

# Bayesian analysis of hierarchical linear mixed modeling using the multivariate $t$ distribution

Tsung I. Lin<sup>a,\*</sup>, Jack C. Lee<sup>b</sup>

<sup>a</sup>Department of Applied Mathematics, National Chung Hsing University, Taichung 402, Taiwan

<sup>b</sup>Institute of Statistics and Graduate Institute of Finance, National Chiao Tung University, Hsinchu 300, Taiwan

Received 23 January 2005; accepted 26 December 2005

Available online 23 February 2006

## Abstract

This article presents a fully Bayesian approach to modeling incomplete longitudinal data using the  $t$  linear mixed model with AR( $p$ ) dependence. Markov chain Monte Carlo (MCMC) techniques are implemented for computing posterior distributions of parameters. To facilitate the computation, two types of auxiliary indicator matrices are incorporated into the model. Meanwhile, the constraints on the parameter space arising from the stationarity conditions for the autoregressive parameters are handled by a reparametrization scheme. Bayesian predictive inferences for the future vector are also investigated. An application is illustrated through a real example from a multiple sclerosis clinical trial.

© 2006 Elsevier B.V. All rights reserved.

*Keywords:* Autoregressive process; Bayesian prediction; Markov chain Monte Carlo; Missing values; Random effects;  $t$  linear mixed models

## 1. Introduction

The most popular analytic tool for longitudinal data analysis with continuous outcomes is the linear mixed model proposed by Laird and Ware (1982). The model assumes that both random effects and within-subject errors are normally distributed for mathematical convenience. However, such normality assumptions are vulnerable to the presence of atypical observations, which may seriously affect the estimation of fixed effects and variance components. To overcome this obstacle, instead of using conventional normal errors, many authors (e.g., Zellner, 1976; Lange et al., 1989) consider heavy-tailed  $t$  errors for linear regression models. A robust extension of the linear mixed models with the multivariate  $t$  distribution, called  $t$  linear mixed models hereafter, is considered by Pinheiro et al. (2001). They present several comparable efficient EM-type algorithms for computing the maximum likelihood (ML) estimates and illustrate the robustness of the model via a real example and some simulations.

Due mainly to recent advances in computing technology, the Bayesian sampling-based approach has been recognized to offer the data-analyst an alternative modeling strategy. The most pragmatic merit of adopting such an approach is the ability to take account of all parameter uncertainties. While simulating directly from the posterior distribution is typically difficult, a number of authors have advocated Markov chain Monte Carlo (MCMC) schemes to deal with intractable posterior integration. For example, the Gibbs sampler (Geman and Geman, 1984) and the Metropolis–Hastings

\* Corresponding author. Tel.: +886 4 22850420; fax: +886 4 22873028.

E-mail addresses: [tilin@amath.nchu.edu.tw](mailto:tilin@amath.nchu.edu.tw), [tilin@mail.thu.edu.tw](mailto:tilin@mail.thu.edu.tw) (T.I. Lin).

(M–H) algorithm (Hastings, 1970) are widely used in many areas of research. For normal linear mixed models, Gelfand et al. (1990) present the Gibbs sampler for Bayesian calculation, and Zeger and Karim (1991) extend the approach to generalized linear mixed models. Chib and Carlin (1999) consider several MCMC sampling schemes for hierarchical longitudinal models. More recently, Schafer and Yucel (2002) present some likelihood-based and Bayesian computational techniques for multivariate longitudinal data with missing values.

In this article, we shall present a fully Bayesian sampling-based approach to the  $t$  linear mixed model when repeated measures are serially correlated and some data are missing. The model involves a natural normal–normal–gamma hierarchy, which is conceptually flexible and would be easily implemented for Bayesian practitioners. As explained by Chi and Reinsel (1989) and Keramidis and Lee (1995), an appropriate dependence structure plays an important role for model fitting as well as predictive power. In longitudinal studies, the repeated measures of each subject are collected over time and hence tend to be serially correlated. Thus, we consider a stationary autoregressive (AR) time series structure to account for the serial correlation for within-subject errors. Note that the pure AR model can be extended to a much richer ARMA family, see Rochon (1992), Lin and Lee (2003) and Lee et al. (2005). Nevertheless, it is appropriate and relatively simple to fit high-order AR models instead of ARMA models due to the fact that longitudinal data are often short time series.

The rest of this article is organized as follows: Section 2 describes the model and the chosen priors. In Section 3, we present Bayesian estimation and posterior predictive inferences. Section 4 illustrates the application of our methods to a set of multiple sclerosis (MS) data. Finally, some concluding remarks are reported in Section 5.

## 2. The $t$ linear mixed model and prior distributions

### 2.1. The model

Assume there are  $N$  subjects in a longitudinal study and the  $i$ th subject,  $\mathbf{Y}_i$ , is repeatedly measured over  $n_i$  times, the underlying formulation of the  $t$  linear mixed model can be represented as:

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad \mathbf{b}_i | \tau_i \sim \mathcal{N}_{m_2} \left( \mathbf{0}, \frac{\sigma^2}{\tau_i} \boldsymbol{\Gamma} \right), \\ \boldsymbol{\varepsilon}_i | \tau_i &\sim \mathcal{N}_{n_i} \left( \mathbf{0}, \frac{\sigma^2}{\tau_i} \mathbf{C}_i \right), \quad \tau_i \sim \text{Gamma} \left( \frac{v}{2}, \frac{v}{2} \right), \end{aligned} \quad (1)$$

where  $\boldsymbol{\beta}$  is an  $m_1 \times 1$  vector of fixed effects with corresponding design matrix  $\mathbf{X}_i$ ,  $\mathbf{b}_i$  is an  $m_2 \times 1$  vector of unobservable random effects with corresponding design matrix  $\mathbf{Z}_i$ , and  $\tau_i$  is an unknown weight assumed to be distributed as gamma with mean 1 and variance  $2/v$ , and  $\mathbf{b}_i | \tau_i$  and  $\boldsymbol{\varepsilon}_i | \tau_i$  are mutually independent random variables having normal densities. Furthermore,  $\boldsymbol{\Gamma}$  is an  $m_2 \times m_2$  unstructured positive definite matrix, while  $\mathbf{C}_i$  is a structured AR( $p$ ) dependence matrix for within-subject errors. Specifically,

$$\mathbf{C}_i = \frac{1}{1 - \phi_1 \rho_1 - \dots - \phi_p \rho_p} [\rho_{|r-s|}],$$

where  $r, s = 1, \dots, n_i$  and  $\rho_k$ 's are implicit function of autoregressive parameters  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)$  and satisfy the Yule–Walker equation (Box et al. (1994)), i.e.,

$$\rho_k = \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p}, \quad \rho_0 = 1, \quad (k = 0, \dots, n_i - 1).$$

In addition, the roots of  $1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p = 0$  must lie outside the unit circle for assuring the stationarity of the model. For the pure AR model, admissible values of  $\boldsymbol{\phi}$  are confined in a  $p$ -dimensional hypercube  $\mathbb{C}_p$ .

The model (1) can be hierarchically formulated as follows:

$$\begin{aligned} \mathbf{Y}_i | \mathbf{b}_i, \tau_i &\sim \mathcal{N}_{n_i} \left( \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \frac{\sigma^2}{\tau_i} \mathbf{C}_i \right), \\ \mathbf{b}_i | \tau_i &\sim \mathcal{N}_{m_2} \left( \mathbf{0}, \frac{\sigma^2}{\tau_i} \boldsymbol{\Gamma} \right), \quad \tau_i \sim \text{Gamma} \left( \frac{v}{2}, \frac{v}{2} \right). \end{aligned} \quad (2)$$

Combining the joint distribution of  $(Y_i, b_i, \tau_i)$  obtained from (2) and integrating out  $b_i$  and  $\tau_i$ , we have

$$f(Y_i) = \frac{\Gamma(v + n_i/2) |\Lambda_i|^{-1/2}}{\Gamma(v/2)(\pi v \sigma^2)^{n_i/2}} \left( 1 + \frac{(Y_i - X_i \beta)^T \Lambda_i^{-1} (Y_i - X_i \beta)}{v \sigma^2} \right)^{-(v+n_i)/2},$$

where  $\Lambda_i = Z_i \Gamma Z_i^T + C_i$ . Let  $t_n(\mu, \Sigma, v)$  denote an  $n$ -dimensional multivariate  $t$  distribution with location vector  $\mu$ , scatter matrix  $\Sigma$  and degrees-of-freedom  $v$ . It follows that the distribution of  $Y_i$  is  $t_{n_i}(X_i \beta, \sigma^2 \Lambda_i, v)$ .

We are concerned with the situation in which some observations are missing. Let  $Y_i$  be partitioned into two components  $(Y_i^o, Y_i^m)$ , where  $Y_i^o$  ( $n_i^o \times 1$ ) and  $Y_i^m$   $((n_i - n_i^o) \times 1)$  denote the observed and missing components of  $Y_i$ , respectively. To facilitate the computation, it is convenient to introduce two auxiliary binary indicator matrices, denoted by  $O_i$  and  $M_i$  henceforth, corresponding to  $Y_i$  such that  $Y_i^o = O_i Y_i$  and  $Y_i^m = M_i Y_i$ , respectively. Note that  $O_i$  and  $M_i$  are  $n_i^o \times n_i$  and  $(n_i - n_i^o) \times n_i$  matrices extracted from an  $n_i$ -dimensional identity matrix  $I_{n_i}$  corresponding to row-positions of  $Y_i^o$  and  $Y_i^m$  in  $Y_i$ , respectively. Moreover, it is easy to see that

$$Y_i = \begin{cases} Y_i^o, & \text{if } n_i^o = n_i, \\ O_i^T Y_i^o + M_i^T Y_i^m, & \text{if } 1 \leq n_i^o < n_i, \end{cases}$$

and  $O_i^T O_i + M_i^T M_i = I_{n_i}$ .

It is crucial to utilize the distributional properties of the marginal distribution of  $Y_i^o$  and the conditional distribution of  $Y_i^m$  given  $Y_i^o$ . Integrating  $Y_i^m$  from the joint density of  $(Y_i^o, Y_i^m)$  leads to the marginal distribution of  $Y_i^o$  which is  $t_{n_i^o}(X_i^o \beta, \sigma^2 \Lambda_i^o, v)$ , where  $X_i^o = O_i X_i$  and  $\Lambda_i^o = O_i \Lambda_i O_i^T$ . Furthermore, we have  $Y_i^m | Y_i^o \sim t_{n_i - n_i^o}(\mu_i^{m \cdot o}, w_i \Lambda_i^{mm \cdot o}, v + n_i^o)$ , where

$$\begin{aligned} \mu_i^{m \cdot o} &= X_i^m \beta + M_i \Lambda_i O_i^T \Lambda_i^{o \cdot o^{-1}} (Y_i^o - X_i^o \beta) = M_i (X_i \beta + \Lambda_i S_i^o (Y_i - X_i \beta)), \\ S_i^o &= O_i^T \Lambda_i^{o \cdot o^{-1}} O_i = O_i^T (O_i \Lambda_i O_i^T)^{-1} O_i, \quad X_i^m = M_i X_i, \\ w_i &= \frac{\sigma^2 v + \delta_i^o}{v + n_i^o}, \quad \delta_i^o = (Y_i - X_i \beta)^T S_i^o (Y_i - X_i \beta), \\ \Lambda_i^{mm \cdot o} &= M_i (I_{n_i} - \Lambda_i S_i^o) \Lambda_i M_i^T. \end{aligned}$$

We will focus on a general approach to estimate parameters  $(\beta, \sigma^2, \Gamma, \phi, v)$  from model (2). Note that the patterns of missingness are assumed to be missing at random (MAR), see Little and Rubin (2002) for more details. In order to facilitate the estimating procedure and achieve the objective of ensuring admissibility of  $\phi$ , we perform a reparameterization on  $\phi$  as in Barndorff-Nielsen and Schou(1973):

$$\phi_k^{(k)} = \gamma_k, \quad \phi_j^{(k)} = \phi_j^{(k-1)} - \gamma_k \phi_{k-j}^{(k-1)}, \quad j = 1, 2, \dots, k - 1, \tag{3}$$

where  $\phi_j^{(p)} = \phi_j = \phi_j^{(j)} - \phi_{j+1}^{(j+1)} \phi_1^{(j)} - \phi_{j+2}^{(j+2)} \phi_2^{(j+1)} - \dots - \phi_p^{(p)} \phi_{p-j}^{(p-1)}$ , for  $j = 1, \dots, p - 1$ . Note that (3) is a one-to-one and onto transformation which reparameterizes  $\phi = (\phi_1, \dots, \phi_p) \in \mathbb{C}^p$  in terms of the partial autocorrelations  $\gamma = (\gamma_1, \dots, \gamma_p) \in \mathbb{R}^p$ , where  $\mathbb{R} = [-1, 1]$ . We treat  $\gamma$  as the reparameterized parameters and all random generation is done in  $\gamma$ , then inverting back to  $\phi$  at the end.

### 2.2. Prior distributions

To complete a Bayesian formulation of model (2), one must specify a prior distribution for  $\theta = (\beta, \sigma^2, \Gamma, \phi, v)$ . Suppose  $(\beta, \sigma^2, \Gamma, \gamma, v)$  are independent a priori, that is

$$\pi(\theta) \propto \pi(\beta) \pi(\sigma^2) \pi(\Gamma) \pi(\gamma) \pi(v). \tag{4}$$

In the absence of good prior information, a convenient strategy of avoiding improper posterior distribution is to use diffuse proper priors. The prior distributions adopted are as follows:

$$\beta \sim \mathcal{N}_{m_1}(\beta_0, \mathbf{B}_0), \quad \sigma^2 \sim \mathcal{IG}\left(\frac{a_0}{2}, \frac{b_0}{2}\right), \quad \Gamma \sim \mathcal{IW}(c_0, \mathbf{\Omega}_0),$$

$$\gamma_i \sim \mathcal{U}(-1, 1) \quad (i = 1, \dots, p), \quad \log(1/v) \sim \mathcal{U}(-10, 10),$$

where  $\mathcal{IG}(\cdot, \cdot)$  denotes the inverse gamma distribution,  $\mathcal{IW}(\cdot, \cdot)$  denotes the inverse Wishart distribution, and  $\mathcal{U}(\cdot, \cdot)$  denotes the uniform distribution. Note that the prior for  $v$  is also considered by Liu and Rubin (1998) on the basis of vagueness. The values of the hyperparameters  $\beta_0, \mathbf{B}_0, a_0, b_0, c_0, \mathbf{\Omega}_0$  can be based on strong prior knowledge or be chosen to reflect diffuse prior information. The hyperparameters  $a_0$  and  $c_0$  are held fixed as small as possible. Meanwhile, we shall set  $\beta_0 = \hat{\beta}, \mathbf{B}_0 = \widehat{\text{cov}}(\hat{\beta}) = \sum_{i=1}^N (\mathbf{X}_i^{\text{oT}} \hat{\Lambda}_i^{\text{o-1}} \mathbf{X}_i^{\text{o}})^{-1}, b_0 = \hat{\sigma}^2$ , and  $\mathbf{\Omega}_0 = \hat{\Gamma}$ , where  $\hat{\beta}, \hat{\sigma}^2, \hat{\Gamma}$  denote the ML estimates for the ordinary normal linear mixed model with missing value imputed by the mean of two adjacent values. Generally, we shall choose  $a_0 = 3$  and  $c_0 = m_2 + 2$ , which will lead to flat distributions, and make  $E_{\pi}(\sigma^2) = \hat{\sigma}^2$  and  $E_{\pi}(\Gamma) = \hat{\Gamma}$ , where  $E_{\pi}(\cdot)$  denotes the expectation taken with respect to the prior distribution.

### 3. Bayesian estimation and predictive inferences

Let  $\mathbf{Y}^{\text{o}} = (\mathbf{Y}_1^{\text{o}}, \dots, \mathbf{Y}_N^{\text{o}}), \mathbf{Y}^{\text{m}} = (\mathbf{Y}_1^{\text{m}}, \dots, \mathbf{Y}_N^{\text{m}}), \mathbf{b} = (\mathbf{b}_1, \dots, \mathbf{b}_N), \boldsymbol{\tau} = (\tau_1, \dots, \tau_N), \mathbf{e}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i, S_i(\boldsymbol{\beta}, \mathbf{b}_i, \gamma) = \mathbf{e}_i^{\text{T}} \mathbf{C}_i \mathbf{e}_i$  and  $n = \sum_{i=1}^N n_i$ . Combining the complete-data likelihood function of model (2) with the prior distribution (4), we have the following joint posterior density of  $(\boldsymbol{\theta}, \mathbf{Y}^{\text{m}}, \mathbf{b}, \boldsymbol{\tau})$ :

$$p(\boldsymbol{\theta}, \mathbf{Y}^{\text{m}}, \mathbf{b}, \boldsymbol{\tau} | \mathbf{Y}^{\text{o}})$$

$$\propto (\sigma^2)^{-(n+Nm_2+a_0+2)/2} \left[ \prod_{i=1}^N \left| \frac{1}{\tau_i} \mathbf{C}_i \right|^{-1/2} \right] \exp \left\{ -\frac{\sum_{i=1}^N \tau_i S_i(\boldsymbol{\beta}, \mathbf{b}_i, \gamma)}{2\sigma^2} \right\}$$

$$\times \left[ \prod_{i=1}^N \left| \frac{1}{\tau_i} \Gamma \right|^{-1/2} \right] \exp \left\{ -\frac{\sum_{i=1}^N \tau_i \mathbf{b}_i^{\text{T}} \Gamma^{-1} \mathbf{b}_i}{2\sigma^2} \right\} \left[ \prod_{i=1}^N \tau_i^{v/2-1} \right] \exp \left\{ -\frac{v}{2} \sum_{i=1}^N \tau_i \right\}$$

$$\times \left[ \frac{(v/2)^{v/2}}{\Gamma(v/2)} \right]^N \exp \left\{ -\frac{(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^{\text{T}} \mathbf{B}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0)}{2} \right\} \exp \left\{ -\frac{b_0}{2\sigma^2} \right\}$$

$$\times |\Gamma|^{-(c_0+m_2+1)/2} \exp \left\{ -\frac{1}{2} \text{tr}(\Gamma^{-1} \mathbf{\Omega}_0) \right\} J_v, \tag{5}$$

where  $J_v = 1/v (0 < v < \infty)$  is the Jacobian of transforming  $\log(1/v)$  to  $v$ . Details on the forms of the full conditions and the implementation of MCMC sampler can be found in the Appendix.

We consider the prediction of  $\mathbf{y}_i$ , a  $q \times 1$  future observations of  $\mathbf{Y}_i$ . Let  $\mathbf{x}_i$  and  $\mathbf{z}_i$  be  $q \times m_1$  and  $q \times m_2$  design matrices of prediction regressors corresponding to  $\mathbf{y}_i$ . We thus have

$$\begin{bmatrix} \mathbf{Y}_i \\ \mathbf{y}_i \end{bmatrix} \sim t_{n_i+q}(\mathbf{X}_i^* \boldsymbol{\beta}, \sigma^2 \mathbf{\Lambda}_i^*, v),$$

where  $\mathbf{X}_i^* = (\mathbf{X}_i^{\text{T}}, \mathbf{x}_i^{\text{T}})^{\text{T}}, \mathbf{Z}_i^* = (\mathbf{Z}_i^{\text{T}}, \mathbf{z}_i^{\text{T}})^{\text{T}}$  and  $\mathbf{\Lambda}_i^* = \mathbf{Z}_i^* \Gamma \mathbf{Z}_i^{*\text{T}} + \mathbf{C}_i^*$  with  $\mathbf{C}_i^* = [\rho_{|r-s|}] (r, s = 1, \dots, n_i + q)$ .

Let  $\mathbf{Y}_i^* = (\mathbf{Y}_i^{\text{T}}, \mathbf{y}_i^{\text{T}})^{\text{T}}$  and  $\mathbf{Y}_i^{\text{o}*} = (\mathbf{Y}_i^{\text{oT}}, \mathbf{y}_i^{\text{oT}})^{\text{T}}$ . We further define four additional auxiliary binary indicator matrices  $\mathbf{O}_i^*, \mathbf{M}_i^*, \mathbf{O}_i^{\dagger}$  and  $\mathbf{M}_i^{\dagger}$  such that  $\mathbf{O}_i^* \mathbf{Y}_i^* = \mathbf{Y}_i^{\text{o}*}, \mathbf{M}_i^* \mathbf{Y}_i^* = \mathbf{Y}_i^{\text{m}}, \mathbf{O}_i^{\dagger} \mathbf{Y}_i^{\text{o}*} = \mathbf{Y}_i^{\text{o}}$  and  $\mathbf{M}_i^{\dagger} \mathbf{Y}_i^{\text{o}*} = \mathbf{y}_i$ , respectively. With arguments similar to those pointed out in Section 2, we have  $\mathbf{Y}_i^{\text{o}*} \sim t_{n_i^{\text{o}}+q}(\mathbf{X}_i^{\text{o}*} \boldsymbol{\beta}, \sigma^2 \mathbf{\Lambda}_i^{\text{o}*}, v)$ , where  $\mathbf{X}_i^{\text{o}*} = \mathbf{O}_i^* \mathbf{X}_i^*$  and  $\mathbf{\Lambda}_i^{\text{o}*} = \mathbf{O}_i^* \mathbf{\Lambda}_i^* \mathbf{O}_i^{*\text{T}}$ .

The posterior predictive distribution of  $\mathbf{y}_i$  is

$$p(\mathbf{y}_i | \mathbf{Y}^{\text{o}}) = \int f(\mathbf{y}_i | \mathbf{Y}^{\text{o}}, \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{Y}^{\text{o}}) d\boldsymbol{\theta}. \tag{6}$$

In (6), we have  $f(y_i|Y, \theta) = t_q(y_i | \mu_{2,1}, w\Omega_{22,1}, v + n_i^o)$ , where

$$\begin{aligned} \mu_{2,1} &= M_i^\dagger X_i^{o*} \beta + M_i^\dagger \Lambda_i^{o*} O_i^{\dagger T} (O_i^\dagger \Lambda_i^{o*} O_i^{\dagger T})^{-1} O_i^\dagger (Y_i^{o*} - X_i^{o*} \beta), \\ w &= \frac{\sigma^2 v + (Y_i^{o*} - X_i^{o*} \beta)^T O_i^{\dagger T} (O_i^\dagger \Lambda_i^{o*} O_i^{\dagger T})^{-1} O_i^\dagger (Y_i^{o*} - X_i^{o*} \beta)}{v + n_i^o}, \\ \Omega_{22,1} &= M_i^\dagger (I_{n_i^o+q} - \Lambda_i^{o*} O_i^{\dagger T} (O_i^\dagger \Lambda_i^{o*} O_i^{\dagger T})^{-1} O_i^\dagger) \Lambda_i^{o*} M_i^{\dagger T}. \end{aligned} \tag{7}$$

Let  $\theta^{(g)}$  be the generated sample at the  $g$ th iteration of the MCMC sampler when the convergence is achieved. We can obtain the approximate predictive distribution of  $y_i$  using the Rao–Blackwellization (Gelfand and Smith, 1990). That is,

$$p(y_i|Y) \approx \frac{1}{G} \sum_{g=1}^G t_q(y_i | \mu_{2,1}^{(g)}, w^{(g)} \Omega_{22,1}^{(g)}, v^{(g)} + n_i).$$

For the future value  $y_i$ , it can be straightforwardly predicted by

$$\hat{y}_i = \frac{1}{G} \sum_{g=1}^G \mu_{2,1}^{(g)}, \tag{8}$$

where  $\mu_{2,1}^{(g)}$  is  $\mu_{2,1}$  in (7) with  $\theta$  replaced by  $\theta^{(g)}$ .

### 4. Example

We apply the methods from previous sections to the MS data from a clinical trial of 52 relapsing–remitting MS (RRMS) patients. The cohort study, conducted at the University of British Columbia site during the period June 1988–May 1990, was a placebo-controlled trial of interferon beta-1b (INFB), which had been approved by the US Food and Drug Administration in mid 1993 for patients with early stage RRMS. All 52 patients were randomized into three treatment groups—a placebo (PL) group, a low-dose (LD) group, and a high-dose (HD) group. The LD and HD treatments correspond to doses of 1.6 and 8 million international units (MIU) of IFNB every other day, respectively. D’yachkova et al. (1997) use the MS data to illustrate the application of the generalized estimating equation (GEE) approach. Gill (2000) separately analyze the three treatment groups based on mixed linear modeling with Huber’s  $\rho$  function. Lin and Lee (2006) present an alternative robust approach using the  $t$  distribution and provide a score test statistic for detecting AR(1) serial correlation.

The response variable in this study is the patient’s “disease burden”, which is measured by the total area of MS lesions on all slices of an cranial magnetic resonance imaging (MRI) scan (in mm<sup>2</sup>). The disease burden for the  $i$ th patient at time point  $j$  is denoted by  $\text{Area}(i, j)$ , with  $j = 0$  as the baseline time point. Due to strong skewness of the untransformed burden measurements, we use the log relative burden (LRB), defined by  $\text{LRB}(i, j) = \log(\text{Area}(i, j)/\text{Area}(i, 0))$ , as the response variable  $Y_{ij}$ .

In this study, three patients were not included in the analysis since two of them (one in each of groups LD and HD) dropped out very early and one in group LD had 3 measurements of zero on MRI scans. Note that each patient was repeatedly measured approximately once every six weeks over the two-year period, so that the maximum number of visits was 17 for each patient. In this data set, all but 5 patients have a complete set of 17 scans: one dropped out from PL after completing 14 visits, two dropped out from LD after completing 13 visits, and two dropped out from HD after completing 12 visits. Assuming these early dropouts are “ignorable” (Rubin, 1976), our analyses are limited to the LRB measurements on the remaining 49 patients. There are 17 patients in PL and 16 patients in both LD and HD. Among these 49 patients, six patients have one or two isolated MRI scans missing. Instead of imputing these missing values via the mean of two adjacent values as in Gill (2000), we simply simulate these missing values using (A.1) via the MCMC approach.

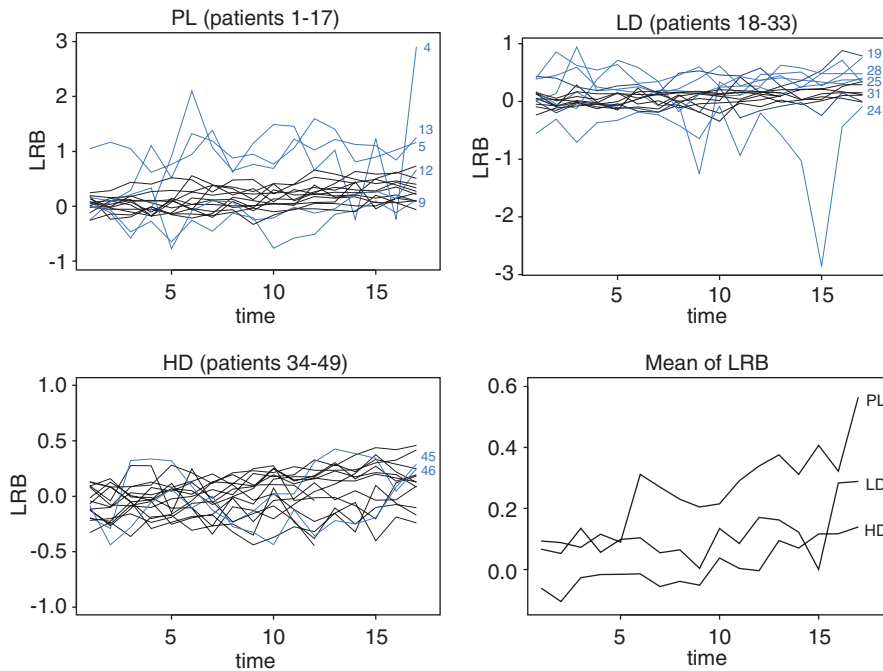


Fig. 1. Longitudinal trends in LRB from three treatment groups and the average LRB versus time

We coded these 49 patients by numbering them from 1 to 49. The identity numbers are 1 to 17 for PL, 18 to 33 for LD and 34 to 49 for HD. Fig. 1 depicts the time evolution of LRB measurements for each patient and the average LRB of each group, indicating that some outlying observations are apparently present for PL and LD groups.

We carry out the analyses for all three treatment groups by fitting a single  $t$  linear mixed model with the variance components and the degrees-of-freedom taken to be common across all treatment groups. In addition, we assume a linear growth function for fixed effects, along with subject-specific random intercepts and slopes and an AR( $p$ ) structure for  $C_i$  for  $p = 0, 1, 2, 3$ . Note that  $p = 0$  indicates the within-subject residuals follow a white noise process. The fitted hierarchical model can be written as

$$\begin{aligned}
 Y_i | b_i, \tau_i &\sim \mathcal{N}_{n_i} \left( X_i \beta + Z_i b_i, \frac{\sigma^2}{\tau_i} C_i \right), \\
 b_i | \tau_i &\sim \mathcal{N}_2 \left( \mathbf{0}, \frac{\sigma^2}{\tau_i} \Gamma \right), \quad \tau_i \sim \text{Gamma} \left( \frac{\nu}{2}, \frac{\nu}{2} \right),
 \end{aligned} \tag{9}$$

with

$$\begin{aligned}
 \beta &= (\beta_0, \beta_1, \beta_2, \beta_3)^T, \quad X_i = [\mathbf{1}_{n_i} \ k_i^{(1)} \ k_i^{(2)} \ k_i^{(3)}], \quad Z_i = [\mathbf{1}_{n_i} \ k_i], \\
 k_i &= (1, 2, \dots, n_i)^T, \quad k_i^{(1)} = k_i \mathbb{1}_{\text{PL}}(Y_i), \quad k_i^{(2)} = k_i \mathbb{1}_{\text{LD}}(Y_i), \quad k_i^{(3)} = k_i \mathbb{1}_{\text{HD}}(Y_i),
 \end{aligned}$$

where  $\mathbf{1}_{n_i}$  is a  $n_i \times 1$  vector of ones, and  $\mathbb{1}_A(u) = 1$  if  $u \in A$  and  $\mathbb{1}_A(u) = 0$  else. In this model,  $\beta_0$  is a fixed intercept effect common to all subjects, and  $\beta_i$  ( $i = 1, 2, 3$ ) are treatment-specific slopes for PL, LD and HD groups, respectively. Furthermore, the random effects  $b_i = (b_{1i}, b_{2i})^T$  have the scale covariance matrix  $\Gamma = \begin{bmatrix} \Gamma_{11} & \Gamma_{21} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}$ , and  $C_i$ 's are implicit functions of  $\phi = (\phi_1, \dots, \phi_p)$ . Of course,  $C_i = I_{n_i}$  if  $p = 0$ .

We ran 10 independent parallel chains with different initial values for each chain started from a point drawn at random from the prior. For each chain, we implemented 10,000 iterations. We monitored the convergence by examining the multivariate potential scale reduction factor (MPSRF) of Brooks and Gelman (1998) using multiple chains. For all the fitted models, convergence to the posterior distribution was quick and the mixing was good. The convergence occurred after 2000 iterations. Discarding the first 2000 iterations as a “burn-in” for each chain, we then stored one imputed

Table 1  
Summarized posterior results of MCMC samples for the  $t$  linear mixed model with four selected dependence structures

Posterior estimates	Parameter											
	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\Gamma_{11}$	$\Gamma_{21}$	$\Gamma_{22}$	$\sigma^2$	$\nu$	$\phi_1$	$\phi_2$	$\phi_3$
White noise												
Mean	-0.0355	0.0193	0.0132	0.0136	1.4407	-0.0036	0.0077	0.0117	2.3703	—	—	—
Standard deviation	0.0216	0.0029	0.0030	0.0027	0.4213	0.0228	0.0026	0.0018	0.4726	—	—	—
Median	-0.0355	0.0192	0.0132	0.0137	1.3803	-0.0021	0.0073	0.0116	2.3259	—	—	—
$Q_{0.025}$	-0.0773	0.0135	0.0074	0.0082	0.7780	-0.0535	0.0037	0.0085	1.5714	—	—	—
$Q_{0.975}$	0.0070	0.0251	0.0190	0.0190	2.4166	0.0382	0.0137	0.0154	3.4067	—	—	—
AR(1)												
Mean	-0.0331	0.0181	0.0133	0.0132	0.9669	0.0234	0.0052	0.0120	2.3964	0.3730	—	—
Standard deviation	0.0208	0.0030	0.0030	0.0028	0.3487	0.0201	0.0024	0.0018	0.4822	0.0449	—	—
Median	-0.0331	0.0181	0.0133	0.0133	0.9173	0.0246	0.0048	0.0118	2.3402	0.3729	—	—
$Q_{0.025}$	-0.0747	0.0122	0.0075	0.0077	0.4295	-0.0204	0.0015	0.0087	1.5670	0.2827	—	—
$Q_{0.975}$	0.0071	0.0241	0.0194	0.0187	1.7647	0.0618	0.0106	0.0159	3.4627	0.4606	—	—
AR(2)												
Mean	-0.0322	0.0178	0.0134	0.0131	0.8890	0.0244	0.0048	0.0120	2.3760	0.3677	0.1267	—
Standard deviation	0.0212	0.0032	0.0030	0.0029	0.3582	0.0192	0.0024	0.0019	0.4741	0.0436	0.0450	—
Median	-0.0324	0.0177	0.0134	0.0131	0.8301	0.0256	0.0043	0.0119	2.3299	0.3686	0.1264	—
$Q_{0.025}$	-0.0724	0.0117	0.0073	0.0072	0.3614	-0.0183	0.0015	0.0087	1.5629	0.2781	0.0406	—
$Q_{0.975}$	0.0092	0.0240	0.0193	0.0188	1.7038	0.0591	0.0104	0.0158	3.4231	0.4519	0.2178	—
AR(3)												
Mean	-0.0314	0.0178	0.0133	0.0131	0.8347	0.0259	0.0044	0.0120	2.3771	0.3677	0.1249	0.0418
Standard deviation	0.0219	0.0032	0.0031	0.0029	0.3455	0.0180	0.0023	0.0018	0.4742	0.0449	0.0453	0.0461
Median	-0.0317	0.0178	0.0134	0.0131	0.7962	0.0270	0.0040	0.0120	2.3385	0.3654	0.1242	0.0411
$Q_{0.025}$	-0.0747	0.0116	0.0070	0.0074	0.2802	-0.0133	0.0010	0.0086	1.5924	0.2788	0.0355	-0.0458
$Q_{0.975}$	0.0118	0.0241	0.0194	0.0187	1.6411	0.0586	0.0100	0.0159	3.4316	0.4593	0.2153	0.1316

parameter values for every 20 iterations to reduce the autocorrelation. We have final 4000 realizations from the target posterior distribution.

Based on the converged simulated realizations under four considered dependence structures, summary statistics of converged MCMC samples, including the mean, standard deviation, median and 95% posterior intervals, are listed in Table 1. As can be seen, all estimates of the  $\nu$ 's are extremely small, indicating the normal-based model is not suitable for the MS data. In addition, the autoregressive parameters  $\phi_1$  and  $\phi_2$  in the AR model are significantly different from zero, while  $\phi_3$  is negligible in the AR(3) model.

We are interested in comparing the appropriateness of the four selected dependence structures using the Bayes factor, which involves the calculation of the marginal likelihoods of the competing models. Assuming the four models are equal-probable a priori, the Bayes factor is equivalent to the ratio between two corresponding marginal densities of the data. Here the marginal density for model  $M_k$  is given by

$$p(\mathbf{Y} | M_k) = \int p(\mathbf{Y} | \boldsymbol{\theta}_k, M_k) \pi(\boldsymbol{\theta}_k | M_k) d\boldsymbol{\theta}_k, \tag{10}$$

where  $\boldsymbol{\theta}_k$  is the vector of parameters of model  $M_k$ ,  $p(\mathbf{Y} | \boldsymbol{\theta}_k, M_k)$  is the likelihood and  $\pi(\boldsymbol{\theta}_k | M_k)$  is the prior density of  $\boldsymbol{\theta}_k$  under model  $M_k$ , respectively. However, the calculation of marginal likelihood remains a computationally intensive task. In this example, after analytically integrating missing values  $\mathbf{Y}^m$  and latent values  $\mathbf{b}_i$ 's and  $\boldsymbol{\tau}_i$ 's, a simple way of calculating the marginal density (10) is to use the harmonic means of sampled likelihoods (Newton and Raftery, 1994). That is,

$$p(\mathbf{Y} | M_k) \approx \left\{ \frac{1}{L} \sum_{\ell=1}^L p(\mathbf{Y} | \boldsymbol{\theta}_k^{(\ell)}, M_k)^{-1} \right\}^{-1}.$$



Table 2  
Estimated marginal densities for selected dependence structures

Dependence structure	$p(Y   \text{Model})$
WN	$1.2290 \times 10^{117}$
AR(1)	$2.1133 \times 10^{131}$
AR(2)	$4.0260 \times 10^{132}$
AR(3)	$6.7018 \times 10^{129}$

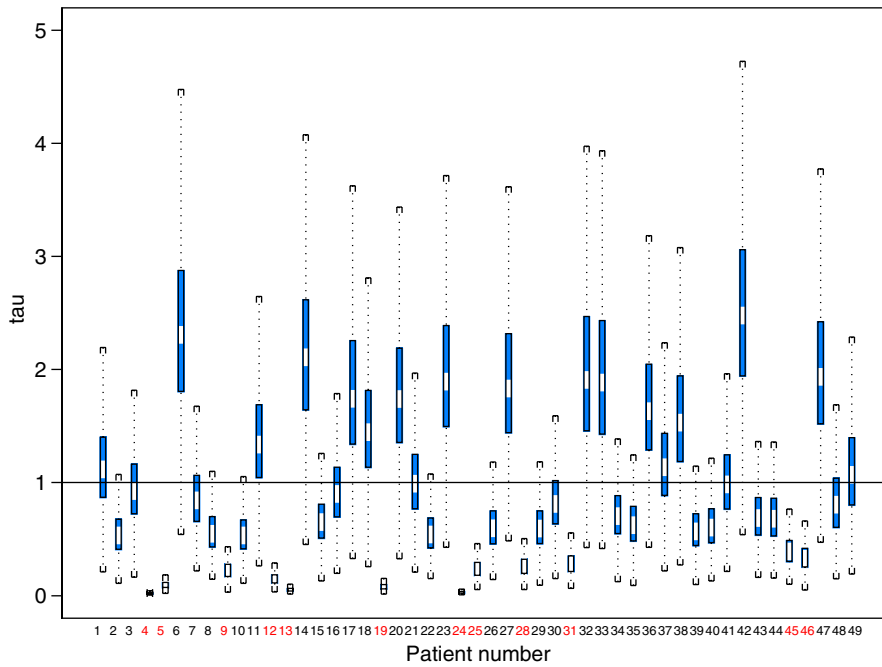


Fig. 2. Marginal posterior distribution of  $\tau_i$  for the 49 patients. The boxplots are drawn containing 2.5%, 25%, 50%, 75%, 97.5% quantiles of the MCMC samples.

The estimated marginal densities for the four selected models are listed in Table 2. All AR models are more plausible than the WN model and the AR(2) model has the largest marginal density. The Bayes factor of AR(2) relative to WN, AR(1) and AR(3) are  $3.2758 \times 10^{15}$ , 19.05 and 600.73, respectively. Based on the analysis so far, it appears that the AR(2) structure is adequate for the MS data.

To detect outlying observations, Fig. 2 shows the boxplots of sampled  $\tau_i$  for  $i = 1, 2, \dots, 49$ . As pointed out by Wakefield et al. (1994), the  $\tau_i$ 's can be used as concise indicators for detecting outliers with prior expectation of 1. Hence, when the value of  $\tau_i$  is substantially lower than 1, it indicates that the  $i$ th patient should be regarded as an outlier in the population. The figure reveals that patients 4, 5, 9, 12, 13, 19, 24, 25, 28, 31, 45 and 46 could be treated as outliers, since they exhibit extremely small values of  $\tau_i$  (none of 95% upper posterior limits exceed 1). We mark these identity numbers in Fig. 1.

For dependent longitudinal data, a more appropriate measure of “fitness” is the predictive accuracy of future observations (Rao, 1987; Lee, 1988). Bayesian predictive accuracy were compared between a normal model and a  $t$  model with the AR(2) structure. The fitted normal model is obtained by setting  $\nu = \infty$  in (9). We use (8) to predict the last measurements of the MS data. To compare the performance of different models, we shall use the mean of square deviations  $|\hat{y}_j - y_j|^2$ , absolute deviations  $|\hat{y}_j - y_j|$  and relative absolute deviations  $|\hat{y}_j - y_j|/y_j$ . We call these measures MSE, MAE and MARE, respectively. We consider one-step-ahead and two-step-ahead forecasts for the last measurements of the MS data. The prediction accuracies are displayed in Table 3. From the prediction results summarized in the table,



Table 3

Comparison of Bayesian predictive accuracies in terms of various discrepancy measures (DM) between the normal linear mixed model (nlmm) and the  $t$  linear mixed model (tlmm) with AR(2) dependence

Discrepancy measure	One-step-ahead			Two-step-ahead		
	nlmm	tlmm	RIP <sup>a</sup> (%)	nlmm	tlmm	RIP <sup>a</sup> (%)
MSE	0.1303	0.1205	7.52	0.1703	0.1554	8.75
MAE	0.1881	0.1785	5.10	0.2070	0.1941	6.23
MARE	0.7328	0.6829	6.81	0.8694	0.8145	6.31

<sup>a</sup>The relative improvement percentage (RIP) is measured by  $[DM(nlmm) - DM(tlmm)]/DM(nlmm) \times 100\%$ .

it appears that the  $t$  model has better prediction results with minimum improvements over 5% for the one-step-ahead forecast and over 6% for the two-step-ahead forecast, respectively.

### 5. Concluding remarks

We provide a fully Bayesian method for handling  $t$  linear mixed models with AR( $p$ ) dependence and simultaneously accommodating the presence of missing values. The proposed approach allows the user to fit longitudinal data in a wide variety of considerations. Under the flexible normal–normal-gamma hierarchy, computational techniques using the MCMC method allow us to generate samples from the posterior and predictive densities in a straightforward manner.

In the illustrated example, the graphical outputs provide both easily understood inferential summaries and informative diagnostic aids for detecting outliers. Furthermore, in terms of Bayesian forecast accuracy, the  $t$  linear mixed model is evidently more adequate than the normal counterpart. Although the proposed approach may still have some limitations, it is worthwhile to note that the situation in which no presence of outlying observations (taking  $\nu = \infty$ ) and no presence of missing values (taking  $O_i = I_{n_i}$ ) are simply special cases.

### Acknowledgements

We gratefully acknowledge the Executive Editor, an Associate Editor and two referees for their valuable comments, which substantially improved the quality of the paper. This research work was supported by the National Science Council of Taiwan (Grant Nos. NSC94-2118-M-005-003 and NSC94-2118-M-009-001).

### Appendix

In this appendix, we provide the detailed forms of full conditionals and strategies for the implementation in the MCMC algorithm.

#### A.1. Full conditional posterior distributions

From (5) the full conditionals are given as follows (using “| . . .” to denote conditional on  $\mathbf{Y}^o$  and all other parameters, missing components  $\mathbf{Y}^m$ , and latent variables  $\mathbf{b}$  and  $\boldsymbol{\tau}$ ):

$$Y_i^m \mid \dots \sim \mathcal{N}_{n-n_i^o} \left( \boldsymbol{\mu}_i^*, \frac{\sigma^2}{\tau_i} \mathbf{C}_i^{\text{mm-o}} \right), \tag{A.1}$$

$$\tau_i \mid \dots \sim \text{Gamma} \left( \frac{n_i + m_2 + \nu}{2}, \frac{\sigma^{-2} (S_i(\boldsymbol{\beta}, \mathbf{b}_i, \gamma) + \mathbf{b}_i^T \boldsymbol{\Gamma}^{-1} \mathbf{b}_i) + \nu}{2} \right), \tag{A.2}$$

$$\mathbf{b}_i \mid \dots \sim \mathcal{N}_{m_2} \left( \mathbf{b}_i^*, \frac{\sigma^2}{\tau_i} \mathbf{W}_i \right), \tag{A.3}$$

$$\boldsymbol{\beta} \mid \dots \sim \mathcal{N}_{m_1}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta), \tag{A.4}$$

$$\sigma^2 \mid \dots \sim \mathcal{IG} \left( \frac{n + Nm_2 + a_0}{2}, \frac{b_0 + \sum_{i=1}^N \tau_i (S_i(\boldsymbol{\beta}, \mathbf{b}_i, \boldsymbol{\gamma}) + \mathbf{b}_i^T \boldsymbol{\Gamma}^{-1} \mathbf{b}_i)}{2} \right), \tag{A.5}$$

$$\boldsymbol{\Gamma} \mid \dots \sim \mathcal{IW} \left( \boldsymbol{\Omega}_0 + \sum_{i=1}^N \tau_i \mathbf{b}_i \mathbf{b}_i^T, N + c_0 \right), \tag{A.6}$$

where  $\mathbf{C}_i = \mathbf{C}_i(\boldsymbol{\gamma})$  and

$$\begin{aligned} \boldsymbol{\mu}_i^* &= \mathbf{M}_i(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \mathbf{C}_i \mathbf{O}_i^T (\mathbf{O}_i \mathbf{C}_i \mathbf{O}_i^T)^{-1} \mathbf{O}_i (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i)), \\ \mathbf{C}_i^{\text{mm}\cdot o} &= \mathbf{M}_i (\mathbf{I}_{n_i} - \mathbf{C}_i \mathbf{O}_i (\mathbf{O}_i \mathbf{C}_i \mathbf{O}_i^T)^{-1} \mathbf{O}_i^T) \mathbf{C}_i \mathbf{M}_i^T, \\ \mathbf{W}_i &= (\boldsymbol{\Gamma}^{-1} + \mathbf{Z}_i^T \mathbf{C}_i^{-1} \mathbf{Z}_i)^{-1}, \quad \mathbf{b}_i^* = \mathbf{W}_i \mathbf{Z}_i^T \mathbf{C}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}), \\ \boldsymbol{\Sigma}_\beta &= \left( \sigma^{-2} \sum_{i=1}^N \tau_i \mathbf{X}_i^T \mathbf{C}_i^{-1} \mathbf{X}_i + \mathbf{B}_0^{-1} \right)^{-1}, \\ \boldsymbol{\mu}_\beta &= \boldsymbol{\Sigma}_\beta \left( \sigma^{-2} \sum_{i=1}^N \tau_i \mathbf{X}_i^T \mathbf{C}_i^{-1} (\mathbf{Y}_i - \mathbf{Z}_i \mathbf{b}_i) + \mathbf{B}_0^{-1} \boldsymbol{\beta}_0 \right). \end{aligned}$$

The full conditional distributions of  $\boldsymbol{\gamma}$  and  $v$  do not have standard forms. They are proportional to the following functional forms:

$$f(\boldsymbol{\gamma} \mid \dots) \propto \left[ \prod_{i=1}^N |\mathbf{C}_i|^{-1/2} \right] \exp \left\{ - \frac{\sum_{i=1}^N \tau_i S_i(\boldsymbol{\beta}, \mathbf{b}_i, \boldsymbol{\gamma})}{2\sigma^2} \right\}, \tag{A.7}$$

$$f(v \mid \dots) \propto \left[ \frac{(v/2)^{v/2}}{\Gamma(v/2)} \right]^N \left[ \prod_{i=1}^N \tau_i^{v/2-1} \right] \exp \left\{ - \frac{v}{2} \sum_{i=1}^N \tau_i \right\} J_v. \tag{A.8}$$

### A.2. Implementation of the MCMC algorithm

Computing posterior model probabilities requires a high dimensional numerical integration over the posterior distribution  $p(\boldsymbol{\theta}, \mathbf{Y}^m, \mathbf{b}, \boldsymbol{\tau} \mid \mathbf{Y})$  in (5), which is analytically intractable. To deal with this difficulty, we use the MCMC approach. One sweep of the algorithm can be implemented as follows:

- Step 1: Using the Gibbs sampler to generate  $\mathbf{Y}_i^m, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\Gamma}, \tau_i$  and  $\mathbf{b}_i, i = 1, \dots, N$ , from the associated full conditional distributions (A.1)–(A.6).
- Step 2: Generate  $\boldsymbol{\gamma}$  from (A.7) using the M–H algorithm.
- Step 3: Generate  $v$  from (A.8) using the M–H algorithm.

To implement the MCMC algorithm at the  $(k + 1)$ st iteration in Step 2, we can transform  $\boldsymbol{\gamma}$  to  $\boldsymbol{\gamma}^* = (\gamma_1^*, \dots, \gamma_p^*) \in \mathbb{R}^p, \mathbb{R} = (-\infty, \infty)$ , where,  $\gamma_i^* = \log((1 + \gamma_i)/(1 - \gamma_i))$  ( $i = 1, \dots, p$ ). We then apply the M–H algorithm to  $g(\boldsymbol{\gamma}^* \mid \dots) = f(\boldsymbol{\gamma} \mid \dots) J_{\boldsymbol{\gamma}^*}$ , where  $J_{\boldsymbol{\gamma}^*} = \prod_{i=1}^p (2e^{\gamma_i^*} / (1 + e^{\gamma_i^*})^2)$  is the Jacobian of transforming  $\boldsymbol{\gamma}$  to  $\boldsymbol{\gamma}^*$ .

A  $p$ -dimensional multivariate normal distribution with mean  $\boldsymbol{\gamma}^{*(k)}$  and covariance matrix  $c^2 \boldsymbol{\Sigma}_{\boldsymbol{\gamma}^*}^{(k)}$  is chosen as the proposal distribution, where the scale  $c \approx 2.4/\sqrt{p}$ , as suggested in Gelman et al. (2004). The covariance matrix  $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}^*}^{(k)}$  can be estimated by inverting the sample information matrix given  $\boldsymbol{\gamma}^{*(k)}$ . Having obtained  $\boldsymbol{\gamma}^*$  from the M–H algorithm, we transform it back to  $\boldsymbol{\gamma}$  by  $\gamma_i = (e^{\gamma_i^*} - 1)/(e^{\gamma_i^*} + 1)$  ( $i = 1, \dots, p$ ), then transform  $\boldsymbol{\gamma}$  back to  $\boldsymbol{\phi}$  by inverting (3). In the case of AR(1),  $\phi_1 = \gamma_i$ . For the AR(2) process,  $\phi_1 = \gamma_1(1 - \gamma_2)$  and  $\phi_2 = \gamma_2$ . For the AR(3) process,  $\phi_1 = \gamma_1 - \gamma_1\gamma_2 - \gamma_2\gamma_3, \phi_2 = \gamma_2 - \gamma_1\gamma_3 + \gamma_1\gamma_2\gamma_3$  and  $\phi_3 = \gamma_3$ .

Similarly, we transform  $v$  to  $v^* = \log(1/v)$  and then apply the M–H algorithm to the function  $g(v^* | \dots) \propto f(v^* | \dots) J_{v^*}$ , where  $J_{v^*} = e^{-v^*}$ . The proposal distribution can be chosen as the truncated normal distribution with mean  $v^{*(k)}$ , variance  $\sigma_{v^*}^{2(k)}$ , and truncated region  $\mathbb{R} = (-10, 10)$ . Furthermore, an approximation of  $\sigma_{v^*}^{2(k)}$  is chosen as  $v^{(k)-2} I^{-1}(v^{(k)})$ . The Fisher information of  $v$ ,  $I(v)$ , is given by:

$$I(v) = \frac{1}{4} \sum_{i=1}^N \left[ \psi\left(\frac{v}{2}\right) - \psi\left(\frac{v+n_i^0}{2}\right) - \frac{2(v+2)}{v(v+n_i^0+2)} - \frac{2}{v} + \frac{4}{v+n_i^0} \right],$$

where  $\psi(x) = d^2/dx^2 \log(\Gamma(x))$  is the trigamma function. At the  $(k+1)$ st iteration, one can generate a truncated normal variate (Gelfand et al., 1992) by

$$v^{*(k+1)} = v^{*(k)} + \sigma_{v^*}^{(k)} \Phi^{-1} \left\{ \Phi\left(\frac{-10 - v^{*(k)}}{\sigma_{v^*}^{(k)}}\right) + U \left[ \Phi\left(\frac{10 - v^{*(k)}}{\sigma_{v^*}^{(k)}}\right) - \Phi\left(\frac{-10 - v^{*(k)}}{\sigma_{v^*}^{(k)}}\right) \right] \right\},$$

where  $\Phi$  denotes the standard normal cumulative distribution function and  $U$  denotes a random uniform  $(0, 1)$  variate.

We repeat Steps 1 to 3 until the sequence becomes stable. The convergence of the MCMC algorithm can be assessed by examining the MPSRF of Brooks and Gelman (1998) using multiple chains. After “burn-in”, we can use the remaining sample to estimate the parameters of interest.

## References

- Barndorff-Nielsen, O., Schou, G., 1973. On the parameterization of autoregressive models by partial autocorrelations. *J. Multivariate Anal.* 3, 408–419.
- Box, G.E.P., Jenkins, G.M., Reinsel, G.C., 1994. *Time Series Analysis Forecasting and Control*, third ed. Prentice-Hall, Englewood Cliffs, NJ.
- Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Statist.* 7, 434–455.
- Chi, E.M., Reinsel, G.C., 1989. Models for longitudinal data with random effects and AR(1) errors. *J. Amer. Statist. Assoc.* 84, 452–459.
- Chib, S., Carlin, E., 1999. On MCMC sampling in hierarchical models. *Statist. Comput.* 9, 17–26.
- D'yachkova, Y.D., Petkau, J., White, R., 1997. Longitudinal analysis for magnetic resonance imaging outcomes in multiple sclerosis clinical trials. *J. Biopharm. Statist.* 7, 501–531.
- Gelfand, A.E., Smith, A.F.M., 1990. Sampling based approaches to calculate marginal densities. *J. Amer. Statist. Assoc.* 85, 398–409.
- Gelfand, A.E., Hills, S.E., Racine-Poon, A., Smith, A.F.M., 1990. Illustration of Bayesian inference in normal data models using Gibbs sampling. *J. Amer. Statist. Assoc.* 85, 972–985.
- Gelfand, A.E., Smith, A.F.M., Lee, T.M., 1992. Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *J. Amer. Statist. Assoc.* 85, 523–532.
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. *Bayesian Data Analysis*. second ed. Chapman & Hall, CRC press, London, Boca Raton, FL.
- Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* 6, 721–741.
- Gill, P.S., 2000. A robust mixed linear model analysis for longitudinal data. *Statist. Med.* 19, 975–987.
- Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97–109.
- Keramidas, E.M., Lee, J.C., 1995. Selection of a covariance structure for growth curves. *Biometrical J.* 37, 783–797.
- Laird, N.M., Ware, J.H., 1982. Random effects models for longitudinal data. *Biometrics* 38, 963–974.
- Lange, K.L., Little, R.J.A., Taylor, J.M.G., 1989. Robust statistical modeling using the  $t$  distribution. *J. Amer. Statist. Assoc.* 84, 881–896.
- Lee, J.C., 1988. Prediction and estimation of growth curve with special covariance structures. *J. Amer. Statist. Assoc.* 83, 432–440.
- Lee, J.C., Lin, T.I., Lee, K.J., Hsu, Y.L., 2005. Bayesian analysis of Box-Cox transformed linear mixed models with ARMA( $p, q$ ) dependence. *J. Statist. Plann. Inference.* 133, 435–451.
- Lin, T.I., Lee, J.C., 2003. On modeling data from degradation sample paths over time. *Austral. J. Statist.* 45, 257–270.
- Lin, T.I., Lee, J.C., 2006. A robust approach to  $t$  linear mixed models applied to multiple sclerosis data. *Statist. Med.*, to appear.
- Little, R.J.A., Rubin, D.B., 2002. *Statistical Analysis With Missing Data*. second ed.. Wiley, New York.
- Liu, C.H., Rubin, D.B., 1998. Ellipsoidally symmetric extensions of the general location model for mixed categorical and continuous data. *Biometrika* 85, 673–688.
- Newton, M.A., Raftery, A.E., 1994. Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion). *J. Roy. Statist. Soc. B* 56, 3–48.
- Pinheiro, J.C., Liu, C.H., Wu, Y.N., 2001. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate  $t$  distribution. *J. Comput. Graph. Statist.* 10, 249–276.
- Rao, C.R., 1987. Prediction of future observations in growth curve models. *Statist. Sci.* 2, 434–471.
- Rochon, J., 1992. ARMA covariance structures with time heteroscedasticity for repeated measures experiments. *J. Amer. Statist. Assoc.* 87, 777–784.

- Rubin, D.B., 1976. Inference and missing data. *Biometrika* 63, 581–592.
- Schafer, J.L., Yucel, R.M., 2002. Computational strategies for multivariate linear mixed-effects models with missing values. *J. Comput. Graph. Statist.* 11, 437–457.
- Wakefield, J.C., Smith, A.F.M., Racine-Poon, A., Gelfand, A.E., 1994. Bayesian analysis of linear and non-linear model by using Gibbs sampler. *Appl. Statist.* 43, 201–221.
- Zeger, S.L., Karim, M.R., 1991. Generalized linear models with random effects; a Gibbs sampling approach. *J. Amer. Statist. Assoc.* 86, 79–86.
- Zellner, A., 1976. Bayesian and non-Bayesian analysis of the regression model with multivariate student-*t* error terms. *J. Amer. Statist. Assoc.* 71, 400–405.