



A semiparametric Bayesian approach to generalized partial linear mixed models for longitudinal data

Nian-Sheng Tang*, Xing-De Duan

Department of Statistics, Yunnan University, Kunming 650091, PR China

ARTICLE INFO

Article history:

Received 15 September 2011

Received in revised form 14 January 2012

Accepted 27 March 2012

Available online 5 April 2012

Keywords:

Bayes factor

Dirichlet process prior

Generalized partial linear mixed models

Gibbs sampler

Metropolis–Hastings algorithm

P-spline

ABSTRACT

A generalized partial linear mixed model (GPLMM) is a natural extension of generalized linear mixed models (GLMMs) and partial linear models (PLMs). Almost all existing methods for analyzing GPLMMs are developed on the basis of the assumption that random effects are distributed as a fully parametric distribution such as normal distribution. In this paper, we extend the GPLMMs by specifying a Dirichlet process prior for a general distribution of random effects, and propose a semiparametric Bayesian approach by simultaneously utilizing an approximation truncation Dirichlet process prior of the random effects and a P -spline approximation of the smoothing function. By combining the block Gibbs sampler and the Metropolis–Hastings algorithm, a hybrid algorithm is presented for sampling observations from the posterior distribution. A procedure for selecting the degree of the polynomial components in nonparametric approximation using Bayes factor is given via path sampling. Some goodness-of-fit statistics are proposed to evaluate the plausibility of the posited model. Several simulation studies and a real example are presented to illustrate the proposed methodologies.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

A generalized partial linear mixed model (GPLMM) is a natural extension of generalized linear mixed models (GLMMs) and partial linear models (PLMs), and is widely used to analyze various correlated data including longitudinal and repeated measurement or clustered data over time by incorporating the between-subject and within-subject variations in many clinical and biomedical studies in recent years. For example, [Qin and Zhu \(2007\)](#) considered the robust estimation of parameters in GPLMMs for longitudinal data via the robust generalized estimating equations on the basis of the regression spline approximation for the nonparametric function; [Qin and Zhu \(2009\)](#) studied the robust estimation of both mean and variance components in generalized partial linear mixed models for longitudinal data based on the construction of robustified likelihood function; [Liang \(2009\)](#) proposed a method to estimate both population and individual parameters and nonparametric curves in a generalized partially linear mixed model with mismeasured covariates by combining local linear regression, penalized quasi-likelihood and local quasi-likelihood techniques and he used the proposed model and method to study the relation between binary viral load measurement and CD4 cell count in AIDS clinical trials. However, all the above mentioned studies assume that the random effects are distributed as a fully parametric distribution such as normal distribution. But, in some practical applications, the random effects may not follow a fully parametric distribution but belong to a nonparametric distribution such as the skewed and bimodal and heavy-tailed distributions. Moreover, violation of the parametric assumption on the random effects may have a large effect on statistical inference on parameters and random effects in the aforementioned models. Therefore, it is of practical importance to consider a flexible distribution

* Corresponding author. Tel.: +86 871 5032416; fax: +86 871 5033700.

E-mail address: nstang@ynu.edu.cn (N.-S. Tang).

for the random effects in GPLMMs. This can be done in a Bayesian framework by specifying a semiparametric hierarchical modeling with an approximate truncation Dirichlet process prior distribution for the random effects.

Nonparametric and semiparametric Bayesian inference for analyzing various random effects models has received considerable attention in past years because of the following reasons: (i) it allows the use of genuine prior information for obtaining better results; (ii) sampling-based Bayesian methods do not depend on asymptotic theory, and hence present more reliable results with small sample sizes (Lee and Tang, 2006a,b); (iii) relaxing the parametric distribution assumption of random effects to be a nonparametric distribution yields better robust results. The Dirichlet process (DP) (Ferguson, 1973) and DP mixtures (DPMs) are the most widely used nonparametric approaches for specifying random effects distribution or unknown distribution of random/latent variable in a Bayesian framework, due to the availability of efficient computational techniques. For example, Kleinman and Ibrahim (1998a) developed a Bayesian approach to analyze the GLMMs on the basis of the DP prior specification for the general distribution of the random effects; Kleinman and Ibrahim (1998b) presented a Bayesian approach to longitudinal random effects models by specifying a DP prior for the distribution of the random effects; Brown and Ibrahim (2003) presented a semiparametric Bayesian hierarchical model for the joint modeling of longitudinal and survival data by relaxing the distributional assumptions for the longitudinal model using DP priors on the parameters defining the longitudinal model; Dunson (2006) used dynamic mixtures of DPs to allow a latent variable distribution to change nonparametrically across groups; Guha (2008) proposed a simulation algorithm for posterior samples from a GLMM with random effects following a mixture of DP priors with normal base measure; Lee et al. (2008) provided a general Bayesian framework for structural equation models with fixed covariates by specifying a semiparametric hierarchical modeling with an approximate truncation DP prior distribution for the latent variables; Song and Lu (2010) gave a Bayesian approach to latent variable models by specifying a DP prior for the distribution of the random effects. Recently, Chow et al. (2011) proposed Bayesian estimation of semiparametric nonlinear dynamic factor analysis models by specifying a DP prior for the vector of person-specific parameters; Song et al. (2010) presented a semiparametric Bayesian approach using the truncated DP with a stick breaking prior to tackle the non-normality of residuals in the measurement equation of structural equation models. However, to the best of our knowledge, there is little work done for Bayesian analysis of GPLMMs with random effects following a DP prior. Hence, a semiparametric Bayesian approach to GPLMMs is developed by using the truncated DP with a stick breaking prior to approximate the non-normality of random effects in GPLMMs on the basis of P -spline approximation of nonparametric function and the hybrid algorithm combining the Gibbs sampler and MH algorithm in this paper.

The rest of this paper is organized as follows. Section 2 describes a GPLMM and presents a truncated Dirichlet process with a stick-breaking prior to approximate distribution of random effects and a P -spline approach to model nonparametric part and log of variance of spline coefficients. A hybrid Bayesian procedure combining the Gibbs sampler and the MH algorithm for simultaneously obtaining Bayesian estimates of unknown parameters, random effects and unknown smoothing function and their standard errors is developed in Section 3. Also, the full conditional distributions are derived, and Bayes factor and some goodness-of-fit statistics are presented in Section 3. Simulation studies and a real example are used to illustrate the proposed methodologies in Section 4. Some concluding remarks are given in Section 5. Technical details are presented in the Appendix.

2. Generalized partial linear mixed models

Consider a longitudinal study with n individuals and n_i observations over time for the i th individual ($i = 1, \dots, n$). Suppose that y_{ij} is observation of the i th individual measured at time t_{ij} ($i = 1, \dots, n, j = 1, \dots, n_i$), and y_{i1}, y_{i2}, \dots , and y_{in_i} given \mathbf{u}_i , which is a $r \times 1$ vector of random effects from the i th individual, are conditionally independent and each $y_{ij}|\mathbf{u}_i$ follows a distribution in the exponential family

$$p(y_{ij}|\mathbf{u}_i, \phi) = \exp[\phi^{-1}\{y_{ij}\theta_{ij} - b(\theta_{ij})\} + c(y_{ij}, \phi)], \quad (2.1)$$

where ϕ is a scale parameter, $c(\cdot, \cdot)$ is a function only depending on y_{ij} and ϕ , θ_{ij} is the (scalar) canonical parameter. It is easily shown that the conditional expectations and variances of y_{ij} given \mathbf{u}_i are given by $\mu_{ij} = E(y_{ij}|\mathbf{u}_i) = \dot{b}(\theta_{ij})$ and $v_{ij} = \text{var}(y_{ij}|\mathbf{u}_i) = \phi \ddot{b}(\theta_{ij})$, respectively, where $\dot{b}(\theta) = \partial b(\theta)/\partial \theta$ and $\ddot{b}(\theta) = \partial^2 b(\theta)/\partial \theta^2$. The conditional mean μ_{ij} satisfies

$$f(\mu_{ij}) \triangleq \eta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{w}_{ij}^T \mathbf{u}_i + g(t_{ij}), \quad i = 1, \dots, n, j = 1, \dots, n_i, \quad (2.2)$$

where $f(\cdot)$ is a known monotonic “link” function, \mathbf{x}_{ij} is a $p \times 1$ vector of explanatory variables, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters of the fixed effects, \mathbf{w}_{ij} is a $r \times 1$ vector of explanatory variables, $g(\cdot)$ is an unknown smooth function which is a twice continuous and differentiable function on some finite interval. Traditional assumption for \mathbf{u}_i in random effects models is that $\mathbf{u}_i \stackrel{\text{i.i.d.}}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma})$. Rather than this traditional assumption, we suppose that \mathbf{u}_i follows an unknown distribution, that is,

$$(\mathbf{u}_i|\mathcal{P}) \stackrel{\text{i.i.d.}}{\sim} \mathcal{P}, \quad \mathcal{P} \sim \mathcal{P}, i = 1, \dots, n, \quad (2.3)$$

where \mathcal{P} is a random probability with an unknown form.

Clearly, it is rather difficult to obtain Bayesian estimates of parameters in $\boldsymbol{\beta}$ based on the above specified models because of unknown forms of \mathcal{P} and $g(\cdot)$. Therefore, to make Bayesian statistical inference on parameters in $\boldsymbol{\beta}$, it is necessary to first

approximate \mathcal{P} and $g(\cdot)$ via some appropriate methods. One common approach is to approximate \mathcal{P} nonparametrically by specifying a DP prior for \mathcal{P} . That is, we let $\mathcal{P} \sim \text{DP}(\tau P_0)$, where P_0 is a base distribution that serves as a starting-point for constructing the nonparametric distribution, and τ is the weight that a researcher assigns a priori to the base distribution and it reflects the researcher's certainty of P_0 as the distribution of \mathcal{P} . The common selection for P_0 is the multivariate normal distribution. According to Sethuraman (1994), \mathcal{P} can be expressed as

$$\mathcal{P}(\cdot) = \sum_{k=1}^{\infty} \pi_k \delta_{Z_k}(\cdot) \quad \text{with } Z_k \stackrel{\text{i.i.d.}}{\sim} P_0,$$

where $\delta_{Z_k}(\cdot)$ denotes a discrete measure concentrated at Z_k , Z_k is the k th matrix consisting of possible values of \mathbf{u}_i and π_k is a random probability weight between 0 and 1 such that $0 \leq \pi_k \leq 1$ and $\sum_{k=1}^{\infty} \pi_k = 1$. For empirical estimation purposes, we consider the following truncation approximation DP for \mathcal{P} (Ishwaran and Zarepour, 2000; Lee et al., 2008):

$$\mathcal{P}(\cdot) = \mathcal{P}_G(\cdot) = \sum_{k=1}^G \pi_k \delta_{Z_k}(\cdot), \quad 1 \leq G < \infty \quad (2.4)$$

where π_k is defined by the following stick-breaking procedure:

$$\pi_1 = v_1 \quad \text{and} \quad \pi_k = v_k \prod_{l=1}^{k-1} (1 - v_l), \quad k = 2, \dots, G, \quad (2.5)$$

with $v_k \stackrel{\text{i.i.d.}}{\sim} \text{Beta}(1, \tau)$ for $k = 1, \dots, G-1$, and $v_G = 1$ so that $\sum_{k=1}^G \pi_k = 1$. In practical applications, G can be chosen empirically or is set to a large value (e.g., $G = 300$). The details for selecting the value of G can refer to Ishwaran and Zarepour (2000), Lee et al. (2008) and Chow et al. (2011). By the construction given by Sethuraman (1994), it is easily shown that $\mathcal{P}_G(\cdot)$ converges almost surely to a Dirichlet process with measure τP_0 , i.e., $\mathcal{P}_G \xrightarrow{\text{a.s.}} \text{DP}(\tau P_0)$.

It is still rather difficult and not efficient to implement the MCMC sampling based on the above hierarchical modeling. An efficient approach to sample \mathbf{u}_i via the MCMC algorithm is to represent \mathbf{u}_i in terms of a latent variable, L_i , which records each \mathbf{u}_i 's cluster membership and conveys its values such that $\mathbf{u}_i = \mathbf{Z}_{L_i}$. Let $\boldsymbol{\pi} = (\pi_1, \dots, \pi_G)$ and $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_G)$. Similar to Lee et al. (2008), the semiparametric hierarchical model defined in (2.3) and (2.4) can be reformulated as

$$(L_i | \boldsymbol{\pi}) \stackrel{\text{i.i.d.}}{\sim} \sum_{k=1}^G \pi_k \delta_k(\cdot) \quad \text{and} \quad (\boldsymbol{\pi}, \mathbf{Z}) \sim g_1(\boldsymbol{\pi}) g_2(\mathbf{Z}), \quad (2.6)$$

where $g_1(\boldsymbol{\pi})$ is given by the stick-breaking prior as defined in (2.5), and the prior distribution for \mathbf{Z}_k involved in $g_2(\mathbf{Z})$ is specified by

$$(\mathbf{Z}_k | \boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z) \stackrel{\text{i.i.d.}}{\sim} N_r(\boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z), \quad k = 1, \dots, G, \quad (2.7)$$

where $\boldsymbol{\mu}_Z = (\mu_{Z1}, \dots, \mu_{Zr})^T$, $\boldsymbol{\Psi}_Z = \text{diag}(\psi_{Z1}, \dots, \psi_{Zr})$, $N_r(\mathbf{u}, \mathbf{D})$ denotes the r -variate normal distribution with mean \mathbf{u} and covariance matrix \mathbf{D} , and r is the dimension of \mathbf{Z}_k or \mathbf{u}_i .

Following Ruppert et al. (2003) and Chen and Tang (2010), we employ the penalized splines to present a spline estimate of $g(\cdot)$, which can be approximated by the following polynomial spline:

$$g(t) = \alpha_0^{(1)} + \alpha_1^{(1)} t + \dots + \alpha_s^{(1)} t^s + \sum_{l=1}^K \alpha_l^{(2)} (t - \mathcal{K}_l)_+^s = \mathbf{B}^T(t) \boldsymbol{\alpha},$$

where s is the degree of the polynomial component, K is the number of knots (K knots define $K+1$ regression intervals because the ending points are not used as knots), $\boldsymbol{\alpha} = (\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)})^T$ is a vector of parameters in which $\boldsymbol{\alpha}^{(1)} = (\alpha_0^{(1)}, \dots, \alpha_s^{(1)})$ and $\boldsymbol{\alpha}^{(2)} = (\alpha_1^{(2)}, \dots, \alpha_K^{(2)})$, $\mathbf{B}(t) = (1, t, \dots, t^s, (t - \mathcal{K}_1)_+^s, \dots, (t - \mathcal{K}_K)_+^s)^T$ with $a_+^s = \{\max(a, 0)\}^s$, \mathcal{K}_l is the location of the l th knot which is chosen to be the $((l+1)/(K+2))$ th quantile of the unique data set $\mathbf{T} = \{t_{ij} : i = 1, \dots, n, j = 1, \dots, n_{ij}\}$. In a Bayesian framework, the amount of smoothness of the penalized splines can be controlled by the variances of coefficients according to the knots $\mathcal{K}_1, \dots, \mathcal{K}_K$. Hence, similar to Chen and Tang (2010), we assume that $\alpha_h^{(1)} \sim N(0, \sigma_0^2)$ for $h = 1, \dots, s$ and $\alpha_l^{(2)} \sim N(0, \sigma_\alpha^2(\mathcal{K}_l))$ for $l = 1, \dots, K$, i.e.,

$$\boldsymbol{\alpha} = (\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)})^T \sim N_{s+K+1}(\mathbf{0}, \boldsymbol{\Lambda}),$$

where $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2)$, $\boldsymbol{\Lambda}_1 = \text{diag}(\sigma_0^2, \dots, \sigma_0^2)$, $\boldsymbol{\Lambda}_2 = \text{diag}(\sigma_\alpha^2(\mathcal{K}_1), \dots, \sigma_\alpha^2(\mathcal{K}_K))$, σ_0^2 is a known constant and $\sigma_\alpha^2(\mathcal{K}_l)$ is the smoothing parameter (shrinkage or ridge parameter).

Following Baladandayuthapani et al. (2005), we model $\sigma_\alpha^2(\mathcal{K}_l)$ by the following hierarchical structure:

$$\begin{cases} \log\{\sigma_\alpha^2(x)\} = \delta_0^{(1)} + \delta_1^{(1)} x + \dots + \delta_q^{(1)} x^q + \sum_{k=1}^{K_\alpha} \delta_k^{(2)} (x - \mathcal{K}_k^\alpha)_+^q + v(x) \triangleq \mathbf{C}^T(x) \boldsymbol{\delta} + v(x), \\ \boldsymbol{\delta} \sim N_{q+K_\alpha+1}(\mathbf{0}, \boldsymbol{\Omega}), \quad v(x) \sim N(0, \sigma_v^2), \end{cases}$$

where $\delta = (\delta^{(1)}, \delta^{(2)})^T$, $\delta^{(1)} = (\delta_0^{(1)}, \dots, \delta_q^{(1)})$, $\delta^{(2)} = (\delta_1^{(2)}, \dots, \delta_{K_\alpha}^{(2)})$, $\mathbf{C}(x) = (1, x, \dots, x^q, (x - \mathcal{K}_1^\alpha)^q, \dots, (x - \mathcal{K}_{K_\alpha}^\alpha)^q)^T$ is a truncated power basis function, q is the spline polynomial degree, $\mathcal{K}_1^\alpha < \dots < \mathcal{K}_{K_\alpha}^\alpha$ are the spline knots often located at equal-spaced sample quantiles for simplicity, and $\mathbf{\Omega} = \text{diag}(\sigma_{0\delta}^2 \mathbf{I}_{p+1}, \gamma \mathbf{I}_{K_\alpha})$ in which γ is an unknown parameter and $\sigma_{0\delta}^2$ is a known constant for identification. Note that $v(x)$ can be viewed as an error term which is introduced to avoid the unstable sampling of MCMC via the multivariate Metropolis–Hastings algorithm and we set σ_v^2 to be some fixed known value for identification, for example, we set it to be 0.01 in simulation studies of this paper. According to our experience, the choice of σ_v^2 has little effect on convergence of MCMC algorithm. The main purpose of this paper is to develop a Bayesian approach to GPLMMs on the basis of the observed data set $\{(y_{ij}, \mathbf{x}_{ij}, \mathbf{w}_{ij}, t_{ij}) : i = 1, \dots, n, j = 1, \dots, n_i\}$.

3. Bayesian analysis of GPLMMs

Let $\mathbf{Y} = \{y_{ij} : i = 1, \dots, n, j = 1, \dots, n_i\}$, $\mathbf{X} = \{\mathbf{x}_{ij} : i = 1, \dots, n, j = 1, \dots, n_i\}$, $\mathbf{W} = \{\mathbf{w}_{ij} : i = 1, \dots, n, j = 1, \dots, n_i\}$, $\mathbf{u} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$. Motivated by recent development in statistical computing, the Gibbs sampler (Geman and Geman, 1984) is used to generate a sequence of random observations from the joint posterior distribution $p(\pi, \mathbf{Z}, \mathbf{L}, \mu_Z, \Psi_Z, \tau, \phi, \beta, \alpha, \Lambda_2, \delta, \gamma | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2)$, and then Bayesian estimates are obtained from the generated sequence of random observations. In this algorithm, observations $\{\pi, \mathbf{Z}, \mathbf{L}, \mu_Z, \Psi_Z, \tau, \phi, \beta, \alpha, \Lambda_2, \delta, \gamma\}$ are drawn iteratively from the following conditional distributions: $p(\mu_Z | \mathbf{Z}, \Psi_Z)$, $p(\Psi_Z | \mathbf{Z}, \mu_Z)$, $p(\tau | \pi)$, $p(\pi, \mathbf{Z} | \mathbf{L}, \mu_Z, \Psi_Z, \tau, \phi, \beta, \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2)$, $p(\mathbf{L} | \pi, \mathbf{Z}, \phi, \beta, \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2)$, $p(\phi | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \beta)$, $p(\beta | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \phi)$, $p(\alpha | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \phi, \beta, \Lambda_2)$, $p(\Lambda_2 | \mathbf{T}, \alpha, \delta, \sigma_v^2)$, $p(\delta | \mathbf{T}, \Lambda_2, \gamma, \sigma_v^2)$ and $p(\gamma | \delta)$.

To derive the conditional distributions for unknown parameters in $\{\mu_Z, \Psi_Z, \tau, \phi, \beta, \gamma\}$, it is necessary to specify their priors. Similar to Chow et al. (2011) and Chen and Tang (2010), we consider the following prior distributions for parameters $\mu_Z, \Psi_Z^{-1}, \tau, \phi, \beta$ and γ :

$$p(\mu_Z) \sim N_r(\mu_{Z_0}, \Psi_{\mu_Z}), \quad p(\Psi_Z^{-1} | c_1, c_2) \stackrel{\text{i.i.d.}}{\sim} \Gamma(c_1, c_2), \quad p(\tau | a_1, a_2) \sim \Gamma(a_1, a_2), \\ p(\phi) \sim \Gamma(a_\phi, b_\phi), \quad p(\beta | \phi) \sim N_p(\beta_0, \phi^{-1} \mathbf{H}_0), \quad p(\gamma) \sim \Gamma(a_\gamma, b_\gamma),$$

where $\Gamma(a, b)$ denotes the gamma distribution with parameters a and b , $\mu_{Z_0}, \Psi_{\mu_Z}, c_1, c_2, a_1, a_2, a_\phi, b_\phi, a_\gamma, b_\gamma, \beta_0$ and \mathbf{H}_0 are hyperparameters whose values are assumed to be known. As suggested by Chow et al. (2011), we set hyperparameters a_1 and a_2 in the prior distribution of τ to be 250 (or 100) and 1 in our simulation studies, which can yield large values of τ (and consequently, more unique \mathbf{u}_i values) to capture some of the more subtle individual difference in the GPLMMs. Furthermore, we set hyperparameter c_1 in the prior distribution of Ψ_Z to be 10 and allow c_2 to be sampled randomly from a $\text{Unif}(3, 7)$ distribution or a constant in our simulation studies. The conditional distributions required in implementing Gibbs sampler are presented in the Appendix. It is easily seen from Appendix that conditional distributions $p(\mu_Z | \mathbf{Z}, \Psi_Z)$, $p(\Psi_Z | \mathbf{Z}, \mu_Z)$, $p(\tau | \pi)$, $p(\mathbf{L} | \pi, \mathbf{Z}, \phi, \beta, \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2)$, $p(\delta | \mathbf{T}, \Lambda_2, \gamma, \sigma_v^2)$ and $p(\gamma | \delta)$ are some familiar distributions, such as the Gamma and normal distributions. Sampling observations from these standard distributions is straightforward and fast. But conditional distributions $p(\pi, \mathbf{Z} | \mathbf{L}, \mu_Z, \Psi_Z, \tau, \phi, \beta, \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2)$, $p(\phi | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \beta)$, $p(\beta | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \phi)$, $p(\alpha | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \phi, \beta, \Lambda_2)$ and $p(\Lambda_2 | \mathbf{T}, \alpha, \delta, \sigma_v^2)$ are some unfamiliar standard distribution and rather complicated, thus drawing observations from these conditional distributions is rather difficult. Hence, the commonly used Metropolis–Hastings (MH) algorithm is employed to sample observations from these conditional distributions. The details for using the MH algorithm to sample observations from these complicated conditional distributions are summarized in Appendix.

3.1. Bayesian estimates

Observations generated from the preceding proposed Gibbs sampler procedure together with the MH algorithm can be used to obtain the joint Bayesian estimates of parameters in $\{\phi, \beta, \gamma\}$ and smoothing function $g(\cdot)$ and random effects \mathbf{u}_i , and their corresponding standard errors.

Let $\{(\phi^{(m)}, \beta^{(m)}, \gamma^{(m)}, \mathbf{u}^{(m)}, \alpha^{(m)}, \Lambda_2^{(m)}, \delta^{(m)}, \sigma_v^{2(m)}) : m = 1, \dots, M\}$ be the generated observations of $(\phi, \beta, \gamma, \mathbf{u}, \alpha, \Lambda_2, \delta, \sigma_v^2)$ from the joint conditional distribution $p(\phi, \beta, \gamma, \mathbf{u}, \alpha, \Lambda_2, \delta, \sigma_v^2 | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T})$ via the preceding proposed hybrid algorithm. Then, Bayesian estimates of $\phi, \beta, \gamma, \mathbf{u}, \alpha, \Lambda_2, \delta$ and σ_v^2 can be obtained from the following expressions

$$\hat{\phi} = \frac{1}{M} \sum_{m=1}^M \phi^{(m)}, \quad \hat{\beta} = \frac{1}{M} \sum_{m=1}^M \beta^{(m)}, \quad \hat{\gamma} = \frac{1}{M} \sum_{m=1}^M \gamma^{(m)}, \quad \hat{\mathbf{u}} = \frac{1}{M} \sum_{m=1}^M \mathbf{u}^{(m)} \\ \hat{\alpha} = \frac{1}{M} \sum_{m=1}^M \alpha^{(m)}, \quad \hat{\Lambda}_2 = \frac{1}{M} \sum_{m=1}^M \Lambda_2^{(m)}, \quad \hat{\delta} = \frac{1}{M} \sum_{m=1}^M \delta^{(m)}, \quad \hat{\sigma}_v^2 = \frac{1}{M} \sum_{m=1}^M \sigma_v^{2(m)}.$$

Also, the smoothing function $g(t)$ can be estimated by $\hat{g}(t) = \mathbf{B}^T(t) \hat{\alpha}$. Similarly, the consistent estimates of the posterior covariance matrices $\text{var}(\phi | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T})$, $\text{var}(\beta | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T})$, $\text{var}(\gamma | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T})$ and $\text{var}(\alpha | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T})$ can be obtained via

the sample covariance matrices of observations generated. For example, $\widehat{\text{var}}(\boldsymbol{\beta}|\mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}) = \frac{1}{M-1} \sum_{m=1}^M (\boldsymbol{\beta}^{(m)} - \hat{\boldsymbol{\beta}})(\boldsymbol{\beta}^{(m)} - \hat{\boldsymbol{\beta}})^T$. Thus, standard errors for components of $\boldsymbol{\beta}$ can be obtained from the diagonal elements of the sample covariance matrix.

3.2. Bayesian model comparison

Bayes factor is an important statistic for comparing several competing models in a Bayesian framework, and it is widely used to make model comparison in various statistical models such as structural equation models (Lee and Song, 2003; Lee and Tang, 2006b) and semiparametric reproductive dispersion mixed-effects models (Chen and Tang, 2010). For some complicated models such as structural equation models, missing data models and random effects models, computing Bayes factor is a rather challenging problem (DiCiccio et al., 1997) because of intractable high-dimensional integrals involved. Hence, some novel methods such as bridge sampling (Meng and Wong, 1996) and path sampling (Gelman and Meng, 1998) have been developed to solve the above mentioned difficulties. In what follows, the path sampling method is used to calculate Bayes factor for comparing two competing models \mathcal{N}_1 and \mathcal{N}_0 :

$$B_{10} = \frac{p(\mathbf{Y}|\mathbf{X}, \mathbf{W}, \mathbf{T}, \mathcal{N}_1)}{p(\mathbf{Y}|\mathbf{X}, \mathbf{W}, \mathbf{T}, \mathcal{N}_0)},$$

where

$$p(\mathbf{Y}|\mathbf{X}, \mathbf{W}, \mathbf{T}, \mathcal{N}_k) = \int p(\mathbf{Y}, \boldsymbol{\pi}, \mathbf{Z}, \mathbf{L}, \boldsymbol{\alpha}, \boldsymbol{\Lambda}_2, \boldsymbol{\delta}|\mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2, \boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k) d\boldsymbol{\pi} d\mathbf{Z} d\mathbf{L} d\boldsymbol{\alpha} d\boldsymbol{\Lambda}_2 d\boldsymbol{\delta} d\boldsymbol{\theta}_k$$

is the marginal probability density of \mathbf{Y} given $(\mathbf{X}, \mathbf{W}, \mathbf{T})$ under \mathcal{N}_k , $p(\mathbf{Y}, \boldsymbol{\pi}, \mathbf{Z}, \mathbf{L}, \boldsymbol{\alpha}, \boldsymbol{\Lambda}_2, \boldsymbol{\delta}|\mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2, \boldsymbol{\theta}_k)$ is the joint probability density of $(\mathbf{Y}, \boldsymbol{\pi}, \mathbf{Z}, \mathbf{L}, \boldsymbol{\alpha}, \boldsymbol{\Lambda}_2, \boldsymbol{\delta})$ given $(\mathbf{X}, \mathbf{W}, \mathbf{T})$ under \mathcal{N}_k and $\boldsymbol{\theta}_k = \{\boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z, \tau, \phi, \boldsymbol{\beta}, \gamma\}$ is the parameters under \mathcal{N}_k for $k = 0$ and 1 .

Similar to Lee and Song (2003), Lee and Tang (2006b) and Chen and Tang (2010), we consider the following class of densities

$$z(h) = \int p(\mathbf{Y}, \boldsymbol{\pi}, \mathbf{Z}, \mathbf{L}, \boldsymbol{\alpha}, \boldsymbol{\Lambda}_2, \boldsymbol{\delta}|\mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2, \boldsymbol{\theta}_h, h) p(\boldsymbol{\theta}_h) d\boldsymbol{\pi} d\mathbf{Z} d\mathbf{L} d\boldsymbol{\alpha} d\boldsymbol{\Lambda}_2 d\boldsymbol{\delta} d\boldsymbol{\theta}_h,$$

where $h \in [0, 1]$ is a continuous parameter, $p(\mathbf{Y}, \boldsymbol{\pi}, \mathbf{Z}, \mathbf{L}, \boldsymbol{\alpha}, \boldsymbol{\Lambda}_2, \boldsymbol{\delta}|\mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2, \boldsymbol{\theta}_h, h)$ is the joint probability density function of $(\mathbf{Y}, \boldsymbol{\pi}, \mathbf{Z}, \mathbf{L}, \boldsymbol{\alpha}, \boldsymbol{\Lambda}_2, \boldsymbol{\delta})$ given $(\mathbf{X}, \mathbf{W}, \mathbf{T})$ under \mathcal{N}_h that links \mathcal{N}_0 and \mathcal{N}_1 with the continuous parameter h such that $\mathcal{N}_h = \mathcal{N}_0$ if $h = 0$, and $\mathcal{N}_h = \mathcal{N}_1$ if $h = 1$. Following Gelman and Meng (1998), it can be shown that

$$\log B_{10} = \log \frac{z(1)}{z(0)} = \int_0^1 E^*[H(\mathbf{Y}, \boldsymbol{\pi}, \mathbf{Z}, \mathbf{L}, \boldsymbol{\alpha}, \boldsymbol{\Lambda}_2, \boldsymbol{\delta}, \boldsymbol{\theta}_h, h|\mathbf{X}, \mathbf{W}, \mathbf{T})] dh,$$

where $H(\mathbf{Y}, \boldsymbol{\pi}, \mathbf{Z}, \mathbf{L}, \boldsymbol{\alpha}, \boldsymbol{\Lambda}_2, \boldsymbol{\delta}, \boldsymbol{\theta}_h, h|\mathbf{X}, \mathbf{W}, \mathbf{T}) = d \log p(\mathbf{Y}, \boldsymbol{\pi}, \mathbf{Z}, \mathbf{L}, \boldsymbol{\alpha}, \boldsymbol{\Lambda}_2, \boldsymbol{\delta}|\mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2, \boldsymbol{\theta}_h, h) / dh$, E^* denotes the expectation with respect to the joint conditional probability density $p(\mathbf{Y}, \boldsymbol{\pi}, \mathbf{Z}, \mathbf{L}, \boldsymbol{\alpha}, \boldsymbol{\Lambda}_2, \boldsymbol{\delta}, \boldsymbol{\theta}_h|\mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2, h)$. Then, $\log B_{10}$ can be estimated by

$$\widehat{\log B_{10}} = \frac{1}{2} \sum_{\ell=0}^h (h_{(\ell+1)} - h_{(\ell)}) (\bar{H}_{(\ell+1)} + \bar{H}_{(\ell)}),$$

where $0 = h_{(0)} < h_{(1)} < h_{(2)} < \dots < h_{(\bar{h})} < h_{(\bar{h}+1)} = 1$, $\bar{H}_{(\ell)} = M^{-1} \sum_{m=1}^M H(\mathbf{Y}, \boldsymbol{\pi}^{(m)}, \mathbf{Z}^{(m)}, \mathbf{L}^{(m)}, \boldsymbol{\alpha}^{(m)}, \boldsymbol{\Lambda}_2^{(m)}, \boldsymbol{\delta}^{(m)}, \boldsymbol{\theta}_{h_{(\ell)}}^{(m)}, h_{(\ell)}|\mathbf{X}, \mathbf{W}, \mathbf{T})$ and $\{(\boldsymbol{\pi}^{(m)}, \mathbf{Z}^{(m)}, \mathbf{L}^{(m)}, \boldsymbol{\alpha}^{(m)}, \boldsymbol{\Lambda}_2^{(m)}, \boldsymbol{\delta}^{(m)}, \boldsymbol{\theta}_{h_{(\ell)}}^{(m)}) : m = 1, \dots, M\}$ are observations generated from the joint conditional probability density $p(\mathbf{Y}, \boldsymbol{\pi}, \mathbf{Z}, \mathbf{L}, \boldsymbol{\alpha}, \boldsymbol{\Lambda}_2, \boldsymbol{\delta}, \boldsymbol{\theta}_h|\mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2, h)$ via the preceding hybrid algorithm. In the simulation studies, we take $h = 10$ and $M = 4000$ after a burn-in of 4000 iterations.

3.3. Some goodness-of-fit statistics

Gelman et al. (1996) developed a posterior predictive (PP) p -value for assessing the plausibility of the posited model in a Bayesian framework. This PP p -value was widely used in Bayesian goodness-of-fit assessment of structural equation models (see e.g., Lee and Zhu, 2000; Lee and Tang, 2006a). To use the PP p -value method to establish the goodness-of-fit assessment, we assume that the posited model defined in (2.1)–(2.4) is plausible, which is referred to as the null hypothesis H_0 . Following Gelman et al. (1996), the PP p -value for our considered model can be defined as

$$p_B = \Pr\{D(\mathbf{Y}^{\text{rep}}|\mathbf{u}, \boldsymbol{\beta}, \phi, \gamma) \geq D(\mathbf{Y}|\mathbf{u}, \boldsymbol{\beta}, \phi, \gamma)|\mathbf{X}, \mathbf{W}, \mathbf{T}, H_0\},$$

where \mathbf{Y}^{rep} denotes a replication of \mathbf{Y} and $D(\cdot|\cdot)$ is a discrepancy variable, the probability is taken with respect to the joint conditional distribution of $(\mathbf{Y}^{\text{rep}}, \mathbf{u}, \boldsymbol{\beta}, \phi, \gamma)$ given $(\mathbf{X}, \mathbf{W}, \mathbf{T}, H_0)$, which is given by

$$p(\mathbf{Y}^{\text{rep}}, \mathbf{u}, \boldsymbol{\beta}, \phi, \gamma|\mathbf{X}, \mathbf{W}, \mathbf{T}, H_0) = p(\mathbf{Y}^{\text{rep}}|\mathbf{X}, \mathbf{W}, \mathbf{T}, H_0) p(\mathbf{u}, \boldsymbol{\beta}, \phi, \gamma|\mathbf{X}, \mathbf{W}, \mathbf{T}).$$

For our considered model H_0 , we choose the following discrepancy variable

$$D(\mathbf{Y}^{\text{rep}}|\mathbf{u}, \boldsymbol{\beta}, \phi, \gamma) = \sum_{i=1}^n (\mathbf{y}_i^{\text{rep}} - \boldsymbol{\mu}_i)^T \mathbf{V}_i^{-1} (\mathbf{y}_i^{\text{rep}} - \boldsymbol{\mu}_i), \quad (3.1)$$

which is asymptotically distributed as a chi-squared distribution with degrees of freedom $N = \sum_{i=1}^n n_i$, where $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{in_i})^T$ and $\mathbf{V}_i = \text{diag}(v_{i1}, \dots, v_{in_i})$. Then, the PP p -value can be rewritten as by

$$p_B = \int \Pr(\chi^2(N) \geq D(\mathbf{Y}|\mathbf{u}, \boldsymbol{\beta}, \phi, \gamma)) p(\mathbf{u}, \boldsymbol{\beta}, \phi, \gamma | \mathbf{X}, \mathbf{W}, \mathbf{T}) d\mathbf{u} d\boldsymbol{\beta} d\phi d\gamma.$$

Evaluating the above integrals is rather difficult, so the commonly used Monte Carlo method is employed to approximate p_B , which can be approximated by

$$\hat{p}_B = M^{-1} \sum_{m=1}^M \Pr(\chi^2(N) \geq D(\mathbf{Y}|\mathbf{u}^{(m)}, \boldsymbol{\beta}^{(m)}, \phi^{(m)}, \gamma^{(m)})).$$

The posited model is plausible if \hat{p}_B is not far from 0.5, for example, within the interval (0.3, 0.7) (Lee and Tang, 2006a).

On the other hand, similar to Lee and Zhu (2000), we consider the following statistic for assessing the plausibility of model H_0 :

$$\text{SRE}_\varepsilon(p_0) = \sum_{i=1}^n I\{\boldsymbol{\varepsilon}_i^T \boldsymbol{\varepsilon}_i > \chi^2(n_i; 1 - p_0)\},$$

where $\boldsymbol{\varepsilon}_i = \mathbf{V}_i^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}_i)$, $I(\cdot)$ is an indicator function. Conditional on \mathbf{Y} and \mathbf{u} , $\text{SRE}_\varepsilon(p_0)$ is asymptotically distributed as the binomial distribution $B(n, p_0)$, which indicates that the posterior probability based on $\text{SRE}_\varepsilon(p_0)$ (Lee and Zhu, 2000) is given by

$$pp(p_0) = \Pr\{\text{SRE}_\varepsilon(p_0) > np_0 + z_\iota(np_0(1 - p_0))^{1/2} | \mathbf{Y}\},$$

where ι can be chosen to be 0.99 or 0.95 and z_ι is the upper 100% percentile of the standard normal distribution. Clearly, it is difficult to obtain the value of $pp(p_0)$ directly from the above equation. Similar to Lee and Zhu (2000), $pp(p_0)$ can be approximated by

$$\hat{pp}(p_0) = M^{-1} \sum_{m=1}^M I\{\text{SRE}_\varepsilon^{(m)}(p_0) > np_0 + z_\iota(np_0(1 - p_0))^{1/2}\},$$

where $\text{SRE}_\varepsilon^{(m)}(p_0) = \sum_{i=1}^n I\{\boldsymbol{\varepsilon}_i^{(m)T} \boldsymbol{\varepsilon}_i^{(m)} > \chi^2(n_i; 1 - p_0)\}$ and $\boldsymbol{\varepsilon}_i^{(m)} = (\mathbf{V}_i^{(m)})^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}_i^{(m)})$ are obtained via observations generated from the preceding hybrid algorithm. The posited model is plausible if $\hat{pp}(p_0)$ is not close to 1.0 (Lee and Tang, 2006a).

To solve the problem of “double use” of the data in calculating the PP p -value, Bayarri and Berger (2000) proposed the partial posterior predictive (PPP) p -value. Similar to Lee and Tang (2006b), the PPP p -value for our considered models can be defined as

$$ppp_B = \int \Pr(D(\mathbf{Y}|\mathbf{u}, \boldsymbol{\beta}, \phi, \gamma) \geq D(\mathbf{Y}^{\text{obs}}|\mathbf{u}, \boldsymbol{\beta}, \phi, \gamma)) p^*(\mathbf{u}, \boldsymbol{\beta}, \phi, \gamma) d\mathbf{u} d\boldsymbol{\beta} d\phi d\gamma, \quad (3.2)$$

where \mathbf{Y}^{obs} is the observed value of \mathbf{Y} , $p^*(\mathbf{u}, \boldsymbol{\beta}, \phi, \gamma) \propto p(\mathbf{Y}^{\text{obs}}|\mathbf{u}, \boldsymbol{\beta}, \phi, \gamma) p(\boldsymbol{\beta}, \phi, \gamma) / p(D^{\text{obs}}|\mathbf{u}, \boldsymbol{\beta}, \phi, \gamma)$, $D(\cdot|\cdot)$ is a discrepancy variable as defined in Eq. (3.1), and D^{obs} is the observed value of $D(\cdot|\cdot)$ based on \mathbf{Y}^{obs} . By the definition of $D(\cdot|\cdot)$, it is easily shown that this discrepancy variable is asymptotically distributed as the chi-squared distribution with degrees of freedom N , which indicates that $p(D^{\text{obs}}|\mathbf{u}, \boldsymbol{\beta}, \phi, \gamma)$ is asymptotically independent of $(\mathbf{u}, \boldsymbol{\beta}, \phi, \gamma)$. This fact shows that

$$p^*(\mathbf{u}, \boldsymbol{\beta}, \phi, \gamma) \propto p(\mathbf{Y}^{\text{obs}}|\mathbf{u}, \boldsymbol{\beta}, \phi, \gamma) p(\boldsymbol{\beta}, \phi, \gamma) \propto \left\{ \prod_{i=1}^n p(\mathbf{y}_i, \mathbf{u}_i | \boldsymbol{\beta}, \phi, \gamma) \right\} p(\boldsymbol{\beta}, \phi, \gamma). \quad (3.3)$$

It is easily seen from Eqs. (3.2) and (3.3) that it is rather difficult to obtain the PPP p -value because of high-dimensional integrals involved. Hence, the Monte Carlo method is used to approximate the above defined PPP p -value. Similar to Lee and Tang (2006b), the above defined PPP p -value can be approximated by

$$\widehat{ppp}_B = M^{-1} \sum_{m=1}^M I\{D(\mathbf{Y}^{(m)}|\mathbf{u}^{(m)}, \boldsymbol{\beta}^{(m)}, \phi^{(m)}, \gamma^{(m)}) \geq D(\mathbf{Y}^{\text{obs}}|\mathbf{u}^{(m)}, \boldsymbol{\beta}^{(m)}, \phi^{(m)}, \gamma^{(m)})\},$$

where $\{(\mathbf{u}^{(m)}, \boldsymbol{\beta}^{(m)}, \phi^{(m)}, \gamma^{(m)}) : m = 1, \dots, M\}$ is a sample generated from $p^*(\mathbf{u}, \boldsymbol{\beta}, \phi, \gamma)$ and $\{\mathbf{Y}^{(m)} : m = 1, \dots, M\}$ is a sample simulated from $p(\mathbf{Y}|\mathbf{u}, \boldsymbol{\beta}, \phi, \gamma)$.

4. Numerical examples

In this section, two simulation studies and a real example relating to respiratory infection in Indonesian children were used to illustrate the preceding proposed methods. Our programs were written in MATLAB language.

4.1. Simulation study

To better investigate the finite performance of the preceding proposed methodologies under known population conditions, we generated data using the GPLMMs defined in Eqs. (2.1)–(2.2) with different distributions for u_i but with the same complete sample size and different number of observations repeated as our empirical example, i.e., $n = 150$ and $n_i = 4$ or 10. A total of 50 replications was conducted for each of the simulation conditions presented as follows.

In the first simulation study, the data set $\{y_{ij} : i = 1, \dots, n, j = 1, \dots, n_i\}$ was generated from a Poisson distribution with the probability density $p(y_{ij}|u_i) = \mu_{ij}^{y_{ij}} \exp(-\mu_{ij})/y_{ij}!$ and $\eta_{ij} = \log(\mu_{ij}) = x_{1ij}\beta_1 + x_{2ij}\beta_2 + u_i + g(t_{ij})$, where $g(t) = \cos(4\pi t)$. In this simulation study, t_{ij} was randomly simulated from the uniform distribution $U[0, 1]$, covariate x_{1ij} was 0 or 1 with probability 0.5, covariate x_{2ij} was generated from the standard normal distribution $N(0, 1)$, and the true value of β was taken to be $\beta^T = (\beta_1, \beta_2) = (-1, 1)$. To test the effectiveness of using the DP prior to approximate distribution of random effects u_i , we considered the following three sets of distributional conditions (see e.g., Chow et al., 2011):

Condition 1. We assumed distribution of u_i to be $u_i \sim \text{Beta}(3, 8)$. This condition was designed to generate positively skewed distribution for the random effects u_i .

Condition 2. We took distribution of u_i to be bimodal, that is,

$$u_i \sim \Gamma(4, 10) \quad \text{for } i = 1, \dots, n/2 \quad \text{and} \quad u_i \sim \Gamma(3, 5) \quad \text{for } i = n/2 + 1, \dots, n.$$

Condition 3. We specified distribution of u_i to be $u_i \sim N(0.6, 0.05)$. This condition was used to illustrate that even when the normality assumption holds, the DP prior is still general enough to capture characteristics of the normal distribution as a special case.

To investigate sensitivity of Bayesian estimates to prior inputs, we considered the following three types of hyperparameter values for unknown parameters Ψ_{zj} , τ , β and γ :

- Typel:** The hyperparameter for β_0 was taken to be the corresponding true value, i.e. $\beta_0^T = (-1, 1)$; $H_0 = 0.25I_2$, $c_1 = 10$ and c_2 is generated from the uniform distribution $U(3, 7)$ or a constant, $a_1 = 250$ (or 100), $a_2 = 1$, $a_\gamma = 1000$ and $b_\gamma = 2$. This can be regarded as a situation with good prior information.
- Typell:** The hyperparameter for β_0 was taken to be $\beta_0^T = 1.5 \times (-1, 1)$ and $H_0 = 0.75I_2$; whilst $c_1, c_2, a_1, a_2, a_\gamma$ and b_γ were taken to be the same as those given in Type I. This can be regarded as a situation with inaccurate prior information.
- Typelll:** The hyperparameter for β_0 was taken to be $\beta_0^T = \mathbf{0}$ and $H_0 = 10I_2$; whilst $c_1, c_2, a_1, a_2, a_\gamma$ and b_γ were taken to be the same as those given in Type I. These hyperparameter values represent a situation with noninformative prior information.

To investigate sensitivity of Bayesian estimate for $g(t)$ to selection of the order of spline, we considered the following two choices of (s, K) : $(s, K) = (2, 10)$, $(3, 10)$ with $(q, K_\alpha) = (1, 3)$.

Based on the above settings and the generated data set, the preceding proposed hybrid algorithm combining the Gibbs sampler and the MH algorithm was used to evaluate the Bayesian estimates of unknown parameters and the random effects and the smoothing function based on 50 replications. For each replication and distributional condition for u_i and type of hyperparameter values, in a few test runs, three parallel chains of observations are generated from three different starting values and the EPSR values (Gelman, 1996) are computed. We observed that in all test runs, the EPSR values are less than 1.2 after 5000 iterations. Observations are collected after 5000 iterations with $M = 5000$ in producing the Bayesian estimates for each replication. In the MH algorithm, we set $\sigma_\beta^2 = 6$, $\sigma_\alpha^2 = 0.25$ and $\sigma_{u_i}^2 = 14$ in the corresponding proposal distributions, which give the approximate acceptance rates 28.4%, 31.7%, 30.6%, respectively. To investigate accuracy of estimate of function $g(t)$, we plotted the true value of function $g(t)$ against its estimate for three types of prior inputs under Conditions 1–3 in Fig. 1, which showed that estimate of $g(t)$ matched well with its true value under our considered cases. Results obtained under various different types of prior inputs via three conditions with different choices of (s, K) are reported in Table 1.

In the second simulation study, we assume that the data set $\{y_{ij} : i = 1, \dots, n, j = 1, \dots, n_i\}$ was generated from the Bernoulli distribution $B(1, p_{ij})$ with $\eta_{ij} = \log \frac{p_{ij}}{1-p_{ij}} = x_{1ij}\beta_1 + x_{2ij}\beta_2 + u_i + g(t_{ij})$, where $g(t) = \cos(4\pi t)$. In this simulation study, the true value of β was taken to be $\beta = (1, 1)^T$, n and n_i were respectively taken to be $n = 150$ and $n_i = 10$, t_{ij} was set to be $t_{ij} = \text{trun}\{5(i + (n/5 - 1))/n\}/50 + 0.1(j - 1)$, covariate x_{1ij} was taken to be -1 or 1 with probability 0.5, and covariate x_{2ij} was taken to be $x_{2ij} = 5t_{ij}^2 + e_{ij}$, where e_{ij} was independently drawn from the normal distribution $N(0, 0.5)$. But u_i was generated from the following two sets of distributional conditions:

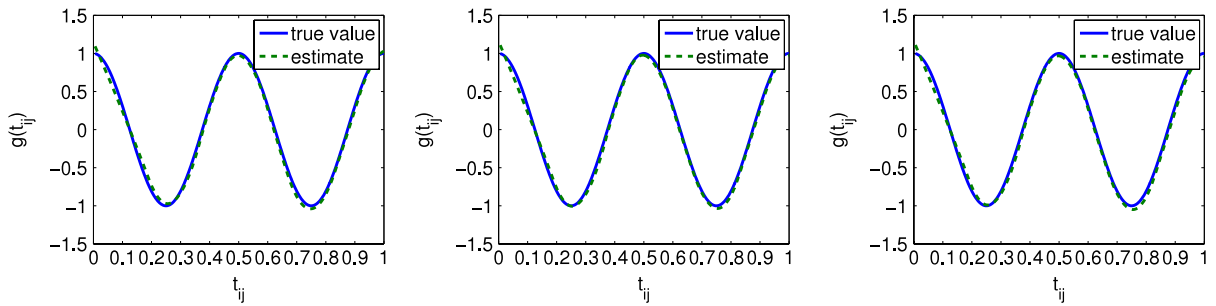


Fig. 1a. Estimate versus true value of $g(t)$ for three priors: type I (left panel), type II (middle panel) and type III (right panel) under Condition 1 in the first simulation.

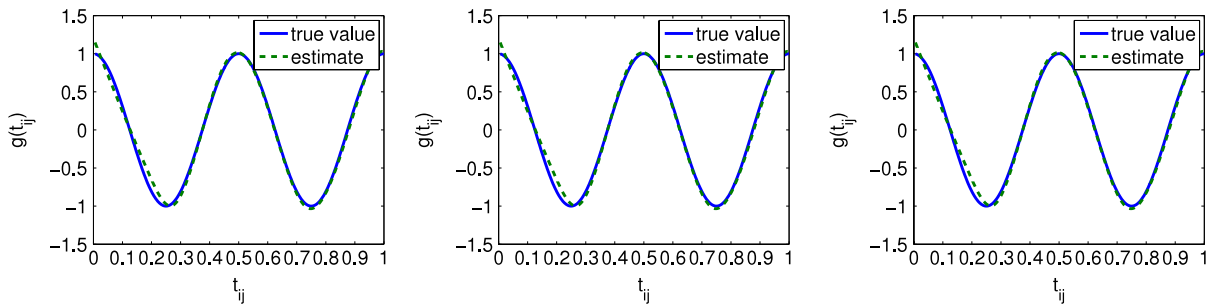


Fig. 1b. Estimate versus true value of $g(t)$ for three priors: type I (left panel), type II (middle panel) and type III (right panel) under Condition 2 in the first simulation.

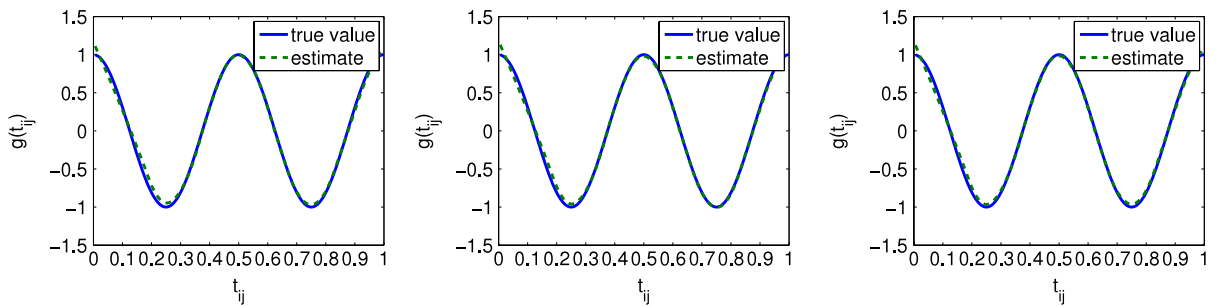


Fig. 1c. Estimate versus true value of $g(t)$ for three priors: type I (left panel), type II (middle panel) and type III (right panel) under Condition 3 in the first simulation.

Table 1
Bayesian estimates of parameters under Conditions 1–5 in the simulation study.

| | (s,K) | Par. | Type I | | Type II | | Type III | |
|-------------|---------|-----------|--------|--------|---------|--------|----------|--------|
| | | | Bias | RMS | Bias | RMS | Bias | RMS |
| Condition 1 | (2, 10) | β_1 | 0.0156 | 0.0861 | 0.0305 | 0.0787 | 0.0075 | 0.0683 |
| | | β_2 | 0.0134 | 0.0466 | 0.0082 | 0.0398 | 0.0122 | 0.0430 |
| | (3,10) | β_1 | 0.0002 | 0.0598 | 0.0312 | 0.0724 | 0.0150 | 0.0745 |
| | | β_2 | 0.0229 | 0.0503 | 0.0147 | 0.0463 | 0.0155 | 0.0387 |
| Condition 2 | (2, 10) | β_1 | 0.0006 | 0.0652 | 0.0084 | 0.0665 | 0.0023 | 0.0795 |
| | | β_2 | 0.0036 | 0.0575 | 0.0092 | 0.0552 | 0.0066 | 0.0308 |
| | (3,10) | β_1 | 0.0113 | 0.0659 | 0.0124 | 0.0762 | 0.0074 | 0.0827 |
| | | β_2 | 0.0002 | 0.0352 | 0.0059 | 0.0505 | 0.0016 | 0.0334 |
| Condition 3 | (2, 10) | β_1 | 0.0118 | 0.0658 | 0.0154 | 0.0673 | 0.0088 | 0.0586 |
| | | β_2 | 0.0030 | 0.0265 | 0.0014 | 0.0348 | 0.0028 | 0.0373 |
| | (3,10) | β_1 | 0.0166 | 0.0669 | 0.0046 | 0.0699 | 0.0079 | 0.0553 |
| | | β_2 | 0.0009 | 0.0337 | 0.0068 | 0.0329 | 0.0119 | 0.0312 |
| Condition 4 | (2,20) | β_1 | 0.0965 | 0.1133 | 0.0971 | 0.1261 | 0.0974 | 0.1081 |
| | | β_2 | 0.0044 | 0.1274 | 0.0612 | 0.1570 | 0.0118 | 0.1626 |
| Condition 5 | (2,20) | β_1 | 0.0105 | 0.0605 | 0.0062 | 0.0581 | 0.0071 | 0.0554 |
| | | β_2 | 0.0440 | 0.1742 | 0.0570 | 0.1531 | 0.0488 | 0.1728 |

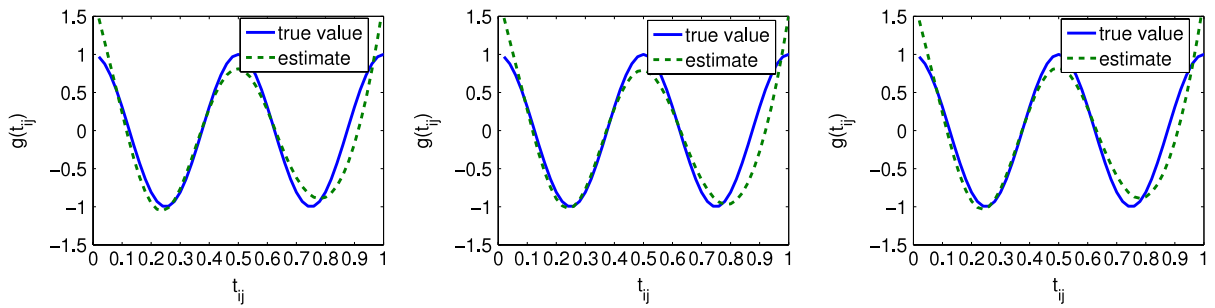


Fig. 2a. Estimate versus true value of $g(t)$ for three priors: type I (left panel), type II (middle panel) and type III (right panel) under Condition 4 in the second simulation.

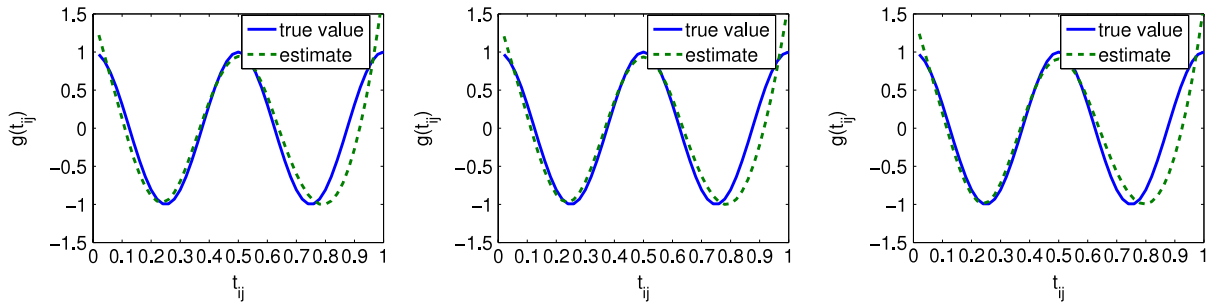


Fig. 2b. Estimate versus true value of $g(t)$ for three priors: type I (left panel), type II (middle panel) and type III (right panel) under Condition 5 in the second simulation.

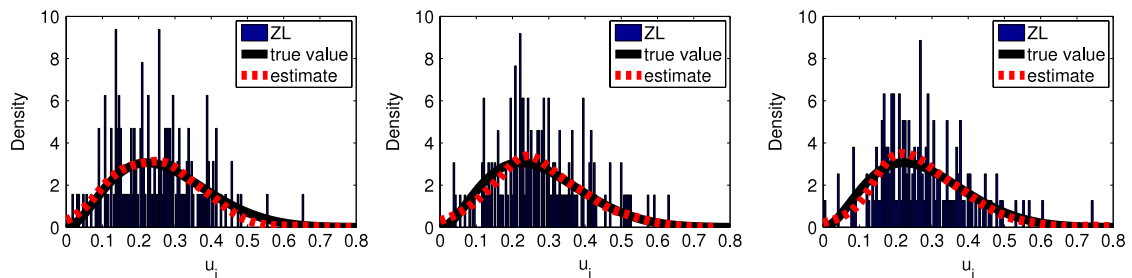


Fig. 3a. Estimated versus true density of u_i for three priors: type I (left panel), type II (middle panel) and type III (right panel) under Condition 1 in the first simulation. 'ZL' denotes the estimated value of random effects u_i .

Condition 4. We defined distribution of u_i as $u_i \sim pN(-(1-p)a, b^2) + (1-p)N(pa, b^2)$, with $p = 0.25$, $a = 0.9$ and $b^2 = 0.0481$.

Condition 5. We specified distribution of u_i to be $u_i \sim N(0, 1)$.

The preceding developed hybrid algorithm together with three types of hyperparameter values as given in the first simulation study and $(s, K) = (2, 20)$ and $(q, K_\alpha) = (1, 3)$ were used to compute Bayesian estimates of parameters in β and random effects u_i and smoothing function $g(t)$ based on 50 replications with 5000 randomly simulated observations after 5000 burn-ins for each replication. Results were presented in Table 1 and the true value of $g(t)$ against its estimate for three different types of prior inputs under Conditions 4 and 5 was displayed in Fig. 2.

In Table 1, 'Bias' denoted the absolute difference between the true value and the mean of the estimates based on 50 replications and 'RMS' was the root mean square between the estimates based on 50 replications and its true value. Examination of Table 1 showed that (i) Bayesian estimates were reasonably accurate regardless of distributional conditions of u_i and prior inputs and the values of s and K in the sense that their Bias values are less than 0.10 and their RMS values are less than 0.2; (ii) Bayesian estimates were not sensitive to prior inputs under all the considered settings; (iii) Conditions 1–3 and 5 showed comparable Bias and RMS values to Condition 4.

To compare distribution of true u_i ($i = 1, \dots, n$) with distribution of \hat{u}_i obtained from the Gibbs sampler, we computed means and standard deviations of \hat{u}_i across individuals. Results were summarized in Table 2. Examination of Table 2 showed that the means and standard deviations of true distributions for u_i can be estimated well by our proposed method under all our considered five conditions although our proposed method underestimates the empirical mean under Conditions 1–3

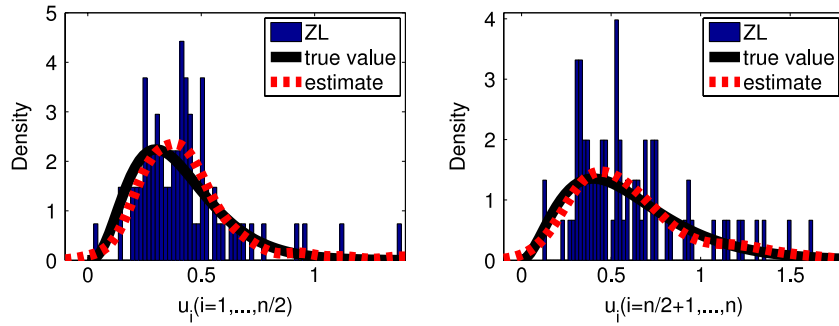


Fig. 3b. Estimated versus true density of u_i for type I prior under Condition 2 in the first simulation. 'ZL' denotes the estimated value of random effects u_i .

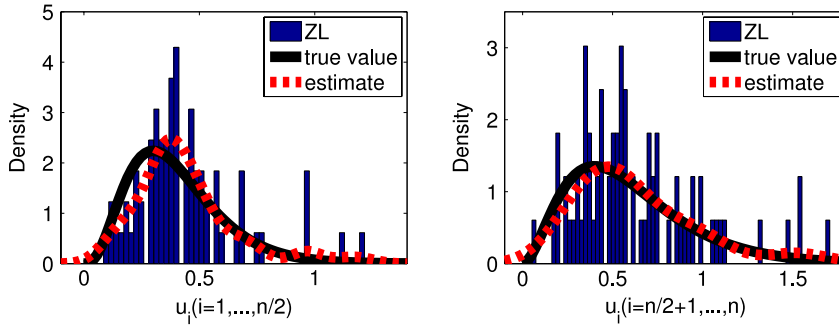


Fig. 3c. Estimated versus true density of u_i for type II under Condition 2 in the first simulation. 'ZL' denotes the estimated value of random effects u_i .

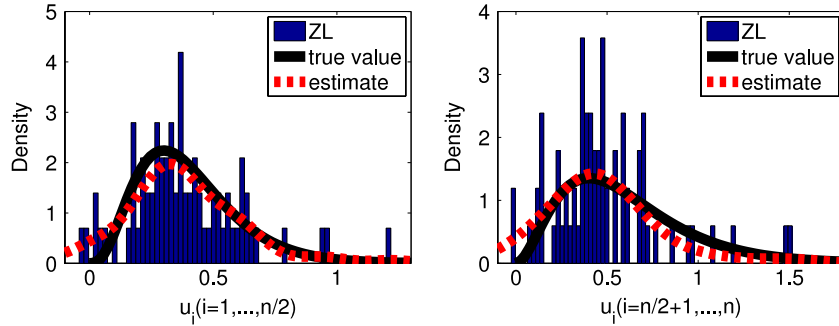


Fig. 3d. Estimated versus true density of u_i for type III under Condition 2 in the first simulation. 'ZL' denotes the estimated value of random effects u_i .

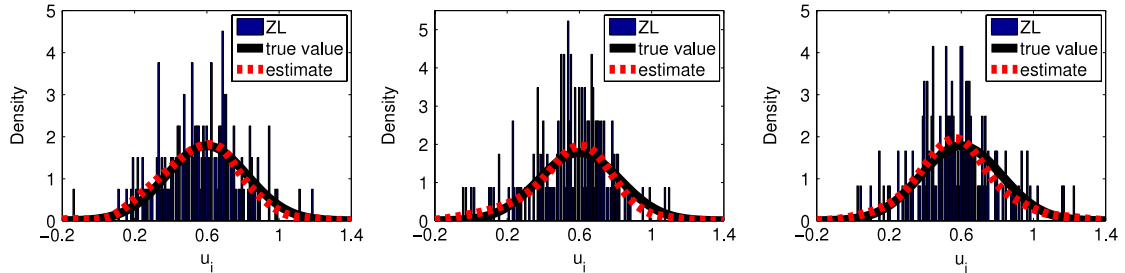


Fig. 3e. Estimated versus true density of u_i for three priors: type I (left panel), type II (middle panel) and type III (right panel) under Condition 3 in the first simulation. 'ZL' denotes the estimated value of random effects u_i .

when empirical means are positive numbers. The reason for underestimating the empirical mean when empirical means are positive numbers is that the base distribution is a normal distribution and the truncated Dirichlet process approximation is used. Fig. 3 plotted the densities of true u_i generated using Conditions 1–5 against their corresponding densities of \hat{u}_i for one randomly selected Monte Carlo run. Examination of Fig. 3 showed that (i) the conditional distributions obtained from

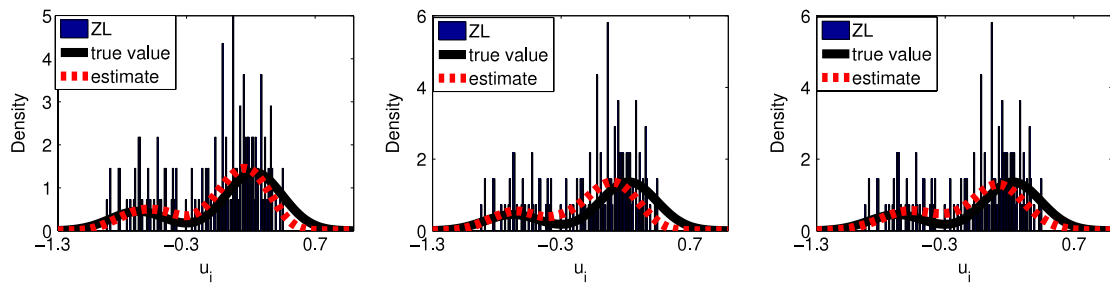


Fig. 3f. Estimated versus true density of u_i for three priors: type I (left panel), type II (middle panel) and type III (right panel) under Condition 4 in the second simulation. 'ZL' denotes the estimated value of random effects u_i .

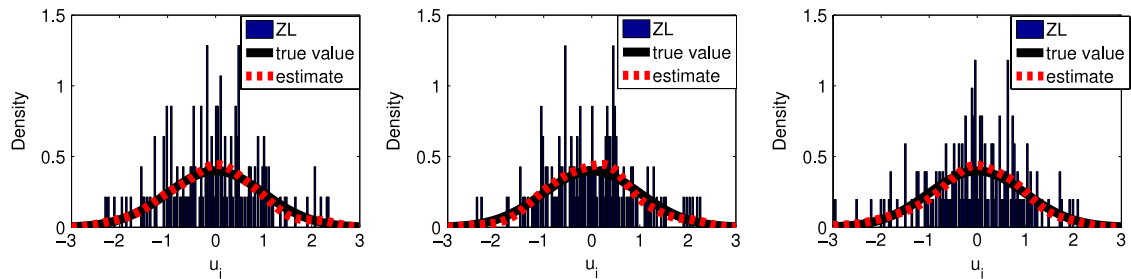


Fig. 3g. Estimated versus true density of u_i for three priors: type I (left panel), type II (middle panel) and type III (right panel) under Condition 5 in the second simulation. 'ZL' denotes the estimated value of random effects u_i .

Table 2

Means and standard deviations of u_i and \hat{u}_i in the simulation study.

| Type | (s,K) | Par. | Condition 1 | | | | Condition 2 | | | | Condition 3 | | | |
|-------------|---------|-------|-------------|----------|------|--------|-------------|----------|------|--------|-------------|----------|------|--------|
| | | | Mean | Est mean | SD | Est SD | Mean | Est mean | SD | Est SD | Mean | Est mean | SD | Est SD |
| I | (2, 10) | u_i | 0.27 | 0.27 | 0.13 | 0.11 | 0.50 | 0.47 | 0.30 | 0.27 | 0.60 | 0.55 | 0.22 | 0.21 |
| | (3, 10) | u_i | 0.27 | 0.27 | 0.13 | 0.11 | 0.50 | 0.46 | 0.30 | 0.27 | 0.60 | 0.55 | 0.22 | 0.22 |
| II | (2, 10) | u_i | 0.27 | 0.27 | 0.13 | 0.11 | 0.50 | 0.46 | 0.30 | 0.26 | 0.60 | 0.56 | 0.22 | 0.20 |
| | (3, 10) | u_i | 0.27 | 0.27 | 0.13 | 0.11 | 0.50 | 0.46 | 0.30 | 0.28 | 0.60 | 0.57 | 0.22 | 0.21 |
| III | (2, 10) | u_i | 0.27 | 0.27 | 0.13 | 0.11 | 0.50 | 0.47 | 0.30 | 0.25 | 0.60 | 0.56 | 0.22 | 0.20 |
| | (3, 10) | u_i | 0.27 | 0.27 | 0.13 | 0.11 | 0.50 | 0.46 | 0.30 | 0.28 | 0.60 | 0.55 | 0.22 | 0.22 |
| Condition 4 | | | Condition 5 | | | | | | | | | | | |
| I | (2, 20) | u_i | 0 | 0.02 | 0.45 | 0.51 | 0 | 0.00 | 1 | 0.96 | | | | |
| II | (2, 20) | u_i | 0 | 0.00 | 0.45 | 0.52 | 0 | 0.01 | 1 | 0.95 | | | | |
| III | (2, 20) | u_i | 0 | 0.06 | 0.45 | 0.52 | 0 | 0.01 | 1 | 0.94 | | | | |

Note: Mean denotes true empirical mean of the distribution; Est mean denotes mean of the posterior samples; SD denotes true empirical standard deviation of the distribution; Est SD denotes standard deviation of the posterior samples.

the truncation DP prior approximations to true distributions of u_i were flexible enough to recover the general shapes of true distribution for u_i under all our considered five conditions; (ii) the shapes of the estimated distributions for u_i were more accurately matched with those for the true distributions of u_i in Conditions 3 and 5 than in the other three conditions because the true densities (i.e., multivariate normal) of u_i were of the same form as the base distribution; (iii) in Conditions 2 and 4, the estimated densities were generally similar with the bimodal densities of the true distribution of u_i , but a slight difference was observed close to the tail areas of the two models. All the above findings showed that the preceding proposed estimation procedures can well recover the true information in a GPLMM under various considered distributional assumptions for u_i .

To illustrate the path sampling procedure in calculating Bayes factor for comparing various parametric part in a GPLMM as given in the first simulation study (i.e., $y_{ij} \sim \text{Poisson}(\mu_{ij})$), we conducted the third simulation study. In this simulation study, we considered the following two competing models:

$$\mathcal{N}_0 : \eta_{ij} = \log(\mu_{ij}) = x_{1ij}\beta_1 + u_i + g(t_{ij}),$$

$$\mathcal{N}_1 : \eta_{ij} = \log(\mu_{ij}) = x_{1ij}\beta_1 + x_{2ij}\beta_2 + u_i + g(t_{ij}).$$

Here, the data set $\{y_{ij} : i = 1, \dots, n, j = 1, \dots, n_i\}$ was generated from a Poisson distribution (e.g., $y_{ij} \sim \text{Poisson}(\mu_{ij})$) with model \mathcal{N}_1 . The true values for β_1 and β_2 were taken to be $\beta_1 = -1$ and $\beta_2 = 1$. Also, we set $g(t) = \cos(4\pi t)$, u_i was generated from the distributions as given in Conditions 1–5, x_{1ij} was 0 or 1 with probability 0.5, and x_{2ij} was generated from the standard normal distribution. Similar to Lee and Tang (2006b), \mathcal{N}_0 and \mathcal{N}_1 can be linked by $\mathcal{N}_{h01} : \eta_{ij} = \log(\mu_{ij}) = x_{1ij}\beta_1 + hx_{2ij}\beta_2 + u_i + g(t_{ij})$, where $h \in [0, 1]$. Clearly, \mathcal{N}_{h01} was equal to model \mathcal{N}_0 when $h = 0$, and \mathcal{N}_{h01} was just model

Table 3Estimated p -values of three goodness-of-fit statistics in the simulation study.

| | p -value | Type I | Type II | Type III |
|-------------|---------------------|--------|---------|----------|
| Condition 1 | \widehat{p}_B | 0.4943 | 0.4588 | 0.4240 |
| | $\widehat{pp}(p_0)$ | 0.4984 | 0.4796 | 0.5860 |
| | \widehat{ppp}_B | 0.4734 | 0.4532 | 0.4140 |
| Condition 2 | \widehat{p}_B | 0.5177 | 0.4950 | 0.5436 |
| | $\widehat{pp}(p_0)$ | 0.4918 | 0.5266 | 0.4646 |
| | \widehat{ppp}_B | 0.5026 | 0.4770 | 0.5324 |
| Condition 3 | \widehat{p}_B | 0.5043 | 0.5169 | 0.4903 |
| | $\widehat{pp}(p_0)$ | 0.5820 | 0.4758 | 0.4592 |
| | \widehat{ppp}_B | 0.5030 | 0.4954 | 0.4764 |
| Condition 4 | \widehat{p}_B | 0.5233 | 0.5607 | |
| | \widehat{ppp}_B | 0.4312 | 0.4390 | |
| Condition 5 | \widehat{p}_B | 0.5233 | 0.5606 | |
| | \widehat{ppp}_B | 0.4572 | 0.4390 | |

\mathcal{N}_1 when $h = 1$. The logarithm Bayes factors calculated by using the path sampling procedure with $(s, K) = (3, 10)$ and $(q, K_\alpha) = (1, 3)$, $(n, n_i) = (150, 4)$, and Type I prior under **Conditions 1–3** were given by $\log B_{10} = 115.56, 66.40$ and 23.80 , respectively. Following Kass and Raftery (1995), the true model \mathcal{N}_1 was selected as expected, which implied that the proposed model comparison procedure is rather effective. Similarly, the logarithm Bayes factors via the path sampling procedure with $(s, K) = (2, 20)$ and $(q, K_\alpha) = (1, 3)$, and Type I prior under **Conditions 4 and 5** for $(n, n_i) = (150, 10)$ were given by $\log B_{10} = 30.2607$ and 27.4053 , respectively, which indicated that the true model \mathcal{N}_1 was selected as expected.

Similar to Chen and Tang (2010), we here used the Bayes factor to select the degree of the polynomial component s in nonparametric approximation of $g(\cdot)$ with the fixed value $K = 10$. For this, we conducted the fourth simulation study. In this simulation study, we considered the following two regression splines approximations for smoothing function $g(t)$:

$$\mathcal{N}_{g0} : g(t) = \alpha_0^{(1)} + \alpha_1^{(1)}t + \alpha_2^{(1)}t^2 + \sum_{l=1}^{10} \alpha_l^{(2)}(t - \mathcal{K}_l)_+^2,$$

$$\mathcal{N}_{g1} : g(t) = \alpha_0^{(1)} + \alpha_1^{(1)}t + \alpha_2^{(1)}t^2 + \alpha_3^{(1)}t^3 + \sum_{l=1}^{10} \alpha_l^{(2)}(t - \mathcal{K}_l)_+^3.$$

The data set $\{y_{ij} : i = 1, \dots, n, j = 1, \dots, n_i\}$ was generated as done in the third simulation study with the same true value of parameters but different sample size, i.e., $(n, n_i) = (150, 4)$. Similarly, \mathcal{N}_{g0} and \mathcal{N}_{g1} was linked by

$$\mathcal{N}_{hg01} : g(t) = \alpha_0^{(1)} + \alpha_1^{(1)}t + \alpha_2^{(1)}t^2 + h\alpha_3^{(1)}t^3 + (1-h) \sum_{l=1}^{10} \alpha_l^{(2)}(t - \mathcal{K}_l)_+^2 + h \sum_{l=1}^{10} \alpha_l^{(2)}(t - \mathcal{K}_l)_+^3,$$

where $h \in [0, 1]$. Clearly, \mathcal{N}_{hg0} was the model \mathcal{N}_0 when $h = 0$, and \mathcal{N}_{hg1} was just the model \mathcal{N}_1 when $h = 1$. The logarithm Bayes factors calculated by the path sampling procedure with Type I prior and $(q, K_\alpha) = (1, 3)$ for **Conditions 1–3** were given by $\log B_{10} = -20.91, -8.84$ and -9.36 , respectively, which indicated that \mathcal{N}_{g0} was selected that was consistent with that given in Table 1.

To investigate the performance of three proposed goodness-of-fit statistics, i.e., the posterior predictive p -value, the partial posterior predictive p -value and the posterior probability based on $\text{SRE}_\varepsilon(p_0)$, we conducted the fifth simulation study. In this simulation study, the data set $\{(y_{ij}, x_{1ij}, x_{2ij}) : i = 1, \dots, n, j = 1, \dots, n_i\}$ was generated as done in the third simulation study with model \mathcal{N}_1 . Based on the generated data set and the above preceding MCMC algorithm, we computed $p_B, pp(p_0)$ and ppp_B values for different prior inputs. Results are displayed in Table 3. Inspecting Table 3 showed that (i) the p -values obtained from these goodness-of-fit statistics were robust to specification of distributions of random effects u_i ; (ii) they were not sensitive to prior inputs.

4.2. A real example

In this subsection, a real example relating to a subset of a cohort studied by Sommer et al. (1984) was used to illustrate the proposed methodologies. In this example, 275 preschool children in Indonesia were examined for up to six consecutive quarters for the presence of respiratory infection. The data set has even been analyzed by Diggle et al. (2002) via a logistic regression model. The response variable y was the prevalence of respiratory infection, the covariates included indicators for Vitamin A deficiency, age, height, Gender, stunting, seasonal cosine, seasonal sine. Following Lin and Zhang (1999) and Lin and Carroll (2001), we regarded age as the cluster-level covariate modeled nonparametrically with the other covariate variables modeled parametrically. Thus, we considered the following model $y_{ij} \sim \text{Bernoulli}(1, p_{ij})$ with $\text{logit}(p_{ij}) = \log\{p_{ij}/(1 - p_{ij})\} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i + g(\text{age}_{ij})$ for $i = 1, \dots, 275$ and $j = 1, \dots, n_i$ with $n_i \in \{1, 2, 3, 4, 5, 6\}$, where u_i was a random effects and was assumed to follow a unknown distribution, $\mathbf{x}_{ij} = (x_{1ij}, x_{2ij}, x_{3ij}, x_{4ij}, x_{5ij}, x_{6ij})$ in which $x_{1ij}, x_{2ij}, x_{3ij}, x_{4ij}, x_{5ij}, x_{6ij}$ were indicators for Vitamin A deficiency, seasonal cosine, seasonal sine, Gender, height,

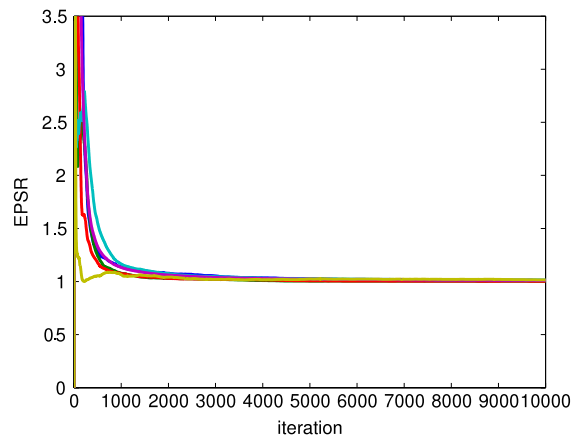


Fig. 4. EPSR values of all parameters against iterations in the real example.

stunting, respectively. The preceding proposed hybrid algorithm was used to obtain Bayesian estimates of parameters in β . For this, we took $K = 20$, that is, we considered 20 knots, and selected \mathcal{K}_l to be the sample $l/(K + 2)$ -quantile of data set $\mathbf{T} = \{\text{age}_{ij} : i = 1, \dots, 275, j = 1, \dots, n_i\}$ for $l = 1, \dots, K$. To select the appropriate regression spline approximation to smoothing function $g(t)$ in obtaining Bayesian estimates of parameters in β , we here considered the following two competing regression spline approximations to $g(t)$, that is,

$$M_0 : g(t) = \alpha_0^{(1)} + \alpha_1^{(1)}t + \alpha_2^{(1)}t^2 + \sum_{l=1}^{20} \alpha_l^{(2)}(t - \mathcal{K}_l)_+^2,$$

$$M_1 : g(t) = \alpha_0^{(1)} + \alpha_1^{(1)}t + \alpha_2^{(1)}t^2 + \alpha_3^{(1)}t^3 + \sum_{l=1}^{20} \alpha_l^{(2)}(t - \mathcal{K}_l)_+^3.$$

The path sampling procedure was used to compute the logarithm Bayes factor for comparing the above two competing models M_0 and M_1 via the following selected prior inputs. The estimates $\tilde{\beta}$ and $\tilde{\mu}_z$ obtained from model M_1 with the noninformative prior were used as prior inputs. The hyperparameter values associated with ψ_{zj} and τ and γ were taken to be $c_1 = 10$ and $c_2 = \text{Uniform}(3, 7)$, $a_1 = 250$ and $a_2 = 1$, $a_\gamma = 1000$ and $b_\gamma = 2$. In the MH algorithm, we set $\sigma_\beta^2 = 1$, $\sigma_\alpha^2 = 10$ and $\sigma_{u_i}^2 = 6$ in their corresponding proposal distributions, which gave approximate acceptance rates 26.5%, 31.0%, 33.2%, respectively. The number of grids was taken to be 10, and for each $h(\ell)$, 4000 simulated observations were used to compute $H(\ell)$ after 4000 burn-in iterations. To test the convergence of the algorithm, plot of the EPSR values for all the unknown parameters against iterations was presented in Fig. 4, which indicated that the algorithm converges about 2000 iterations because EPSR values of all unknown parameters were less than 1.2 about 2000 iterations. The estimated logarithm Bayes factor via the path sampling procedure was equal to $\log \widehat{B}_{10} = -0.1353$, which indicated that M_0 was significantly better than M_1 . That is, $S = 2$ was clearly better than $S = 3$. The posterior predictive probability corresponding to the model M_0 was equal to 0.36, which indicated that the posited model provides a reasonable fit to the considered data set. Fig. 5 displayed Bayesian estimate of nonparametric function $g(\text{age}_{ij})$, which indicated that the P -spline estimate of function $g(\text{age}_{ij})$ has a significant nonlinear trend. To compare the performance of our proposed method with those given in He et al. (2005) and Lin and Carroll (2001), we calculated Bayesian estimates (EST), standard error estimates (SD) of the unknown parameters in β . Results were given in Table 4, which showed that (i) Vitamin A, Seasonal cosine, Seasonal sine, Gender and Stunted have a strong effect on the risk of respiratory infection; (ii) Vitamin A and Stunted have a positive effect on the risk of respiratory infection, whilst Seasonal cosine, Seasonal sine, Gender and Height have a negative effect on the risk of respiratory infection; (iii) SDs of parameters obtained from our proposed method are smaller than those obtained from Lin and Carroll (2001), which is referred to as kernel method, and He et al. (2005), which is called robust estimation method.

5. Discussion

Generalized semiparametric linear mixed models (GPLMMs) are a natural extension of generalized linear mixed models (GLMMs) and nonparametric regression models (NRMs). At present, almost all existing methods in GPLMMs are developed on the basis of the assumption that the random effects are distributed as a normal distribution or the nonparametric part is approximated by kernel method (Lin and Carroll, 2001). In some practical applications, the normality assumption for the random effects may be unreasonable or kernel method involves selection of bandwidth. Hence, in this paper, to relax the unreasonable assumption for the random effects and avoid the selection of bandwidth, we develop a semiparametric Bayesian method for GPLMMs by using a truncation Dirichlet process prior to approximate distribution of random effects and the P -splines to approximate nonparametric function $g(t)$.

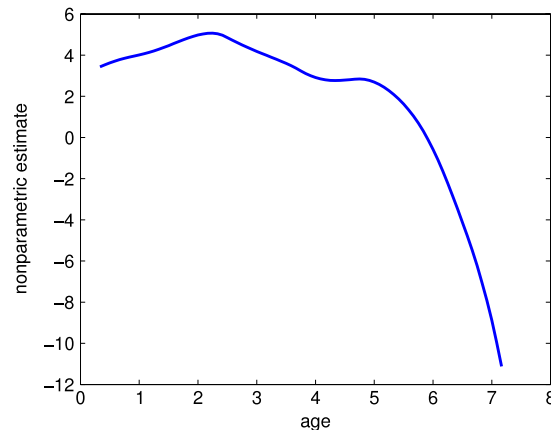


Fig. 5. Nonparametric estimate of effects of age on the prevalence of respiratory infection.

Table 4

Bayesian estimates and standard errors in the real example.

| Parameter | Bayesian method | | Robust method | | Kernel method | |
|-----------------|-----------------|-------|---------------|-------|---------------|-------|
| | EST | SD | EST | SD | EST | SD |
| Vitamin A | 0.725 | 0.478 | 0.715 | 0.367 | 0.611 | 0.529 |
| Seasonal cosine | −0.716 | 0.179 | −0.698 | 0.188 | −0.587 | 0.210 |
| Seasonal sine | −0.255 | 0.181 | −0.079 | 0.166 | −0.161 | 0.183 |
| Gender | −0.557 | 0.235 | −0.611 | 0.269 | −0.508 | 0.295 |
| Height | −0.025 | 0.025 | −0.069 | 0.039 | −0.026 | 0.035 |
| Stunted | 0.784 | 0.427 | 0.071 | 0.466 | 0.463 | 0.525 |

Although the idea using Dirichlet process prior to approximate distribution of random effects and the P -spline to approximate the nonparametric function is not new, the idea simultaneously utilizing the truncation Dirichlet process approximation to distribution of random effects and the P -spline approximation to nonparametric function to develop Bayesian inference for a random effects model, including Bayesian estimates of parameters and goodness-of-fit assessment and selection of the degree of the polynomial components in nonparametric approximation via model comparison is new. Particularly, the novel contributions of this paper include that (i) we considered a more complicated model than GLMMs and NRMs, whose Bayesian analysis has not been done, and investigated the effect of selection of prior and hyperparameters for the mean and variance components on estimates of parameters; (ii) a Bayesian approach is proposed to analyze GPLMMs, a hybrid algorithm combining the Gibbs sampler and the Metropolis–Hastings (MH) algorithm is developed to simultaneously obtain the joint Bayesian estimates of unknown parameters, the random effects and nonparametric function in the considered models on the basis of the truncated power series basis approximating nonparametric function and a nonparametric prior (i.e., a truncation Dirichlet prior) for the distribution of the random effects; (iii) the required conditional distributions, which are more complicated than those in analysis of GLMMs and NRMs, are derived and implementation of the MH algorithm for sampling observations from the complicated conditional distribution is presented; (iv) computations and applications of three p -values for addressing the goodness-of-fit of the posited model and the Bayes factor for selecting the degree of the polynomial component s in nonparametric approximation of smoothing function are given; (v) results obtained from simulation studies and the real examples show that our proposed approach cannot only produce accurate Bayesian estimates but also recover the true distribution of the random effects.

Because the Gibbs sampling method with the Dirichlet process for the random effects suffers from the following three limitations: (i) it involves sampling the infinite collection of $\mathbf{V} = \{v_1, v_2, \dots\}$ and $\pi = \{\pi_1, \pi_2, \dots\}$; (ii) the resulting Markov chain from \mathcal{P} tends to mix slowly because of the Gibbs sampler use of one-coordinate-at-a-time updates; (iii) it has the undesirable side effect that it allows inference for the posterior of \mathcal{P} to be based only on the values of the random effects \mathbf{u}_i (Ishwaran and Zarepour, 2000). To overcome the above mentioned problems for the Dirichlet process, the truncated Dirichlet process is employed to approximate \mathcal{P} because of its advantage of exact sampling for τ and directly drawing values from the posterior of \mathcal{P}_G .

Acknowledgments

The authors are grateful to the Editor, an Associate Editor and two referees for constructive suggestions that greatly improved the paper. The research was fully supported by grants from the National Natural Science Foundation of China (10961026, 11171293).

Appendix. Conditional distributions used in the Gibbs sampling procedures

To obtain Bayesian estimates of parameters in the proposed GPLMMs, the Gibbs sampler is used to generate a sequence of random observations from the joint posterior distribution $p(\pi, \mathbf{Z}, \mathbf{L}, \mu_Z, \Psi_Z, \tau, \phi, \beta, \alpha, \Lambda_2, \delta, \gamma | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2)$. The Gibbs sampler (Geman and Geman, 1984) is implemented by iteratively drawing observations from the conditional distributions: $p(\mu_Z | \mathbf{Z}, \Psi_Z)$, $p(\Psi_Z | \mathbf{Z}, \mu_Z)$, $p(\pi, \mathbf{Z} | \mathbf{L}, \mu_Z, \Psi_Z, \tau, \phi, \beta, \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2)$, $p(\mathbf{L} | \pi, \mathbf{Z}, \phi, \beta, \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2)$, $p(\phi | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \beta)$, $p(\beta | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \phi)$, $p(\alpha | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \phi, \beta, \Lambda_2)$, $p(\Lambda_2 | \mathbf{T}, \alpha, \delta, \sigma_v^2)$, $p(\delta | \mathbf{T}, \Lambda_2, \gamma, \sigma_v^2)$ and $p(\gamma | \delta)$. The conditional distributions required in implementing the above Gibbs sampler are summarized as follows.

Steps (a)–(e): *Conditional distributions related to the nonparametric components*

To sample random effects \mathbf{u}_i in terms of the latent variable L_i , $i = 1, \dots, n$, we first generate $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_G)$ from its posterior distribution and then take $\mathbf{u}_i = \mathbf{Z}_{L_i}$. Here, the base distribution is taken to be an r -variate normal distribution with mean vector μ_Z and covariance matrix Ψ_Z . Because it is rather difficult to sample observations from the posterior distribution of $(\pi, \mathbf{Z}, \mathbf{L}, \mu_Z, \Psi_Z, \tau)$, the blocked Gibbs sampler is employed to solve the above difficult. That is, Gibbs sampling of the nonparametric components was implemented via the following five steps relating to the following five conditional distributions $p(\pi, \mathbf{Z} | \mathbf{L}, \mu_Z, \Psi_Z, \tau, \phi, \beta, \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2)$, $p(\mathbf{L} | \pi, \mathbf{Z}, \phi, \beta, \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2)$, $p(\mu_Z | \mathbf{Z}, \Psi_Z)$, $p(\Psi_Z | \mathbf{Z}, \mu_Z)$ and $p(\tau | \pi)$. These five conditional distributions are summarized as follows.

Step (a). Posterior samples for $[\mu_Z | \mathbf{Z}, \Psi_Z]$ can be obtained by sampling from

$$p(\mu_Z | \mathbf{Z}, \Psi_Z) \sim N_r(\mu_\mu, \Sigma_\mu), \quad (\text{A.1})$$

where $\Sigma_\mu = (G\Psi_Z^{-1} + \Psi_Z^{-1})^{-1}$ and $\mu_\mu = \Sigma_\mu[\Psi_Z^{-1}\mu_{Z_0} + \Psi_Z^{-1}\sum_{g=1}^G \mathbf{Z}_g]$.

Step (b). For $j = 1, \dots, r$, the diagonal elements of Ψ_Z given (\mathbf{Z}, μ_Z) are distributed as

$$p(\Psi_{Z_j}^{-1} | \mathbf{Z}, \mu_Z) \sim \Gamma\left(c_1 + \frac{G}{2}, c_2 + \frac{1}{2} \sum_{g=1}^G (Z_{gj} - \mu_{Z_j})^2\right), \quad (\text{A.2})$$

where Z_{gj} is the j th element of \mathbf{Z}_g and μ_{Z_j} is the j th element of μ_Z .

Step (c). Similar to Ishwaran and Zarepour (2000), the conditional distribution $p(\tau | \pi)$ can be shown to be

$$p(\tau | \pi) \sim \Gamma\left(a_1 + G - 1, a_2 - \sum_{g=1}^G \log(1 - v_g^*)\right), \quad (\text{A.3})$$

where v_g^* is a random weight sampled from the beta distribution and is sampled with Block 4.

Step (d). As π and τ given $(\mathbf{Z}, \phi, \beta, \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2)$ are independent, the distribution $(\pi, \mathbf{Z} | \mathbf{L}, \mu_Z, \Psi_Z, \tau, \phi, \beta, \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2)$ is proportional to $p(\pi | \mathbf{L}, \tau) p(\mathbf{Z} | \mathbf{L}, \mu_Z, \Psi_Z, \phi, \beta, \alpha, \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2)$. Thus, the conditional distribution can be decomposed into two independent components.

Conditional distribution $p(\pi | \mathbf{L}, \tau)$. It can be shown that the conditional distribution $p(\pi | \mathbf{L}, \tau)$ follows a generalized Dirichlet distribution, i.e.,

$$p(\pi | \mathbf{L}, \tau) \sim \wp(a_1^*, b_1^*, \dots, a_{G-1}^*, b_{G-1}^*), \quad (\text{A.4})$$

where $a_g^* = 1 + d_g$, $b_g^* = \tau + \sum_{j=g+1}^G d_j$ for $g = 1, \dots, G-1$, and d_g is the number of L_i s (and thus individuals) whose value equals g . Sampling observations from the conditional distribution $p(\pi | \mathbf{L}, \tau)$ can be implemented by the following steps: (1) sample v_g^* from a Beta (a_g^*, b_g^*) distribution; (2) π_1, \dots, π_G are obtained with the following expression

$$\pi_1 = v_1^*, \quad \pi_G = 1 - \sum_{g=1}^{G-1} \pi_g \quad \text{and} \quad \pi_g = \prod_{j=1}^{g-1} (1 - v_j^*) v_g^* \quad \text{for } g \neq 1 \text{ or } G. \quad (\text{A.5})$$

Conditional distribution $p(\mathbf{Z} | \mathbf{L}, \mu_Z, \Psi_Z, \phi, \beta, \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2)$. Let L_1^*, \dots, L_d^* be the d unique values of $\{L_1, \dots, L_n\}$ (i.e., unique number of “clusters”), $\mathbf{Z}^L = (\mathbf{Z}_{L_1^*}^L, \dots, \mathbf{Z}_{L_d^*}^L)$, and let $\mathbf{Z}^{[L]}$ be components in $\mathbf{Z}^{[L]} = (\mathbf{Z}_1, \dots, \mathbf{Z}_G)$ other than \mathbf{Z}^L . Then

$$p(\mathbf{Z} | \mathbf{L}, \mu_Z, \Psi_Z, \phi, \beta, \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2) = p(\mathbf{Z}^{[L]} | \mu_Z, \Psi_Z) p(\mathbf{Z}^L | \mathbf{L}, \mu_Z, \Psi_Z, \phi, \beta, \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2),$$

where $p(\mathbf{Z}^{[L]} | \mu_Z, \Psi_Z)$ is the r -variate normal distribution $N_r(\mu_Z, \Psi_Z)$ and

$$p(\mathbf{Z}^L | \mathbf{L}, \mu_Z, \Psi_Z, \phi, \beta, \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2) = \prod_{k=1}^d p(\mathbf{Z}_{L_k^*}^L | \mathbf{L}, \mu_Z, \Psi_Z, \phi, \beta, \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2).$$

It can be shown that the conditional distribution $p(\mathbf{Z}_{L_k^*}^L | \mathbf{L}, \mu_Z, \Psi_Z, \phi, \beta, \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2)$ is non-standard and cannot be derived directly via Gibbs sampling. Specifically,

$$p(\mathbf{Z}_{L_k^*}^L | \mathbf{L}, \mu_Z, \Psi_Z, \phi, \beta, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2) \propto p(\mathbf{Z}_{L_k^*}^L | \mathbf{L}, \mu_Z) \prod_{\{i: L_i = L_k^*\}} [p(\mathbf{y}_i | \mathbf{u}_i = \mathbf{Z}_{L_k^*}^L, \phi, \beta, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2)],$$

which is a non-standard or familiar distribution because of exponential distribution of y_{ij} . So an MH algorithm is employed to sample observation $\mathbf{Z}_{L_k}^*$. At the m th iteration with a current value $\mathbf{Z}_{L_k}^{(m)}$, a new candidate $\mathbf{Z}_{L_k}^*$ is generated from the normal distribution $N(\mathbf{Z}_{L_k}^{(m)}, \sigma_u^2 \boldsymbol{\Omega}_u)$, where $\boldsymbol{\Omega}_u = (\boldsymbol{\Psi}_Z^{-1} + \sum_{\{i:L_i=L_k^*\}} \sum_{j=1}^{n_i} \boldsymbol{\Delta}_{uij})$ and $\boldsymbol{\Delta}_{uij} = \partial^2 b(\theta_{ij}) / \partial \mathbf{u}_i \partial \mathbf{u}_i^T |_{\mathbf{u}_i=\mathbf{Z}_{L_k}^{(m)}}$. The new $\mathbf{Z}_{L_k}^*$ is accepted with probability

$$\min \left\{ 1, \frac{p(\mathbf{Z}_{L_k}^* | \boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z) \prod_{\{i:L_i=L_k^*\}} \prod_{j=1}^{n_i} p(y_{ij} | \mathbf{u}_i = \mathbf{Z}_{L_k}^*, \phi, \boldsymbol{\beta})}{p(\mathbf{Z}_{L_k}^{(m)} | \boldsymbol{\mu}_Z, \boldsymbol{\Psi}_Z) \prod_{\{i:L_i=L_k^*\}} \prod_{j=1}^{n_i} p(y_{ij} | \mathbf{u}_i = \mathbf{Z}_{L_k}^{(m)}, \phi, \boldsymbol{\beta})} \right\}. \quad (\text{A.6})$$

The variance σ_u^2 can be chosen such that the average acceptance rate is approximately 0.25 or more.

Step (e). The conditional distribution $p(L_i | \boldsymbol{\pi}, \mathbf{Z}, \phi, \boldsymbol{\beta}, \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2)$ is given by

$$p(L_i | \boldsymbol{\pi}, \mathbf{Z}, \phi, \boldsymbol{\beta}, \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2) \stackrel{\text{i.i.d.}}{\sim} \text{Multinomial}(\pi_{i1}^*, \dots, \pi_{iG}^*), \quad (\text{A.7})$$

where π_{ik}^* is proportional to $\pi_k p(\mathbf{y}_i | \mathbf{u}_i = \mathbf{Z}_{L_k}^*, \phi, \boldsymbol{\beta}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \sigma_v^2)$ and $\pi_k (k = 1, \dots, G)$ are given in Eq. (A.5).

Step (f). The conditional distribution $p(\phi | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \boldsymbol{\beta})$ is given by

$$p(\phi | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \boldsymbol{\beta}) \propto \phi^{\frac{p}{2} + a_\phi - 1} \exp \left\{ -\phi \left[\tau_\phi + \frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{H}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \right] \right. \\ \left. + \phi^{-1} \sum_{i=1}^n \sum_{j=1}^{n_i} (y_{ij} \theta_{ij} - b(\theta_{ij})) + \sum_{i=1}^n \sum_{j=1}^{n_i} c(y_{ij}, \phi) \right\}, \quad (\text{A.8})$$

which is a non-standard or familiar distribution. Hence, similar to Liu (2001), we adopt the following Metropolized independence sampler (MIS) algorithm to sample observations from the posterior $p(\phi | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \boldsymbol{\beta})$. At the m th iteration with a current value $\phi^{(m)}$, a new candidate ϕ is draw from $h(\phi) \stackrel{D}{=} N(\phi^{(m)}, \sigma_\phi^2) I(0, \infty)$ and is accepted with probability

$$\min \left\{ 1, \frac{p(\phi | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \boldsymbol{\beta}) h(\phi^{(m)})}{p(\phi^{(m)} | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \boldsymbol{\beta}) h(\phi)} \right\}.$$

The variance σ_ϕ^2 can be chosen such that the average acceptance rate is approximately 0.25 or more.

Step (g). The conditional distribution $p(\boldsymbol{\beta} | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \phi)$ is given by

$$p(\boldsymbol{\beta} | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \phi) \propto \exp \left\{ \phi^{-1} \sum_{i=1}^n \sum_{j=1}^{n_i} (y_{ij} \theta_{ij} - b(\theta_{ij})) - \frac{\phi}{2} (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{H}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \right\}. \quad (\text{A.9})$$

Similarly, the MH algorithm for simulating observations from the conditional distribution $p(\boldsymbol{\beta} | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \phi)$ is implemented as follows. Given the current value $\boldsymbol{\beta}^{(m)}$, a new candidate $\boldsymbol{\beta}$ is generated from $N(\boldsymbol{\beta}^{(m)}, \sigma_\beta^2 \boldsymbol{\Omega}_\beta)$ and is accepted with probability

$$\min \left\{ 1, \frac{p(\boldsymbol{\beta} | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \phi)}{p(\boldsymbol{\beta}^{(m)} | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \phi)} \right\},$$

where

$$\boldsymbol{\Omega}_\beta = \phi^{-1} \left(\frac{1}{2} \sum_{i=1}^n \mathbf{X}_i^T \bar{\mathbf{V}}_i \mathbf{X}_i + \mathbf{H}_0^{-1} \right)^{-1}$$

with $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i})^T$, $\boldsymbol{\theta}_i = (\theta_{i1}, \dots, \theta_{in_i})^T$ and $\bar{\mathbf{V}}_i = E_{\mathbf{y}_i}(\partial^2 b(\boldsymbol{\theta}_i) / \partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_i^T) |_{\boldsymbol{\beta}=\mathbf{0}}$.

Step (h). The conditional distribution $p(\boldsymbol{\alpha} | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \phi, \boldsymbol{\beta}, \boldsymbol{\Lambda}_2)$ is given by

$$p(\boldsymbol{\alpha} | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \phi, \boldsymbol{\beta}, \boldsymbol{\Lambda}_2) \propto \exp \left\{ \sum_{i=1}^n \sum_{j=1}^{n_i} \phi^{-1} [y_{ij} \theta_{ij} - b(\theta_{ij})] - \frac{1}{2} \boldsymbol{\alpha}^T \boldsymbol{\Lambda}^{-1} \boldsymbol{\alpha} \right\}, \quad (\text{A.10})$$

where $f(\dot{b}(\theta_{ij})) = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{u}_i + \mathbf{B}^T(t_{ij}) \boldsymbol{\alpha}$. Again, the MH algorithm for simulating observations from the conditional distribution $p(\boldsymbol{\alpha} | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \phi, \boldsymbol{\beta}, \boldsymbol{\Lambda}_2)$ is implemented as follows. Given the current value $\boldsymbol{\alpha}^{(m)}$, a new candidate $\boldsymbol{\alpha}$ is generated from $N(\boldsymbol{\alpha}^{(m)}, \sigma_\alpha^2 \boldsymbol{\Omega}_\alpha)$ and is accepted with probability

$$\min \left\{ 1, \frac{p(\boldsymbol{\alpha} | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \phi, \boldsymbol{\beta}, \boldsymbol{\Lambda}_2)}{p(\boldsymbol{\alpha}^{(m)} | \mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{T}, \mathbf{u}, \phi, \boldsymbol{\beta}, \boldsymbol{\Lambda}_2)} \right\},$$

where

$$\mathbf{\Omega}_\alpha = \left(\frac{\phi}{2} \sum_{i=1}^n \sum_{j=1}^{n_i} \ddot{b}(\theta_{ij}) \mathbf{B}(t_{ij}) \mathbf{B}(t_{ij})^T + \mathbf{\Lambda}^{-1} \right)^{-1}$$

with $\ddot{b}(\theta_{ij}) = E_{y_{ij}} (\partial^2 b(\theta_{ij}) / \partial \mu_{ij}^2) |_{\alpha=0}$.

Step (i). The conditional distribution $p(\mathbf{\Lambda}_2 | \mathbf{T}, \alpha, \delta, \sigma_v^2)$ is given by

$$p(\mathbf{\Lambda}_2 | \mathbf{T}, \alpha, \delta, \sigma_v^2) \propto p(\alpha^{(2)} | \mathbf{\Lambda}_2) p(\mathbf{\Lambda}_2) \propto \prod_{l=1}^K p(\alpha_l^{(2)} | \sigma_\alpha^2(\kappa_l)) p(\sigma_\alpha^2(\kappa_l) | \delta, \mathbf{T}, \sigma_v^2),$$

which yields

$$p(\sigma_\alpha^2(\kappa_l) | \mathbf{T}, \alpha_l^{(2)}, \delta, \sigma_v^2) \propto (\sigma_\alpha^2(\kappa_l))^{-1/2} \exp \left\{ -\frac{\alpha_l^{(2)2}}{2\sigma_\alpha^2(\kappa_l)} - \frac{(\log \sigma_\alpha^2(\kappa_l) - \mathbf{C}^T(\kappa_l)\delta)^2}{2\sigma_v^2} \right\} \quad (\text{A.11})$$

for $l = 1, \dots, K$. The MIS algorithm for sampling observation from the conditional distribution $p(\sigma_\alpha^2(\mathcal{K}_l) | \mathbf{T}, \alpha_l^{(2)}, \delta, \sigma_v^2)$ can be implemented as follows. Given the current value $\sigma_\alpha^{2(m)}(\mathcal{K}_l)$, a new candidate $\sigma_\alpha^2(\mathcal{K}_l)$ is drawn from $h(\sigma_\alpha^2(\mathcal{K}_l)) \stackrel{D}{=} N(\sigma_\alpha^{2(m)}(\mathcal{K}_l), \sigma_{\mathcal{K}_l}^2) I(0, \infty)$ and is accepted with probability

$$\min \left\{ 1, \frac{p(\sigma_\alpha^2(\mathcal{K}_l) | \mathbf{T}, \alpha_l^{(2)}, \delta, \sigma_v^2) h(\sigma_\alpha^{2(m)}(\mathcal{K}_l))}{p(\sigma_\alpha^{2(m)}(\mathcal{K}_l) | \mathbf{T}, \alpha_l^{(2)}, \delta, \sigma_v^2) h(\sigma_\alpha^2(\mathcal{K}_l))} \right\},$$

where $\sigma_{\mathcal{K}_l}^2$ is chosen such that the average acceptance rate is about 0.25 or more.

Step (j). The conditional distribution $p(\delta | \mathbf{T}, \mathbf{\Lambda}_2, \gamma, \sigma_v^2)$ is given by

$$p(\delta | \mathbf{T}, \mathbf{\Lambda}_2, \gamma, \sigma_v^2) \propto \exp \left\{ -\frac{1}{2\sigma_v^2} \sum_{l=1}^K (\log \sigma_\alpha^2(\kappa_l) - \mathbf{C}^T(\kappa_l)\delta)^2 - \frac{1}{2} \delta^T \mathbf{\Omega}^{-1} \delta \right\},$$

which leads to

$$p(\delta | \mathbf{T}, \mathbf{\Lambda}_2, \gamma, \sigma_v^2) \sim N_{q+K_\alpha+1}(\boldsymbol{\mu}_\delta, \mathbf{M}_\delta), \quad (\text{A.12})$$

where $\mathbf{M}_\delta = (\mathbf{\Omega}^{-1} + \sum_{l=1}^K \mathbf{C}(\kappa_l) \mathbf{C}^T(\kappa_l) / \sigma_v^2)^{-1}$ and $\boldsymbol{\mu}_\delta = \mathbf{M}_\delta \sum_{l=1}^K \mathbf{C}(\kappa_l) \log \sigma_\alpha^2(\kappa_l) / \sigma_v^2$.

Step (k). For the conditional distribution of γ^{-1} , we have $p(\gamma^{-1} | \delta) \propto p(\delta | \gamma) p(\gamma^{-1})$, which yields

$$p(\gamma^{-1} | \delta) \sim \Gamma(a_\gamma + K_\alpha/2, \tau_\gamma + \delta^{(2)T} \delta^{(2)} / 2). \quad (\text{A.13})$$

References

- Baladandayuthapani, V., Mallick, B.K., Carroll, R.J., 2005. Spatially adaptive Bayesian penalized regression splines (*P*-splines). *Journal of Computational and Graphical Statistics* 14, 378–394.
- Bayarri, M.J., Berger, J.O., 2000. *P* values for composite null models. *Journal of the American Statistical Association* 95, 1127–1142.
- Brown, E.R., Ibrahim, J.G., 2003. A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics* 59, 221–228.
- Chen, X.D., Tang, N.S., 2010. Bayesian Analysis of semiparametric reproductive dispersion mixed-effects models. *Computational Statistics and Data Analysis* 54, 2145–2158.
- Chow, S.M., Tang, N.S., Yuan, Y., Song, X.Y., Zhu, H.T., 2011. Bayesian estimation of semiparametric nonlinear dynamic factor analysis models using the Dirichlet process prior. *British Journal of Mathematical and Statistical Psychology* 64, 69–106.
- DiCiccio, T.J., Kass, R.E., Raftery, A., Wasserman, L., 1997. Computing Bayes factors by combining simulation and asymptotic approximations. *Journal of the American Statistical Association* 18, 251–262.
- Diggle, P., Liang, K.L., Zeger, S., 2002. *Analysis of Longitudinal Data*, second ed. Oxford University Press.
- Dunson, D.B., 2006. Bayesian dynamic modeling of latent trait distributions. *Biostatistics* 7, 551–568.
- Ferguson, T.S., 1973. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* 1, 209–230.
- Gelman, A., 1996. Inference and monitoring convergence. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.
- Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 721–741.
- Gelman, A., Meng, X.L., 1998. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statistical Science* 13, 163–185.
- Gelman, A., Meng, X.L., Stern, H., 1996. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* 6, 733–807.
- Guha, S., 2008. Posterior simulation in the generalized linear mixed model with semiparametric random effects. *Journal of Computational and Graphical Statistics* 17, 410–425.
- He, X.M., Fung, W.K., Zhu, Z.Y., 2005. Robust estimation in generalized partial linear models for clustered data. *Journal of the American Statistical Association* 100, 1176–1184.
- Ishwaran, H., Zarepour, M., 2000. Markov chain monte carlo in approximate dirichlet and beta two-parameter process hierarchical models. *Biometrika* 87, 371–390.

- Kass, R.E., Raftery, A.E., 1995. Bayes factor. *Journal of the American Statistical Association* 90, 773–795.
- Kleinman, K.P., Ibrahim, J.G., 1998a. A semi-parametric Bayesian approach to generalized linear mixed models. *Statistics in Medicine* 17, 2579–2596.
- Kleinman, K.P., Ibrahim, J.G., 1998b. A semiparametric Bayesian approach to the random effects model. *Biometrics* 54, 921–938.
- Lee, S.Y., Lu, B., Song, X.Y., 2008. Semiparametric Bayesian analysis of structural equation models with fixed covariates. *Statistics in Medicine* 15, 2341–2360.
- Lee, S.Y., Song, X.Y., 2003. Statistical analysis of nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology* 53, 209–232.
- Lee, S.Y., Tang, N.S., 2006a. Bayesian analysis of structural equation models with mixed exponential family and ordered categorical data. *British Journal of Mathematical and Statistical Psychology* 59, 151–172.
- Lee, S.Y., Tang, N.S., 2006b. Bayesian analysis of nonlinear structural equation models with nonignorable missing data. *Psychometrika* 71, 541–564.
- Lee, S.Y., Zhu, H.T., 2000. Maximum likelihood estimation and model comparison for mixtures of structural equation models with ignorable missing data. *Journal of Classification* 20, 221–255.
- Liang, H., 2009. Generalized partially linear mixed-effects models incorporating mismeasured covariates. *Annals of the Institute of Statistical Mathematics* 61, 27–46.
- Lin, X., Carroll, R.J., 2001. Semiparametric regression for clustered data. *Biometrika* 88, 1179–1185.
- Lin, X.H., Zhang, D.W., 1999. Inference in Generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society, Series B* 61, 381–400.
- Liu, J.S., 2001. *Monte Carlo Strategies in Scientific Computing*. Springer-Verlag, New York.
- Meng, X.L., Wong, Y.X., 1996. Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica* 6, 831–860.
- Qin, G.Y., Zhu, Z.Y., 2007. Robust estimation in generalized semiparametric mixed models for longitudinal data. *Journal of Multivariate Analysis* 98, 1658–1683.
- Qin, G.Y., Zhu, Z.Y., 2009. Robustified maximum likelihood estimation in generalized partial linear mixed model for longitudinal data. *Biometrics* 65, 52–59.
- Ruppert, D., Wand, M.P., Carroll, R.J., 2003. *Semiparametric Regression*. Cambridge University Press, New York.
- Sethuraman, J., 1994. A constructive definition of dirichlet priors. *Statistica Sinica* 4, 639–650.
- Sommer, A., Katz, J., Tarwotjo, I., 1984. Increased risk of respiratory infection and diarrhea in children with pre-existing mild vitamin a deficiency. *American Journal of Clinical Nutrition* 40, 1090–1095.
- Song, X.Y., Lu, Z.H., 2010. Semiparametric latent variable models with Bayesian P -splines. *Journal of Computational and Graphical Statistics* 19, 590–608.
- Song, X.Y., Pan, J.H., Kwok, T., Vandenput, L., Ohlsson, C., Leung, P.C., 2010. A semiparametric Bayesian approach for structural equation models. *Biometrical Journal* 52, 314–332.