Statistica Sinica **21** (2011), 1857-1880
doi:http://dx.doi.org/10.5705/ss.2009.306

ESTIMATION IN MULTIVARIATE t LINEAR MIXED MODELS FOR MULTIPLE LONGITUDINAL DATA

Wan-Lun Wang and Tsai-Hung Fan

Feng Chia University and National Central University

Abstract: The multivariate linear mixed model (MLMM) is a frequently used tool for a joint analysis of more than one series of longitudinal data. Motivated by a concern of sensitivity to potential outliers or data with longer-than-normal tails and possible serial correlation, we develop a robust generalization of the MLMM that is constructed by using the multivariate t distribution and a parsimonious AR(p) dependence structure for the within-subject errors. A score test for the inspection of autocorrelation among within-subject errors is derived. A hybrid ECME-scoring procedure is developed for computing the maximum likelihood estimates with standard errors as a by-product. The methodology is illustrated through an application to a set of AIDS data and several simulation studies.

Key words and phrases: AR(p), ECME algorithm, outliers, random effects, score test.

1. Introduction

The linear mixed model (LMM), originally proposed by Laird and Ware (1982), has been broadly studied and widely used for the analysis of single-outcome longitudinal data. The popularity of such a model arises from its ability to account for between- and within-subjects correlations, together with ready implementation through commonly available software, e.g., SAS (SAS (2001)), procedure `MIXED`, and R (R Development Core Team (2008)), library `NLME`. Comprehensive reviews that cover methodological and computational aspects of the LMM are contained in books by Fitzmaurice, Laird, and Ware (2004) and Hedeker and Gibbons (2006), among others. Diggle et al. (2002) studied the GEE methods for fitting marginal linear models to clustered longitudinal data.

In many biomedical studies and clinical trials, it is quite common that repeated measures are collected on two or more response variables and are often referred to as multivariate longitudinal data. There has been growing interest in pursuing a multivariate generalization of LMM for dealing with such data. Shah, Laird, and Schoenfeld (1997) extended the LMM to the multivariate linear mixed model (MLMM), to allow the analysis of multiple longitudinal outcomes, and developed an EM algorithm (Dempster, Laird, and Rubin (1977)) to estimate the

model parameters. Schafer and Yucel (2002) described several improved EM-type and Markov chain Monte Carlo (MCMC) procedures for multiple imputation of missing values in MLMM. Fieuws and Verbeke (2004) applied a joint modelling strategy to investigate questions about the "association of the evolutions" and the "evolution of the association". Roy (2006) discussed how to estimate the correlation coefficient between two response variables with repeated measurements. More recently, Wang and Fan (2010) provided some additional tools, including estimation, testing, and prediction, for the MLMM with autoregressive (AR) errors.

In the framework of LMMs and MLMMs, the random effects and the within-subject errors are routinely assumed normal for mathematical convenience. However, such an assumption is not always realistic because of the presence of atypical observations. To remedy this weakness, such authors as Zellner (1976) and Lange, Little, and Taylor (1989) considered the use of the multivariate t distribution (Hogan and Laird (1997)) that contains a harmonizing parameter ν (called the degrees of freedom), for robust estimation of linear regression models. The value of ν, which controls the thickness of the tails of the distribution, is directly related to the degree of robustness of inference, and smaller ν yields higher robustness. Recently, Pinheiro, Liu, and Wu (2001) proposed a t linear mixed model (tLMM) and demonstrated its robustness against outliers through an application to orthodontic data and extensive simulations. Further work in this direction is Lin and Lee (2006, 2007), and Song, Zhang, and Qu (2007).

Our main objective is to extend the existing tLMM to a multivariate version, called the multivariate t linear mixed model (MtLMM) for properly modeling multi-response longitudinal data with thick tails. To our knowledge, the proposed model, including the MLMM as a limiting/special case ($\nu \to \infty$), has never been considered in the literature. In addition, a stationary AR process of order p for the within-subject errors is considered to account for the extra autocorrelation not caused by random effects. It is worthwhile mentioning that the pure AR structure can be easily extended to a much richer autoregressive moving average (ARMA) family, but a high-order AR model is usually enough due to the fact that longitudinal data are often short time series. Besides, the white noise process is included as a special case of AR models when the AR parameters approach zero.

The rest of this paper is organized as follows. In Section 2, we describe a set of bivariate repeated measures from the AIDS Clinical Trials Group (ACTG) 175 study in which some outliers appear to be present. In Section 3, we define notation, formulate the model, and derive a score test statistic to assess the existence of within-subject autocorrelation. In Section 4, ML estimation is carried out by using a hybrid maximization scheme, which combines the stability of the Expectation Conditional Maximization Either (ECME) algorithm

(Liu and Rubin (1994)) with the rapid convergence feature of the Fisher scoring procedure. Standard errors are obtained by inverting the expected information matrix. The model selection procedure for assessing the fits of candidate models is also addressed. In Section 5, we illustrate these techniques with the clinical study preliminarily analyzed in Section 2, and demonstrate the robust property of MtLMM under various levels of perturbation to the data. Concluding remarks are in Section 6. Proofs of the theoretical results, along with a small simulation study, are deferred to the supplementary document available on-line at `http://www.stat.sinica.edu.tw/statistica`.

2. A Motivating Example: the ACTG 175 Data

To motivate the proposed model, consider a preliminary analysis of the ACTG 175 data originally reported by Hammer et al. (1996). This clinical trial study involved a total of 2467 HIV-1-infected participants who were recruited from 43 AIDS Clinical Trials Units and 9 National Hemophilia Foundation sites in the United States and Puerto Rico. The recruitment period was from December 1991 to October 1992, and the recruited patients were measured repeatedly until the end of the follow-up period, November 1994. On each patient visit, two important immunologic measurements, namely CD4 and CD8 cell counts (per cubic millimeter), were collected at 2, 4, and 8 weeks, and then per protocol about every 12 weeks thereafter. Patients were randomly assigned to one of the four treatments: zidovudine alone, zidovudine plus didanosine, zidovudine plus zalcitabine, and didanosine alone. The primary objective of this clinical study was to understand the amount of derivative tracking in the measurements of immunologic markers; and to compare monotherapy with these four treatments in HIV-1-infected patients on the basis of time to progression to AIDS or death. Many analyses have been done on the natural history of CD4 repeated measures as well as the time-to-event data by using the model-based or semiparametric likelihood-based approaches (see Hogan and Laird (1997); Song, Davidian, and Tsiatis (2002), for example).

For illustration, a total of $N = 30$ patients were randomly selected, and missingness was not considered in this. As pointed out by Song, Davidian, and Tsiatis (2002), only nine events occurred before week 12 and the observed profiles were followed by a decline after week 12, so we considered the post-week-12 data henceforth. We focused on samples measured from 20 to 128 weeks. Hence, the measurement times were commonly across patients for the selected data set. To achieve constant variance of CD4 and CD8 cell counts, we take the base 10 logarithm of all repeated measures and write $\mathbf{y}_{i1} = \log_{10}(\text{CD4})_i$ and $\mathbf{y}_{i2} = \log_{10}(\text{CD8})_i$ for patient i, respectively.

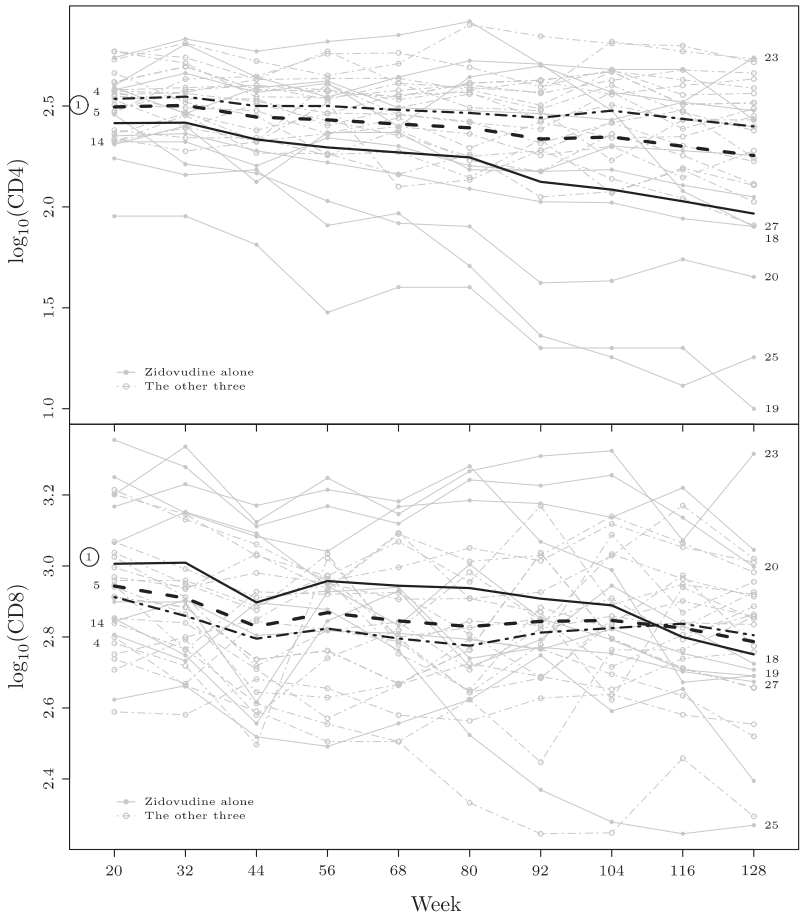

Figure 1. Trajectories of $\log_{10}$(CD4) and $\log_{10}$(CD8) for 30 randomly selected patients along with their mean profiles based on different treatment groups (thick dashed line for the entire group; solid lines for zidovudine alone; and dot-dashed lines for the other three medicines).

Figure 1 depicts the trajectories of patients' 10 scheduled visits along with their mean profiles. It can be observed that the trend of population mean profiles decline linearly over time. Moreover, we found that patients differ in their initial levels and time trends. Besides, the mean profiles of zidovudine alone are different from the other three therapies. Thus, the treatment and time-varying covariates should be considered in evaluating the effects of depletion of CD4 and CD8 cell counts.

Assuming a linear trend for the population average, and subject-specific intercepts and slopes for the random effects, $\mathbf{y}_{i1}$ and $\mathbf{y}_{i2}$ are fitted separately by

the LMMs

$$\mathbf{y}_{ij} = [\mathbf{1}_{10} : \text{treat}_i\mathbf{1}_{10} : \mathbf{t}_i : \text{treat}_i\mathbf{t}_i]\boldsymbol{\beta}_j + [\mathbf{1}_{10} : \mathbf{t}_i]\mathbf{b}_{ij} + \mathbf{e}_{ij},\ i = 1, \ldots, 30,\ j = 1, 2. \tag{2.1}$$

Here $\mathbf{1}_{10} = (1, \ldots, 1)^{\mathrm{T}}$; treat_i is a treatment indicator (1 = zidovudine alone; 0 = the other three therapies); $\mathbf{t}_i = (t_{i1}, \ldots, t_{i10})^{\mathrm{T}}$ with $t_{it} = (\text{week}_{it} - 8)/12$; $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, \beta_{j2}, \beta_{j3})^{\mathrm{T}}$ is the regression-coefficient vector in fixed effects on $\mathbf{y}_{ij}$, in which β_{j0} and β_{j2} denote the intercepts and slopes, respectively, for the other three therapies; β_{j1} and β_{j3} denote the difference in intercepts and slopes between zidovudine alone and the other three therapies; $\mathbf{b}_{ij} = (b_{ij0}, b_{ij1})^{\mathrm{T}}$ and $\mathbf{e}_{ij}$ are the normally distributed random effects and within-subject errors, respectively, on $\mathbf{y}_{ij}$.

Figure 2 depicts scatter plots of estimated empirical Bayes estimates (Laird and Ware (1982)) for the random slopes and standardized residuals together with their normal quantile-quantile (Q-Q) plots. From this figure, we have three comments.

(i) The sample correlation coefficient for the estimates of two random slopes, b_{i11} and b_{i21}, is 0.61, revealing that patients with a larger decline in CD4 also have a larger decline in CD8, and vice versa.

(ii) The sample correlation coefficient for residuals, $\mathbf{e}_{i1}$ and $\mathbf{e}_{i2}$, is 0.64, implying that CD4 and CD8 should be analyzed jointly.

(iii) The Q-Q plots for the estimates of random slopes and residuals exhibit heavy-tailed behavior, suggesting that the normality assumption for random effects and within-subject errors might be inappropriate.

Accordingly, we look to establish a robust generalization of MLMM by considering a multivariate t distribution, denoted by $t_a(\boldsymbol{\mu}, \boldsymbol{\Omega}, \nu)$, with density

$$f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Omega}, \nu) = \frac{\Gamma\big((\nu+a)/2\big)|\boldsymbol{\Omega}|^{-1/2}}{\Gamma(\nu/2)(\pi\nu)^{a/2}}\Big(1+\frac{(\mathbf{y}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Omega}^{-1}(\mathbf{y}-\boldsymbol{\mu})}{\nu}\Big)^{-(\nu+a)/2}, \mathbf{y} \in \mathcal{R}^a.$$

A detailed account of mathematical properties and estimation methods for this distribution can be found in Kotz and Nadarajah (2004) and Nadarajah and Kotz (2005, 2008). In Section 5, we return to these data.

3. Notation and Setting

3.1. Model formulation

Let $\mathbf{Y}_i = [\mathbf{y}_{i1} : \cdots : \mathbf{y}_{ir}]$ be an $n_i \times r$ response matrix containing r observable vectors $\mathbf{y}_{ij} = (y_{ij1}, \ldots, y_{ijn_i})^{\mathrm{T}}$ of length n_i, the jth response variable from subject i over occasions $t = 1, \ldots, n_i$ ($i = 1, \ldots, N$; $j = 1, \ldots, r$). Let $\mathbf{E}_i = [\mathbf{e}_{i1} : \cdots : \mathbf{e}_{ir}]$

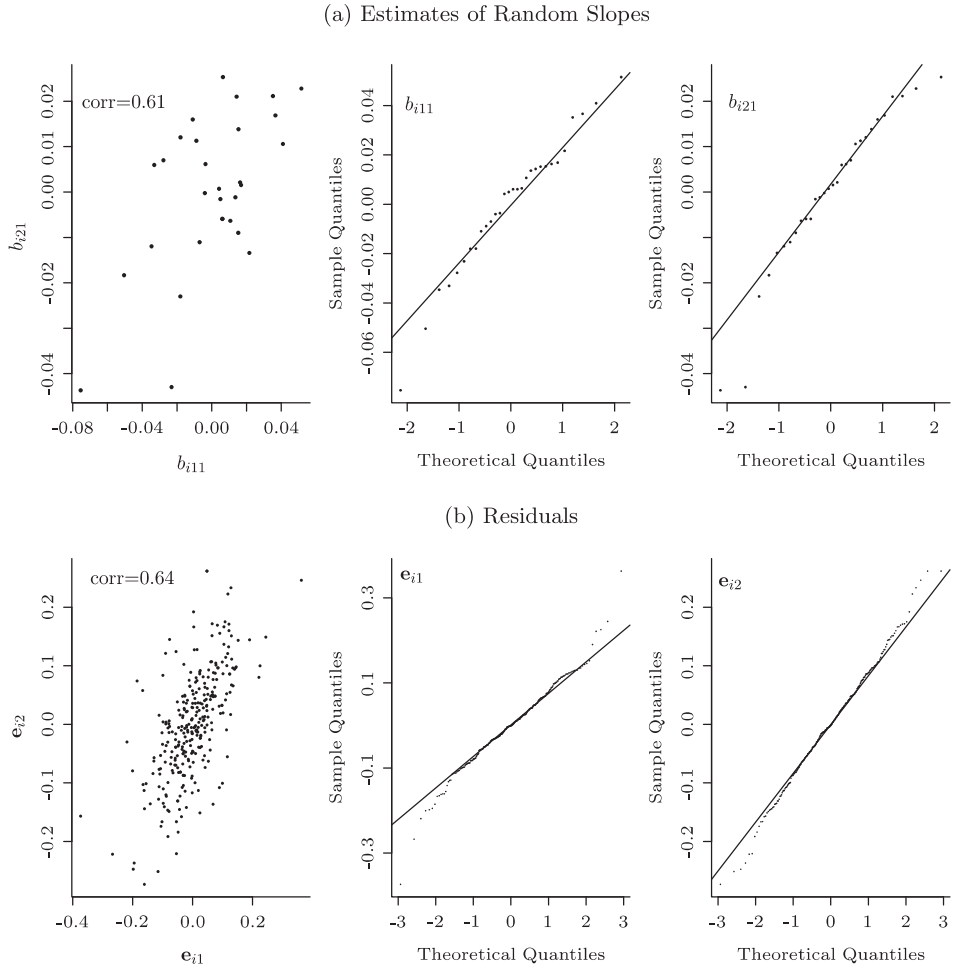

Figure 2. Scatter plots (left panel) and Q-Q plots (right panel) of (a) estimates of random slopes and (b) residuals for $\log_{10}$(CD4) and $\log_{10}$(CD8), respectively, based on LMMs with random intercepts, random slopes and uncorrelated errors.

be an $n_i \times r$ error-component matrix, where the jth column $\mathbf{e}_{ij}$ is the within-subject error corresponding to $\mathbf{y}_{ij}$. Let $\mathbf{X}_{ij}$, which is obtained as functions of the basic covariates, be an $n_i \times q_1$ design matrix for fixed effects, and let $\mathbf{Z}_{ij}$ be an $n_i \times q_2$ design matrix for random effects. To analyze the unbalanced longitudinal data, different numbers of measurements and/or unequal sets of occasions of each subject, we write $\mathbf{X}_i = \text{diag}\{\mathbf{X}_{i1}, \ldots, \mathbf{X}_{ir}\}$ and $\mathbf{Z}_i = \text{diag}\{\mathbf{Z}_{i1}, \ldots, \mathbf{Z}_{ir}\}$, both known full-rank block-diagonal matrices. We set out to model the relationship between repeated measures and covariates, as well as investigate the association of evolutions among all outcomes and the evolutions of each outcome.

We utilize the vec$(\cdot)$ operator that vectorizes a matrix by stacking its columns vertically, to obtain $\mathbf{y}_i = \mathrm{vec}(\mathbf{Y}_i) = (\mathbf{y}_{i1}^{\mathrm{T}}, \ldots, \mathbf{y}_{ir}^{\mathrm{T}})^{\mathrm{T}}$ and $\mathbf{e}_i = \mathrm{vec}(\mathbf{E}_i) = (\mathbf{e}_{i1}^{\mathrm{T}}, \ldots, \mathbf{e}_{ir}^{\mathrm{T}})^{\mathrm{T}}$. Note that, without balanced data, there is indeed no advantage in modeling $\mathbf{Y}_i$ directly. We take the MtLMM for the response vector $\mathbf{y}_i$ to be of the form

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i \quad \text{with} \quad \begin{bmatrix}\mathbf{b}_i \\ \mathbf{e}_i\end{bmatrix} \sim t_{(q_2+n_i)r}\left(\begin{bmatrix}\mathbf{0} \\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_i\end{bmatrix}, \nu_i\right), \tag{3.1}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^{\mathrm{T}}, \ldots, \boldsymbol{\beta}_r^{\mathrm{T}})^{\mathrm{T}}$ is the regression parameter, with each q_1-vector $\boldsymbol{\beta}_j$ used to describe the fixed effects of the jth outcome, and $\mathbf{b}_i = (\mathbf{b}_{i1}^{\mathrm{T}}, \ldots, \mathbf{b}_{ir}^{\mathrm{T}})^{\mathrm{T}}$ is a q_2r-vector of random effects. We assume that $\mathbf{D} = [D_{jj'}]$ is a $q_2r \times q_2r$ symmetric positive-definite matrix, where $D_{jj'}$ is a $q_2 \times q_2$ partition matrix, and in particular for $j = j'$, D_{jj} is a unstructured covariance matrix in the random effects for the jth outcome only, and $\mathbf{R}_i$ is an $n_ir \times n_ir$ structured covariance matrix in error components. We take $\nu_i = \nu$, for all i, and further assume that the joint distributions of $(\mathbf{b}_i^{\mathrm{T}}, \mathbf{e}_i^{\mathrm{T}})^{\mathrm{T}}$ for distinct subjects are independent.

For the within-subject error $\mathbf{e}_i$, if we do not make any special assumption about the covariance matrix for each row and column of $\mathbf{E}_i$, then the unknown parameters in $\mathbf{R}_i$ are too numerous to be accurately estimated. Accordingly, it is convenient to impose a parsimonious structure on $\mathbf{R}_i$. Here, the row vectors of $\mathbf{E}_i$, the errors at different measurement times, are assumed to be serially correlated and to have distribution $t_r(\mathbf{0}, \boldsymbol{\Sigma}, \nu)$, where $\boldsymbol{\Sigma} = [\sigma_{jj'}]$ describes the unstructured covariance among r outcomes. If the r outcomes come from different study environments, then it might be appropriate to suppose $\boldsymbol{\Sigma}$ to be diagonal; this is not common, however. Each column of $\mathbf{E}_i$ has distribution $t_{n_i}(\mathbf{0}, \sigma_{jj}\mathbf{C}_i, \nu)$, where $\mathbf{C}_i$ captures a time-dependence structure over n_i occasions on each outcome. The AR(p) dependence structure $\mathbf{C}_i = \mathbf{C}_i(\boldsymbol{\phi}) = [\rho_{|t-t'|}(\boldsymbol{\phi})]$, $t, t' = 1, \ldots, n_i$, is used to address autocorrelation over time, where

$$\rho_s(\boldsymbol{\phi}) = \rho_s = \phi_1\rho_{s-1} + \cdots + \phi_p\rho_{s-p}, \quad \rho_0 = 1, \quad (s = 0, \ldots, n_i - 1),$$

see Box, Jenkins, and Reinsel (1994). For the pure AR model, the admissible values of $\boldsymbol{\phi}$ are restricted in a p-dimensional hypercube $\mathbb{C}^p$; to ensure the stationarity of the AR model, the roots of $1 - \phi_1\mathbb{B} - \phi_2\mathbb{B}^2 - \cdots - \phi_p\mathbb{B}^p = 0$ must lie outside the unit circle, where $\mathbb{B}$ is a backward shift operator such that $\mathbb{B}^v\rho_s = \rho_{s-v}$, $v = 0, \ldots, p$.

Under this consideration, the covariance structure of $\mathbf{e}_i$ is $\boldsymbol{\Sigma} \otimes \mathbf{C}_i$, and the response $\mathbf{y}_i \sim t_{n_ir}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Lambda}_i, \nu)$, where $\boldsymbol{\Lambda}_i = \boldsymbol{\Lambda}_i(\mathbf{D}, \boldsymbol{\Sigma}, \boldsymbol{\phi}) = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^{\mathrm{T}} + \boldsymbol{\Sigma} \otimes \mathbf{C}_i$ are implicit functions depending on $\boldsymbol{\phi}$ and distinct elements of $\mathbf{D}$ and $\boldsymbol{\Sigma}$. If $\nu > 1$, then $\mathbf{X}_i\boldsymbol{\beta}$ is the mean of $\mathbf{y}_i$, and if $\nu > 2$, then $\nu(\nu-2)^{-1}\boldsymbol{\Lambda}_i$ is its variance-covariance matrix. Note that the tLMM of Lin and Lee (2006), specified by

$\mathbf{y}_{ij} \sim t_{n_i}(\mathbf{X}_{ij}\boldsymbol{\beta}_j, \mathbf{Z}_{ij}D_{jj}\mathbf{Z}_{ij}^{\mathrm{T}} + \sigma_{jj}\mathbf{C}_i, \nu)$, can be viewed as a univariate case of our model. Let

$$\Delta_i = \Delta_i(\boldsymbol{\beta}, \mathbf{D}, \boldsymbol{\Sigma}, \boldsymbol{\phi}) = \boldsymbol{\epsilon}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\epsilon}_i \tag{3.2}$$

denote the Mahalanobis distance between $\mathbf{y}_i$ and $\mathbf{X}_i\boldsymbol{\beta}$, where $\boldsymbol{\epsilon}_i = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}$. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\alpha}^{\mathrm{T}})^{\mathrm{T}}$ be the vector of unknown parameters, where $\boldsymbol{\alpha} = (\boldsymbol{\omega}^{\mathrm{T}}, \nu)^{\mathrm{T}}$ with $\boldsymbol{\omega} = (\mathrm{vech}(\mathbf{D})^{\mathrm{T}}, \mathrm{vech}(\boldsymbol{\Sigma})^{\mathrm{T}}, \boldsymbol{\phi}^{\mathrm{T}})^{\mathrm{T}}$. Then the log-likelihood function of $\boldsymbol{\theta}$ can be formed by summing over the logarithms of multivariate t densities at each $\mathbf{y}_i$, written as $\ell = \sum_{i=1}^{N} \ell_i$, where

$$\ell_i = \log\Gamma\Big(\frac{\nu + n_i r}{2}\Big) - \log\Gamma(\frac{\nu}{2}) - \frac{n_i r}{2}\log(\pi\nu) - \frac{1}{2}\log|\boldsymbol{\Lambda}_i| - \frac{\nu + n_i r}{2}\log\Big(1 + \frac{\Delta_i}{\nu}\Big).$$

The first theorem is useful for obtaining the Fisher information matrix. The score vector $\mathbf{s}_{\boldsymbol{\theta}}$, the information matrix $\mathbf{J}_{\boldsymbol{\theta}\boldsymbol{\theta}}$, and the proof of Theorem 1 are sketched in the supplementary material.

Theorem 1. *Under* (3.1)*, we have* $\Delta_i \sim n_i r\mathcal{F}(n_i r, \nu)$ *and* $\nu/(\nu+\Delta_i) \sim$ *Beta*$(\nu/2, n_i r/2)$*. Then*

(i) $E\big[(\nu + \Delta_i)^{-1}\big] = (\nu + n_i r)^{-1}$;

(ii) $E\big[(\nu + \Delta_i)^{-2}\big] = \big[\nu(\nu + n_i r)(\nu + n_i r + 2)\big]^{-1}(\nu + 2)$;

(iii) $E\big[(\nu + \Delta_i)^{-1}\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i^{\mathrm{T}}\big] = (\nu + n_i r)^{-1}\boldsymbol{\Lambda}_i$;

(iv) $E\big[(\nu + \Delta_i)^{-2}\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i^{\mathrm{T}}\big] = \big[(\nu + n_i r)(\nu + n_i r + 2)\big]^{-1}\boldsymbol{\Lambda}_i$.

3.2. The score test for autocorrelation

One might be interested in whether autocorrelation exists among the within-subject errors of each outcome. A direct check for possible autocorrelation concentrates on the simplest case in which $\mathbf{C}_i = \mathbf{C}_i(\phi_1)$ has an AR(1) dependence of the form $(1-\phi_1^2)^{-1}[\phi_1^{|t-t'|}]$, $|\phi_1| < 1$. We provide a score test procedure to test the null hypothesis $H_0 : \phi_1 = 0$ against the alternative hypothesis $H_1 : \phi_1 \neq 0$. The reason for administering the score test is that it only requires an evaluation of the ML estimates under the null model and, like the likelihood ratio test (LRT) and Wald's test, it is also Chernoff-consistent (Shao (2003)).

Rewrite $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\eta}^{\mathrm{T}}, \phi_1)^{\mathrm{T}}$, where $\boldsymbol{\eta} = (\mathrm{vech}(\mathbf{D})^{\mathrm{T}}, \mathrm{vech}(\boldsymbol{\Sigma})^{\mathrm{T}}, \nu)^{\mathrm{T}}$ represents a vector that includes the distinct parameters of all covariance components and the degrees of freedom. The block-partitioned matrix $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}$ is a submatrix of (S2.1) rearranged according to the columns and rows of the respective parameters $(\boldsymbol{\eta}^{\mathrm{T}}, \phi_1)$, written as

$$\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} = \begin{bmatrix} \mathbf{J}_{\boldsymbol{\eta}\boldsymbol{\eta}} & \mathbf{J}_{\boldsymbol{\eta}\phi_1} \\ \mathbf{J}_{\boldsymbol{\eta}\phi_1}^{\mathrm{T}} & J_{\phi_1\phi_1} \end{bmatrix}.$$

Let $\hat{\boldsymbol{\theta}}_0 = (\hat{\boldsymbol{\beta}}_0^{\mathrm{T}}, \hat{\boldsymbol{\eta}}_0^{\mathrm{T}}, 0)^{\mathrm{T}}$ be the ML estimates of $\boldsymbol{\theta}$ under the null model, where $\hat{\boldsymbol{\eta}}_0 = (\mathrm{vech}(\hat{\mathbf{D}}_0)^{\mathrm{T}}, \mathrm{vech}(\hat{\boldsymbol{\Sigma}}_0)^{\mathrm{T}}, \hat{\nu}_0)^{\mathrm{T}}$. Then the score test statistic is

$$\lambda_s = \Big[\frac{\partial \ell}{\partial \boldsymbol{\theta}}\Big]_{\hat{\boldsymbol{\theta}}_0}^{\mathrm{T}} \Big[\mathbf{J}_{\boldsymbol{\theta}\boldsymbol{\theta}}\Big]_{\hat{\boldsymbol{\theta}}_0}^{-1} \Big[\frac{\partial \ell}{\partial \boldsymbol{\theta}}\Big]_{\hat{\boldsymbol{\theta}}_0} = \frac{[\partial \ell / \partial \phi_1]^2_{\hat{\boldsymbol{\theta}}_0}}{[J_{\phi_1 \phi_1 \cdot \boldsymbol{\eta}}]_{\hat{\boldsymbol{\theta}}_0}}, \tag{3.3}$$

in which the denominator of the right hand side can be calculated as $J^0_{\phi_1\phi_1} - \mathbf{J}^{0^{\mathrm{T}}}_{\boldsymbol{\eta}\phi_1} \mathbf{J}^{0^{-1}}_{\boldsymbol{\eta}\boldsymbol{\eta}} \mathbf{J}^0_{\boldsymbol{\eta}\phi_1}$, where $J^0_{\phi_1\phi_1}$, $\mathbf{J}^0_{\boldsymbol{\eta}\phi_1}$, and $\mathbf{J}^0_{\boldsymbol{\eta}\boldsymbol{\eta}}$ are $J_{\phi_1\phi_1}$, $\mathbf{J}_{\boldsymbol{\eta}\phi_1}$, and $\mathbf{J}_{\boldsymbol{\eta}\boldsymbol{\eta}}$ evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_0$. To evaluate the numerator of the right hand side of (3.3), we first let $\hat{\mathbf{e}}^0_i = \mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}_0 - \mathbf{Z}_i\hat{\mathbf{b}}^0_i$ and $\hat{\mathbf{b}}^0_i = \hat{\mathbf{D}}_0\mathbf{Z}_i^{\mathrm{T}}\hat{\boldsymbol{\Lambda}}^{-1}_{i0}(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}_0)$ denote the estimated "residuals" and "random effects" for subject i under the null model. We further define $\hat{\mathbf{u}}_i = (\hat{\boldsymbol{\Sigma}}_0^{-1} \otimes \mathbf{I}_{n_i})\hat{\mathbf{e}}^0_i$, $\hat{\Delta}_{i0} = (\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}_0)^{\mathrm{T}}\hat{\boldsymbol{\Lambda}}^{-1}_{i0}(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}_0)$ and $\dot{\mathbf{C}}_i(0) = \mathbf{L}_i + \mathbf{L}_i^{\mathrm{T}}$ with $\hat{\boldsymbol{\Lambda}}_{i0} = \mathbf{Z}_i\hat{\mathbf{D}}_0\mathbf{Z}_i^{\mathrm{T}} + (\hat{\boldsymbol{\Sigma}}_0 \otimes \mathbf{I}_{n_i})$, $\mathbf{I}_{n_i}$ an identity matrix of order n_i, and $\mathbf{L}_i$ an $n_i \times n_i$ matrix with the entries of 1 on the first super-diagonal and 0 otherwise. Since $[\partial \ell / \partial \boldsymbol{\theta}]_{\hat{\boldsymbol{\theta}}_0}$ has all zero entries except for the first partial derivative of ℓ with respect to ϕ_1, we get

$$\Big[\frac{\partial \ell}{\partial \phi_1}\Big]_{\hat{\boldsymbol{\theta}}_0} = \frac{1}{2}\sum_{i=1}^{N}\Big\{\frac{(\hat{\nu}_0 + n_i r)}{(\hat{\nu}_0 + \hat{\Delta}_{i0})}\hat{\mathbf{u}}_i^{\mathrm{T}}(\hat{\boldsymbol{\Sigma}}_0 \otimes \dot{\mathbf{C}}_i(0))\hat{\mathbf{u}}_i - \mathrm{tr}\big(\hat{\boldsymbol{\Lambda}}^{-1}_{i0}(\hat{\boldsymbol{\Sigma}}_0 \otimes \dot{\mathbf{C}}_i(0))\big)\Big\}.$$

To study the asymptotic properties of the score test as well as the ML estimates $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^{\mathrm{T}}, \hat{\boldsymbol{\alpha}}^{\mathrm{T}})^{\mathrm{T}}$, obtained by the proposed algorithm described in the next section, we make the following assumptions:

(i) Model (3.1) is correct, and the n_i's are bounded above.
(ii) The parameter spaces for $\boldsymbol{\beta}$, $\boldsymbol{\omega}$, and ν are compact sets of $\mathbb{R}^{q_1 r}$, $\mathbb{R}^g$, and $\mathbb{R}^+$, and the true value of $\boldsymbol{\theta}$ is in the interior of the parameter space of $\boldsymbol{\theta}$.
(iii) As $N \to \infty$, $N^{-1}\mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\beta}} \to \mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}$, $N^{-1}\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} \to \mathbf{I}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}$, and for any $q_1 r$-vector $\mathbf{a} \neq \mathbf{0}$,

$$\frac{\max_i \{\mathbf{a}^{\mathrm{T}}\mathbf{X}_i^{\mathrm{T}}\mathbf{X}_i\mathbf{a}\}}{\mathbf{a}^{\mathrm{T}}\left(\sum_{i=1}^N \mathbf{X}_i^{\mathrm{T}}\mathbf{X}_i\right)\mathbf{a}} \to 0.$$

Theorem 2. *Under Assumptions* (i)–(iii),

(i) $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}$, $\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}})$, *and* $\sqrt{N}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}^{-1}_{\boldsymbol{\alpha}\boldsymbol{\alpha}})$, *as* $N \to \infty$.
(ii) *If* H_0 *is true, then the score test statistic* λ_s *converges in distribution to a chi-square with one degree of freedom.*

Theorem 2 follows directly from the Weak Law of Large Numbers, the Central Limit Theorem, and Slutsky's Theorem. A value of λ_s that rejects the null model does not imply the alternative model is enough, but suggests that autocorrelation may be present in the $\mathbf{e}_{ij}$'s. Then, a higher-order AR model might

be more suitable. Moreover, if the autoregressive parameter ϕ_1 is close to zero instead, the within-subject errors of each outcome could be uncorrelated.

4. Computational Aspects

4.1. Maximum likelihood estimation

To estimate the unknown parameters at (3.1), we first indicate how to carry out ML estimation via EM-type algorithms. The EM algorithm originally proposed by Dempster, Laird, and Rubin (1977) has several appealing features including monotone convergence, with each iteration increasing the likelihood, and simplicity of implementation. However, ML estimation at (3.1) is complicated enough that the EM algorithm is computationally difficult at the M-step. Besides, the degrees of freedom ν, as well as the AR parameters $\boldsymbol{\phi}$ in the multivariate t version of linear mixed models, are difficult to estimate. To go further, we apply an extension of the EM algorithm, called the ECME (Liu and Rubin (1994)) algorithm, which shares the appealing features of both EM and Expectation Conditional Maximization (ECM, Meng and Rubin (1993)) and has typically a faster convergence rate than either EM or ECM measured in iterations or computer time. The ECME algorithm proceeds to estimate parameters by replacing the M-steps of EM with *either* CM-steps that maximize a sequence of constrained Q functions, as in ECM, *or* CML-steps that maximize the correspondingly constrained actual likelihood function.

Note that the model at (3.1) can utilize random τ_i's in a three-level normal-normal-gamma hierarchical specification as

$$\begin{aligned} \mathbf{y}_i|(\mathbf{b}_i,\tau_i) &\sim N_{n_i r}\big(\mathbf{X}_i\boldsymbol{\beta}+\mathbf{Z}_i\mathbf{b}_i,\tau_i^{-1}\mathbf{R}_i\big),\\ \mathbf{b}_i|\tau_i &\sim N_{q_2 r}(\mathbf{0},\tau_i^{-1}\mathbf{D}), \quad \tau_i \sim \text{Gamma}(\frac{\nu}{2},\frac{\nu}{2}). \end{aligned} \tag{4.1}$$

Recall from (3.1) that $\mathbf{b}_i|\tau_i$ and $\mathbf{e}_i|\tau_i$ are implicitly assumed to be independent. If the conditional distribution of $\mathbf{y}_i$ given τ_i is taken to be $N_{n_i r}(\mathbf{X}_i\boldsymbol{\beta},\boldsymbol{\Lambda}_i/\tau_i)$, the unconditional distribution of $\mathbf{y}_i$ is $t_{n_i r}(\mathbf{X}_i\boldsymbol{\beta},\boldsymbol{\Lambda}_i,\nu)$.

Treating the unobservable random effects $\mathbf{b}_i$ and latent variables τ_i as "missing" data, we have the complete data $\{(\mathbf{y}_i,\mathbf{b}_i,\tau_i),\ i=1,\ldots,N\}$. The complete-data log-likelihood function for all subjects, omitting constant terms, is

$$\begin{aligned} \ell_c(\boldsymbol{\theta}) = \sum_{i=1}^{N}\ell_c^{[i]}(\boldsymbol{\theta}) = \sum_{i=1}^{N}\frac{1}{2}\Big\{ &\log|\mathbf{R}_i^{-1}| + \log|\mathbf{D}^{-1}| - \tau_i\big[\mathbf{e}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\mathbf{e}_i + \mathbf{b}_i^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{b}_i\big]\\ &+\nu\log(\frac{\nu}{2}) - 2\log\Gamma(\frac{\nu}{2}) + \nu\big(\log\tau_i - \tau_i\big)\Big\}. \end{aligned}$$

To ensure the admissibility of $\boldsymbol{\phi}$ and stabilize the estimating procedure, we follow Barndorff-Nielsen and Schou (1973) to reparameterize $\boldsymbol{\phi}$ as

$$\begin{aligned}&\phi_p^{(p)} = \pi_p,\\ &\phi_v^{(p)} = \phi_v^{(p-1)} - \pi_p\phi_{p-v}^{(p-1)} = \pi_v - \pi_{v+1}\phi_1^{(v)} - \pi_{v+2}\phi_2^{(v+1)} - \cdots - \pi_p\phi_{p-v}^{(p-1)},\end{aligned} \tag{4.2}$$

where $\phi_v^{(p)}$ is the vth AR parameter under AR(p) model, and $\pi_v = \phi_v^{(v)}$ is the partial autocorrelation at lag v, for $v = 1, \ldots, p-1$. The reparameterization (4.2) is a one-to-one transformation that maps the AR parameters $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p) \in \mathbb{C}^p$ onto the partial autocorrelations $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_p) \in [-1, 1]^p$ such that $\mathbf{C}_i(\boldsymbol{\phi}) = \mathbf{C}_i(\boldsymbol{\pi})$.

Let $\hat{\boldsymbol{\theta}}^{(h)} = (\hat{\boldsymbol{\beta}}^{(h)}, \hat{\mathbf{D}}^{(h)}, \hat{\boldsymbol{\Sigma}}^{(h)}, \hat{\boldsymbol{\phi}}^{(h)}, \hat{\nu}^{(h)})$ be the parameter estimates in the hth iteration. The E-step for subject i is

$$Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(h)}) = E(\ell_c^{[i]}(\boldsymbol{\theta})|\mathbf{y}_i, \hat{\boldsymbol{\theta}}^{(h)}) = \int\int \ell_c^{[i]}(\boldsymbol{\theta}) f(\mathbf{b}_i, \tau_i|\mathbf{y}_i, \hat{\boldsymbol{\theta}}^{(h)}) d\mathbf{b}_i d\tau_i, \tag{4.3}$$

integrating out $\mathbf{b}_i$ and τ_i in the complete-data log-likelihood function. The next result allows the evaluation of this conditional expectation.

Theorem 3. *Under the three-level hierarchical form of* (4.1)*, the conditional distribution of* $(\mathbf{y}_i^{\mathrm{T}}, \mathbf{b}_i^{\mathrm{T}})^{\mathrm{T}}$ *given the latent variable* τ_i *is*

$$\begin{bmatrix}\mathbf{y}_i\\ \mathbf{b}_i\end{bmatrix}\Bigg|\tau_i \sim N_{(n_i+q_2)r}\left(\begin{bmatrix}\mathbf{X}_i\boldsymbol{\beta}\\ \mathbf{0}\end{bmatrix}, \tau_i^{-1}\begin{bmatrix}\boldsymbol{\Lambda}_i & \mathbf{Z}_i\mathbf{D}\\ \mathbf{D}\mathbf{Z}_i^{\mathrm{T}} & \mathbf{D}\end{bmatrix}\right).$$

(i) *The conditional distribution of* $\mathbf{b}_i$ *given* $\mathbf{y}_i$ *and* τ_i *is multivariate normal with mean vector* $E(\mathbf{b}_i|\mathbf{y}_i, \tau_i) = \mathbf{D}\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$ *and variance-covariance matrix* $\mathrm{cov}(\mathbf{b}_i|\mathbf{y}_i, \tau_i) = \tau_i^{-1}\big(\mathbf{D}^{-1} + \mathbf{Z}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\mathbf{Z}_i\big)^{-1}$.

(ii) *The conditional distribution of* τ_i *given* $\mathbf{y}_i$ *is a gamma distribution with shape parameter* $(\nu + n_i r)/2$ *and inverse scale parameter* $(\nu + \Delta_i)/2$.

The proof of Theorem 3 is outlined in the supplementary material.
Now, (4.3) can be obtained as

$$\begin{aligned}Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(h)}) = \frac{1}{2}\Big\{ &\log|\mathbf{R}_i^{-1}| + \log|\mathbf{D}^{-1}| - \mathrm{tr}\big(\mathbf{D}^{-1}\hat{\mathbf{B}}_i^{(h)}\big) - \mathrm{tr}\big(\mathbf{R}_i^{-1}\hat{\boldsymbol{\Psi}}_i^{(h)}(\boldsymbol{\beta})\big)\\ &+\nu\big(\log(\frac{\nu}{2}) + \hat{\kappa}_i^{(h)} - \hat{\tau}_i^{(h)}\big)\Big\} - \log\Gamma(\frac{\nu}{2}),\end{aligned}$$

where

$$\begin{aligned}
\hat{\tau}_i^{(h)} &= E(\tau_i|\mathbf{y}_i,\hat{\boldsymbol{\theta}}^{(h)}) = (\hat{\nu}^{(h)}+n_i r)/(\hat{\nu}^{(h)}+\hat{\Delta}_i^{(h)}),\\
\hat{\kappa}_i^{(h)} &= E(\log\tau_i|\mathbf{y}_i,\hat{\boldsymbol{\theta}}^{(h)}) = \mathcal{D}_{\mathcal{G}}\Big(\frac{\hat{\nu}^{(h)}+n_i r}{2}\Big)-\log\Big(\frac{\hat{\nu}^{(h)}+\hat{\Delta}_i^{(h)}}{2}\Big),\\
\hat{\mathbf{B}}_i^{(h)} &= E(\tau_i\mathbf{b}_i\mathbf{b}_i^{\mathrm{T}}|\mathbf{y}_i,\hat{\boldsymbol{\theta}}^{(h)}) = \hat{\tau}_i^{(h)}\hat{\mathbf{b}}_i^{(h)}\hat{\mathbf{b}}_i^{(h)\mathrm{T}}+\hat{\mathbf{V}}_{\mathbf{b}_i}^{(h)},\\
\hat{\boldsymbol{\Psi}}_i^{(h)}(\boldsymbol{\beta}) &= E(\tau_i\mathbf{e}_i\mathbf{e}_i^{\mathrm{T}}|\mathbf{y}_i,\hat{\boldsymbol{\theta}}^{(h)})\\
&= \hat{\tau}_i^{(h)}\big(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{Z}_i\hat{\mathbf{b}}_i^{(h)}\big)\big(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{Z}_i\hat{\mathbf{b}}_i^{(h)}\big)^{\mathrm{T}}+\mathbf{Z}_i\hat{\mathbf{V}}_{\mathbf{b}_i}^{(h)}\mathbf{Z}_i^{\mathrm{T}},
\end{aligned}$$

with $\hat{\boldsymbol{\Lambda}}_i^{(h)} = \mathbf{Z}_i\hat{\mathbf{D}}^{(h)}\mathbf{Z}_i^{\mathrm{T}}+\hat{\mathbf{R}}_i^{(h)}$, $\hat{\mathbf{R}}_i^{(h)} = \hat{\boldsymbol{\Sigma}}^{(h)}\otimes\mathbf{C}_i(\hat{\boldsymbol{\phi}}^{(h)})$, $\hat{\Delta}_i^{(h)} = (\mathbf{y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}}^{(h)})^{\mathrm{T}}\hat{\boldsymbol{\Lambda}}_i^{(h)^{-1}}$ $(\mathbf{y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}}^{(h)})$, $\hat{\mathbf{b}}_i^{(h)} = E(\mathbf{b}_i|\mathbf{y}_i,\hat{\boldsymbol{\theta}}^{(h)}) = \hat{\mathbf{D}}^{(h)}\mathbf{Z}_i^{\mathrm{T}}\hat{\boldsymbol{\Lambda}}_i^{(h)^{-1}}(\mathbf{y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}}^{(h)})$, and

$$\hat{\mathbf{V}}_{\mathbf{b}_i}^{(h)} = \hat{\tau}_i^{(h)}\mathrm{cov}(\mathbf{b}_i|\mathbf{y}_i,\hat{\boldsymbol{\theta}}^{(h)}) = \big(\hat{\mathbf{D}}^{(h)^{-1}}+\mathbf{Z}_i^{\mathrm{T}}\hat{\mathbf{R}}_i^{(h)^{-1}}\mathbf{Z}_i\big)^{-1}.$$

The Q-function (Dempster, Laird, and Rubin (1977)) is given by $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(h)}) = \sum_{i=1}^N Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(h)})$. To obtain analytic expressions for the CM-steps, we let $\hat{\mathbf{e}}_{ij}^{(h)} = \mathbf{y}_{ij}-\mathbf{X}_i\boldsymbol{\beta}_j-\mathbf{Z}_i\hat{\mathbf{b}}_{ij}^{(h)}$ and $\hat{\mathbf{e}}_{il}^{(h)} = \mathbf{y}_{il}-\mathbf{X}_i\boldsymbol{\beta}_l-\mathbf{Z}_i\hat{\mathbf{b}}_{il}^{(h)}$, where $\hat{\mathbf{b}}_{ij}^{(h)}$ is a $q_2\times 1$ subvector consisting of the $((j-1)q_2+1)$th to (jq_2)th entries of $\hat{\mathbf{b}}_i^{(h)}$. It follows that $\hat{\boldsymbol{\Psi}}_i^{(h)}(\boldsymbol{\beta}) = \big[\hat{\boldsymbol{\psi}}_{ijl}^{(h)}(\boldsymbol{\beta})\big]$, where $\hat{\boldsymbol{\psi}}_{ijl}^{(h)}(\boldsymbol{\beta}) = E(\tau_i\mathbf{e}_{ij}\mathbf{e}_{il}^{\mathrm{T}}|\mathbf{y}_i,\hat{\boldsymbol{\theta}}^{(h)}) = \hat{\tau}_i^{(h)}\hat{\mathbf{e}}_{ij}^{(h)}\hat{\mathbf{e}}_{il}^{(h)\mathrm{T}}+\mathbf{Z}_i\hat{\mathbf{V}}_{\mathbf{b}_{ijl}}^{(h)}\mathbf{Z}_i^{\mathrm{T}}$ is a square matrix of order n_i with $\hat{\mathbf{V}}_{\mathbf{b}_{ijl}}^{(h)}$ a $q_2\times q_2$ submatrix consisting of the $((j-1)q_2+1)$th to (jq_2)th rows and columns of $\hat{\mathbf{V}}_{\mathbf{b}_i}^{(h)}$, for $j,l=1,\ldots,r$. In summary, the CM-steps for ECM and ECME algorithms can proceed as follows.

CM-step 1 *for ECM and ECME.* Fix $\boldsymbol{\phi} = \hat{\boldsymbol{\phi}}^{(h)}$ and $\nu = \hat{\nu}^{(h)}$, and update $\hat{\boldsymbol{\beta}}^{(h)}$, $\hat{\mathbf{D}}^{(h)}$, and $\hat{\boldsymbol{\Sigma}}^{(h)}$ by maximizing $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(h)})$ to get

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^{(h+1)} &= \Big(\sum_{i=1}^N\hat{\tau}_i^{(h)}\mathbf{X}_i^{\mathrm{T}}\hat{\mathbf{R}}_i^{(h)^{-1}}\mathbf{X}_i\Big)^{-1}\sum_{i=1}^N\hat{\tau}_i^{(h)}\mathbf{X}_i^{\mathrm{T}}\hat{\mathbf{R}}_i^{(h)^{-1}}\big(\mathbf{y}_i-\mathbf{Z}_i\hat{\mathbf{b}}_i^{(h)}\big),\\
\hat{\mathbf{D}}^{(h+1)} &= N^{-1}\sum_{i=1}^N\hat{\mathbf{B}}_i^{(h)};\\
\hat{\sigma}_{jl}^{(h+1)} &= \begin{cases}
\Big(\sum_{i=1}^N n_i\Big)^{-1}\sum_{i=1}^N\mathrm{tr}\big(\mathbf{C}_i^{-1}(\hat{\boldsymbol{\phi}}^{(h)})\hat{\boldsymbol{\psi}}_{ijl}^{(h)}(\hat{\boldsymbol{\beta}}^{(h)})\big) & \text{for } j=l,\\
\Big(2\sum_{i=1}^N n_i\Big)^{-1}\sum_{i=1}^N\mathrm{tr}\big(\mathbf{C}_i^{-1}(\hat{\boldsymbol{\phi}}^{(h)})(\hat{\boldsymbol{\psi}}_{ijl}^{(h)}(\hat{\boldsymbol{\beta}}^{(h)})+\hat{\boldsymbol{\psi}}_{ilj}^{(h)}(\hat{\boldsymbol{\beta}}^{(h)}))\big) & \text{for } j\neq l,
\end{cases}
\end{aligned}$$

the estimates of the distinct elements in $\boldsymbol{\Sigma}$.

CM-step 2 *for ECM.* Fix $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{(h+1)}$, $\mathbf{D} = \hat{\mathbf{D}}^{(h+1)}$, and $\boldsymbol{\Sigma} = \hat{\boldsymbol{\Sigma}}^{(h+1)}$. Calculate $(\hat{\boldsymbol{\pi}}^{(h+1)}, \hat{\nu}^{(h+1)})$ by maximizing the constrained Q function

$$(\hat{\boldsymbol{\pi}}^{(h+1)}, \hat{\nu}^{(h+1)}) = \arg\max_{(\boldsymbol{\pi},\nu)} \Big\{ \sum_{i=1}^{N} \Big(r \log|\mathbf{C}_i^{-1}(\boldsymbol{\pi})| - \operatorname{tr}\big((\hat{\boldsymbol{\Sigma}}^{-1^{(h+1)}} \otimes \mathbf{C}_i^{-1}(\boldsymbol{\pi})) \times \hat{\boldsymbol{\Psi}}_i^{(h+1/2)}\big) + \nu \log(\frac{\nu}{2}) - 2\log\Gamma(\frac{\nu}{2}) + \nu\big(\hat{\kappa}_i^{(h)} - \hat{\tau}_i^{(h)}\big)\Big)\Big\},$$

where $\hat{\boldsymbol{\Psi}}_i^{(h+1/2)} = \hat{\boldsymbol{\Psi}}_i^{(h)}(\hat{\boldsymbol{\beta}}^{(h+1)})$.

CM-step 2 *for ECME.* Given current estimates, calculate $(\hat{\boldsymbol{\pi}}^{(h+1)}, \hat{\nu}^{(h+1)})$ by maximizing the constrained log-likelihood function

$$(\hat{\boldsymbol{\pi}}^{(h+1)}, \hat{\nu}^{(h+1)}) = \arg\max_{(\boldsymbol{\pi},\nu)} \Big\{ \sum_{i=1}^{N} \Big(\log\Gamma\Big(\frac{\nu + n_i r}{2}\Big) - \log\Gamma\Big(\frac{\nu}{2}\Big) - \frac{n_i r}{2}\log(\nu) - \frac{1}{2}\log\big|\hat{\boldsymbol{\Lambda}}_i^{(h+1)}(\boldsymbol{\pi})\big| - \frac{\nu + n_i r}{2}\log\Big(1 + \frac{\hat{\Delta}_i^{(h+1)}(\boldsymbol{\pi})}{\nu}\Big)\Big)\Big\}.$$

In CM-step 2, we use the *nlminb* routine in R to do a $(p+1)$-dimensional search of $(\hat{\boldsymbol{\pi}}^{(h+1)^{\mathrm{T}}}, \hat{\nu}^{(h+1)})$ subject to box constraints: $\boldsymbol{\pi} \in [-1,1]^p$, $\nu \in (0,\infty)$. Then update $\hat{\boldsymbol{\phi}}^{(h)}$ by inverting $\hat{\boldsymbol{\pi}}^{(h+1)}$ back to $\hat{\boldsymbol{\phi}}^{(h+1)}$ according to (4.2).

In some situations, the EM-type algorithms may be too slow to be of any practical use. To speed up the convergence, a hybrid ECME-scoring procedure that starts running a moderate number of ECME iterations, followed by the scoring method, is recommended. To perform the scoring procedure, first update $\hat{\boldsymbol{\beta}}^{(h)}$ by the generalized least squares step

$$\hat{\boldsymbol{\beta}}^{(h+1)} = \Big(\sum_{i=1}^{N} \hat{\tau}_i^{(h)} \mathbf{X}_i^{\mathrm{T}} \hat{\boldsymbol{\Lambda}}_i^{(h)^{-1}} \mathbf{X}_i\Big)^{-1} \sum_{i=1}^{N} \hat{\tau}_i^{(h)} \mathbf{X}_i^{\mathrm{T}} \hat{\boldsymbol{\Lambda}}_i^{(h)^{-1}} \mathbf{y}_i. \tag{4.4}$$

Given the current estimates $\hat{\boldsymbol{\beta}}^{(h+1)}$ and $\hat{\boldsymbol{\alpha}}^{(h)}$, one iteration of the scoring procedure for updating $\hat{\boldsymbol{\alpha}}^{(h)}$ gives

$$\hat{\boldsymbol{\alpha}}^{(h+1)} = \hat{\boldsymbol{\alpha}}^{(h)} + \hat{\mathbf{J}}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{(h)^{-1}} \hat{\mathbf{s}}_{\boldsymbol{\alpha}}^{(h+1/2)}, \tag{4.5}$$

where $\hat{\mathbf{s}}_{\boldsymbol{\alpha}}^{(h+1/2)}$ and $\hat{\mathbf{J}}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{(h)}$ are $\mathbf{s}_{\boldsymbol{\alpha}} = (\mathbf{s}_{\boldsymbol{\omega}}^{\mathrm{T}}, s_\nu)^{\mathrm{T}}$ and $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}$ evaluated at $\hat{\boldsymbol{\beta}}^{(h+1)}$ and $\hat{\boldsymbol{\alpha}}^{(h)}$.

As recommended in Chi and Reinsel (1989), when the increment $\hat{\mathbf{J}}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{(h)^{-1}} \hat{\mathbf{s}}_{\boldsymbol{\alpha}}^{(h+1/2)}$ is too large, (4.5) can be modified as $\hat{\boldsymbol{\alpha}}^{(h+1)} = \hat{\boldsymbol{\alpha}}^{(h)} + \hat{\varrho}^{(h)} \hat{\mathbf{J}}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{(h)^{-1}} \hat{\mathbf{s}}_{\boldsymbol{\alpha}}^{(h+1/2)}$, where $\hat{\varrho}^{(h)}$ is such that minus twice the log-likelihood at the $(h+1)$th iteration is less

than the one at the hth iteration, say $-2\hat{\ell}^{(h+1)} \leq -2\hat{\ell}^{(h)}$. Given the initial value $\hat{\boldsymbol{\theta}}^{(0)}$, the ML estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}$ can be obtained by iterating the ECME steps and then (4.4) and (4.5) until an increase in the log-likelihood is less than a pre-specified tolerance. Upon convergence, the asymptotic variance-covariance matrix of estimates can be approximated by $\text{cov}(\hat{\boldsymbol{\beta}}) = \hat{\mathbf{J}}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ and $\text{cov}(\hat{\boldsymbol{\alpha}}) = \hat{\mathbf{J}}^{-1}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}$, respectively. In general, the iteration-based procedure converges to local or global modes. A convenient way to circumvent such a limitation is to try a variety of initial values that are representatives of the parameter space, and compare their relative log-likelihood values when employing the algorithm.

4.2. Model choice

A variety of information theoretic criteria exist to properly determine the best choice among competing models. To best identify a model supported by the data, we adopt the Akaike information criterion (AIC; Akaike (1973)) and the Bayesian information criterion (BIC; Schwarz (1978)). For a model with parameters $\boldsymbol{\theta}$, they are defined as

$$\text{AIC} = 2m - 2\ell_{\max} \quad \text{and} \quad \text{BIC} = m \log N - 2\ell_{\max},$$

where $\ell_{\max}$ is the maximized log-likelihood and m is the number of free parameters in the model. There is no consensus regarding which criterion is better, and combined use of AIC and BIC could be of help in screening reasonable candidate models.

A formal test concerning the appropriateness of using the normal model $H_0 : \nu^{-1} = 0$ versus t model $H_1 : \nu^{-1} > 0$ is nontrivial since the null hypothesis is on the boundary of the parameter space. For testing parameters under non-standard settings, Self and Liang (1987) have shown the limiting distribution of the LRT statistic follows a mixture of chi-square distributions. Referring to Case 5 of Self and Liang (1987), the LRT statistic under $H_0 : \nu^{-1} = 0$ is an equally weighted mixture of χ^2_0 and χ^2_1 distributions, where χ^2_0 denotes a degenerate distribution with all of its mass at zero. In this case, the critical values are 2.71 and 5.41 at the 5% and 1% significance levels, respectively.

5. Analytical Results

In this section, we describe how the proposed MtLMM can be used to improve the fitting performance on the set of ACTG 175 data pre-analyzed in Section 2. Several graphical diagnostic tools for investigating the underlying model assumption and detection of outliers, together with the robustness of MtLMM, are presented.

Table 1. Summary of model selection criteria.

Criterion	Model	RI				RIS			
		UNC	AR(1)	AR(2)	AR(3)	UNC	AR(1)	AR(2)	AR(3)
$\ell_{\max}$	N	383.396	470.794	478.266	480.003	470.338	496.333	498.052	499.177
	T	413.474	484.981	491.353	493.215	494.857	511.999	512.431	513.407
AIC	N	-738.792	-911.589	-924.532	-926.006	-898.675	-948.666	-950.104	-950.354
	T	-796.949	-937.962	-948.705	-950.430	-945.714	-977.998	-976.861	-976.814
BIC	N	-719.175	-890.571	-902.113	-902.185	-869.250	-917.839	-917.877	-916.725
	T	-775.930	-915.543	-924.885	-925.208	-914.886	-945.769	-943.233	-941.785

N: MLMM; T: MtLMM

Table 2. ML estimation for the fitted MtLMM with RIS and AR(1).

Fixed effects	β_{10}	β_{11}	β_{12}	β_{13}	β_{20}	β_{21}	β_{22}	β_{23}		
Est	2.5629	-0.1017	-0.0119	-0.0321	2.8800	0.0975	-0.0076	-0.0114		
Sd	0.0304	0.0526	0.0055	0.0096	0.0387	0.0670	0.0041	0.0071		
Random effects	d_{11}	d_{21}	d_{22}	d_{31}	d_{32}	d_{33}	d_{41}	d_{42}	d_{43}	d_{44}
Est	0.0107	0.0015	0.0004	-0.0001	-0.0003	0.0195	0.0003	0.0001	0.0000	0.0002
Sd	0.0049	0.0006	0.0002	0.0044	0.0008	0.0079	0.0005	0.0001	0.0006	0.0001
Within-subject errors	σ_{11}	σ_{21}	σ_{22}	ϕ_1	ν					
Est	0.0071	0.0055	0.0091	0.3838	7.3395					
Sd	0.0010	0.0009	0.0013	0.0671	2.4010					

d_{ls}, for $l, s = 1, \ldots, q_2 r$ are the distinct elements of $\mathbf{D}$.

5.1. Model fitting

We apply our methods to the ACTG 175 data. In the MtLMM setting, the design matrix for the fixed effects is

$$\mathbf{X}_i = \mathbf{I}_2 \otimes \big[\mathbf{1}_{10} : \text{treat}_i \mathbf{1}_{10} : \mathbf{t}_i : \text{treat}_i \mathbf{t}_i\big],$$

where $\mathbf{I}_2$ is a 2×2 identity matrix, and $\mathbf{1}_{10}$, treat_i, and $\mathbf{t}_i$ are defined in Section 2. We compare results obtained by the random intercept (RI) model and the random intercept plus slope (RIS) model. Specifically, the design matrix for RI and RIS are $\mathbf{Z}_i = \mathbf{I}_2 \otimes \mathbf{1}_{10}$ and $\mathbf{Z}_i = \mathbf{I}_2 \otimes [\mathbf{1}_{10} : \mathbf{t}_i]$, respectively.

To explore the presence of autocorrelation among within-subject errors, we employ the score test procedure with the uncorrelated (UNC) covariance structure, say $\mathbf{C}_i = \mathbf{I}_{10}$. The values of λ_s for RI and RIS scenarios are 158.40 and 40.36, respectively. The results, being highly significant, signify that the autocorrelation cannot be completely explained by the inclusion of random slopes. Consequently, three selected AR(p) ($p = 1, 2, 3$) structures for the within-subject errors are considered in the later analysis. For the sake of model comparison, we also fit the MLMM counterparts, which can be treated as the reduced MtLMMs as ν tends to infinity.

Table 1 reports the values of the maximized log-likelihood $\ell_{\max}$ together with the associated AIC and BIC values. To fit the models via the ML method,

we employed the hybrid ECME-scoring algorithm developed in Section 4 under different starting values. The algorithm was terminated when an increase in the log-likelihood was smaller than 10^{-6}. When comparing fitted objects, the model with the smallest AIC or BIC value was taken to be the best one. In light of these two criteria, the results show that the t models are all superior to their normal counterparts. The best choice is the MtLMM with RIS and a parsimonious AR(1) structure (although AR(2) and AR(3) are somewhat comparable); the resulting ML estimates of parameters along with their standard errors are listed in Table 2.

Consider the estimates of parameters in the table. For the fixed effects, the significance of the intercepts (β_{10}, β_{20}) indicates that CD4 and CD8 are different from zero at baseline, and the difference of the intercepts between the zidovudine alone group and the other three therapies (β_{11}, β_{21}) being insignificant indicates that the two groups have no statistically significant difference in CD4 and CD8 at baseline. Meanwhile, the negative slopes (β_{12}, β_{22}) reveal that both CD4 and CD8 significantly fall across time, and the negative differences of time slopes between the two groups (β_{13}, β_{23}) suggest there are significant increases in the rates of progression to primary endpoint, say the occurrence of AIDS or death, for the zidovudine alone group. Although only slight evidence of significance is presented for the estimates of slopes in CD8, this is consistent with the result of Hammer et al. (1996) in which treatments, except for the group of zidovudine alone, offer advantages to slow the progression of AIDS symptoms.

For the estimates of σ_{jl}'s, they are apparently significant being far larger than twice the associated standard errors. To investigate the existence of the relationship between two response variables, the estimate of the correlation $\rho_{21} = \sigma_{21}/\sqrt{\sigma_{11}\sigma_{22}}$ for the "best" model is 0.684. As a result, there is strong evidence that the CD4 and CD8 cell counts are positively related. Furthermore, the estimate of the AR parameter ϕ_1 is highly significant, supported also by the score test statistic. It is noteworthy that the estimate of degrees of freedom is somewhat small ($\hat{\nu} = 7.34$), and the observed LRT statistic for testing MLMM-RIS-AR(1) versus MtLMM-RIS-AR(1) is 31.33. Here the LRT statistic follows an equally weighted mixture of χ^2_0 and χ^2_1 distributions. The resulting p-value 1.087×10^{-8} guarantees the appropriateness of the use of multivariate t distribution.

5.2. Diagnostics

We consider diagnostics to assess the validity of the underlying distributional assumptions on error terms. To examine the adequacy of the fitted models, a formal measure for judging the distributional assumptions of residuals is to use the Mahalanobis distance $\hat{\Delta}_i$, which has an asymptotic χ^2_{df} under MLMM and a scale F distribution $df\mathcal{F}(df, \hat{\nu})$ under MtLMM, with $df = 20$. Here $\hat{\Delta}_i$ is Δ_i

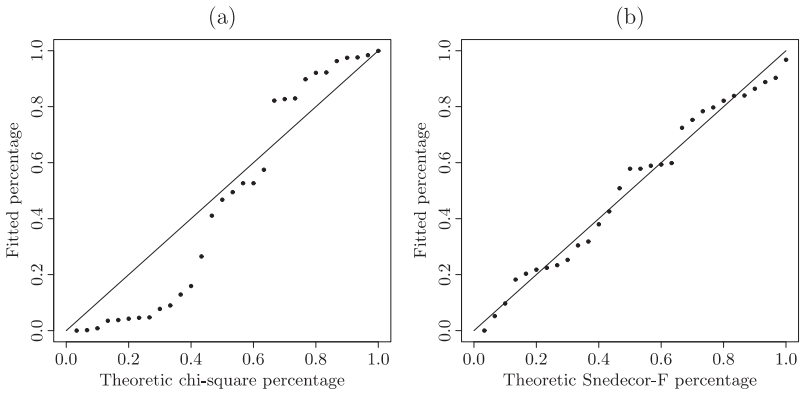

Figure 3. Healy's plot when (a) MLMM with RIS–AR(3) or (b) MtLMM with RIS–AR(1) is fitted to the 30 selected AIDS patients.

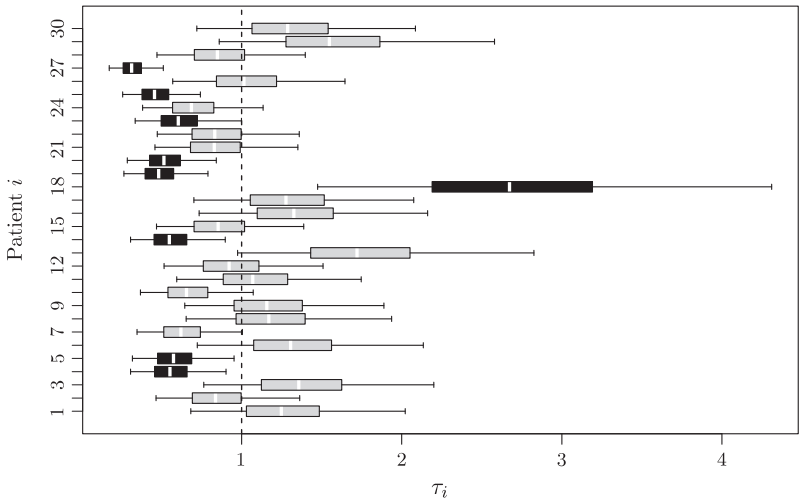

Figure 4. Empirical conditional distributions of τ_i for the 30 selected AIDS patients. The 2.5%, 25%, 50%, 75% and 97.5% quantiles of the "empirical" samples are drawn on the boxplots.

in (3.2) with parameters replaced by their ML estimates. Then, the Healy-type (1968) plots are constructed by plotting the cumulative probabilities of χ^2_{df} and $df\mathcal{F}(df,\nu)$ associated with the ordered values of $\hat{\Delta}_i$ against their nominal values $(i-0.5)/N$ for $i=1,\ldots,N$. One can examine whether the corresponding Healy's plot resembles a straight line through the origin having unit slope; a poor fitting of the model is suggested if the scatter curve has a serious departure from the 45-degree line. Figure 3 displays the Healy-type plots for the choices of MLMM and MtLMM selected by the smallest AIC. The patterns are analogous to the choices selected by the smallest BIC. A comparison of the two plots indicates the t model

tracks the identity line more closely than does the normal counterpart, revealing a substantial improvement provided by the use of multivariate t distribution.

Furthermore, it is of interest to detect outliers in the multi-outcome longitudinal data. As pointed out by Wakefield et al. (1994), the performance of the τ_i can be used as sensitive indicators for outlying observations against the prior expectation of one. Thus, if the value of τ_i is substantially lower or higher than one, then the ith patient might be regarded as an outlier in the population. By Theorem 3 (ii), we draw the "empirical" samples of τ_i from the gamma distribution with shape $(\hat{\nu} + df)/2$ and inverse scale $(\hat{\nu} + \hat{\Delta}_i)/2$, where $\hat{\nu}$ and $\hat{\Delta}_i$ are obtained from the best fitted MtLMM. Figure 4 indicates patients 4, 5, 14, 18, 19, 20, 23, 25 and 27 could be treated as potential outliers, since none of those 95% confidence intervals cover one. We also mark these identity numbers in Figure 1.

5.3. Robustness

We now illustrate the effect of robustness of MtLMM with respect to MLMM after some perturbed values are introduced to the original data. Such procedure can be done by (i) adding a contaminated vector $\boldsymbol{\xi} = (\xi_1, \xi_2)^{\mathrm{T}}$ to one observation vector, denoted by $\mathbf{y}_{i\cdot t}(\boldsymbol{\xi}) = \mathbf{y}_{i\cdot t} + \boldsymbol{\xi}$, where $\mathbf{y}_{i\cdot t} = (y_{i1t}, y_{i2t})^{\mathrm{T}}$; (ii) re-estimating the parameters twice under the RIS-AR(1) assumption of MLMM and MtLMM; (iii) computing the relative change in the estimates, i.e., $\hat{\boldsymbol{\theta}}(\boldsymbol{\xi})/\hat{\boldsymbol{\theta}} - 1$. Here $\hat{\boldsymbol{\theta}}(\boldsymbol{\xi})$ denotes the estimate for the contaminated data and $\hat{\boldsymbol{\theta}}$ is the original estimate. In this study, we vary each element of $\boldsymbol{\xi}$ between -2.5 to 2.5 units by increments of 0.25 unit and add it to the last time point ($t = 128$ week) of the first patient; this patient has $\mathrm{treat}_i = 0$ and responses crossing around the middle profile. Notice that the identity number of patient 1 is marked in Figure 1.

We concentrate on the ML estimates of fixed effects $\boldsymbol{\beta}$ and autocorrelation coefficient ϕ_1, since the other estimates for $\mathbf{D}$ and $\boldsymbol{\Sigma}$ themselves have different interpretations under different model assumptions. Instead, with the marginal correlation (Fieuws and Verbeke (2004)) that depends on the estimates of $\mathbf{D}$ and $\boldsymbol{\Sigma}$ is a function of time t, given by

$$r_m(t) = \frac{d_{31} + td_{41} + td_{32} + t^2 d_{42} + \sigma_{21}}{\sqrt{d_{11} + 2td_{21} + t^2 d_{22} + \sigma_{11}}\sqrt{d_{33} + 2td_{43} + t^2 d_{44} + \sigma_{22}}},$$

can be used to explore the question how the evolution of CD4 cell counts is associated with the evolution of CD8 cell counts. Thus, we are also interested in observing the change of correlation $r_m(t)$ between both evolutions due to outliers.

Figure 5 exhibits the curves of the percentage changes for the estimates of $\boldsymbol{\beta}$ and ϕ_1, together with their estimated standard errors, for various contaminants $\boldsymbol{\xi}$. Figure 6 presents the originally estimated marginal correlation compared with the marginal correlations implied by the contaminated data. As seen in these

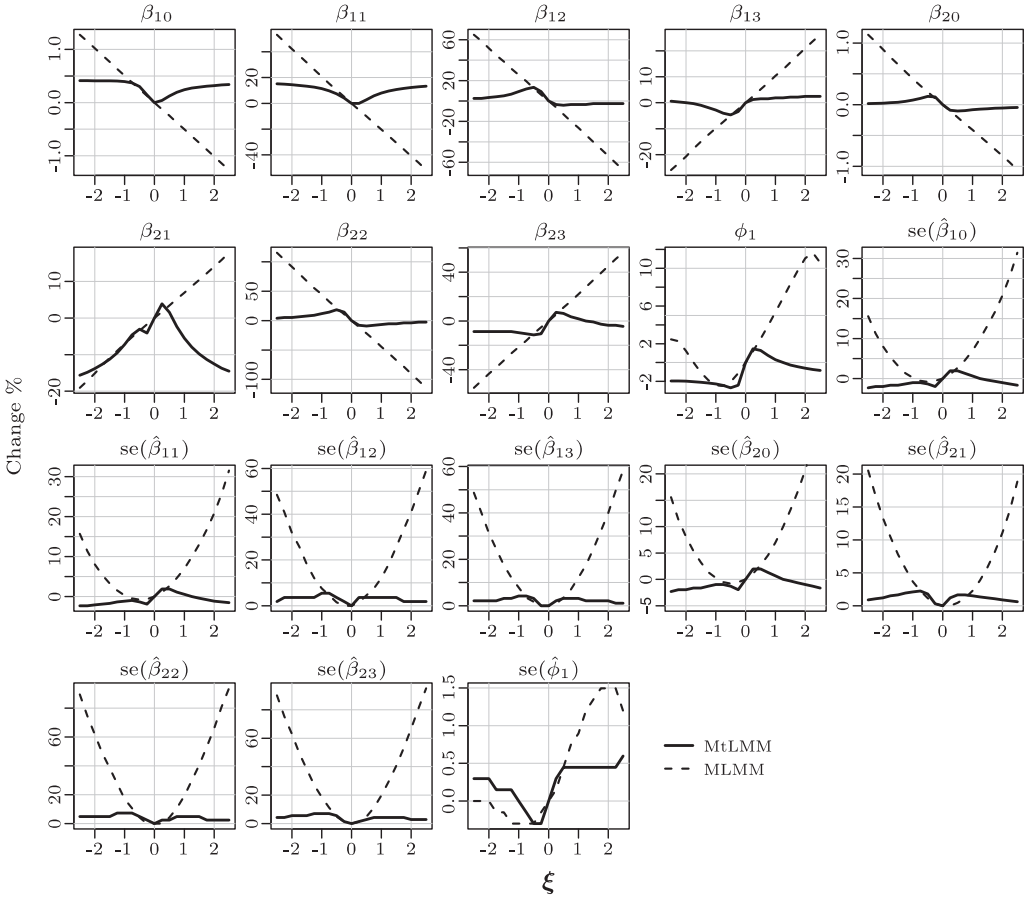

Figure 5. Percentage relative changes in the ML estimates under the MtLMM and MLMM (RIS–AR(1)) with varying contamination $\boldsymbol{\xi}$ in a pair of observations.

figures, the influence on parameter estimation of the single outlier is unbounded in the case of the MLMM, whereas it is obviously bounded in the MtLMM. More specifically, outliers in the MLMM effect changes of up to $\pm 115\%$ in the estimates of model parameters and from -1% to 95% in their standard errors. Such changes in normal fits can have an enormous impact on the correctness of statistical inferences. While in the MtLMM, effects are limited to $\pm 18\%$ on estimates and between -2% and 7% on their standard errors. In addition, with the interference of outliers, the tendency of marginal correlation under MtLMM is to maintain similar scale, while under MLMM it tends to be more volatile. This suggests that MtLMM, which downweights the influence of outliers and heavy-tailed noise, provides an appropriate way for achieving robust inference.

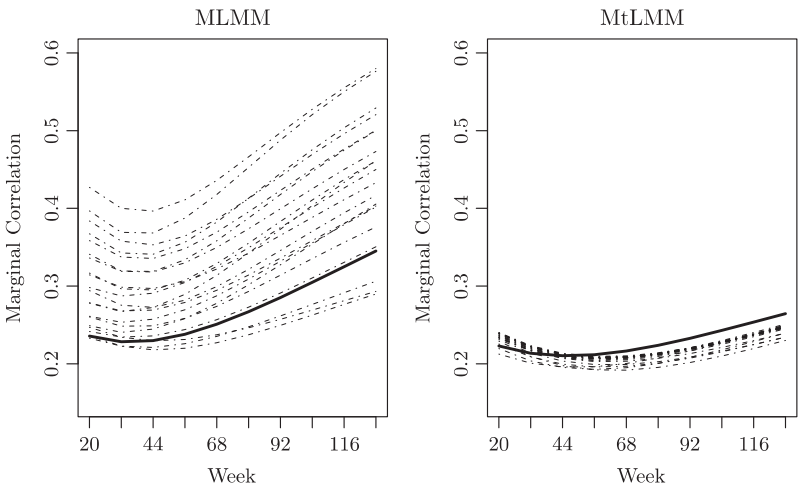

Figure 6. Comparison of marginal correlations derived from the fitted MLMM and MtLMM for the original data (solid lines) and the data with varying contamination $\boldsymbol{\xi}$ in a pair of observations (dot-dashed lines).

6. Conclusion

We have proposed a new robust approach to MLMM based on the multivariate t distribution, called the MtLMM. It offers a great deal of flexibility in dealing with multi-outcome longitudinal data in the presence of outliers or heavy-tailed noise. The proposed methodologies allow the practitioner to fit longitudinal data in a broad variety of situations. We have presented a three-level hierarchical representation for MtLMM and an efficient hybrid ECME-scoring algorithm for carrying out ML estimation. When the random effects are of high dimension, the computational expense can be prohibitively large, and some parsimonious restrictions for the covariance structure of random effects are needed. With our approach, the computational burden associated with low-dimensional random effects for each outcome, say $q_2 \leq 3$, is not heavy and reliable inferences for parameter estimates are obtainable.

Numerical results reveal that the fitted MtLMMs substantially outperforms the MLMM counterparts on the basis of likelihood-based model selection criteria. The graphical outputs provide both easily understood inferential summaries and informative diagnostic aids for checking model assumption and detecting outliers. Simulation results show that the influence of any outlying observations on the parameter estimates has a limited range for the MtLMM, while misleading estimates obtained from the MLMM can have a serious impact on statistical inferences.

Recently, Vaida and Blanchard (2005) proposed the conditional AIC (cAIC) and marginal AIC (mAIC) for model selection when the focus lies in the choice

of random effects. They concluded that the use of cAIC and mAIC might lead to different model specifications. Liang, Wu, and Zou (2008) relaxed Vaida and Blanchard's assumptions to provide a corrected version of cAIC that accounts for the uncertainty of the unknown covariance matrix of the random effects. Greven and Kneib (2008) commented that the calculation of the corrected cAIC carries high computational cost in carrying out numerical approximations, and is prohibitive in settings with large sample sizes and numbers of candidate models. Nevertheless, the cAIC approaches and their modifications are potentially useful for the choice of random effects in univariate LMMs. To our knowledge, no such modification has been developed concerning the choice of the serial correlation structure. Furthermore, extensions to a more general context, such as our proposed MtLMMs (or MLMMs), are not straightforward, but is an interesting topic.

Some possibilities for future research along these lines are as follows. The methodology can be extended to consider more general parametric distributions, such as the normal/independent distribution (Lange and Sinsheimer (1993)) discussed in Rosa, Padovani, and Gianola (2003); Rosa, Gianola, and Padovani (2004), the skew-elliptical distribution (Sahu, Dey, and Branco (2003)) utilized in Jara, Quintana and Martín (2008), the skew normal and skew t distributions exploited in Lin and Lee (2008) and Ho and Lin (2010), and the skew normal/independent distribution studied in Lachos, Ghosh, and Arellano-Vallec (2010). It is of interest to compare, both theoretically and empirically, the robustness of statistical inference among these competing models. Sometimes, ML estimation for the t-based model may be problematic because the likelihood function is ill-behaved when the degrees of freedom approaches zero or infinity. With the fast development of computational techniques, it is a worthwhile task to develop a fully Bayesian approach to inferring MtLMM via the Markov chain Monte Carlo method.

Acknowledgements

We gratefully acknowledge an associate editor and an anonymous referee for their valuable comments and insightful suggestions that substantially improved the quality of the paper.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In 2*nd Int. Symp. on Information Theory* (Edited by B. N. Petrov and F. Csaki), 267-281. Akademiai Kiado, Budapest.

Barndorff-Nielsen, O. and 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. and Reinsel, G. C. (1994). *Time Series Analysis Forecasting and Control.* 3rd edition. Holden-Day.

Chi, E. M. and Reinsel, G. C. (1989). Models for longitudinal data with random effects and AR(1) errors. *J. Amer. Statist. Assoc.* **84**, 452-459.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B* **39**, 1-38.

Diggle, P. J., Heagerty, P., Liang, K. Y. and Zeger, S. L. (2002). *Analysis of Longitudinal Data.* 2nd edition. Clarendon Press, New York.

Fieuws, S. and Verbeke, G. (2004). Joint modelling of multivariate longitudinal profiles: pitfalls of the random-effects approach. *Statist. Med.* **23**, 3093-3104.

Fitzmaurice, G. M., Laird, N. M. and Ware, J. (2004). *Applied Longitudinal Analysis.* Wiley, New York.

Greven, S. and Kneib, T. (2008). On the behavior of marginal and conditional Akaike information criteria in linear mixed models. Technical Report **179**, Dept. of Biostatistics, Johns Hopkins University.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M., Phair, J. P., Niu, M., Hirsch, M. S. and Merigan, T. C. (1996). A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *The New England J. Medicine* **335**, 1081-1090.

Healy, M. J. R. (1968). Multivariate normal plotting. *Appl. Statist.* **17**, 157-161.

Hedeker, D. and Gibbons, R. D. (2006). *Longitudinal Data Analysis.* Wiley, Hoboken, New Jersey.

Ho, H. J. and Lin, T. I. (2010). Robust linear mixed models using the skew t distribution with application to schizophrenia data. *Biom. J.* **52**, 449-469.

Hogan, J. W. and Laird, N. M. (1997). Model-based approaches to analysing incomplete longitudinal and failure time data. *Statist. Med.* **16**, 259-272.

Jara, A., Quintana, F. and Martín, E. S. (2008). Linear mixed models with skew-elliptical distributions: A Bayesian approach. *Comput. Statist. Data Anal.* **52**, 5033-5045.

Kotz, S. and Nadarajah, S. (2004). *Multivariate t Distributions and Their Applications.* Cambridge University Press.

Lachos, V. H., Ghosh, P. and Arellano-Vallec, R. B. (2010). Likelihood based inference for skew-normal independent linear mixed models. *Statist. Sinica* **20**. 303-322.

Laird, N. M. and Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics* **38**, 963-974.

Lange, K. and Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *J. Comput. Graph. Statist.* **2**, 175-198.

Lange, K. L., Little, R. J. A. and Taylor, J. M. G. (1989). Robust statistical modeling using the t distribution. *J. Amer. Statist. Assoc.* **84**, 881-896.

Liang, H., Wu, H. and Zou, G. (2008). A note on conditional AIC for linear mixed-effects models. *Biometrika* **95**, 773-778.

Lin, T. I. (2008). Longitudinal data analysis using t linear mixed models with autoregressive dependence structures. *J. Data Sci.* **6**, 333-355.

Lin, T. I. and Lee, J. C. (2006). A robust approach to t linear mixed models applied to multiple sclerosis data. *Statist. Med.* **25**, 1397-1412.

Lin, T. I. and Lee, J. C. (2007). Bayesian analysis of hierarchical linear mixed modeling using the multivariate t distribution. *J. Statist. Plann. Inference* **137**, 484-495.

Lin, T. I. and Lee, J. C. (2008). Estimation and prediction in linear mixed models with skew normal random effects for longitudinal data. *Statist. Med.* **27**, 1490-1507.

Liu, C. and Rubin, D. B. (1994). The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. *Biometrika* **81**, 633-648.

Meng, X. L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* **80**, 267-278.

Nadarajah, S. and Kotz, S. (2005). Mathematical properties of the multivariate t distribution. *Acta Applicandae Mathematicae* **89**, 53-84.

Nadarajah, S. and Kotz, S. (2008). Estimation methods for the multivariate t distribution. *Acta Applicandae Mathematicae* **102**, 99-118.

Pinheiro, J. C., Liu, C. H. and Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects model using the multivariate t distribution. *J. Comput. Graph. Statist.* **10**, 249-276.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna.

Rosa, G. J. M., Padovani, C. R. and Gianola, D. (2003). Robust linear mixed models with normal/independent distributions and Bayesian MCMC implementation. *Biom. J.* **45**, 573-590.

Rosa, G. J. M., Gianola, D. and Padovani, C. R. (2004). Bayesian longitudinal data analysis with mixed models and thick-tailed distributions using MCMC. *J. Appl. Stat.* **31**, 855-873.

Roy, A. (2006). Estimating correlation coefficient between two variables with repeated observations using mixed effects model. *Biom. J.* **48**, 286-301.

Sahu, S. K., Dey, D. K. and Branco, M. D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *Can. J. Stat.* **31**, 129-150

SAS (2001). *SAS/Stat User's Guide, Version 8.2.* SAS Institute Inc., Cary, NC.

Schafer, J. L. and Yucel, R. M. (2002). Computational strategies for multivariate linear mixed-effects models with missing values. *J. Comput. Graph. Statist.* **11**, 437-457.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6**, 461-464.

Self, S. G. and Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Statist. Assoc.* **82**, 605-610.

Shah, A., Laird, N. and Schoenfeld, D. (1997). A random-effects model for multiple characteristics with possibly missing data. *J. Amer. Statist. Assoc.* **92**, 775-779.

Shao, J. (2003). *Mathematical Statistics.* 2nd edition. Springer-Verlag, New York.

Song, P. X. K., Zhang, P. and Qu, A. (2007). Maximum likelihood inference in robust linear mixed-effects models using multivariate t distributions. *Statist. Sinica* **17**, 929-943.

Song, X., Davidian, M. and Tsiatis, A. A. (2002). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* **58**, 742-753.

Vaida, F. and Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika* **92**, 351-370.

Wakefield, J. C., Smith, A. F. M., Racine-Poon, A. and Gelfand, A. E. (1994). Bayesian analysis of linear and non-linear model by using Gibbs sampler. *Appl. Statist.* **43**, 201-221.

Wang, W. L. and Fan, T. H. (2010). ECM-based maximum likelihood inference for multivariate linear mixed models with autoregressive errors. *Comput. Statist. Data Anal.* **54**, 1328-1341.

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.

Department of Statistics, Feng Chia University, Taichung 40724, Taiwan.

E-mail: wlunwang@fcu.edu.tw

Graduate Institute of Statistics, National Central University, Jhongli 32001, Taiwan.

E-mail: thfan@stat.ncu.edu.tw

(Received November 2009; accepted August 2010)