

Longitudinal Data Analysis Using t Linear Mixed Models with Autoregressive Dependence Structures

Tsung-I Lin

National Chung Hsing University

Abstract: The t linear mixed model with AR(p) dependence structure is proposed for the analysis of longitudinal data in which the underlying repeated measures contain thick tails and serial correlations simultaneously. For parameter estimation, I develop a hybrid maximization scheme that combines the stability of the Expectation Conditional Maximization Either (ECME) algorithm with the rapid convergence property of the scoring method. Empirical Bayes estimation of random effects and prediction of future values for the proposed model are also considered. The proposed methodologies are applied to a real example from a tumor growth study on twenty-two mice. Numerical comparisons indicate that the proposed model outperforms the normal model from both inferential and predictive perspectives.

Key words: Conditional prediction, ECME algorithm, maximum likelihood estimation, outliers, random effects, reparameterization.

1. Introduction and Historical Perspective

In most theoretical research methods of linear mixed models as well as their applications, the error terms are routinely assumed to be normally distributed for mathematical convenience. However, such a normality assumption could be violated and in turn may affect the estimates of regression coefficients and variance components when the experimental data involve thicker than normal tails or atypical observations. Over the past three decades, the multivariate t distribution has been recognized as a useful generalization of the normal distribution for robustifying linear regression models (Zellner, 1976; Lange *et al.*, 1989) and linear mixed models (Pinheiroit *et al.*, 2001; Lin and Lee, 2006; Lin and Lee, 2007).

A n -dimensional random vector \mathbf{Y} is said to follow a multivariate t distribution with location vector $\boldsymbol{\mu}$, scaling covariance matrix $\boldsymbol{\Sigma}$ and degrees-of-freedom ν , denoted by $\mathbf{Y} \sim t_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, if its density function is given by

$$f(\mathbf{Y}) = \frac{\Gamma((\nu + n)/2)}{\Gamma(\nu/2)(\pi\nu)^{n/2}} |\boldsymbol{\Sigma}|^{-1/2} \left(1 + \frac{(\mathbf{Y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu})}{\nu} \right)^{-(\nu+n)/2}.$$

For a comprehensive overview of the fundamental theories and characterizations of the multivariate t distribution along with its recent advances and applications, the interested reader can refer to the monograph by Kotz and Nadarajah (2004).

The t linear mixed model can be written as

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad (1.1)$$

along with the assumption of

$$\begin{bmatrix} \mathbf{b}_i \\ \boldsymbol{\varepsilon}_i \end{bmatrix} \sim t_{n_i+m_2} \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \sigma^2 \begin{bmatrix} \boldsymbol{\Gamma} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_i \end{bmatrix}, \nu \right), \quad (1.2)$$

where \mathbf{Y}_i is an n_i -dimensional vector made on subject i ($i = 1, \dots, N$), \mathbf{X}_i and \mathbf{Z}_i are known full-rank covariate matrices of dimensions $n_i \times m_1$ and $n_i \times m_2$, respectively, $\boldsymbol{\beta}$ is an $m_1 \times 1$ vector of fixed effects used to describe the population mean, \mathbf{b}_i is an $m_2 \times 1$ vector of unobservable random effects, $\boldsymbol{\varepsilon}_i$ is an $n_i \times 1$ vector of errors.

As pointed out by Pinheiro *et al.* (2001), it is crucial to emphasize that \mathbf{b}_i and $\boldsymbol{\varepsilon}_i$ are uncorrelated but dependent. From (1.2), there exists a latent variable $\tau_i \sim \text{Gamma}(\nu/2, \nu/2)$, where $\text{Gamma}(\alpha, \beta)$ stands for the gamma distribution with mean α/β , such that $\mathbf{b}_i | \tau_i$ and $\boldsymbol{\varepsilon}_i | \tau_i$ are independent. Moreover, $\boldsymbol{\Gamma}$ is an $m_2 \times m_2$ unstructured positive definite matrix. Note that the scaling covariance matrices of \mathbf{b}_i and $\boldsymbol{\varepsilon}_i$ share with the same scaling factor σ^2 for computational convenience (Lindstrom and Bates, 1988). The structured AR(p) dependence matrix, $\mathbf{C}_i = \mathbf{C}_i(\boldsymbol{\phi}) = [\rho_{|r-s|}(\boldsymbol{\phi})]$, $r, s = 1, \dots, n_i$, is considered for within-subject errors, where ρ_k 's are implicit functions of the autoregressive parameters $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)$ and satisfy the Yule-Walker equation (Box *et al.*, 1994), i.e.,

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

For the pure AR model, the admissible values of $\boldsymbol{\phi}$ are restricted in a p -dimensional hypercube \mathbb{C}_p . To enforce uniqueness of the model, the roots of $1 - \phi_1B - \phi_2B^2 - \dots - \phi_pB^p = 0$ must lie outside the unit circle, say $\boldsymbol{\phi} \in \mathbb{C}_p$.

It follows from the essential property of the multivariate t distribution that if $\mathbf{Y}_i \sim t_{n_i}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, it can be hierarchically expressed as $\mathbf{Y}_i | \tau_i \sim N_{n_i}(\boldsymbol{\mu}, \boldsymbol{\Sigma}/\tau_i)$ and $\tau_i \sim \text{Gamma}(\nu/2, \nu/2)$. Under the complete data framework, the random effects \mathbf{b}_i 's and the unknown precision scales τ_i 's can be viewed as latent variables. A three-level specification of model (1.1) is represented by

$$\begin{aligned} \mathbf{Y}_i | \mathbf{b}_i, \tau_i &\sim N_{n_i} \left(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \frac{\sigma^2}{\tau_i} \mathbf{C}_i(\boldsymbol{\phi}) \right), \\ \mathbf{b}_i | \tau_i &\sim N_{m_2} \left(\mathbf{0}, \frac{\sigma^2}{\tau_i} \boldsymbol{\Gamma} \right), \quad \tau_i \sim \text{Gamma}(\nu/2, \nu/2). \end{aligned} \quad (1.3)$$

Pinheiro *et al.* (2001) provided some efficient EM-type algorithms for maximum likelihood (ML) estimation in t linear mixed models with heteroscedastic errors. Since longitudinal data are occasionally collected over time, observations within each subject tend to be autocorrelated. To account for the happened serial correlation, I exploit a stationary AR(p) dependence structure for the within-subject errors. Notice that the pure AR model can be extended to a richer ARMA family (Lin and Lee, 2003; Lee *et al.*, 2005). Nevertheless, it is appropriate and relatively simple to fit high-order AR models instead of using complicated ARMA models due to the fact that longitudinal data are often collections of short time series.

In (1.2), if one replaces ν with ∞ and \mathbf{C}_i with \mathbf{I}_{n_i} , then model (1.1) reduces to the normal linear mixed model (Hartely and Rao, 1967). Laird and Ware (1982) discussed in detail how the EM algorithm (Dempster *et al.*, 1977) can be applied to estimate the parameters of this model. To tackle the problem of slow convergence with EM, Schafer (1998) described a hybrid procedure which combines the stability of EM with the rapid convergence of Fisher scoring method. The improved procedures developed in Schafer (1998) are now bundled in the R package `lmm`, which can be freely downloaded from the web site (<http://cran.r-project.org/>).

In the next section, computational aspects of ML estimation are described. The estimation of random effects and prediction of future values are discussed in Section 3. The proposed methodologies are illustrated in Section 4 with a real data set and a brief discussion is given in the final section.

2. Computational Aspects

2.1 Likelihood inference

It follows from (1.3) that the marginal density of \mathbf{Y}_i is

$$f(\mathbf{Y}_i) = \frac{\Gamma(\frac{\nu+n_i}{2}) |\mathbf{\Lambda}_i|^{-1/2}}{\Gamma(\frac{\nu}{2})(\pi\nu\sigma^2)^{n_i/2}} \left(1 + \frac{(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{\Lambda}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})}{\nu\sigma^2} \right)^{-(\nu+n_i)/2}, \quad (2.1)$$

where $\mathbf{\Lambda}_i = \mathbf{\Lambda}_i(\boldsymbol{\Gamma}, \boldsymbol{\phi}) = \mathbf{Z}_i\boldsymbol{\Gamma}\mathbf{Z}_i^T + \mathbf{C}_i$ and $\mathbf{C}_i = \mathbf{C}_i(\boldsymbol{\phi})$. This implies that $\mathbf{Y}_i \sim t_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \sigma^2\mathbf{\Lambda}_i, \nu)$.

For notational simplicity, I denote $\mathbf{e}_i = \mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}$ and $\Delta_i = \Delta_i(\boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\phi}) = \mathbf{e}_i^T \mathbf{\Lambda}_i^{-1} \mathbf{e}_i$. Letting $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \sigma^2, \nu, \boldsymbol{\Gamma}, \boldsymbol{\phi})$ be the model parameters, the log-likelihood

function for $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_N)$, omitting the constant term, is

$$\begin{aligned} \ell(\boldsymbol{\alpha} | \mathbf{Y}) &= \sum_{i=1}^N \left(\log \Gamma\left(\frac{\nu + n_i}{2}\right) - \log \Gamma\left(\frac{\nu}{2}\right) \right) - \frac{n}{2} \log(\nu \sigma^2) \\ &\quad - \frac{1}{2} \sum_{i=1}^N \log |\boldsymbol{\Lambda}_i| - \frac{1}{2} \sum_{i=1}^N (\nu + n_i) \log \left(1 + \frac{\Delta_i}{\sigma^2 \nu} \right), \end{aligned} \quad (2.2)$$

where $n = \sum_{i=1}^N n_i$ is the total number of observations from all subjects in the study.

To ensure the nonnegative definiteness of $\boldsymbol{\Gamma}$ in the estimating procedure, I reparameterize $\boldsymbol{\Gamma} = \mathbf{U}^T \mathbf{U}$ by the Cholesky decomposition, where \mathbf{U} is an upper triangular matrix. In order to ensure uniqueness of $\boldsymbol{\Gamma}$, I restrict the diagonal elements of \mathbf{U} to be positive. Let $\boldsymbol{\omega}$ be a $g \times 1$ vector that contains $m_2(m_2 + 1)/2$ distinct entries in \mathbf{U} and the autoregressive coefficients in $\boldsymbol{\phi}$. Hence, $g = (m_2^2 + m_2 + 2p)/2$ and $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \sigma^2, \nu, \boldsymbol{\omega})$. The following proposition is useful in the derivation of the information matrix.

Proposition 1. For model (1.1), the following holds:

- (a) $\sigma^{-2} \Delta_i \sim n_i F(n_i, \nu)$.
- (b) $\nu / (\nu + \sigma^{-2} \Delta_i) \sim \text{Beta}(\nu/2, n_i/2)$.
- (c) $E \left(\frac{\Delta_i}{(\sigma^2 \nu + \Delta_i)^2} \right) = \frac{n_i}{\sigma^2 (\nu + n_i) (\nu + n_i + 2)}$.

The proof follows directly from the essential properties of the multivariate t distribution (see, e.g., Nadarajah and Kotz, 2005) and hence is omitted. \square

2.2 The Fisher-scoring algorithm

In light of (2.2), there is no closed-form solution available for the ML estimates. Instead, some certain numerical techniques such as the scoring method can be used to find optimal parameter estimates. Explicit expressions of the score vector $\mathbf{s}_{\boldsymbol{\alpha}}$ and the Fisher information matrix $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}$ required for the developed Fisher scoring algorithm are sketched in Appendix A.

Let $\boldsymbol{\theta} = (\sigma^2, \nu, \boldsymbol{\omega})$ denote the set of parameters excluding the fixed effects $\boldsymbol{\beta}$, then $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \boldsymbol{\theta})$, $\mathbf{s}_{\boldsymbol{\alpha}} = (\mathbf{s}_{\boldsymbol{\beta}}, \mathbf{s}_{\boldsymbol{\theta}})$ and

$$\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} = \begin{bmatrix} \mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\theta}} \\ \mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\theta}}^T & \mathbf{J}_{\boldsymbol{\theta}\boldsymbol{\theta}} \end{bmatrix}, \quad (2.3)$$

which is a block partitioned matrix since $\mathbf{J}_{\beta\theta}$ is an $m_2 \times (g + 2)$ zero matrix.

The update of the current estimate $\hat{\theta}^{(k)}$ at the k th iteration is obtained through the following recursive equation

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + \hat{\mathbf{J}}_{\theta\theta}^{(k)-1} \hat{\mathbf{s}}_{\theta}^{(k)}, \tag{2.4}$$

where $\hat{\mathbf{J}}_{\theta\theta}^{(k)}$ and $\hat{\mathbf{s}}_{\theta}^{(k)}$ are $\mathbf{J}_{\theta\theta}$ and \mathbf{s}_{θ} evaluated at $\hat{\beta}^{(k)}$ and $\hat{\theta}^{(k)}$, respectively.

Meanwhile, the update of the current estimate $\hat{\beta}^{(k)}$ is obtained by a generalized-least-squares step as follows:

$$\hat{\beta}^{(k+1)} = \left(\sum_{i=1}^N \hat{w}_i^{(k+1)} \mathbf{X}_i^T \hat{\Lambda}_i^{(k+1)-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \hat{w}_i^{(k+1)} \mathbf{X}_i^T \hat{\Lambda}_i^{(k+1)-1} \mathbf{Y}_i, \tag{2.5}$$

where $\hat{w}_i^{(k+1)} = (\hat{\nu}^{(k+1)} + n_i) / (\hat{\nu}^{(k+1)} + \Delta_i(\hat{\beta}^{(k)}, \hat{\Gamma}^{(k+1)}, \hat{\phi}^{(k+1)}))$.

The ML estimates of β and θ are obtained by iterating (2.4) and (2.5) until $\|\hat{\alpha}^{(k+1)} - \hat{\alpha}^{(k)}\| / \|\hat{\alpha}^{(k)}\|$ converges according to a default tolerance, say 10^{-8} .

In order to facilitate the estimating procedure and achieve the objective of ensuring admissibility of ϕ , we perform a reparameterization on ϕ as in Barndorff-Nielsen and Schou (1973):

$$\phi_k^{(k)} = \pi_k, \quad \phi_j^{(k)} = \phi_j^{(k-1)} - \pi_k \phi_{k-j}^{(k-1)}, \quad j = 1, 2, \dots, k - 1, \tag{2.6}$$

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 $j = 1, \dots, p - 1$.

Notice that (2.6) 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 $\pi = (\pi_1, \dots, \pi_p) \in \mathbb{R}^p$, where $\mathbb{R} = [-1, 1]$.

In the scoring procedure, one needs to evaluate the inverse of $\mathbf{C}_i(\phi)$, denoted by $\mathbf{C}_i^{-1}(\phi)$, and its differentiation with respect to ϕ_k for $k = 1, \dots, p$. However, it is very difficult to obtain $\partial \mathbf{C}_i(\phi) / \partial \phi_k$ directly. If the inverse of $\mathbf{C}_i(\phi)$ can be explicitly expressed, its differentiation can be obtained by the formula $\partial \mathbf{C}_i(\phi) / \partial \phi_k = -\mathbf{C}_i(\phi) [\partial \mathbf{C}_i^{-1}(\phi) / \partial \phi_k] \mathbf{C}_i(\phi)$. To get the inverse as well as the differentiation of the AR(p) autocovariance matrix in analytical forms, I need the following proposition:

Proposition 2. The inverse of the autocovariance matrix for the AR(p) process is

$$\mathbf{C}_i^{-1}(\phi) = \sigma^{-2} (\mathbf{L}_{\phi}^T \mathbf{L}_{\phi} - \mathbf{H}_{\phi} \mathbf{H}_{\phi}^T), n_i \geq p,$$

where \mathbf{H}_ϕ and \mathbf{L}_ϕ are $n_i \times p$ and $n_i \times n_i$ matrices, respectively, and $[\mathbf{H}_\phi, \mathbf{L}_\phi] = [\eta_{ij}]$ with

$$\eta_{ij} = \begin{cases} -\phi_k, & j = i + p - k, k = 1, 2, \dots, p, \\ 1, & j = i + p, \\ 0, & \text{otherwise.} \end{cases}$$

Proof. See Lin and Ho (2008). \square

By Proposition 2, the derivative of \mathbf{C}_i^{-1} with respect to ϕ_k is

$$\begin{aligned} \frac{\partial \mathbf{C}_i^{-1}(\phi)}{\partial \phi_k} &= \frac{\partial}{\partial \phi_k} (\mathbf{L}_\phi^\top \mathbf{L}_\phi - \mathbf{H}_\phi \mathbf{H}_\phi^\top) \\ &= \left(\frac{\partial \mathbf{L}_\phi}{\partial \phi_k} \right)^\top \mathbf{L}_\phi + \mathbf{L}_\phi^\top \left(\frac{\partial \mathbf{L}_\phi}{\partial \phi_k} \right) \\ &\quad - \left(\frac{\partial \mathbf{H}_\phi}{\partial \phi_k} \right) \mathbf{H}_\phi^\top - \mathbf{H}_\phi \left(\frac{\partial \mathbf{H}_\phi}{\partial \phi_k} \right)^\top, \end{aligned} \quad (2.7)$$

where

$$\left[\frac{\partial \mathbf{H}_\phi}{\partial \phi_k} \mid \frac{\partial \mathbf{L}_\phi}{\partial \phi_k} \right] = \frac{\partial}{\partial \phi_k} [\mathbf{H}_\phi \mid \mathbf{L}_\phi] = [b_{ij(k)}],$$

with

$$b_{ij(k)} = \begin{cases} -1, & \text{if } j = i + p - k; \\ 0, & \text{otherwise.} \end{cases}, k = 1, \dots, p.$$

When data arise from the t model, one can choose the initial values by fitting a normal counterpart with a relatively large starting value for the degrees of freedom, e.g., $\hat{\nu}^{(0)} = 100$. However, they may be far from optimum and subsequently cause divergence in the scoring procedure. To prevent this obstacle, it is suggested that the user can run the ECME algorithm (described in the next subsection) with moderate iterations to seek ideal starting values before implementing the scoring procedure.

2.3 An efficient ECME algorithm for ML estimation

The EM algorithm (Dempster *et al.*, 1977) is a popular iterative algorithm for ML estimation in models with incomplete data and has several appealing features such as stability of monotone convergence and simplicity of implementation. However, EM loses some of its attraction when its M-step becomes computationally intractable. To cope with this problem, Meng and Rubin (1993)

introduced the ECM algorithm, which is itself an extension of the EM algorithm, that replaces the M-step of EM with a sequence of computationally simpler conditional maximization (CM) steps. The ECME algorithm (Liu and Rubin, 1994), a further generalization of EM, extends ECM with the CM-steps by maximizing *either* the expected complete data log-likelihood function or the correspondingly constrained actual log-likelihood function, called the ‘CML-step’. The main advantage of the ECME algorithm is that it not only preserves the nice features of EM and ECM, but also converges substantially faster than EM and ECM, as demonstrated by Liu and Rubin (1995) for ML estimation of multivariate t distribution with unknown degrees of freedom.

Sometimes, the CML-step in ECME might need a high-dimensional search when many parameters are involved in the constrained log-likelihood function. To circumvent such a difficulty, the CML-step can be implemented by incorporating the Fisher-scoring algorithm with *step-halving* to ensure increasing the log-likelihood. In practice, it is not necessary to iterate until the scoring procedure converges. Instead, a few scoring steps are enough provided that the constrained log-likelihood function increases at each iteration.

I next present a modified version of the ECME algorithm which is computationally feasible and delicate. It follows from (1.3) that, conditional on the missing weight τ_i , the joint distribution of \mathbf{Y}_i and \mathbf{b}_i is

$$\begin{aligned} \begin{bmatrix} \mathbf{Y}_i \\ \mathbf{b}_i \end{bmatrix} \Big| \tau_i &\sim N_{n_i+m_2} \left(\begin{bmatrix} \mathbf{X}_i \boldsymbol{\beta} \\ \mathbf{0} \end{bmatrix}, \frac{\sigma^2}{\tau_i} \begin{bmatrix} \mathbf{Z}_i \boldsymbol{\Gamma} \mathbf{Z}_i^T + \mathbf{C}_i(\phi) & \mathbf{Z}_i \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma} \mathbf{Z}_i^T & \boldsymbol{\Gamma} \end{bmatrix} \right), \\ \tau_i &\sim \text{Gamma} \left(\frac{\nu}{2}, \frac{\nu}{2} \right). \end{aligned} \quad (2.8)$$

From (2.8), it is not difficult to verify that

$$E(\mathbf{b}_i | \mathbf{Y}_i) = E(E(\mathbf{b}_i | \mathbf{Y}_i, \tau_i)) = \boldsymbol{\Gamma} \mathbf{Z}_i^T \boldsymbol{\Lambda}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}) \quad (2.9)$$

and

$$\text{cov}(\mathbf{b}_i | \mathbf{Y}_i, \tau_i) = \frac{\sigma^2}{\tau_i} (\mathbf{Z}_i^T \mathbf{C}_i^{-1} \mathbf{Z}_i + \boldsymbol{\Gamma}^{-1})^{-1}. \quad (2.10)$$

Let $\mathbf{b} = (\mathbf{b}_1, \dots, \mathbf{b}_N)$ and $\boldsymbol{\tau} = (\tau_1, \dots, \tau_N)$. In the ECME algorithm, \mathbf{b} and $\boldsymbol{\tau}$ are treated as missing values. Accordingly, the complete-data log-likelihood

function of $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Gamma}, \boldsymbol{\phi}, \nu)$, omitting the constant term, is given by

$$\begin{aligned} \ell_c(\boldsymbol{\alpha}|\mathbf{Y}) &= -\frac{n + Nm_2}{2} \log(\sigma^2) - \frac{1}{2} \sum_{i=1}^N \log |\mathbf{C}_i| - \frac{N}{2} \log |\boldsymbol{\Gamma}| \\ &\quad - \frac{1}{2\sigma^2} \text{tr} \left(\boldsymbol{\Gamma}^{-1} \sum_{i=1}^N \tau_i \mathbf{b}_i \mathbf{b}_i^T \right) \\ &\quad - \frac{1}{2\sigma^2} \sum_{i=1}^N \text{tr} \left(\mathbf{C}_i^{-1} \tau_i (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i) (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i)^T \right) \\ &\quad + N \left(\frac{\nu}{2} \log \left(\frac{\nu}{2} \right) - \log \Gamma \left(\frac{\nu}{2} \right) \right) + \frac{\nu}{2} \sum_{i=1}^N (\log(\tau_i) - \tau_i). \end{aligned} \quad (2.11)$$

Based on the property of the multivariate normal distribution concerning its conjugacy for the prior distributions of \mathbf{b}_i and τ_i , applying Bayes' rule yields the following posterior distributions

$$\mathbf{b}_i | \mathbf{Y}_i, \tau_i \sim N_{m_2} \left(\boldsymbol{\Gamma} \mathbf{Z}_i^T \boldsymbol{\Lambda}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}), \frac{\sigma^2}{\tau_i} (\mathbf{Z}_i^T \mathbf{C}_i^{-1} \mathbf{Z}_i + \boldsymbol{\Gamma}^{-1})^{-1} \right) \quad (2.12)$$

$$\tau_i | \mathbf{Y}_i \sim \text{Gamma} \left(\frac{\nu + n_i}{2}, \frac{\nu + \sigma^{-2} \Delta_i}{2} \right). \quad (2.13)$$

Let $\hat{\boldsymbol{\alpha}}^{(k)}$ be the estimate of $\boldsymbol{\alpha}$ at the k th iteration. At the $(k+1)$ st iteration, it needs to calculate the so-called Q -function, i.e., $Q(\boldsymbol{\alpha} | \hat{\boldsymbol{\alpha}}^{(k)}) = E(\ell_c(\boldsymbol{\alpha}|\mathbf{Y}) | \mathbf{Y}, \hat{\boldsymbol{\alpha}}^{(k)})$, which is the conditional expectation of (2.11) given \mathbf{Y} and $\hat{\boldsymbol{\alpha}}^{(k)}$. It can be observed from (2.11) that the conditional expectations required for the calculation of the Q -function are $E(\tau_i | \mathbf{Y}_i, \hat{\boldsymbol{\alpha}}^{(k)})$, $E(\log \tau_i | \mathbf{Y}_i, \hat{\boldsymbol{\alpha}}^{(k)})$, $E(\tau_i \mathbf{b}_i | \mathbf{Y}_i, \hat{\boldsymbol{\alpha}}^{(k)})$ and $E(\tau_i \mathbf{b}_i \mathbf{b}_i^T | \mathbf{Y}_i, \hat{\boldsymbol{\alpha}}^{(k)})$.

Proposition 3. Taking the conditional expectation of the complete-data log-likelihood (2.11) with respect to the missing data \mathbf{b} and $\boldsymbol{\tau}$ given the observed data \mathbf{Y} at its current estimate $\hat{\boldsymbol{\alpha}}^{(k)}$, we have

$$\begin{aligned} Q(\boldsymbol{\alpha} | \hat{\boldsymbol{\alpha}}^{(k)}) &= -\frac{(n + Nm_2)}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N \text{tr} \left(\mathbf{C}_i^{-1} \hat{\boldsymbol{\Psi}}_{1i}^{(k)}(\boldsymbol{\beta}) \right) - \frac{1}{2} \sum_{i=1}^N \log |\mathbf{C}_i| \\ &\quad - \frac{N}{2} \log |\boldsymbol{\Gamma}| - \frac{1}{2\sigma^2} \text{tr} \left(\boldsymbol{\Gamma}^{-1} \sum_{i=1}^N \hat{\boldsymbol{\Psi}}_{2i}^{(k)} \right) + N \left(\frac{\nu}{2} \log \left(\frac{\nu}{2} \right) - \log \Gamma \left(\frac{\nu}{2} \right) \right) \\ &\quad + \frac{\nu}{2} \sum_{i=1}^N (\hat{\kappa}_i^{(k)} - \hat{\tau}_i^{(k)}), \end{aligned} \quad (2.14)$$

where

$$\hat{\tau}_i^{(k)} = E\left(\tau_i \mid \mathbf{Y}_i, \hat{\boldsymbol{\alpha}}^{(k)}\right) = \frac{\hat{\nu}^{(k)} + n_i}{\hat{\nu}^{(k)} + \hat{\sigma}^{-2(k)} \hat{\Delta}^{(k)}}, \quad (2.15)$$

$$\begin{aligned} \hat{\kappa}_i^{(k)} &= E\left(\log \tau_i \mid \mathbf{Y}_i, \hat{\boldsymbol{\alpha}}^{(k)}\right) \\ &= \text{DG}\left(\frac{\hat{\nu}^{(k)} + n_i}{2}\right) - \log\left(\frac{\hat{\nu}^{(k)} + \hat{\sigma}^{-2(k)} \hat{\Delta}^{(k)}}{2}\right), \end{aligned} \quad (2.16)$$

$$\begin{aligned} \hat{\Psi}_{1i}^{(k)}(\boldsymbol{\beta}) &= E\left(\tau_i(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top \mid \mathbf{Y}_i, \hat{\boldsymbol{\alpha}}^{(k)}\right) \\ &= \hat{\tau}_i^{(k)}\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\hat{\mathbf{b}}_i^{(k)}\right)\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\hat{\mathbf{b}}_i^{(k)}\right)^\top \\ &\quad + \hat{\sigma}^{2(k)}\mathbf{Z}_i\left(\mathbf{Z}_i^\top \hat{\mathbf{C}}_i^{(k)-1}\mathbf{Z}_i + \hat{\Gamma}^{(k)-1}\right)^{-1}\mathbf{Z}_i^\top, \end{aligned} \quad (2.17)$$

$$\begin{aligned} \hat{\Psi}_{2i}^{(k)} &= E(\tau_i\mathbf{b}_i\mathbf{b}_i^\top \mid \mathbf{Y}_i, \hat{\boldsymbol{\alpha}}^{(k)}) \\ &= \hat{\tau}_i^{(k)}\hat{\mathbf{b}}_i^{(k)}\hat{\mathbf{b}}_i^{(k)\top} + \hat{\sigma}^2\left(\mathbf{Z}_i^\top \hat{\mathbf{C}}_i^{(k)-1}\mathbf{Z}_i + \hat{\Gamma}^{(k)-1}\right)^{-1}, \end{aligned} \quad (2.18)$$

with $\hat{\mathbf{b}}_i^{(k)} = E(\mathbf{b}_i \mid \mathbf{Y}_i, \hat{\boldsymbol{\alpha}}^{(k)}) = \hat{\Gamma}^{(k)}\mathbf{Z}_i^\top \hat{\Lambda}_i^{(k)-1}(\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}^{(k)})$ and $\hat{\Delta}_i^{(k)}$ being $\Delta_i(\boldsymbol{\beta}, \Gamma, \phi)$ evaluated at $\hat{\boldsymbol{\beta}}^{(k)}$, $\hat{\Gamma}^{(k)}$ and $\hat{\phi}^{(k)}$.

Proof: See Appendix B. □

Applying Proposition 3 leads to the following ECME algorithm:

E-step: Given $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}^{(k)}$, impute $\hat{\tau}_i^{(k)}$, $\hat{\kappa}_i^{(k)}$, $\hat{\Psi}_{1i}^{(k)}(\boldsymbol{\beta})$ and $\hat{\Psi}_{2i}^{(k)}$ for $i = 1, \dots, N$ by using Eqs (2.15)-(2.18).

CM-steps:

CM-step 1: Update $\hat{\boldsymbol{\beta}}^{(k)}$ by

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \left(\sum_{i=1}^N \hat{\tau}_i^{(k)}\mathbf{X}_i^\top \hat{\mathbf{C}}_i^{(k)-1}\mathbf{X}_i\right)^{-1}\sum_{i=1}^N \hat{\tau}_i^{(k)}\mathbf{X}_i^\top \hat{\mathbf{C}}_i^{(k)-1}(\hat{\mathbf{Y}}_i^{(k)} - \mathbf{Z}_i\hat{\mathbf{b}}_i^{(k)}), \quad (2.19)$$

which is obtained by maximizing (2.14) over $\boldsymbol{\beta}$ given $\phi = \hat{\phi}^{(k)}$.

CM-step 2: Fix $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{(k+1)}$, $\Gamma = \hat{\Gamma}^{(k)}$ and $\phi = \hat{\phi}^{(k)}$, and update $\hat{\sigma}^{2(k)}$ by maximizing (2.14) over σ^2 , which gives

$$\hat{\sigma}^{2(k+1)} = \frac{1}{n + Nm_2}\sum_{i=1}^N \left(\text{tr}(\hat{\mathbf{C}}_i^{(k)-1}\hat{\Psi}_{1i}^{(k)}(\hat{\boldsymbol{\beta}}^{(k+1)})) + \text{tr}(\hat{\Gamma}^{(k)-1}\hat{\Psi}_{2i}^{(k)})\right) \quad (2.20)$$

CM-step 3: Update $\hat{\Gamma}^{(k)}$ by

$$\hat{\Gamma}^{(k+1)} = \frac{1}{N\hat{\sigma}^{2(k+1)}}\sum_{i=1}^N \hat{\Psi}_{2i}^{(k)}, \quad (2.21)$$

which is obtained by maximizing (2.14) over $\mathbf{\Gamma}$ given $\sigma^2 = \hat{\sigma}^{2(k+1)}$.

CML-step 4: Let $\boldsymbol{\eta} = (\nu, \boldsymbol{\phi})$ and update $\hat{\boldsymbol{\eta}}^{(k+1)} = (\hat{\nu}^{(k+1)}, \hat{\boldsymbol{\phi}}^{(k+1)})$ by maximizing the constrained log-likelihood (2.2) using the most current estimates. Therefore, I have

$$h(\boldsymbol{\eta}) = \sum_{i=1}^N \left(\log \Gamma \left(\frac{\nu + n_i}{2} \right) - \log \Gamma \left(\frac{\nu}{2} \right) \right) - \frac{n}{2} \log(\nu) - \frac{1}{2} \sum_{i=1}^N \log | \mathbf{\Lambda}_i(\hat{\mathbf{\Gamma}}^{(k+1)}, \boldsymbol{\phi}) | - \frac{1}{2} \sum_{i=1}^N (\nu + n_i) \log \left(1 + \frac{\Delta_i(\hat{\boldsymbol{\beta}}^{(k+1)}, \hat{\mathbf{\Gamma}}^{(k+1)}, \boldsymbol{\phi})}{\nu \hat{\sigma}^{2(k+1)}} \right). \quad (2.22)$$

To guarantee the increase of $h(\hat{\boldsymbol{\eta}}^{(k)})$ in the spirit of GEM (Dempster *et al.*, 1977), a simple way is to perform one-step scoring with a simple step-halving operation, which can be implemented by updating $\hat{\boldsymbol{\eta}}^{(k+1)} = \hat{\boldsymbol{\eta}}^{(k)} + (1/2)^{m-1} \hat{\mathbf{J}}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{(k)-1} \hat{\mathbf{s}}_{\boldsymbol{\eta}}^{(k)}$, where m is the smallest positive integer k satisfying $h(\hat{\boldsymbol{\eta}}^{(k+1)}) > h(\hat{\boldsymbol{\eta}}^{(k)})$.

For the multivariate t model, the ECME algorithm might be too painfully slow to be of any practical use. A remedy is to employ a hybrid ECME-scoring algorithm by running a moderate number of ECME iterations and then switch to the scoring algorithm. Ideally, this hybrid procedure can enhance the convergence rate and offer the asymptotic standard errors at convergence.

2.4 Large sample inferences

To make inference on the parameters such as asymptotic standard errors, confidence regions and hypothesis testing, one usually relies on the approximate distributions of the ML estimators $\hat{\boldsymbol{\alpha}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{\theta}} = (\hat{\sigma}^2, \hat{\nu}, \hat{\boldsymbol{\omega}})$ are estimators of variance components. For model (1.1), the consistency and asymptotic normality can be established under some mild regularity conditions sketched in Zacks (1971). Further, I make the following assumptions:

- (a) Model (1.1) is correctly specified.
- (b) The parameter spaces Θ for $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \boldsymbol{\theta})$ are compact.
- (c) The true values of $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \boldsymbol{\theta})$, denoted as $\boldsymbol{\alpha}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$, are all interiors points of Θ .
- (d) The design matrices \mathbf{X}_i and \mathbf{Z}_i in model (1.1) are full rank and all of their elements are bounded by a single finite real number.

With technical skills described in Miller (1977) and Demidenko and Stukel (2002, Sec. 5), it can be proved that the ML estimator $\hat{\boldsymbol{\alpha}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$ is consistent

for $\alpha_0 = (\beta_0, \theta_0)$ and is asymptotically normally distributed, say

$$\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta_0 \\ \hat{\theta} - \theta_0 \end{bmatrix} \xrightarrow{D} N_d(\mathbf{0}, \mathbf{J}_{\alpha\alpha}^{-1}(\alpha_0)),$$

where d is the number of distinct parameters in α , the symbol ' \xrightarrow{D} ' stands for convergence in distribution and $\mathbf{J}_{\alpha\alpha}(\alpha_0)$ is the Fisher information matrix $\mathbf{J}_{\alpha\alpha}$ in (2.3) evaluated at $\alpha = \alpha_0$.

From the block-diagonal form of $\mathbf{J}_{\alpha\alpha}$ (i.e. $\mathbf{J}_{\beta\theta} = \mathbf{J}_{\theta\beta}^T = \mathbf{0}$), it follows that $\hat{\beta}$ and $\hat{\theta}$ are asymptotically independent. Therefore, the asymptotic confidence regions and hypothesis tests for fixed effects β and variance components θ can be obtained separately through their approximate $N_{m_1}(\beta, \hat{\mathbf{J}}_{\beta\beta}^{-1})$ and $N_{d-m_1}(\theta, \hat{\mathbf{J}}_{\theta\theta}^{-1})$ distributions. Using the normal approximation, an approximate $(1 - \alpha)\%$ confidence region for β is thus provided by

$$\left\{ \beta \in \mathbb{R}^{m_1} \mid (\beta - \hat{\beta})^T \hat{\mathbf{J}}_{\beta\beta} (\beta - \hat{\beta}) \leq \chi_{m_1}^2(\alpha) \right\},$$

where $\chi_{m_1}^2(\alpha)$ is the $(1 - \alpha)$ quantile of $\chi_{m_1}^2$.

In practice, the normal approximation could be more accurate when the parameter space is unrestricted. Hence, it is recommend to use $\log \sigma^2$ in place of σ^2 , $\log(1/\nu)$ in place of ν and $\text{logit}(\pi_i)$ in place of $\pi_i \in [-1, 1]$ (π_i : reparameterization of ϕ_i) for $i = 1, \dots, p$. One may obtain the confidence intervals of the parameters by computing their unrestricted scale, then inverting them back to the original scale.

3. Estimation of Random Effects and Prediction

One usually focuses on estimating the parameters in a marginal model. However, inference for random effects may sometimes be attractive. Since the random effects \mathbf{b}_i 's are unobservable, their estimates can be obtained using the empirical Bayes technique (Laird and Ware, 1982). Combining distributions in (1.3) and then integrating with respect to τ_i , the joint distribution of $(\mathbf{Y}_i, \mathbf{b}_i)$ is

$$\begin{bmatrix} \mathbf{Y}_i \\ \mathbf{b}_i \end{bmatrix} \sim t_{n_i+m_2} \left(\begin{bmatrix} \mathbf{X}_i \beta \\ \mathbf{0} \end{bmatrix}, \sigma^2 \begin{bmatrix} \Lambda_i & \mathbf{Z}_i \Gamma \\ \Gamma \mathbf{Z}_i^T & \Gamma \end{bmatrix}, \nu \right). \quad (3.1)$$

Following Laird and Ware (1982), the empirical Bayes estimate of \mathbf{b}_i can be obtained by the expectation of the posterior density of \mathbf{b}_i given \mathbf{Y}_i with unknown parameters replaced by their ML estimates.

From (2.9), the empirical Bayes estimate of \mathbf{b}_i is given by

$$\hat{\mathbf{b}}_i = E(\mathbf{b}_i | \mathbf{Y}_i, \hat{\alpha}) = \hat{\Gamma} \mathbf{Z}_i^T \hat{\Lambda}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}), \quad (3.2)$$

where $\hat{\boldsymbol{\alpha}} = (\hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \hat{\boldsymbol{\Gamma}}, \hat{\boldsymbol{\phi}}, \hat{\nu})$ represents the ML estimates of parameters at convergence.

I next consider the prediction of \mathbf{y}_i , a future $q \times 1$ vector of measurements \mathbf{Y}_i , given the observed unbalanced repeated measurements $\mathbf{Y} = (\mathbf{Y}_{(i)}^T, \mathbf{Y}_i^T)^T$, where $\mathbf{Y}_{(i)} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_{i-1}^T, \mathbf{Y}_{i+1}^T, \dots, \mathbf{Y}_N^T)^T$. This is called the *conditional prediction* by Lee and Geisser (1972). To make this type of prediction, the dependence structure must be extendable to future values of all observed subjects. The AR(p) model considered in this paper satisfies this requirement.

Let \mathbf{x}_i and \mathbf{z}_i denote $q \times m_1$ and $q \times m_2$ design matrices of prediction regressors corresponding to \mathbf{y}_i . Further, I assume that

$$\begin{bmatrix} \mathbf{Y}_i \\ \mathbf{y}_i \end{bmatrix} \sim t_{n_i+q} \left(\begin{bmatrix} \mathbf{X}_i \boldsymbol{\beta} \\ \mathbf{x}_i \boldsymbol{\beta} \end{bmatrix}, \sigma^2 \left(\mathbf{Z}_i^* \boldsymbol{\Gamma} \mathbf{Z}_i^{*T} + \mathbf{C}_i^*(\boldsymbol{\phi}) \right), \nu \right), \quad (3.3)$$

where $\mathbf{Z}_i^* = (\mathbf{Z}_i^T, \mathbf{z}_i^T)^T$ and $\mathbf{C}_i^*(\boldsymbol{\phi}) = [\rho_{|r-s|}(\boldsymbol{\phi})]$ for $r, s = 1, \dots, n_i + q$.

Let $\boldsymbol{\Omega} = \mathbf{Z}_i^* \boldsymbol{\Gamma} \mathbf{Z}_i^{*T} + \mathbf{C}_i^*(\boldsymbol{\phi}) = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{bmatrix}$, where $\boldsymbol{\Omega}_{11} = \boldsymbol{\Lambda}_i$ is an $n_i \times n_i$ matrix, $\boldsymbol{\Omega}_{21} = \boldsymbol{\Omega}_{12}^T$ is a $q \times n_i$ matrix, and $\boldsymbol{\Omega}_{22}$ is a $q \times q$ matrix. Here subscript i of $\boldsymbol{\Omega}$ and the partitioned components $\boldsymbol{\Omega}_{11}$, $\boldsymbol{\Omega}_{12}$, $\boldsymbol{\Omega}_{21}$ and $\boldsymbol{\Omega}_{22}$ are suppressed for notational convenience.

Recalling the property concerning the conditional distribution of the multivariate t distribution, it suffices to have

$$\mathbf{y}_i | \mathbf{Y}_i \sim t_q \left(\boldsymbol{\mu}_{2 \cdot 1}, \frac{\sigma^2 \nu + \Delta_i}{\nu + n_i} \boldsymbol{\Omega}_{22 \cdot 1}, \nu + n_i \right), \quad (3.4)$$

where $\boldsymbol{\mu}_{2 \cdot 1} = \mathbf{x}_i \boldsymbol{\beta} + \boldsymbol{\Omega}_{21} \boldsymbol{\Omega}_{11}^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta})$ and $\boldsymbol{\Omega}_{22 \cdot 1} = \boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21} \boldsymbol{\Omega}_{11}^{-1} \boldsymbol{\Omega}_{12}$. The estimated minimum mean square error (MSE) predictor of \mathbf{y}_i is obtained from the conditional expectation of \mathbf{y}_i given \mathbf{Y}_i , i.e.,

$$\hat{\mathbf{y}}_i(\boldsymbol{\alpha}) = \boldsymbol{\mu}_{2 \cdot 1} = \mathbf{x}_i \boldsymbol{\beta} + \boldsymbol{\Omega}_{21} \boldsymbol{\Omega}_{11}^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}). \quad (3.5)$$

Consequently, the MSE covariance matrix of the predictor (3.5) is determined as

$$\begin{aligned} E(\hat{\mathbf{y}}_i(\boldsymbol{\alpha}) - \mathbf{y}_i)(\hat{\mathbf{y}}_i(\boldsymbol{\alpha}) - \mathbf{y}_i)^T &= E(\text{cov}(\mathbf{y}_i | \mathbf{Y}_i)) \\ &= E \left(\frac{\sigma^2 \nu + \Delta_i}{\nu + n_i - 2} \boldsymbol{\Omega}_{22 \cdot 1} \right) \\ &= \frac{\sigma^2 \nu}{\nu - 2} \boldsymbol{\Omega}_{22 \cdot 1}. \end{aligned} \quad (3.6)$$

Typically, the predicted value of \mathbf{y}_i can be obtained by substituting the ML estimate $\hat{\boldsymbol{\alpha}}$ into (3.5). It is noted that expression (3.6) does not account for the

variation of the estimation of unknown parameters, and hence it will underestimate the true value for small size N . A better approximation of (3.6) may be employed by the bootstrap approach (Efron and Tibshirani, 1986), whereas it requires vast amounts of computing power.

4. An Illustrative Example

I applied the results developed in Sections 2 and 3 to the *in vivo* growth of lung tumor for 22 xenografted nude mice allocated in the control group. The data was originally reported by Rygaard and Spang-Thomsen (1997) and subsequently analyzed by Demidenko (2004, Ch. 10). This longitudinal study is to investigate the tumor growth rates for the immune-deficient nude mice with human tumor xenografts implanted after 14 days, defined as the baseline (day 0).

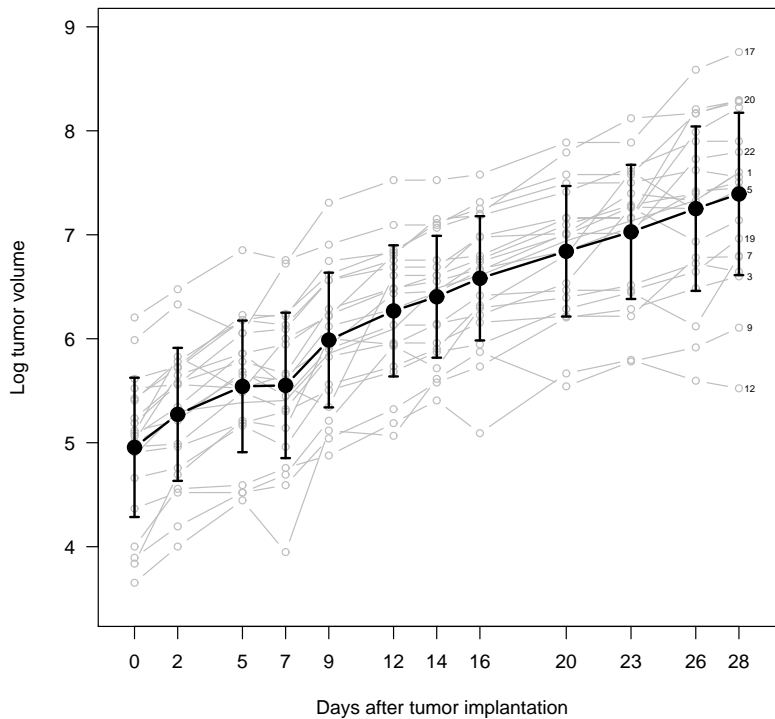


Figure 1: Tumor growth curves of 22 xenografted nude mice. The darker lines indicate the mean profile responses and ± 1 sample standard deviations across time. The numbers on the right of curves are subject indices.

Figure 1 depicts the logarithm of tumor growth volumes over an unevenly

spaced 28-day period for the 22 nude mice assigned to receive the control treatment. The mean level exhibits a linear pattern over time. To diagnose whether there exist outliers in the data, I employed the discordant outlier detection procedure of Pan and Fang (2002, Eq. 4.30) and identified that the 12th mouse is a discordant outlier at the 5% significant level. In other words, the normal linear mixed model could be inappropriate for this data set.

To analyze this data set, I fit model (1.1) with random effects on both the intercepts and slopes. The t linear mixed model for the data is specified by

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad \mathbf{b}_i \sim t_2(\mathbf{0}, \sigma^2\boldsymbol{\Gamma}, \nu), \\ \boldsymbol{\varepsilon}_i &\sim t_{12}(\mathbf{0}, \sigma^2\mathbf{C}_i(\boldsymbol{\phi}), \nu), \quad \mathbf{b}_i \mid \tau_i \perp \boldsymbol{\varepsilon}_i \mid \tau_i, \quad (i = 1, \dots, 22). \end{aligned} \quad (4.1)$$

Here the design matrices are $\mathbf{X} = [\mathbf{1} \ \mathbf{k}]$ and $\mathbf{Z} = \mathbf{X}$, where $\mathbf{1} = (1, \dots, 1)^T$ is a 12×1 unitary vector and $\mathbf{k} = (0, 1, 2.5, 3.5, 4.5, 6, 7, 8, 10, 11.5, 13, 14)^T$. I fit model (4.1) using the entire set of measurements with a white noise (WN) dependence structure, i.e., $\boldsymbol{\phi} = \mathbf{0}$, and three selected AR(p) dependence structures for $p = 1, 2, 3$. Preliminary analysis suggests the correlation of the random intercepts and the random slopes is very weak. Hence, the Cholesky factor \mathbf{U} is taken as a diagonal matrix and satisfies $\boldsymbol{\Gamma} = \mathbf{U}^2$. The resulting ML estimates together with the values of the maximized log-likelihood and AICs, are given in Table 1. As seen in the table, the AR(2) model is the favorite choice since it has the smallest AIC.

Table 1: ML estimation results for various dependence structures

Parameter	WN		AR(1)		AR(2)		AR(3)	
	mle	se	mle	se	mle	se	mle	se
β_0	5.1781	0.1219	5.1644	0.1237	5.1429	0.1290	5.1407	0.1296
β_1	0.1677	0.0070	0.1685	0.0067	0.1670	0.0068	0.1662	0.0069
σ^2	0.0355	0.0054	0.0354	0.0054	0.0407	0.0071	0.0412	0.0077
u_{11}	2.8517	0.4857	2.7352	0.4865	2.5673	0.5583	2.4879	0.7368
u_{22}	0.1542	0.0299	0.1210	0.0336	0.0797	0.0319	0.0786	0.0305
ϕ_1	—	—	0.4333	0.0944	0.4987	0.1165	0.4984	0.1317
ϕ_2	—	—	—	—	0.2614	0.0977	0.2392	0.1011
ϕ_3	—	—	—	—	—	—	0.0812	0.0937
ν	9.1585	4.4354	9.084	4.3800	8.6018	4.1360	8.6243	4.0508
$\ell(\hat{\boldsymbol{\alpha}})$	-30.26		-18.89		-14.11		-13.69	
AIC	72.52		51.78		44.22		45.38	

AIC = $-2(\text{maximized log-likelihood} - \text{number of parameters})$.

When t models are favorable, the profile log-likelihood function for the degrees of freedom ν should have a significant drop around the ML estimate $\hat{\nu}$. To illustrate this, Figure 2 depicts the profile log-likelihood functions of ν for the four selected dependence structures. It is apparent from Figure 2 that all t models appear to be more suitable than their normal counterparts for this data set.

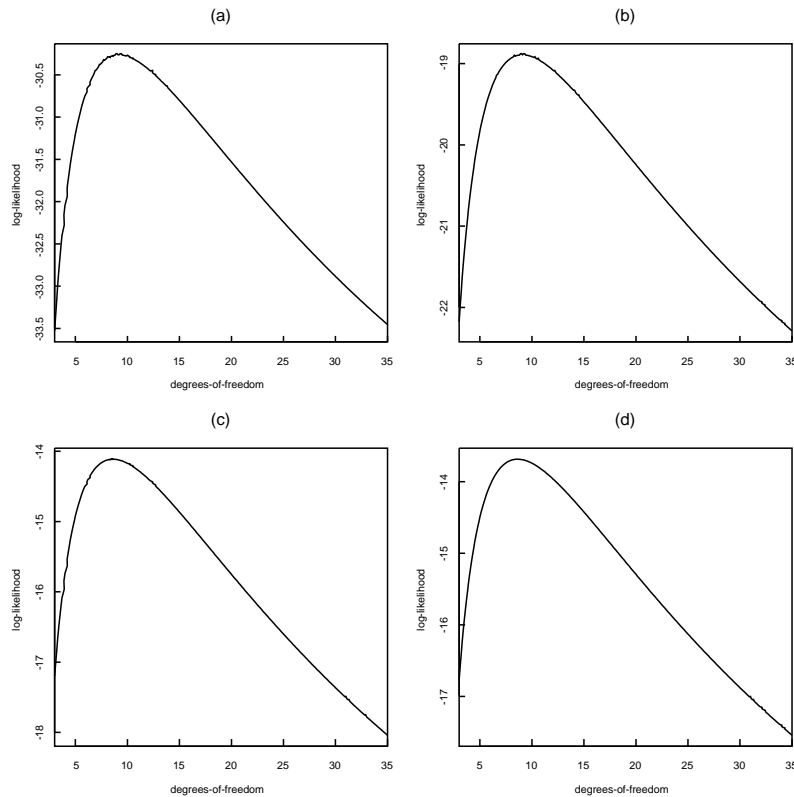


Figure 2: Profile log-likelihood of ν for the t linear mixed model with selected (a) WN; (b) AR(1); (c) AR(2); (d) AR(3) dependence structures.

For dependent longitudinal data, a more appropriate measure of ‘fitness’ is the predictive accuracy of future observations (Rao, 1987; Lee, 1988). I next consider the conditional prediction for the future values, which are usually of practical interest for dependent data. Following the *pseudo cross validation* (PCV) method utilized by Keramidas and Lee (1990), the technique proceeds as follows: (i) holdout the last q measurements on the i th participant; (ii) compute ML estimates using the remaining data as the sample; (iii) predict the q true measurements $(y_{i,12-q+1}, \dots, y_{i,12})^T$, denoted by $(\hat{y}_{i,12-q+1}, \dots, \hat{y}_{i,12})^T$, using formula (3.5). The procedure is repeated across subjects $i = 1, \dots, 22$.

To compare the performance of different models, I use the mean of square deviations $|\hat{y}_{ij} - y_{ij}|^2$, mean of absolute deviations $|\hat{y}_{ij} - y_{ij}|$ and mean of absolute relative deviations $|\hat{y}_{ij} - y_{ij}|/y_{ij}$, abbreviated as MSD, MAD and MARD hereinafter, respectively. To assess the prediction accuracy via PCV, I consider one-step-ahead and two-step-ahead forecasts for the last measurement of each mouse. Table 2 compares the prediction accuracy for the t linear mixed model (4.1) with the normal-based counterpart under the AR(2) structure. The relative improvement percentage (RIP) listed in Table 2 is defined as the percentage decrease in the relative prediction accuracy when using t predictors. From the prediction results summarized in Table 2, it appears that the t model has better prediction performances.

Table 2: A comparison of predictive accuracy in terms of various discrepancy measures between the normal and t linear mixed models with AR(2) dependence. The relative improvement percentage (RIP) is measured by $(\text{normal} - t)/\text{normal} \times 100\%$.

Discrepancy Measure	One-step-ahead			Two-step-ahead		
	normal	t	RIP(%)	normal	t	RIP(%)
MSD	0.0405	0.0384	5.19	0.1242	0.1163	6.36
MAD	0.1702	0.1624	4.58	0.3020	0.2843	5.86
MARD	0.0235	0.0224	4.68	0.0421	0.0396	5.94

It is interesting to compare the prediction accuracies of the t and normal linear mixed models after some contaminations are introduced into the original data set, which is a common way of demonstrating the robustness of a model, see, e.g., Peel and McLachlan (2000). The procedure is implemented by adding various constants $-10, -8, \dots, 8, 10$ to the sixth observation of the first mouse, which is around the middle point of the data. I use the first eleven measurements of each mouse as the sample to predict the last measurement. It is readily seen from Table 3 that the parameters obtained by the normal model are highly affected by a single outlier, whereas the influence for the t model is limited in a short range. The last column also shows that the t model can produce substantial gains in prediction accuracy as the level of perturbation becomes larger. This suggests the t model provides a favorable way for achieving robust statistical inferences.

Table 3: Summary of the comparison of parameter estimates and prediction accuracy when fitting the normal and t linear mixed models using perturbed samples. The first column shows various perturbation constants.

constant	model	ML estimates					MARD	RIP(%)
		β_1	β_2	ϕ_1	ϕ_2	ν		
-10	normal	5.059	0.173	0.018	-0.011	—	0.0431	44.8
	t	5.173	0.167	0.414	0.295	2.921	0.0238	
-8	normal	5.064	0.173	0.045	0.007	—	0.0402	41.3
	t	5.172	0.168	0.413	0.294	3.107	0.0236	
-6	normal	5.068	0.174	0.096	0.042	—	0.0362	36.4
	t	5.171	0.168	0.411	0.293	3.398	0.0235	
-4	normal	5.072	0.174	0.147	0.079	—	0.0287	18.8
	t	5.168	0.168	0.408	0.290	3.950	0.0233	
-2	normal	5.074	0.174	0.257	0.166	—	0.0251	12.6
	t	5.160	0.168	0.400	0.283	5.708	0.0232	
+2	normal	5.083	0.173	0.303	0.211	—	0.0237	4.64
	t	5.162	0.168	0.413	0.295	5.236	0.0226	
+4	normal	5.093	0.175	0.203	0.136	—	0.0275	17.8
	t	5.168	0.168	0.414	0.295	3.840	0.0226	
+6	normal	5.101	0.175	0.130	0.082	—	0.344	34.0
	t	5.171	0.167	0.415	0.296	3.346	0.0227	
+8	normal	5.108	0.176	0.080	0.045	—	0.0393	42.2
	t	5.172	0.167	0.416	0.296	3.076	0.0227	
+10	normal	5.114	0.177	0.051	0.024	—	0.0433	47.3
	t	5.173	0.167	0.416	0.297	2.899	0.0228	

5. Discussion

In this paper, I provide a robust approach to the linear mixed model based on the multivariate t distribution and utilize the convenient $AR(p)$ dependence structure to precisely capture the within-subject dependence, which allows practitioners to analyze longitudinal data in a wide variety of considerations. Besides, the proposed computational technique and prediction method are very easy to implement. Numerical results illustrated in Section 4 indicates the proposed model for this data set is evidently more adequate than the conventional normal linear mixed model in the comparison of model fitting as well as prediction of future values.

A number of competing approaches to the robust estimation of linear mixed models have been proposed in the literature, e.g., robust estimating equations of Huggins (1992, 1993), robust REML of Richardson and Welsh (1995), bounded

influence estimation of Richardson (1997) and robust Huber's ρ method of Gill (2000). It is a worthwhile task to compare the t linear mixed model with alternative competing approaches.

Recently, Lin and Lee (2008) have proposed a novel linear mixed model in which the random effects are assumed to follow a multivariate skew normal distribution (Azzalini and Dalla Valle, 1996; Azzalini and Capitanio, 1999) as an alternative to generalization and mentioned an appropriate specification of random effects may enhance predictive abilities. A unified approach to modeling random effects for longitudinal data within the framework of multivariate skew t distribution (Azzalini and Capitanio, 2003) is in progress and will be reported in a follow-up paper.

Appendix A. The Score Vector and Information Matrix

The score vector, $\mathbf{s}\boldsymbol{\alpha} = \partial\ell(\boldsymbol{\alpha}|\mathbf{Y})/\partial\boldsymbol{\alpha}$, has the following elements:

$$\begin{aligned} \mathbf{s}\boldsymbol{\beta} &= \sum_{i=1}^N (\nu + n_i) \frac{\mathbf{X}_i^T \boldsymbol{\Lambda}_i^{-1} \mathbf{e}_i}{\sigma^2 \nu + \Delta_i}, \\ \mathbf{s}_{\sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^2} \sum_{i=1}^N (\nu + n_i) \left(\frac{\Delta_i}{\sigma^2 \nu + \Delta_i} \right), \\ \mathbf{s}_{\nu} &= \frac{1}{2} \sum_{i=1}^N \left(\text{DG} \left(\frac{\nu + n_i}{2} \right) - \text{DG} \left(\frac{\nu}{2} \right) - \frac{n_i}{\nu} - \log \left(1 + \frac{\Delta_i}{\sigma^2 \nu} \right) + \frac{(\nu + n_i)}{\nu} \frac{\Delta_i}{\sigma^2 \nu + \Delta_i} \right), \\ [\mathbf{s}\boldsymbol{\omega}]_r &= -\frac{1}{2} \sum_{i=1}^N \left(\text{tr}(\boldsymbol{\Lambda}_i^{-1} \dot{\boldsymbol{\Lambda}}_{ir}) - (\nu + n_i) \left(\frac{\mathbf{e}_i^T \boldsymbol{\Lambda}_i^{-1} \dot{\boldsymbol{\Lambda}}_{ir} \boldsymbol{\Lambda}_i^{-1} \mathbf{e}_i}{\sigma^2 \nu + \Delta_i} \right) \right), \end{aligned}$$

where $\Delta_i = \Delta_i(\boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\phi})$ for $r = 1, \dots, g$, $\text{DG}(x) = d \log \Gamma(x)/dx$ is the digamma function and

$$\begin{aligned} \dot{\boldsymbol{\Lambda}}_{ir} &= \frac{\partial \boldsymbol{\Lambda}_i(\boldsymbol{\omega})}{\partial \omega_r} \\ &= \frac{\partial}{\partial \omega_r} (\mathbf{Z}_i^T \boldsymbol{\Gamma} \mathbf{Z}_i + \mathbf{C}_i) \\ &= \begin{cases} \mathbf{Z}_i^T \left(\frac{\partial \mathbf{U}^T}{\partial u_{ij}} \mathbf{U} + \mathbf{U}^T \frac{\partial \mathbf{U}}{\partial u_{ij}} \right) \mathbf{Z}_i & \text{if } \omega_r = \gamma_{ij}; \\ \frac{\partial \mathbf{C}_i}{\partial \phi_k} & \text{if } \omega_r = \phi_k, \end{cases} \end{aligned}$$

with

$$\frac{\partial \mathbf{U}}{\partial u_{ij}} = [a_{st}]_{m_2 \times m_2} \quad \text{and} \quad a_{st} = a_{ts} = \begin{cases} 1 & \text{if } s = i, t = j; \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\frac{\partial \mathbf{C}_i}{\partial \phi_k} = \frac{\partial (\mathbf{C}_i^{-1})^{-1}}{\partial \phi_k} = -\mathbf{C}_i \frac{\partial \mathbf{C}_i^{-1}}{\partial \phi_k} \mathbf{C}_i,$$

where $\partial \mathbf{C}_i^{-1} / \partial \phi_k$ is given in (2.7).

The information matrix, $\mathbf{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} = E(-\partial^2 \ell(\boldsymbol{\alpha} | \mathbf{Y}) / \partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^T)$, has the following elements:

$$\begin{aligned} \mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\beta}} &= \sum_{i=1}^N \frac{\nu + n_i}{\sigma^2 (\nu + n_i + 2)} \mathbf{X}_i^T \boldsymbol{\Lambda}_i^{-1} \mathbf{X}_i, \\ \mathbf{J}_{\boldsymbol{\beta}\sigma} &= \mathbf{0}_{m_1 \times 1}, \quad \mathbf{J}_{\boldsymbol{\beta}\nu} = \mathbf{0}_{m_1 \times 1}, \quad \mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\omega}} = \mathbf{0}_{m_1 \times g}, \\ \mathbf{J}_{\sigma^2 \sigma^2} &= \frac{\nu}{2\sigma^4} \sum_{i=1}^N \frac{n_i}{\nu + n_i + 2}, \quad \mathbf{J}_{\sigma^2 \nu} = -\frac{1}{\sigma^2} \sum_{i=1}^N \frac{n_i}{(\nu + n_i)(\nu + n_i + 2)}, \\ [\mathbf{J}_{\sigma^2 \boldsymbol{\omega}}]_r &= \frac{\nu}{2\sigma^2} \sum_{i=1}^N \frac{1}{\nu + n_i + 2} \text{tr}(\boldsymbol{\Lambda}_i^{-1} \dot{\boldsymbol{\Lambda}}_{ir}), \\ \mathbf{J}_{\nu\nu} &= \frac{1}{4} \sum_{i=1}^N \left(\text{TG}\left(\frac{\nu}{2}\right) - \text{TG}\left(\frac{\nu + n_i}{2}\right) - \frac{2n_i(\nu + n_i + 4)}{\nu(\nu + n_i)(\nu + n_i + 2)} \right), \\ [\mathbf{J}_{\nu \boldsymbol{\omega}}]_r &= -\sum_{i=1}^N \frac{1}{(\nu + n_i)(\nu + n_i + 2)} \text{tr}(\boldsymbol{\Lambda}_i^{-1} \dot{\boldsymbol{\Lambda}}_{ir}), \\ [\mathbf{J}_{\boldsymbol{\omega}\boldsymbol{\omega}}]_{rs} &= \frac{1}{2} \sum_{i=1}^N \frac{1}{\nu + n_i + 2} \left((\nu + n_i) \text{tr}(\boldsymbol{\Lambda}_i^{-1} \dot{\boldsymbol{\Lambda}}_{ir} \boldsymbol{\Lambda}_i^{-1} \dot{\boldsymbol{\Lambda}}_{is}) - \text{tr}(\boldsymbol{\Lambda}_i^{-1} \dot{\boldsymbol{\Lambda}}_{ir}) \text{tr}(\boldsymbol{\Lambda}_i^{-1} \dot{\boldsymbol{\Lambda}}_{is}) \right), \end{aligned}$$

for $r, s = 1, \dots, g$, where $\text{TG}(x) = \frac{d^2}{dx^2} \log \Gamma(x)$ denotes the trigamma function.

Appendix B. Proof of Proposition 3

The conditional density $f(\tau_i | \mathbf{Y}_i)$ is

$$\begin{aligned} f(\tau_i | \mathbf{Y}_i) &\propto f(\mathbf{Y}_i, \tau_i) \\ &\propto \tau_i^{\frac{\nu+n_i}{2}-1} \exp \left\{ -\left(\frac{\nu + \sigma^{-2} \Delta_i}{2}\right) \tau_i \right\}. \end{aligned}$$

It follows that the conditional distribution of $\tau_i | \mathbf{Y}_i$ is Gamma $\left(\frac{\nu + n_i}{2}, \frac{\nu + \sigma^{-2} \Delta_i}{2}\right)$.

Hence, it suffices to show

$$\hat{\tau}_i^{(k)} = E(\tau_i | \mathbf{Y}_i, \hat{\boldsymbol{\alpha}}^{(k)}) = \frac{\hat{\nu}^{(k)} + n_i}{\hat{\nu}^{(k)} + \hat{\sigma}^{-2(k)} \hat{\Delta}^{(k)}},$$

and

$$\hat{\kappa}_i^{(k)} = E(\log \tau_i | \mathbf{Y}_i, \hat{\boldsymbol{\alpha}}^{(k)}) = \text{DG}\left(\frac{\hat{\nu}^{(k)} + n_i}{2}\right) - \log\left(\frac{\hat{\nu}^{(k)} + \hat{\sigma}^{-2(k)} \hat{\Delta}^{(k)}}{2}\right).$$

From (2.9) and (2.10), the following results are obtained.

$$\begin{aligned} \hat{\Psi}_{1i}^{(k)}(\boldsymbol{\beta}) &= E(\tau_i(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top | \mathbf{Y}_i, \hat{\boldsymbol{\alpha}}^{(k)}) \\ &= \hat{\tau}_i^{(k)}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top - (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})E(\tau_i(\mathbf{Z}_i\mathbf{b}_i)^\top | \mathbf{Y}_i, \hat{\boldsymbol{\alpha}}^{(k)}) \\ &\quad - E(\tau_i\mathbf{Z}_i\mathbf{b}_i | \mathbf{Y}_i, \hat{\boldsymbol{\alpha}}^{(k)})(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top + \mathbf{Z}_iE(\tau_i\mathbf{b}_i\mathbf{b}_i^\top | \mathbf{Y}_i, \hat{\boldsymbol{\alpha}}^{(k)})\mathbf{Z}_i^\top \\ &= \hat{\tau}_i^{(k)}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top - \hat{\tau}_i^{(k)}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})\hat{\mathbf{b}}_i^{(k)\top}\mathbf{Z}_i^\top \\ &\quad - \hat{\tau}_i^{(k)}\mathbf{Z}_i\hat{\mathbf{b}}_i^{(k)}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top + \hat{\tau}_i^{(k)}\mathbf{Z}_i\hat{\mathbf{b}}_i^{(k)}\hat{\mathbf{b}}_i^{(k)\top}\mathbf{Z}_i^\top \\ &\quad + \hat{\sigma}^{2(k)}\mathbf{Z}_i(\mathbf{Z}_i^\top\hat{\mathbf{C}}_i^{(k)-1}\mathbf{Z}_i + \hat{\Gamma}^{(k)-1})^{-1}\mathbf{Z}_i^\top \\ &= \hat{\tau}_i^{(k)}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\hat{\mathbf{b}}_i^{(k)})(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\hat{\mathbf{b}}_i^{(k)})^\top \\ &\quad + \hat{\sigma}^{2(k)}\mathbf{Z}_i(\mathbf{Z}_i^\top\hat{\mathbf{C}}_i^{(k)-1}\mathbf{Z}_i + \hat{\Gamma}^{(k)-1})^{-1}\mathbf{Z}_i^\top \\ \hat{\Psi}_{2i}^{(k)} &= E(\tau_i\mathbf{b}_i\mathbf{b}_i^\top | \mathbf{Y}_i, \hat{\boldsymbol{\alpha}}^{(k)}) \\ &= E(\tau_iE(\mathbf{b}_i\mathbf{b}_i^\top | \mathbf{Y}_i, \tau_i, \hat{\boldsymbol{\alpha}}^{(k)}) | \mathbf{Y}_i, \hat{\boldsymbol{\alpha}}^{(k)}) \\ &= E(\tau_iE(\mathbf{b}_i | \mathbf{Y}_i, \tau_i, \hat{\boldsymbol{\alpha}}^{(k)})E(\mathbf{b}_i^\top | \mathbf{Y}_i, \tau_i, \hat{\boldsymbol{\alpha}}^{(k)}) + \tau_i\text{cov}(\mathbf{b}_i | \mathbf{Y}_i, \tau_i, \hat{\boldsymbol{\alpha}}^{(k)}) | \mathbf{Y}_i, \hat{\boldsymbol{\alpha}}^{(k)}) \\ &= \hat{\tau}_i^{(k)}\hat{\mathbf{b}}_i^{(k)}\hat{\mathbf{b}}_i^{(k)\top} + \hat{\sigma}^{2(k)}(\mathbf{Z}_i^\top\hat{\mathbf{C}}_i^{(k)-1}\mathbf{Z}_i + \hat{\Gamma}^{(k)-1})^{-1}. \end{aligned}$$

Acknowledgements

The author are grateful to the Editors and two anonymous referees for their insightful comments and valuable suggestions, which led to substantial improvements in the presentation of this work. This research was partly supported by the National Science Council of Taiwan (Grant NO. NSC95-2118-M-005-001-MY2).

References

- Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society, Series B* **61**, 579-602.
- Azzalini A. and Capitanio A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t -distribution. *Journal of the Royal Statistical Society, Series B* **65**, 367-389.

-
- Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715-26.
- Barndorff-Nielsen, O. E. and Schou, G. (1973). On the reparameterization of autoregressive models by partial autocorrelations. *Journal of Multivariate Analysis* **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.
- Demidenko, E. (2004). *Mixed Models: Theory and Applications*. Wiley.
- Demidenko, E. and Stukel, T. (2002). Efficient estimation of general linear mixed effects models *Journal of Statistical Planning Inference* **104** 197-219.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussions). *Journal of the Royal Statistical Society, Series B* **39**, 1-38.
- Efron, B. and Tibshirani, R. (1986). Bootstrap method for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* **1**, 54-77.
- Gill, P.S. (2000). A robust mixed linear model analysis for longitudinal data. *Statistics and Medicine* **19**, 975-987.
- Hartley, H. O. and Rao, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* **54**, 93-108.
- Huggins, R. M. (1992). On the robust analysis of variance components models for pedigree data. *Australian Journal of Statistics* **35**, 43-57.
- Huggins, R. M. (1993). A robust approach to the analysis of repeated measures. *Biometrics* **49**, 715-720.
- Keramidas, E. M. and Lee, J. C. (1990). Forecasting technological substitutions with concurrent short time series. *Journal of the American Statistical Association* **85**, 625-632.
- Kotz, S, Nadarajah, S. (2004). *Multivariate t Distributions and Their Applications*. Cambridge University Press.
- Laird, N. M. and Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics* **38**, 963-974.
- Lange, K. L., Little, R. J. A. and Taylor, J. M. G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association* **84**, 881-896.
- Lee, J. C. (1988). Prediction and estimation of growth curve with special covariance structures. *Journal of the American Statistical Association* **83**, 432-440.
- Lee, J. C. and Geisser, S. (1972). Growth curve prediction. *Sankhya Series A* **34**, 393-412.

- Lee, J. C., Lin T. I., Lee, K. J. and Hsu, Y. L. (2005). Bayesian analysis of Box-Cox transformed linear mixed models with ARMA(p, q) dependence. *Journal of Statistical Planning Inference* **133**, 435-451.
- Lin, T. I. and Ho, J. C. (2008). A simplified approach to inverting the autocovariance matrix of a general ARMA(p, q) process. *Statistics & Probability Letters* **78**, 36-41.
- Lin, T. I. and Lee, J. C. (2003). On modelling data from degradation sample paths over time. *Australian & New Zealand Journal of Statistics* **45**, 257-270.
- Lin, T. I. and Lee, J. C. (2006). A robust approach to t linear mixed models applied to multiple sclerosis data. *Statistics in Medicine* **25**, 1397-1412.
- Lin, T. I. and Lee, J. C. (2007). Bayesian analysis of hierarchical linear mixed modeling using the multivariate t distribution, *Journal of Statistical Planning 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. *Statistics in Medicine* **27**, 1490-1507.
- Lindstrom, M. J. and Bates, D. M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated measures data. *Journal of the American Statistical Association* **83**, 432-440.
- Liu, C. H. and Rubin, D. B. (1994). The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. *Biometrika* **81**, 633-648.
- Liu, C. H. and Rubin, D. B. (1995). ML estimation of the t distribution using EM and its extensions, ECM and ECME. *Statistica Sinica* **5**, 19-39.
- Meng, X. L. and Rubin, D.B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **80**, 267-278.
- Miller, J. J. (1977). Asymptotic Properties of Maximum Likelihood Estimates in the Mixed Model of the Analysis of Variance. *The Annals of Statistics* **5** 746-762.
- Nadarajah, S., Kotz, S. (2005). Mathematical properties of the multivariate t distribution. *Acta Applicandae Mathematicae* **89**, 53-84.
- Pan, J. X. and Fang, K. T. (2002). *Growth Curve Models and Statistical Diagnostics*. Springer.
- Peel, D. and McLachlan, G. J. (2000). Robust mixture modelling using the t distribution. *Statistics and Computing* **10**, 339-348.
- Pinheiro, J. C., Liu, C. H. and Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics* **10**, 249-276.
- Rao, C. R. (1987). Prediction of future observations in growth curve models. *Statistical Science* **2**, 434-471.
- Richardson, A. M. (1997). Bounded influence estimation in the mixed linear model. *Journal of the American Statistical Association* **92**, 154-161.

-
- Richardson, A. M. and Welsh, A. H. (1995). Robust restricted maximum likelihood in mixed linear models. *Biometrics* **51**, 1429-1439.
- Ryggard, K. and Spang-Thomsen, M. (1997). Quantitation and Gompertzian analysis of tumor growth data. *Breast Cancer Research and Treatment* **46**, 303-312.
- Schafer, J. L. (1998). Some improved procedures for linear mixed models. Technical Report, Department of Statistics, The Pennsylvania State University.
- Zacks, S. (1971). *The Theory of Statistical Inference*. Wiley.
- Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate student- t error terms. *Journal of the American Statistical Association* **71**, 400-405.

Received January 21, 2008; accepted May 19, 2008.

Tsung-I Lin
Department of Applied Mathematics
National Chung Hsing University
Taichung 402, Taiwan
tilin@amath.nchu.edu.tw