Statistics and Its Interface Volume 11 (2018) 251–264

Bayesian inference on multivariate-t nonlinear mixed-effects models for multiple longitudinal data with missing values

Wan-Lun Wang* and Luis M. Castro

The multivariate-t nonlinear mixed-effects model (MtNLMM) has been shown to be a promising robust tool for analyzing multiple longitudinal trajectories following arbitrary growth patterns in the presence of outliers and possible missing responses. Owing to intractable likelihood function of the model, we devise a fully Bayesian estimating procedure to account for the uncertainties of model parameters, random effects, and missing responses via the Markov chain Monte Carlo method. Posterior predictive inferences for the future values and missing responses are also investigated. We conduct a simulation study to demonstrate the feasibility of our Bayesian sampling schemes. The proposed techniques are illustrated through applications to two case studies.

Keywords and phrases: Missing responses, Multivariate longitudinal data, Nonlinear mean profiles, Posterior distributions, Taylor series expansion.

1. INTRODUCTION

The multivariate nonlinear mixed-effects model (MNLMM), as proposed by Marshall et al. (2006), has been shown to be a promising tool for analyzing multivariate longitudinal data with arbitrary patterns of continuous responses collected from many research fields such as biomedical, psychological, environmental science and clinical studies. The multivariate linear mixed-effects model (MLMM) (Shah et al., 1997) can be viewed as a special case of the MNLMM when the link function is specified as a simple identity function that is linear in fixed effects and random effects. There is a large amount of literature describing methods for the estimation of MLMM and MNLMM along with their applications. From a maximum likelihood (ML) perspective, the related work can be found, for example, in Sammel et al. (1999), Roy and Lin (2002), Roy (2006), Marshall et al. (2009), Wang and Fan (2010), and Wang (2015). On the other hand, De la Cruz-Mesía and Marshall (2006) developed a Bayesian treatment of the nonlinear mixed-effects model (Lindstrom and Bates, 1990, NLMM). Schafer and Yucel (2002) proposed Markov chain Monte Carlo (MCMC) sampling-based strategies for Bayesian analysis of the MLMM with missing values. Recently, De la Cruz (2014) investigated the Bayesian analysis of the NLMM under a class of fat-tailed multivariate distributions, including the Student's t and slash and contaminated normal.

In the framework of MLMM and MNLMM, the random effects and within-subject errors are routinely assumed to follow a multivariate normal distribution for mathematical tractability and computational simplicity. However, such a normality assumption can usually cause a lack of robustness against outliers and subsequently result in invalid inference. To overcome this weakness, a number of authors have considered different extensions of mixed-effects models based on the multivariate-t distribution (Kotz and Nadarajah, 2004) for robust modeling both sources of variability. Pinheiro et al. (2001) introduced the t linear mixed-effects model (tLMM) constructed by assuming multivariate-t distributed random effects and errors to accommodate outlying responses within subjects or unusual subjects. Wang and Fan (2011) presented a multi-outcome version of tLMM, namely the multivariate-t linear mixed-effects model (MtLMM), for analyzing multivariate longitudinal data. More recently, Wang and Lin (2014) proposed the multivariate-t nonlinear mixed-effects model (MtNLMM) as an extension of the MNLMM for robust inference.

In this paper, we adopt a variant of MCMC method for a Bayesian treatment of the MtNLMM to assess the uncertainties of model parameters. The proposed sampling procedure allows to vary the specification of joint posterior densities which are formulated by incorporating the joint prior density with the approximate likelihood function, obtained by using the first-order Taylor series expansion on the model around the individual parameters. The proposed Bayesian sampling procedure provides a tremendous flexibility in generating posterior samples from their full conditional posterior distributions.

The occurrence of missing responses with arbitrary patterns is an unavoidable problem in multi-outcome longitudinal studies due to a variety of reasons, for example, missed visits, dropouts, loss to follow-up, death or disabling conditions, and so forth. A comprehensive study that covers methodological and computational aspects of handling missing data can be found in Schafer and Yucel

*Corresponding author.

(2002), Marshall et al. (2006), Wang (2013), Wang and Lin (2014), and Wang and Lin (2015), among others. To reflect extra uncertainty due to missing data, this paper also offers a modified MCMC procedure for the MtNLMM with missing responses at random. The proposed fully Bayesian treatment of MtNLMM allows for taking the uncertainties of model parameters and missing outcomes into account through appropriate prior choices and is shown to provide more accurate inference than does the likelihood-based approach (Wang and Lin, 2014), especially when the small size is not large. Further, our proposed Bayesian model includes that of Wang and Lin (2015) as a special case if the mean functions relating to covariates are made linear in parameters.

The organization of this paper is as follows. In Section 2, we establish notation and present the Bayesian formulation of the MtNLMM together with the prior specifications. In Section 3, we describe the implementation of the MCMC algorithms and discuss posterior predictive inferences for missing values and future responses. Section 4 illustrates our methodology through two real-data examples from HIV/AIDS and pregnant women studies. A simulation study is conducted in Section 5 to study whether the Bayesian treatment of MtNLMM can give reasonable results under various missingness settings. Section 6 concludes with a short summary of issues raised by our methods and some directions for possible future research. The required full conditional posterior distributions and the implementation for Metropolis-Hastings (M-H) algorithm (Hastings, 1970) are relegated in the appendices. R code for implementation of the proposed Bayesian approach is available online (see Supplemental Materials Section, http://intlpress.com/site/pub/pages/journals/items/sii/content/vols/0011/0002/s001).

2. BAYESIAN MODEL SPECIFICATION

2.1 Multivariate-t nonlinear mixed-effects model

Suppose that there are N subjects in the study. Each subject has his/her own response matrix $\boldsymbol{Y}_i$ which is composed of r column vectors $\boldsymbol{y}_{ij} = (y_{ij,1}, \ldots, y_{ij,s_i})^{\mathrm{T}}$ for the jth characteristic ($j = 1, \ldots, r$) and, from the other side, s_i row vectors $\boldsymbol{y}_{i,k} = (y_{i1,k}, \ldots, y_{ir,k})$ for the kth occasion ($k = 1, \ldots, s_i$). The response matrix with dimension $s_i \times r$ can be formed as $\boldsymbol{Y}_i = [\boldsymbol{y}_{i1} : \cdots : \boldsymbol{y}_{ir}] = [\boldsymbol{y}_{i,1}^{\mathrm{T}} : \cdots : \boldsymbol{y}_{i,s_i}^{\mathrm{T}}]^{\mathrm{T}}$ for the ith subject ($i = 1, \ldots, N$). Let $\boldsymbol{x}_i$ collect a set of covariates for the ith subject. The relationship between responses $\boldsymbol{Y}_i$ and covariates $\boldsymbol{x}_i$ cannot be completely modeled by the regression mean function, so the within-subject errors are needed and defined by a $s_i \times r$ matrix $\boldsymbol{E}_i = [\boldsymbol{e}_{i1} : \cdots : \boldsymbol{e}_{ir}] = [\boldsymbol{e}_{i,1}^{\mathrm{T}} : \cdots, \boldsymbol{e}_{i,s_i}^{\mathrm{T}}]^{\mathrm{T}}$. Here $\boldsymbol{e}_{ij} = (e_{ij,1}, \ldots, e_{ij,s_i})^{\mathrm{T}}$ is a column vector corresponding to $\boldsymbol{y}_{ij}$, and $\boldsymbol{e}_{i,k} = (e_{i1,k}, \ldots, e_{ir,k})$ is a row vector corresponding to $\boldsymbol{y}_{i,k}$. For convenience of model formulation, the vec($\cdot$) operator, which strings out all columns of a matrix vertically into a stacked vector, is utilized on $\boldsymbol{Y}_i$ and $\boldsymbol{E}_i$ such that $\boldsymbol{y}_i = \mathrm{vec}(\boldsymbol{Y}_i)$ and $\boldsymbol{e}_i = \mathrm{vec}(\boldsymbol{E}_i)$.

The MtNLMM for the ith subject takes the form:

$$\boldsymbol{y}_i = \boldsymbol{\mu}_i(\boldsymbol{\eta}_i, \boldsymbol{x}_i) + \boldsymbol{e}_i, \tag{1}$$

where $\boldsymbol{\mu}_i$ is a nonlinear, vector-valued and differentiable function used to link the relationship between the responses $\boldsymbol{y}_i$ and covariates $\boldsymbol{x}_i$ through a vector-valued individual parameters $\boldsymbol{\eta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i$. In the way, the fixed effects $\boldsymbol{\beta}$ and random effects $\boldsymbol{b}_i$ can be incorporated into the model such that $\boldsymbol{\mu}_i(\boldsymbol{\eta}_i, \boldsymbol{X}_i) = \boldsymbol{\mu}_i(\boldsymbol{\beta}, \boldsymbol{b}_i)$, where $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ are design matrices of size $g \times p$ and $g \times q$ for the fixed effects and random effects, respectively. We further assume that the random effects and within-subject errors are jointly distributed as

$$\begin{bmatrix} \boldsymbol{b}_i \\ \boldsymbol{e}_i \end{bmatrix} \sim t_{q+n_i}\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{D} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R}_i \end{bmatrix}, \nu \right), \tag{2}$$

where $t_a(\boldsymbol{\mu}, \boldsymbol{\Omega}, \nu)$ represents the a-variate t distribution with location vector $\boldsymbol{\mu}$, scale-covariance matrix $\boldsymbol{\Omega}$ and degrees of freedom (DOF) ν, and $n_i = s_i r$. Under the assumption of (2), the two-level hierarchy of the MtNLMM is

$$\boldsymbol{y}_i | \boldsymbol{b}_i \sim t_{n_i}(\boldsymbol{\mu}_i(\boldsymbol{\beta}, \boldsymbol{b}_i), \boldsymbol{R}_i, \nu), \qquad \boldsymbol{b}_i \sim t_q(\boldsymbol{0}, \boldsymbol{D}, \nu). \tag{3}$$

For computational convenience, we consider $\boldsymbol{R}_i = \boldsymbol{\Sigma} \otimes \boldsymbol{C}_i$ for describing the among-characteristic and among-occasion correlations simultaneously, where $\boldsymbol{\Sigma} = [\sigma_{jl}] \in \mathcal{R}^{r \times r}$, for $j, l = 1, \ldots, r$, and $\boldsymbol{C}_i \in [-1, 1]^{s_i \times s_i}$. According to the assumption of $\boldsymbol{e}_{ij}$ and $\boldsymbol{e}_{i,k}$, the jth column (outcome) and the kth row (occasion) of $\boldsymbol{Y}_i$, say $\boldsymbol{y}_{ij}$ and $\boldsymbol{y}_{i,k}$, respectively, can be written as

$$\boldsymbol{y}_{ij} = \boldsymbol{\mu}_{ij}(\boldsymbol{\eta}_i, \boldsymbol{x}_{ij}) + \boldsymbol{e}_{ij}, \quad \text{and} \quad \boldsymbol{y}_{i,k} = \boldsymbol{\mu}_i^k(\boldsymbol{\eta}_i, \boldsymbol{x}_{ik}) + \boldsymbol{e}_{i,k},$$

where $\boldsymbol{\mu}_{ij}(\boldsymbol{\eta}_i, \boldsymbol{x}_{ij}) = (\mu_j(\boldsymbol{\eta}_i, \boldsymbol{x}_{ij,1}), \ldots, \mu_j(\boldsymbol{\eta}_i, \boldsymbol{x}_{ij,s_i}))^{\mathrm{T}}$ represents the vector of a link function relating the jth outcome $\boldsymbol{y}_{ij}$ over s_i occasions to the covariates $\boldsymbol{x}_{ij}$, and $\boldsymbol{\mu}_i^k(\boldsymbol{\eta}_i, \boldsymbol{x}_{ik}) = (\mu_1(\boldsymbol{\eta}_i, \boldsymbol{x}_{i1,k}), \ldots, \mu_j(\boldsymbol{\eta}_i, \boldsymbol{x}_{ij,k}), \ldots, \mu_r(\boldsymbol{\eta}_i, \boldsymbol{x}_{ir,k}))$ represents a vector of r link functions with each relating the corresponding outcome variable at the same time to the covariates $\boldsymbol{x}_{i,k}$.

To avoid the non-identifiability problem (Galecki, 1994), the specification of variance component for within-subject errors which is used to describe the serial correlation among occasions, denoted by $\boldsymbol{C}_i$, should be specified as a correlation matrix rather than a covariance matrix. Toward this end, we exploit a parsimonious damped exponential correlation (Muñoz et al., 1992, DEC) structure which is simple and flexible for handling the observations measured at irregularly visited occasions using a function of time and few parameters. The DEC structure is defined as

$$\boldsymbol{C}_i = \boldsymbol{C}_i(\boldsymbol{\phi}; \boldsymbol{t}_i) = \left[\phi_1^{|t_{ik} - t_{ik'}|^{\phi_2}}\right], \tag{4}$$

where the parameter vector $\boldsymbol{\phi}$ contains the autoregressive (AR) coefficient $\phi_1 \in [0, 1)$, which describes the autocor-

relation between observations separated by the absolute length of two time-points, and the damping parameter $\phi_2 \in [0, \infty)$, which permits acceleration of the exponential decay of the autocorrelation function. The DEC structure in (4) includes the compound symmetry, the continuous-type first-order autoregressive (AR(1)) model and the first-order moving average (MA(1)) model as special cases formed by setting $\phi_2 = 0$, $\phi_2 = 1$ and $\phi_2 \to \infty$, respectively.

2.2 Prior and posterior distributions

In a Bayesian treatment, one must specify prior distributions for model parameters $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{D}, \boldsymbol{\Sigma}, \phi_1, \phi_2, \nu\}$. Assuming that the model parameters $\boldsymbol{\theta}$ are independent a priori, the joint prior density is

$$\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\beta})\pi(\boldsymbol{D})\pi(\boldsymbol{\Sigma})\pi(\phi_1)\pi(\phi_2)\pi(\nu). \tag{5}$$

The prior distributions adopted are as follows:

$$\begin{aligned}&\boldsymbol{\beta} \sim \mathcal{N}_p(\boldsymbol{\beta}_0, \boldsymbol{F}_0),\ \boldsymbol{D} \sim \mathcal{IW}(d_0, \boldsymbol{G}_0),\ \boldsymbol{\Sigma} \sim \mathcal{IW}(s_0, \mathbf{H}_0),\\ &\phi_1 \sim \mathcal{U}(0,1),\ (1+\phi_2)^{-1} \sim \mathcal{U}(0,1),\ (1+\nu)^{-1} \sim \mathcal{U}(0,1),\end{aligned}$$

where $\mathcal{U}(0,1)$ denotes a uniform distribution between 0 and 1, and $\mathcal{IW}(s, \boldsymbol{\Omega})$ denotes an inverse Wishart distribution with scale-covariance matrix $\boldsymbol{\Omega}$ and DOF s. The hyperparameters $\boldsymbol{\beta}_0$, $\boldsymbol{F}_0$, d_0, $\boldsymbol{G}_0$, s_0 and $\mathbf{H}_0$ are specified to reflect vague prior information.

In order to conduct the Bayesian computation of MtNLMM, we formulate *four joint posterior densities* used across *four cycles* of the procedure. Firstly, we treat $\boldsymbol{b} = \{\boldsymbol{b}_i\}_{i=1}^N$ as latent data and combine them with observed data $\boldsymbol{y} = \{\boldsymbol{y}_i\}_{i=1}^N$ as the complete data. Multiplying the joint prior density (5) by the complete-data likelihood function sketched based on hierarchy (3) leads to the *first* joint posterior density, given by

$$p(\boldsymbol{\theta}, \boldsymbol{b}|\boldsymbol{y}) = \pi(\boldsymbol{\theta}) \prod_{i=1}^{N} t_{n_i}(\boldsymbol{y}_i|\boldsymbol{\mu}_i(\boldsymbol{\beta}, \boldsymbol{b}_i), \boldsymbol{R}_i, \nu) t_q(\boldsymbol{b}_i|\mathbf{0}, \boldsymbol{D}, \nu), \tag{6}$$

where $t_a(\cdot|\boldsymbol{\mu}, \boldsymbol{\Omega}, \nu)$ represents the probability density function (pdf) of $t_a(\boldsymbol{\mu}, \boldsymbol{\Omega}, \nu)$.

In view of (6), none of full conditional posterior has a recognizable distribution. Due to the high dimensionality of parameters $\boldsymbol{\theta}$ and random effects $\boldsymbol{b}$, the implementation of the M-H algorithm for simulating each entry of $\boldsymbol{\theta}$ and $\{\boldsymbol{b}_i\}_{i=1}^N$ is painfully difficult. The difficulty we suffer from is "how to select the suitable proposal distributions which can provide stable convergence for each parameter". To remit this difficulty, we utilize the stochastic representation of multivariate-t distribution and a Taylor approximation of the model. Accordingly, the full conditional posteriors of $\boldsymbol{b}_i$ and all entries in $\boldsymbol{\theta}$ except for $\boldsymbol{\phi}$ and ν show standard distributional forms. Thus the Gibbs sampler (Geman and Geman, 1984) can be implemented straightforwardly to simulate posterior samples of most of the parameters and random effects, while the M-H algorithm is performed merely for $\boldsymbol{\phi}$ and ν.

Based on the essential property of multivariate-t distribution, we introduce a set of scaling weights (latent data) $\boldsymbol{\tau} = \{\tau_i\}_{i=1}^N$, where $\tau_i \sim \text{Gamma}(\nu/2, \nu/2)$. The *second* joint posterior density of $(\boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{\tau})$ is

$$\begin{aligned}p(\boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{\tau}|\boldsymbol{y}) \quad = \quad & \pi(\boldsymbol{\theta}) \prod_{i=1}^{N} \phi_{n_i}\big(\boldsymbol{y}_i \mid \boldsymbol{\mu}_i(\boldsymbol{\beta}_i, \boldsymbol{b}_i), \tau_i^{-1}\boldsymbol{R}_i\big)\\ & \times \phi_q\big(\boldsymbol{b}_i \mid \mathbf{0}, \tau_i^{-1}\boldsymbol{D}\big)\mathcal{G}\big(\tau_i|\nu/2, \nu/2\big),\end{aligned} \tag{7}$$

where $\phi_a(\cdot|\boldsymbol{\mu}, \boldsymbol{\Omega})$ denotes the pdf of a-variate normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Omega}$, and $\mathcal{G}(\cdot|a, b)$ denotes the pdf of gamma distribution with mean a/b and variance a/b^2.

Then, we apply the first-order Taylor expansion on model (1) around the unobservable individual parameter $\boldsymbol{\eta}_i^{(s)} = \boldsymbol{A}_i\boldsymbol{\beta}_i^{(s)} + \boldsymbol{B}_i\boldsymbol{b}_i^{(s)}$ to linearize the MtNLMM, where $\boldsymbol{\beta}^{(s)}$ and $\{\boldsymbol{b}_i^{(s)}\}_{i=1}^N$ are the posterior samples of fixed effects and random effects at the sth iteration of the MCMC procedure. Model (1) becomes

$$\tilde{\boldsymbol{y}}_i = \tilde{\boldsymbol{X}}_i\boldsymbol{\beta} + \tilde{\boldsymbol{Z}}_i\boldsymbol{b}_i + \boldsymbol{e}_i, \tag{8}$$

which is called the *pseudo-data* model henceforth. The notation $\tilde{\boldsymbol{y}}_i$ is an $n_i \times 1$ vector composed of r *pseudo-response* vectors $\tilde{\boldsymbol{y}}_{ij} = (\tilde{y}_{ij,1}, \cdots, \tilde{y}_{ij,s_i})^{\mathrm{T}}$ in which

$$\tilde{y}_{ij,k} = y_{ij,k} - \mu_j(\boldsymbol{\eta}_i^{(s)}, \boldsymbol{x}_{ij,k}) + \tilde{\boldsymbol{x}}_{ij,k}\boldsymbol{\beta}^{(s)} + \tilde{\boldsymbol{z}}_{ij,k}\boldsymbol{b}_i^{(s)},$$

$\tilde{\boldsymbol{X}}_i$ is an $n_i \times p$ matrix with rows made up of $p \times 1$ vector $\tilde{\boldsymbol{x}}_{ij,k} = \dot{\mu}_j(\boldsymbol{\eta}_i^{(s)}, \boldsymbol{x}_{ij,k})^{\mathrm{T}}\boldsymbol{A}_i$, and $\tilde{\boldsymbol{Z}}_i$ is an $n_i \times q$ matrix with rows made up of $q \times 1$ vector $\tilde{\boldsymbol{z}}_{ij,k} = \dot{\mu}_j(\boldsymbol{\eta}_i^{(s)}, \boldsymbol{x}_{ij,k})^{\mathrm{T}}\boldsymbol{B}_i$, where $\dot{\mu}_j(\boldsymbol{\eta}_i^{(s)}, \boldsymbol{x}_{ij,k})$ is the first partial derivative of $\mu_j(\boldsymbol{\eta}_i^{(s)}, \boldsymbol{x}_{ij,k})$ with respect to $\boldsymbol{\eta}_i$.

In light of the pseudo-data model specified in (8) along with assumption (2), we obtain $\tilde{\boldsymbol{y}}_i \sim t_{n_i}(\tilde{\boldsymbol{X}}_i\boldsymbol{\beta}, \tilde{\boldsymbol{\Lambda}}_i, \nu)$, where $\tilde{\boldsymbol{\Lambda}}_i = \tilde{\boldsymbol{Z}}_i\boldsymbol{D}\tilde{\boldsymbol{Z}}_i^{\mathrm{T}} + \boldsymbol{\Sigma} \otimes \boldsymbol{C}_i$. Using the scaling weights $\boldsymbol{\tau}$, the two-level hierarchy for pseudo-data model (8) is given by

$$\tilde{\boldsymbol{y}}_i|\tau_i \sim \mathcal{N}_{n_i}\big(\tilde{\boldsymbol{X}}_i\boldsymbol{\beta}, \tau_i^{-1}\tilde{\boldsymbol{\Lambda}}_i\big), \quad \tau_i \sim \text{Gamma}(\nu/2, \nu/2). \tag{9}$$

Multiplying the joint prior density in (5) by the complete-data likelihood, which is the product of the pdfs of (9) for all subjects, yields the *third* joint posterior, denoted by $p(\boldsymbol{\theta}, \boldsymbol{\tau}|\tilde{\boldsymbol{y}})$, where $\tilde{\boldsymbol{y}} = \{\tilde{\boldsymbol{y}}_i\}_{i=1}^N$.

Model (8) can be alternatively represented as

$$\begin{aligned}\tilde{\boldsymbol{y}}_i|(\boldsymbol{b}_i, \tau_i) \quad &\sim \quad \mathcal{N}_{n_i}\big(\tilde{\boldsymbol{X}}_i\boldsymbol{\beta} + \tilde{\boldsymbol{Z}}_i\boldsymbol{b}_i, \tau_i^{-1}\boldsymbol{R}_i\big),\\ \boldsymbol{b}_i|\tau_i \quad &\sim \quad \mathcal{N}_q(\mathbf{0}, \tau_i^{-1}\boldsymbol{D}),\\ \tau_i \quad &\sim \quad \text{Gamma}(\nu/2, \nu/2).\end{aligned} \tag{10}$$

Combining the data information from (10) with the prior density in (5) gives rise to the *fourth* joint posterior

$p(\boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{\tau}|\tilde{\boldsymbol{y}})$. Following the above four joint posteriors, the full conditional posterior distributions of parameters as well as latent data, which are useful for the MCMC procedures, are sketched in Appendix A.

3. COMPUTATIONAL STRATEGIES

3.1 Implementation of MCMC algorithm

The MCMC methods are primarily used for calculating numerical approximations of multi-dimensional integrals through sampling from a probability distribution to create a Markov chain that has the target distribution as its equilibrium distribution. The two most popular tools of MCMC methods in Bayesian inference are the Gibbs sampler (Geman and Geman, 1984) and M-H algorithm (Hastings, 1970) which draw posterior samples sequentially from the full conditional posterior distributions and then correct these samples to achieve an approximation of the target posterior distribution. However, the Bayesian computation for the MtNLMM is complicated such that the conventional MCMC procedure is computationally inefficient when all the full conditional posteriors are derived from (6). To go further, we exploit a variant of MCMC method that allows to vary the specification of the joint posterior densities with the *data augmentation (DA)* steps for random effects as well as scaling weights and *posterior (P)* steps for parameters at each iteration. Therefore, the MCMC method offers enormous flexibility in formulating the procedure and typical efficiency in sampling.

Let $\boldsymbol{\theta}^{(s)} = \{\boldsymbol{\beta}^{(s)}, \boldsymbol{D}^{(s)}, \boldsymbol{\Sigma}^{(s)}, \boldsymbol{\phi}^{(s)}, \nu^{(s)}\}$, $\{\boldsymbol{b}_i^{(s)}\}_{i=1}^N$, and $\{\tau_i^{(s)}\}_{i=1}^N$ be the generated posterior samples of parameters, random effects and scaling weights, respectively, at the sth iteration. The MCMC algorithm, which consists of four cycles of DA-steps and/or P-steps with each cycle being developed under one of the four joint posterior densities, proceeds as Algorithm 1.

Algorithm 1. *(MCMC algorithm for the MtNLMM)*

P-Step for the 1st cycle. *Generate $\boldsymbol{\phi} = (\phi_1, \phi_2)$ and ν from (A.1) and (A.2), respectively, via the M-H algorithm described in Appendix B.*

DA-Step for the 2nd cycle. *Generate τ_i from (A.3) via the Gibbs sampler, for $i = 1, \ldots, N$.*

P-Step for the 2nd cycle. *Generate $\boldsymbol{D}$ and $\boldsymbol{\Sigma}$ from (A.4) and (A.5), respectively, via the Gibbs sampler.*

P-Step for the 3rd cycle. *Generate $\boldsymbol{\beta}$ from (A.6) via the Gibbs sampler.*

DA-Step for the 4th cycle. *Generate $\boldsymbol{b}_i$ from (A.8) via the Gibbs sampler, for $i = 1, \cdots, N$.*

Having the post-convergence posterior samples of size L, say $\{\boldsymbol{\theta}^{(l)}\}_{l=1}^L$ that should be a subset of $\{\boldsymbol{\theta}^{(s)}\}_{s=1}^S = \{\boldsymbol{\beta}^{(s)}, \boldsymbol{D}^{(s)}, \boldsymbol{\Sigma}^{(s)}, \boldsymbol{\phi}^{(s)}, \nu^{(s)}\}_{s=1}^S$ after removing the 'burn-in' samples, denoted by $\{\boldsymbol{\theta}^{(s)}, s \geq s_0\}$, the posterior means of parameters can be estimated by $\hat{\boldsymbol{\beta}} = \frac{1}{L}\sum_{l=1}^L \boldsymbol{\beta}^{(l)}$, $\hat{\boldsymbol{D}} = \frac{1}{L}\sum_{l=1}^L \boldsymbol{D}^{(l)}$, $\hat{\boldsymbol{\Sigma}} = \frac{1}{L}\sum_{l=1}^L \boldsymbol{\Sigma}^{(l)}$, and $\hat{\boldsymbol{\phi}} = \frac{1}{L}\sum_{l=1}^L \boldsymbol{\phi}^{(l)}$. Due to the heavy-tailed behavior, the posterior median is an appropriate estimator for ν. The $100(1-\alpha)\%$ posterior intervals for $\boldsymbol{\theta}$ can be constructed by $[\boldsymbol{\theta}^{[\alpha/2]}, \boldsymbol{\theta}^{[1-\alpha/2]}]$, where $\boldsymbol{\theta}^{[\alpha]}$ denotes the α-percentile of posterior samples of $\boldsymbol{\theta}^{(l)}$.

Furthermore, it is of interest to interpret subject-specific variability by the estimation of unobservable random effects and obtain the fitted values of repeated measures. Using the post-convergence samples of $\{\boldsymbol{b}^{(l)}\}_{l=1}^L$, the posterior estimates of random effects and fitted responses can be calculated as

$$\hat{\boldsymbol{b}}_i = \frac{1}{L}\sum_{l=1}^L \boldsymbol{b}_i^{(l)}, \quad \text{and} \quad \hat{\boldsymbol{y}}_i = \frac{1}{L}\sum_{l=1}^L \boldsymbol{\mu}_i(\boldsymbol{\beta}^{(l)}, \boldsymbol{b}_i^{(l)}). \tag{11}$$

3.2 Imputation for missing responses

The occurrence of missing data due to a variety of reasons becomes a ubiquitous problem for researchers in practice. For fitting the (univariate) NLMM and tNLMM with missing values, the estimation process is straightforward without too much complexity added to the computational burden because it can be done through distinct subject-specific design matrices. However, missingness on multivariate longitudinal data can produce more complex patterns such as an *intermittent* structure, which indicates that one characteristic could be measured but the other could be missing for a subject during his/her scheduled visits. To handle such kind of data, the Bayesian methodology for fitting the MtNLMM under an incomplete-data framework is also developed. Most statistical procedures for conducting missing data depend on conditions of missing-data mechanisms. In what follows, we assume that the missingness of data is under missing at random (Rubin, 1976, MAR) with an ignorability, meaning that the missing-data process relies only upon the observed values themselves.

Following the notation introduced in Wang and Lin (2014) especially for the permutation matrices $\boldsymbol{O}_i$ and $\boldsymbol{M}_i$, which extract the observed and missing parts of observations of each subject, model (1) for partially observed data can be written as

$$\boldsymbol{y}_i^{\mathrm{o}} = \boldsymbol{\mu}_i^{\mathrm{o}}(\boldsymbol{\eta}_i, \boldsymbol{x}_i) + \boldsymbol{e}_i^{\mathrm{o}}.$$

The *first two* joint posterior densities are modified as $p(\boldsymbol{\theta}, \boldsymbol{b}|\boldsymbol{y}^{\mathrm{o}})$ and $p(\boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{\tau}|\boldsymbol{y}^{\mathrm{o}})$ accordingly.

Similarly, pseudo-data model (8) can be rewritten as

$$\tilde{\boldsymbol{y}}_i^{\mathrm{o}} = \tilde{\boldsymbol{X}}_i^{\mathrm{o}}\boldsymbol{\beta} + \tilde{\boldsymbol{Z}}_i^{\mathrm{o}}\boldsymbol{b}_i + \boldsymbol{e}_i^{\mathrm{o}}$$

in which only the observed responses and their corresponding covariates are included in $\tilde{\boldsymbol{y}}_i$, $\tilde{\boldsymbol{X}}_i$ and $\tilde{\boldsymbol{Z}}_i$. Then we have $\tilde{\boldsymbol{y}}_i^{\mathrm{o}} \sim t_{n_i^{\mathrm{o}}}(\tilde{\boldsymbol{X}}_i^{\mathrm{o}}\boldsymbol{\beta}, \tilde{\boldsymbol{\Lambda}}_i^{\mathrm{oo}}, \nu)$ where $\tilde{\boldsymbol{\Lambda}}_i^{\mathrm{oo}} = \boldsymbol{O}_i\tilde{\boldsymbol{\Lambda}}_i\boldsymbol{O}_i^{\mathrm{T}}$. Treating $\{\tilde{\boldsymbol{y}}_i^{\mathrm{m}}, \tau_i\}_{i=1}^N$ as the latent data and then combining them with observed pseudo data $\tilde{\boldsymbol{y}}^{\mathrm{o}} = \{\tilde{\boldsymbol{y}}_i^{\mathrm{o}}\}_{i=1}^N$ as the complete data, we obtain the complete-data likelihood function. Multiplying this likelihood function by the joint prior $\pi(\boldsymbol{\theta})$

in (5) gives the modified *third* joint posterior, denoted by $p(\boldsymbol{\theta}, \tilde{\boldsymbol{y}}^{\mathrm{m}}, \boldsymbol{\tau}|\tilde{\boldsymbol{y}}^{\mathrm{o}})$. Subsequently, we can derive the full conditional posterior density for missing responses $\tilde{\boldsymbol{y}}_i^{\mathrm{m}}$ given observed responses $\tilde{\boldsymbol{y}}_i^{\mathrm{o}}$, scaling weight τ_i and parameters $\boldsymbol{\theta}$, denoted by $p(\tilde{\boldsymbol{y}}_i^{\mathrm{m}}|\tilde{\boldsymbol{y}}_i^{\mathrm{o}}, \tau_i, \boldsymbol{\theta})$. Multiplying it by $p(\tau_i|\tilde{\boldsymbol{y}}_i^{\mathrm{o}}, \boldsymbol{\theta})$ and then integrating out τ_i yields

(12)
$$[\tilde{\boldsymbol{y}}_i^{\mathrm{m}}|\tilde{\boldsymbol{y}}_i^{\mathrm{o}}, \boldsymbol{\theta}] \sim t_{n_i - n_i^{\mathrm{o}}}\left(\tilde{\boldsymbol{y}}_i^{\mathrm{m\cdot o}}, \left(\frac{\nu + \Delta_{\tilde{\boldsymbol{y}}_i^{\mathrm{o}}}}{\nu + n_i^{\mathrm{o}}}\right)\tilde{\boldsymbol{\Lambda}}_i^{\mathrm{mm.o}}, \nu + n_i^{\mathrm{o}}\right),$$

where $\tilde{\boldsymbol{y}}_i^{\mathrm{m\cdot o}} = \tilde{\boldsymbol{X}}_i^{\mathrm{m}}\boldsymbol{\beta} + \tilde{\boldsymbol{\Lambda}}_i^{\mathrm{mo}}\tilde{\boldsymbol{\Lambda}}_i^{\mathrm{oo}^{-1}}(\tilde{\boldsymbol{y}}_i^{\mathrm{o}} - \tilde{\boldsymbol{X}}_i^{\mathrm{o}}\boldsymbol{\beta})$, $\Delta_{\tilde{\boldsymbol{y}}_i^{\mathrm{o}}} = (\tilde{\boldsymbol{y}}_i^{\mathrm{o}} - \tilde{\boldsymbol{X}}_i^{\mathrm{o}}\boldsymbol{\beta})^{\mathrm{T}}\tilde{\boldsymbol{\Lambda}}_i^{\mathrm{oo}^{-1}}(\tilde{\boldsymbol{y}}_i^{\mathrm{o}} - \tilde{\boldsymbol{X}}_i^{\mathrm{o}}\boldsymbol{\beta})$, and $\tilde{\boldsymbol{\Lambda}}_i^{\mathrm{mm\cdot o}} = \boldsymbol{M}_i(\boldsymbol{I}_{n_i} - \tilde{\boldsymbol{\Lambda}}_i\boldsymbol{O}_i^{\mathrm{T}}\tilde{\boldsymbol{\Lambda}}_i^{\mathrm{oo}^{-1}}\boldsymbol{O}_i)\tilde{\boldsymbol{\Lambda}}_i\boldsymbol{M}_i^{\mathrm{T}}$ with $\tilde{\boldsymbol{\Lambda}}_i^{\mathrm{mo}} = \boldsymbol{M}_i\tilde{\boldsymbol{\Lambda}}_i\boldsymbol{O}_i^{\mathrm{T}}$. Having posterior samples of pseudo missing responses $\{\tilde{\boldsymbol{y}}_i^{\mathrm{m}^{(s)}}\}_{s=1}^{S}$ generated from (12), we transform them back to

$$\boldsymbol{y}_i^{\mathrm{m}^{(s)}} = \tilde{\boldsymbol{y}}_i^{\mathrm{m}^{(s)}} + \boldsymbol{\mu}_i(\boldsymbol{\beta}^{(s)}, \boldsymbol{b}_i^{(s)}) - \tilde{\boldsymbol{X}}_i^{\mathrm{m}}\boldsymbol{\beta}^{(s)} - \tilde{\boldsymbol{Z}}_i^{\mathrm{m}}\boldsymbol{b}_i^{(s)}. \quad (13)$$

After a sufficiently long 'burn-in' period, the Markov chain $\{\boldsymbol{y}_i^{\mathrm{m}^{(s)}}, s \geq s_0\}$ tends to converge to the posterior predictive distribution of missing responses $p(\boldsymbol{y}_i^{\mathrm{m}}|\boldsymbol{y}_i^{\mathrm{o}})$.

As the three-level hierarchy given in (10), its version of the partially observed data, which comprises the conditional distribution of $\tilde{\boldsymbol{y}}_i^{\mathrm{o}}|(\boldsymbol{b}_i, \tau_i)$, that of $\boldsymbol{b}_i|\tau_i$, and the marginal distribution of τ_i, can be determined. Hence the modified *fourth* joint posterior density $p(\boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{\tau}|\tilde{\boldsymbol{y}}^{\mathrm{o}})$ is obtained. Using the above four modified joint posterior densities, the full conditional posterior distributions for parameters and latent data are given in Appendix A.

As a consequence, the MCMC algorithm for Bayesian inference of the MtNLMM with missing responses proceeds as Algorithm 2 sequentially.

Algorithm 2. *(MCMC algorithm for the MtNLMM with missing responses)*

P-step for the 1st cycle. *Generate* $\boldsymbol{\phi} = (\phi_1, \phi_2)$ *and* ν *from (A.9) and (A.10), respectively, via the M-H algorithm.*

DA-step for the 2nd cycle. *Generate* τ_i *from (A.11) via the Gibbs sampler, for* $i = 1, \ldots, N$.

P-step for the 2nd cycle. *Generate* $\boldsymbol{D}$ *and* $\boldsymbol{\Sigma}$ *from (A.4) and (A.5), respectively, via the Gibbs sampler.*

DA-step for the 3rd cycle. *Generate* $\tilde{\boldsymbol{y}}_i^{\mathrm{m}}$ *from (12) via the Gibbs sampler, and then obtain the samples of* $\boldsymbol{y}_i^{\mathrm{m}}$ *based on (13), for* $i = 1, \ldots, N$.

P-step for the 3rd cycle. *Generate* $\boldsymbol{\beta}$ *from (A.12) via the Gibbs sampler.*

DA-step for the 4th cycle. *Generate* $\boldsymbol{b}_i$ *from (A.13) via the Gibbs sampler, for* $i = 1, \cdots, N$.

3.3 Predictive inference for future values

We now turn our attention to the predictive inference for future values of a new subject. First, let $\boldsymbol{Y}_{01}(s_{01} \times r)$ be an observed response matrix for a new subject over the first portion of time and $\boldsymbol{Y}_{02}(s_{02} \times r)$ the corresponding response matrix over the future portion of time. Suppose that $\boldsymbol{y}_0 = \mathrm{vec}([\boldsymbol{Y}_{01}^{\mathrm{T}} : \boldsymbol{Y}_{02}^{\mathrm{T}}]^{\mathrm{T}})$ satisfies the specification of model (1) in which the mean vector is defined as $\boldsymbol{\mu}_0(\boldsymbol{\beta}, \boldsymbol{b}_0)$, and $(\boldsymbol{b}_0^{\mathrm{T}}, \boldsymbol{e}_0^{\mathrm{T}})^{\mathrm{T}} \sim t_{q+n_{01}+n_{02}}(\boldsymbol{0}, \mathrm{diag}\{\boldsymbol{D}, \boldsymbol{R}_0\}, \nu)$, where $\boldsymbol{R}_0 = \boldsymbol{\Sigma} \otimes \boldsymbol{C}_0$. Define $\boldsymbol{y}_{01} = \mathrm{vec}(\boldsymbol{Y}_{01})$ and $\boldsymbol{y}_{02} = \mathrm{vec}(\boldsymbol{Y}_{02})$. To deal with missing values possibly existing in $\boldsymbol{y}_{01}$, two permutation matrices $\boldsymbol{O}_{01}$ and $\boldsymbol{O}_{02}$, which are of dimensions $n_{01}^{\mathrm{o}} \times (n_{01} + n_{02})$ and $n_{02} \times (n_{01} + n_{02})$ such that $\boldsymbol{O}_{01}\boldsymbol{y}_0 = \boldsymbol{y}_{01}^{\mathrm{o}}$ (observed values) and $\boldsymbol{O}_{02}\boldsymbol{y}_0 = \boldsymbol{y}_{02}$ (future values to be predicted), are introduced for ease of notation. Let $\tilde{\boldsymbol{\Lambda}}_{11}^{\mathrm{oo}} = \boldsymbol{O}_{01}\tilde{\boldsymbol{\Lambda}}_0\boldsymbol{O}_{01}^{\mathrm{T}}$, $\tilde{\boldsymbol{\Lambda}}_{12}^{\mathrm{o}} = \boldsymbol{O}_{01}\tilde{\boldsymbol{\Lambda}}_0\boldsymbol{O}_{02}^{\mathrm{T}}$, and $\tilde{\boldsymbol{\Lambda}}_{22} = \boldsymbol{O}_{02}\tilde{\boldsymbol{\Lambda}}_0\boldsymbol{O}_{02}^{\mathrm{T}}$, where $\tilde{\boldsymbol{\Lambda}}_0 = \tilde{\boldsymbol{Z}}_0\boldsymbol{D}\tilde{\boldsymbol{Z}}_0^{\mathrm{T}} + \boldsymbol{R}_0$. The pseudo-data model for $\boldsymbol{y}_0$ is

$$\begin{bmatrix}\tilde{\boldsymbol{y}}_{01}^{\mathrm{o}} \\ \tilde{\boldsymbol{y}}_{02}\end{bmatrix} \sim t_{n_{01}^{\mathrm{o}}+n_{02}}\left(\begin{bmatrix}\tilde{\boldsymbol{X}}_{01} \\ \tilde{\boldsymbol{X}}_{02}\end{bmatrix}\boldsymbol{\beta}, \begin{bmatrix}\tilde{\boldsymbol{\Lambda}}_{11}^{\mathrm{oo}} & \tilde{\boldsymbol{\Lambda}}_{12}^{\mathrm{o}} \\ \tilde{\boldsymbol{\Lambda}}_{12}^{\mathrm{o}^{\mathrm{T}}} & \tilde{\boldsymbol{\Lambda}}_{22}\end{bmatrix}, \nu\right).$$

Subsequently, we draw the posterior predictive inference on future pseudo responses $\tilde{\boldsymbol{y}}_{02}$ based on the observed pseudo responses $\tilde{\boldsymbol{y}}_{01}^{\mathrm{o}}$ (extracted the observed components from $\tilde{\boldsymbol{y}}_{01}$). Define $\boldsymbol{R}_{11} = \boldsymbol{O}_{01}\boldsymbol{R}_0\boldsymbol{O}_{01}^{\mathrm{T}}$, $\boldsymbol{R}_{21} = \boldsymbol{O}_{02}\boldsymbol{R}_0\boldsymbol{O}_{01}^{\mathrm{T}}$, $\boldsymbol{R}_{12} = \boldsymbol{R}_{21}^{\mathrm{T}}$, $\boldsymbol{R}_{22} = \boldsymbol{O}_{02}\boldsymbol{R}_0\boldsymbol{O}_{02}^{\mathrm{T}}$, $\boldsymbol{G} = \tilde{\boldsymbol{Z}}_{02} - \boldsymbol{R}_{21}\boldsymbol{R}_{11}^{-1}\tilde{\boldsymbol{Z}}_{01}$, $\tilde{\boldsymbol{Z}}_{01} = \boldsymbol{O}_{01}\tilde{\boldsymbol{Z}}_0$ and $\tilde{\boldsymbol{Z}}_{02} = \boldsymbol{O}_{02}\tilde{\boldsymbol{Z}}_0$. By Bayes' theorem, simple matrix algebra yields the conditional posterior distribution of $\tilde{\boldsymbol{y}}_{02}$, given by

(14)
$$[\tilde{\boldsymbol{y}}_{02}|\tilde{\boldsymbol{y}}_{01}^{\mathrm{o}}, \boldsymbol{\theta}] \sim t_{n_{02}}\left(\tilde{\boldsymbol{\mu}}_{2\cdot1}, \left(\frac{\nu + \tilde{\Delta}_{01}^{\mathrm{o}}}{\nu + n_{01}^{\mathrm{o}}}\right)\tilde{\boldsymbol{\Lambda}}_{22\cdot1}, n_{01}^{\mathrm{o}} + \nu\right),$$

where

$$\begin{aligned}
\tilde{\boldsymbol{\mu}}_{2\cdot1} &= \tilde{\boldsymbol{X}}_{02}\boldsymbol{\beta} + \tilde{\boldsymbol{Z}}_{02}\boldsymbol{b}_{2\cdot1} + \boldsymbol{R}_{21}\boldsymbol{R}_{11}^{-1}(\tilde{\boldsymbol{y}}_{01}^{\mathrm{o}} - \tilde{\boldsymbol{X}}_{01}\boldsymbol{\beta} - \tilde{\boldsymbol{Z}}_{01}\boldsymbol{b}_{2\cdot1}), \\
\tilde{\boldsymbol{\Lambda}}_{22\cdot1} &= (\boldsymbol{R}_{22} - \boldsymbol{R}_{21}\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}) \\
&\quad + \boldsymbol{G}(\boldsymbol{W}_{01} - \boldsymbol{W}_{01}(\boldsymbol{W}_{01} + \boldsymbol{D})^{-1}\boldsymbol{W}_{01})\boldsymbol{G}^{\mathrm{T}}, \\
\tilde{\Delta}_{01}^{\mathrm{o}} &= (\tilde{\boldsymbol{y}}_{01}^{\mathrm{o}} - \tilde{\boldsymbol{X}}_{01}\boldsymbol{\beta})^{\mathrm{T}}\tilde{\boldsymbol{\Lambda}}_{11}^{\mathrm{oo}^{-1}}(\tilde{\boldsymbol{y}}_{01}^{\mathrm{o}} - \tilde{\boldsymbol{X}}_{01}\boldsymbol{\beta})
\end{aligned}$$

with $\boldsymbol{b}_{2\cdot1} = \tilde{\boldsymbol{b}}_{01} - \boldsymbol{W}_{01}(\boldsymbol{W}_{01} + \boldsymbol{D})^{-1}\tilde{\boldsymbol{b}}_{01}$ in which $\tilde{\boldsymbol{b}}_{01} = \boldsymbol{W}_{01}\tilde{\boldsymbol{Z}}_{01}^{\mathrm{T}}\boldsymbol{R}_{11}^{-1}(\tilde{\boldsymbol{y}}_{01} - \tilde{\boldsymbol{X}}_{01}\boldsymbol{\beta})$, and $\boldsymbol{W}_{01} = (\tilde{\boldsymbol{Z}}_{01}^{\mathrm{T}}\boldsymbol{R}_{11}^{-1}\tilde{\boldsymbol{Z}}_{01})^{-1}$.

Finally, we generate posterior samples of pseudo future responses $\{\tilde{\boldsymbol{y}}_{02}^{(s)}\}_{s=1}^{S}$ from (14), in which the model parameters are replaced by their posterior samples $\{\boldsymbol{\theta}^{(s)}, s \geq s_0\}$. The future responses can be predicted as

$$\boldsymbol{y}_{02}^{(s)} = \tilde{\boldsymbol{y}}_{02}^{(s)} + \boldsymbol{\mu}_0(\hat{\boldsymbol{\eta}}_0, \boldsymbol{X}_{02}) - \tilde{\boldsymbol{X}}_{02}\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{Z}}_{02}\hat{\boldsymbol{b}}_{2\cdot1}, \quad (15)$$

where $\hat{\boldsymbol{\eta}}_0$, $\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{b}}_{2\cdot1}$ are the posterior means of $\boldsymbol{\eta}_0$, $\boldsymbol{\beta}$, and $\boldsymbol{b}_{2\cdot1}$, respectively. The Markov chain $\{\boldsymbol{y}_{02}^{(s)}, s \geq s_0\}$ tends to converge stationarily to the posterior predictive distribution of future responses $p(\boldsymbol{y}_{02}|\boldsymbol{y}_{01}^{\mathrm{o}})$.

4. APPLICATIONS

We apply our proposed method to analyze two real datasets. Both datasets are fitted by the MtNLMM and MNLMM with specific mean curves and various autocorrelation structures. For Bayesian model selection, we use the expected Akaike information criterion (Brooks, 2002, EAIC), the expected Bayesian information criterion (Brooks, 2002, EBIC), and the deviance information criterion (Spiegelhalter et al., 2002, DIC), defined as EAIC $= 2m + 2\overline{D(\boldsymbol{\theta})}$, EBIC $= m \log N + 2\overline{D(\boldsymbol{\theta})}$, and DIC $= 2\overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}})$, where m is the number of model parameters, $\overline{D(\boldsymbol{\theta})} = E^{\boldsymbol{\theta}|\tilde{\boldsymbol{y}}}[-2\ell(\boldsymbol{\theta}|\tilde{\boldsymbol{y}})]$ is the posterior expectation of the deviance, and $D(\bar{\boldsymbol{\theta}})$ is the deviance evaluated at the posterior means of parameters with $\ell(\boldsymbol{\theta}|\tilde{\boldsymbol{y}})$ being the observed log-likelihood function of $\boldsymbol{\theta}$ for the pseudo responses. Having the posterior samples $\{\boldsymbol{\theta}^{(l)}\}_{l=1}^{L}$, $\overline{D(\boldsymbol{\theta})}$ and $D(\bar{\boldsymbol{\theta}})$ can be approximated by $L^{-1}\sum_{l=1}^{L} D(\boldsymbol{\theta}^{(l)})$ and $D(L^{-1}\sum_{l=1}^{L}\boldsymbol{\theta}^{(l)})$, respectively. Models with smaller EAIC, EBIC, and DIC values are better supported by the data. Another commonly used criterion is the logarithm of pseudo-marginal likelihood (LPML), defined as

$$\text{LPML} = \sum_{i=1}^{N} \log(\widehat{\text{CPO}}_i),$$

where $\widehat{\text{CPO}}_i = \{L^{-1}\sum_{l=1}^{L} 1/\pi(\tilde{\boldsymbol{y}}_i|\boldsymbol{\theta}^{(l)})\}^{-1}$ is the Monte Carlo estimate of conditional predictive ordinate (CPO) statistic (Carlin and Louis, 2006) using a harmonic-mean approximation (Dey and Chang, 1997). Here the term $\pi(\tilde{\boldsymbol{y}}_i|\boldsymbol{\theta}^{(l)})$ is evaluated by the individual observed likelihood for the pseudo data $f(\tilde{\boldsymbol{y}}_i|\boldsymbol{\theta})$ at each posterior sample $\boldsymbol{\theta}^{(l)}$. Models with larger LPML values should be preferred.

4.1 ACTG 315 study

The first example concerns the ACTG 315 study developed by the Immunology Research Agenda Committee of the US National Institute of Allergy and Infectious Disease, the ACTG sponsor. In the study, 53 human immunodeficiency virus type 1 (HIV-1) infected patients (participants) were recruited by University Hospitals of Cleveland, Rush-Presbyterian-St. Luke's Medical Center and University of Colorado Health Science Center. After the recruitment and start of antiviral therapy (ART), patients were repeatedly measured their plasma HIV-1 RNA (viral load) copies and CD4$^+$ T cell counts at days 0, 2, 7, 10, 14, 28, 56, 84, 168 and 196. HIV-1 infection is associated with progressive and profound loss of immune function that places infected persons at enhanced risk for opportunistic infections, and even death. A reaction in HIV-1 related immune deficiency can be characterized by decreases in numbers of circulating CD4$^+$ T helper lymphocytes. In other words, CD4$^+$ T cells in blood decline to a lower level after HIV-1 infection and may recover to a high level after ART suppress viral load. During ART treatments, the virologic marker (measured by HIV-1 RNA) and the immunologic marker (measured by CD4$^+$ T cells) generally exhibit a negative correlation. Accordingly, a joint analysis of HIV-1 RNA and CD4$^+$ counts is helpful to take into account the evolution of the correlation among responses across occasions. For a more detailed account of the study, the reader is referred to Lederman et al. (1998) and Connick et al. (2000).

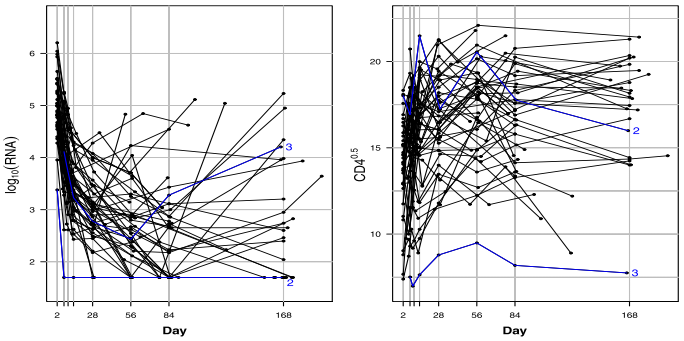

Figure 1. Trajectories of $\log_{10}$(RNA) *and CD4*$^{0.5}$ *for 48 HIV-1 infected patients.*

After excluding four early drop-out patients and one due to a plasma HIV-1 RNA pattern that suggested intermittent adherence to study therapy, a total of 48 patients as a part of the clinical trial on 53 patients were recruited in the later analysis. To stabilize the variances and reduce the strong skewness among the two makers, the HIV-1 RNA copies and CD4$^+$ T cells are transformed by a log-base-10 and a square-root function, respectively, which are widely used in HIV-AIDS clinical trials. The time trajectories of the two transformed markers, namely $\log_{10}$(RNA) and CD4$^{0.5}$, across visited days are shown in Figure 1. The data have been previously analyzed by a number of authors using different approaches, for example, Wu and Ding (1999), Liang et al. (2003), Wu and Liang (2004), Lachos et al. (2013), Lin and Wang (2013), and Wang (2015).

Let $y_{i1,k}$ and $y_{i2,k}$ be $\log_{10}$(RNA) and CD4$^{0.5}$ markers, respectively, at the kth occasion for patient i. Consider the MtNLMM with nonlinear mean functions for $y_{i1,k}$ and $y_{i2,k}$:

$$\begin{aligned} y_{i1,k} &= \log_{10}\left(e^{(\beta_1+b_{i1}+\beta_2 t_{ik})} + e^{\beta_3 \text{rna}_i}\right) + e_{i1,k}, \\ y_{i2,k} &= (\beta_4 + b_{i2})/(1 + e^{(\beta_5 - t_{ik})/\beta_6}) + e_{i2,k}, \end{aligned} \tag{16}$$

where $t_{ik} = \text{day}_{ik}/7$ is the kth visited time point (week) for patient i, and rna_i is the baseline $\log_{10}$-transformed RNA levels for patient i. The random effects and within-subject errors $(b_{i1}, b_{i2}, \boldsymbol{e}_{i1}^{\mathrm{T}}, \boldsymbol{e}_{i2}^{\mathrm{T}})^{\mathrm{T}}$, where $\boldsymbol{e}_{ij} = (e_{ij,1}, \ldots, e_{ij,s_i})^{\mathrm{T}}$, for $j = 1, 2$, are assumed to follow the multivariate-t distribution with zero location vector and scale-covariance matrix in block-diagonal form, *i.e.*, $\text{diag}\{\boldsymbol{D}, \boldsymbol{\Sigma} \otimes \boldsymbol{C}_i\}$. The considered autocorrelation functions on $\boldsymbol{C}_i$ include the uncorrelated structure (UNC), the continuous-type AR(1), and the DEC dependence. For the purpose of comparison, the normal counterparts of MtNLMMs are also fitted to the data.

Table 1. Model selection criteria under the six competing models for the ACTG 315 data

Criteria	MNLMM			MtNLMM		
	UNC	AR(1)	DEC	UNC	AR(1)	DEC
EAIC	2089.613	2038.391	2054.598	2113.664	2008.104	2047.705
EBIC	2112.068	2062.717	2080.795	2137.990	2034.301	2075.773
DIC	2075.034	2020.283	2034.591	2096.839	1991.061	2028.670
LPML	-1040.873	-1012.435	-1020.065	-952.615	-818.533	-881.764

We implement Algorithm 1 by running five parallel chains with 30,000 iterations per chain and different initial values extracted randomly from the prior distributions. The convergence is monitored through the use of multivariate potential scale reduction factor (Brooks and Gelman, 1998, MPSRF), suggesting that the convergence achieves after 20,000 iterations for all of the fitted models. We therefore discard the first 20,000 iterations as a 'burn-in' period for each chain and then store one imputed sample per 10 iterations to reduce the autocorrelation within each chain. We have a sample size of $L = 5{,}000$ realizations used to approximate the posterior distributions of interest.

Table 1 presents the EAIC, EBIC, DIC and LPML scores for six competing models. We found that the MtNLMMs provide better performance than their normal counterparts in terms of DIC and LPML. According to the EAIC and EBIC, except for the models with UNC errors, the MtNLMMs are superior to their normal counterparts. Overall, the best fit model is the MtNLMM with AR(1) errors. Table 2 reports the summary statistics for the posterior inference of parameters, including the posterior means (Mean), posterior standard deviations (SD), and posterior 2.5% and 97.5% quantiles that create the 95% posterior intervals, under the best model and the MNLMM-AR(1) which is the best one under the normal models. For the sake of comparison, the ML estimates (Est) of parameters, obtained by the pseudo expectation conditional maximization (ECM) algorithm (Wang and Lin, 2014), together with their standard errors (SE) and 95% confidence intervals, showing the lower and upper confidence limits (LCL and UCL), are also listed in Table 2.

We now turn our attention to the parameter estimates for the fitted MNLMM-AR(1) and MtNLMM-AR(1) from the Bayesian and ML approaches. It can be seen that the posterior and ML point estimates of model parameters show slight differences between the normal and t models, especially for the variance components, say the elements in $\boldsymbol{D}$, $\boldsymbol{\Sigma}$ and ϕ_1. From the estimates of $\boldsymbol{\Sigma}$, we can compute the correlations between $\log_{10}$(RNA) and $\text{CD4}^{0.5}$ levels. Under Bayesian and ML approaches respectively, the correlations are -0.245 and -0.134 for the fitted MNLMM-AR(1) and -0.138 and -0.149 for the fitted MtNLMM-AR(1), implying a negative relationship between the virologic marker (measured by HIV-1 RNA copies) and the immunologic marker (measured by CD4^+ T cells) during the ART treatments. This finding is consistent with the earlier studies by

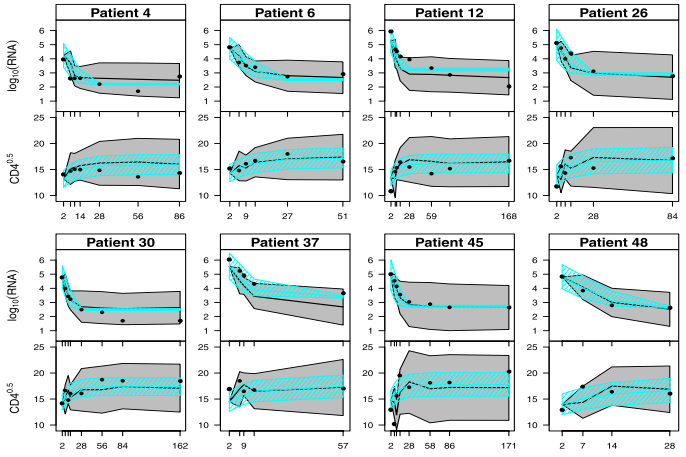

Figure 2. The $\log_{10}(RNA)$ and $CD4^{0.5}$ observations (•) along with the 95% posterior intervals for the mean curves (blue shadow slash) and the 95% predictive intervals for one-step ahead predictors (gray region) for eight randomly selected patients from the ACTG 315 study.

Orendi et al. (1998) and Sachsenberg et al. (1998). The autoregressive parameter ϕ_1 is also significantly different from zero, revealing the existence of autocorrelation among occasions on each characteristic. From the Bayesian viewpoint, the posterior median of ν is 13.527 with the 95% posterior interval $[5.958, 29.228]$, while from the ML inference, the point estimate of ν is 18.973 with the 95% confidence interval $[8.585, 35.799]$, indicating a moderate degree of fat tails. As can be seen, the parameters for the fixed effects are significantly different from zero, and the posterior standard deviations are quite smaller than the corresponding standard errors obtained by the inverse of observed information matrix. It implies that Bayesian approach gives more precise estimates than the ML inference in the study. In particular for the variance of random effects d_{11}, it reveals that the variation of random effects for $\log_{10}$(RNA) appears statistically significant under a Bayesian approach, while that under ML estimation does not.

As an illustration, Figure 2 exhibits the observations of $\log_{10}$(RNA) and $\text{CD4}^{0.5}$, the fitted values calculated by (11), and the one-step-ahead predictors calculated by (15) together with their 95% posterior (predictive) intervals for eight randomly selected patients. Generally, the fitted and predicted curves adapt the trend along the observed repeated measures. However, some of the features are not ideally captured because the viral load (RNA copies) and CD4^+ T cells are highly variable immune system markers which are sometimes difficult to fit. Apparently, the posterior bands for fitted values are narrower than those for one-step-ahead predictors.

Figure 3 displays the scatter plots of the posterior estimates of random effects (b_{i1}, b_{i2}) via (11) superimposed on a set of contour dashed lines obtained by the bivariate

Table 2. Posterior estimates obtained by the MCMC algorithm and ML estimates obtained by the pseudo ECM algorithm under the fitted MNLMM with AR(1) errors and MtNLMM with AR(1) errors, respectively, for the ACTG 315 data

Model	Method		Parameter													
			β_1	β_2	β_3	β_4	β_5	β_6	d_{11}	d_{21}	d_{22}	σ_{11}	σ_{21}	σ_{22}	ϕ_1	ν
N	Bayes	Mean	12.117	-2.634	1.306	16.879	-1.741	1.332	0.005	0.034	5.868	0.455	-0.283	2.941	0.155	–
		SD	0.182	0.144	0.018	0.272	0.338	0.206	0.007	0.126	1.283	0.036	0.082	0.300	0.033	–
		2.5%	11.766	-2.916	1.270	16.345	-2.404	0.927	0.001	-0.206	3.792	0.390	-0.445	2.430	0.096	–
		97.5%	12.473	-2.352	1.343	17.427	-1.075	1.731	0.024	0.295	8.806	0.527	-0.125	3.602	0.222	–
	ML	Est	12.048	-2.656	1.304	16.860	-1.731	1.308	0.006	0.016	4.754	0.463	-0.218	5.701	0.682	–
		SE	0.251	0.178	0.027	0.391	0.493	0.326	0.466	0.541	1.383	0.046	0.096	0.599	0.031	–
		LCL	11.556	-3.004	1.250	16.093	-2.698	0.669	-0.906	-1.045	2.044	0.374	-0.406	4.528	0.621	–
		UCL	12.540	-2.308	1.357	17.628	-0.765	1.947	0.918	1.076	7.464	0.552	-0.029	6.874	0.743	–
T	Bayes	Mean	11.945	-2.557	1.290	16.935	-1.825	1.329	2.584	0.069	1.945	0.410	-0.207	5.497	0.684	13.527
		SD	0.201	0.124	0.020	0.248	0.380	0.242	1.026	0.442	0.713	0.049	0.102	1.012	0.051	6.416
		2.5%	11.554	-2.798	1.252	16.453	-2.580	0.868	1.033	-0.794	0.876	0.322	-0.415	3.809	0.580	5.958
		97.5%	12.345	-2.310	1.328	17.429	-1.065	1.809	4.974	0.989	3.635	0.516	-0.013	7.705	0.772	29.228
	ML	Est	11.970	-2.593	1.291	16.916	-1.856	1.361	0.008	0.161	4.590	0.421	-0.218	5.080	0.684	18.973
		SE	0.247	0.171	0.027	0.392	0.526	0.341	0.434	0.515	1.367	0.049	0.090	0.620	0.032	8.585
		LCL	11.487	-2.928	1.238	16.148	-2.886	0.693	-0.844	-0.849	1.909	0.325	-0.394	3.865	0.621	2.147
		UCL	12.453	-2.258	1.344	17.684	-0.826	2.029	0.859	1.171	7.270	0.517	-0.041	6.296	0.747	35.799

* N: MNLMM with AR(1) errors; T: MtNLMM with AR(1) errors

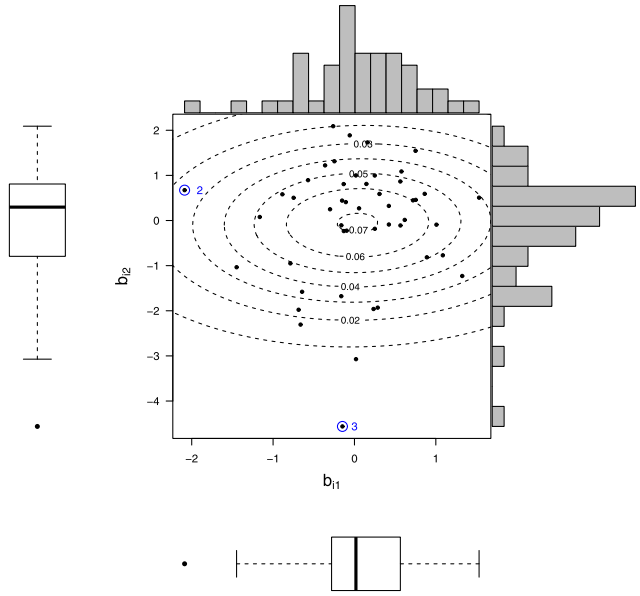

Figure 3. Scatter plot for posterior estimates of random effects (b_{i1}, b_{i2}) together with the bivariate t contour, histograms and boxplots.

t density under the fitted MtNLMM-AR(1), together with two summary histograms and boxplots. It reveals that the contour plot ideally adapts the shape of scattering pattern. From the histograms and boxplots, the heavy-tailed phenomena appear due to the extremely estimated $\hat{b}_{21}$ and $\hat{b}_{32}$ of patients 2 and 3, respectively, which can be treated as potential outliers. The two patients are pointed out in the scatter plot of estimated random effects and also marked in Figure 1. It is readily seen that the trajectory pattern of $\log_{10}$(RNA) of patient 2 and that of $\text{CD4}^{0.5}$ of patient 3 deviate systematically from other patients, confirming the extremely far-from-zero estimated $\hat{b}_{21}$ and $\hat{b}_{32}$.

4.2 Pregnant women data

The second example concerns a 2-year follow-up study of 124 women diagnosed with normal pregnancies and 37 women with abnormal pregnancies in a private fertilization obstetrics clinic in Santiago de Chile, Chile. The women were classified into the normal group if they had a terminal delivery, and the abnormal group if they had spontaneous abortions or other types of adverse pregnancy outcomes. To detect complications or a high risk of losing the fetus, the 161 young women were repeatedly measured their beta-subunit human chorionic gonadotropin (β-HCG) and estradiol concentrations during the first trimester of pregnancy. Figure 4 shows the trajectories in $\log_{10}$ scale for the β-HCG and estradiol across time (in days) for the pregnant women. The number of observations on each outcome measured at irregular occasions for the pregnant women ranges from 0 to 6. For normal and abnormal groups, there are around 2.61% and 0% missing β-HCG values, and 26.49% and 57.5% missing estradiol values, respectively. More detailed information about the clinical background can be found in Marshall et al. (2006) and Marshall et al. (2009).

Let $\boldsymbol{y}_{i1} = (y_{i1,1}, \ldots, y_{i1,s_i})^{\mathrm{T}}$ and $\boldsymbol{y}_{i2} = (y_{i2,1}, \ldots, y_{i2,s_i})^{\mathrm{T}}$ be the $\log_{10}(\beta$-HCG) and $\log_{10}$(Estradiol) for woman i, respectively, and define $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}^{\mathrm{T}}, \boldsymbol{y}_{i2}^{\mathrm{T}})^{\mathrm{T}}$. The data are fitted by the MtNLMM with two distinct mean curves for $\log_{10}(\beta$-HCG) and $\log_{10}$(Estradiol) of the ith woman at time t_{ik},

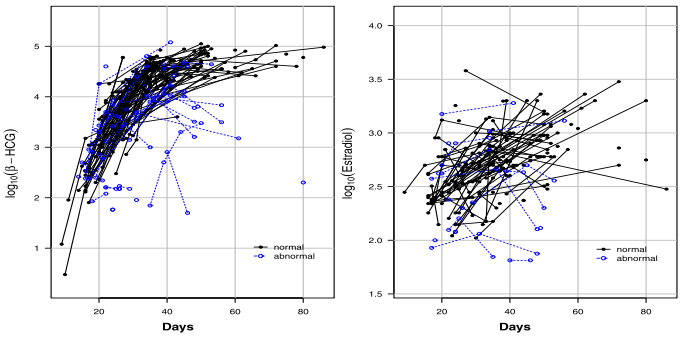

Figure 4. Trajectories of $\log_{10}(\beta\text{-HCG})$ *and* $\log_{10}$*(Estradiol) for 161 pregnant women dataset.*

Table 3. Model selection criteria under the six competing models for the pregnant women data

Criteria	MNLMM			MtNLMM		
	UNC	AR(1)	DEC	UNC	AR(1)	DEC
EAIC	651.559	616.333	661.298	485.150	488.181	496.992
EBIC	691.617	659.473	707.519	528.289	534.402	546.294
DIC	628.900	592.090	630.148	467.019	473.548	473.146
LPML	-331.880	-306.442	-334.314	-241.136	-291.027	-251.939

given by

$$(17)\quad \begin{aligned} y_{i1,k} &= \frac{\beta_1 + b_{i1} + \beta_2 \text{group}_i}{1 + e^{(\beta_3 - t_{ik})/\beta_4}} + e_{i1,k}, \\ y_{i2,k} &= \beta_5 + (\beta_6 + \beta_7 \text{group}_i) t_{ik} + b_{i2} + e_{i2,k}, \end{aligned}$$

where group_i is a group indicator (0 = normal group, 1 = abnormal group) used to investigate whether there exists any difference of the evolutions of responses between the normal and abnormal pregnant women. The random effects and within-subject errors are assumed to be $\boldsymbol{b}_i = (b_{i1}, b_{i2}) \sim t_2(\boldsymbol{0}, \boldsymbol{D}, \nu)$ and $\boldsymbol{e}_i = (\boldsymbol{e}_{i1}^{\mathrm{T}}, \boldsymbol{e}_{i2}^{\mathrm{T}})^{\mathrm{T}} \sim t_{2s_i}(\boldsymbol{0}, \boldsymbol{R}_i, \nu)$, where $\boldsymbol{R}_i = \boldsymbol{\Sigma} \otimes \boldsymbol{C}_i$, and the UNC, continuous-type AR(1), and DEC structures are considered for $\boldsymbol{C}_i$. The normal counterparts with mean functions specified in (17), where $\boldsymbol{b}_i = (b_{i1}, b_{i2}) \sim \mathcal{N}_2(\boldsymbol{0}, \boldsymbol{D})$ and $\boldsymbol{e}_i = (\boldsymbol{e}_{i1}^{\mathrm{T}}, \boldsymbol{e}_{i2}^{\mathrm{T}})^{\mathrm{T}} \sim \mathcal{N}_{2s_i}(\boldsymbol{0}, \boldsymbol{R}_i)$, are also fitted to the data for comparison purpose.

To conduct the Bayesian computation, we perform Algorithm 2 using five parallel chains with 20,000 iterations per chain and random initialization selected from the priors. After diagnosing the convergence via the MPSRF, we delete the first 16,000 iterations as a 'burn-in' period for each chain and then store one imputed sample per 20 iterations to reduce the autocorrelation within each chain. As a result, there are $L = 1{,}000$ realizations used to draw the desired posterior inference.

Observing the model selection criteria listed in Table 3, all MtNLMMs outperform over their normal counterparts. The preferred models under the normal and t scenarios are MNLMM-AR(1) and MtNLMM-UNC, respectively, and the later model is better than the former. Table 4 shows the point and interval estimates for all parameters of interest obtained from the Bayesian and ML methods under the best chosen normal and t models. Focusing on the posterior inference, all the estimates of fixed effects are significantly different from zero because the corresponding posterior intervals do not contain zero. The estimates of β_2 and β_7 suggest that the change in responses over time exists significantly between the normal and abnormal groups. We compute the estimates of correlations of $\log_{10}(\beta\text{-HCG})$ and $\log_{10}(\text{Estradiol})$, given by $\sigma_{21}/\sqrt{\sigma_{11}\sigma_{22}} = 0.677$ and 0.163 under the best MNLMM and MtNLMM, respectively, and they are somewhat different. Furthermore, the estimates of AR parameter $\phi_1 = 0.723$ are highly significant but only under the MNLMM. The posterior median of the DOF $\nu = 3.010$ is quite small, confirming the presence of a heavy-tailed behavior among the random effects and errors. When comparing the ML and Bayesian inference, the posterior means of MCMC samples are quite similar to ML estimates for fixed effects, whereas the Bayesian estimation provides smaller variabilities, leading to a shorter range for the resulting interval estimate. On the other hand, the two methods usually give somewhat different estimates for variance components ($\boldsymbol{D}$ and $\boldsymbol{\Sigma}$), as was the case of univariate normal/t linear mixed models (Browne and Drapery, 2006; Lin and Lee, 2007).

As an illustration, Figure 5 displays the observed responses, their fitted curves, and the imputed values along with the 95% posterior intervals of missing responses under the best model for six pregnant women who had at least four measurement occasions and more than one missing β-HCG and/or estradiol concentrations. As anticipated, the fitted curves are almost close to the observations, and the imputed values show a reasonable pattern. The 95% posterior intervals give more precise results (narrower intervals) when the woman had more observed measurements.

5. SIMULATIONS

We conduct simulations to examine the performance of MNLMM and MtNLMM approaches under conditions resulting from the fully observed data and two missing mechanisms. As suggested by one referee, we also compare the accuracy of Bayesian and ML estimation methods for parameter recovery. The data are generated from a similar model defined in (16) with the presumed model parameters, given by $\beta_1 = 12$, $\beta_2 = -3$, $\beta_3 = 1$, $\beta_4 = 17$, $\beta_5 = -2$, $\beta_6 = 1$,

$$\boldsymbol{D} = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & -0.75 \\ -0.75 & 1 \end{bmatrix}, \quad \boldsymbol{C}_i = \boldsymbol{I}_{10},$$

and $\nu = 5$. For each 'simulated' subject, the time points t_{ik} are set as the values 0, 2, 7, 10, 14, 28, 56, 84, 168 and 196 (days) divided by 7 (week) as the temporal scales used in the ACTG 315 example.

Table 4. Posterior estimates obtained by the MCMC algorithm and ML estimates obtained by the pseudo ECM algorithm under the fitted MNLMM with AR(1) errors and MtNLMM with UNC errors, respectively, the pregnant women data

Model	Method		Parameter														
			β_1	β_2	β_3	β_4	β_5	β_6	β_7	d_{11}	d_{21}	d_{22}	σ_{11}	σ_{21}	σ_{22}	ϕ_1	ν
N	Bayes	Mean	4.673	-0.767	14.683	7.285	2.336	0.011	-0.008	0.010	0.006	0.004	0.494	0.173	0.132	0.723	–
		SD	0.046	0.070	0.387	0.456	0.041	0.001	0.001	0.001	0.001	<0.000	0.066	0.032	0.022	0.068	–
		2.5%	4.581	-0.899	13.937	6.333	2.259	0.009	-0.010	0.008	0.005	0.003	0.383	0.117	0.095	0.588	–
		97.5%	4.761	-0.628	15.464	8.156	2.415	0.013	-0.005	0.013	0.008	0.005	0.642	0.237	0.176	0.810	–
	ML	Est	4.680	-0.787	14.853	7.194	2.332	0.011	-0.008	0.041	0.020	0.017	0.201	0.038	0.074	0.854	–
		SE	0.059	0.086	0.483	0.574	0.055	0.001	0.002	0.032	0.013	0.011	0.028	0.012	0.011	0.033	–
		LCL	4.565	-0.955	13.912	6.074	2.224	0.008	-0.011	-0.022	-0.005	-0.004	0.147	0.015	0.052	0.790	–
		UCL	4.795	-0.618	15.794	8.314	2.440	0.014	-0.004	0.104	0.045	0.038	0.254	0.061	0.095	0.918	–
T	Bayes	Mean	4.713	-0.601	15.315	6.781	2.271	0.013	-0.006	0.035	0.006	0.027	0.067	0.009	0.045	–	3.010
		SD	0.028	0.049	0.207	0.262	0.032	0.001	0.001	0.008	0.003	0.005	0.008	0.004	0.006	–	0.984
		2.5%	4.659	-0.697	14.881	6.268	2.206	0.011	-0.009	0.022	0.001	0.018	0.053	0.001	0.035	–	2.000
		97.5%	4.768	-0.510	15.695	7.284	2.333	0.015	-0.004	0.055	0.014	0.039	0.084	0.018	0.058	–	5.000
	ML	Est	4.704	-0.596	15.283	6.825	2.275	0.013	-0.006	0.039	0.010	0.015	0.082	0.014	0.039	–	3.574
		SE	0.045	0.067	0.315	0.396	0.046	0.001	0.002	0.013	0.007	0.006	0.012	0.006	0.007	–	0.643
		LCL	4.616	-0.726	14.668	6.053	2.185	0.010	-0.009	0.013	-0.003	0.004	0.059	0.002	0.026	–	2.321
		UCL	4.791	-0.466	15.897	7.598	2.364	0.015	-0.003	0.065	0.023	0.027	0.106	0.025	0.052	–	4.828

* N: MNLMM with AR(1) errors; T: MtNLMM with UNC errors

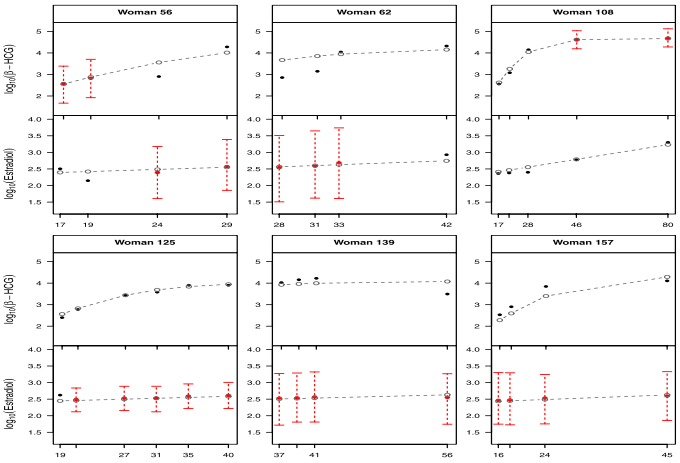

Figure 5. The $\log_{10}(\beta$-HCG) and $\log_{10}$(Estradiol) observations (•), the fitted curves (gray dashed lines), and the imputed values (red dots) together with the 95% posterior intervals (red dashed lines) for missing responses for six selected pregnant women.

Once a sample of complete data is generated, artificial missing values are introduced to the sample under MCAR and MAR mechanisms. In the MCAR experiments, 15% sub-samples of values are randomly selected and then deleted. As far as the MAR mechanism is concerned, nonresponse items are generated for each outcome variable $\boldsymbol{y}_j$, $j = 1, 2$, depending on the value of y_{jk}, which is the observed value of $\boldsymbol{y}_j$ at time k, for $k = 1, \ldots, 10$. Let q_{ji} be the ith quartile of the empirical distribution of $\boldsymbol{y}_j$. The nonresponse probabilities for $\boldsymbol{y}_j$ are 1% if $y_{jk} < q_{j1}$, 5% if $q_{j1} \le y_{jk} < q_{j2}$, 10% if $q_{j2} \le y_{jk} < q_{j3}$, and 15% if $y_{jk} \ge q_{j3}$.

Table 5. Model selection under various sample sizes and missing proportions based on 100 replications

Missingness	Criteria		$N = 25$		$N = 50$	
			MNLMM	MtNLMM	MNLMM	MtNLMM
Full Data	DIC	Mean	1556.352	1473.348	3160.725	2981.424
		Freq	(1)	(99)	(0)	(100)
	LPML	Mean	-812.589	-748.355	-1643.631	-1476.624
		Freq	(0)	(100)	(0)	(100)
MCAR	DIC	Mean	1454.250	1377.558	2926.504	2833.399
		Freq	(11)	(89)	(3)	(97)
	LPML	Mean	-753.870	-712.742	-1504.381	-1432.803
		Freq	(3)	(97)	(2)	(98)
MAR	DIC	Mean	1520.686	1491.557	3084.063	3021.791
		Freq	(39)	(61)	(24)	(76)
	LPML	Mean	-801.570	-776.133	-1600.044	-1550.067
		Freq	(16)	(84)	(24)	(76)

For each simulated data set, the MNLMM and MtNLMM approaches are fitted using the MCMC procedure described in Section 3 and the ECM algorithm of Wang and Lin (2014) to draw inference from both Bayesian and ML perspectives. A total of 100 replications are undertaken across sample sizes $N = 25$ and 50 for each of the fully observed and two missing data patterns. One thing worth noting is that the ECM procedure for ML estimation may fail, though seldom, to achieve convergence for a particularly simulated missing pattern. To ensure that the comparison among different methods is evaluated based on the same simulated data, an additional data set is regenerated using the R `try()` function in the case of non-convergence.

Table 5 lists the averages of DIC and LPML together with the frequencies (Freq) of the models supported by DIC and LPML. Observing the table, the MtNLMM leads to smaller DIC and larger LPML values. Moreover, the frequency of

Table 6. Comparison of parameter estimation obtained by the Bayesian MCMC and pseudo ECM-based ML methods for N =25 and 50 under full data, MCAR and MAR mechanisms

Missingness	Parameter	True value	$N=25$ Bayes Est (Bias)	$N=25$ Bayes 2.5% 97.5% (Length)	$N=25$ ML Est (Bias)	$N=25$ ML LCI UCI (Length)	$N=50$ Bayes Est (Bias)	$N=50$ Bayes 2.5% 97.5% (Length)	$N=50$ ML Est (Bias)	$N=50$ ML LCI UCI (Length)
Full Data	β_1	12	11.958 (-0.042)	11.396 12.520 (1.125)	11.922 (-0.078)	11.150 12.694 (1.545)	11.996 (-0.004)	11.586 12.403 (0.816)	12.009 (0.009)	11.450 12.567 (1.116)
	β_2	-3	-2.954 (0.046)	-3.384 -2.529 (0.856)	-2.920 (0.080)	-3.514 -2.327 (1.187)	-2.948 (0.052)	-3.252 -2.645 (0.607)	-2.914 (0.086)	-3.337 -2.490 (0.847)
	β_3	1	0.997 (-0.003)	0.940 1.055 (0.116)	0.993 (-0.007)	0.912 1.074 (0.162)	1.000 (< 0.000)	0.959 1.041 (0.082)	0.998 (-0.002)	0.940 1.056 (0.116)
	β_4	17	17.002 (0.002)	16.675 17.329 (0.654)	17.003 (0.003)	16.578 17.427 (0.849)	17.001 (0.001)	16.759 17.244 (0.485)	16.997 (-0.003)	16.683 17.311 (0.627)
	β_5	-2	-2.039 (-0.039)	-2.677 -1.412 (1.265)	-2.102 (-0.102)	-2.992 -1.213 (1.779)	-2.066 (-0.066)	-2.516 -1.624 (0.892)	-2.135 (-0.135)	-2.771 -1.499 (1.272)
	β_6	1	1.015 (0.015)	0.728 1.304 (0.575)	1.042 (0.042)	0.637 1.446 (0.809)	1.031 (0.031)	0.827 1.237 (0.410)	1.064 (0.064)	0.774 1.355 (0.581)
MCAR	β_1	12	11.968 (-0.032)	11.362 12.582 (1.220)	11.921 (-0.079)	11.113 12.730 (1.617)	12.007 (0.007)	11.559 12.456 (0.897)	12.030 (0.030)	11.445 12.615 (1.169)
	β_2	-3	-2.962 (0.038)	-3.464 -2.459 (1.005)	-2.923 (0.077)	-3.563 -2.283 (1.280)	-2.954 (0.046)	-3.307 -2.600 (0.707)	-2.927 (0.073)	-3.384 -2.469 (0.915)
	β_3	1	0.998 (-0.002)	0.934 1.064 (0.130)	0.992 (-0.008)	0.906 1.077 (0.170)	1.001 (0.001)	0.954 1.047 (0.092)	0.999 (-0.001)	0.938 1.060 (0.122)
	β_4	17	17.001 (0.001)	16.674 17.331 (0.656)	17.004 (0.004)	16.575 17.434 (0.859)	16.999 (-0.001)	16.753 17.246 (0.493)	17.000 (< 0.000)	16.681 17.318 (0.637)
	β_5	-2	-1.994 (0.006)	-2.805 -1.159 (1.647)	-2.093 (-0.093)	-3.059 -1.128 (1.931)	-2.050 (-0.050)	-2.599 -1.508 (1.090)	-2.128 (-0.128)	-2.815 -1.442 (1.373)
	β_6	1	0.995 (-0.005)	0.619 1.364 (0.744)	1.037 (0.037)	0.598 1.475 (0.877)	1.022 (0.022)	0.775 1.272 (0.497)	1.062 (0.062)	0.748 1.375 (0.627)
MAR	β_1	12	11.954 (-0.046)	11.344 12.556 (1.212)	11.909 (-0.091)	11.110 12.708 (1.597)	11.986 (-0.014)	11.549 12.418 (0.869)	11.988 (-0.012)	11.410 12.565 (1.155)
	β_2	-3	-2.969 (0.031)	-3.449 -2.486 (0.963)	-2.938 (0.062)	-3.557 -2.318 (1.240)	-2.959 (0.041)	-3.294 -2.624 (0.670)	-2.929 (0.071)	-3.372 -2.487 (0.886)
	β_3	1	0.994 (-0.006)	0.931 1.057 (0.125)	0.984 (-0.016)	0.902 1.066 (0.164)	0.995 (-0.005)	0.951 1.040 (0.089)	0.990 (-0.010)	0.932 1.049 (0.118)
	β_4	17	16.981 (-0.019)	16.653 17.312 (0.658)	16.984 (-0.016)	16.555 17.412 (0.857)	16.984 (-0.016)	16.740 17.229 (0.489)	16.978 (-0.022)	16.661 17.295 (0.634)
	β_5	-2	-2.015 (-0.015)	-2.776 -1.262 (1.513)	-2.109 (-0.109)	-3.029 -1.188 (1.841)	-2.059 (-0.059)	-2.571 -1.549 (1.023)	-2.151 (-0.151)	-2.820 -1.483 (1.337)
	β_6	1	1.007 (0.007)	0.664 1.350 (0.686)	1.047 (0.047)	0.627 1.467 (0.839)	1.027 (0.027)	0.794 1.263 (0.469)	1.070 (0.070)	0.765 1.376 (0.611)

MtNLMM supported by the two criteria becomes larger as the sample size N increases. The results apparently reveal the benefit of using the t distribution to handle data sets with outliers.

When comparing the performance between the Bayesian and ML methods, we focus on the estimation accuracies of fixed effects. Table 6 shows the estimation results, including averages of posterior means and ML estimates (Est) of fixed effects, together with the corresponding 95% posterior intervals (2.5%, 97.5%), the 95% confidence intervals (UCL, LCL), average differences between the estimators and the true values (Bias, in parentheses), and average lengths of posterior (confidence) intervals (Length, in parentheses) over 100 replicates. Judging from this table, both Bayesian and ML methods can provide reliable estimates of model parameters because the posterior (confidence) intervals absolutely contain the true values of corresponding parameters and the biases are quite small. In almost all cases, the Bayes estimates obtained as the mean of posterior samples show relatively smaller biases as compared to the ML estimates. Moreover, the Bayesian credible intervals created by the 2.5% and 97.5% quantiles of the relevant posterior distributions are shorter than the 95% confidence intervals through the large-sample normal approximation. As anticipated, the lengths of posterior (confidence) intervals decrease as the sample size increases. In summary, this simulation confirms

that the proposed MCMC scheme actually outperforms over the ECM algorithm in providing more precise inference for fixed effects.

6. CONCLUSION

We have presented a fully Bayesian treatment of the MtNLMM with a parsimonious DEC structure accounting for the dependence of within-subject errors. Several Bayesian hierarchical structures of the MtNLMM are constructed by using the first-order Taylor expansion to linearize the model and incorporating the weakly informative prior distributions for model parameters, which lead to full conditional posteriors following standard distributions. The proposed MCMC algorithms vary the specification of joint posteriors alternately and thus provide great efficiency in carrying out Bayesian posterior inference. Besides, the predictive inference for future responses and missing values can be drawn based on the corresponding posterior distributions with parameters replaced by their posterior estimates. Numerical results demonstrate that the proposed method performs reasonably well for the simulated and real datasets.

In the case where MAR with ignorability is not realistic, the relationship between the unobserved measurements and missingness process should be further investigated. Therefore, it would be interesting to generalize the Bayesian MtNLMM under non-ignorable missingness.

ACKNOWLEDGMENTS

The authors would like to express their deepest gratitude to the Co-editor, the Associate editor and the reviewer for their insightful comments and suggestions that greatly improved this paper. W.L. Wang acknowledges the support of the Ministry of Science and Technology of Taiwan under Grant no. MOST 105-2118-M-035-004-MY2. L.M. Castro acknowledges support from Grant FONDECYT 1170258 and CONICYT-Chile through BASAL project CMM, Universidad de Chile.

APPENDIX A. FULL CONDITIONAL POSTERIOR DISTRIBUTIONS

We state the full conditional posterior distributions for model parameters as well as the latent variables under both fully and partially observed-data frameworks of the MtNLMM. From the *first* joint posterior given in (6), the full conditional posterior densities of $\boldsymbol{\phi}$ and ν are

(A.1) $p(\boldsymbol{\phi}|\boldsymbol{y},\boldsymbol{b},\boldsymbol{\theta}_{(-\boldsymbol{\phi})})$

$$\propto \prod_{i=1}^{N} |\boldsymbol{C}_i(\boldsymbol{\phi})|^{-r/2}\Big(1+\frac{\Delta_{\boldsymbol{y}_i}}{\nu}\Big)^{-(\nu+n_i)/2}(1+\phi_2)^{-2},$$

and

(A.2) $p(\nu|\boldsymbol{y},\boldsymbol{b},\boldsymbol{\theta}_{(-\nu)})$

$$\propto \prod_{i=1}^{N} \frac{\Gamma(\frac{\nu+n_i}{2})}{\Gamma(\frac{\nu}{2})(\pi\nu)^{n_i/2}}\Big(1+\frac{\Delta_{\boldsymbol{y}_i}}{\nu}\Big)^{-(\nu+n_i)/2}(1+\nu)^{-2},$$

respectively, where $\Delta_{\boldsymbol{y}_i} = \big(\boldsymbol{y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta},\boldsymbol{b}_i)\big)^{\mathrm{T}}\boldsymbol{R}_i^{-1}\big(\boldsymbol{y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta},\boldsymbol{b}_i)\big)$. The symbol "$\boldsymbol{\theta}_{(-\boldsymbol{\alpha})}$" comprises all parameters contained in $\boldsymbol{\theta}$ except for $\boldsymbol{\alpha}$. From the *second* joint posterior given in (7), we obtain the full conditional posterior distributions of τ_i, $\boldsymbol{D}$ and $\boldsymbol{\Sigma}$, which show the standard distributions and are given by

(A.3) $\big[\tau_i|\boldsymbol{y}_i,\boldsymbol{b}_i,\boldsymbol{\theta}\big]$

$$\sim \text{Gamma}\left(\frac{n_i+q+\nu}{2},\frac{\Delta_{\boldsymbol{y}_i}+\Delta_{\boldsymbol{b}_i}+\nu}{2}\right),$$

(A.4) $\big[\boldsymbol{D}|\boldsymbol{y},\boldsymbol{b},\boldsymbol{\tau},\boldsymbol{\theta}_{(-\boldsymbol{D})}\big]$

$$\sim \mathcal{IW}\big(N+d_0,\sum_{i=1}^{N}\tau_i\boldsymbol{b}_i\boldsymbol{b}_i^{\mathrm{T}}+\boldsymbol{G}_0\big),$$

(A.5) $\big[\boldsymbol{\Sigma}|\boldsymbol{y},\boldsymbol{b},\boldsymbol{\tau},\boldsymbol{\theta}_{(-\boldsymbol{\Sigma})}\big]$

$$\sim \mathcal{IW}\big(\sum_{i=1}^{N}s_i+s_0,\sum_{i=1}^{N}\boldsymbol{\psi}_i(\boldsymbol{\beta},\boldsymbol{b}_i)+\mathbf{H}_0\big),$$

where $\Delta_{\boldsymbol{b}_i}=\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{D}^{-1}\boldsymbol{b}_i$ and $\boldsymbol{\psi}_i=\big[\mathrm{tr}\big(\boldsymbol{C}_i^{-1}(\boldsymbol{\phi})E_{ijl}\big)\big]$ with

$$E_{ijl}=\tau_i\big(\boldsymbol{y}_{ij}-\boldsymbol{\mu}_{ij}(\boldsymbol{\beta},\boldsymbol{b}_{ij})\big)\big(\boldsymbol{y}_{il}-\boldsymbol{\mu}_{il}(\boldsymbol{\beta},\boldsymbol{b}_{il})\big)^{\mathrm{T}},$$

for $j,l=1,2,\cdots,r$. Besides, the full conditional posterior distribution of $\boldsymbol{\beta}$ drawn from the *third* joint posterior $p(\boldsymbol{\theta},\boldsymbol{\tau}|\tilde{\boldsymbol{y}})$ is

(A.6)

$$\big[\boldsymbol{\beta}|\tilde{\boldsymbol{y}},\boldsymbol{\tau},\boldsymbol{\theta}_{(-\boldsymbol{\beta})}\big]\sim\mathcal{N}_p\Big(\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\big(\sum_{i=1}^{N}\tau_i\tilde{\boldsymbol{X}}_i^{\mathrm{T}}\tilde{\boldsymbol{\Lambda}}_i^{-1}\tilde{\boldsymbol{y}}_i+\boldsymbol{F}_0^{-1}\boldsymbol{\beta}_0\big),\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\Big),$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}=\big(\sum_{i=1}^{N}\tau_i\tilde{\boldsymbol{X}}_i^{\mathrm{T}}\tilde{\boldsymbol{\Lambda}}_i^{-1}\tilde{\boldsymbol{X}}_i+\boldsymbol{F}_0^{-1}\big)^{-1}$. From the *fourth* joint posterior $p(\boldsymbol{\theta},\boldsymbol{b},\boldsymbol{\tau}|\tilde{\boldsymbol{y}})$, the full conditional posterior distribution of $\boldsymbol{b}_i$ can be derived and given by

(A.7) $\big[\boldsymbol{b}_i|\tilde{\boldsymbol{y}}_i,\tau_i,\boldsymbol{\theta}\big]$

$$\sim\mathcal{N}_q\left(\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i}\tilde{\boldsymbol{Z}}_i^{\mathrm{T}}\boldsymbol{R}_i^{-1}(\tilde{\boldsymbol{y}}_i-\tilde{\boldsymbol{X}}_i\boldsymbol{\beta}),\tau_i^{-1}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i}\right),$$

where $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i}=\big(\tilde{\boldsymbol{Z}}_i^{\mathrm{T}}\boldsymbol{R}_i^{-1}\tilde{\boldsymbol{Z}}_i+\boldsymbol{D}^{-1}\big)^{-1}$. It follows straightforwardly from (A.7) that

(A.8)

$$\big[\boldsymbol{b}_i|\tilde{\boldsymbol{y}}_i,\boldsymbol{\theta}\big]\sim t_q\Big(\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i}\tilde{\boldsymbol{Z}}_i^{\mathrm{T}}\boldsymbol{R}_i^{-1}(\tilde{\boldsymbol{y}}_i-\tilde{\boldsymbol{X}}_i\boldsymbol{\beta}),\Big(\frac{\nu+\Delta_{\tilde{\boldsymbol{y}}_i}}{\nu+n_i}\Big)\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i},\nu+n_i\Big).$$

Under the partially observed-data framework of the MtNLMM as described in Section 3.2, the required full conditional posterior distributions are stated as follows. Using $p(\boldsymbol{\theta},\boldsymbol{b}|\boldsymbol{y}^{\mathrm{o}})$, the conditional posterior densities of $\boldsymbol{\phi}$ and ν are

(A.9) $p(\boldsymbol{\phi}|\boldsymbol{y}^{\mathrm{o}},\boldsymbol{b},\boldsymbol{\theta}_{(-\boldsymbol{\phi})})$

$$\propto \prod_{i=1}^{N} |\boldsymbol{R}_i^{\mathrm{oo}}|^{-1/2}\Big(1+\frac{\Delta_{\boldsymbol{y}_i^{\mathrm{o}}}}{\nu}\Big)^{-(\nu+n_i^{\mathrm{o}})/2}(1+\phi_2)^{-2},$$

(A.10) $$p(\nu|\boldsymbol{y}^{\mathrm{o}},\boldsymbol{b},\boldsymbol{\theta}_{(-\nu)})$$
$$\propto \prod_{i=1}^{N} \frac{\Gamma(\frac{\nu+n_i^{\mathrm{o}}}{2})}{\Gamma(\frac{\nu}{2})(\pi\nu)^{n_i^{\mathrm{o}}/2}} \times \Big(1+\frac{\Delta_{\boldsymbol{y}_i^{\mathrm{o}}}}{\nu}\Big)^{-(\nu+n_i^{\mathrm{o}})/2}(1+\nu)^{-2}.$$

Using $p(\boldsymbol{\theta},\boldsymbol{b},\boldsymbol{\tau}|\boldsymbol{y}^{\mathrm{o}})$, the full conditional posterior distribution of τ_i is

(A.11) $$\big[\tau_i|\boldsymbol{y}_i^{\mathrm{o}},\boldsymbol{b}_i,\boldsymbol{\theta}\big] \sim \mathrm{Gamma}\left(\frac{n_i^{\mathrm{o}}+q+\nu}{2},\frac{\Delta_{\boldsymbol{y}_i^{\mathrm{o}}}+\Delta_{\boldsymbol{b}_i}+\nu}{2}\right),$$

where $\Delta_{\boldsymbol{y}_i^{\mathrm{o}}}=(\boldsymbol{y}_i^{\mathrm{o}}-\boldsymbol{\mu}_i^{\mathrm{o}}(\boldsymbol{\beta},\boldsymbol{b}_i))^{\mathrm{T}}\boldsymbol{R}_i^{\mathrm{oo}^{-1}}(\boldsymbol{y}_i^{\mathrm{o}}-\boldsymbol{\mu}_i^{\mathrm{o}}(\boldsymbol{\beta},\boldsymbol{b}_i))$ with $\boldsymbol{\mu}_i^{\mathrm{o}}(\boldsymbol{\beta},\boldsymbol{b}_i)=\boldsymbol{O}_i\boldsymbol{\mu}_i(\boldsymbol{\beta},\boldsymbol{b}_i)$ and $\boldsymbol{R}_i^{\mathrm{oo}}=\boldsymbol{O}_i\boldsymbol{R}_i\boldsymbol{O}_i^{\mathrm{T}}$. Similarly, the full conditional posterior distributions of $\boldsymbol{D}$ and $\boldsymbol{\Sigma}$ are expressed as (A.4) and (A.5), respectively, in which the term $\boldsymbol{y}_i$ should be replaced by $\boldsymbol{O}_i^{\mathrm{T}}\boldsymbol{y}_i^{\mathrm{o}}+\boldsymbol{M}_i^{\mathrm{T}}\boldsymbol{y}_i^{\mathrm{m}^{(s)}}$, where $\boldsymbol{y}_i^{\mathrm{m}^{(s)}}$ are the generated posterior samples of missing responses via (13). In addition, it follows from the modified *third* and *fourth* joint posterior densities that

(A.12) $$\big[\boldsymbol{\beta}|\tilde{\boldsymbol{y}}^{\mathrm{o}},\boldsymbol{\tau},\boldsymbol{\theta}_{(-\boldsymbol{\beta})}\big] \sim \mathcal{N}_p\Big(\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{\mathrm{o}}\big(\sum_{i=1}^{N}\tau_i\tilde{\boldsymbol{X}}_i^{\mathrm{o}^{\mathrm{T}}}\tilde{\boldsymbol{\Lambda}}_i^{\mathrm{oo}^{-1}}\tilde{\boldsymbol{y}}_i^{\mathrm{o}}+\boldsymbol{F}_0^{-1}\boldsymbol{\beta}_0\big),\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{\mathrm{o}}\Big),$$

and

(A.13) $$[\boldsymbol{b}_i|\tilde{\boldsymbol{y}}_i^{\mathrm{o}},\boldsymbol{\theta}] \sim t_q\Big(\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i}^{\mathrm{o}}\tilde{\boldsymbol{Z}}_i^{\mathrm{o}^{\mathrm{T}}}\boldsymbol{R}_i^{\mathrm{oo}^{-1}}(\tilde{\boldsymbol{y}}_i^{\mathrm{o}}-\tilde{\boldsymbol{X}}_i^{\mathrm{o}}\boldsymbol{\beta}), \Big(\frac{\nu+\Delta_{\tilde{\boldsymbol{y}}_i^{\mathrm{o}}}}{\nu+n_i^{\mathrm{o}}}\Big)\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i}^{\mathrm{o}},\nu+n_i^{\mathrm{o}}\Big),$$

respectively, where $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{\mathrm{o}}=\big(\sum_{i=1}^{N}\tau_i\tilde{\boldsymbol{X}}_i^{\mathrm{o}^{\mathrm{T}}}\tilde{\boldsymbol{\Lambda}}_i^{\mathrm{oo}^{-1}}\tilde{\boldsymbol{X}}_i^{\mathrm{o}}+\boldsymbol{F}_0^{-1}\big)^{-1}$, and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i}^{\mathrm{o}}=\big(\tilde{\boldsymbol{Z}}_i^{\mathrm{o}^{\mathrm{T}}}\boldsymbol{R}_i^{\mathrm{oo}^{-1}}\tilde{\boldsymbol{Z}}_i^{\mathrm{o}}+\boldsymbol{D}^{-1}\big)^{-1}$.

APPENDIX B. IMPLEMENTATION OF M-H ALGORITHM FOR $\boldsymbol{\phi}$ AND ν

Let $\phi_1^*=\log(\phi_1/(1-\phi_1))$ and $\phi_2^*=\log(\phi_2)$ such that $\boldsymbol{\phi}^*=(\phi_1^*,\phi_2^*)$ is located within a $\mathbb{R}^2$ space. We generate $\boldsymbol{\phi}^*=(\phi_1^*,\phi_2^*)$ through a bivariate normal distribution with mean vector $\boldsymbol{\phi}^{*^{(s)}}=(\phi_1^{*^{(s)}},\phi_2^{*^{(s)}})$ and variance-covariance matrix $c\mathbf{J}_{\boldsymbol{\phi}^*}$ which is chosen to be the asymptotic covariance matrix for $\boldsymbol{\phi}$ (Wang and Lin, 2014), where c is a tuning suitable value such that the acceptance of the M-H algorithm is between 0.2 and 0.4. The conditional posterior density of $\boldsymbol{\phi}^*$ is

$$f(\boldsymbol{\phi}^*|\boldsymbol{y},\boldsymbol{b},\boldsymbol{\theta}_{(-\boldsymbol{\phi}^*)})=p(\boldsymbol{\phi}|\boldsymbol{y},\boldsymbol{b},\boldsymbol{\theta}_{(-\boldsymbol{\phi})})e^{\phi_1^*+\phi_2^*}/(1+e^{\phi_1^*})^2,$$

and the resulting acceptance rate is

$$\alpha_1(\boldsymbol{\phi}^{*^{(s)}},\boldsymbol{\phi}^{*^{(s+1)}})=\min\Big\{\frac{f(\boldsymbol{\phi}^{*^{(s+1)}}|\boldsymbol{y},\boldsymbol{b}^{(s+1)},\boldsymbol{\theta}_{(-\boldsymbol{\phi}^*)}^{(s+1)})}{f(\boldsymbol{\phi}^{*^{(s)}}|\boldsymbol{y},\boldsymbol{b}^{(s+1)},\boldsymbol{\theta}_{(-\boldsymbol{\phi}^*)}^{(s+1)})},1\Big\}.$$

Having obtained $\boldsymbol{\phi}^*$, we transform them back to $(\phi_1,\phi_2)=(e^{\phi_1^*}/(1+e^{\phi_1^*}),e^{\phi_2^*})$.

For obtaining the posterior samples of ν from (A.2), we let $\nu^*=\log(\frac{1}{\nu-2})$ such that $\nu^*\in(-\infty,\infty)$, and then generate $\nu^{*^{(s+1)}}$ through a normal distribution with mean $\nu^{*^{(s)}}$ and variance $\sigma^2_{\hat{\nu}^*}$, where an approximation of $\sigma^2_{\hat{\nu}^*}$ is chosen as the asymptotic variance of $\hat{\nu}^*$. It follows from (A.2) that the conditional posterior density of ν^* is

$$f(\nu^*|\boldsymbol{y},\boldsymbol{b},\boldsymbol{\theta}_{(-\nu^*)})=p(\nu|\boldsymbol{y},\boldsymbol{b},\boldsymbol{\theta}_{(-\nu)})e^{-\nu^*}.$$

After having a candidate sample of ν^*, it is accepted with probability

$$\alpha_2(\nu^{*^{(s)}},\nu^{*^{(s+1)}})=\min\Big\{\frac{f(\nu^{*^{(s+1)}}|\boldsymbol{y},\boldsymbol{b}^{(s+1)},\boldsymbol{\theta}_{(-\nu^*)}^{(s+1)})}{f(\nu^{*^{(s)}}|\boldsymbol{y},\boldsymbol{b}^{(s+1)},\boldsymbol{\theta}_{(-\nu^*)}^{(s+1)})},1\Big\}.$$

Finally we transform the obtained ν^* back to $\nu=(1+2e^{-\nu^*})/e^{\nu^*}$.

Received 6 August 2016

REFERENCES

Brooks, S. P., 2002. Discussion on the paper by D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde. Journal of the Royal Statistical Society, Series B 64 (3), 616–618.

Brooks, S. P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7, 434–455. MR1665662

Browne, W. J., Drapery, D., 2006. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. Bayesian Analysis 1, 473–514. MR2221283

Carlin, B. P., Louis, T. A., 2006. Bayesian Methods for Data Analysis (Texts in Statistical Science). Chapman & Hall/CRC, New York. MR2027492

Connick, E., Lederman, M. M., Kotzin, B. L., Spritzler, J., Kuritzkes, D. R., Clair, M. S., Sevin, A. D., Fox, L., Chiozzi, M. H., Leonard, J. M., Rousseau, F., Roe, J. D., Martinez, A., Kessler, H., Landay, A., 2000. Immune reconstitution in the first year of potent antiretroviral therapy and its relationship to virologic response. Journal of Infectious Diseases 181, 358–363.

De la Cruz, R., 2014. Bayesian analysis for nonlinear mixed-effects models under heavy-tailed distributions. Pharmaceutical Statistics 13, 81–93.

De la Cruz-Mesía, R., Marshall, G., 2006. Non-linear random effects models with continuous time autoregressive errors: a Bayesian approach. Statistics in Medicine 25, 1471–1484. MR2227289

Dey, D. K., Chang, H., 1997. Bayesian approach for the nonlinear random effects models. Biometrics 53, 1239–1252.

Galecki, A. T., 1994. General class of covariance structures for two or more repeated factors in longitudinal data analysis. Communications in Statistics - Theory and Methods 23, 3105–3119.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 721–741.

Hastings, W. K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109. MR3363437
Kotz, S., Nadarajah, S., 2004. Multivariate t Distributions and Their Applications. Cambridge University Press, Cambridge. MR2038227
Lachos, V. H., Castro, L. M., Dey, D., 2013. Bayesian inference in nonlinear mixed-effects models using normal independent distributions. Computational Statistics & Data Analysis 64, 237–252. MR3061901
Lederman, M., Connick, E., Landay, A., Kuritzkes, D., Spritzler, J., Clair, M. S., Kotzin, B. L., Fox, L., Chiozzi, M. H., Leonard, J. M., Rousseau, F., Wade, M., Roe, J. D., Martinez, A., Harold, K., 1998. Immunologic responses associated with 12 weeks of combination antiretroviral therapy consisting of zidovudine, lamivudine, and ritonavir: results of AIDS clinical trials group protocol 315. Journal of Infectious Diseases 178, 70–79.
Liang, H., Wu, H., Carroll, R. J., 2003. The relationship between virologic responses in AIDS clinical research using mixed-effects varying-coefficient models with measurement error. Biostatistics 4, 297–312.
Lin, T. I., Lee, J. C., 2007. Bayesian analysis of hierarchical linear mixed modeling using the multivariate t distribution. Journal of Statistical Planning and Inference 137, 484–495. MR2298952
Lin, T. I., Wang, W. L., 2013. Multivariate skew-normal linear mixed models for multi-outcome longitudinal data. Statistical Modelling 13 (3), 199–221. MR3179524
Lindstrom, M. J., Bates, D. M., 1990. Nonlinear mixed effects models for repeated measures data. Biometrics 46, 673–687. MR1085815
Marshall, G., De la Cruz-Mesía, R., Barón, A. E., Rutledge, J. H., Zerbe, G. O., 2006. Non-linear random effects model for multivariate responses with missing data. Statistics in Medicine 25, 2817–2830. MR2242205
Marshall, G., De la Cruz-Mesía, R., Quintana, F. A., Barón, A. E., 2009. Discriminant analysis for longitudinal data with multiple continuous responses and possibly missing data. Biometrics 65, 69–80. MR2665847
Muñoz, A., Carey, V., Schouten, J. P., Segal, M., Rosner, B., 1992. A parametric family of correlation structures for the analysis of longitudinal data. Biometrics 48, 733–742.
Orendi, J., Bloem, A., Borleffs, J., Wijnholds, F. J., de Vos, M., Nottet, H., Visser, M., Snippe, H., Verhoef, J., Boucher, C., 1998. Activation and cell cycle antigens in CD4 and CD8 T cells correlate with plasma human immunodeficiency virus (HIV-1) RNA level in HIV-1 infection. Journal of Infectious Diseases 178, 1279–1287.
Pinheiro, J. C., Liu, C. H., Wu, Y., 2001. Efficient algorithms for robust estimation in linear mixed-effects model using the multivariate t distribution. Journal of Computational and Graphical Statistics 10, 249–276. MR1939700
Roy, A., 2006. Estimating correlation coefficient between two variables with repeated observations using mixed effects model. Biometrical Journal 48, 286–301. MR2224259
Roy, J., Lin, X., 2002. Analysis of multivariate longitudinal outcomes with nonignorable dropouts and missing covariates: changes in methadone treatment practices. Journal of the American Statistical Association 97, 40–52. MR1973718
Rubin, D., 1976. Inference and missing data. Biometrika 63, 581–592. MR0455196
Sachsenberg, N., Perelson, A., Yerly, S., Schockmel, G., Leduc, D., Hirschel, B., Perrin, L., 1998. Turnover of CD4 and CD8 T lymphocytes in HIV-1 infection as measured by Ki-67 antigen. Journal of Experimental Medicine 187, 1295–1303.
Sammel, M., Lin, X., Ryan, L., 1999. Multivariate linear mixed models for multiple outcomes. Statistics in Medicine 18, 2479–2492.
Schafer, J. L., Yucel, R. M., 2002. Computational strategies for multivariate linear mixed-effects models with missing values. Journal of Computational and Graphical Statistics 11, 437–457. MR1938143
Shah, A., Laird, N., Schoenfeld, D., 1997. A random-effects model for multiple characteristics with possibly missing data. Journal of the American Statistical Association 92, 775–779. MR1467867
Spiegelhalter, D. J., Best, N. G., Carlin, B. P., Linde, A. V. D., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B 64, 583–639. MR1979380
Wang, W. L., 2013. Multivariate t linear mixed models for irregularly observed multiple repeated measures with missing outcomes. Biometrical Journal 55, 554–571. MR3079990
Wang, W. L., 2015. Approximate methods for maximum likelihood estimation of multivariate nonlinear mixed-effects models. Entropy 17, 5353–5381.
Wang, W. L., Fan, T. H., 2010. ECM-based maximum likelihood inference for multivariate linear mixed models with autoregressive errors. Computational Statistics & Data Analysis 54, 1328–1341. MR2600835
Wang, W. L., Fan, T. H., 2011. Estimation in multivariate t linear mixed models for multiple longitudinal data. Statistica Sinica 21, 1857–1880. MR2896002
Wang, W. L., Lin, T. I., 2014. Multivariate t nonlinear mixed-effects models for multi-outcome longitudinal data with missing values. Statistics in Medicine 33, 3029–3046. MR3260520
Wang, W. L., Lin, T. I., 2015. Bayesian analysis of multivariate t linear mixed models with missing responses at random. Journal of Statistical Computation and Simulation 85 (17), 3594–3612. MR3395686
Wu, H., Ding, A., 1999. Population HIV-1 dynamics in vivo: applicable models and inferential tools for virological data from AIDS clinical trials. Biometrics 55, 410–418.
Wu, H., Liang, H., 2004. Backfitting random varying-coefficient models with time-dependent smoothing covariates. Scandinavian Journal of Statistics 31, 3–19. MR2042595

Wan-Lun Wang
Department of Statistics
Graduate Institute of Statistics and Actuarial Science
Feng Chia University
Taichung 40724
Taiwan
E-mail address: wlunwang@fcu.edu.tw

Luis M. Castro
Department of Statistics
Pontificia Universidad Católica de Chile
Casilla 306, Correo 22
Santiago
Chile
E-mail address: lmcastro@uc.cl