



A robust approach to joint modeling of mean and scale covariance for longitudinal data

Tsung-I. Lin^{a,*}, Yun-Jen Wang^b

^aDepartment of Applied Mathematics and Institute of Statistics, National Chung Hsing University, Taichung 402, Taiwan

^bGraduate Institute of Finance, National Chiao Tung University, Hsinchu 300, Taiwan

ARTICLE INFO

Article history:

Received 7 November 2007

Received in revised form

4 November 2008

Accepted 12 February 2009

Available online 3 March 2009

Keywords:

Covariance structure

Maximum likelihood estimates

Reparameterization

Robustness

Outliers

Prediction

ABSTRACT

In this paper, we propose a multivariate t regression model with its mean and scale covariance modeled jointly for the analysis of longitudinal data. A modified Cholesky decomposition is adopted to factorize the dependence structure in terms of unconstrained autoregressive and scale innovation parameters. We present three distinct representations of the log-likelihood function of the model and study the associated properties. A computationally efficient Fisher scoring algorithm is developed for carrying out maximum likelihood estimation. The technique for the prediction of future responses in this context is also investigated. The implementation of the proposed methodology is illustrated through two real-life examples and extensive simulation studies.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

In recent years, the method of joint modeling of mean and covariance on the general linear model with multivariate normal errors, called the normal joint modeling model (NJMM) hereafter, was heuristically introduced by Pourahmadi (1999, 2000). The key advantages of such normal-error models include the convenience in statistical interpretation and computational ease in parameter estimation. Yet it still exists several weaknesses. For instance, the assumption of normality for the error terms may be questionable in many practical situations when atypical points exist or the underlying data exhibit thick tails. A number of authors in the literature have used a more thick-tailed distribution, like the multivariate t distribution, in place of normal for robust estimation of general linear models (Zellner, 1976; Lange et al., 1989; He et al., 2004). Robust estimation for linear mixed models using the multivariate t distribution has been studied by Welsh and Richardson (1997) and Pinheiro et al. (2001), among others.

Specifically, a p -dimensional random vector \mathbf{Y} is said to follow a multivariate t distribution with degrees of freedom (df) ν , mean vector $\boldsymbol{\mu}$ and scale covariance matrix $\boldsymbol{\Sigma}$ if its probability density function is

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

We shall use the notation $\mathbf{Y} \sim T_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ to denote that \mathbf{Y} follows the above distribution. The multivariate t distribution has attracted considerable attention over the past 20–30 years. It has been applied in a wide variety of research fields, see Kotz and Nadarajah (2004) and the references therein.

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

E-mail address: tilin@amath.nchu.edu.tw (T.-I. Lin).

In this paper, we extend Pourahmadi’s approach of joint mean–covariance parameterization to general linear models with the error term distributed according to a multivariate t distribution, also called the t joint modeling model (TJMM), as a robust approach to the analysis of longitudinal data.

Suppose that the repeated measurements of a continuous random variable are observed over time on each of m subjects. Let $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$ be the response vector for the i th subject measured at time points $\mathbf{t}_i = (t_{i1}, \dots, t_{in_i})^T$, which are allowed unevenly spaced, and let the associate covariates $\mathbf{X}_i = [\mathbf{x}_{i1} \dots \mathbf{x}_{ip}]$ be an $n_i \times p$ full rank design matrix.

The TJMM is defined as

$$\mathbf{Y}_i \sim T_{n_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \nu) \quad (i = 1, \dots, m), \tag{1}$$

where $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{in_i})^T = \mathbf{X}_i \boldsymbol{\beta}$ is the mean response vector for subject i . Moreover, to ensure positive definiteness of $\boldsymbol{\Sigma}_i = [\sigma_{ij}]$, we reparameterize it via the modified Cholesky decomposition as

$$\mathbf{L}_i \boldsymbol{\Sigma}_i \mathbf{L}_i^T = \mathbf{D}_i, \tag{2}$$

where $\mathbf{D}_i = \text{diag}\{\sigma_1^2, \dots, \sigma_{n_i}^2\}$ and $\mathbf{L}_i = [\ell_{jk}]$ is a unit lower triangular matrix with the (j, k) th entry being $-\phi_{jk}$. Obviously, $\boldsymbol{\Sigma}_i^{-1} = \mathbf{L}_i^T \mathbf{D}_i^{-1} \mathbf{L}_i$. The parameters ϕ_{jk} and σ_j^2 in \mathbf{L}_i and \mathbf{D}_i are referred to as the *autoregressive parameters* and *scale innovation variances* of $\boldsymbol{\Sigma}_i$, respectively. Note that such decomposition in (2) is unique and has several nice features. For a detailed discussion on the modified Cholesky decomposition, interested readers are referred to Pourahmadi (2001, Section 3.5).

Statistical interpretations for such reparameterization include (a) the below-diagonal entries of \mathbf{L}_i are the negatives of the autoregressive parameters, namely $-\phi_{jk}$, in

$$\hat{Y}_{ij} = \mu_{ij} + \sum_{k=1}^{j-1} \phi_{jk}(Y_{ik} - \mu_{ik}),$$

which is the linear least-squares predictor of Y_{ij} based on its predecessors; (b) the diagonal entries of \mathbf{D}_i are the scale innovation variances $\sigma_j^2 = c_v^{-1} \text{var}(Y_{ij} - \hat{Y}_{ij})$, where $c_v = \nu/(\nu - 2)$.

To make the dimension of unconstrained parameters ϕ_{jk} and $\log \sigma_j^2$ more parsimonious, we model them using covariates in the spirit of Pourahmadi (1999), namely, for $j = 1, \dots, n_i$ and $k = 1, \dots, j - 1$,

$$\phi_{jk} = \mathbf{z}_{jk}^T \boldsymbol{\gamma}, \quad \log \sigma_j^2 = \mathbf{w}_j^T \boldsymbol{\lambda}, \tag{3}$$

where \mathbf{z}_{jk} and \mathbf{w}_j are $d \times 1$ and $q \times 1$ covariate vectors, which can usually be determined in terms of polynomial of measurement time t_{ij} 's with degrees of $d - 1$ and $q - 1$, respectively, and $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ are $d \times 1$ and $q \times 1$ vectors of unknown parameters. Note that $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ are assumed to be common for all $\boldsymbol{\Sigma}_i$'s for exhibiting the same covariance structure. In other words, $\boldsymbol{\Sigma}_i$ depends on i only through its dimension $n_i \times n_i$.

The rest of the paper is organized as follows. In the next section, we present three distinct representations of the log-likelihood function of TJMM and describe a Fisher scoring algorithm for the implementation of ML estimation. Section 3 is devoted to addressing the prediction issue. For illustration purposes, two real examples are presented in Section 4 and extensive simulation results are reported in Section 5. Finally, some concluding remarks are briefly summarized in Section 6, and the technical derivations are sketched in Appendices.

2. Computational aspects of parameter estimation

2.1. The log-likelihood function

For notational convenience, let $\mathbf{r}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} = (r_{ij})_{j=1}^{n_i}$, $\Delta_i(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}) = \mathbf{r}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i$ and $n = \sum_{i=1}^m n_i$ be the total number of observations. Denote by $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \nu)$ the population model parameters vector, where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_d)$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_q)$. Given independent observations $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_m)$, the log-likelihood function of $\boldsymbol{\alpha}$ corresponding to TJMM can be written in three distinct representations as follows:

$$\begin{aligned} \ell(\boldsymbol{\alpha} | \mathbf{Y}) &= \sum_{i=1}^m \log \Gamma\left(\frac{\nu + n_i}{2}\right) - m \log \Gamma\left(\frac{\nu}{2}\right) - \frac{n}{2} \log(\pi\nu) - \frac{1}{2} \sum_{i=1}^m \log |\boldsymbol{\Sigma}_i| - \frac{1}{2} \sum_{i=1}^m (\nu + n_i) \log\left(1 + \frac{\Delta_i(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})}{\nu}\right) \\ &= \sum_{i=1}^m \log \Gamma\left(\frac{\nu + n_i}{2}\right) - m \log \Gamma\left(\frac{\nu}{2}\right) - \frac{n}{2} \log(\pi\nu) - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^{n_i} \log \sigma_j^2 - \frac{1}{2} \sum_{i=1}^m (\nu + n_i) \log\left(1 + \frac{1}{\nu} \sum_{j=1}^{n_i} \eta_{ij}^2\right) \\ &= \sum_{i=1}^m \log \Gamma\left(\frac{\nu + n_i}{2}\right) - m \log \Gamma\left(\frac{\nu}{2}\right) - \frac{n}{2} \log(\pi\nu) - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^{n_i} \mathbf{w}_j^T \boldsymbol{\lambda} - \frac{1}{2} \sum_{i=1}^m (\nu + n_i) \log\left(1 + \frac{(\mathbf{r}_i - \mathbf{Z}_i \boldsymbol{\gamma})^T \mathbf{D}_i^{-1} (\mathbf{r}_i - \mathbf{Z}_i \boldsymbol{\gamma})}{\nu}\right), \end{aligned} \tag{4}$$

where $\hat{\mathbf{r}}_i = \{\hat{r}_{ij}\}_{j=1}^{n_i} = \{\sum_{k=1}^{j-1} r_{ik} \mathbf{z}_{jk}^T \boldsymbol{\gamma}\}_{j=1}^{n_i}$ is the predictor of r_{ij} based on its predecessors $r_{i,j-1}, \dots, r_{i,1}$, $\eta_{ij} = \sigma_j^{-1}(r_{ij} - \hat{r}_{ij})$ and \mathbf{Z}_i is an $n_i \times d$ matrix defined by

$$\mathbf{Z}_i = [\mathbf{z}(i, 1) \ \dots \ \mathbf{z}(i, n_i)]^T, \quad \mathbf{z}(i, j) = \sum_{k=1}^{j-1} r_{ik} \mathbf{z}_{jk}.$$

Following Pourahmadi (1999), it can readily be seen that $\mathbf{L}_i \mathbf{r}_i = \mathbf{r}_i - \hat{\mathbf{r}}_i$. Some useful properties concerning this consequence are summarized in the following proposition:

Proposition 1. With notations η_{ij} and $\Delta_i(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$ as defined in the previous paragraph, we have

- (a) $\eta_{ij} = \sigma_j^{-1}(r_{ij} - \hat{r}_{ij}) \sim t(v)$ for $j = 1, \dots, n_i$.
- (b) $n_i^{-1} \Delta_i(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}) = n_i^{-1} \sum_{j=1}^{n_i} \eta_{ij}^2 \sim F(n_i, v)$.
- (c) $B = (1 + v^{-1} \sum_{j=1}^{n_i} \eta_{ij}^2)^{-1} \sim \text{Beta}(v/2, n_i/2)$.

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

2.2. The scoring algorithm for maximum likelihood estimation

In light of (4), there is no closed-form solution available for the ML estimates and certain numerical techniques such as the iterative Newton–Raphson or Fisher scoring algorithms are used instead in order to find optimal parameter estimates. Explicit expressions of the score vector and the Fisher information required for the developed Fisher scoring algorithm are shown in this subsection. Taking the first derivative of (4) with respect to $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \boldsymbol{\lambda}, v)$, a $(p + d + q + 1) \times 1$ vector, leads to the score vector \mathbf{s}_α . Expressions for the elements of \mathbf{s}_α are

$$\begin{aligned} \mathbf{s}_\beta &= \sum_{i=1}^m \tau_i \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i, & \mathbf{s}_\gamma &= \sum_{i=1}^m \tau_i \mathbf{Z}_i^T \mathbf{D}_i^{-1} (\mathbf{r}_i - \mathbf{Z}_i \boldsymbol{\gamma}), \\ \mathbf{s}_\lambda &= -\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^{n_i} \mathbf{w}_j + \frac{1}{2} \sum_{i=1}^m \left(\tau_i \sum_{j=1}^{n_i} \eta_{ij}^2 \mathbf{w}_j \right), \\ s_v &= \frac{1}{2} \sum_{i=1}^m \left\{ \phi\left(\frac{v+n_i}{2}\right) - \phi\left(\frac{v}{2}\right) - \frac{n_i}{v} - \log\left(1 + \frac{\Delta_i(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\gamma})}{v}\right) + \frac{\tau_i}{v} \Delta_i(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}) \right\}, \end{aligned} \tag{5}$$

where

$$\tau_i = \frac{v + n_i}{v + \Delta_i(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\gamma})} \tag{6}$$

and $\phi(x) = (d/dx) \log \Gamma(x)$ is the digamma function.

Computation of the Hessian matrix $\mathbf{H}_{\alpha\alpha}$, obtained by the second derivative of (4) with respect to $\boldsymbol{\alpha}$, is given in Appendix A. To obtain the elements of the Fisher information matrix

$$\mathbf{I}_{\alpha\alpha} = -E(\mathbf{H}_{\alpha\alpha}) = -E\left(\frac{\partial^2 \ell(\boldsymbol{\alpha}|\mathbf{Y})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^T}\right), \tag{7}$$

we summarize the necessary formulae for calculating the expectations of (A.1)–(A.10), the elements of Hessian matrix, in the following proposition.

Proposition 2. For the multivariate t model (1) with reparameterization (3), the following holds:

- (a) $E\left(\frac{1}{v + \Delta_i(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\gamma})}\right) = \frac{1}{v + n_i}$, (b) $E\left(\frac{\tau_i}{v + \Delta_i(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\gamma})}\right) = \frac{v + 2}{v(v + n_i + 2)}$,
- (c) $E(\tau_i^2 \mathbf{r}_i \mathbf{r}_i^T) = \frac{v + n_i}{v + n_i + 2} \boldsymbol{\Sigma}_i$, (d) $E\left(\tau_i \sum_{j=1}^{n_i} \eta_{ij}^2 \mathbf{w}_j \mathbf{w}_j^T\right) = \sum_{j=1}^{n_i} \mathbf{w}_j \mathbf{w}_j^T$,
- (e) $E\left(\tau_i^2 \left(\sum_{j=1}^{n_i} \eta_{ij}^2 \mathbf{w}_j\right) \left(\sum_{j=1}^{n_i} \eta_{ij}^2 \mathbf{w}_j^T\right)\right) = \frac{v + n_i}{v + n_i + 2} \left\{ \left(\sum_{j=1}^{n_i} \mathbf{w}_j\right) \left(\sum_{j=1}^{n_i} \mathbf{w}_j^T\right) + 2 \sum_{j=1}^{n_i} \mathbf{w}_j \mathbf{w}_j^T \right\}$,

$$(f) E(\tau_i \mathbf{Z}_i^T \mathbf{D}_i^{-1} \mathbf{Z}_i) = \sum_{j=1}^{n_i} \sigma_j^{-2} \boldsymbol{\kappa}_j \boldsymbol{\Sigma}_i \boldsymbol{\kappa}_j^T, \quad (g) E(\tau_i \mathbf{z}(i, j) r_{ij}) = \sum_{k=1}^{j-1} \sigma_{kj} \mathbf{z}_{jk},$$

$$(h) E(\tau_i \mathbf{Z}_i^T \mathbf{D}_i^{-1} \boldsymbol{\Psi}_i) = \sum_{j=1}^{n_i} \sigma_j^{-2} \left(\sum_{k=1}^{j-1} \sigma_{kj} \mathbf{z}_{jk} \mathbf{w}_j^T - \boldsymbol{\kappa}_j \boldsymbol{\Sigma}_i \boldsymbol{\kappa}_j^T \boldsymbol{\gamma} \mathbf{w}_j^T \right),$$

where $\boldsymbol{\kappa}_j = [\mathbf{z}_{j1} \cdots \mathbf{z}_{jj-1} \mathbf{0} \cdots \mathbf{0}]$ is a $d \times n_i$ matrix and

$$\boldsymbol{\Psi}_i = [\mathbf{w}_1(r_{i1} - \boldsymbol{\gamma}^T \mathbf{z}(i, 1)) \cdots \mathbf{w}_{n_i}(r_{in_i} - \boldsymbol{\gamma}^T \mathbf{z}(i, n_i))]^T. \tag{8}$$

The detailed proof is sketched in Appendix B. Applying Proposition 2, the exact expressions for the elements of (7) are given by

$$\mathbf{I}_{\beta\beta} = \sum_{i=1}^m \frac{v + n_i}{v + n_i + 2} \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i, \quad \mathbf{I}_{\beta\gamma} = \mathbf{0}_{p \times d}, \quad \mathbf{I}_{\beta\lambda} = \mathbf{0}_{p \times q}, \quad \mathbf{I}_{\beta v} = \mathbf{0}_{p \times 1},$$

$$\mathbf{I}_{\gamma\gamma} = \sum_{i=1}^m \frac{v + n_i}{v + n_i + 2} \left\{ \sum_{j=1}^{n_i} \sigma_j^{-2} \left(\sum_{k=1}^{j-1} \sum_{\ell=1}^{j-1} \sigma_{k\ell} \mathbf{z}_{jk} \mathbf{z}_{j\ell}^T \right) \right\},$$

$$\mathbf{I}_{\gamma\lambda} = \sum_{i=1}^m \frac{v + n_i}{v + n_i + 2} \left\{ \sum_{j=1}^{n_i} \sigma_j^{-2} \left(\sum_{k=1}^{j-1} \sigma_{kj} \mathbf{z}_{jk} \mathbf{w}_j^T - \sum_{k=1}^{j-1} \sum_{\ell=1}^{j-1} \sigma_{k\ell} \mathbf{z}_{jk} \mathbf{z}_{j\ell}^T \boldsymbol{\gamma} \mathbf{w}_j^T \right) \right\},$$

$$\mathbf{I}_{\gamma v} = \mathbf{0}_{d \times 1},$$

$$\mathbf{I}_{\lambda\lambda} = \frac{1}{2} \sum_{i=1}^m \frac{v + n_i}{v + n_i + 2} \left\{ \sum_{j=1}^{n_i} \mathbf{w}_j \mathbf{w}_j^T - \frac{(\sum_{j=1}^{n_i} \mathbf{w}_j)(\sum_{j=1}^{n_i} \mathbf{w}_j^T)}{v + n_i} \right\},$$

$$\mathbf{I}_{\lambda v} = - \sum_{i=1}^m \left\{ \frac{1}{(v + n_i)(v + n_i + 2)} \sum_{j=1}^{n_i} \mathbf{w}_j \right\},$$

$$\mathbf{I}_{vv} = \frac{1}{4} \sum_{i=1}^m \left\{ \psi\left(\frac{v}{2}\right) - \psi\left(\frac{v + n_i}{2}\right) - \frac{2n_i(v + n_i + 4)}{v(v + n_i)(v + n_i + 2)} \right\},$$

where $\psi(x) = (d^2/dx^2) \log \Gamma(x)$ stands for the trigamma function.

Let $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \boldsymbol{\lambda}, v)$ denote the set of parameters excluding regression coefficients $\boldsymbol{\beta}$, so that $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \boldsymbol{\theta})$. Owing to the fact that $\mathbf{I}_{\beta\theta} = \mathbf{0}$, the Fisher information matrix (7) has a block-diagonal form

$$\mathbf{I}_{\alpha\alpha} = \begin{bmatrix} \mathbf{I}_{\beta\beta} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\theta\theta} \end{bmatrix} \quad \text{where } \mathbf{I}_{\theta\theta} = \begin{bmatrix} \mathbf{I}_{\gamma\gamma} & \mathbf{I}_{\gamma\lambda} & \mathbf{0} \\ \mathbf{I}_{\gamma\lambda}^T & \mathbf{I}_{\lambda\lambda} & \mathbf{I}_{\lambda v} \\ \mathbf{0} & \mathbf{I}_{\lambda v}^T & I_{vv} \end{bmatrix}. \tag{9}$$

Furthermore, with regard to $\mathbf{I}_{\gamma v} = \mathbf{0}$, this implies the ML estimates of $\boldsymbol{\gamma}$ and v are asymptotically uncorrelated.

Thus one can proceed the scoring method to calculate the ML estimates of parameters through the following steps:

Step 1: Choose a starting value $\hat{\boldsymbol{\alpha}}^{(0)} = (\hat{\boldsymbol{\beta}}^{(0)}, \hat{\boldsymbol{\gamma}}^{(0)}, \hat{\boldsymbol{\lambda}}^{(0)}, \hat{v}^{(0)})$ and form the scale covariance matrix $\hat{\boldsymbol{\Sigma}}_i^{(0)} = [\hat{\sigma}_{ij}^{(0)}]$ using the modified Cholesky decomposition for $i = 1, \dots, m$.

Step 2: With the current estimates $\hat{\boldsymbol{\beta}}^{(h)}$ and $\hat{\boldsymbol{\theta}}^{(h)} = (\hat{\boldsymbol{\gamma}}^{(h)}, \hat{\boldsymbol{\lambda}}^{(h)}, \hat{v}^{(h)})$ at the h th iteration, update the value $\hat{\boldsymbol{\theta}}^{(h)}$ using the scoring procedure

$$\hat{\boldsymbol{\theta}}^{(h+1)} = \hat{\boldsymbol{\theta}}^{(h)} + \hat{\mathbf{I}}_{\theta\theta}^{(h-1)} \hat{\mathbf{s}}_{\theta}^{(h)},$$

where $\hat{\mathbf{s}}_{\theta}^{(h)} = (\hat{\mathbf{s}}_{\boldsymbol{\gamma}}^{(h)}, \hat{\mathbf{s}}_{\boldsymbol{\lambda}}^{(h)}, \hat{\mathbf{s}}_v^{(h)})$ and $\hat{\mathbf{I}}_{\theta\theta}^{(h)}$ are $\mathbf{s}_{\boldsymbol{\gamma}}, \mathbf{s}_{\boldsymbol{\lambda}}$ and s_v in (5) and $\mathbf{I}_{\theta\theta}$ in (9) evaluated at $\hat{\boldsymbol{\alpha}}^{(h)} = (\hat{\boldsymbol{\beta}}^{(h)}, \hat{\boldsymbol{\theta}}^{(h)})$. The scale covariance matrix can be computed by

$$\hat{\boldsymbol{\Sigma}}_i^{(h+1)-1} = \mathbf{L}_i(\hat{\boldsymbol{\gamma}}^{(h+1)})^T \mathbf{D}_i^{-1}(\hat{\boldsymbol{\lambda}}^{(h+1)}) \mathbf{L}_i(\hat{\boldsymbol{\gamma}}^{(h+1)}).$$

Step 3: The value of $\hat{\beta}^{(h+1)}$ can be obtained by a generalized-least-squares step as

$$\hat{\beta}^{(h+1)} = \left(\sum_{i=1}^m \hat{\tau}_i^{(h+1)} \mathbf{X}_i^T \hat{\Sigma}_i^{(h+1)-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^m \hat{\tau}_i^{(h+1)} \mathbf{X}_i^T \hat{\Sigma}_i^{(h+1)-1} \mathbf{Y}_i \right),$$

where $\hat{\tau}_i^{(h+1)}$ is τ_i in (6) with β, γ, λ and v replaced by $\hat{\beta}^{(h)}, \hat{\gamma}^{(h+1)}, \hat{\lambda}^{(h+1)}$ and $\hat{v}^{(h+1)}$, respectively.

Iterative optimization algorithm such as the scoring method must select good starting values $\hat{\alpha}^{(0)}$. For the starting values of β, γ and λ , a convenient choice is to use the ML estimates obtained by fitting the unbalanced joint mean–covariance normal regression model as described in Pan and MacKenzie (2003). As for v , it is general to use a relative large starting value, say $\hat{v}^{(0)} = 50$, which corresponds to an initial assumption of near-normality for the error terms. The iterations are repeated until a suitable convergence rule is satisfied, e.g., $\|\hat{\alpha}^{(h+1)} - \hat{\alpha}^{(h)}\|/\|\hat{\alpha}^{(h)}\|$ is smaller than the specified tolerance.

After convergence of successive iterations, the converged solutions are the ML estimates, denoted by $\hat{\alpha} = (\hat{\beta}, \hat{\lambda}, \hat{\gamma}, \hat{v})$. Under suitable regularity conditions (Zacks, 1971), the ML estimator $\hat{\alpha}$ is asymptotically normally distributed. Because of $\mathbf{I}_{\beta\theta} = \mathbf{0}$ and $\mathbf{I}_{\gamma v} = \mathbf{0}$, it follows immediately that $\hat{\beta}$ and $\hat{\theta} = (\hat{\gamma}, \hat{\lambda}, \hat{v})$ as well as $\hat{\gamma}$ and \hat{v} are asymptotically independent. In other words, the asymptotic confidence regions and hypothesis tests for the regression coefficients β as well as the variance components θ can be obtained by assuming that the ML estimators $\hat{\beta}$ and $\hat{\theta}$ have approximate $N_p(\beta, \mathbf{V}_\beta)$ and $N_{d+q+1}(\theta, \mathbf{V}_\theta)$ distributions, respectively. In practice, the asymptotic covariance matrix \mathbf{V}_β and \mathbf{V}_θ are unknown and they can be empirically estimated by substituting ML estimates into $\mathbf{I}_{\beta\beta}^{-1}$ and $\mathbf{I}_{\theta\theta}^{-1}$, respectively.

3. Prediction

In this section we are interested in deriving the computing formulae for predicting the future values of \mathbf{Y}_i , say the i th individual, which is one of the m individuals under consideration. Let $\tilde{\mathbf{y}}_i$ be a future $g \times 1$ vector of measurements of \mathbf{Y}_i . We further assume \mathbf{Y}_i and $\tilde{\mathbf{y}}_i$ jointly follows an $(n_i + g)$ -variate t distribution. Let $\tilde{\mathbf{x}}_i$ be a $g \times p$ matrix of prediction regressors corresponding to $\tilde{\mathbf{y}}_i$. That is, $E(\tilde{\mathbf{y}}_i) = \tilde{\mathbf{x}}_i \beta$. Let $\tilde{\mathbf{z}}_{jk}$ and $\tilde{\mathbf{w}}_j$, respectively, denote a $d \times 1$ and a $q \times 1$ vector such that $\phi_{jk} = \tilde{\mathbf{z}}_{jk}^T \gamma$, and $\log \sigma_j^2 = \tilde{\mathbf{w}}_j^T \lambda$ for $j = n_i + 1, \dots, n_i + g$ and $k = 1, \dots, j - 1$.

In the above setting, we then have

$$\begin{bmatrix} \mathbf{Y}_i \\ \tilde{\mathbf{y}}_i \end{bmatrix} \sim T_{n_i+g}(\mathbf{X}_i^* \beta, \Sigma_i^*, v)$$

with $\mathbf{X}_i^* = (\mathbf{X}_i^T, \tilde{\mathbf{x}}_i^T)^T$ and $\Sigma_i^{*-1} = \mathbf{L}_i^{*T} \mathbf{D}_i^{*-1} \mathbf{L}_i^*$, where \mathbf{L}_i^* is an $(n_i + g) \times (n_i + g)$ lower triangular matrix with $-\phi_{jk}$ at the (j, k) th position, $k < j, j = 1, \dots, n_i + g$, and \mathbf{D}_i^* is a diagonal matrix with σ_j^2 as its diagonal entry, i.e.,

$$\mathbf{L}_i^* = \begin{bmatrix} 1 & & & & & & & \\ -\mathbf{z}_{21}^T \gamma & 1 & & & & & & \mathbf{0} \\ \vdots & \vdots & \ddots & & & & & \\ -\mathbf{z}_{n_i,1}^T \gamma & \dots & -\mathbf{z}_{n_i, n_i-1}^T \gamma & & & & & 1 \\ -\tilde{\mathbf{z}}_{n_i+1,1}^T \gamma & \ddots & \dots & & -\tilde{\mathbf{z}}_{n_i+1, n_i}^T \gamma & & & 1 \\ \vdots & \ddots & \dots & & \dots & \ddots & & \ddots \\ -\tilde{\mathbf{z}}_{n_i+g,1}^T \gamma & \vdots & \dots & & \dots & & & -\tilde{\mathbf{z}}_{n_i+g, n_i+g-1}^T \gamma 1 \end{bmatrix}$$

and

$$\mathbf{D}_i^* = \text{diag}\{\mathbf{w}_1^T \lambda, \dots, \mathbf{w}_{n_i}^T \lambda, \tilde{\mathbf{w}}_{n_i+1}^T \lambda, \dots, \tilde{\mathbf{w}}_{n_i+g}^T \lambda\}.$$

Let Σ_i^* be partitioned conformably with the dimension of $\mathbf{Y}_i^* = (\mathbf{Y}_i^T, \tilde{\mathbf{y}}_i^T)^T$. Therefore, we have

$$\Sigma_i^* = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}. \tag{10}$$

Here we suppress the subscript i of the partitioned matrices in (10) for notational convenience. Note that $\Sigma_{11} = \Sigma_i$ and $\Sigma_{21} = \Sigma_{12}^T$. Making use of the conditional property concerning the multivariate t distribution, we have

$$\tilde{\mathbf{y}}_i | \mathbf{Y}_i, \alpha \sim T_g \left(\mu_{2,1}, \frac{v + A_i(\beta, \gamma, \lambda)}{v + n_i} \Sigma_{22,1}, v + n_i \right), \tag{11}$$

where $\mu_{2,1} = E(\mathbf{y}_i | \mathbf{Y}_i, \alpha)$ and $\Sigma_{22,1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$.

From (11), the minimum mean-squared error (MSE) predictor of $\tilde{\mathbf{y}}$ is

$$\hat{\mathbf{y}}_i(\boldsymbol{\alpha}) = \boldsymbol{\mu}_{2,1} = \tilde{\mathbf{x}}_i\boldsymbol{\beta} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}). \tag{12}$$

Consequently, the MSE covariance matrix for the predictor (12) is determined by

$$\begin{aligned} E(\hat{\mathbf{y}}_i(\boldsymbol{\alpha}) - \tilde{\mathbf{y}}_i)(\hat{\mathbf{y}}_i(\boldsymbol{\alpha}) - \tilde{\mathbf{y}}_i)^T &= E(\text{cov}(\tilde{\mathbf{y}}_i|\mathbf{Y}_i)) \\ &= E\left(\frac{v + A_i(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})}{v + n_i - 2}\boldsymbol{\Sigma}_{22,1}\right) \\ &= \frac{v}{v - 2}\boldsymbol{\Sigma}_{22,1}. \end{aligned} \tag{13}$$

Typically, the predicted value of $\tilde{\mathbf{y}}_i$ can be treated by substituting the ML estimate $\hat{\boldsymbol{\alpha}}$ into (12), leading to $\hat{\mathbf{y}} = \hat{\mathbf{y}}(\hat{\boldsymbol{\alpha}})$. Meanwhile, it should be noted that the variance estimate of (13) does not account for errors due to estimation of unknown $\boldsymbol{\alpha}$ and hence it will underestimate the actual variance for smaller sample sizes.

4. Numerical illustrations

4.1. The tumor growth data

We apply the results developed in Sections 2 and 3 to the *in vivo* growth of lung tumor for the control group of 22 xenografted nude mice, which is originally reported by Ryggard and Spang-Thomsen (1997) and subsequently analyzed by Demidenko (2004, Chapter 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).

Fig. 1 depicts the logarithm of tumor growth volumes over an unequally spaced 28-day period for the 22 mice together with the associated *sample regressograms* for the generalized autoregressive parameters and sample innovation variances. The profile plot in Fig. 1(a) reveals that the mean level exhibits a linear growth trend over time. The design matrix for the mean response for the 22 mice takes the form

$$\mathbf{X}_i = [\mathbf{1} \ \mathbf{k}], \quad \mathbf{1} = (1, 1, \dots, 1)^T, \quad \mathbf{k} = (0, 1, 2.5, 3.5, 4.5, 6, 7, 8, 10, 11.5, 13, 14)^T, \quad i = 1, \dots, 22.$$

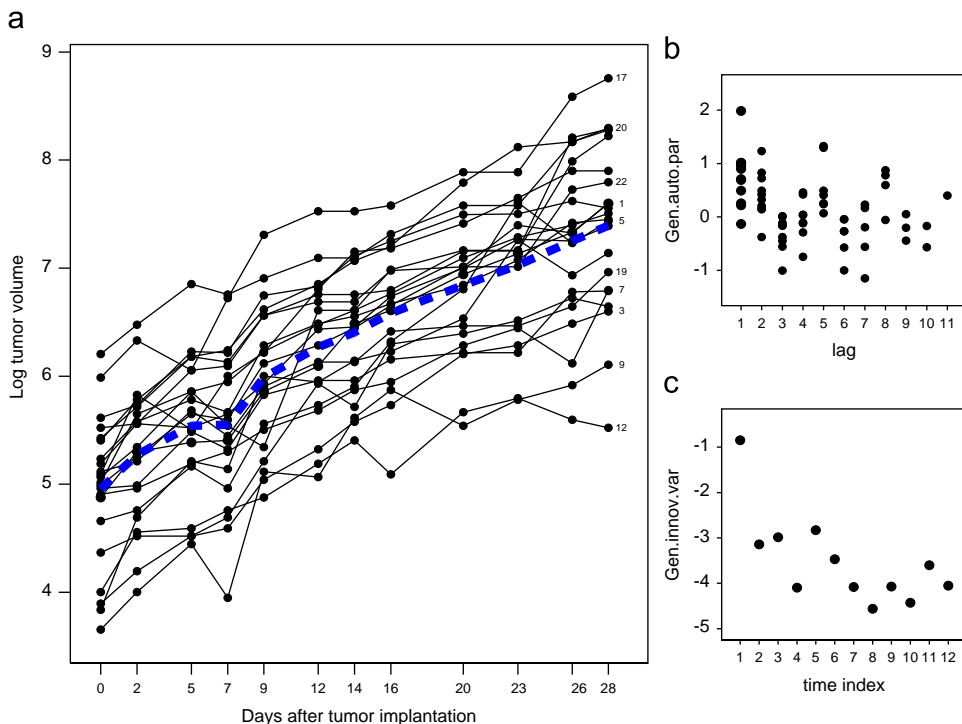


Fig. 1. (a) Trajectories of tumor growths for 22 xenografted nude mice; (b) sample generalized autoregressive parameters; (c) sample log-innovation variances.

Table 1
Comparison of $\hat{\ell}_{\max}$, number of parameters, and BIC values for some Poly(d, q) choices of NJMM and TJMM.

(d, q)	$\hat{\ell}_{\max}$		Number of parameters		BIC	
	NJMM	TJMM	NJMM	TJMM	NJMM	TJMM
(2,5)	-25.1867	-15.9188	11	12	3.8352	3.1332
(2,6)	-24.7836	-15.1324	12	13	3.9391	3.2022
(3,3)	-23.0706	-13.3374	10	11	3.5023	2.7580
(3,4)	-23.0593	-13.3187	11	12	3.6418	2.8968
(3,5) ^b	-8.9331	-1.3094	12	13	2.4981	1.9456
(3,6)	-8.9254	-1.3092	13	14	2.6379	2.0860
(4,3)	-22.9333	-12.9712	11	12	3.6304	2.8652
(4,4)	-22.9260	-12.9443	12	13	3.7702	3.0033
(4,5) ^a	-6.0243	1.7060	13	14	2.3742	1.8119
(4,6)	-6.0180	1.7060	14	15	2.5141	1.9524
(5,3)	-22.9331	-12.9345	12	13	3.7709	3.0024
(5,4)	-22.9257	-12.9066	13	14	3.9107	3.1404
(5,5)	-6.0200	1.7065	14	15	2.5143	1.9524
(5,6)	-6.0102	1.7065	15	16	2.6539	2.0929

^aPoly(4,5) is the best model.
^bPoly(3,5) is the second-best model.

Table 2
Parameter estimates for the best two Poly(d, q) choices of NJMM and TJMM.

	Poly(4,5)				Poly(3,5)			
	NJMM		TJMM		NJMM		TJMM	
	MLE	SE	MLE	SE	MLE	SE	MLE	SE
$\hat{\beta}_0$	5.2960	0.1071	5.3535	0.1062	5.2575	0.0944	5.3145	0.0917
$\hat{\beta}_1$	0.1628	0.0071	0.1617	0.0066	0.1628	0.0073	0.1615	0.0067
$\hat{\gamma}_0$	1.5646	0.1626	1.5787	0.1720	1.1649	0.0827	1.1745	0.0882
$\hat{\gamma}_1$	-1.0816	0.1806	-1.0928	0.1899	-0.6121	0.0637	-0.6170	0.0678
$\hat{\gamma}_2$	0.2555	0.0594	0.2587	0.0622	0.0970	0.0127	0.0975	0.0135
$\hat{\gamma}_3$	-0.0248	0.0075	-0.0253	0.0078	-0.0047	0.0007	-0.0047	0.0008
$\hat{\gamma}_4$	0.0009	0.0003	0.0009	0.0003	-	-	-	-
$\hat{\lambda}_0$	6.8278	1.3174	6.0870	1.3882	6.1052	1.3174	5.3054	1.3880
$\hat{\lambda}_1$	-11.3551	1.7462	-10.5501	1.8328	-10.2234	1.7462	-9.3982	1.8326
$\hat{\lambda}_2$	4.7445	0.7534	4.4029	0.7908	4.2148	0.7534	3.8697	0.7907
$\hat{\lambda}_3$	-0.8808	0.1402	-0.8241	0.1472	-0.7779	0.1402	-0.7199	0.1471
$\hat{\lambda}_4$	0.0737	0.0117	0.0695	0.0123	0.0649	0.0117	0.0606	0.0123
$\hat{\lambda}_5$	-0.0023	0.0004	-0.0022	0.0004	-0.0020	0.0004	-0.0019	0.0004
$\hat{\nu}$	-	-	7.6832	3.4450	-	-	7.7313	3.4756

Furthermore, Figs. 1(b) and (c) suggest both ϕ_{jk} 's and $\log \sigma_j^2$'s could be well suited to polynomial functions in lags of degree greater than or equal to three. As in Pourahmadi (2000), we use the Poly(d, q) as a shorthand for imposing two distinct polynomials of lagged times $j - k$ and j with degrees d and q for ϕ_{jk} and $\log \sigma_j^2$, respectively. Specifically, the covariates \mathbf{z}_{jk} and \mathbf{w}_j are chosen as

$$\mathbf{z}_{jk} = (1, (j - k), (j - k)^2, \dots, (j - k)^d)^T,$$

$$\mathbf{w}_j = (1, j, j^2, \dots, j^q)^T, \quad j = 1 \dots, 12; \quad k = 1, \dots, j - 1.$$

To identify the optimal degrees of (d, q), we use the BIC-based model selection criterion. The best pair is said to be $(d^*, q^*) = \text{argmin}_{(d,q)} \{BIC(d, q)\}$ with d and q lying within the range 0 to n^* , where $n^* = \max_{1 \leq i \leq m} n_i$,

$$BIC(d, q) = -\frac{2}{m} \hat{\ell}_{\max}(d, q) + \frac{p + d + q + 3}{m} \log(m),$$

and $\hat{\ell}_{\max}(d, q)$ is the maximized log-likelihood with respect to Poly(d, q).

For comparison purposes, we fit the tumor data by using NJMM and TJMM under various possible choices of Poly(d, q). A stopping criterion used for the scoring algorithm was performed to terminate iterations when $\|\hat{\boldsymbol{\alpha}}^{(h+1)} - \hat{\boldsymbol{\alpha}}^{(h)}\| / \|\hat{\boldsymbol{\alpha}}^{(h)}\| < 10^{-8}$ has met. The values of $\hat{\ell}_{\max}$, together with the corresponding number of parameters and BIC values for selected pairs (d, q) with better fittings are listed in Table 1. Judging from the BIC values, Poly(4, 5) and Poly(3, 5) are the best two choices for both NJMM and TJMM. Table 2 shows the ML estimates and the associated standard errors for the best two fitting NJMM and TJMM. As seen in the table, it is noteworthy that the estimates of the df for the two fitted TJMM are somewhat small, reflecting the presence of longer-than-normal errors.

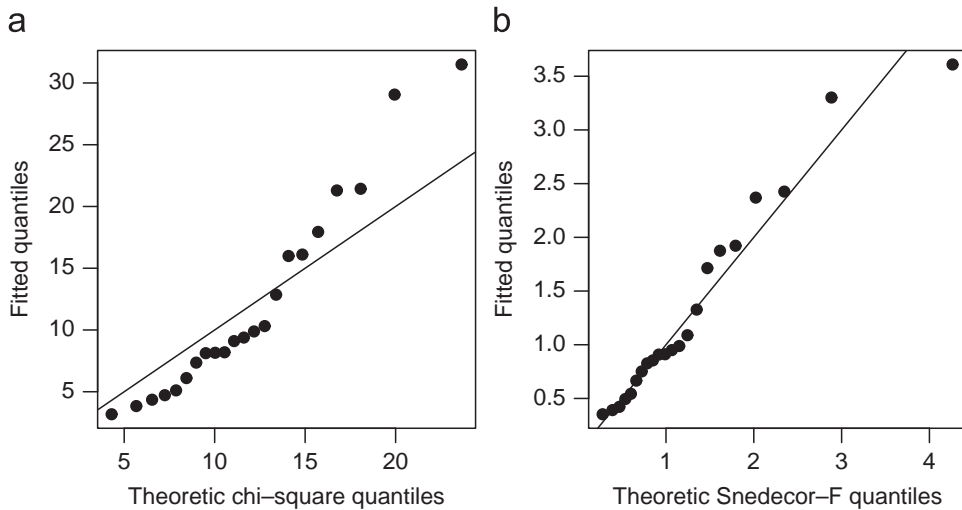


Fig. 2. Healy plot when either (a) an NJMM-Poly(4,5) model or (b) a TJMM-Poly(4,5) model is fitted to the tumor growth data.

We consider diagnostics to assess the validity of the underlying assumptions for the error term. For NJMM, a formal measure for judging the joint normality of residuals is the Mahalanobis-like distance $\Delta_i(\hat{\beta}, \hat{\gamma}, \hat{\lambda})$, which has an asymptotic chi-square distribution with n_i df. Checking the normality can be achieved by constructing Healy's (1968) plot (or the chi-square plot), which is a well-known diagnostic tool widely used in the context of multivariate normal theory.

As for TJMM, it is known from Proposition 1(b) that $n_i^{-1} \Delta_i(\hat{\beta}, \hat{\gamma}, \hat{\lambda})$ follows the F distribution with n_i and ν df. To assess the fitness of TJMM, one can construct another Healy-type plot (or the Snedecor-F plot) by plotting the ordered F statistics against the quantiles of $F(n_i, \nu)$ distribution for nominal values $(i - 0.5)/m, i = 1, \dots, m$. To compare NJMM and TJMM fits with graphical diagnostics, one can examine whether the corresponding Healy's plot resembles a straight line through the origin having unit slope. In other words, the scatter curve with serious departure from the 45° line indicates a poor fitting of the model.

Fig. 2 displays Healy's plots for Poly(4,5) of NJMM and TJMM. A comparison of these two plots indicates the t model tracks the identity line more closely than the normal counterpart, revealing a substantial improvement provided by the use of multivariate t distribution. Note that Healy's plots for Poly(3,5) of NJMM and TJMM are similar and thus are not shown in the present paper.

One essential characteristic of longitudinal data is that measurements within each experimental unit are collected repeatedly over time, so they typically exhibit serial correlations. Therefore, the predictive accuracy of future observations (Rao, 1987) can be treated as an alternative measure of fitness of the data. Now, we are interested in comparing the predictive performance of the proposed TJMM with that of t linear mixed models (TLMM), a generalization of Laird and Ware's (1982) linear mixed models with the random effects and within-subject errors distributed as a multivariate t distribution. We conduct a preliminary analysis to select appropriate TLMM for the tumor growth curve data. Results suggest that TLMM with random intercepts and slopes plus AR(1) dependence ($\hat{\nu} = 9.06, \text{BIC} = 2.84$), denoted by TLMM-RIS-AR(1), is the most preferred choice among some selected normal/ t mixed models. The prediction approach for TLMM with autoregressive dependence was described in Lin and Lee (2006).

For ease of comparison, we use the pseudo-cross-validation approach to assess the relative predictive performance of TJMM-Poly(4,5) and TLMM-RIS-AR(1). This approach of comparing forecasts with the corresponding actual values is in spirit to holdout cross-validation of Stone (1974) and predictive sample reuse of Geisser (1975). 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) prediction of the q true measurements $\mathbf{y}_i = (y_{i,12-q+1}, \dots, y_{i,12})^T$, denoted by $\hat{\mathbf{y}}_i = (\hat{y}_{i,12-q+1}, \dots, \hat{y}_{i,12})^T$, is made by using formula (12). The procedure is repeated across m subjects.

As a measure of predictive accuracies we consider the empirical mean squared forecast error (MSFE), defined as

$$\text{MSFE} = \frac{1}{mq} \sum_{i=1}^m (\hat{\mathbf{y}}_i - \mathbf{y}_i)^T (\hat{\mathbf{y}}_i - \mathbf{y}_i).$$

The prediction results for the first six step-ahead forecasts are shown in Table 3. It appears that the predictive performance of the TJMM predictor is substantially better than that of TLMM with relative improvement percentages (RIPs) ranging from 12.3% to 24.2%.

4.2. The orthodontic growth data

This example, originally analyzed by Potthoff and Roy (1964), studied the dental growth of distances (in millimeters) from the center of the pituitary to the pterygomaxillary fissure measured repeatedly at 8, 10, 12, and 14 years for 11 girls and 16 boys.

Table 3
Comparison of predictive accuracies in terms of MSFE for two robust *t* models.

Model	<i>q</i> step-ahead forecasts					
	<i>q</i> = 1	<i>q</i> = 2	<i>q</i> = 3	<i>q</i> = 4	<i>q</i> = 5	<i>q</i> = 6
TLMM	0.0372	0.1181	0.1021	0.1213	0.1158	0.1112
TJMM	0.0313	0.0971	0.0774	0.0991	0.0982	0.0975
RIP (%)	16.1	17.8	24.2	18.3	15.2	12.3

RIP means the relative improvement percentage (rate of MSFE improvement) of the TJMM predictor over the TLMM predictor.

Table 4
Parameter estimates of the Poly(1,1) model for the orthodontic growth data.

Parameter	NJMM		TJMM	
	MLE	SE	MLE	SE
$\hat{\beta}_0$	16.0707	0.9829	16.5863	0.6851
$\hat{\delta}_0$	1.3198	1.5398	0.9819	0.9689
$\hat{\beta}_1$	0.8122	0.0839	0.7713	0.0601
$\hat{\delta}_1$	-0.3341	0.1314	-0.2999	0.0850
$\hat{\gamma}_0$	0.7337	0.1653	1.0051	0.1243
$\hat{\gamma}_1$	-0.2188	0.0890	-0.3551	0.0679
$\hat{\lambda}_0$	1.8898	0.3333	1.5295	0.2853
$\hat{\lambda}_1$	-0.3145	0.1217	-0.3612	0.0947
$\hat{\nu}$	-	-	5.5165	2.0384
$\hat{\ell}_{\max}$	-212.8414	-	-205.4788	-
BIC	16.7426	-	16.3193	-

Pinheiro et al. (2001) applied TLMM to this data set and investigated the robustness of the *t* distribution via simulations. As in the preceding subsection, we compare the ML results under NJMM and TJMM for this example. Similar to Pinheiro et al. (2001), we also illustrate the robustness of the proposed model by using a single contamination on this data set to see the influence for the estimated parameters.

Assuming a linear growth function for describing the mean response, we have

$$\mu_{ij} = \beta_0 + \delta_0 I_i(F) + (\beta_1 + \delta_1 I_i(F)) t_i, \quad i = 1, \dots, 27, j = 1, \dots, 4,$$

where $\mu_{ij} = E(Y_{ij})$ is the mean orthodontic distance for the *i*th subject at age *t_j*; β_0 and β_1 denote the intercept and the slope for boys, respectively; δ_0 and δ_1 denote the difference in intercept and slope between girls and boys, respectively, and $I_i(F)$ is an indicator variable for the female.

Since there are only four measurements for each boy and each girl, we assume a parsimonious Poly(1,1) model for describing the dependence structure. The corresponding ML estimates along with standard errors, $\hat{\ell}_{\max}$ and BIC under the normal and *t* settings are shown in Table 4. In the table, there is strong evidence that the residuals exhibit a leptokurtic distribution ($\hat{\nu} = 5.52$). Furthermore, the *t* model has a smaller BIC value and all the standard errors of parameter estimates under the *t* model are much smaller than the corresponding normal model. From the above studies, it definitely indicates that the fitting results of TJMM are significantly superior to those obtained by using the NJMM in this example.

To further see the robustness of TJMM estimates against perturbation, it can be done by (a) adding a contaminated value ζ to a single observation y_{ij} , denoted by $y_{ij}(\zeta) = y_{ij} + \zeta$; (b) re-estimating the parameters twice under the Poly(1,1) assumption of NJMM and TJMM; (c) computing the relative change in the estimates, i.e., $\hat{\alpha}(\zeta)/\hat{\alpha} - 1$. Here $\hat{\alpha}$ denotes one of the original estimates in Table 4 and $\hat{\alpha}(\zeta)$ the corresponding estimate for the contaminated data. In the orthodontic growth example, we illustrate such procedure by adding a contaminated value to the last observation (age = 14 years) of the first girl with various ζ between -20 and 20 mm by increments of 2 mm. Fig. 3 displays the percentage changes in ML estimates under both models for different contaminations ζ . It is readily seen that the parameter estimates made by NJMM are highly affected by a single outlier, whereas the influence for TJMM is limited in a short range. This suggests TJMM provides a favorable way for achieving robust statistical inference.

5. Simulation studies

We first undertook a simulation study to examine the performance and flexibility of the proposed TJMM approach. The two main objectives of the study are: (1) when data are simulated from normal distribution, whether the proposed model will be over-fitted? (2) when data are simulated from the *t* distribution, how bad can the normal models behave? For the sake of convenience, the true parameters are chosen to mimic the tumor data fitted with TJMM-Poly(3,3), except that the df is specified at two different settings. The ML estimates of parameters, treated as true parameters, are $\hat{\beta} = (5.27, 0.16)^T$, $\hat{\gamma} = (1.12, -0.58, 0.09, -0.0043)^T$ and

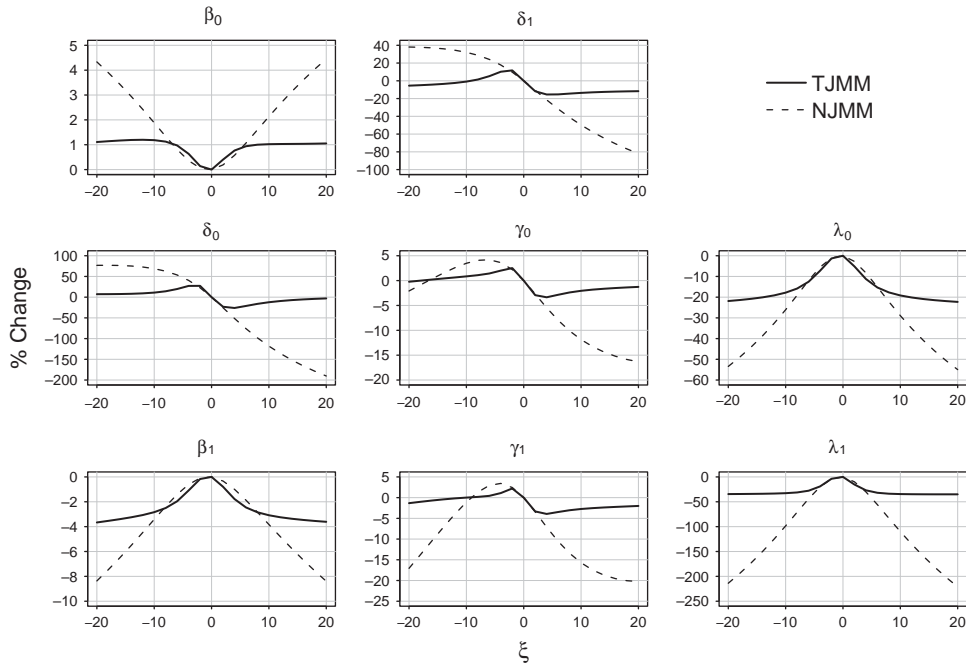


Fig. 3. Percentage of change in the ML estimates under the normal and *t* joint modeling models with varying contamination ξ of a single measurement.

$\hat{\lambda} = (-0.4, -1.37, 0.18, -0.007)^T$. For the values of df's used in the study, we take a low value ($\nu = 4$) for yielding a heavy-tailed distribution and a high value ($\nu = 50$) for approaching to normality. Note that the simulated longitudinal data are ‘unbalanced’ in the sense of unevenly spaced measurement times, although each subject has the same number of observations. We also note that if the df is set too high, say far more than 50, the estimate of df is hardly to converge and tends to infinity.

To examine the finite sample properties, the evaluated sample size ranges from small ($m = 25$) to relatively large ($m = 100$). Simulations were run with a total of 500 replications for each combination of ν and m and each simulated data set was fitted under TJMM and NJMM scenarios. The detailed numerical results, including the average ML estimates for the fixed effects (β_0, β_1), the autoregressive parameters ($\gamma_0, \gamma_1, \gamma_2, \gamma_3$) and the scale innovation variances ($\lambda_0, \lambda_1, \lambda_2, \lambda_3$), the average of maximized log-likelihood values ℓ_{\max} , the average of associated BIC values and the median estimates for the df, together with their standard errors in parentheses, are summarized in Table 5. Following Taylor and Verbyla (2004), due to the strong skewness of their Monte Carlo ML estimates, the median value was chosen as an appropriate estimator for ν in the table.

When the data are generated near normality ($\nu = 50$), both models produce similar estimates and standard errors, although the TJMM is slightly over-fitted because of large standard errors of ν 's and slightly smaller BIC. Overall, we may say both models are equally better. When the data are generated from the *t* distribution with heavy tails ($\nu = 4$), the TJMM leads to significantly smaller biases and standard errors as well as the BIC than those made by NJMM, reflecting its superiority for outlier resistance. Table 5 also suggests the bias and standard errors can be decreased as the sample size increases, confirming the consistence of estimators.

With the same data sets generated from the setting $\nu = 4$, we conduct a second simulation study to compare the modeling adequacy and prediction ability between TJMM and TLMM. The primary goal of this study is to investigate how much the TLMM will miss if the scale covariance depends on covariates, i.e., Σ_i has a non-stationary dependence structure. Table 6 displays the comparisons of average MSFEs for the first six step-ahead forecasts and the associated BIC values based on the 500 generated samples and each of the sample sizes $m = 25, 100$. As anticipated, the numerical results indicate the TJMM produces substantial gains in prediction accuracy as well as modeling adequacy.

6. Concluding remarks

We present a robust approach to the analysis of longitudinal data with the mean and the scale covariance modeled jointly in a framework of multivariate *t* errors. In particular, the reparameterization scheme in (2) suggests a flexible way of modeling the variance components as a function of covariates, making the model to be more parsimonious than the arbitrary covariance structure. We develop an explicit procedure for ML estimation based on the scoring method, which can be easily performed by practitioners. Numerical results illustrated in Section 4 indicate that the proposed TJMM is suitable for these two real-life applications and yields better performance on the prediction of future values. The simulation results suggest the proposed TJMM is a promising technique for longitudinal data with nonstationary dependence structure. As a note in passing, the situation in

Table 5

Simulation results: average estimates for $\beta, \gamma, \lambda, \ell_{\max}$ and BIC and median estimates for ν based on 500 replications. Values within parentheses are empirical standard errors.

Parameter	True value	$m = 25$				$m = 100$			
		$\nu = 4$		$\nu = 50$		$\nu = 4$		$\nu = 50$	
		TJMM	NJMM	TJMM	NJMM	TJMM	NJMM	TJMM	NJMM
β_0	5.27	5.2701 (0.0841)	5.2732 (0.1786)	5.2802 (0.0788)	5.2721 (0.0892)	5.2700 (0.0497)	5.2674 (0.0841)	5.2720 (0.0372)	5.2696 (0.0375)
β_1	0.16	0.1601 (0.0070)	0.1602 (0.0089)	0.1598 (0.0065)	0.1599 (0.0071)	0.1601 (0.0035)	0.1601 (0.0044)	0.1600 (0.0033)	0.1599 (0.0033)
γ_0	1.12	1.0984 (0.1134)	1.0820 (0.1566)	1.1038 (0.1049)	1.0956 (0.1095)	1.1141 (0.0543)	1.1067 (0.0851)	1.1184 (0.0483)	1.1154 (0.0517)
γ_1	-0.58	-0.5649 (0.0887)	-0.5540 (0.1227)	-0.5705 (0.0832)	-0.5640 (0.0852)	-0.5760 (0.0423)	-0.5702 (0.0665)	-0.5797 (0.0377)	-0.5766 (0.0401)
γ_2	0.09	0.0872 (0.0181)	0.0852 (0.0250)	0.0884 (0.0171)	0.0872 (0.0173)	0.0892 (0.0086)	0.0881 (0.0135)	0.0901 (0.0077)	0.0894 (0.0081)
γ_3	-0.0043	-0.0042 (0.0011)	-0.0040 (0.0015)	-0.0042 (0.0010)	-0.0042 (0.0010)	-0.0043 (0.0005)	-0.0042 (0.0008)	-0.0043 (0.0005)	-0.0043 (0.0005)
λ_0	-0.4	-0.4175 (0.5387)	-0.1527 (0.7327)	-0.3997 (0.4802)	-0.3981 (0.4516)	-0.4032 (0.2501)	-0.2801 (0.4531)	-0.4024 (0.2332)	-0.3801 (0.2285)
λ_1	-1.37	-1.3636 (0.3400)	-1.3525 (0.4620)	-1.4005 (0.3101)	-1.3776 (0.3069)	-1.3666 (0.1505)	-1.3741 (0.2427)	-1.3713 (0.1487)	-1.3643 (0.1443)
λ_2	0.18	0.1797 (0.0598)	0.1786 (0.0814)	0.1858 (0.0546)	0.1831 (0.0553)	0.1800 (0.0262)	0.1809 (0.0418)	0.1805 (0.0261)	0.1796 (0.0255)
λ_3	-0.007	-0.0071 (0.0030)	-0.0072 (0.0041)	-0.0073 (0.0028)	-0.0072 (0.0028)	-0.0070 (0.0013)	-0.0071 (0.0021)	-0.0070 (0.0013)	-0.0070 (0.0013)
ν		4.31 (2.25)	-	39.52 (45.33)	-	4.09 (0.77)	-	49.62 (38.38)	-
ℓ_{\max}		-14.27 (28.56)	-45.78 (41.73)	39.13 (14.17)	39.87 (14.32)	-73.93 (58.27)	-237.84 (117.55)	138.29 (26.60)	141.18 (27.85)
BIC		2.56 (2.28)	4.95 (3.34)	-1.71 (1.15)	-2.39 (1.13)	1.98 (1.16)	5.22 (2.35)	-2.26 (0.53)	-2.62 (0.56)

Table 6

Comparison of predictive accuracies in terms of MSFE and the BIC values for two robust t models.

m	Model	q step-ahead forecasts						BIC
		$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$	$q = 6$	
25	TLMM	0.1042 (0.0611)	0.1208 (0.0608)	0.1345 (0.0661)	0.1472 (0.0626)	0.1582 (0.0710)	0.1631 (0.0711)	3.76 (2.33)
	TJMM	0.0920 (0.0583)	0.1098 (0.0598)	0.1225 (0.0619)	0.1338 (0.0659)	0.1411 (0.0705)	0.1426 (0.0646)	2.56 (2.28)
	RIP (%)	11.7	9.1	8.9	9.1	10.8	12.5	31.9
100	TLMM	0.1068 (0.0400)	0.1285 (0.0405)	0.1437 (0.0413)	0.1556 (0.0437)	0.1652 (0.0464)	0.1724 (0.0511)	3.43 (1.17)
	TJMM	0.0961 (0.0363)	0.1182 (0.0394)	0.1322 (0.0405)	0.1408 (0.0382)	0.1477 (0.0408)	0.1522 (0.0487)	1.98 (1.16)
	RIP (%)	10.1	8.1	8.0	9.5	10.6	11.7	42.3

The relative improvement percentage (RIP) is measured by $(TLMM - TJMM)/TLMM \times 100\%$. (Replications = 500).

which subjects containing intermittent missing values can be easily handled via the EM-type algorithms (Dempster et al., 1977; Meng and Dyk, 1997).

Some possibilities for the future research along this line are as follows. One possible extension is to apply the joint mean-covariance modeling approach to other non-normal random errors such as the skew normal distribution (Azzalini and Dalla Valle, 1996; Azzalini and Capitanio, 1999) and the skew t distribution (Jones and Faddy, 2003; Azzalini and Capitanio, 2003). With the development of rapid computational techniques and low-cost computer powers, another worthwhile task is to develop a fully Bayesian inference for gaining more reliable inferences via the Markov chain Monte Carlo method.

Acknowledgments

The authors are grateful to the Executive Editor, the Associate Editor and the referee for their valuable comments and helpful suggestions which led to substantial improvements of the paper. This research work was supported by the National Science Council of Taiwan (Grant no. NSC97-2118-M-005-001-MY2).

Appendix A. Elements of the Hessian matrix

$$\mathbf{H}_{\beta\beta} = - \sum_{i=1}^m \tau_i \left\{ \mathbf{X}_i^T \Sigma_i^{-1} \mathbf{X}_i - 2 \frac{\mathbf{X}_i^T \Sigma_i^{-1} \mathbf{r}_i \mathbf{r}_i^T \Sigma_i^{-1} \mathbf{X}_i}{v + \Delta_i(\boldsymbol{\beta}, \boldsymbol{\lambda}, \gamma)} \right\}, \tag{A.1}$$

$$\mathbf{H}_{\beta\gamma} = \sum_{i=1}^n \tau_i \left\{ \frac{\partial \mathbf{X}_i^T \mathbf{L}_i^T}{\partial \boldsymbol{\gamma}^T} (\mathbf{I}_d \otimes \mathbf{D}_i^{-1} \mathbf{L}_i \mathbf{r}_i) + \mathbf{X}_i^T \mathbf{L}_i^T \frac{\partial \mathbf{D}_i^{-1} \mathbf{L}_i \mathbf{r}_i}{\partial \gamma} + 2 \frac{\mathbf{X}_i^T \Sigma_i^{-1} \mathbf{r}_i (\mathbf{r}_i - \mathbf{Z}_i \boldsymbol{\gamma})^T \mathbf{D}_i^{-1} \mathbf{Z}_i}{v + \Delta_i(\boldsymbol{\beta}, \boldsymbol{\lambda}, \gamma)} \right\}, \tag{A.2}$$

$$\mathbf{H}_{\beta\lambda} = \sum_{i=1}^m \tau_i \left\{ \frac{\partial \mathbf{X}_i^T \mathbf{L}_i^T \mathbf{D}_i^{-1} \mathbf{L}_i \mathbf{r}_i}{\partial \boldsymbol{\lambda}^T} - \frac{\mathbf{X}_i^T \Sigma_i^{-1} \mathbf{r}_i (\sum_{j=1}^{n_i} \eta_{ij}^2 \mathbf{w}_j^T)}{v + \Delta_i(\boldsymbol{\beta}, \boldsymbol{\lambda}, \gamma)} \right\}, \tag{A.3}$$

$$\mathbf{H}_{\beta v} = \sum_{i=1}^m \left\{ \frac{1}{v + \Delta_i(\boldsymbol{\beta}, \boldsymbol{\lambda}, \gamma)} - \frac{\tau_i}{v + \Delta_i(\boldsymbol{\beta}, \boldsymbol{\lambda}, \gamma)} \right\} \mathbf{X}_i^T \Sigma_i^{-1} \mathbf{r}_i, \tag{A.4}$$

$$\mathbf{H}_{\gamma\gamma} = - \sum_{i=1}^m \tau_i \left\{ \mathbf{Z}_i^T \mathbf{D}_i^{-1} \mathbf{Z}_i - \frac{2 \mathbf{Z}_i^T \mathbf{D}_i^{-1} (\mathbf{r}_i - \mathbf{Z}_i \boldsymbol{\gamma}) (\mathbf{r}_i - \mathbf{Z}_i \boldsymbol{\gamma})^T \mathbf{D}_i^{-1} \mathbf{Z}_i}{v + \Delta_i(\boldsymbol{\beta}, \boldsymbol{\lambda}, \gamma)} \right\}, \tag{A.5}$$

$$\mathbf{H}_{\gamma\lambda} = - \sum_{i=1}^m \tau_i \left\{ \mathbf{Z}_i^T \mathbf{D}_i^{-1} \boldsymbol{\Psi}_i - \frac{\mathbf{Z}_i^T \mathbf{D}_i^{-1} (\mathbf{r}_i - \mathbf{Z}_i \boldsymbol{\gamma}) (\sum_{j=1}^{n_i} \eta_{ij}^2 \mathbf{w}_j^T)}{v + \Delta_i(\boldsymbol{\beta}, \boldsymbol{\lambda}, \gamma)} \right\}, \tag{A.6}$$

$$\mathbf{H}_{\lambda\lambda} = - \frac{1}{2} \sum_{i=1}^m \tau_i \left\{ \sum_{j=1}^{n_i} \eta_{ij}^2 \mathbf{w}_j \mathbf{w}_j^T - \frac{(\sum_{j=1}^{n_i} \eta_{ij}^2 \mathbf{w}_j) (\sum_{j=1}^{n_i} \eta_{ij}^2 \mathbf{w}_j^T)}{v + \Delta_i(\boldsymbol{\beta}, \boldsymbol{\lambda}, \gamma)} \right\}, \tag{A.7}$$

$$\mathbf{H}_{\lambda v} = \frac{1}{2} \sum_{i=1}^m \left\{ \frac{1}{v + \Delta_i(\boldsymbol{\beta}, \boldsymbol{\lambda}, \gamma)} - \frac{\tau_i}{v + \Delta_i(\boldsymbol{\beta}, \boldsymbol{\lambda}, \gamma)} \right\} \sum_{j=1}^{n_i} \eta_{ij}^2 \mathbf{w}_j, \tag{A.8}$$

$$\mathbf{H}_{\gamma v} = \sum_{i=1}^m \left\{ \frac{1}{v + \Delta_i(\boldsymbol{\beta}, \boldsymbol{\lambda}, \gamma)} - \frac{\tau_i}{v + \Delta_i(\boldsymbol{\beta}, \boldsymbol{\lambda}, \gamma)} \right\} \mathbf{Z}_i^T \mathbf{D}_i^{-1} (\mathbf{r}_i - \mathbf{Z}_i \boldsymbol{\gamma}), \tag{A.9}$$

$$\mathbf{H}_{v v} = \frac{1}{2} \sum_{i=1}^m \left\{ \frac{1}{2} \psi \left(\frac{v + n_i}{2} \right) - \frac{1}{2} \psi \left(\frac{v}{2} \right) + \frac{n_i}{v^2} + \frac{\Delta_i(\boldsymbol{\beta}, \boldsymbol{\lambda}, \gamma)}{v + \Delta_i(\boldsymbol{\beta}, \boldsymbol{\lambda}, \gamma)} \left(\frac{1 - \tau_i}{v} - \frac{n_i}{v^2} \right) \right\}, \tag{A.10}$$

where $\boldsymbol{\Psi}_i$ is defined in (8), \mathbf{I}_d is the identity matrix of order d and \otimes is the Kronecker product. Note that $\partial \mathbf{D}_i^{-1} / \partial \lambda_k = \text{diag}\{\mathbf{w}_1^T \mathbf{e}_k \sigma_1^{-2}, \dots, \mathbf{w}_{n_i}^T \mathbf{e}_k \sigma_{n_i}^{-2}\}$ with \mathbf{e}_k being a $q \times 1$ binary vector whose k th element is one and is zero elsewhere; $\partial \mathbf{L}_i / \partial \gamma_k$ is a lower triangular matrix whose (j, k) th entry is $\mathbf{z}_{js}^T \mathbf{e}_k^*$, for $j = 2, \dots, n_i$ and $s = 1, \dots, (i - 1)$, where \mathbf{e}_k^* is a $d \times 1$ binary vector with the k th element being one and zero elsewhere.

Appendix B. Proof of Proposition 2

(a) By Proposition 1(c), we have

$$E \left(\frac{1}{v + \Delta_i(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})} \right) = E \left(\frac{B}{v} \right) = \frac{1}{v + n_i}.$$

(b) Since $\tau_i = (n_i + v) / (v + \Delta_i(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})) = ((n_i + v) / v) B$, we have

$$E \left(\frac{\tau_i}{v + \Delta_i(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})} \right) = \frac{v + n_i}{v^2} E(B^2) = \frac{v + 2}{v(v + n_i + 2)}.$$

(c) Denote by $t_p(\mathbf{r}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ the density of $T_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. Note that $\mathbf{r}_i \sim T_{n_i}(\mathbf{0}, \boldsymbol{\Sigma}_i, \nu)$, we then have

$$\begin{aligned} E(\tau_i^2 \mathbf{r}_i \mathbf{r}_i^T) &= \frac{(\nu + n_i)^2}{\nu^2} \int \mathbf{r}_i \mathbf{r}_i^T \frac{\Gamma\left(\frac{\nu + n_i}{2}\right) |\boldsymbol{\Sigma}_i|^{-1/2}}{\Gamma\left(\frac{\nu}{2}\right) (\pi\nu)^{n_i/2}} \left(1 + \frac{\mathbf{r}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i}{\nu}\right)^{-((\nu+4)+n_i)/2} d\mathbf{r}_i \\ &= \frac{(\nu + n_i)^2}{\nu^2} \frac{\Gamma\left(\frac{\nu + n_i}{2}\right) |\boldsymbol{\Sigma}_i|^{-1/2}}{\Gamma\left(\frac{\nu}{2}\right) (\pi\nu)^{n_i/2}} \frac{\frac{\nu + 2}{2} \frac{\nu}{2} \Gamma\left(\frac{\nu}{2}\right) (\pi(\nu + 4))^{n_i/2}}{\frac{\nu + n_i + 2}{2} \frac{\nu + n_i}{2} \Gamma\left(\frac{\nu + n_i}{2}\right) \left(\frac{\nu + 4}{\nu}\right)^{n_i/2} |\boldsymbol{\Sigma}_i|^{-1/2}} \\ &\quad \times \int \mathbf{r}_i \mathbf{r}_i^T t_{n_i}\left(\mathbf{r}_i|\mathbf{0}, \frac{\nu}{\nu + 4} \boldsymbol{\Sigma}_i, \nu + 4\right) d\mathbf{r}_i \\ &= \frac{(\nu + n_i)}{(\nu + n_i + 2)} \boldsymbol{\Sigma}_i. \end{aligned}$$

(d) Define $\mathbf{L}_i = [\ell_{i1} \ \ell_{i2} \ \dots \ \ell_{in_i}]^T$. Since $\mathbf{L}_i \mathbf{r}_i = (\mathbf{r}_i - \hat{\mathbf{r}}_i)$, it implies that $(r_{ij} - \hat{r}_{ij}) = \ell_{ij}^T \mathbf{r}_i$. We then have

$$\begin{aligned} E\left(\tau_i \sum_{j=1}^{n_i} \eta_{ij}^2 \mathbf{w}_j \mathbf{w}_j^T\right) &= E\left\{\tau_i \sum_{j=1}^{n_i} \sigma_j^{-2} \mathbf{w}_j \mathbf{r}_i^T \ell_{ij} \ell_{ij}^T \mathbf{r}_i \mathbf{w}_j^T\right\} \\ &= \sum_{j=1}^{n_i} \sigma_j^{-2} \mathbf{w}_j E(\tau_i \mathbf{r}_i^T \ell_{ij} \ell_{ij}^T \mathbf{r}_i) \mathbf{w}_j^T \\ &= \sum_{j=1}^{n_i} \sigma_j^{-2} \mathbf{w}_j (\text{tr}(\ell_{ij} \ell_{ij}^T \boldsymbol{\Sigma}_i)) \mathbf{w}_j^T \\ &= \sum_{j=1}^{n_i} \mathbf{w}_j \mathbf{w}_j^T, \end{aligned}$$

where the last equality follows directly from $\text{tr}(\ell_{ij} \ell_{ij}^T \boldsymbol{\Sigma}_i) = \ell_{ij}^T \boldsymbol{\Sigma}_i \ell_{ij} = \sigma_j^2$.

(e) The result follows directly from the fact of

$$\begin{aligned} E\left(\frac{\partial^2 \log f(\mathbf{Y}_i)}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T}\right) &= -\frac{1}{2} \left\{ E\left(\tau_i \sum_{j=1}^{n_i} \eta_{ij}^2 \mathbf{w}_j \mathbf{w}_j^T\right) - \frac{1}{\nu + n_i} E\left(\tau_i^2 \left(\sum_{j=1}^{n_i} \eta_{ij}^2 \mathbf{w}_j\right) \left(\sum_{j=1}^{n_i} \eta_{ij}^2 \mathbf{w}_j^T\right)\right) \right\} \\ &= -E\left(\frac{\partial \log f(\mathbf{Y}_i)}{\partial \boldsymbol{\lambda}} \frac{\partial \log f(\mathbf{Y}_i)}{\partial \boldsymbol{\lambda}^T}\right) \\ &= \frac{1}{4} \left(\sum_{j=1}^{n_i} \mathbf{w}_j\right) \left(\sum_{j=1}^{n_i} \mathbf{w}_j^T\right) - \frac{1}{4} E\left(\tau_i^2 \left(\sum_{j=1}^{n_i} \eta_{ij}^2 \mathbf{w}_j\right) \left(\sum_{j=1}^{n_i} \eta_{ij}^2 \mathbf{w}_j^T\right)\right). \end{aligned}$$

(f) Since $\mathbf{z}(i, j) = \sum_{k=1}^{j-1} r_{ik} \mathbf{z}_{jk} = \boldsymbol{\kappa}_j \mathbf{r}_i$, the following trivially holds:

$$\mathbf{Z}_i^T \mathbf{D}_i^{-1} \mathbf{Z}_i = \sum_{j=1}^{n_i} \sigma_j^{-2} \mathbf{z}(i, j) \mathbf{z}^T(i, j) = \sum_{j=1}^{n_i} \sigma_j^{-2} \boldsymbol{\kappa}_j \mathbf{r}_i \mathbf{r}_i^T \boldsymbol{\kappa}_j^T.$$

Therefore,

$$E(\tau_i \mathbf{Z}_i^T \mathbf{D}_i^{-1} \mathbf{Z}_i) = E\left(\tau_i \sum_{j=1}^{n_i} \sigma_j^{-2} \boldsymbol{\kappa}_j \mathbf{r}_i \mathbf{r}_i^T \boldsymbol{\kappa}_j^T\right) = \sum_{j=1}^{n_i} \sigma_j^{-2} \boldsymbol{\kappa}_j \boldsymbol{\Sigma}_i \boldsymbol{\kappa}_j^T.$$

(g) Since $E(\tau_i \mathbf{r}_i \mathbf{r}_i^T) = \boldsymbol{\Sigma}_i$, it is straightforward to show that

$$E(\tau_i \mathbf{z}(i, j) \mathbf{r}_{ij}) = E\left(\tau_i \sum_{k=1}^{j-1} r_{ik} \mathbf{z}_{jk} \mathbf{r}_{ij}\right) = \sum_{k=1}^{j-1} E(\tau_i r_{ik} \mathbf{r}_{ij}) \mathbf{z}_{jk} = \sum_{k=1}^{j-1} \sigma_{kj} \mathbf{z}_{jk}.$$

(h) Since $\mathbf{Z}_i^T \mathbf{D}_i^{-1} \boldsymbol{\Psi}_i = \sum_{j=1}^{n_i} \sigma_j^{-2} \mathbf{z}(i, j) r_{ij} \mathbf{w}_j^T - \sum_{j=1}^{n_i} \sigma_j^{-2} \mathbf{z}(i, j) \mathbf{z}^T(i, j) \boldsymbol{\gamma} \mathbf{w}_j^T$, it suffices to verify that

$$\begin{aligned} E(\tau_i \mathbf{Z}_i^T \mathbf{D}_i^{-1} \boldsymbol{\Psi}_i) &= \sum_{j=1}^{n_i} \sigma_j^{-2} E(\tau_i \mathbf{z}(i, j) r_{ij}) \mathbf{w}_j^T - \sum_{j=1}^{n_i} \sigma_j^{-2} E(\tau_i \mathbf{z}(i, j) \mathbf{z}^T(i, j)) \boldsymbol{\gamma} \mathbf{w}_j^T \\ &= \sum_{j=1}^{n_i} \sigma_j^{-2} \left(\sum_{k=1}^{j-1} \sigma_{kj} \mathbf{z}_{jk} \right) \mathbf{w}_j^T - \sum_{j=1}^{n_i} \sigma_j^{-2} (\boldsymbol{\kappa}_j \boldsymbol{\Sigma}_i \boldsymbol{\kappa}_j^T) \boldsymbol{\gamma} \mathbf{w}_j^T \\ &= \sum_{j=1}^{n_i} \sigma_j^{-2} \left(\sum_{k=1}^{j-1} \sigma_{kj} \mathbf{z}_{jk} \mathbf{w}_j^T - \boldsymbol{\kappa}_j \boldsymbol{\Sigma}_i \boldsymbol{\kappa}_j^T \boldsymbol{\gamma} \mathbf{w}_j^T \right). \end{aligned}$$

References

- Azzalini, A., Capitaino, A., 1999. Statistical applications of the multivariate skew-normal distribution. *J. Roy. Statist. Soc. B* 61, 579–602.
- Azzalini, A., Capitaino, A., 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew *t*-distribution. *J. Roy. Statist. Soc. B* 65, 367–389.
- Azzalini, A., Dalla Valle, A., 1996. The multivariate skew-normal distribution. *Biometrika* 83, 715–726.
- Demidenko, E., 2004. *Mixed Models: Theory and Applications*. Wiley, New York.
- Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. B* 39, 1–38.
- Geisser, S., 1975. The predictive sample reuse method with applications. *J. Am. Statist. Assoc.* 70, 320–328.
- He, X., Cui, H., Simpson, D.G., 2004. Longitudinal data analysis using *t*-type regression. *J. Statist. Plann. Inference* 122, 253–269.
- Healy, M.J.R., 1968. Multivariate normal plotting. *Appl. Statist.* 17, 157–161.
- Jones, M.C., Faddy, M.J., 2003. A skew extension of the *t*-distribution, with applications. *J. Roy. Statist. Soc. B* 65, 159–174.
- Kotz, S., Nadarajah, S., 2004. *Multivariate *t* Distributions and their Applications*. Cambridge University Press, Cambridge.
- Laird, N.M., Ware, J.H., 1982. Random effects models for longitudinal data. *Biometrics* 38, 963–974.
- Lange, K.L., Little, R.J.A., Taylor, J.M.G., 1989. Robust statistical modeling using the *t* distribution. *J. Amer. Statist. Assoc.* 84, 881–896.
- Lin, T.I., Lee, J.C., 2006. A robust approach to *t* linear mixed models applied to multiple sclerosis data. *Statist. Med.* 25, 1397–1412.
- Meng, X.L., Dyk, V.D., 1997. The EM algorithm—an old folk-song sung to a fast new tune. *J. Roy. Statist. Soc. B* 59, 511–567.
- Nadarajah, S., Kotz, S., 2005. Mathematical properties of the multivariate *t* distribution. *Acta Appl. Math.* 89, 53–84.
- Pan, J., MacKenzie, G., 2003. On modelling mean-covariance structures in longitudinal studies. *Biometrika* 90, 239–244.
- Pinheiro, J.C., Liu, C.H., Wu, Y.N., 2001. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate *t* distribution. *J. Comput. Graph. Statist.* 10, 249–276.
- Potthoff, R.F., Roy, S.N., 1964. A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* 51, 313–326.
- Pourahmadi, M., 1999. Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika* 86, 677–690.
- Pourahmadi, M., 2000. Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika* 87, 425–435.
- Pourahmadi, M., 2001. *Foundations of Time Series Analysis and Prediction Theory*. Wiley, New York.
- Rao, C.R., 1987. Prediction of future observations in growth curve models (with discussion). *Statist. Sci.* 2, 434–471.
- Ryggard, K., Spang-Thomsen, M., 1997. Quantitation and Gompertzian analysis of tumor growth data. *Breast Cancer Res. Treat.* 46, 303–312.
- Stone, M., 1974. Cross-validatory choice and assessment of statistical prediction (with discussion). *J. R. Statist. Soc. B* 36, 111–147.
- Taylor, J., Verbyla, A., 2004. Joint modelling of location and scale parameters of the *t* distribution. *Statist. Modelling* 4, 91–112.
- Welsh, A.H., Richardson, A.M., 1997. Approaches to the robust estimation of mixed models. In: Maddala, G.S., Rao, C.R. (Eds.), *Handbooks of Statistics*, vol. 15. Elsevier Science, Amsterdam, pp. 343–383.
- Zacks, S., 1971. *The Theory of Statistical Inference*. Wiley, New York.
- 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.