sim N(\mu_i^*, 1)$ right truncated at 0 if $w_l^{(i)} = 0$

In addition, the full conditional distribution for τ_l is given by

$$\begin{aligned} p(\tau_l \mid \cdot) &\propto \pi(\boldsymbol{\epsilon_l} \mid \tau_l) \cdot \pi(\tau_l) \\ &\propto |V_l|^{\frac{1}{2}} \exp\{-\frac{1}{2}(\boldsymbol{\lambda_l} - \boldsymbol{\eta_l})^T V_l (\boldsymbol{\lambda_l} - \boldsymbol{\eta_l})\} \times \mathbf{I}\{\tau_{min}^{-1}, \tau_{max}^{-1}\} \end{aligned} \tag{2.41}$$

where $\tau_l \in (\tau_{min}^{-1}, \tau_{max}^{-1})$, so it is not reducible to a standard distribution (LeSage (2000)) and then we might adopt a Metropolis-Hastings algorithm during the sampling procedures. We suggest a normal distribution be used as a transition kernel in the M-H step. Additionally, the restriction of τ_l to the interval $(\tau_{min}^{-1}, \tau_{max}^{-1})$ can be implemented using a rejection-sampling step during the MCMC sampling.

Given that the proposed function is $\tau_l^* \sim N(\tau_l^{[j-1]}, \sigma_l^2)\mathbf{I}\{\tau_{min}^{-1}, \tau_{max}^{-1}\}$, where σ_l^2 is chosen to have the acceptance probability around 20%∼40%, we then have

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

where

$$\Phi(\tau_l^{[j-1]}, \tau_l^*) = \min\left\{1, \frac{p(\tau_l^* \mid \boldsymbol{\eta_l}, \boldsymbol{\lambda_l})}{p(\tau_l^{[j-1]} \mid \boldsymbol{\eta_l}, \boldsymbol{\lambda_l})}\right\}.$$

2.2.3.4 Posterior Distributions for Overdispersion Parameters

Because

$$\pi(\sigma_i^2|.) \propto (\sigma_i^2)^{-(a_\mu^{(i)} + MN/2)-1} \exp\left[-\frac{1}{\sigma_i^2}\left(b_\mu^{(i)} + \frac{1}{2}\sum_{x\in\Theta_{\text{age}}}\sum_{t\in\Theta_{\text{time}}}(\nu_{x,t}^{(i)})^2\right)\right], \tag{2.42}$$

$$\begin{aligned} \pi(\nu_{x,t}^{(i)}|.) \propto &\exp\left[-E_{x,t}^{(i)}\exp(\alpha_x^{(i)} + \beta_x\kappa_t + \beta_x^{(i)}\kappa_t^{(i)} + \nu_{x,t}^{(i)})\right] \times \exp(\nu_{x,t}^{(i)} D_{x,t}^{(i)}) \\ &\times \exp\left[-\frac{(\nu_{x,t}^{(i)})^2}{2\sigma_i^2}\right], \end{aligned} \tag{2.43}$$

we have

$$\sigma_i^2|. \sim \text{InvGamma}\left(a_\mu^{(i)} + \frac{MN}{2}, b_\mu^{(i)} + \frac{1}{2}\sum_{x\in\Theta_{\text{age}}}\sum_{t\in\Theta_{\text{time}}}(\nu_{x,t}^{(i)})^2\right), \tag{2.44}$$

and update $\nu_{x,t}^{(i)}$ via the MH sampling. Given that $\boldsymbol{\theta}^{[j]}\backslash(\nu_{x,t}^{(i)})^{[j]}$ and $(\nu_{x,t}^{(i)})^{[j-1]}$ are available, the proposed function is $(\nu_{x,t}^{(i)})^* \sim N((\nu_{x,t}^{(i)})^{[j-1]}, \sigma_q^2)$, where σ_q^2 is chosen to have the acceptance probability around 20%∼40%, we then have

$$(\nu_{x,t}^{(i)})^{[j]} = \begin{cases} (\nu_{x,t}^{(i)})^* & \text{if } u \le \Phi((\nu_{x,t}^{(i)})^{[j-1]}, (\nu_{x,t}^{(i)})^*) \\ (\nu_{x,t}^{(i)})^{[j-1]} & \text{o.w.} \end{cases},$$

where

$$\Phi((\nu_{x,t}^{(i)})^{[j-1]}, (\nu_{x,t}^{(i)})^*) = \min\left\{1, \frac{\pi((\nu_{x,t}^{(i)})^*|\boldsymbol{\theta}^{[j]}\backslash\{(\nu_{x,t}^{(i)})^{[j]}\}, G)}{\pi((\nu_{x,t}^{(i)})^{[j-1]}|\boldsymbol{\theta}^{[j]}\backslash\{(\nu_{x,t}^{(i)})^{[j]}\}, G)}\right\}.$$

2.3 Numerical Analysis

2.3.1 Data Description

Two Japanese mortality datasets from the Japanese Mortality Database (JMD) were selected to illustrate our proposed methods. One contains national gender-specific deaths and exposures from 1951 to 2016; the other contains deaths and exposures in six adjacent counties (Akita, Aomori, Fukushima, Iwate, Miyagi and Yamagata) in the northeast Japan region and two counties (Tokyo and Kanagawa) in the central Japan from 1975 to 2016.

For the first dataset, we consider each gender as a single population, and calibrate the model using the data from 1951 to 2000 for ages 0-99, and separate the data from 2001 to 2016 for the validation purpose. In the second case of eight regions, we establish the model based on the data from 1975 to 2004 and choose the data from 2005 to 2016 as the validation set. Hence, we have $i = F$ or M, $x_1 = 0$, $x_M = 99$, $t_1 = 1951$, $t_N = 2000$, $M = 100$, and $N = 50$ for national data and have $i = \{1, \cdots, 8\}$, $x_1 = 0$, $x_M = 99$, $t_1 = 1975$, $t_N = 2004$, $M = 100$, and $N = 30$ for regional data.

2.3.2 Initial Settings of Prior Distributions

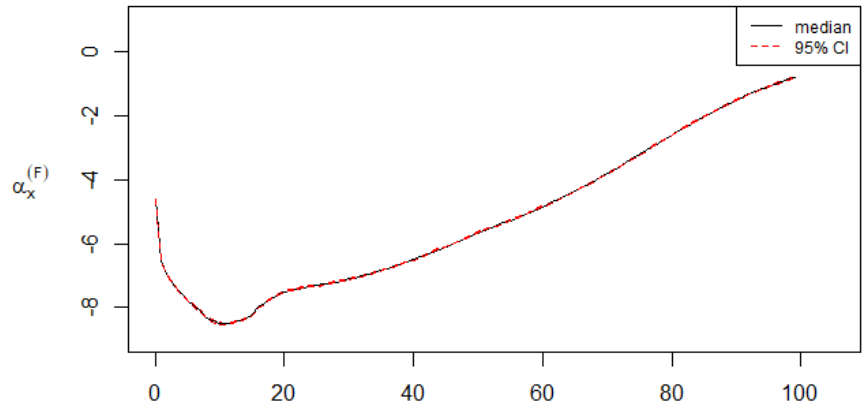
Following the prior specifications in section 2.2.2, we consider $a_x^{(i)} = b_x^{(i)} = 1$, $a_\beta = b_\beta = 0.01$, and $a_\beta^{(i)} = b_\beta^{(i)} = 0.001$ as age-related hyperparameters, and set $a_\mu^{(i)} = b_\mu^{(i)} = 2.5$ for overdispersion parameters. For those hyperparameters related to the time factors, they are set as $a = b = 1$, $a_\kappa = b_\kappa = 0.001$, $a_\kappa^{(i)} = b_\kappa^{(i)} = 0.001$, $\sigma_\rho^2 = 1$, $\sigma_{\rho^{(i)}}^2 = 0.1$, $\boldsymbol{\varphi}_0 = (0,0)'$, $\Sigma_0 = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$, 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 ($50 \times 100 \times 2$ and $30 \times 100 \times 8$) of our analyzing data sets.

2.3.3 Estimation and Model Performance

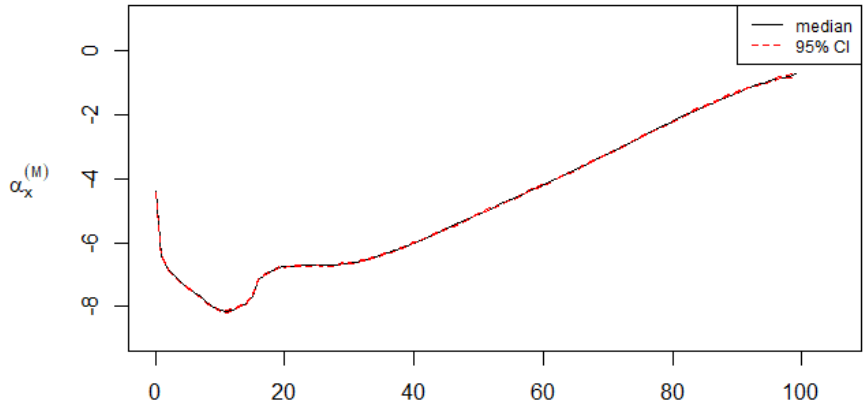
To evaluate the performance of BPLNLCrm, we generate an MCMC sample of 20000 iterations with the first 10000 as burn-in (so that $j = 1, 2, \ldots, 10000$), and summarize the posterior medians of $\boldsymbol{\alpha}^{(i)}$, $\boldsymbol{\beta}$, $\boldsymbol{\beta}^{(i)}$, $\boldsymbol{\kappa}$ and $\boldsymbol{\kappa}^{(i)}$ along with the 95% highest posterior density (HPD) intervals in Sections 2.3.3.1 and 2.3.3.2, respectively. We also assess the overall model fitting by comparing the posterior predictive distributions of death tolls with the observed counts, and compare the 20-year gender-specific mortality projection with the method by Antonio et al. (2015), and discuss performance of two proposed models with regularized time structure for regional data in Section 2.3.4.

2.3.3.1 Estimation for Age Parameters

Figures 2.1(a) and 2.1(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 that there are 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 related to aging.



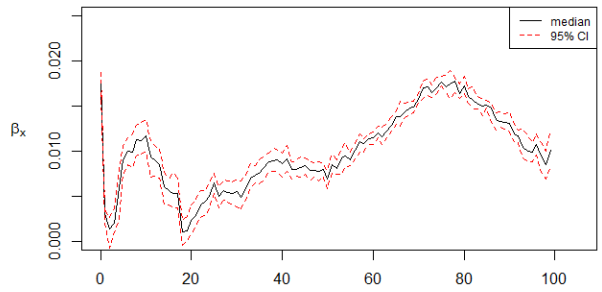
(a)



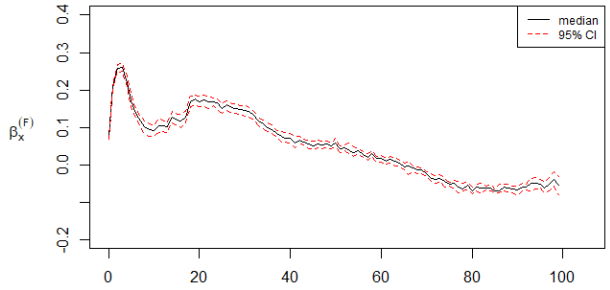
(b)

Fig. 2.1. Plots of the posterior medians of $\alpha_x^{(F)}$ and $\alpha_x^{(M)}$ with their 95% HDP intervals.

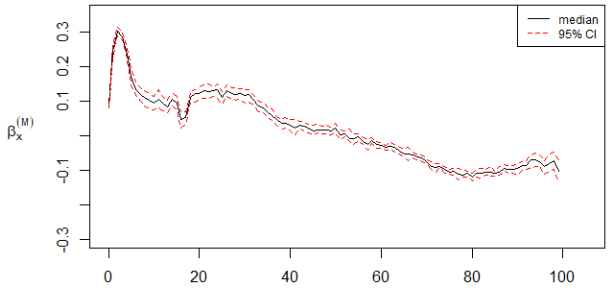
Similarly, Figure 2.2(a) shows the posterior median and 95% HPD interval of common factor β_x while Figures 2.2(b) and 2.2(c) are for the population-specific parameters $\beta_x^{(F)}$ and $\beta_x^{(M)}$, respectively. It can be seen that the corresponding posterior distributions are concentrated, indicating that the effect sizes of β_x and $\beta_x^{(i)}$ are less sensitive to all time change at any ages.



(a)



(b)



(c)

Fig. 2.2. Plots of the posterior medians of β_x, $\beta_x^{(F)}$ and $\beta_x^{(M)}$ with their 95% HDP intervals.

2.3.3.2 Estimation for Time Parameters

Before looking at the results of $\boldsymbol{\kappa}$, $\boldsymbol{\kappa}^{(F)}$, and $\boldsymbol{\kappa}^{(M)}$, we first discuss the selected models for the dependence structures of gender-specific time parameters. Out of 10000 MCMC iterations, we observe that the non-zero proportions of $w_1^{(F)}$, $w_2^{(F)}$, $w_1^{(M)}$, and $w_2^{(M)}$ are described in the following table:

Table 2.1: Frequency of variable selection in the model

Variables	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
$w_1^{(F)}=1, w_1^{(M)}=1, w_2^{(F)}=0, w_2^{(M)}=0$	0.01

Accordingly, using this table, the time series models for $\boldsymbol{\kappa}^{(F)}$ and $\boldsymbol{\kappa}^{(M)}$ in (2.8) are determined to

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

and

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

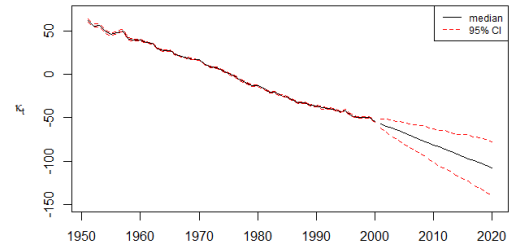
respectively.

In Figures 2.3(a)-2.3(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 (2.8) iteratively. Specifically, we have

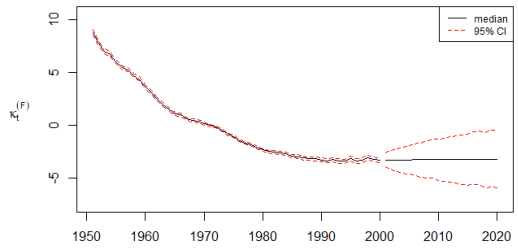
$$\kappa_{N+t'}^{[j]} \sim N \left(\varphi_1^{[j]} + \varphi_2^{[j]} (N+t') + \rho^{[j]} [\kappa_{N+t'-1}^{[j]} - \varphi_1^{[j]} - \varphi_2^{[j]} (N+t'-1)], (\sigma_\kappa^2)^{[j]} \right)$$

for $j = 1, 2, \ldots, 10000$ and $t' = 1, 2, \ldots, 20$. A similar procedure on the third equation in (2.8) is implemented for $\kappa_t^{(i)}$. From Figure 3, we observe decreasing trends in most of time windows, which

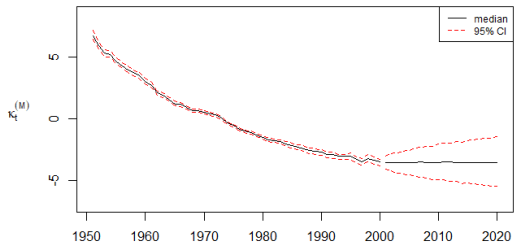
might be attributed to the advances in medical technology and social welfare. We also observe that $\kappa_t^{(F)}$ and $\kappa_t^{(M)}$ have similar estimated curves and will converge to the same size when time passes, meaning that the gender-specific time effects will reach a stable status in the long run as mentioned in Li and Lee (2005).



(a)



(b)



(c)

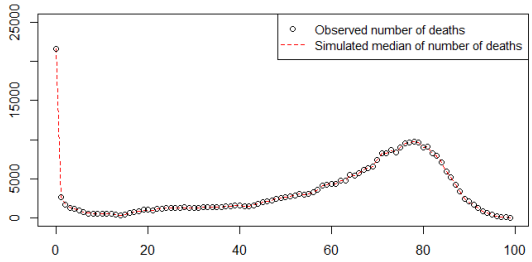
Fig. 2.3. Plots of the posterior medians of κ_t, $\kappa_t^{(F)}$ and $\kappa_t^{(M)}$ with their 95% HDP intervals and the corresponding 20-year ahead projection.

2.3.4 Assessments of Model Fitting and Prediction

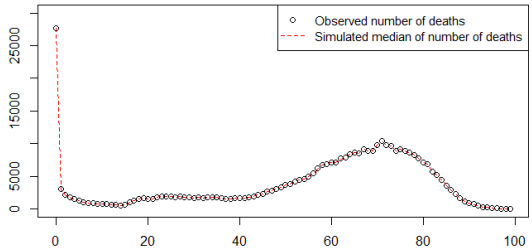
2.3.4.1 The Proposed Model with Genders

The way to empirically peek the posterior distributions of future time effects (that is, $\kappa_{2001}, \kappa_{2002}, \ldots, \kappa_{2020}$) in the previous section can also be used to assess the overall model fitting and its prediction ability. Particularly, we compare the observed death tolls $D_{x,t}^{(i)}$ and mortality rates $\mu_{x,t}^{(i)}$ with the ones $(D_{x,t}^{(i)})^{[j]}$ and $(\mu_{x,t}^{(i)})^{[j]}$ that are simulated from the fitted model, where $(D_{x,t}^{(i)})^{[j]}$ follows the Poisson distribution with mean equal to $E_{x,t}^{(i)}(\mu_{x,t}^{(i)})^{[j]}$, and $(\mu_{x,t}^{(i)})^{[j]} = \exp\left[(\alpha_x^{(i)})^{[j]} + \beta_x^{[j]}\kappa_t^{[j]} + (\beta_x^{(i)})^{[j]}(\kappa_t^{(i)})^{[j]} + (\nu_{x,t}^{(i)})^{[j]}\right]$ for $j = 1, 2, \ldots, 10000$.

Figures 2.4-2.6 present the medians of simulated number of deaths for each gender at any ages in three selected years 1960, 1980, and 2000, respectively. It is clear that the simulated values are close to the observed ones, implying that our proposed model fits well to the Japanese mortality data.



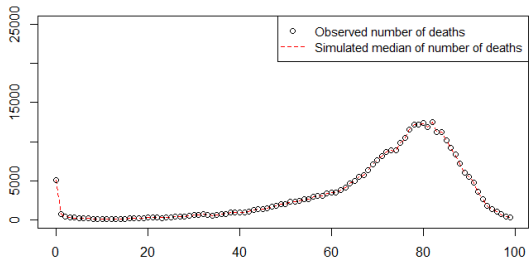
(a)



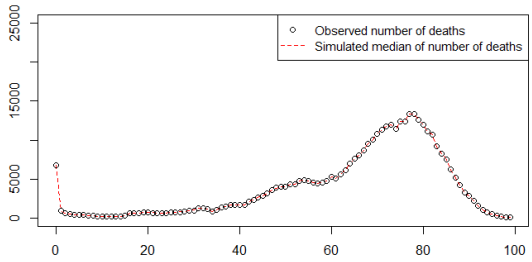
(b)

Fig. 2.4. Plots of the observed and simulated number of deaths for female (top) and male (bottom) in 1960.

Also, the figures show the test statistics median is much close to observed data points in each examples. We use this approach to conduct a model validation to verify the model specification under our hierarchical model assumptions. Usually the analogous integral does not contains a closed form solution, and about the review of approaches to estimate the posterior predictive distribution, readers are referred to Krüger et al. (2020).

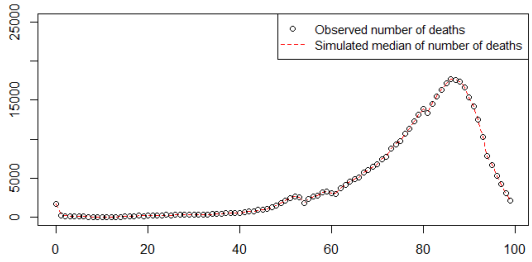


(a)

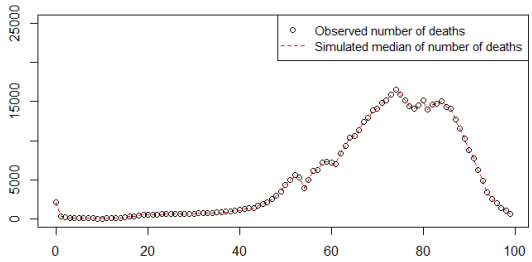


(b)

Fig. 2.5. Plots of the observed and simulated number of deaths for female (top) and male (bottom) in 1980.



(a)



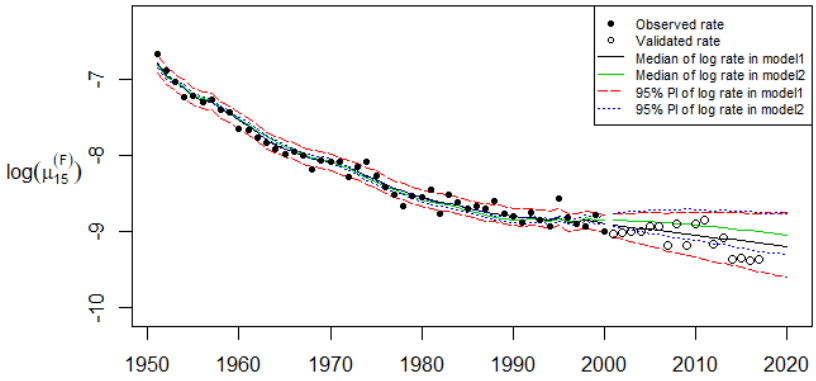
(b)

Fig. 2.6. Plots of the observed and simulated number of deaths for female (top) and male (bottom) in 2000.

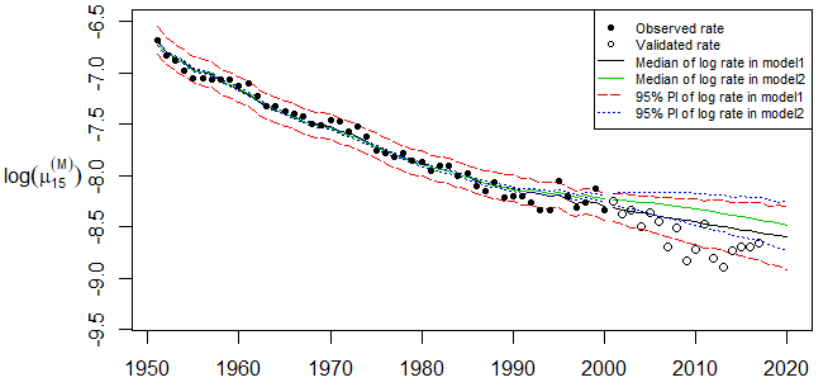
Figures 2.7-2.9 present the medians and 95% HPD intervals of simulated log mortality rates at three selected ages 15, 55, and 70, respectively. In addition to the training time window (years 1950-2000), 20-year ahead projections are provided to assess the prediction ability of BPLNLCrm (marked as "model 1"). We also include the simulated results of the method by (Antonio et al., 2015) (marked as "model 2") as a comparison. For the improvement of the model, the criterion we suggest is more observed and holdout data are expected to close to the projected medians and in the reasonable credible intervals.

From Figures 2.7-2.9, 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 models provides better 20-years ahead projections because validated log rates are closer to the black 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. In general, the prediction ability to future mortality rate shows the overdispersion model has a better performance by the validation approach for 2001 to 2016 span.

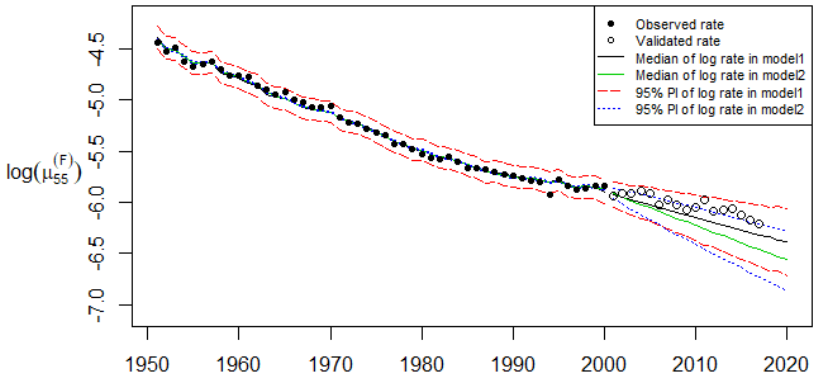


(a)

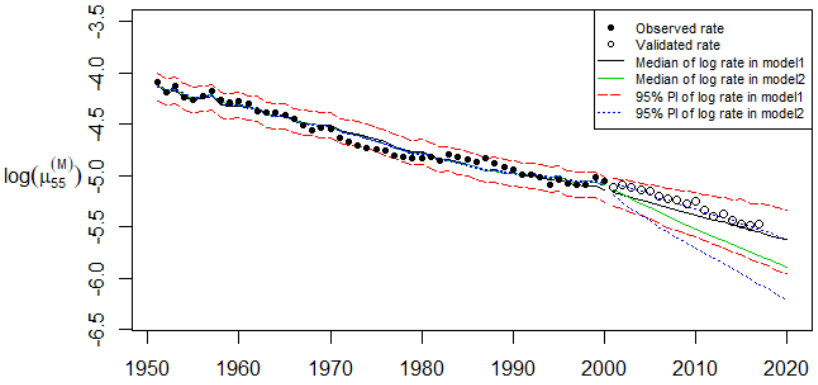


(b)

Fig. 2.7. Plots of the observed and simulated log death rates at age 15 along with 20-year ahead projections and 95% HDP intervals for female (top) and male (bottom).

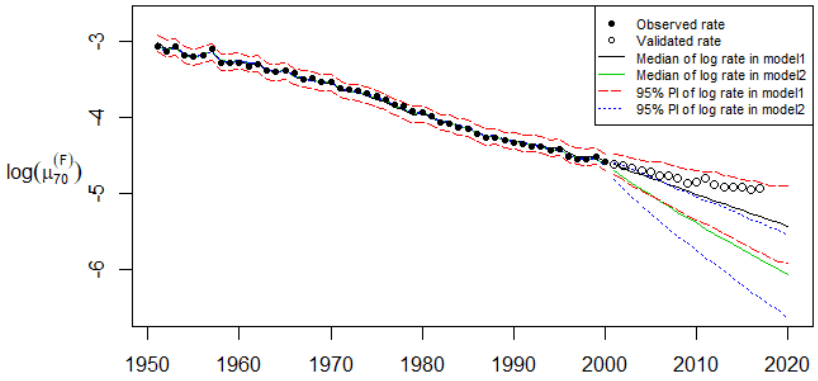


(a)

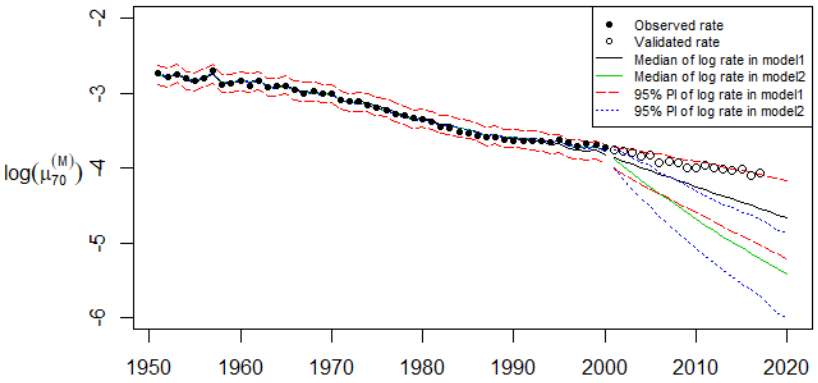


(b)

Fig. 2.8. Plots of the observed and simulated log death rates at age 55 along with 20-year ahead projections and 95% HDP intervals for female (top) and male (bottom).



(a)



(b)

Fig. 2.9. Plots of the observed and simulated log death rates at age 70 along with 20-year ahead projections and 95% HDP intervals for female (top) and male (bottom).

2.3.4.2 The Proposed Model with Spatial Information

The dominant variables structure can be determined with the higher frequency in Table 2.2 compared to those in Table 2.3. It is mentioned that $\boldsymbol{w_1} = (w_1^{(8)}, \ldots, w_1^{(1)})$ and $w_2^{(i)}$ denote the non-zero selected latent model variable in the tables. The largest proportion increased since the posterior selection behavior would affect other adjacent places in the proposed spatial probit model.

Therefore, most of binary variables tend to synchronize to one consequence in neighboring regions during each update process.

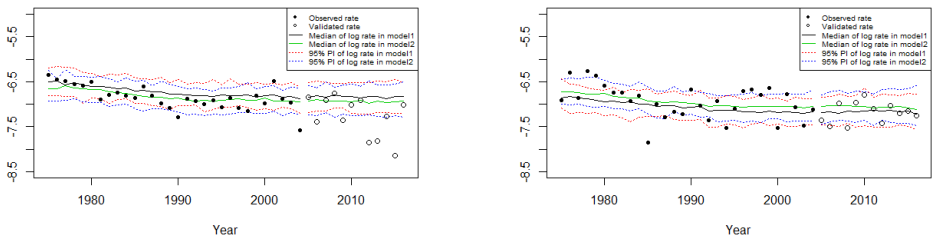
Table 2.2: High frequency selections in the model without spatial information

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 2.3: High frequency selections in the model with spatial information

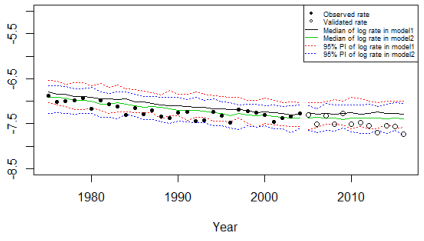
Variables	Proportion
$\boldsymbol{w_1}; w_2^{(1)}, w_2^{(2)}, w_2^{(3)}, w_2^{(4)}, w_2^{(5)}, w_2^{(6)}$	0.38
$\boldsymbol{w_1}; w_2^{(1)}, w_2^{(2)}, w_2^{(3)}, w_2^{(4)}, w_2^{(5)}, w_2^{(6)}, w_2^{(7)}$	0.07
$\boldsymbol{w_1}; w_2^{(1)}, w_2^{(2)}, w_2^{(3)}, w_2^{(4)}, w_2^{(5)}, w_2^{(6)}, w_2^{(8)}$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

The Markov chains underlying Bayesian spatial probit model are efficient at handling mixed problems such as model selection. This approach can deal with a large number of spatial sets and considers the interactions among these sets. Instead of first finding the combination with significant effects and then assessing their overlaps and making projections among them, this approach would accomplish this process in just one step. The focus now is to explore the dominant variable combinations performed by variable selection procedure and how this affect dependence with other spaces.

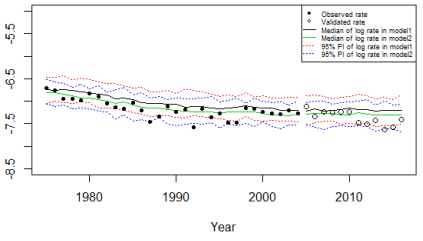


(a)

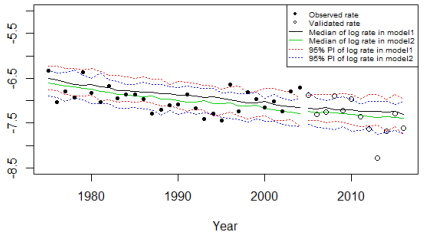
(b)



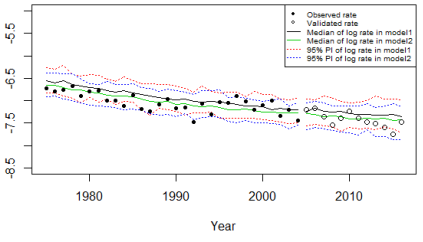
(c)



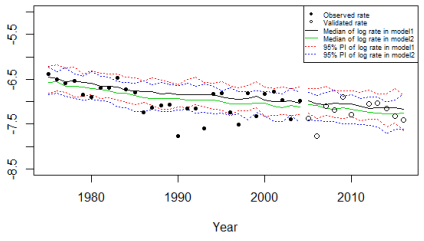
(d)



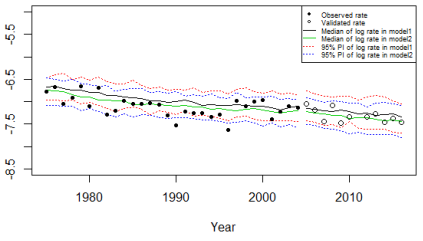
(e)



(f)

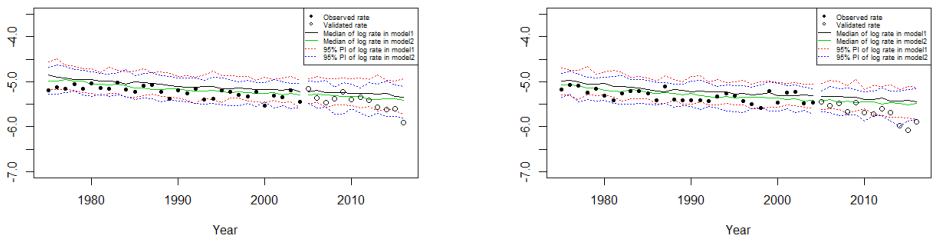


(g)



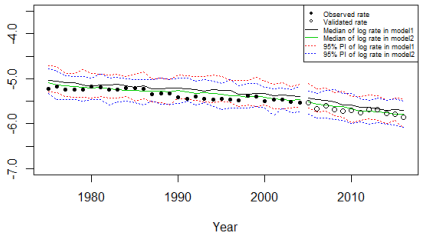
(h)

Fig. 2.10. Plots of the observed log death rates, fitted log death rates and 95% HDP intervals aged 35

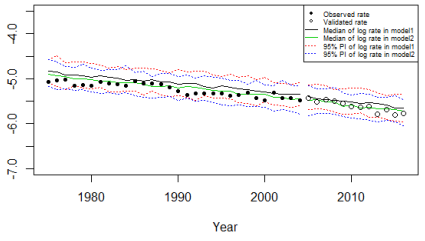


(a)

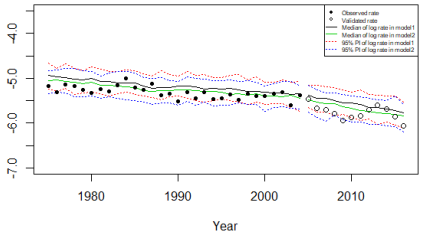
(b)



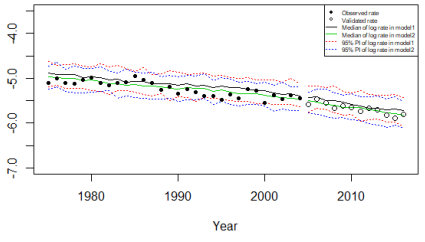
(c)



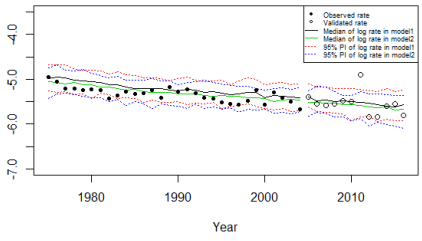
(d)



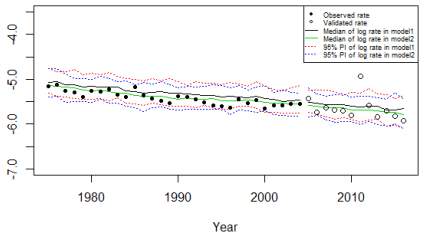
(e)



(f)

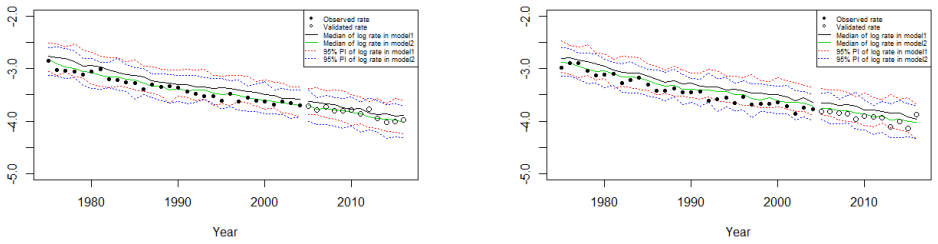


(g)

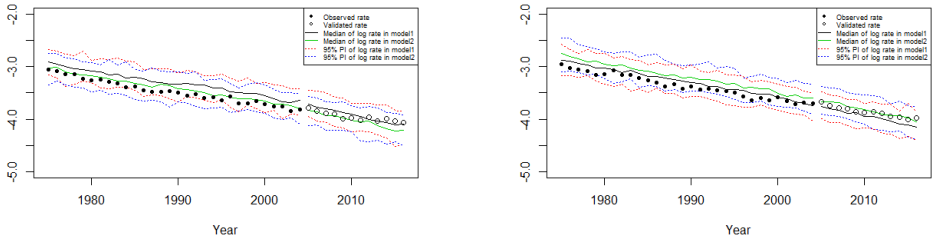


(h)

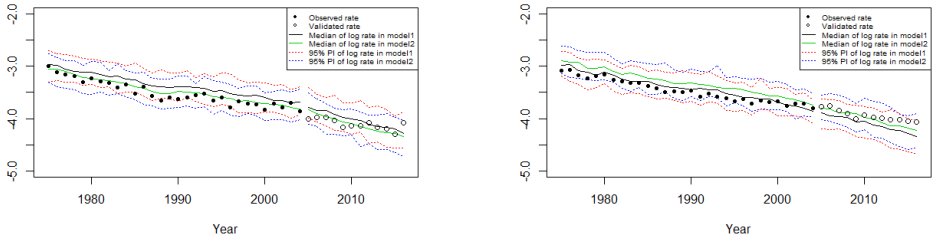
Fig. 2.11. Plots of the observed log death rates, fitted log death rates and 95% HDP intervals aged 55



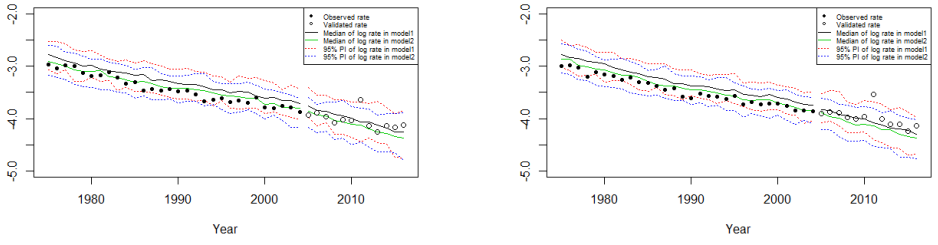
(a) (b)



(c) (d)



(e) (f)



(g) (h)

Fig. 2.12. Plots of the observed log death rates, fitted log death rates and 95% HDP intervals aged 75

Further, we evaluate our spatial model by the comparison of variability for previous three models. Figures 2.10 to 2.12 present the 95% HPD intervals of simulated log mortality rates at three selected ages 35, 55, and 75 for two models, respectively. In addition to the training time window (years 1975-2004), 12-year ahead projections are provided to assess the prediction ability of our spatial model. For convenience, model 1 and model 2 denote the traditional Dirac spike model and the proposed probit model in the following discussion, respectively. In figures, each subfigure represents sequentially mortality projections of Akita, Aomori, Fukushima, Iwate, Miyagi, Yamagata, Tokyo and Kanagawa eight regions from (a) - (h). From Figure 2.10 to 2.12, we see that almost every observed points of 1975-2004 fall in the 95% prediction intervals (blue curves). Moreover, the spatial model provides solid 12-years ahead projections because validated log rates are down the middle of those blue curves. We also notice that the prediction intervals may properly present the variability of data by introducing additional spatial terms.

The greater flexibility of the probit models allows the fitted values to generate an appropriate time series structure for parameters, where the variations due to the adjacent correlation are absorbed into the spatial structure. These projection plots clearly favor the proposed spatial models here, where their credible intervals provide reasonably high coverage of the observed rates across the ages, with most points dispersing around the medians and lying within the intervals, while the credible intervals for the new model do not appear to be overly much wider.

2.4 Discussion

This chapter 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 Dirac spike function, the new model can simultaneously conduct model selection and estimation of population-specific time effects. As a result, with this affordable computation even when n is large, it can avoid unnecessary assumptions on dependence structures of $\kappa_t^{(i)}$.

Further, we go to another direction to improve BPLNLCrm with considering the geographical dependence structure of death tolls among populations. For our proposed method, the number of death in each population is treated as conditional independence. However, say using the Japanese

mortality data as an example, it is reasonable to believe there are some unmeasurable factors such as culture and dietary habits affecting both adjacent regional moralities. Similarly, when multi-population refers to multi-county, interactions among counties can also be measured and quantified by our proposed Bayesian spatial probit model. Hence, the assumption of conditional independence may be relaxed.

Chapter 3

Bayesian Hierarchical Spatial Modeling for Morality Forecasting

3.1 Introduction

The use of Bayesian methods in the fields of disease mapping, epidemiology, and regional health applications is well established. Clayton and Kaldor (1987) proposed an empirical Bayes estimation to map mortality from diseases such as cancer. Besag et al. (1991) discuss a spatial modeling method to estimate relative risks of diseases under a Bayesian framework. The idea is based on the decomposition of spatial structure component and specified regional random effects for a risk parameter η_i. It can be depicted as

$$\eta_i = \mu + \phi_i + \theta_i, \tag{3.1}$$

where μ is a fixed effect representing the overall risk level across all areas, ϕ_i is a conditional autoregressive spatial component and θ_i is an ordinary random effect on diversity. The spatial structured random effects are specified via an adjacent matrix of the counties by conditional autoregressive priors. Lack of temporal parameters in the model, however, space-time interaction effects cannot be measured gradually. To explain how spatial heterogeneity and patterns evolved over time, Knorr-Held (2000) introduced a nested spatio-temporal model, which calibrated log odds as a combination

of temporal and spatial main effects and spatio-temporal interactions. In detail, the author assumed counts of deaths y_{it} in county i and year t, and counts of persons at risk n_{it} in county i and year t. Then, $y_{it} \sim Bin(n_{it} \cdot \pi_{it}); i = 1, \ldots, m$ counties and $t = 1, \ldots, T$ years, where π_{it} is the probability of an outcome of interest in county i at time t.

The Bayesian hierarchical model structure for log-odds η_{it} can be specified as

$$\eta_{it} = \ln\{\pi_{it}/(1-\pi_{it})\} \qquad \text{with} \qquad \eta_{it} = \alpha_t + \theta_i + \phi_i + \delta_{it}, \tag{3.2}$$

where α_t is a temporal effect of year t, θ_i and ϕ_i represent features of county i respectively without or with spatial structure, and δ_{it} is a spatio-temporal parameter. The interaction terms are assumed to be Gaussian with precision matrix, which are specified as the Kronecker product of the structure matrices of those main effects. Specifically, two combinations are possibly depending on which of the temporal effects is assumed to interact with which of the two spatial effects. While the object of analysis is for the combined data of white male between 55 and 64 in this model, hierarchical modeling and extensions stratified by age are not described in detail.

Different from these spatial structure modeling, a log-bilinear mortality projection model, known as the Lee-Carter (LC) model (Lee and Carter, 1992), incorporated separate age-specified parameters and time-series structure parameters, which represented the trend of mortality variation over time. The LC model implied that the future mortality rate would follow the same tendency as past periods, but neglected the accidental effects and improvements of society, since only the stochastic temporal component was accountable for variation in the estimation. Taking into account the drawbacks in the LC model, Brouhns et al. (2002) suggested a hierarchical Poisson log-bilinear model. The extended model, the Poisson Lee-Carter (PLC) model, kept the basic form of mortality risk and captured more information of variability including death and exposure to risk. However, this Poisson framework had a limitation in mean–variance equality assumption. Wong et al. (2018) proposed a Poisson Log-normal Lee Carter (PLNLC) model to resolve this issue. They introduced an overdispersion term with a normal distribution to the structure of mortality rate by adding more mortality variations across units. For other extensions of log normal structure in mortality projection, readers are referred to Cairns et al. (2006), Pedroza (2006), Delwarde et al. (2007), and Hyndman and Ullah (2007).

To reflect more possible uncertainties that affect predicted values and prediction intervals,

Czado et al. (2005) implemented the Poisson LC model in the Bayesian framework by Markov chain Monte Carlo (MCMC) sampling. In the light of borrowing strength among populations, Antonio et al. (2015) further focused on simultaneously projecting the mortality rates within multiple groups by combining the common with population-specific age and time effects in the model. Some researchers also discussed the related extension of the LC model for multiple populations cases (Li and Lee, 2005; Cairns et al., 2011; Li and Hardy, 2011; Liu et al., 2020).

Inspired by Knorr-Held (2000) and Antonio et al. (2015), we propose a new Bayesian modeling to mortality projection for geographic related multiple populations. This approach takes advantage of the time series structure for prediction over time in Poisson hierarchical models and breaks through the limitation of non-stochastic spatial model. By introducing an additive spatial parameter, it not only takes account of the instinct geographic characteristics for each population, but also develops the role of overdispersion term discussed by Wong et al. (2018). Specifically, by incorporating the conditional autoregressive (CAR) model to spatial components, the geographical dependence structures relax the independent assumption of overdispersion in the PLNLC model and reveal the unmeasured interactions of variation among adjacent regions.

The remainder of this chapter is organized as follows. In Section 3.2, we present a brief description of mortality trends in every county of Japan. Section 3.3 introduces the proposed model along with the prior settings and detailed steps of a Markov chain Monte Carlo (MCMC) sampling. The real data is applied to illustrate the advantage of the proposed approach in Section 3.4. Finally, we draw a conclusion in Section 3.5.

3.2 Geographical Mortality Data

In order to demonstrate age-specific mortality change at regional level, we take into account mortality rates for representative age groups in three separate years. The adopted data are collected by the Japan Mortality Database (JMD) referring to 45 Japanese counties.

In Figure 3.1, the classified mortality rates illustrate levels of regional death in age group of 30, 50 and 70 for three given years (1975, 1990 and 2005). A general declining trend that occurred in each age group over time can be showed in these figures. Indeed, we can easily see that the first map of each row is dominated by the darkest mortality category, since almost all the counties fell in the highest intervals. On the other hand, we can clearly see the lowest mortality categories in 2005.

Spatial correlation of mortality is supported by the fact that higher death rates are concentrated in some visible cluster of counties. In this respect, it indicates the importance of spatial structure components when modeling mortality rates for sub-populations in terms of county.

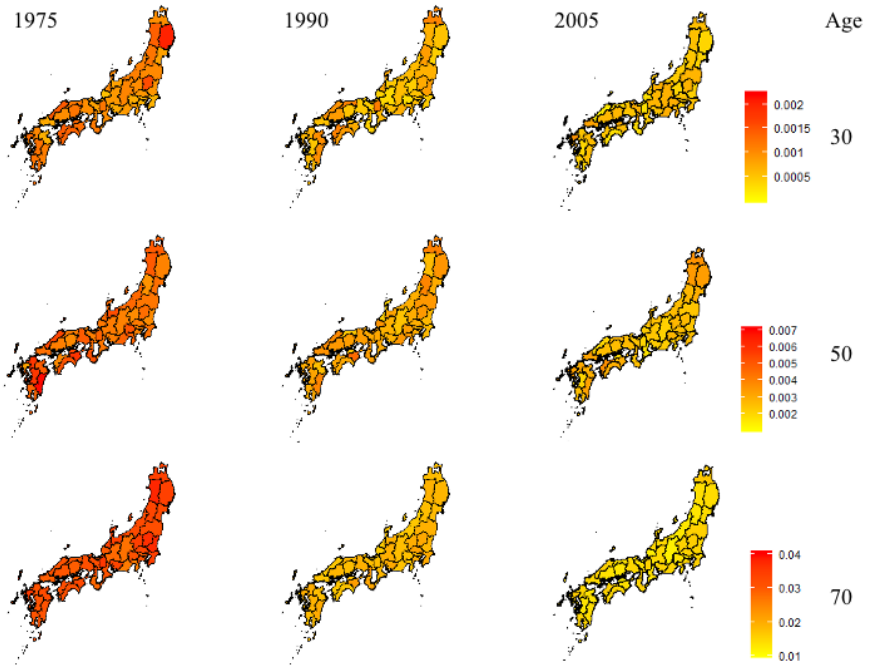


Fig. 3.1. Maps of observed mortality rates in age groups 30, 50 and 70 by Japanese county in 1975 (left), 1990 (middle) and 2005 (right).

3.3 The Proposed Model

3.3.1 Bayesian Poisson Log-bilinear Model with Spatial Structure

Let $\Theta_{\text{age}} = \{x_1, x_1+1, \ldots, x_1+M-1\} \equiv \{x_1, x_2, \ldots, x_M\}$, $\Theta_{\text{time}} = \{t_1, t_1+1, \ldots, t_1+N-1\} \equiv \{t_1, t_2, \ldots, t_N\}$ and $\Theta_{\text{space}} = \{1, 2, \ldots, p\}$ denote the sets of age, time and space considered in the training dataset, respectively. By introducing a spatial random factor θ_i, we propose a Bayesian Poisson log-bilinear with spatial structure (PLBS) model as follows,

$$D_{x,t}^{(i)} \sim \text{Poi}(E_{x,t}^{(i)} \cdot \mu_{x,t}^{(i)}), \tag{3.3}$$

$$\text{with} \quad \log \mu_{x,t}^{(i)} = \alpha_x^{(i)} + \beta_x \kappa_t + \beta_x^{(i)} \kappa_t^{(i)} + \theta_i, \tag{3.4}$$

$$\kappa_t = \varphi_1 + \varphi_2 t + \rho[\kappa_{t-1} - \varphi_1 - \varphi_2(t-1)] + \epsilon_t, \tag{3.5}$$

$$\kappa_t^{(i)} = \varphi_1^{(i)} + \varphi_2^{(i)} t + \rho^{(i)}[\kappa_{t-1}^{(i)} - \varphi_1^{(i)} - \varphi_2^{(i)}(t-1)] + \epsilon_t^{(i)}, \tag{3.6}$$

where $\epsilon_t \overset{i.i.d.}{\sim} N(0, \sigma_\kappa^2)$ and $\epsilon_t^{(i)} \overset{i.i.d.}{\sim} N(0, \sigma_{\kappa^{(i)}}^2)$; $x \in \Theta_{\text{age}}$, $t \in \Theta_{\text{time}}$ and $i \in \Theta_{\text{space}}$; the superscript (i) is the spatial index for the i-th spatial region.

Due to the existence of θ_i, $\alpha_x^{(i)}$ is approximate to the mean of log rates at age x across time in the i-th population, although the interpretation of each common or specific parameter in our PLBS model is similar to the one in Liu et al. (2020).

The equations 3.3 and 3.4 can be viewed as a generalization of PLC model (2.2) to a multi-population problem at geographical level, and the equations 3.5 and 3.6 describe the autoregressive of order one AR (1) with drift structure of κ_t and $\kappa_t^{(i)}$.

Let

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

and $\boldsymbol{\varphi} = (\varphi_1, \varphi_2)'$, $\boldsymbol{Q} = \boldsymbol{U'U}$. Note that $n+1$ dependence structures of common time effects $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \ldots, \kappa_N)' \sim N(\boldsymbol{W\varphi}, \sigma_\kappa^2 \boldsymbol{Q}^{-1})$.
Similarly, $\boldsymbol{\kappa}^{(i)} \sim N(\boldsymbol{W\varphi}^{(i)}, \sigma_{\kappa^{(i)}}^2 (\boldsymbol{Q}^{(i)})^{-1})$, where $\boldsymbol{\kappa}^{(i)} = (\kappa_1^{(i)}, \kappa_2^{(i)}, \ldots, \kappa_N^{(i)})'$. Also let

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

and $\boldsymbol{\varphi^{(i)}} = (\varphi_1^{(i)}, \varphi_2^{(i)})', \boldsymbol{Q^{(i)}} = \boldsymbol{U^{(i)'}U^{(i)}}$ for the population specific time parameter.

For model identifiability, we need to specify some constraints:

$$\sum_{x\in\Theta_{\text{age}}} \beta_x = 1, \quad \sum_{t\in\Theta_{\text{time}}} \kappa_t = 0, \quad \sum_{x\in\Theta_{\text{age}}} ||\beta_x^{(i)}||_2 = 1 \quad \text{and} \sum_{t\in\Theta_{\text{time}}} \kappa_t^{(i)} = 0. \tag{3.7}$$

Turning next to another unobserved factor of the model, it is postulated that all unobserved variations among the mortality differences for individuals in separate regions can be captured by dependencies of the spatial effect θ_i. For all sets of spatial parameters, we assume a Conditional AutoRegressive (CAR) structure, which can be formulated as a multivariate normal distribution.

As for $\boldsymbol{\theta} = \{\theta_1, \theta_2, \cdots, \theta_p\}$, assume

$$\boldsymbol{\theta} \sim N_p\left(\mathbf{0}, \sigma_\theta^2(\boldsymbol{I_p} - \lambda \boldsymbol{W_\theta})^{-1}\right), \qquad \text{where} \qquad \lambda \in (\lambda_{min}^{-1}, \lambda_{max}^{-1}) \tag{3.8}$$

and $\lambda_{min}, \lambda_{max}$ are the smallest and largest eigenvalues of $\boldsymbol{W_\theta}$,

Denote $\boldsymbol{Q_\theta} = \boldsymbol{I_p} - \lambda \boldsymbol{W_\theta}$, then

$$\boldsymbol{\theta} \sim N_p\left(\mathbf{0}, \sigma_\theta^2 \boldsymbol{Q_\theta^{-1}}\right), \tag{3.9}$$

where $\boldsymbol{W_\theta}$ is the adjacency matrix whose the generic (i, j) entry $w_{ij} = 1$ if i and j are considered as neighbors, while $w_{ij} = 0$ otherwise; and λ is the unknown parameter of interest that captures the strength of spatial dependence.

The pattern of unobserved mortality difference existing among groups in a specific region i is probably similar to that in neighboring regions. This is normalized by assuming an interaction effect vector $\boldsymbol{\theta}$, which exhibits the conditional autoregressive structure. Derived from the Brook's Lemma (Brook, 1964), the conditional distribution of θ_i is given by:

$$\theta_i \mid \theta_{-i} \sim N(\lambda \sum_{i\neq j} w_{ij}\theta_j, \sigma_\theta^2), \tag{3.10}$$

where $\theta_{-i} = \{\theta_1, \cdots, \theta_{i-1}, \theta_{i+1}, \cdots, \theta_p\}$, and

$$w_{ij} = \begin{cases} 1 & \text{if i and j are neighbors,} \\ 0 & \text{otherwise.} \end{cases} \tag{3.11}$$

We need to point out that this paradigm is first introduced by Besag (1974) for analysis of the similar properties of neighboring regions. We also refer to extensions of CAR model about Bayesian sampling contributed by Banerjee et al. (2014), Song and De Oliveira (2012) and De Oliveira (2012).

3.3.2 Prior Specification

3.3.2.1 Prior Distributions for Age Parameters

Referring to Czado et al. (2005) and Liu et al. (2020), we carry out the same transformation technique $e_x^{(i)} = \exp(\alpha_x^{(i)})$ to facilitate the computation and propose

$$e_x^{(i)} \sim \text{Gamma}(a_x^{(i)}, b_x^{(i)}), \tag{3.12}$$

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

$$\begin{aligned}
\boldsymbol{\beta} &\sim N\left(\frac{1}{M}\boldsymbol{J}_M, \sigma_\beta^2 \boldsymbol{I}_M\right), \\
\boldsymbol{\beta}^{(i)} &\sim N\left(\frac{1}{M}\boldsymbol{J}_M, \sigma_{\beta^{(i)}}^2 \boldsymbol{I}_M\right),
\end{aligned}$$

where $\boldsymbol{J}_M$ is a $M \times 1$ vector with all elements equal to 1, and $\boldsymbol{I}_M$ is an identity matrix of size M.

Moreover, we adopt conjugate priors for hyperparameters σ_β^2 and $\sigma_{\beta^{(i)}}^2$ following the inverse Gamma distributions, namely,

$$\begin{aligned}
\sigma_\beta^2 &\sim \text{InvGamma}(a_\beta, b_\beta), \\
\sigma_{\beta^{(i)}}^2 &\sim \text{InvGamma}(a_\beta^{(i)}, b_\beta^{(i)}).
\end{aligned}$$

3.3.2.2 Prior Distributions for Time Parameters

We consider the prior distributions for unknown $\boldsymbol{\varphi}$, ρ and σ_κ^2, which are treated as random variables involved in the prior distribution of κ as hyperparameters. The corresponding priors are

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

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, and 0 otherwise.
Analogously, we specify prior distributions for the population specific parameters:

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

where $\sigma_{\rho^{(i)}}^2$, $a_\kappa^{(i)}$, and $b_\kappa^{(i)}$ are pre-specified hyperparameters.

3.3.2.3 Prior Distributions for Spatial Parameters

We employ a uniform prior on the overall degree of spatial influence λ diffused over the relevant range of its value. In particular, since it is well known that $\lambda_{min} < 0$ and $\lambda_{max} > 0$ under our assumptions for the adjacent matrix $\boldsymbol{W_\theta}$, λ must lie in the interval $(\lambda_{min}^{-1}, \lambda_{max}^{-1})$. The settings of the priors corresponding to λ and σ_θ^2 are given by

$$\begin{aligned}
&\lambda \sim \text{Unif}(\lambda_{min}^{-1}, \lambda_{max}^{-1}), \\
&\sigma_\theta^2 \sim \text{InvGamma}(a_\theta, b_\theta),
\end{aligned}$$

where $\lambda_{min}, \lambda_{max}$ are the minimum and maximum eigenvalues respectively of $\boldsymbol{W_\theta}$; and
a_θ and b_θ denote constant hyperparameters.

3.3.3 Posterior Computation

In this section, we focus on the posterior sampling approaches for the PLBS model. The main strategy is a mixture of the Gibbs and Metropolis-Hasting (M-H) samplings in the MCMC procedure.

3.3.3.1 Posterior Distributions for Age Parameters

The full conditional distributions of age parameters are given by

$$\begin{aligned}\pi(e_x^{(i)}|.) &\propto \exp(-c_x^{(i)}e_x^{(i)})(e_x^{(i)})^{D_{x,\cdot}^{(i)}}\left|\frac{d}{de_x^{(i)}}g^{-1}(\alpha_x^{(i)})\right|\pi(e_x^{(i)}) \\ &\propto \exp\left[-(b_x^{(i)}+c_x^{(i)})e_x^{(i)}\right](e_x^{(i)})^{a_x^{(i)}+D_{x,\cdot}^{(i)}-1}, \qquad (3.13)\\ \pi(\beta_x|.) &\propto \prod_{i=1}^{n}\prod_{t\in\Theta_{\text{time}}}\exp\left[-E_{x,t}^{(i)}\exp(\alpha_x^{(i)}+\beta_x\kappa_t+\beta_x^{(i)}\kappa_t^{(i)}+\theta_i)\right]\times\exp(\beta_x\kappa_t D_{x,t}^{(i)}) \\ &\quad\times\exp\left[-\frac{(\beta_x-\frac{1}{M})^2}{2\sigma_\beta^2}\right], \qquad (3.14)\\ \pi(\beta_x^{(i)}|.) &\propto \prod_{t\in\Theta_{\text{time}}}\exp\left[-E_{x,t}^{(i)}\exp(\alpha_x^{(i)}+\beta_x\kappa_t+\beta_x^{(i)}\kappa_t^{(i)}+\theta_i)\right]\times\exp(\beta_x^{(i)}\kappa_t^{(i)}D_{x,t}^{(i)}) \\ &\quad\times\exp\left[-\frac{(\beta_x^{(i)}-\frac{1}{M})^2}{2\sigma_{\beta^{(i)}}^2}\right], \qquad (3.15)\end{aligned}$$

and

$$\begin{aligned}\pi(\sigma_\beta^2|.) &\propto (\sigma_\beta^2)^{-\tilde{a}_\beta-1}\exp(-\tilde{b}_\beta/\sigma_\beta^2), \qquad (3.16)\\ \pi(\sigma_{\beta^{(i)}}^2|.) &\propto (\sigma_{\beta^{(i)}}^2)^{-\tilde{a}_\beta^{(i)}-1}\exp(-\tilde{b}_\beta^{(i)}/\sigma_{\beta^{(i)}}^2), \qquad (3.17)\end{aligned}$$

where the notation "|." represents "conditional on the data G and all other parameters ", $c_x^{(i)} = \sum_{t\in\Theta_{\text{time}}}E_{x,t}^{(i)}\exp(\beta_x\kappa_t+\beta_x^{(i)}\kappa_t^{(i)}+\theta_i)$, $D_{x,\cdot}^{(i)} = \sum_{t\in\Theta_{\text{time}}}D_{x,t}^{(i)}-1$, $\tilde{a}_\beta = a_\beta+\frac{M}{2}$, $\tilde{b}_\beta = b_\beta + \frac{1}{2}\left(\boldsymbol{\beta}-\frac{1}{M}\boldsymbol{J}_M\right)'\left(\boldsymbol{\beta}-\frac{1}{M}\boldsymbol{J}_M\right)$, $\tilde{a}_\beta^{(i)} = a_\beta^{(i)}+\frac{M}{2}$, and $\tilde{b}_\beta^{(i)} = b_\beta^{(i)}+\frac{1}{2}\left(\boldsymbol{\beta}^{(i)}-\frac{1}{M}\boldsymbol{J}_M\right)'\left(\boldsymbol{\beta}^{(i)}-\frac{1}{M}\boldsymbol{J}_M\right)$.

We can directly simulate using the Gibbs sampler from the full conditional distribution of $e_x^{(i)}$ since the density of Gamma distribution is tractable in (3.13). Further, the Metropolis-Hastings sampling is applied to update β_x and $\beta_x^{(i)}$.

From (3.13), (3.16), and (3.17), we have

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

thus they can be easily updated in each iteration by the Gibbs sampling.

3.3.3.2 Posterior Distributions for Time Parameters

In this part, we discuss the sampling algorithms for common and population-specific time parameters. Let $\boldsymbol{\kappa}_{-t} = \boldsymbol{\kappa}\backslash\{\kappa_t\} = (\kappa_1, \ldots, \kappa_{t-1}, \kappa_{t+1}, \ldots, \kappa_{t_N})'$ 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

$$
\begin{aligned}
\pi(\kappa_t|.) \propto \prod_{i=1}^{n} \prod_{x\in\Theta_{\text{age}}} &\exp\left[-E_{x,t}^{(i)} \exp(\alpha_x^{(i)} + \beta_x \kappa_t + \beta_x^{(i)} \kappa_t^{(i)} + \theta_i)\right]\\
&\times \exp(\beta_x \kappa_t D_{x,t}^{(i)}) \times f(\kappa_t|\boldsymbol{\kappa}_{-t}), \qquad (3.18)\\
\pi(\boldsymbol{\varphi}|.) \propto \exp&\left[-\frac{1}{2\sigma_\kappa^2}(\boldsymbol{\varphi}'(\boldsymbol{\Sigma}^*)^{-1}\boldsymbol{\varphi} - 2(\boldsymbol{\kappa}'\boldsymbol{Q}\boldsymbol{W} + \sigma_\kappa^2 \boldsymbol{\varphi}_0' \boldsymbol{\Sigma}_0^{-1})\boldsymbol{\varphi})\right], \qquad (3.19)\\
\pi(\rho|.) \propto \exp&\left[-\frac{1}{2\sigma_\kappa^2}\left(a_\rho \rho^2 + \frac{\sigma_\kappa^2}{\sigma_\rho^2}\rho^2 - 2b_\rho \rho\right)\right] \mathbf{1}\{\rho \in (-1,1)\}, \qquad (3.20)\\
\pi(\sigma_\kappa^2|.) \propto (\sigma_\kappa^2)^{-(a_\kappa + N/2)-1} \exp&\left[-\frac{1}{\sigma_\kappa^2}\left(b_\kappa + \frac{1}{2}(\boldsymbol{\kappa} - \boldsymbol{W}\boldsymbol{\varphi})'\boldsymbol{Q}(\boldsymbol{\kappa} - \boldsymbol{W}\boldsymbol{\varphi})\right)\right], \qquad (3.21)
\end{aligned}
$$

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

$$
\begin{aligned}
f(\kappa_t \mid \boldsymbol{\kappa}_{-t}) &\propto f(\kappa_t) f(\kappa_{t+1} \mid \kappa_t)\\
&\propto \exp\left[-\frac{1}{2\sigma_\kappa^2}[(\kappa_t - \eta_t)^2 + (\kappa_{t+1} - \eta_{t+1} - \rho(\kappa_t - \eta_t))^2]\right];
\end{aligned} \qquad (3.22)
$$

when $t_1 < t < t_N$,

$$
\begin{aligned}
f(\kappa_t \mid \boldsymbol{\kappa}_{-t}) &\propto f(\kappa_{t+1} \mid \kappa_t) f(\kappa_t \mid \kappa_{t-1})\\
&\propto \exp\left[-\frac{1}{2\sigma_\kappa^2}[(\kappa_t - \eta_t - \rho(\kappa_{t-1} - \eta_{t-1}))^2 + (\kappa_{t+1} - \eta_{t+1} - \rho(\kappa_t - \eta_t))^2]\right]; \qquad (3.23)
\end{aligned}
$$

when $t = t_N$,

$$f(\kappa_t \mid \boldsymbol{\kappa}_{-t}) \propto f(\kappa_t \mid \kappa_{t-1}) \propto \exp\left[-\frac{1}{2\sigma_\kappa^2}(\kappa_t - \eta_t - \rho(\kappa_{t-1} - \eta_{t-1}))^2\right]. \tag{3.24}$$

The updated time parameter κ_t are required to be drawn using an M-H approach since its conditional posterior distribution is not of standard form.

From (3.19), (3.20), and (3.21), $\boldsymbol{\varphi}, \rho$ and σ_κ^2 are updated by

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

In a similar way, let $\boldsymbol{\kappa}_{-t}^{(i)} = \boldsymbol{\kappa}^{(i)} \backslash \{\kappa_t^{(i)}\} = (\kappa_1^{(i)}, \ldots, \kappa_{t-1}^{(i)}, \kappa_{t+1}^{(i)}, \ldots, \kappa_{t_N}^{(i)})'$ and $\eta_t^{(i)} = \varphi_1^{(i)} + \varphi_2^{(i)}t$. When $t = t_1$,

$$f(\kappa_t^{(i)} \mid \boldsymbol{\kappa}_{-t}^{(i)}) \propto \exp\left[-\frac{1}{2\sigma^2_{\kappa^{(i)}}}[(\kappa_t^{(i)} - \eta_t^{(i)})^2 + (\kappa_{t+1}^{(i)} - \eta_{t+1}^{(i)} - \rho^{(i)}(\kappa_t^{(i)} - \eta_t^{(i)}))^2]\right]; \tag{3.25}$$

when $t_1 < t < t_N$,

$$f(\kappa_t^{(i)} \mid \boldsymbol{\kappa}_{-t}^{(i)}) \propto \exp\left[-\frac{1}{2\sigma^2_{\kappa^{(i)}}}[(\kappa_t^{(i)} - \eta_t^{(i)} - \rho(\kappa_{t-1}^{(i)} - \eta_{t-1}^{(i)}))^2 + (\kappa_{t+1}^{(i)} - \eta_{t+1}^{(i)} - \rho(\kappa_t^{(i)} - \eta_t^{(i)}))^2]\right]; \tag{3.26}$$

when $t = t_N$,

$$f(\kappa_t^{(i)} \mid \boldsymbol{\kappa}_{-t}^{(i)}) \propto \exp\left[-\frac{1}{2\sigma^2_{\kappa^{(i)}}}(\kappa_t^{(i)} - \eta_t^{(i)} - \rho(\kappa_{t-1}^{(i)} - \eta_{t-1}^{(i)}))^2\right]. \tag{3.27}$$

We obtain full conditional distributions for specific parameters:

$$\begin{aligned}
\pi(\kappa_t^{(i)}|.) \propto \prod_{x \in \Theta_{\text{age}}} &\exp\left[-E_{x,t}^{(i)} \exp(\alpha_x^{(i)} + \beta_x \kappa_t + \beta_x^{(i)}\kappa_t^{(i)} + \theta_i)\right]\\
&\times \exp(\beta_x^{(i)}\kappa_t^{(i)} D_{x,t}^{(i)}) \times f(\kappa_t^{(i)}|\boldsymbol{\kappa}_{-t}^{(i)}),
\end{aligned} \tag{3.28}$$

$$\boldsymbol{\varphi}^{(i)} \mid \cdot \sim N(\boldsymbol{\Sigma}_i^*(\boldsymbol{W}'\boldsymbol{Q}^{(i)}\boldsymbol{\kappa}^{(i)} + \sigma^2_{\kappa^{(i)}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\varphi}_0), \sigma^2_{\kappa^{(i)}}\boldsymbol{\Sigma}_i^*),$$

$$\rho^{(i)} \mid \cdot \sim N\left(\frac{b_\rho^{(i)}}{a_\rho^{(i)} + \frac{\sigma^2_{\kappa^{(i)}}}{\sigma^2_{\rho^{(i)}}}}, \frac{\sigma^2_{\kappa^{(i)}}}{a_\rho^{(i)} + \frac{\sigma^2_{\kappa^{(i)}}}{\sigma^2_{\rho^{(i)}}}}\right)\mathbf{1}\left\{\rho \in (-1,1)\right\},$$

$$\sigma^2_{\kappa^{(i)}} \mid \cdot \sim \text{InvGamma}\left(a_\kappa^{(i)} + \frac{N}{2}, b_\kappa^{(i)} + \frac{1}{2}(\boldsymbol{\kappa}^{(i)} - \boldsymbol{W}\boldsymbol{\varphi}^{(i)})'\boldsymbol{Q}^{(i)}(\boldsymbol{\kappa} - \boldsymbol{W}\boldsymbol{\varphi}^{(i)})\right),$$

where $\boldsymbol{\Sigma}_i^* = (\boldsymbol{W}'\boldsymbol{Q}^{(i)}\boldsymbol{W} + \sigma^2_{\kappa^{(i)}}\boldsymbol{\Sigma}_0^{-1})^{-1}$, $a_\rho^{(i)} = \sum_{t=t_2}^{t_N}(\kappa_{t-1}^{(i)} - \eta_{t-1}^{(i)})^2$, and $b_\rho^{(i)} = \sum_{t=t_2}^{t_N}(\kappa_t^{(i)} - \eta_t^{(i)})(\kappa_{t-1}^{(i)} - \eta_{t-1}^{(i)})$.

3.3.3.3 Posterior Distributions for Spatial Parameters

Spatial interaction effects due to the varying geographic location can be well exhibited by the structure of spatial dependencies, where θ_i depends on a group of other updated spatial parameters.

In light of this, the conditional posterior distribution of θ_i is

$$\pi(\theta_i \mid \cdot) \propto \prod_x \prod_t \exp(-E_{x,t}^{(i)} \exp(\alpha_x^{(i)} + \beta_x \kappa_t + \beta_x^{(i)}\kappa_t^{(i)} + \theta_i)) \cdot \exp(\theta_i D_{x,t}^{(i)}) \times \exp(-\frac{1}{2\sigma_i^2}(\theta_i - \lambda \sum_{i \neq j} w_{ij}\theta_j)^2). \tag{3.29}$$

Also, λ and σ_θ^2 can be updated via the full conditional distributions,

$$\pi(\lambda \mid \cdot) \propto \det(\Sigma_\theta)^{-\frac{1}{2}} \exp(-\frac{1}{2}\boldsymbol{\theta}'\Sigma_\theta^{-1}\boldsymbol{\theta}) \times \mathbf{1}\{\lambda \in (\lambda_{min}^{-1}, \lambda_{max}^{-1})\},$$

$$\sigma_\theta^2 \sim \text{InvGamma}(a_\theta + \frac{p}{2}, b_\theta + \frac{1}{2}\boldsymbol{\theta}\boldsymbol{Q}_{\boldsymbol{\theta}}\boldsymbol{\theta}'),$$

where $\Sigma_\theta = \sigma_\theta^2 Q_\theta^{-1}$.

3.4 Data Analysis

3.4.1 Dataset Description

The data to illustrate our proposed method is deaths and exposures to risk in 45 counties of Japan from the Japanese Mortality Database (JMD). We consider each county as a single population, and calibrate the model on the data from 1975 to 2004 for ages 0-99 while the data from 2005 to

2016 is separated for the validation purpose. Hence, we have $i = \{1, \cdots, 45\}$, $x_1 = 0$, $x_M = 99$, $t_1 = 1975$, $t_N = 2004$, $M = 100$, and $N = 30$ in the model.

3.4.2 Initial Values of Prior Distributions

Following the prior specifications in Sections 3.3.2.1-3.3.2.3, we consider $a_x = b_x = 1$, and $a_\beta = b_\beta = 0.01$ as age-related hyperparameters. For those hyperparameters related to the spacial and time factors, they are set as $a_\theta = b_\theta = 0.1$, $\lambda = 0.2$, $a_\kappa = b_\kappa = 0.001$, $\sigma_\rho^2 = 1$, $\boldsymbol{\varphi}_0 = (0, 0)^{'}$, $\Sigma_0 = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$, respectively. It is worthy of mentioning that the pre-specified values here are similar to the ones in (Liu et al., 2020) and non-informative relative to the size ($30 \times 100 \times 45$) of our analyzing data set.

3.4.3 Model Evaluation

Through this part, we are working on performances of the PLBS model and Poisson log-normal Lee–Carter (PLNLC) model. To evaluate the performances of two models, we generate an MCMC sample of 20000 iterations with the first 10000 as burn-ins. With a common measurement, we list the squared Pearson residuals under the PLNLC model for each population given by

$$r_{x,t}^{(i)2} = \frac{(D_{x,t}^{(i)} - E_{x,t}^{(i)}\hat{\mu}_{x,t}^{(i)})^2}{E_{x,t}^{(i)}\hat{\mu}_{x,t}^{(i)}}, \tag{3.30}$$

where $\hat{\mu}_{x,t}^{(i)} = \exp(\hat{\alpha}_x^{(i)} + \hat{\beta}_x^{(i)}\hat{\kappa}_t^{(i)})$ and estiamtes $\hat{\alpha}_x^{(i)}$, $\hat{\beta}_x^{(i)}$ and $\hat{\kappa}_t^{(i)}$ are posterior medians via MCMC sampling. Here, heat maps of $r_{x,t}^{(i)2}$ are opted to visualize the lack of fit of the PLNLC model for mortality data in three Japanese counties, as depicted in Fig 3.2(a) to Fig 3.4(a).

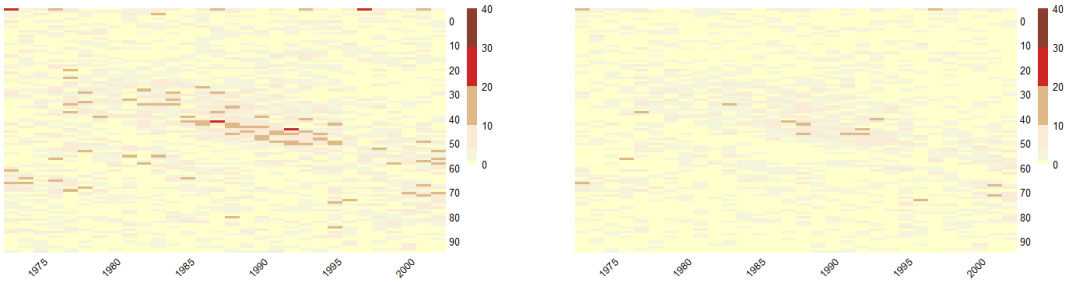
The existence of a considerable amount of dark color cells dispersing around various regions in the heat maps is one evidence showing the lack of fit in the PLNLC model with ignoring the spatial effect.

Similarly, heat maps of the squared Pearson residuals for our model can be constructed. The expression of squared residuals for our spatial model follows

$$r_{x,t}^{(i)2} = \frac{(D_{x,t}^{(i)} - E_{x,t}^{(i)}\hat{\mu}_{x,t}^{(i)})^2}{E_{x,t}^{(i)}\hat{\mu}_{x,t}^{(i)}}, \tag{3.31}$$

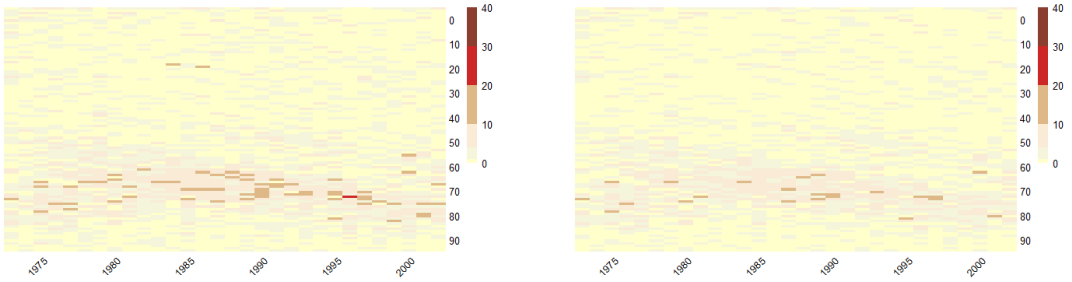
where $\hat{\mu}_{x,t}^{(i)} = \exp(\hat{\alpha}_x^{(i)} + \hat{\beta}_x\hat{\kappa}_t + \hat{\beta}_x^{(i)}\hat{\kappa}_t^{(i)} + \theta_i)$ and here the posterior medians of the parameters α_x, β_x, $\beta_x^{(i)}$, κ_t, $\kappa_t^{(i)}$and θ_i are plugged into the expression for an estimate.

Fig 3.2(b) to 3.4(b) show that there exist more lighter color cells in the heat maps of the spatial model than those in the other model, which implies an overall improved goodness of fit. In general, an appropriate explanation to overdispersion provides the fitted values with a higher flexibility and greater potential smoothing, which better serves the purpose of illustrating the mortality trend. Thus the unexplained variation are better calibrated and we can obtain a much more proper credible interval for the associated mortality prediction. On the other hand, failure to account for spatial effect typically leads to inaccurate evaluation of the variation in the model. It causes undesirable fluctuation of fitted values around the observations that prevents a reasonable estimation of the underlying process. Furthermore, lack of spatial effect also results in biased forecast since the extra variation sourced from geographic interaction is ignored.



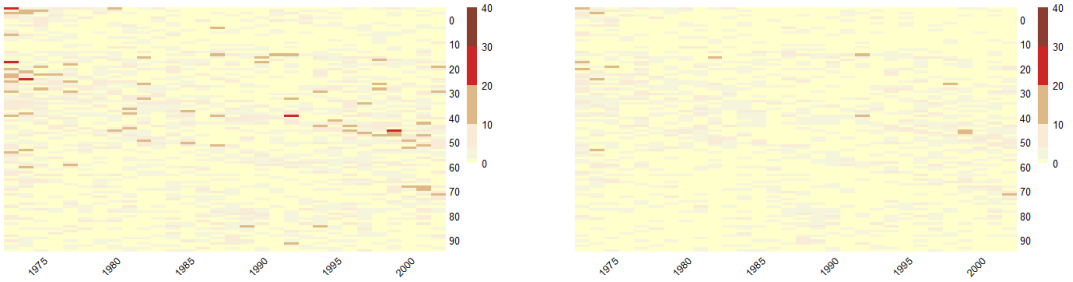
(a) (b)

Fig. 3.2. Heat map of squared Pearson residuals under the PLNLC model (top) and the PLBS model (bottom) for Miyagi.



(a) (b)

Fig. 3.3. Heat map of squared Pearson residuals under the PLNLC model (top) and the PLBS model (bottom) for Osaka.

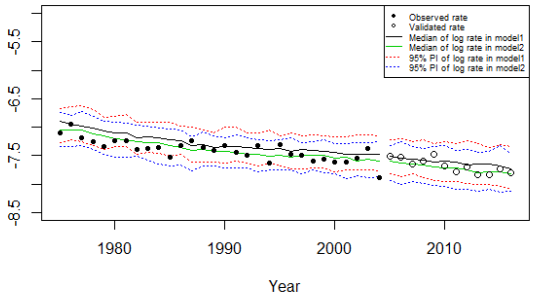


(a) (b)

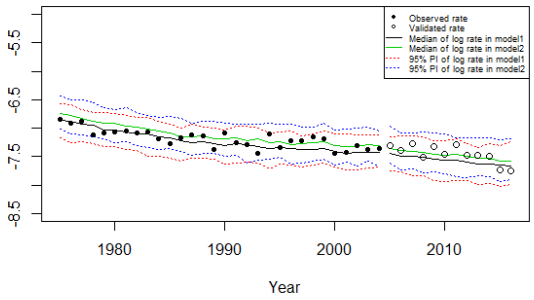
Fig. 3.4. Heat map of squared Pearson residuals under the PLNLC model (top) and the PLBS model (bottom) for Ehime.

Further, we evaluate our spatial model by comparing the variability with that of the PLNLC model. Figures 3.5 to 3.7 present the 95% HPD intervals of simulated log mortality rates at three selected ages 30, 50, and 70 for two models, respectively. In addition to the training time window (years 1975-2004), 12-year ahead projections are provided to assess the prediction ability of our spatial model. For convenience, model 1 and model 2 denote the PLNLC model and the spatial model respectively in the following discussion.

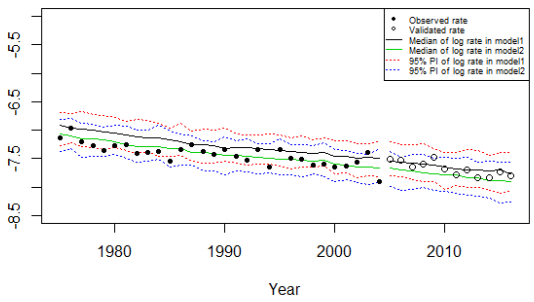
Apparently in Figures 3.5 to 3.7, almost every observed point of 1975-2004 falls in the 95% prediction intervals (blue curves). Moreover, the spatial model provides solid 12-years ahead projections because validated log rates are scattered around the projected medians within those blue curves. We also notice that the prediction intervals may properly present the variability of data by introducing additional spatial terms. Due to the more flexible dependence structure of spatial overdispersion terms, the variations among adjacent regions are absorbed into the geographic related model. The elder deaths retain more robust trend, possibly ascribe to biological components benefited from environmental factors and health systems. From the projection plots of the proposed models, it is indicated that the credible intervals provide reasonably high coverage of the observed rates across the ages without getting much wider. Hence, we can conclude that the spatial model edges the compared PLNLC model for mortality forecast in our dataset.



(a)

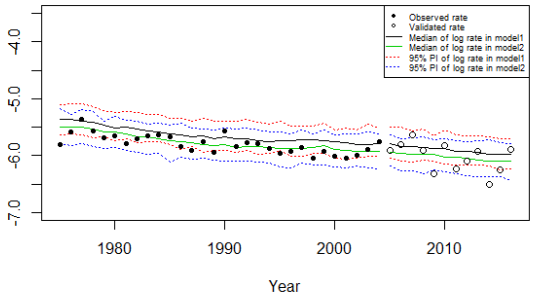


(b)

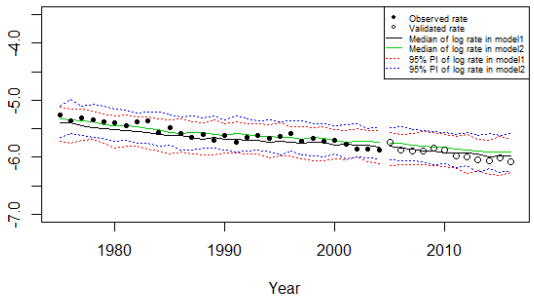


(c)

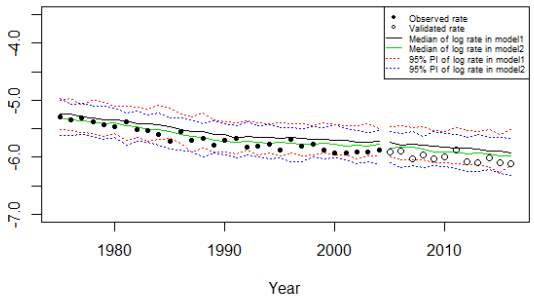
Fig. 3.5. Plots of the observed log death rates, fitted log death rates, the associated 12-years ahead projection of the log death rates and 95% HDP intervals aged 30 under two models for Miyagi (upper), Osaka (middle) and Ehime (lower).



(a)

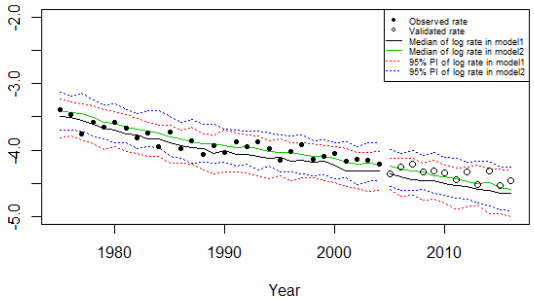


(b)

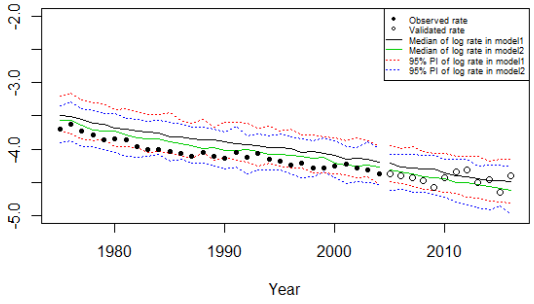


(c)

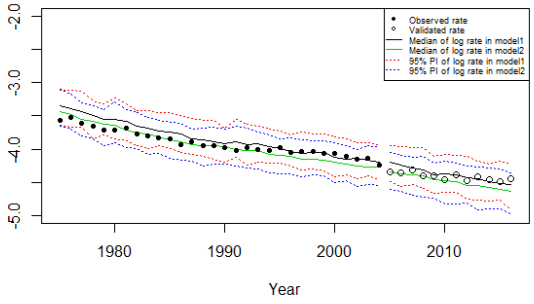
Fig. 3.6. Plots of the observed log death rates, fitted log death rates, the associated 12-years ahead projection of the log death rates and 95% HDP intervals aged 50 under two models for Miyagi (upper), Osaka (middle) and Ehime (lower).



(a)



(b)



(c)

Fig. 3.7. Plots of the observed log death rates, fitted log death rates, the associated 12-years ahead projection of the log death rates and 95% HDP intervals aged 70 under two models for Miyagi (upper), Osaka (middle) and Ehime (lower).

3.5 Conclusions

In this chapter, we apply a Bayesian spatial modeling for estimation of mortality in multiple neighbor regions. From illustrations of nationwide cases, we can observe the evidences of spatio-temporal effect at county level in Japan. The incorporation of the conditional dependence structure in the Poisson hierarchical framework enables our model to simultaneously conduct the observed spatial association and temporal correlation trends. Following this line, estimation of mortality rates can borrow more information from each other about the county-level data. As a result, the issue of large uncertainty of mortality rates via separate estimation can be mitigated at specific age groups in some areas. We demonstrate the advantage of the proposed model within one realistic study for the Japanese mortality in multiple regions from 1975 to 2016. The calibrated stochastic spatial models stick to the holdouts closely, whereas the compared PLNLC model does not capture well the ones at some calibration periods and specific ages.

In our extension of projection modeling, the number of death in each population is treated as conditional independence. It is reasonable in our proposed approach since estimated log-mortality rates are mostly in accord with the observed results. Further development of the present work can involve the dependence between mortality rates by age in the projection model. Some researchers consider cohort effect models to solve the dependence structure of death tolls among populations (Renshaw and Haberman, 2006; Booth and Tickle, 2008). Similarly, we can incorporate this paradigm to strengthen the projection power of our model.

Chapter 4

Future Work

In Chapter 2, we assume overdispersion terms $\nu_{x,t}^{(i)} \sim N(0, \sigma_i^2)$ in the equation (2.8). However, by observing the distribution of residuals against to the assumption above in specific populations, we need to reconsider the distribution of overdispersion.

One application of the Poisson distribution to incorporate overdispersion is developed by Wong et al. (2018). The prior of overdispersion term $\nu_{x,t}$ was designated as a Gamma distribution. Consequently, the structure of Poisson LC model is transformed to a negative binomial model for number of death $D_{x,t}$. One advantage of the model is computational convenience for some parameters; however, it appears a large deviation between the previous assumption and real error terms within some specific populations. In order to obtain a better description of overdispersion, we consider another direction to utilize more Bayesian approaches. Inspired by the frameworks of Efron (2014), we suggest an empirical prior for it.

For example, $\nu_{x,t}^{(i)}$ can be formulated as

$$\nu_{x,t}^{(i)} = \lambda f_a + (1 - \lambda) f_b, \tag{4.1}$$

where f_a and f_b can be some empirical distributions in the exponential family, and $\lambda \in (0, 1)$.

Specifically, it is a common empirical selection as follows:

$$f_a \sim N(0, \sigma_a^2) \quad \text{and} \quad f_b \sim N(0, \sigma_b^2) \tag{4.2}$$

$$\lambda \sim \text{Unif } (0, 1) \tag{4.3}$$

If the model is not fitted well, we can use higher degree to restructure the overdispersion terms in (4.1). Thus, the new proposed priors would be in accordance with our Bayesian hierarchical Poisson model. The subsequent update procedure can be easily derived by the MCMC algorithm.

Another direction of our work in future is exploring methodologies for resolving missing value problem. In Chapter 3, we discuss some advantages of borrowing information among adjacent areas. One favorable aspect under the spatial dependence structure is that we can potentially fill the missing values for specific regions if we have obtained or estimated the corresponding data for certain age and time groups in their neighbors. This method is appropriate for some small ranges of missing data issue; however, the estimation based on these imputations would not be accurate and effective for samples involving numerous missing values. The major drawback of this method is that time series estimates (especially variance and covariance parameters) are often biased by so many similar values in the new sample.

To overcome this issue, there are imputation methods that fit a reasonable model for the mean trend of the time series. There are several modern approaches, but we are looking forward to focusing on Kalman and sequential Kalman filters incorporating MCMC sampling procedures for future study. The Kalman recursion method is ideally suited to modeling of an Autoregressive process with missing data. The method is the evaluation of the (Gaussian) likelihood based on time series sequence $\{y_1, \cdots, y_i, \cdots, y_n\}$, where $i's$ are positive ordered integers. Note that the $y's$ can occur at irregular intervals, or equivalently for the possibility that $n-r$ observations are missing from the sequence $\{y_1, \cdots, y_i, \cdots, y_r, \cdots, y_n\}$. The solution of this problem will enable us to carry out maximum likelihood estimation for Autoregressive processes with missing values. Some researchers have contributed on the topic of mortality projection with missing value for single population under the Bayesian framework. We hope to extend those works to a wider range of interest. Specifically, we expect to explore more research of mortality projection for multiple geographically related population including missing data.

Bibliography

[1] Albert, J. H. and Chib, S. "Bayesian analysis of binary and polychotomous response data". In: *Journal of the American Statistical Association* (1993), pp. 669–679.

[2] Antonio, K., Bardoutsos, A., and Ouburg, W. "Bayesian Poisson log-bilinear models for mortality projections with multiple populations". In: *European Actuarial Journal* 5.2 (2015), pp. 245–281.

[3] Arató, N. M., Dryden, I. L., and Taylor, C. C. "Hierarchical Bayesian modelling of spatial age-dependent mortality". In: *Computational Statistics and & Data Analysis* 51.2 (2006), pp. 1347–1363.

[4] Banerjee, S., Carlin, B. P., and Gelfand, A. E. *Hierarchical modeling and analysis for spatial data*. CRC press, 2014.

[5] Bergeron-Boucher, M.-P., Simonacci, V., Oeppen, J., and Gallo, M. "Coherent modeling and forecasting of mortality patterns for subpopulations using multiway analysis of compositions: an application to Canadian provinces and territories". In: *North American Actuarial Journal* 22.1 (2018), pp. 92–118.

[6] Besag, J. "Spatial interaction and the statistical analysis of lattice systems". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 36.2 (1974), pp. 192–225.

[7] Besag, J. "Statistical analysis of non-lattice data". In: *Journal of the Royal Statistical Society: Series D (The Statistician)* 24.3 (1975), pp. 179–195.

[8] Besag, J., York, J., and Mollié, A. "Bayesian image restoration, with two applications in spatial statistics". In: *Annals of the institute of statistical mathematics* 43.1 (1991), pp. 1–20.

[9] Booth, H. and Tickle, L. "Mortality modelling and forecasting: A review of methods". In: *Annals of actuarial science* 3.1-2 (2008), pp. 3–43.

[10] Brook, D. "On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems". In: *Biometrika* 51.3/4 (1964), pp. 481–483.

[11] Brouhns, N., Denuit, M., and Vermunt, J. K. "A Poisson log-bilinear regression approach to the construction of projected lifetables". In: *Insurance: Mathematics and economics* 31.3 (2002), pp. 373–393.

[12] Cairns, A. J., Blake, D., and Dowd, K. "A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration". In: *Journal of Risk and Insurance* 73.4 (2006), pp. 687–718.

[13] Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., and Khalaf-Allah, M. "Bayesian stochastic mortality modelling for two populations". In: *ASTIN Bulletin: The Journal of the IAA* 41.1 (2011), pp. 29–59.

[14] Clayton, D. and Kaldor, J. "Empirical Bayes estimates of age-standardized relative risks for use in disease mapping". In: *Biometrics* (1987), pp. 671–681.

[15] Congdon, P. "A model for spatial variations in life expectancy; mortality in Chinese regions in 2000". In: *International journal of health geographics* 6.1 (2007), pp. 1–13.

[16] Czado, C., Delwarde, A., and Denuit, M. "Bayesian Poisson log-bilinear mortality projections". In: *Insurance: Mathematics and Economics* 36.3 (2005), pp. 260–284.

[17] De Oliveira, V. "Bayesian analysis of conditional autoregressive models". In: *Annals of the Institute of Statistical Mathematics* 64.1 (2012), pp. 107–133.

[18] Delwarde, A., Denuit, M., and Partrat, C. "Negative binomial version of the Lee-Carter model for mortality forecasting". In: *Applied Stochastic Models in Business and Industry* 23.5 (2007), pp. 385–401.

[19] Efron, B. "Two modeling strategies for empirical Bayes estimation". In: *Statistical science: a review journal of the Institute of Mathematical Statistics* 29.2 (2014), pp. 285–301.

[20] Gelman, A. "Prior distributions for variance parameters in hierarchical models". In: *Bayesian analysis* 1.3 (2006), pp. 515–534.

[21] George, E. I. and McCulloch, R. E. "Variable selection via Gibbs sampling". In: *Journal of the American Statistical Association* 88.423 (1993), pp. 881–889.

[22] Girosi, F. and King, G. "Demographic forecasting". PhD thesis. Harvard University, 2003.

[23] Greco, F. and Scalone, F. "A space-time extension of the Lee-Carter model in a hierachical Bayesian frame-work: modeling provincial morality in Italy". In: *Measuring Uncertainty in Population Forecasts: A New Approach.* (2013), pp. 412–423.

[24] Hoff, P. D. *A First Course in Bayesian Statistical Methods*. 1st. Springer Publishing Company, Incorporated, 2009. ISBN: 0387922997, 9780387922997.

[25] Hyndman, R. J. and Ullah, M. S. "Robust forecasting of mortality and fertility rates: a functional data approach". In: *Computational Statistics & Data Analysis* 51.10 (2007), pp. 4942–4956.

[26] Ishwaran, H., Rao, J. S., et al. "Spike and slab variable selection: frequentist and Bayesian strategies". In: *The Annals of Statistics* 33.2 (2005), pp. 730–773.

[27] Khana, D., Rossen, L. M., Hedegaard, H., and Warner, M. "A Bayesian spatial and temporal modeling approach to mapping geographic variation in mortality rates for subnational areas with R-INLA". In: *Journal of data science: JDS* 16.1 (2018), pp. 147–182.

[28] Knorr-Held, L. "Bayesian modelling of inseparable space-time variation in disease risk". In: *Statistics in medicine* 19.17-18 (2000), pp. 2555–2567.

[29] Krüger, F., Lerch, S., Thorarinsdottir, T., and Gneiting, T. "Predictive inference based on Markov chain Monte Carlo output". In: *International Statistical Review* (2020).

[30] Kuo, L. and Mallick, B. "Variable selection for regression models". In: *Sankhyā: The Indian Journal of Statistics, Series B* (1998), pp. 65–81.

[31] Lee, R. D. and Carter, L. R. "Modeling and forecasting US mortality". In: *Journal of the American statistical association* 87.419 (1992), pp. 659–671.

[32] LeSage, J. P. "Bayesian estimation of limited dependent variable spatial autoregressive models". In: *Geographical Analysis* 32.1 (2000), pp. 19–35.

[33] Li, J. S.-H. and Hardy, M. R. "Measuring basis risk in longevity hedges". In: *North American Actuarial Journal* 15.2 (2011), pp. 177–200.

[34] Li, N. and Lee, R. "Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method". In: *Demography* 42.3 (2005), pp. 575–594.

[35] Liu, Z., Sun, X., Liu, L., and Wang, Y.-B. "Bayesian Poisson log-normal model with regularized time structure for mortality projection of multi-population". In: *arXiv preprint arXiv:2010.04775* (2020).

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

[37] Pedroza, C. "A Bayesian forecasting model: predicting US male mortality". In: *Biostatistics* 7.4 (2006), pp. 530–550.

[38] Plat, R. "On stochastic mortality modeling". In: *Insurance: Mathematics and Economics* 45.3 (2009), pp. 393–404.

[39] Rau, R. and Schmertmann, C. "Bayesian modeling of small-area mortality with relational models schedules and spatially varying parameters". In: *2017 International Population Conference*. IUSSP. 2017.

[40] Renshaw, A. E. and Haberman, S. "A cohort-based extension to the Lee-Carter model for mortality reduction factors". In: *Insurance: Mathematics and economics* 38.3 (2006), pp. 556–570.

[41] Renshaw, A. E. and Haberman, S. "Lee-Carter mortality forecasting with age-specific enhancement". In: *Insurance: Mathematics and Economics* 33.2 (2003), pp. 255–272.

[42] Smith, T. E. and LeSage, J. P. "A Bayesian probit model with spatial dependencies". In: *Spatial and spatiotemporal econometrics*. Emerald Group Publishing Limited, 2004.

[43] Song, J. J. and De Oliveira, V. "Bayesian model selection in spatial lattice models". In: *Statistical Methodology* 9.1-2 (2012), pp. 228–238.

[44] Tuljapurkar, S., Li, N., and Boe, C. "A universal pattern of mortality decline in the G7 countries". In: *Nature* 405.6788 (2000), pp. 789–792.

[45] Waller, L. A. and Gotway, C. A. *Applied spatial statistics for public health data*. Vol. 368. John Wiley & Sons, 2004.

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