### Title

Mixture of linear mixed models using multivariate t distribution




### Journal

Journal of Statistical Computation and Simulation, 86(4)

### ISSN

0094-9655



### Authors

Bai, Xiuqin
Chen, Kun
Yao, Weixin



### Publication Date

2016-03-03

### DOI

10.1080/00949655.2015.1036431

Peer reviewed

# Mixture of Linear Mixed Models Using Multivariate $t$ Distribution

Xiuqin Bai, * Kun Chen,† Weixin Yao, ‡

**Abstract**

Linear mixed models are widely used when multiple correlated measurements are made on each unit of interest. In many applications, the units may form several distinct clusters, and such heterogeneity can be more appropriately modeled by a *finite mixture linear mixed model*. The classical estimation approach, in which both the random effects and the error parts are assumed to follow normal distribution, is sensitive to outliers, and failure to accommodate outliers may greatly jeopardize the model estimation and inference. We propose a new mixture linear mixed model using multivariate $t$ distribution. For each mixture component, we assume the response and the random effects jointly follow a multivariate $t$ distribution, to conveniently robustify the estimation procedure. An efficient ECM algorithm is developed for conducting maximum likelihood estimation. The degrees of freedom parameters of the $t$ distributions are chosen data adaptively, for achieving flexible tradeoff between estimation robustness and efficiency. Simulation studies and an application on analyzing lung growth longitudinal data showcase the efficacy of the proposed approach.

*Xiuqin Bai is PhD student, Department of Statistics, Kansas State University, Manhattan, Kansas 66506. Email: bxq@ksu.edu.

†Kun Chen is Assistant Professor, Department of Statistics, University of Connecticut, Storrs, CT, 06269. Email: kun.chen@uconn.edu.

‡Weixin Yao is Associate Professor, Department of Statistics, University of California, Riverside, California 92521, U.S.A. Email: weixin.yao@ucr.edu.

**Key words**: ECM algorithm; Linear mixed models; Longitudinal data; Mixture models; Multivariate $t$ distribution.

# 1 Introduction

The classical linear mixed model can be expressed as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\mathbf{b} + \boldsymbol{\epsilon} \tag{1.1}$$

where $\mathbf{y}$ is a $N \times 1$ response vector, $\mathbf{X}$ is a $N \times p$ design matrix for the fixed effects, $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed-effect coefficients, $\mathbf{U}$ is a $N \times q$ design matrix for the random effects, $\mathbf{b} \sim N_q(0, \boldsymbol{\Psi})$ is a $q \times 1$ vector of random effect coefficients, and $\boldsymbol{\epsilon}$ is a $N \times 1$ vector of errors for observations and assumed to have multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{\Lambda}$. Based on the above model setup, $\mathbf{X}\boldsymbol{\beta}$ models the fixed effects and $\mathbf{U}\mathbf{b}$ models the random effects. It follows that $\mathbf{y}$ has a multivariate normal distribution with mean $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and covariance matrix $\mathbf{V} = \text{cov}(\mathbf{y}) = \mathbf{U}\boldsymbol{\Psi}\mathbf{U}^T + \boldsymbol{\Lambda}$. ~~The main goal here is still to model the relationship between response variable and predictor variables. Linear mixed models are thus considered important extensions of the conventional linear regression models for handling dependent data, which arise in various problems, e.g., when the observations are taken on groups of related individuals, or when repeated measurements are made over time on the same set of individuals.~~For clarity and without loss of generality, in the sequel we shall mainly refer to the repeated measurement setup when presenting our proposed methodology, similar to Celeux et al. (2005).

In many applications, however, the underlying assumption that the regression relationship is homogeneous across all the subjects could be violated. Of particular interest here is the situation that the subjects may form several distinct clusters, indicating mixed regression relationships. Such heterogeneity can be modeled by a finite mixture of linear mixed regression models, consisting of, say, $m$ homogeneous groups/components (Celeux et al., 2005; Yau et al., 2002; Ng et al., 2006). Suppose there are $I$ subjects under study,

and $n_i$ repeated measurements are gathered on the $i$th subjects, for $i = 1, \dots I$. We consider a mixture linear mixed model setup as follows. For each $i = 1, \dots I$, let $Z_i$ be a latent variable with $P(Z_i = j) = \pi_j$, $j = 1, \dots m$. Given $Z_i = j$, we assume that the response $\mathbf{y}_i \in \mathbb{R}^{n_i}$ follows a linear mixed model, i.e.,

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta}_j + \mathbf{U}_i \mathbf{b}_{ij} + \mathbf{e}_{ij} \tag{1.2}$$

where $\mathbf{X}_i \in \mathbb{R}^{n_i \times p}$ is the fixed-effect covariate matrix, $\boldsymbol{\beta}_j \in \mathbb{R}^p$ a fixed-effect coefficient vector, $\mathbf{U}_i \in \mathbb{R}^{n_i \times q}$ the random-effect covariate matrix, $\mathbf{b}_{ij}$ the random-effect coefficient vector which is thought as random, and $\mathbf{e}_{ij}$ the random error vector. Following the conventional formulations of the normal mixture model and the mixed model, it is natural to assume that

$$\mathbf{b}_{ij} \sim N_q(0, \boldsymbol{\Psi}_j), \qquad \mathbf{e}_{ij} \sim N_{n_i}(0, \boldsymbol{\Lambda}_{ij}),$$

and all $\mathbf{b}_{ij}$s, $\mathbf{e}_{ij}$s, for $i = 1, \dots, I$ and $j = 1, \dots, m$ are independent. Usually each error covariance matrix $\boldsymbol{\Lambda}_{ij}$ is assumed to be dependent on $i$ only through its dimension, e.g., an AR(1) correlation structure with some correlation parameter $\rho$ so that $\boldsymbol{\Lambda}_{ij} = \boldsymbol{\Lambda}(\rho, i)$. The correlation structure among each $n_i$ observations on subject $i$ is induced and modeled by the random component $\mathbf{U}_i \mathbf{b}_{ij}$. Conditional on $Z_i = j$, the joint distribution of $(\mathbf{y}_i, \mathbf{b}_{ij})$ is

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{b}_{ij} \end{pmatrix} \Big| Z_i = j \sim N_{n_i+q}\left( \begin{pmatrix} \mathbf{X}_i \boldsymbol{\beta}_j \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{U}_i \boldsymbol{\Psi}_j \mathbf{U}_i^T + \boldsymbol{\Lambda}_{ij} & \mathbf{U}_i \boldsymbol{\Psi}_j \\ \boldsymbol{\Psi}_j \mathbf{U}_i^T & \boldsymbol{\Psi}_j \end{pmatrix} \right), \tag{1.3}$$

and the mixture distribution of $\mathbf{y}_i$ itself, without observing $Z_i$, is

$$\mathbf{y}_i \sim \sum_{j=1}^{m} \pi_j N_{n_i}(\mathbf{X}_i \boldsymbol{\beta}_j, \mathbf{U}_i \boldsymbol{\Psi}_j \mathbf{U}_i^T + \boldsymbol{\Lambda}_{ij}). \tag{1.4}$$

Although the above normal mixture linear mixed model is quite appealing in modeling the regression relationship with the aforementioned hierarchically clustered data, one

drawback of the model is that it can be very sensitive to outliers, an undesirable property inherited from the normal mixture model.

Motivated by Lange et al. (1989), Welsh and Richardson (1997), and Pinheiro et al. (2001), we propose a new mixture linear mixed model by replacing the normal distribution with the multivariate $t$ distribution. For each mixture component, we assume the response and the random effects jointly follow a multivariate $t$ distribution, in a similar fashion as (1.3), to conveniently robustify the estimation procedure. An efficient ECM algorithm is developed for conducting maximum likelihood estimation. The degrees of freedom parameters of the $t$ distributions are chosen data adaptively for achieving flexible tradeoff between estimation robustness and efficiency. We demonstrate via simulation study that the proposed approach is indeed robust and can be much more efficient than the traditional normal mixture model when outliers are present in the data, and in the absence of outliers the proposed approach leads to comparable performance to that of the normal mixture model. An application on lung growth of children further showcases the efficacy of the proposed approach.

The rest of this paper is organized as follows. In Section 2, we introduce our new method of using the multivariate $t$ distribution in mixture of linear mixed-effects models. In Section 3, we provide a simulation study to compare our new method with the traditional normal based estimation method. In Section 4, An application of the new method to a real data set is provided. We conclude this paper with some discussion in Section 5.

# 2 Robust $t$-Mixture Linear Mixed Models

## 2.1 The $t$-Mixture of Linear Mixed Models

In practice, outliers and anomalies are bounded to occur, and failure to accommodate outliers may put both the model estimation and inference in jeopardy. This motivates us to construct a robust $t$-mixture of linear mixed model. Given $Z_i = j$, we start by

assuming that the joint distribution of $(\mathbf{y}_i, \mathbf{b}_{ij})$ is

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{b}_{ij} \end{pmatrix} \mid Z_i = j \sim t_{n_i+q}\left( \begin{pmatrix} \mathbf{X}_i\boldsymbol{\beta}_j \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{U}_i\boldsymbol{\Psi}_j\mathbf{U}_i^T + \boldsymbol{\Lambda}_{ij} & \mathbf{U}_i\boldsymbol{\Psi}_j \\ \boldsymbol{\Psi}_j\mathbf{U}_i^T & \boldsymbol{\Psi}_j \end{pmatrix}, \nu_j \right), \tag{2.1}$$

where we use $t_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ to denotes a $n$-dimensional multivariate $t$ distribution with mean vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$ and degrees of freedom $\nu$; in the sequel we also use $t_n(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ to denote its probability density function. Throughout, the error covariance matrices are assumed to take the form $\boldsymbol{\Lambda}_{ij} = \sigma_j^2\mathbf{R}_i$, for $i = 1, \ldots, I$, $j = 1, \ldots, m$, where $\mathbf{R}_i$ are known matrices taken to be the identity matrix, unless ~~otherwisely~~ otherwise noted.

The proposed approach essentially assumes that $\mathbf{y}_i$ follows a mixture distribution,

$$\mathbf{y}_i \sim \sum_{j=1}^{m} \pi_j t_{n_i}(\mathbf{X}_i\boldsymbol{\beta}_j, \mathbf{U}_i\boldsymbol{\Psi}_j\mathbf{U}_i^T + \boldsymbol{\Lambda}_{ij}, \nu_j), \tag{2.2}$$

and given the observed data for $i = 1, \ldots, I$, the log-likelihood function is

$$\sum_{i=1}^{I} \ln \left\{ \sum_{j=1}^{m} \pi_j t_{n_i}(\mathbf{y}_i; \mathbf{X}_i\boldsymbol{\beta}_j, \mathbf{U}_i\boldsymbol{\Psi}_j\mathbf{U}_i^T + \boldsymbol{\Lambda}_{ij}, \nu_j) \right\}. \tag{2.3}$$

Comparing to model (1.4), we have used the multivariate $t$ distribution to replace the multivariate normal distribution, following similar idea in Lange et al. (1989). This extension allows us to carry out the mixture mixed effect model analysis for any data involving errors with longer-than-normal tails. The degrees of freedom parameters of the $t$ distributed components are unknown and estimated from the data, and this provides a convenient way for achieving flexible tradeoff between robustness and efficiency, i.e., in the special case $\nu = 1$, the distribution becomes a multivariate Cauchy distribution, and as $\nu \to \infty$, the distribution rolls back to the multivariate normal. Also note that in the above model we have directly specified the distribution of $\mathbf{y}_i$ as the multivariate $t$, instead of separately specifying the distributions of the random effects and the error terms, as the latter is unnecessary and may lead to untractable or inconvenient marginal distribution

of $\mathbf{y}_i$.

To understand better about model (2.2), we shall discuss several of its alternative representations, which may ultimately facilitate the maximum likelihood estimation to be elaborated in the next section. It is known that the multivariate $t$ distribution can be written as a normal scale mixture distribution, i.e., its probability density function $t(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ can be expressed as

$$t(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \int_0^\infty f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}/u) g(u; \frac{\nu}{2}, \frac{\nu}{2}) du,$$

where $f$ denotes the normal density and $g$ the Gamma density. In light of the above characterization, it is convenient to express model (2.2) as a hierarchical model,

$$\begin{aligned} \mathbf{y}_i \mid b_{ij}, \tau_{ij}, j = 1, \ldots, m &\sim \sum_{j=1}^{m} \pi_j N(\mathbf{X}_i \boldsymbol{\beta}_j + \mathbf{U}_i \mathbf{b}_{ij}, \frac{1}{\tau_{ij}} \boldsymbol{\Lambda}_{ij}), \\ b_{ij} \mid \tau_{ij} &\sim N(\mathbf{0}, \frac{1}{\tau_{ij}} \boldsymbol{\Psi}_j), \text{ for } j = 1, \ldots, m, \\ \tau_{ij} &\sim \text{Gamma}(\frac{\nu_j}{2}, \frac{\nu_j}{2}), \text{ for } j = 1, \ldots, m. \end{aligned} \tag{2.4}$$

Model (2.2) could also be written in a conventional linear mixed model form. Given $Z_i = j$,

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta}_j + \mathbf{U}_i \mathbf{b}_{ij} + \mathbf{e}_{ij}, \qquad i = 1, \ldots, I,$$

where $\mathbf{b}_{ij} \sim t_q(\mathbf{0}, \boldsymbol{\Psi}_j, \nu_j)$, and $\mathbf{e}_{ij} \sim t_{n_i}(\mathbf{0}, \boldsymbol{\Lambda}_{ij}, \nu_j)$. Condition on $\tau_{ij}$, $\mathbf{b}_{ij}$ is independent of $\mathbf{e}_{ij}$, which means that in general $\mathbf{b}_{ij}$ and $\mathbf{e}_{ij}$ are uncorrelated but not independent, for any $\nu_j < \infty$ (Pinheiro et al., 2001). It is now clear that in our proposed method, both $\mathbf{b}_{ij}$ and $\mathbf{e}_{ij}$ follow multivariate $t$ distribution, and thus the method is robust against potential outliers in both the random effects or the within-subject random errors.

By integrating out $\mathbf{b}_{ij}$, the hierarchical model can be equivalently expressed as

$$\mathbf{y}_i \mid \tau_{ij}, j = 1, \ldots, m \sim \sum_{j=1}^{m} \pi_j N(\mathbf{X}_i\boldsymbol{\beta}_j, \frac{1}{\tau_{ij}}(\mathbf{U}_i\boldsymbol{\Psi}_j\mathbf{U}_i^T + \boldsymbol{\Lambda}_{ij})),$$
$$\tau_{ij} \sim \text{Gamma}(\frac{\nu_j}{2}, \frac{\nu_j}{2}), \text{ for } j = 1, \ldots, m.$$

The conditional distribution of $\tau_{ij}$ can then be readily derived,

$$\tau_{ij} \mid \mathbf{y}_i, Z_i = j \sim \text{Gamma}\left(\frac{\nu_j + n_i}{2}, \frac{\nu_j + \delta_{ij}^2(\boldsymbol{\beta}_j, \boldsymbol{\Psi}_j, \sigma_j^2)}{2}\right),$$

where

$$\delta_{ij}^2(\boldsymbol{\beta}_j, \boldsymbol{\Psi}_j, \sigma_j^2) = (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}_j)^T(\mathbf{U}_i\boldsymbol{\Psi}_j\mathbf{U}_i^T + \boldsymbol{\Lambda}_{ij})^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}_j). \tag{2.5}$$

Therefore,

$$E(\tau_{ij} \mid \mathbf{y}_i, Z_i = j) = \frac{\nu_j + n_i}{\nu_j + \delta_{ij}^2(\boldsymbol{\beta}_j, \boldsymbol{\Psi}_j, \sigma_j^2)}. \tag{2.6}$$

The above results will be useful in the proposed ECM algorithm in next section.

## 2.2 An Efficient ECM Algorithm For Maximum Likelihood Estimation

We propose to conduct maximum likelihood estimation and inference of the proposed robust $t$-mixture linear mixed model. Direct maximization of the log-likelihood function (2.3) constructed from mixture multivariate $t$ distributions is quite difficult. In this section, we derive an efficient ECM algorithm to solve the problem, extending the works by Lange et al. (1989) and Pinheiro et al. (2001) in a more general context of mixture model. The EM algorithm is commonly applied in problems with missing or incomplete data, which is particularly suitable here, in view of the alternative hierarchical model representation of the $t$-mixture model in (2.4).

Let $\mathbf{y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_I\}, \mathbf{b} = \{\mathbf{b}_{ij}; i = 1, \ldots, I, j = 1, \ldots, m\}$, and $\boldsymbol{\tau} = \{\tau_{ij}; i = 1, \ldots, I, j = 1, \ldots, m\}$. Let

$$Z_{ij} = \begin{cases} 1 & \text{if the } i\text{th subject is from the } j\text{th mixture component,} \\ 0 & \text{otherwise,} \end{cases}$$

and $\mathbf{Z} = \{Z_{ij}; i = 1, \ldots, I, j = 1, \ldots, m\}$. Similarly, let $\boldsymbol{\pi} = \{\pi_j; j = 1, \ldots, m\}$, $\boldsymbol{\beta} = \{\boldsymbol{\beta}_j; j = 1, \ldots, m\}$, $\boldsymbol{\Psi} = \{\boldsymbol{\Psi}_j; j = 1, \ldots, m\}$, $\boldsymbol{\sigma}^2 = \{\sigma_j^2; j = 1, \ldots, m\}$, and $\boldsymbol{\nu} = \{\nu_j; j = 1, \ldots, m\}$.

In our problem, $\mathbf{y}$ is the observed response vector, while $(\mathbf{b}, \boldsymbol{\tau}, \mathbf{Z})$ can be viewed as the missing data. Based on the hierarchical model formulation in (2.4), the likelihood of the complete data $(\mathbf{y}, \mathbf{b}, \boldsymbol{\tau}, \mathbf{Z})$ given the covariates $(\mathbf{X}_i, \mathbf{U}_i)$ is,

$$\prod_{i=1}^{I} \prod_{j=1}^{m} \left\{ \pi_j f(\mathbf{y}_i; \mathbf{X}_i \boldsymbol{\beta}_j + \mathbf{U}_i \mathbf{b}_{ij}, \frac{1}{\tau_{ij}} \boldsymbol{\Lambda}_{ij}) f(\mathbf{b}_{ij}; \mathbf{0}, \frac{1}{\tau_{ij}} \boldsymbol{\Psi}_j) g(\tau_{ij}; \frac{\nu_j}{2}, \frac{\nu_j}{2}) \right\}^{Z_{ij}}.$$

It follows that the complete log-likelihood function is

$$\begin{aligned}
&\ell(\boldsymbol{\pi}, \boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\sigma}^2, \boldsymbol{\nu} \mid \mathbf{y}, \mathbf{b}, \boldsymbol{\tau}, \mathbf{Z}) \\
=& \sum_{i=1}^{I} \sum_{j=1}^{m} Z_{ij} \ln(\pi_j) \\
&+ \sum_{i=1}^{I} \sum_{j=1}^{m} Z_{ij} \left\{ -\frac{1}{2} \ln |\frac{1}{\tau_{ij}} \sigma_j^2 \mathbf{R}_i| - \frac{1}{2} \mathbf{E}_{ij}^T \left( \frac{1}{\tau_{ij}} \sigma_j^2 \mathbf{R}_i \right)^{-1} \mathbf{E}_{ij} + \text{const} \right\} \\
&+ \sum_{i=1}^{I} \sum_{j=1}^{m} Z_{ij} \left\{ -\frac{1}{2} \ln |\frac{1}{\tau_{ij}} \boldsymbol{\Psi}_j| - \frac{1}{2} (\mathbf{b}_{ij})^T \left[ \frac{1}{\tau_{ij}} \boldsymbol{\Psi}_j \right]^{-1} \mathbf{b}_{ij} + \text{const} \right\} \\
&+ \sum_{i=1}^{I} \sum_{j=1}^{m} Z_{ij} \left\{ \left( \frac{\nu_j}{2} - 1 \right) \ln(\tau_{ij}) - \frac{\tau_{ij}}{2} \nu_j - \ln \left( \Gamma \left( \frac{\nu_j}{2} \right) \right) + \frac{\nu_j}{2} \ln \left( \frac{\nu_j}{2} \right) \right\},
\end{aligned}$$

where $\mathbf{E}_{ij} = \mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_j - \mathbf{U}_i \mathbf{b}_{ij}$, and we have adopted the setting that $\boldsymbol{\Lambda}_{ij} = \sigma_j^2 \mathbf{R}_i$. Based on the idea of ECM algorithm, we shall separate the above log-likelihood function into

four parts, based on the parameters involved, i.e., let

$$\begin{aligned}\ell(\boldsymbol{\pi}, \boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\sigma}^2, \boldsymbol{\nu} \mid \mathbf{y}, \mathbf{b}, \boldsymbol{\tau}, \mathbf{Z}) =& \ell_0(\boldsymbol{\pi} \mid \mathbf{y}, \mathbf{b}, \boldsymbol{\tau}, \mathbf{Z}) + \ell_1(\boldsymbol{\beta}, \boldsymbol{\sigma}^2 \mid \mathbf{y}, \mathbf{b}, \boldsymbol{\tau}, \mathbf{Z}) \\ &+ \ell_2(\boldsymbol{\Psi} \mid \mathbf{y}, \mathbf{b}, \boldsymbol{\tau}, \mathbf{Z}) + \ell_3(\boldsymbol{\nu} \mid \mathbf{y}, \mathbf{b}, \boldsymbol{\tau}, \mathbf{Z}),\end{aligned}$$

where

$$\begin{aligned}\ell_0(\boldsymbol{\pi} \mid \mathbf{y}, \mathbf{b}, \boldsymbol{\tau}, \mathbf{Z}) =& \sum_{i=1}^{I}\sum_{j=1}^{m} Z_{ij} \ln(\pi_j), \\ \ell_1(\boldsymbol{\beta}, \boldsymbol{\sigma}^2 \mid \mathbf{y}, \mathbf{b}, \boldsymbol{\tau}, \mathbf{Z}) =& \sum_{i=1}^{I}\sum_{j=1}^{m} Z_{ij}\left( \left\{ -\frac{n_i}{2}\ln\sigma_j^2 - \frac{\tau_{ij}}{2\sigma_j^2}\mathbf{E}_{ij}^T \boldsymbol{R_i}^{-1}\mathbf{E}_{ij} \right\} \right) \\ =& -\sum_{i=1}^{I}\sum_{j=1}^{m} Z_{ij}\frac{n_i}{2}\ln\sigma_j^2 \\ &- \sum_{i=1}^{I}\sum_{j=1}^{m} Z_{ij}\left[ \frac{\tau_{ij}}{2\sigma_j^2}\mathrm{tr}\left\{ \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{U}_i\mathbf{b}_{ij})(\mathbf{y}_i - \mathbf{U}_i\mathbf{b}_{ij})^T \right\} \right] \\ &+ \sum_{i=1}^{I}\sum_{j=1}^{m} Z_{ij}\left\{ \frac{\tau_{ij}}{\sigma_j^2}\boldsymbol{\beta}_j^T\mathbf{X}_i^T\mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{U}_i\mathbf{b}_{ij}) \right\} \\ &- \sum_{i=1}^{I}\sum_{j=1}^{m} Z_{ij}\left( \frac{\tau_{ij}}{2\sigma_j^2}\boldsymbol{\beta}_j^T\mathbf{X}_i^T\mathbf{R}_i^{-1}\mathbf{X}_i\boldsymbol{\beta}_j \right), \\ \ell_2(\boldsymbol{\Psi} \mid \mathbf{y}, \mathbf{b}, \boldsymbol{\tau}, \mathbf{Z}) =& \sum_{i=1}^{I}\sum_{j=1}^{m} Z_{ij}\left( -\frac{1}{2}\ln|\boldsymbol{\Psi}_j| \right) - \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{m} Z_{ij}\left( \tau_{ij}\mathbf{b}_{ij}^T\boldsymbol{\Psi}_j^{-1}\mathbf{b}_{ij} \right) \\ =& -\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{m} Z_{ij}\ln|\boldsymbol{\Psi}_j| - \frac{1}{2}\mathrm{tr}\left( \boldsymbol{\Psi}_j^{-1}\sum_{i=1}^{I}\sum_{j=1}^{m} Z_{ij}\tau_{ij}\mathbf{b}_{ij}\mathbf{b}_{ij}^T \right),\end{aligned}$$

and

$$\ell_3(\boldsymbol{\nu} \mid \mathbf{y}, \mathbf{b}, \boldsymbol{\tau}, \mathbf{Z}) = \sum_{i=1}^{I}\sum_{j=1}^{m} Z_{ij}\left[ \left\{ \frac{\nu_j}{2}\left( \ln(\frac{\nu_j}{2}) + \ln(\tau_{ij}) - \tau_{ij} \right) - \ln(\tau_{ij}) - \ln\left( \Gamma(\frac{\nu_j}{2}) \right) \right\} \right].$$

Let $\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\sigma}^2, \boldsymbol{\nu})$, collecting all the unknown parameters. Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, we now derive the expected complete data log-likelihood, $E\{\ell(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{b}, \boldsymbol{\tau}, \mathbf{Z}) \mid \mathbf{y}, \widehat{\boldsymbol{\theta}}\}$, with respect

to the missing data $(\mathbf{b}, \boldsymbol{\tau}, \mathbf{Z})$ and conditional on the observed data $\mathbf{y}$. This simplifies to the calculations of the following quantities,

$$
\begin{aligned}
p_{ij} &= E(Z_{ij} = 1 \mid \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{y}), \\
\widehat{\tau}_{ij} &= E(\tau_{ij} \mid \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{y}, Z_{ij} = 1), \\
\widehat{\mathbf{b}}_{ij} &= E(\mathbf{b}_{ij} \mid \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{y}, Z_{ij} = 1, \tau_{ij}), \\
\widehat{\boldsymbol{\Omega}}_{ij} &= \tau_{ij} \mathrm{cov}(\mathbf{b}_{ij} \mid \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{y}, Z_{ij} = 1, \tau_{ij}).
\end{aligned}
$$

From (2.2), it is easy to show that

$$
p_{ij} = \frac{\pi_j t_{n_i}(\mathbf{y}_i; \mathbf{X}_i \widehat{\boldsymbol{\beta}}_j, \mathbf{U}_i \widehat{\boldsymbol{\Psi}}_j \mathbf{U}_i^T, \widehat{\nu}_j)}{\sum_{j=1}^m \pi_j t_{n_i}(\mathbf{y}_i; \mathbf{X}_i \widehat{\boldsymbol{\beta}}_j, \mathbf{U}_i \widehat{\boldsymbol{\Psi}}_j \mathbf{U}_i^T, \widehat{\nu}_j)}. \tag{2.7}
$$

By (2.6) we have

$$
\widehat{\tau}_{ij} = \frac{\widehat{\nu}_j + n_i}{\widehat{\nu}_j + \delta_{ij}^2(\boldsymbol{\beta}_j, \widehat{\boldsymbol{\Psi}}_j, \widehat{\sigma}_j^2)}, \tag{2.8}
$$

where $\delta_{ij}^2(\boldsymbol{\beta}_j, \widehat{\boldsymbol{\Psi}}_j, \widehat{\sigma}_j^2)$ is defined as in (2.5). Next, based on the assumed multivariate $t$ model (2.1) and the normal scale mixture representation,

$$
\mathbf{b}_{ij} \mid \mathbf{y}_i, Z_{ij} = 1, \tau_{ij} \sim N_q\bigg( \mathbf{A}(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_j), \frac{1}{\tau_{ij}} (\boldsymbol{\Psi}_j - \mathbf{A}\mathbf{U}_i \boldsymbol{\Psi}_j) \bigg),
$$

where $\mathbf{A} = \boldsymbol{\Psi}_j \mathbf{U}_i^T (\mathbf{U}_i \boldsymbol{\Psi}_j \mathbf{U}_i^T + \sigma_j^2 \mathbf{R}_i)^{-1}$. It follows that

$$
\begin{aligned}
\widehat{\mathbf{b}}_{ij} &= \widehat{\boldsymbol{\Psi}}_j \mathbf{U}_i^T \bigg( \mathbf{U}_i \widehat{\boldsymbol{\Psi}}_j \mathbf{U}_i^T + \widehat{\sigma}_j^2 \mathbf{R}_i \bigg)^{-1} (\mathbf{y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}_j) \\
&= \bigg( \widehat{\boldsymbol{\Psi}}_j^{-1} + \frac{1}{\widehat{\sigma}_j^2} \mathbf{U}_i^T \mathbf{R}_i^{-1} \mathbf{U}_i \bigg)^{-1} \frac{1}{\widehat{\sigma}_j^2} \mathbf{U}_i^T \mathbf{R}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}_j),
\end{aligned} \tag{2.9}
$$

and

$$\widehat{\mathbf{\Omega}}_{ij} = \widehat{\mathbf{\Psi}}_j - \widehat{\mathbf{\Psi}}_j \mathbf{U}_i^T (\mathbf{U}_i \widehat{\mathbf{\Psi}}_j \mathbf{U}_i^T + \widehat{\sigma}_j^2 \mathbf{R}_i)^{-1} \mathbf{U}_i \widehat{\mathbf{\Psi}}_j = \left( \widehat{\mathbf{\Psi}}_j^{-1} + \frac{1}{\widehat{\sigma}_j^2} \mathbf{U}_i^T \mathbf{R}_i^{-1} \mathbf{U}_i \right)^{-1}. \tag{2.10}$$

With all the above results, we have

$$E\left(\ell_0(\boldsymbol{\pi} \mid \mathbf{y}, \mathbf{b}, \boldsymbol{\tau}, \mathbf{Z}) \mid \mathbf{y}, \widehat{\boldsymbol{\theta}}\right) = \sum_{i=1}^{I} \sum_{j=1}^{m} p_{ij} \ln(\pi_j), \tag{2.11}$$

$$\begin{aligned} &E\left(\ell_1(\boldsymbol{\beta}, \boldsymbol{\sigma}^2 \mid \mathbf{y}, \mathbf{b}, \boldsymbol{\tau}, \mathbf{Z}) \mid \mathbf{y}, \widehat{\boldsymbol{\theta}}\right) \\ &= -\sum_{i=1}^{I} \sum_{j=1}^{m} p_{ij} \frac{n_i}{2} \ln \sigma_j^2 - \sum_{i=1}^{I} \sum_{j=1}^{m} p_{ij} \frac{1}{2\sigma_j^2} \mathrm{tr}\left[ \mathbf{R}_i^{-1} \left\{ \mathbf{U}_i \widehat{\mathbf{\Omega}}_{ij} \mathbf{U}_i^T + \widehat{\tau}_{ij} (\mathbf{y}_i - \mathbf{U}_i \widehat{\mathbf{b}}_{ij})(\mathbf{y}_i - \mathbf{U}_i \widehat{\mathbf{b}}_{ij})^T \right\} \right] \\ &\quad + \sum_{i=1}^{I} \sum_{j=1}^{m} p_{ij} \frac{1}{\sigma_j^2} \widehat{\tau}_{ij} \boldsymbol{\beta}_j^T \mathbf{X}_i^T \mathbf{R}_i^{-1} (\mathbf{y}_i - \mathbf{U}_i \widehat{\mathbf{b}}_{ij}) - \sum_{i=1}^{I} \sum_{j=1}^{m} p_{ij} \frac{1}{2\sigma_j^2} \widehat{\tau}_{ij} \boldsymbol{\beta}_j^T \mathbf{X}_i^T \mathbf{R}_i^{-1} \mathbf{X}_i \boldsymbol{\beta}_j, \end{aligned} \tag{2.12}$$

$$\begin{aligned} &E\left(\ell_2(\mathbf{\Psi} \mid \mathbf{y}, \mathbf{b}, \boldsymbol{\tau}, \mathbf{Z}) \mid \mathbf{y}, \widehat{\boldsymbol{\theta}}\right) \\ &= -\frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{m} p_{ij} \ln |\mathbf{\Psi}_j| - \frac{1}{2} \mathrm{tr}\left\{ \mathbf{\Psi}_j^{-1} \sum_{i=1}^{I} \sum_{j=1}^{m} p_{ij} (\widehat{\tau}_{ij} \widehat{\mathbf{b}}_{ij} \widehat{\mathbf{b}}_{ij}^T + \widehat{\mathbf{\Omega}}_{ij}) \right\}, \end{aligned} \tag{2.13}$$

and

$$\begin{aligned} E\left(\ell_3(\boldsymbol{\nu} \mid \mathbf{y}, \mathbf{b}, \boldsymbol{\tau}, \mathbf{Z}) \mid \mathbf{y}, \widehat{\boldsymbol{\theta}}\right) = \sum_{i=1}^{I} \sum_{j=1}^{m} p_{ij} \Bigg[ &\frac{\nu_j}{2} \left\{ \ln(\frac{\nu_j}{2}) + E[\ln(\tau_{ij}) \mid \mathbf{y}, \widehat{\boldsymbol{\theta}}, Z_{ij} = 1] - \widehat{\tau}_{ij} \right\} \\ &- E[\ln(\tau_{ij}) \mid \mathbf{y}, \widehat{\boldsymbol{\theta}}, Z_{ij} = 1] - \ln\{\Gamma(\frac{\nu_j}{2})\} \Bigg]. \end{aligned} \tag{2.14}$$

Following Khodabina and Alireza (2010) and based on properties of generalized Gamma distribution,

$$E\left(\ln(\tau_{ij}) \mid \mathbf{y}, \widehat{\boldsymbol{\theta}}, Z_{ij} = 1\right) = \ln \widehat{\tau}_{ij} + \left\{ \psi\left(\frac{\nu_j + n_i}{2}\right) - \ln\left(\frac{\nu_j + n_i}{2}\right) \right\},$$

where

$$\psi\left(\frac{\nu_j + n_i}{2}\right) = \frac{\partial \Gamma(\frac{\nu_j + n_i}{2})}{\partial(\frac{\nu_j + n_i}{2})} / \Gamma(\frac{\nu_j + n_i}{2}).$$

Now we are ready to fully describe our proposed ECM algorithm (Meng and Rubin 1993)for conducting maximum likelihood estimation.

**Initialization**: Set $k = 0$; obtain some initial estimates of the parameters $\boldsymbol{\theta}^{(0)}$, including $\pi_j^{(0)}$, $\boldsymbol{\beta}_j^{(0)}$, $\boldsymbol{\Psi}_j^{(0)}$, $\nu_j^{(0)}$, and $\sigma_j^{2(0)}$, for $j = 1, \ldots, m$.

**Initial values:** There are many ways to find the initial values for $\{\pi_j^{(0)}, \boldsymbol{\beta}_j^{(0)}, \sigma^{(0)}, j = 1, \ldots, m\}$. One method is to use trimmed likelihood estimates (TLE) (Neykov, et al. 2007). Note that the TLE is robust to both low leverage and high leverage outliers under certain general conditions (Neykov, et al. 2007). Another possible method is that we first randomly partition the data or a subset of the data into $m$ groups. For each group, we use some robust regression method, such as the MM-estimate (Yohai, 1987), to estimate the component regression parameters. Similar partition ideas have been used to find the initial values for finite mixture models (McLachlan and Peel, 2000). In addition, we can also apply the robust linear clustering method to find the initial regression parameter values. See, for example, Hennig (2002, 2003), and García-Escudero, et al. (2009). Note that though, technically, the robust linear clustering methods do not produce consistent regression component estimators. But in many cases, they are close enough to provide good initial values, since the proposed algorithm doesn't require the initial values to be consistent.

**E-step**: At $(k+1)^{th}$ iteration, given $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$, compute $p_{ij}^{(k+1)}, \mathbf{b}_{ij}^{(k+1)}, \tau_{ij}^{(k+1)}$ and $\boldsymbol{\Omega}_{ij}^{(k+1)}$ based on (2.7), (2.9), (2.8) and (2.10), respectively, for $i = 1, \ldots, I$ and $j = 1, \ldots, m$. Subsequently, the four components of the expected complete log-likelihood can be constructed from (2.11), (2.12), (2.13), and (2.14), respectively.

**CM-step**:

**M-0:** Obtain $\pi_j^{(k+1)}$, $j = 1, \ldots, m$, by maximizing $E\left(\ell_0(\boldsymbol{\pi} \mid \mathbf{y}, \mathbf{b}, \boldsymbol{\tau}, \mathbf{Z}) \mid \mathbf{y}, \boldsymbol{\theta}^{(k)}\right)$, with respect to $\boldsymbol{\pi}$,

$$\pi_j^{(k+1)} = \frac{1}{I} \sum_{i=1}^{I} p_{ij}^{(k+1)}.$$

**M-1:** Given $\sigma_j^2 = \sigma_j^{2(k)}$, $j = 1, \ldots, m$, obtain $\boldsymbol{\beta}_j^{(k+1)}$, $j = 1, \ldots, m$, by maximizing $E\left(\ell_1(\boldsymbol{\beta}, \boldsymbol{\sigma}^{2(k)} \mid \mathbf{y}, \mathbf{b}, \boldsymbol{\tau}, \mathbf{Z}) \mid \mathbf{y}, \boldsymbol{\theta}^{(k)}\right)$ with respect to $\boldsymbol{\beta}$,

$$\boldsymbol{\beta}_j^{(k+1)} = \left\{ \sum_{i=1}^{I} p_{ij}^{(k+1)} \frac{\tau_{ij}^{(k+1)}}{\sigma_j^{2(k)}} \mathbf{X}_i^T \mathbf{R}_i^{-1} \mathbf{X}_i \right\}^{-1} \left\{ \sum_{i=1}^{I} p_{ij}^{(k+1)} \frac{\tau_{ij}^{(k+1)}}{\sigma_j^{2(k)}} \mathbf{X}_i^T \mathbf{R}_i^{-1} (\mathbf{y}_i - \mathbf{U}_i \mathbf{b}_{ij}^{(k+1)}) \right\}.$$

**M-2:** Given $\boldsymbol{\beta}_j = \boldsymbol{\beta}_j^{(k+1)}$, $j = 1, \ldots, m$, obtain $\sigma_j^{2(k+1)}$, $j = 1, \ldots, m$, by maximizing $E\left(\ell_1(\boldsymbol{\beta}^{(k+1)}, \boldsymbol{\sigma}^2 \mid \mathbf{y}, \mathbf{b}, \boldsymbol{\tau}, \mathbf{Z}) \mid \mathbf{y}, \boldsymbol{\theta}^{(k)}\right)$ with respect to $\boldsymbol{\sigma}^2$,

$$\sigma_j^{2(k+1)} = \frac{\sum_{i=1}^{I} p_{ij}^{(k+1)} \left\{ \tau_{ij}^{(k+1)} \mathbf{E}_{ij}^T \mathbf{R}_i^{-1} \mathbf{E}_{ij} + \mathrm{tr}(\boldsymbol{\Omega}_{ij}^{(k+1)} \mathbf{U}_i^T \mathbf{R}_i^{-1} \mathbf{U}_i) \right\}}{\sum_{i=1}^{I} p_{ij}^{(k+1)} n_i},$$

**M-3:** Obtain $\boldsymbol{\Psi}_j^{(k+1)}$, $j = 1, \ldots, m$, by maximizing $E\left(\ell_2(\boldsymbol{\Psi} \mid \mathbf{y}, \mathbf{b}, \boldsymbol{\tau}, \mathbf{Z}) \mid \mathbf{y}, \boldsymbol{\theta}^{(k)}\right)$ with respect to $\boldsymbol{\Psi}$,

$$\boldsymbol{\Psi}_j^{(k+1)} = \frac{\sum_{i=1}^{I} p_{ij}^{(k+1)} (\tau_{ij}^{(k+1)} \mathbf{b}_{ij}^{(k+1)} (\mathbf{b}_{ij}^{(k+1)})^T + \boldsymbol{\Omega}_{ij}^{(k+1)})}{\sum_{i=1}^{I} p_{ij}^{(k+1)}}.$$

**M-4:** Obtain $\nu_j^{(k+1)}$, $j = 1, \ldots, m$, by maximizing $E\left(\ell_3(\boldsymbol{\nu} \mid \mathbf{y}, \mathbf{b}, \boldsymbol{\tau}, \mathbf{Z}) \mid \mathbf{y}, \boldsymbol{\theta}^{(k)}\right)$ with respect to $\boldsymbol{\nu}$.

$$\nu_j^{(k+1)} = \arg\max_{\nu_j} \left[ \sum_{i=1}^{I} p_{ij}^{(k+1)} \frac{\nu_j}{2} \left\{ \ln(\frac{\nu_j}{2}) + E[\ln(\tau_{ij}) \mid \mathbf{y}, \boldsymbol{\theta}^{(k)}, Z_{ij} = 1] - \tau_{ij}^{(k+1)} \right\} - \ln\{\Gamma(\frac{\nu_j}{2})\} \right].$$

The problem is separable in each $\nu_j$. Although these one-dimensional problems do not admit explicit solutions, they can be solved by numerical optimization methods, e.g., the Newton-Raphson algorithm or the secant method. However, we find that the above approach may not be always stable, partly due to the high nonlinearity of the objective function. Alternatively, we can replace M-4 by carrying out constrained estimation of the actual log-likelihood (2.3) with respect to the unknown degrees of freedom parameters, with all the other parameters held fixed at their currently updated values (Pinheiro et al.,

2001).

**M-4*:** Obtain $\nu_j^{(k+1)}$, $j = 1, \ldots, m$, by maximizing (2.3) with respect to $\boldsymbol{\nu}$, with $\boldsymbol{\pi} = \boldsymbol{\pi}^{(k+1)}$, $\boldsymbol{\beta} = \boldsymbol{\beta}^{(k+1)}$, $\boldsymbol{\Psi} = \boldsymbol{\Psi}^{(k+1)}$, and $\boldsymbol{\sigma}^2 = \boldsymbol{\sigma}^{2(k+1)}$.

In the case that $\nu_j = \nu$, $j = 1, \ldots, m$, it is convenient to use a profile likelihood approach to avoid either M-4 or M-4* step entirely in the ECM algorithm, i.e., conduct maximum likelihood estimation with $\nu$ held fixed, for a grid of $\nu$ values, say, $\nu = 1, \ldots, 20$, and then the final estimate of the degrees of freedom is selected as the one that gives the largest log-likelihood.

In the M-step, we do not aim to fully maximize the expected log-likelihood, as it requires iteratively solving M-1 and M-2, which may be computationally inefficient. Nevertheless, solving each of the five subproblems once in the M-step monotonically increases the expected log-likelihood, which implies that the stable monotone convergence property of the ECM algorithm is preserved. The E-step and M-step are carried out alternatingly, until convergence is reached, i.e., the log-likelihood function (2.3) stops increasing up to a small tolerance value. Based on our limited experience, the proposed algorithm works well in terms of both computational stability and efficiency.

# 3 Simulation Study

We generate data from the following model

$$\mathbf{y}_i = \begin{cases} \mathbf{X}_i\boldsymbol{\beta}_1 + \mathbf{U}_i\mathbf{b}_{i1} + \mathbf{e}_{i1}, & \text{if } Z_i = 1; \\ \mathbf{X}_i\boldsymbol{\beta}_2 + \mathbf{U}_i\mathbf{b}_{i2} + \mathbf{e}_{i2}, & \text{if } Z_i = 2, \end{cases}$$

where $i = 1, \ldots, I$, $\boldsymbol{\beta}_1 = (1, 1, 0, 0)^T$, $\boldsymbol{\beta}_2 = (0, 0, 1, 1)^T$, and $\pi_1 = P(Z_i = 1) = 0.4$. The rows of the covariates $\mathbf{X}_i \in \mathbb{R}^{n_i \times 4}$ are independently generated from $N_4(\mathbf{0}, \mathbf{I})$. The rows of $\mathbf{U}_i \in \mathbb{R}^{n_i \times 2}$ are independently generated from $N_2(\mathbf{0}, \mathbf{I})$.

We consider the following three types of random effects and error distributions.

1. $t$ distribution: $\mathbf{e}_{ij} \sim t_{n_i}(\mathbf{0}, \boldsymbol{\Lambda}_{ij}, \nu)$, $\mathbf{b}_{ij} \sim t_q(\mathbf{0}, \boldsymbol{\Psi}_j, \nu)$, and given $\tau_{ij}$, $\mathbf{b}_{ij}$ and $\mathbf{e}_{ij}$

are conditionally independent. That is, $\mathbf{b}_{ij} \mid \tau_{ij} \sim N(0, \frac{1}{\tau_{ij}}\mathbf{\Psi}_j)$ and $\mathbf{e}_{ij} \mid \tau_{ij} \sim N(0, \frac{1}{\tau_{ij}}\mathbf{\Lambda}_{ij})$. We set $\mathbf{\Lambda}_{ij}$ as an identity matrix and $\mathbf{\Psi}_j$ as a diagonal matrix with diagonal elements 1 and off-diagonal elements 0.5. We consider three degrees of freedom values, i.e., $\nu \in \{1, 3, 5\}$.

2. Normal distribution: $\mathbf{e}_{ij} \sim N_{n_i}(\mathbf{0}, \mathbf{\Lambda}_{ij})$ and $\mathbf{b}_{ij} \sim N_q(\mathbf{0}, \mathbf{\Psi}_j)$, where we set $\mathbf{\Lambda}_{ij}$ as an identity matrix and $\mathbf{\Psi}_j$ as a diagonal matrix with diagonal elements 1 and off-diagonal elements 0.5.

3. Contaminated normal distribution: $\mathbf{e}_{ij} \sim 0.95N_{n_i}(\mathbf{0}, \mathbf{I}) + 0.05N_{n_i}(\mathbf{0}, 25\mathbf{I})$ and $\mathbf{b}_{ij} \sim 0.95N_q(\mathbf{0}, \mathbf{I}) + 0.05N_q(\mathbf{0}, 25\mathbf{I})$.

We have experimented with various sample sizes and numbers of replicates per sample. In particular, the following four cases are considered herein,

Case 1: $n_i = 8$, $I = 100$.

Case 2: $n_i = 8$, $I = 200$.

Case 3: $n_i = 4$, $I = 200$.

Case 4: $n_i = 4$, $I = 400$.

The simulation is replicated 500 times under each setting.

We mainly compare our proposed robust $t$-mixture method to the normal mixture mixed effect approach. Our implemented EM algorithm can be readily modified to fit the normal mixture model; a more straightforward approach is to use our $t$-mixture EM algorithm with the degrees of freedom parameters held fixed at large number values, say, 1000, so that the $t$ distribution becomes very close to normal. Similar to Bordes et al. (2007) and Hunter and Young (2012), we use the true parameter values as the initial values to start the EM algorithm, in order to avoid the possible bias introduced by different starting values among replications or label switching issues (Celeux, et al., 2000; Stephens, 2000; Yao and Lindsay, 2009), so as to compare the "best-case" results of the

various estimation methods. The degrees of freedom estimates in the $t$-mixture model is determined based on the aforementioned profile likelihood approach.

In Table 1, we report the average degrees of freedom estimates using the $t$-mixture model under the aforementioned five mixed effect and error structures. As the tail of the assumed mixed effect and error distribution becomes heavier, the estimated degrees of freedom becomes smaller on average as expected. Therefore, the proposed approach captures the tail behavior of the mixed effect and error distributions quite well.

In Tables 2–5, we report the median squared errors (MedSE) for parameter estimates and the relative efficiencies of our proposed $t$-mixture method as compared to the conventional normal mixture model. In Figures 1–2, we also show the MedSE for some of parameter estimates for cases 1 and 2. Our $t$-mixture approach works very well and consistently outperforms the normal mixture model when the random effects and error distributions are of heavy tail or are contaminated by outliers. Even when the random effects and the error terms follow normal distribution, the performance of the $t$-mixture model is comparable to that of the normal mixture model. This is essentially because the latter method can be treated as a special case of our proposed robust $t$-mixture model, and thus the efficiency loss is minimal when no outlier presents in the data. When the true model has $t$-distributed random effects and errors, the relative efficiency estimates may be very high. This is because the normal mixture model may fail miserably when applied to heavy-tailed Cauchy or close-to-Cauchy distributions.

# 4 Lung Growth Data Analysis

We consider a longitudinal dataset on lung growth of girls, from a study of air pollution and health in six cities across the U.S.; see Dockery et al. (1983) for the details of the study. Here we focus on the records gathered from Topeka, Kansas. The lung growth status of 300 girls in Topeka were tracked. Most of them were enrolled in the first or second grade and between the ages of six and seven, and measurements of participants

were obtained annually until graduation from high school or loss to follow-up (Dockery et al., 1983). We have omitted the subjects with only one record, and now the number of observations gathered on each of the remaining 252 subjects over time ranges from 2 to 12.

We use the logarithmic forced expiratory volume in one second (fev1) as the response variable. Specially, this variable measures the volume of air that can be forcibly exhaled from the lungs in the first second of a forced expiratory maneuver, and it is critically important in the diagnosis of obstructive and restrictive diseases and is a commonly-used measure of lung function from the pulmonary function tests. We are interested in modeling the lung growth pattern over time, and thus the age variable is used as both the fixed-effect covariate and the random-effect covariate. It is also of great interest to investigate whether the subjects form several distinct clusters or groups that exhibit different behaviors on lung growth. We thus fit the data based on the traditional normal mixture of linear mixed models and the proposed robust $t$-mixture of linear mixed models. Following Heinzl et al. (2013), a three-component mixture model is used.

Table 6 shows the estimated parameters. $\widehat{\beta}_{0i}$'s and $\widehat{\beta}_{1i}$'s are relative intercepts and slopes in the mixture of linear mixed model with response "fev1" and predictor "age". Following the three components in the original paper, we got $\widehat{\beta}_{0i}$, $\widehat{\beta}_{1i}$, where $i = 1, 2, 3$. and two probabilities $\widehat{\pi}$, where $\widehat{\pi}_3 = 1 - \widehat{\pi}_1 - \widehat{\pi}_2$ . Based on the profile likelihood approach, the degrees of freedom of the $t$-mixture model is estimated to be $\widehat{\nu} = 28$, which is quite large. This result suggests that the random effects and the errors may be approximately normally distributed in this application. To test our robust estimation approach, however, we add some artificial outliers for some arbitrarily selected subjects in the dataset and refit the $t$-mixture model. Using contaminated datasets with outliers in *one* subject (add 10 to the response (fev1) values in 1st subject.), the estimated degrees of freedom is $\nu = 9$, and using the contaminated datasets with outliers in *two* subjects (add 10 to the response (fev1) values in 1st and 2nd subjects.), the estimate becomes $\nu = 6$. The decrease in the estimated degrees of freedom as the number of outliers increases clearly demonstrates

the robustness of the proposed approach. In addition, compared to the estimates of the traditional normal mixture of linear mixed models, the parameter estimates for the new method does not change much when the outliers are added into the data set.

Our analysis based on original datareveals some interesting cluster structure. In Figure 3, three distinct groups can be clearly distinguished by the intercept and slope estimates bases on the mixed effects. Girls assigned to different clusters are marked with different colors and symbols. It appears that cluster 1 (blue, triangle) consists of the girls who had initial low-level lung function and then experienced relatively fast lung growth to their adulthood. In contrast, cluster 2 (red, circle) consists of the girls who had relatively high level of initial lung development and then experienced relatively slow lung growth to their adulthood. Cluster 3 (black, cross) is the smallest cluster of the three, which appears to consist of the girls who had relatively low level of initial lung development and also experienced relatively slow lung growth over time.

# 5 Discussion

In this article, we have proposed a robust mixture linear mixed model, using multivariate $t$ distribution to robustify the model estimation and inference. An ECM algorithm is proposed to maximize the mixture likelihood. The simulation study and real data application demonstrated that the proposed method has comparable performance to normal based method when there are no outliers but has much better performance when the error has heavy tail or there are outliers.

However, based on our limited empirical experience, the estimates of degrees of freedom are not very accurate. Its consistency might require much larger sample size than we used in our simulation study. Although Hennig (2004) and Yao et al. (2014) pointed out that the mixture of t-distributio has a very small breakdown point, the breakdown occurs only when the outliers are very extreme. Therefore, the t-distribution has been widely used to provide a robust estimation for mixture models (Peel and McLachlan, 2000).

It is of interest to extend the proposed model to other distributions that possess certain robustness properties, e.g., mixture of Laplace distributed mixed effects and random errors. Note that the proposed method can only handle moderate outliers in $y$ direction and is not robust to outliers in $x$ direction. It is worthwhile to apply the trimmed-likelihood idea to the mixture linear mixed model setups. The recently developed penalized estimation approaches may also be adopted to directly capture and accommodate potential outliers.

# Acknowledgements

The authors thank the editor, the associate editor, and reviewers for their constructive comments that have led to a dramatic improvement of the earlier version of this article. Yao's research is supported by NSF grant DMS-1461677.

# References

Bordes, L., Chauveau, D., and Vandekerkhove, P. (2007). A stochastic EM algorithm for a semiparametric mixture model. *Computational Statistics and Data Analysis*, 51, 5429-5443

Celeux, G., Hurn, M., and Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, 95, 957-970.

Celeux, G., Martin, O., and Lavergne, Ch. (2005). Mixture of linear mixed models for clustering gene expression profiles from repeated microarray experiments. *Statistical Modeling*, 5, 1-25.

Dockery, D. W., Berkery, C. S., Ware, J. H., Speizer, F. E., and Ferris, B. G. (1983). Dis-

tribution of fvc and fev1 in children 6 to 11 years old. *American Review of Respiratory Disease*, 128, 405-12.

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52(2): 399-433.

Hartley, H. O. and Rao, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika*, 54, 93108.

Heinzl, F., and Tutz, G. (2013). Clustering in linear mixed models with approximate Dirichlet process mixtures using EM algorithm. *Statistical Modeling*, 13, 41-67.

Henderson, C. R., Kempthorne, O., Searle, S. R., and von Krosigk, C. M. (1959). The estimation of environmental and genetic trends from records subject to culling. *International Biometric Society*, 15(2): 192-218.

Hunter, D. R., and Young, D. S. (2012). Semiparametric mixtures of regressions. *Journal of Nonparametric Statistics*, 24 (1): 19-38

Khodabina, M., and Alireza, A. (2010). Some properties of generalized gamma distribution. *Mathematical Sciences*, 4, 9-28.

Lange, K. L., Little, R. J. A., and Taylor, J. M. G. (1989). Robust statitical modeling using the t distribution. *Journal of the American Statistical Association*, 84, 881-896.

Mclean, R. A., Sanders, W. L., Stroup, W. W. (1991). A unified approach to mixed linear models. *The American Statistician*, 45(1): 54-64.

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

Ng, S. K., Mclachlan, G. J., Wang, K., Jones, L. B. T , and Ng, S.-W. (2006). A mixture model with random-effects components for clustering correlated gene-expression profiles. *Bioinformatrics*, 22, 1745-1752.

Peel, D. and McLachlan, G. J. (2000). Robust mixture modelling using the t distribution *Statistics and Computing*, 10, 339-348.

Pinheiro, J. C., Liu, CH. H., 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 (2), 249-276.

Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical Science*, 6, 15-32.

Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of Royal Statistical Society*, B62, 795-809.

Yau, K. K. W., Lee, A. H., and Ng, S. K. (2002). Finite mixture regression model with random effects: application to neonatal hospital length of stay. *Computational Statistics & Data Analysis*, 41, 359-366.

Welsh, A. H. and Richardson, A. M. (1997). Approaches to the robust estimation of mixed models. *Handbook of Statistics*, Vol. 15 of Maddala, G. S., and Rao, C. R. (1997), chapter 13, 343-384.

Yao, W. and Lindsay, B. G. (2009). Bayesian mixture labeling by highest posterior density, *Journal of American Statistical Association*, 104, 758-767.

Yao, W., Wei, Y., and Yu, C. (2014). Robust Mixture Regression Using T-Distribution. *Computational Statistics and Data Analysis*, 71, 116-127.

| #replicates | #subjects | Estimated degrees of freedom | | | | |
|---|---|---|---|---|---|---|
| | | $t_1$ | $t_3$ | $t_5$ | Normal | Contaminated Normal |
| $n_i = 8$ | $I = 100$ | 1.605 | 5.665 | 6.716 | 12.46 | 4.839 |
| | $I = 200$ | 1.605 | 3.832 | 9.868 | 12.46 | 4.133 |
| $n_i = 4$ | $I = 200$ | 2.575 | 6.149 | 7.199 | 10.65 | 5.665 |
| | $I = 400$ | 1.488 | 5.252 | 7.766 | 12.46 | 3.015 |

Table 1: Degrees of freedom estimation results, based on 500 simulation runs.

| Estimator | | Random Effects and Error Distribution | | | | |
|---|---|---|---|---|---|---|
| | | $t_1$ | $t_3$ | $t_5$ | Normal | Contaminated N. |
| $\widehat{\pi}_1$ | MedSE(NMM) | 0.196 | 0.014 | 0.014 | 0.009 | 0.014 |
| | MedSE($t$MM) | 0.064 | 0.010 | 0.014 | 0.009 | 0.009 |
| | Efficiency | 3.063 | 1.400 | 1.000 | 1.000 | 1.556 |
| $\widehat{\beta}_{11}$ | MedSE(NMM) | 0.265 | 0.106 | 0.005 | 0.004 | 0.010 |
| | MedSE($t$MM) | 0.197 | 0.006 | 0.003 | 0.004 | 0.004 |
| | Efficiency | 1.345 | 17.667 | 1.667 | 1.000 | 2.500 |
| $\widehat{\beta}_{21}$ | MedSE(NMM) | 0.279 | 0.110 | 0.005 | 0.004 | 0.012 |
| | MedSE($t$MM) | 0.216 | 0.007 | 0.004 | 0.004 | 0.004 |
| | Efficiency | 1.292 | 15.714 | 1.250 | 1.000 | 3.000 |
| $\widehat{\beta}_{31}$ | MedSE(NMM) | 0.276 | 0.094 | 0.006 | 0.003 | 0.010 |
| | Median($t$MM) | 0.237 | 0.008 | 0.005 | 0.003 | 0.004 |
| | Efficiency | 1.165 | 11.750 | 1.200 | 1.000 | 2.500 |
| $\widehat{\beta}_{41}$ | MedSE(NMM) | 0.265 | 0.118 | 0.005 | 0.004 | 0.012 |
| | Median($t$MM) | 0.192 | 0.008 | 0.004 | 0.004 | 0.004 |
| | Efficiency | 1.380 | 14.750 | 1.250 | 1.000 | 3.000 |
| $\widehat{\beta}_{12}$ | MedSE(NMM) | 7.871 | 0.014 | 0.005 | 0.003 | 0.011 |
| | MedSE($t$MM) | 0.085 | 0.005 | 0.004 | 0.003 | 0.004 |
| | Efficiency | 92.600 | 1.280 | 1.250 | 1.000 | 2.750 |
| $\widehat{\beta}_{22}$ | MedSE(NMM) | 7.516 | 0.015 | 0.006 | 0.003 | 0.012 |
| | MedSE($t$MM) | 0.078 | 0.004 | 0.004 | 0.003 | 0.004 |
| | Efficiency | 92.600 | 3.750 | 1.500 | 1.000 | 3.000 |
| $\widehat{\beta}_{32}$ | MedSE(NMM) | 7.235 | 0.017 | 0.008 | 0.003 | 0.012 |
| | MedSE($t$MM) | 0.081 | 0.005 | 0.004 | 0.003 | 0.004 |
| | Efficiency | 89.321 | 3.400 | 2.000 | 1.000 | 3.000 |
| $\widehat{\beta}_{42}$ | MedSE(NMM) | 5.869 | 0.017 | 0.006 | 0.003 | 0.012 |
| | MedSE($t$MM) | 0.081 | 0.005 | 0.005 | 0.003 | 0.004 |
| | Efficiency | 72.457 | 3.400 | 1.200 | 1.000 | 3.000 |

Table 2: Simulation results for Case 1: $n_i = 8$, $I = 100$.

| Estimator | | Random Effects and Error Distribution | | | | |
|---|---|---|---|---|---|---|
| | | $t_1$ | $t_3$ | $t_5$ | Normal | Contaminated N. |
| $\widehat{\pi}_1$ | MedSE(NMM) | 0.211 | 0.010 | 0.011 | 0.012 | 0.014 |
| | MedSE($t$MM) | 0.054 | 0.010 | 0.011 | 0.012 | 0.011 |
| | Efficiency | 3.907 | 1.000 | 1.000 | 1.000 | 1.273 |
| $\widehat{\beta}_{11}$ | MedSE(NMM) | 0.289 | 0.041 | 0.004 | 0.001 | 0.004 |
| | MedSE($t$MM) | 0.151 | 0.002 | 0.002 | 0.001 | 0.002 |
| | Efficiency | 1.914 | 20.500 | 2.000 | 1.000 | 2.000 |
| $\widehat{\beta}_{21}$ | MedSE(NMM) | 0.270 | 0.040 | 0.003 | 0.001 | 0.005 |
| | MedSE($t$MM) | 0.139 | 0.002 | 0.002 | 0.002 | 0.002 |
| | Efficiency | 1.942 | 20 | 1.500 | 0.500 | 2.500 |
| $\widehat{\beta}_{31}$ | MedSE(NMM) | 0.266 | 0.034 | 0.003 | 0.001 | 0.005 |
| | MedSE($t$MM) | 0.161 | 0.003 | 0.002 | 0.001 | 0.002 |
| | Efficiency | 1.652 | 11.333 | 1.500 | 1.000 | 2.500 |
| $\widehat{\beta}_{41}$ | MedSE(NMM) | 0.271 | 0.040 | 0.004 | 0.002 | 0.004 |
| | MedSE($t$MM) | 0.155 | 0.002 | 0.002 | 0.002 | 0.002 |
| | Efficiency | 1.748 | 20.000 | 2.000 | 1.000 | 2.000 |
| $\widehat{\beta}_{12}$ | MedSE(NMM) | 7.753 | 0.008 | 0.004 | 0.002 | 0.007 |
| | MedSE($t$MM) | 0.031 | 0.002 | 0.002 | 0.002 | 0.002 |
| | Efficiency | 250.097 | 4.000 | 2.00 | 1.000 | 3.500 |
| $\widehat{\beta}_{22}$ | MedSE(NMM) | 5.797 | 0.008 | 0.004 | 0.001 | 0.008 |
| | MedSE($t$MM) | 0.028 | 0.002 | 0.002 | 0.001 | 0.002 |
| | Efficiency | 207.036 | 4.000 | 2.000 | 1.000 | 4.000 |
| $\widehat{\beta}_{32}$ | MedSE(NMM) | 6.116 | 0.008 | 0.004 | 0.001 | 0.009 |
| | MedSE($t$MM) | 0.035 | 0.002 | 0.002 | 0.002 | 0.002 |
| | Efficiency | 175.029 | 4.000 | 2.000 | 0.500 | 4.500 |
| $\widehat{\beta}_{42}$ | MedSE(NMM) | 6.783 | 0.009 | 0.003 | 0.002 | 0.008 |
| | MedSE($t$MM) | 0.033 | 0.002 | 0.002 | 0.002 | 0.002 |
| | Efficiency | 204.545 | 4.500 | 1.500 | 1.000 | 4.000 |

Table 3: Simulation results for Case 2: $n_i = 8$, $I = 200$.

| Estimator | | Random Effects and Error Distribution | | | | |
|---|---|---|---|---|---|---|
| | | $t_1$ | $t_3$ | $t_5$ | Normal | Contaminated N. |
| $\widehat{\pi}_1$ | MedSE(NMM) | 0.208 | 0.039 | 0.012 | 0.011 | 0.068 |
| | MedSE($t$MM) | 0.079 | 0.013 | 0.012 | 0.011 | 0.008 |
| | Efficiency | 2.633 | 3.000 | 1.000 | 1.000 | 8.500 |
| $\widehat{\beta}_{11}$ | MedSE(NMM) | 0.253 | 0.181 | 0.007 | 0.004 | 0.095 |
| | MedSE($t$MM) | 0.253 | 0.008 | 0.006 | 0.002 | 0.005 |
| | Efficiency | 1.000 | 2.130 | 1.167 | 2.000 | 19.000 |
| $\widehat{\beta}_{21}$ | MedSE(NMM) | 0.228 | 0.201 | 0.006 | 0.003 | 0.104 |
| | MedSE($t$MM) | 0.246 | 0.010 | 0.005 | 0.002 | 0.006 |
| | Efficiency | 0.927 | 20.100 | 1.200 | 1.500 | 5.567 |
| $\widehat{\beta}_{31}$ | MedSE(NMM) | 0.251 | 0.194 | 0.007 | 0.003 | 0.106 |
| | MedSE($t$MM) | 0.250 | 0.009 | 0.005 | 0.002 | 0.005 |
| | Efficiency | 1.008 | 21.556 | 1.400 | 1.500 | 21.200 |
| $\widehat{\beta}_{41}$ | MedSE(NMM) | 0.249 | 0.203 | 0.009 | 0.004 | 0.113 |
| | MedSE($t$MM) | 0.241 | 0.008 | 0.006 | 0.002 | 0.006 |
| | Efficiency | 1.029 | 25.375 | 1.500 | 2.000 | 18.833 |
| $\widehat{\beta}_{12}$ | MedSE(NMM) | 11.846 | 0.335 | 0.007 | 0.004 | 0.043 |
| | MedSE($t$MM) | 0.405 | 0.011 | 0.005 | 0.002 | 0.004 |
| | Efficiency | 29.249 | 30.455 | 1.400 | 2.000 | 10.750 |
| $\widehat{\beta}_{22}$ | MedSE(NMM) | 16.726 | 0.209 | 0.007 | 0.004 | 0.048 |
| | MedSE($t$MM) | 0.443 | 0.009 | 0.006 | 0.002 | 0.004 |
| | Efficiency | 37.756 | 23.222 | 1.167 | 2.000 | 12.000 |
| $\widehat{\beta}_{32}$ | MedSE(NMM) | 15.735 | 0.270 | 0.007 | 0.004 | 0.045 |
| | MedSE($t$MM) | 0.337 | 0.009 | 0.005 | 0.002 | 0.004 |
| | Efficiency | 46.691 | 30.000 | 1.400 | 2.000 | 11.250 |
| $\widehat{\beta}_{42}$ | MedSE(NMM) | 15.323 | 0.275 | 0.008 | 0.003 | 0.035 |
| | MedSE($t$MM) | 0.379 | 0.009 | 0.006 | 0.002 | 0.004 |
| | Efficiency | 40.456 | 30.556 | 1.333 | 1.500 | 8.750 |

Table 4: Simulation results for Case 3: $n_i = 4$, $I = 200$.

| Estimator | | Random Effects and Error Distribution | | | | |
|---|---|---|---|---|---|---|
| | | $t_1$ | $t_3$ | $t_5$ | Normal | Contaminated N. |
| $\widehat{\pi}_1$ | MedSE(NMM) | 0.222 | 0.211 | 0.012 | 0.009 | 0.224 |
| | MedSE($t$MM) | 0.044 | 0.054 | 0.012 | 0.009 | 0.010 |
| | Efficiency | 5.045 | 3.907 | 1.000 | 1.000 | 22.400 |
| $\widehat{\beta}_{11}$ | MedSE(NMM) | 0.275 | 0.289 | 0.007 | 0.002 | 0.177 |
| | MedSE($t$MM) | 0.083 | 0.151 | 0.006 | 0.002 | 0.002 |
| | Efficiency | 3.313 | 1.914 | 1.167 | 1.000 | 88.500 |
| $\widehat{\beta}_{21}$ | MedSE(NMM) | 0.280 | 0.270 | 0.006 | 0.002 | 0.174 |
| | MedSE($t$MM) | 0.084 | 0.139 | 0.005 | 0.002 | 0.002 |
| | Efficiency | 3.333 | 1.942 | 1.200 | 1.000 | 87.000 |
| $\widehat{\beta}_{31}$ | MedSE(NMM) | 0.279 | 0.266 | 0.007 | 0.002 | 0.181 |
| | MedSE($t$MM) | 0.079 | 0.161 | 0.005 | 0.002 | 0.002 |
| | Efficiency | 3.532 | 1.652 | 1.400 | 1.000 | 90.500 |
| $\widehat{\beta}_{41}$ | MedSE(NMM) | 0.276 | 0.271 | 0.009 | 0.002 | 0.180 |
| | MedSE($t$MM) | 0.075 | 0.155 | 0.006 | 0.002 | 0.002 |
| | Efficiency | 3.680 | 1.748 | 1.600 | 1.000 | 90.000 |
| $\widehat{\beta}_{12}$ | MedSE(NMM) | 14.856 | 7.753 | 0.007 | 0.002 | 0.042 |
| | MedSE($t$MM) | 0.024 | 0.031 | 0.005 | 0.002 | 0.002 |
| | Efficiency | 619.000 | 250.097 | 1.400 | 1.000 | 21.000 |
| $\widehat{\beta}_{22}$ | MedSE(NMM) | 17.778 | 5.797 | 0.007 | 0.002 | 0.059 |
| | MedSE($t$MM) | 0.025 | 0.028 | 0.006 | 0.002 | 0.002 |
| | Efficiency | 711.120 | 207.036 | 1.167 | 1.000 | 29.500 |
| $\widehat{\beta}_{32}$ | MedSE(NMM) | 12.837 | 6.116 | 0.007 | 0.002 | 0.043 |
| | MedSE($t$MM) | 0.030 | 0.035 | 0.005 | 0.002 | 0.002 |
| | Efficiency | 427.900 | 175.029 | 1.400 | 1.000 | 21.500 |
| $\widehat{\beta}_{42}$ | MedSE(NMM) | 18.654 | 6.783 | 0.008 | 0.002 | 0.041 |
| | MedSE($t$MM) | 0.030 | 0.033 | 0.006 | 0.001 | 0.002 |
| | Efficiency | 621.8 | 205.545 | 1.333 | 2.000 | 20.500 |

Table 5: Simulation results for Case 4: $n_i = 4$, $I = 400$.

| | Original | | With 1 outlier | | With 2 outliers | |
|---|---|---|---|---|---|---|
| | $t_{28}$ | Normal | $t_9$ | Normal | $t_6$ | Normal |
| $\widehat{\pi}_1$ | 0.248 | 0.235 | 0.281 | 0.569 | 0.297 | 0.196 |
| $\widehat{\pi}_2$ | 0.688 | 0.704 | 0.652 | 0.402 | 0.630 | 0.765 |
| $\widehat{\beta}_{01}$ | -0.010 | -0.010 | -0.041 | -0.126 | -0.056 | -0.175 |
| $\widehat{\beta}_{11}$ | 0.074 | 0.074 | 0.074 | 0.083 | 0.075 | 0.083 |
| $\widehat{\beta}_{02}$ | -0.350 | -0.341 | -0.361 | -0.336 | -0.368 | -0.274 |
| $\widehat{\beta}_{12}$ | 0.092 | 0.091 | 0.093 | 0.090 | 0.093 | 0.087 |
| $\widehat{\beta}_{03}$ | -0.307 | -0.296 | -0.293 | -0.418 | -0.279 | -0.335 |
| $\widehat{\beta}_{13}$ | 0.074 | 0.073 | 0.074 | 0.090 | 0.075 | 0.088 |

Table 6: Estimation results for the Topeka girls lung function data analysis.

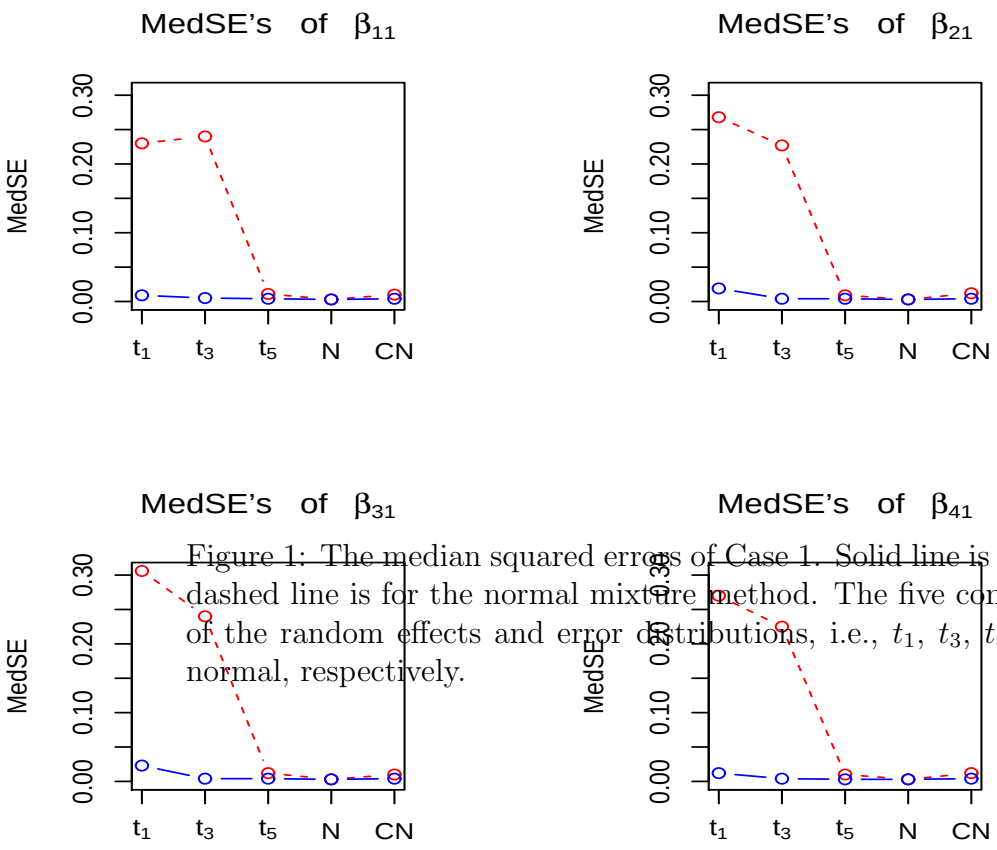


Figure 1: The median squared errors of Case 1. Solid line is for the $t$-mixture method and dashed line is for the normal mixture method. The five conditions refer to five scenarios of the random effects and error distributions, i.e., $t_1$, $t_3$, $t_5$, normal, and contaminated normal, respectively.

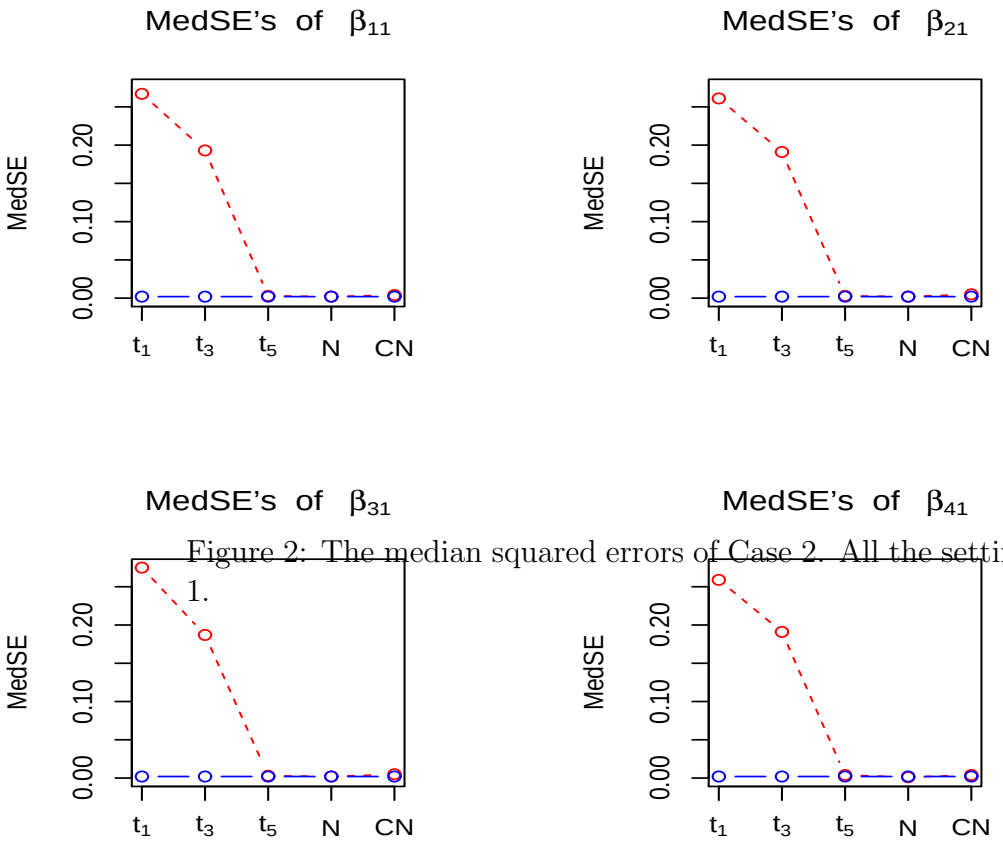


Figure 2: The median squared errors of Case 2. All the settings are the same as in Figure 1.

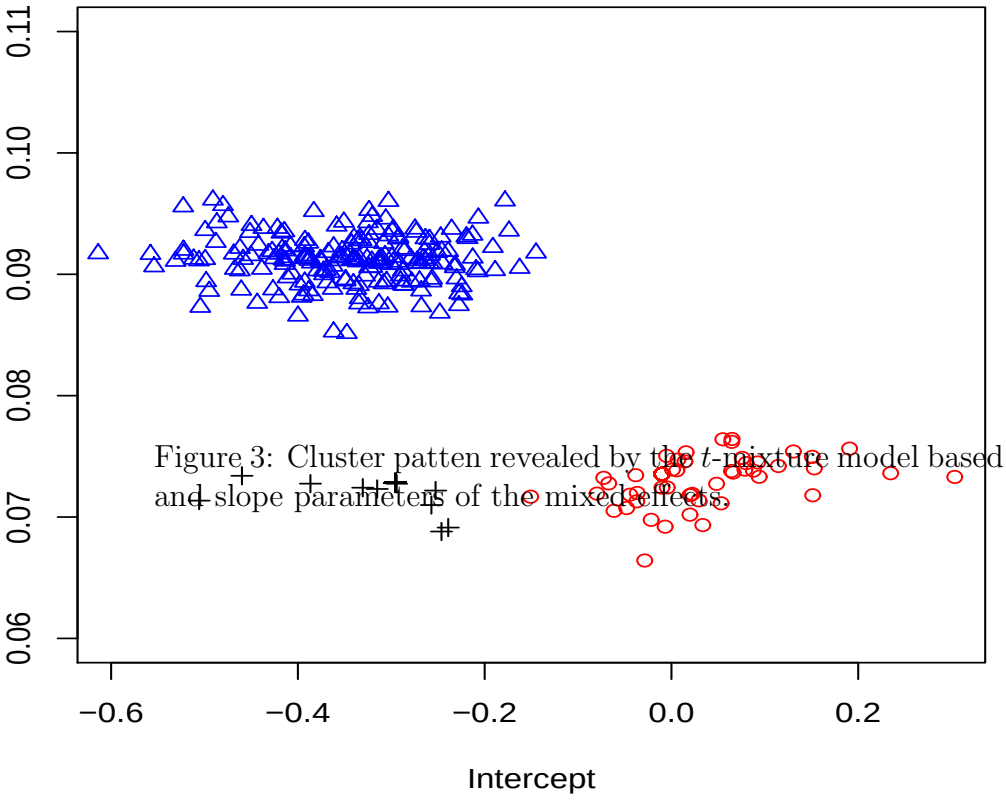


Figure 3: Cluster patten revealed by the $t$-mixture model based on the estimated intercept and slope parameters of the mixed effects.