Linear mixed models for multiple outcomes using extended multivariate skew-t distributions

BINBING YU*, A. JAMES O'MALLEY, AND PULAK GHOSH

Multivariate outcomes with heavy skewness and thick tails often arise from clustered experiments or longitudinal studies. Linear mixed models with multivariate skew-t (MST) distributions for the random effects and the error terms is a popular tool of robust modeling for such outcomes. However the usual MST distribution only allows a common degree of freedom for all marginal distributions, which is only appropriate when each marginal has the same amount of tail heaviness. In this paper, we introduce a new class of extended MST distributions, which allow different degrees of freedom and thereby can accommodate heterogeneity in tail-heaviness across outcomes. The extended MST distributions yield a flexible family of models for multivariate outcomes. The hierarchical representation of the MST distribution allows MCMC methods to be easily applied to compute the parameter estimates. The proposed model is applied to data from two biomedical studies: one on bivariate markers of AIDS progression and the other on sexual behavior from a longitudinal study.

KEYWORDS AND PHRASES: Multivariate skew-t, Robust method, Scale-mixture representation.

1. INTRODUCTION

Skew-normal and skew-t distributions were originally proposed by Azzalini and Capitanio (1999, 2003) for modeling data with heavy tails and strong skewness without transformation. Since then, the statistical properties of the skew distributions and their extensions have been further studied by many other researchers. The multivariate skew-t (MST) distributions, including the multivariate skew normal (MSN) distributions as a special case, have become increasingly popular. Branco and Dey (2001) discussed a general class of multivariate skew-elliptical distributions. Lin (2010) presented a robust mixture modeling framework using the MST distribution. Fruhwirth-Schnatter and Pyne (2010) developed Bayesian inference for finite mixtures of univariate and MSN and MST distributions. Applications and case studies using multivariate skew distributions can be found in many research areas, including economics, engineering, epidemiology and environmental sciences. For example, Panagiotelis and Smith

*Corresponding author.

(2008) used the MST in the forecasting of intraday electricity prices; Wang, Ng and McLachlan (2009) applied the MST to fluorescence-activated cell sorting data; Ghosh, Branco and Chakraborty (2006) developed a bivariate MSN random-effects model to model the CD4 counts and the viral RNA in an HIV-RNA study. The estimation of the MST parameters has been implemented in the R package `sn` (Azzalini and Genton, 2008).

Linear mixed models (LMM) are a popular tool for analyzing continuous outcome variables with repeated measurements or clustering (Verbeke and Molenberghs, 2001). Correlations among repeated observations are taken into account through random effects. In standard LMM, both the random effects and the errors are assumed to be normally distributed. In many situations, the normality assumptions may be questionable or totally incorrect. Misspecification of the random-effects distributions may introduce biases to parameter estimates and lead to incorrect inference (Verbeke and Lesaffre, 1996) or may increase or decrease the power of statistical tests (Litiere, Alonso and Molenberghs, 2007). Transformation of the outcome variables is often used as a remedy to improve normality. For example, Gurka et al. (2006) extended the use of Box-Cox transformation to LMM. O'Malley and Zou (2006) proposed a Bayesian multivariate hierarchical transformation model to estimate the ROC curves for two prostate cancer markers, where the Box-Cox transformation was applied to clustered marker values. Although the transformation methods provide a better fit to the data, the assumption of joint normality is rarely satisfied and the transformed variables are usually difficult to interpret. Thus, statistical models without transformation that are less sensitive to departures from normality and related assumptions are of practical interest.

Several extensions of the standard LMM have been developed by replacing the normal distributions of the random effects and errors with more general distributions. See Laird (1978); Müller and Rosner (1997); Zhang and Davidian (2001), among others. Recently, Ho and Lin (2010); Jara, Quintana and San Martin (2008); Lin and Lee (2008) and Lachos, Ghosh and Arellano-Valle (2010) proposed to use various multivariate skew distributions to model the random-effects and the error terms in the LMMs. Bayesian approaches or EM-type algorithms were used to obtain parameter estimates and to draw inferences (Wang and Fan, 2010, 2011, 2012). The MSN or MST distributions are more

flexible parametric families with capability of accommodating various departures from normality. By using the skew distributions, we avoid the need of ad hoc transformations and the models remain robust to outliers.

To date, most models only considered a univariate outcome or multiple outcomes of the same type (Sammel, Lin and Ryan, 1999; Shah, Laird and Schoenfeld, 1997). Therefore all marginal distributions of the multiple outcomes are univariate skew-t with the same degrees of freedom (DF). When multiple outcomes are considered simultaneously (Boscardin, Taylor and Law, 1998; Pantazis et al., 2005), their marginal distributions should have different DF because they are different in nature. Furthermore, when both intercept and slope are considered as random effects, it is expected that the random intercepts and the random slopes should have different tail heaviness. The objective of this paper is to extend the MST to accommodate multiple outcomes with different skewnesses and tails. The special case of bivariate t distributions with variable marginal DF has been considered by Jones (2002) and Shaw and Lee (2008). The extended MST distributions are used for random effects to create a more general LMM for multiple outcomes. Markov chain Monte Carlo (MCMC) method are used for parameter estimation and inferences. The proposed method can be implemented in the freely available software WinBUGS or OpenBUGS (Lunn et al., 2009).

The rest of the paper is organized as follows. We describe the statistical models and the estimation method in Section 2. Simulation results are presented in Section 3. Two applications, a longitudinal HIV marker study and a sexual behavior study, are presented in Section 4. Other potential applications and suggestions for additional research are discussed in Section 5.

2. LINEAR MIXED MODEL WITH EXTENDED MULTIVARIATE SKEW-t DISTRIBUTIONS

We consider K correlated outcomes with different skewnesses and tails for N subjects. This could occur in longitudinal studies, for example, multiple markers are assessed at the follow-up visits for each subject. It could also arise in clinical settings, for example, several cognitive tests are performed at different locations for each individual. We assume that all K outcomes are measured n_i times for the ith subject. In general, each outcome variable may have a different number of observations or may be measured at irregular and a different number of occasions. In the notations hereafter, vectors of parameters or observations are denoted as bold symbols. Let $\boldsymbol{y}_i^{(k)} = (y_{i1}^{(k)}, \ldots, y_{in_i}^{(k)})'$ be a column vector of n_i continuous measurements of the kth outcome for the ith subject, $i = 1, \ldots, N, k = 1, \ldots, K$. We assume that $\boldsymbol{y}_i^{(k)}$ follows an LMM:

$$(1) \qquad \mathbf{y}_i^{(k)} = \boldsymbol{X}_i^{(k)} \boldsymbol{\beta}^{(k)} + \boldsymbol{Z}_i^{(k)} \boldsymbol{b}_i^{(k)} + \boldsymbol{e}_i^{(k)},$$

where $\boldsymbol{X}_i^{(k)}$ and $\boldsymbol{Z}_i^{(k)}$ are pre-specified $n_i \times p_k$ and $n_i \times q_k$ full rank design matrices corresponding to the p_k dimensional fixed-effects vector $\boldsymbol{\beta}^{(k)}$ and the q_k-dimensional random-effects vector $\boldsymbol{b}_i^{(k)}$, respectively, and $\boldsymbol{e}_i^{(k)}$ is an n_i-dimensional vector for the within-subject errors.

Jara, Quintana and San Martin (2008) proposed to use multivariate skew elliptical distributions to model the error terms and the random effects. Specifically, they considered the MST distributions for one outcome with repeated measurements ($K = 1$). We assume that the error terms and the random effects for the kth outcome

$$(2) \qquad \boldsymbol{e}_i^{(k)} \sim MST(\boldsymbol{\mu}_{e_i}^{(k)}, \boldsymbol{\Sigma}_{e_i}^{(k)}, \boldsymbol{\Lambda}_{e_i}^{(k)}, \nu_e^{(k)}),$$

$$(3) \qquad \boldsymbol{b}_i^{(k)} \sim MST(\boldsymbol{\mu}_b^{(k)}, \boldsymbol{\Sigma}_b^{(k)}, \boldsymbol{\Lambda}_b^{(k)}, \nu_b^{(k)}),$$

where $MST(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \nu)$ denotes a MST distribution with location vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$, diagonal matrix $\boldsymbol{\Lambda}$ for skewness and scale parameter ν for DF. When $\nu \to \infty$, the skew-t distribution converges to the skew-normal distribution. The DF parameters $\nu_e^{(k)}$ and $\nu_b^{(k)}$ are the same for different individuals. The scale matrix $\boldsymbol{\Lambda}_b^{(k)}$ and the skewness matrix $\boldsymbol{\Sigma}_b^{(k)}$ for the random effects $\boldsymbol{b}_i^{(k)}$ have dimensions $q_k \times q_k$. The matrices $\boldsymbol{\Lambda}_{e_i}^{(k)}$ and $\boldsymbol{\Sigma}_{e_i}^{(k)}$ for the error terms $\boldsymbol{e}_i^{(k)}$ have dimensions $n_i \times n_i$. Therefore, the dimensions of $\boldsymbol{\Lambda}_b^{(k)}$ and $\boldsymbol{\Sigma}_b^{(k)}$ only depend on the number of random effects q_k. On the other hand, the dimensions of $\boldsymbol{\Lambda}_{e_i}^{(k)}$ and $\boldsymbol{\Sigma}_{e_i}^{(k)}$ depend on the cluster size n_i for the ith subject. If the matrices $\boldsymbol{\Lambda}_{e_i}^{(k)}$ and $\boldsymbol{\Sigma}_{e_i}^{(k)}$ are totally unspecified, the number of unknown parameters increases with the cluster size n_i. Unlike the Generalized Estimating Equation (GEE) where the correlation matrices are treated as nuisance parameters, the correlation matrices of an LMM should be estimated based on observed data. Over-parametrization may cause estimation problems for the correlation matrices. Therefore, more parsimonious parameterizations have been used. For example, Jara, Quintana and San Martin (2008) assumed that $\boldsymbol{\Lambda}_{e_i}^{(k)} = \lambda_e^{(k)} \mathbf{I}_{n_i}$ and $\boldsymbol{\Sigma}_{e_i}^{(k)} = (\sigma_e^{(k)})^2 \mathbf{I}_{n_i}$, where $\mathbf{I}_{n_i}$ is an $n_i \times n_i$ identity matrix. More generally, one can assume $\boldsymbol{\Sigma}_{e_i}^{(k)}$ to be $n_i \times n_i$ dimensional AR(1) or compound symmetry matrices. By using these formulations, the number of unknown parameters do not increase with the number of subjects or outcomes.

When multiple outcomes are considered simultaneously, the correlation may be taken into account by incorporating the random effects $\boldsymbol{b}_i^{(1)}, \ldots, \boldsymbol{b}_i^{(K)}$. In the analysis, we rely on the random effects to provide correlations among multiple outcomes, so that we assume that $\boldsymbol{e}_i^{(1)}, \ldots, \boldsymbol{e}_i^{(K)}$ are mutually independent across different outcomes and individuals. Therefore, the conditional distributions $\boldsymbol{y}_i^{(k)} | \boldsymbol{b}_i^{(k)}, k = 1, \ldots, K$, in (1) are independent for different outcomes. Thus,

$$(4) \qquad \mathbf{y}_i^{(k)} | \boldsymbol{b}_i^{(k)} \overset{\text{ind}}{\sim} MST(\boldsymbol{X}_i^{(k)} \boldsymbol{\beta}^{(k)} + \boldsymbol{Z}_i^{(k)} \boldsymbol{b}_i^{(k)}, \boldsymbol{\Sigma}_e^{(k)}, \boldsymbol{\Lambda}_e^{(k)}, \nu_e^{(k)}).$$

Meanwhile, the random effects $\boldsymbol{b}_i^{(1)}, \ldots, \boldsymbol{b}_i^{(K)}$ should be correlated in order to introduce correlations among multiple outcomes. In addition, the distributions of the random effects for different outcomes should be allowed to have unique DF.

2.1 Extended MST for random effects

To solve this problem, we propose an extended MST (EMST) distribution with different marginal DFs for the random effects. Let $\boldsymbol{b}_i = (\boldsymbol{b}_i^{(1)}, \ldots, \boldsymbol{b}_i^{(K)})$ be the q dimensional vector of combined random effects for the ith subject, where $\boldsymbol{b}_i^{(k)}$ is the vector of random effects of the kth outcome and $q = q_1 + q_2 + \cdots + q_K$. Instead of a single scale parameter ν for the MST distribution, we use a vector $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_q)$. The q-dimensional random vector of the extended MST $\boldsymbol{r}$, denoted by $\boldsymbol{r} \sim EMST(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \boldsymbol{\nu})$ can be generated hierarchically (Lin, 2010),

$$
\begin{aligned}
\boldsymbol{r}|\boldsymbol{\gamma}, \boldsymbol{\tau} &\sim MVN_q(\boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\gamma}, \boldsymbol{\tau}^{-\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\tau}^{-\frac{1}{2}}), \\
\boldsymbol{\gamma}|\boldsymbol{\tau} &\sim HN_q(\boldsymbol{0}, \boldsymbol{\tau}^{-1}), \\
\boldsymbol{\tau} &= \mathrm{Diag}(\tau_1, \ldots, \tau_q), \\
\tau_l &\sim \Gamma(\nu_l/2, \nu_l/2), \quad l = 1, \ldots, q,
\end{aligned}
$$

where $MVN_q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, HN_q stands for a q-dimensional half-normal distribution and $\Gamma(a, b)$ denotes a Gamma distribution with mean a/b. The extended MST distribution has the following properties:

- The marginal distributions are univariate skew-t with ν_l DF.
- When $\nu_1 = \cdots = \nu_q \equiv \nu$, the extended MST reduces to a regular MST distribution with common DF. Furthermore, if $\boldsymbol{\Lambda} = \boldsymbol{0}$, it is a symmetric multivariate t distribution with ν DF.
- When $\min(\nu_1, \ldots, \nu_q) \to \infty$, this reduces to a MSN distribution.

For the proposed LMM for multiple outcomes, we may assume that the random effects

$$
(5) \qquad \boldsymbol{b}_i \sim EMST(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b, \boldsymbol{\Lambda}_b, \boldsymbol{\nu}_b),
$$

where the location parameters $\boldsymbol{\mu}_b = (\boldsymbol{\mu}_b^{(1)'}, \ldots, \boldsymbol{\mu}_b^{(K)'})'$ with $\boldsymbol{\mu}_b^{(k)}$ defined in Section 2.2, $\boldsymbol{\Sigma}_b$ is the covariance matrix with block diagonal elements $\boldsymbol{\Sigma}_b^{(k)}$, $\boldsymbol{\Lambda}_b = \mathrm{Diag}(\boldsymbol{\Lambda}_b^{(1)}, \ldots, \boldsymbol{\Lambda}_b^{(K)})$ with diagonal matrices $\boldsymbol{\Lambda}_b^{(k)}$, and $\boldsymbol{\nu}_b = (\boldsymbol{\nu}_b^{(1)}, \ldots, \boldsymbol{\nu}_b^{(K)})'$ is the vector of DF. The DF parameters $\boldsymbol{\nu}_b^{(k)} = (\nu_{b_1}^{(k)}, \ldots, \nu_{b_{q_k}}^{(k)}), k = 1, \ldots, K$. One typical example is that $\boldsymbol{b}_i^{(k)} = (b_{i1}^{(k)}, b_{i2}^{(k)})'$ is the vector of random intercept and random slope for the kth outcome. This formulation allows different DFs for the random intercepts and the random slopes.

2.2 Bayesian estimation

The usual estimation method for MST is based on the likelihood function, which could be very cumbersome. Based on the hierarchical representation in Section 2.1, one can use the MCMC method to obtain parameter estimates. Following Jara, Quintana and San Martin (2008), we assume that the error terms follow MST distributions as specified in (2) with $\boldsymbol{\mu}_{e_i}^{(k)} = \mu_e^{(k)}\mathbf{1}_{n_i}, \boldsymbol{\Sigma}_{e_i}^{(k)} = (\sigma_e^{(k)})^2\mathbf{I}_{n_i}$ and $\boldsymbol{\Lambda}_{e_i}^{(k)} = \lambda_e^{(k)}\mathbf{I}_{n_i}$, where $\mathbf{1}_{n_i}$ is a column vector of 1's and $\mathbf{I}_{n_i}$ is an $n_i \times n_i$ identity matrix. The random effects $\boldsymbol{b}_i$ follow extended MST distributions as specified in (5).

We set the restrictions that

$$
\mu_e^{(k)} = -\sqrt{\frac{\nu_e^{(k)}}{\pi}} \frac{\Gamma[(\nu_e^{(k)} - 1)/2]}{\Gamma(\nu_e^{(k)}/2)} \lambda_e^{(k)}, \quad \text{and}
$$

$$
(\boldsymbol{\mu}_b^{(k)})' = \left(-\sqrt{\frac{\nu_{b_1}^{(k)}}{\pi}} \frac{\Gamma[(\nu_{b_1}^{(k)} - 1)/2]}{\Gamma(\nu_{b_1}^{(k)}/2)} \lambda_{b_1}^{(k)}, \ldots, \right.
$$
$$
\left. -\sqrt{\frac{\nu_{b_{q_k}}^{(k)}}{\pi}} \frac{\Gamma[(\nu_{b_{q_k}}^{(k)} - 1)/2]}{\Gamma(\nu_{b_{q_k}}^{(k)}/2)} \lambda_{b_{q_k}}^{(k)} \right).
$$

By this parametrization, the marginal expectation $E(\boldsymbol{e}_i^{(1)}) = \cdots = E(\boldsymbol{e}_i^{(K)}) = E(\boldsymbol{b}_i) = \boldsymbol{0}$ and hence the unconditional expectation of $\boldsymbol{y}_i^{(k)}$ is $E(\boldsymbol{y}_i^{(k)}) = \boldsymbol{X}_i^{(k)}\boldsymbol{\beta}_i^{(k)}$. In this way, the parameters $\boldsymbol{\beta}$ control the expectation (location) of outcomes and the parameters $\boldsymbol{\Lambda}$ and $\boldsymbol{\nu}$ controls the skewness and tails of the random effects and the error terms, avoiding confounding between the location and skewness parameters. However, as $\boldsymbol{y}_i^{(k)}$ is a linear combination of two skew-t variables, the marginal distribution of $\boldsymbol{y}_i^{(k)}$ does not have a close form.

As described by Lin (2010), the outcome $\boldsymbol{y}_i^{(k)}$ can be generated as a hierarchical mixture of normal distributions:

$$
E(\boldsymbol{y}_i^{(k)}|\boldsymbol{b}_i^{(k)}) = \boldsymbol{X}_i^{(k)}\boldsymbol{\beta}^{(k)} + \boldsymbol{Z}_i^{(k)}\boldsymbol{b}_i^{(k)} + \boldsymbol{\mu}_{e_i}^{(k)} + \boldsymbol{\Lambda}_{e_i}^{(k)}\boldsymbol{\gamma}_{e_i}^{(k)},
$$

$$
(6) \quad \boldsymbol{y}_i^{(k)}|\boldsymbol{b}_i^{(k)} \sim MVN_{p_k}\big(E(\boldsymbol{y}_i^{(k)}|\boldsymbol{b}_i^{(k)}), \boldsymbol{\Sigma}_{e_i}^{(k)}/\tau_{e_i}^{(k)}\big),
$$

$$
(7) \quad \boldsymbol{\gamma}_{e_i}^{(k)}|\tau_{e_i}^{(k)} \sim MVN_{n_i}\big(\boldsymbol{0}, \mathbf{I}_{n_i}/\tau_{e_i}^{(k)}\big)I\big(\boldsymbol{\gamma}_{e_i}^{(k)} > \boldsymbol{0}\big),
$$

$$
(8) \qquad \tau_{e_i}^{(k)} \sim \Gamma\big(\nu_e^{(k)}/2, \nu_e^{(k)}/2\big),
$$

for $k = 1, \ldots, K$. Similarly, the random effects $\boldsymbol{b}_i$ can be generated as

$$
(9) \quad \boldsymbol{b}_i|\boldsymbol{\gamma}_b, \boldsymbol{\tau}_b \sim MVN_q\big(\boldsymbol{\mu}_b + \boldsymbol{\Lambda}_b\boldsymbol{\gamma}_b, \boldsymbol{\tau}_b^{-\frac{1}{2}}\boldsymbol{\Sigma}_b\boldsymbol{\tau}_b^{-\frac{1}{2}}\big),
$$

$$
(10) \qquad \boldsymbol{\gamma}_b|\boldsymbol{\tau}_b \sim MVN_q\big(\boldsymbol{0}, \boldsymbol{\tau}_b^{-1}\big)I(\boldsymbol{\gamma}_b > \boldsymbol{0}),
$$

$$
(11) \qquad \tau_{b_i}^{(k)} \sim \Gamma\big(\nu_{b_i}^{(k)}/2, \nu_{b_i}^{(k)}/2\big), \quad i = 1, \ldots, q_k,
$$

where $\boldsymbol{\tau}_b = \mathrm{Diag}(\tau_{b_1}^{(1)}, \ldots, \tau_{b_{q_1}}^{(1)}, \ldots, \tau_{b_1}^{(K)}, \ldots, \tau_{b_{q_K}}^{(K)})$.

To complete specification of a Bayesian model it is necessary to choose prior distributions for parameters in the

model that do not have a distribution assigned to them (referred to as fixed effects in frequentist inference). Fixed effects are usually modeled using a MVN distribution. Usually vague priors are specified by taking the covariance matrix to be a diagonal matrix with large elements. Prior distributions for the DF can be discrete as in Albert and Chib (1993) and Besag et al. (1995), or continuous as in Geweke (1993). When a discrete prior with too coarse a grid is used, there may be numerical problems or poor mixing, unless good initial values are provided. Spiegelhalter et al. (2003) advised to either use a continuous prior or a discrete prior on a fine grid.

Following Jara, Quintana and San Martin (2008), we used the independent truncated exponential distributions. The truncation point was chosen to ensure that the third-order moment associated with the skewness of the MST distribution was finite. Specifically, the parameters of the error terms included $\{((\sigma_e^{(k)})^2, \lambda_e^{(k)}, \nu_e^{(k)}), k = 1, \ldots, K\}$ and the parameters for the random-effects were $\{(\mathbf{\Sigma}_b^{(k)}, \mathbf{\Lambda}_b^{(k)}, \nu_b^{(k)}), k = 1, \ldots, K\}$. Note that $\mathbf{\Lambda}_b^{(k)}$ was a diagonal matrix. Let $\boldsymbol{\lambda}_b^{(k)}$ be the vector of the diagonal elements of $\mathbf{\Lambda}_b^{(k)}$. We assumed mutual independence across the parameters. The priors for the parameters were specified as:

$$(12) \qquad \boldsymbol{\beta}_k \sim MVN(\mathbf{0}, \mathbf{\Sigma}_\beta),$$

$$(13) \qquad \lambda_e^{(k)} \sim N(0, \sigma_{\lambda_e}^2),$$

$$(14) \qquad \nu_e^{(k)} \sim \exp(m_{\nu_e})I(\nu_e^{(k)} > 3),$$

$$(15) \qquad \boldsymbol{\lambda}_b^{(k)} \sim MVN(0, \mathbf{\Sigma}_{\lambda_b}),$$

$$(16) \qquad \nu_{b_i}^{(k)} \sim \exp(m_{\nu_b})I(\nu_b^{(k)} > 3), \quad i = 1, \ldots, q_k,$$

for $k = 1, \ldots, K$. Besides the truncated exponential distribution for the DF parameters, (Gelman and Hill, 2006, pp. 372) suggested to use uniform distributions for the inverse of the DF parameters. Specifically, we may use the priors $\nu_e^{(k)} \sim \text{Uniform}(0, 1/3)$ and $\nu_{b_i}^{(k)} \sim \text{Uniform}(0, 1/3)$. As a sensitivity analysis on prior distributions, we tried both types of priors for the DF parameters. The hyperparameters were chosen such that the priors are weakly informative. For example, we let the variances for each component of $\boldsymbol{\beta}_k, \lambda_e^{(k)}, \boldsymbol{\lambda}_b^{(k)}$ be 1,000 and the mean be $m_{\nu_e} = m_{\nu_b} = 100$.

Inverted gamma distributions were used for the variance parameters $(\sigma_e^{(k)})^2$ and inverted Wishart distributions (Odell and Feiveson, 1966) were used for the covariance matrix $\mathbf{\Sigma}_b$:

$$(17) \qquad 1/(\sigma_e^{(k)})^2 \sim \Gamma(a_e, b_e), \quad k = 1, \ldots, K,$$

$$(18) \qquad (\mathbf{\Sigma}_b)^{-1} \sim \text{Wishart}(d_b, \mathbf{\Omega}_b),$$

where $\text{Wishart}(d, \mathbf{\Omega})$ is a Wishart distribution with d degrees of freedom and mean $d\mathbf{\Omega}$. The augmented joint posterior density of the unknown parameters can be found by combining expressions (6)–(18). The full conditional distributions for the MCMC were straightforward to derive and

sample from except for $\boldsymbol{\nu}$ and $\boldsymbol{\gamma}$, for which the Metropolis-Hastings algorithm (Chib and Greenberg, 1995) was used.

After a sufficient number of burn-in iterations, we used the remaining samples to estimate any function of the parameters of interest. In order to see how stable the final estimates are, two independent MCMC chains were run with different initial values and starting points. The convergence of the MCMC samples of the parameters after excluding the initial burn-in samples were monitored using the R package CODA. For example, Gelman (2006) used a 'potential scale reduction factor' (PSRF) for each parameter in $\boldsymbol{\theta}$, together with upper and lower confidence limits. Approximate convergence was achieved when the upper limits were close to 1.

3. SIMULATION

Simulation studies were conducted to assess the performance of the proposed method. We considered bivariate outcomes $\boldsymbol{y} = (y^{(1)}, y^{(2)})$ from 200 subjects and each subject had 5 measurements for both outcomes. Let $y_{ij}^{(k)}$ denote the jth measurement of the kth outcome for the ith individual, $i = 1, \ldots, 200, j = 1, \ldots, 5$ and $k = 1, 2$. Two covariates were used, i.e., $t_{ij} = j$ and $x_i = \mathbf{I}(i > 100)$. We considered two scenarios. In Scenario 1, a bivariate LMM with correlated random intercepts was given as:

$$y_{ij}^{(k)} = \beta_1^{(k)} + \beta_2^{(k)} t_{ij} + \beta_3^{(k)} x_i + b_i^{(k)} + e_{ij}^{(k)}.$$

The vector of correlated random intercepts were denoted by $\boldsymbol{b}_i = (b_i^{(1)}, b_i^{(2)})$. In Scenario 2, the two outcomes were independent and each outcome followed an LMM with both random intercepts and random slopes for t_{ij}:

$$y_{ij}^{(k)} = \beta_1^{(k)} + \beta_2^{(k)} t_{ij} + \beta_3^{(k)} x_i + b_{i1}^{(k)} + b_{i2}^{(k)} t_{ij} + e_{ij}^{(k)}.$$

Let $\boldsymbol{b}_i^{(1)} = (b_{i1}^{(1)}, b_{i2}^{(1)})$ and $\boldsymbol{b}_i^{(2)} = (b_{i1}^{(2)}, b_{i2}^{(2)})$ be the random effects for the two outcomes. We assumed that $\boldsymbol{b}_i^{(1)}$ and $\boldsymbol{b}_i^{(2)}$ were independent but for each outcome the random intercept and random slope were correlated.

In both scenarios, the regression coefficients were $\beta^{(1)} = (\beta_1^{(1)}, \beta_2^{(1)}, \beta_3^{(1)}) = (2, 0, 1)$ and $\beta^{(2)} = (\beta_1^{(2)}, \beta_2^{(2)}, \beta_3^{(2)}) = (0, 1, 1)$. The parameter $\beta_2^{(k)}$ measured the effect of time and the parameter $\beta_3^{(k)}$ measured the effect of the individual-level covariate. To demonstrate the advantage of using a MST distribution for the random effects, we assumed the error terms $e_{ij}^{(k)}$ were independent normal $N(0, 0.25)$ across different outcomes and individuals, but generated the random effects $\boldsymbol{b}_i, \boldsymbol{b}_i^{(1)}$ and $\boldsymbol{b}_i^{(2)}$ from three bivariate distributions with different skewnesses and tails. The three bivariate distributions for the random effects were the McKay's bivariate Gamma, the bivariate t_3 distribution and the bivariate normal. All three distributions were standardized such that the marginal distributions had mean 0 and standard deviation 2. The correlation between the two marginals was 0.5.
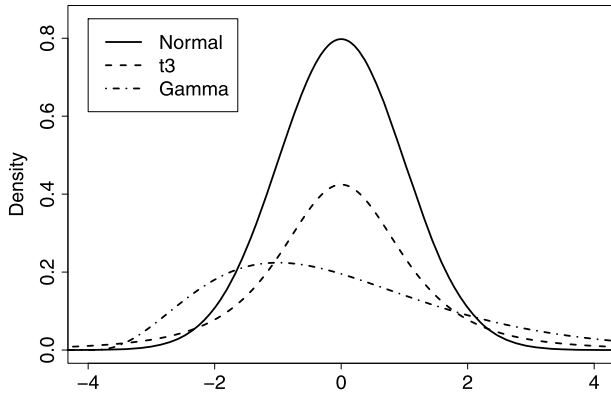
Figure 1. Distributions of random effects used in simulation.

Table 1. Summary of the parameter estimates for bivariate random-intercept model (Scenario 1)

Parameter	Normal		Gamma		t_3	
	Bias	ESD†	Bias	ESD	Bias	ESD
	LMM with normal random effects					
$\beta_1^{(1)}$	0.010	0.190	0.003	0.210	0.004	0.207
$\beta_2^{(1)}$	0.000	0.012	-0.001	0.011	-0.001	0.011
$\beta_3^{(1)}$	0.007	0.280	0.016	0.270	0.003	0.283
$\beta_1^{(2)}$	0.001	0.206	0.018	0.224	-0.011	0.194
$\beta_2^{(2)}$	0.000	0.011	0.001	0.012	0.001	0.010
$\beta_3^{(2)}$	0.014	0.310	0.018	0.300	0.022	0.269
$SD(b_i^{(1)})$	-0.021	0.105	-0.039	0.134	-0.108	0.320
$SD(b_i^{(2)})$	-0.016	0.097	-0.022	0.114	-0.092	0.304
	LMM with MST random effects					
$\beta_1^{(1)}$	0.010	0.195	0.003	0.201	0.002	0.180
$\beta_2^{(1)}$	0.000	0.012	-0.001	0.011	-0.001	0.011
$\beta_3^{(1)}$	0.004	0.289	0.018	0.252	0.016	0.219
$\beta_1^{(2)}$	0.003	0.210	0.010	0.215	-0.007	0.169
$\beta_2^{(2)}$	0.000	0.011	0.001	0.012	0.001	0.010
$\beta_3^{(2)}$	0.025	0.313	0.022	0.279	0.021	0.203
$SD(b_i^{(1)})$	-0.003	0.105	-0.022	0.134	-0.091	0.319
$SD(b_i^{(2)})$	0.002	0.097	-0.005	0.114	-0.074	0.303

† ESD is defined as the standard deviation of the posterior estimates from the 250 simulations.

The normal distribution was used as the benchmark. The density functions of the three bivariate functions were plotted in Figure 1. The t_3 distribution had heavier tails and the Gamma distribution presented a certain degrees of skewness. After the multivariate outcomes were generated from the LMM with random effects from various distributions, the LMMs with normal or MST random effects were fit to the data. The two methods were referred to as LMM-NOR and LMM-MST hereafter. Here we focused on the fixed parameters of regression coefficients $\beta_i^{(k)}$ and the standard deviation (SD) of the random effects. The results were based on 250 simulated datasets for each case. For the LMM-MST, the estimates were based on 10,000 MCMC samples after the initial 10,000 runs were discarded. Converge diagnostics showed good mixing of the MCMC simulation as the PSRF values were close to 1 for all fixed parameters. We tried both truncated Exponential and Uniform prior distributions for the DF parameters. The resulting summary statistics for the regression coefficients and the SD of the random effects from both priors were very close. Therefore, the posterior inference with regard to the regression parameters were insensitive to the prior specifications of the DF parameters. We reported the results based on the truncated Exponential priors only. Summary of the parameter estimates for scenarios 1 and 2 were shown in Table 1 and Table 2, respectively.

In Table 1, we present the bias and empirical standard deviation (ESD) of the posterior estimates for each parameter from the bivariate LMM with correlated random intercepts. The ESD is defined as the standard error of the posterior estimates from the 250 simulations. We see that the biases of the fixed parameters $\beta_i^{(k)}$ from both types of LMM are very close to 0 and are comparable. This shows that the estimates of the fixed effect from the LMM-NOR are robust when the errors are non-gaussian. Using the criterion of the actual coverage rate of the 95% confidence interval, the robustness of the LMM-NOR has been confirmed by Jacqmin-Gadda et al. (2007) when the errors are independent. Although the estimates of fixed effects are unbiased and maintain the nominal coverage rate, incorrect specifica-

tion of the distributions of random effects and error terms may lead to efficiency loss. When the underlying true distribution for the random effects is normal, the ESD of the fixed parameters from the two methods are close. The ESDs from the LMM-MST are slightly larger than those from the LMM-NOR. The differences are all within 0.01. When the underlying distributions for random effects are Gamma or t_3, the ESD for the fixed parameters from both methods are close for the most part. However, the efficiencies of $\beta_1^{(k)}$ and $\beta_3^{(k)}, k = 1, 2$, are degraded when LMM-NOR is used. For example, in the last column for the t_3 distribution, the ESDs for $\beta_3^{(1)}$ and $\beta_3^{(2)}$ are 0.283 and 0.269, respectively. The corresponding ESDs from the LMM-MST are 0.219 and 0.203, which is a reduction of more than 0.06. This suggests that using LMM-NOR routinely may lead to inefficient inferences on fixed effects of primary interest, such as the treatment effects. Using simulations with a logistic random-intercept model, Litiere, Alonso and Molenberghs (2007, 2011) reached a similar conclusion that the misspecification of random-effects may distort the power calculation. In addition, we present the estimates of the SD of the random effects in the bottom of Table 1. The estimated SD for random-effects from the LMM-MST have smaller biases than those from the LMM-NOR, even when the true underlying distribution is normal. This is in part due to the joint simultaneous estimation of bivariate outcomes.

In Table 2, we present the summary statistics of the parameter estimates for Scenario 2, where the LMMs with both

Table 2. Summary of the parameter estimates for LMM with random intercepts and random slopes (Scenario 2)

	Normal		Gamma		t_3	
Parameter	Bias	ESD†	Bias	ESD	Bias	ESD
LMM with normal random effects						
$\beta_1^{(1)}$	−0.016	0.184	−0.003	0.201	0.002	0.192
$\beta_2^{(1)}$	−0.008	0.135	0.006	0.151	−0.002	0.138
$\beta_3^{(1)}$	0.022	0.248	−0.002	0.244	0.002	0.251
$\beta_1^{(2)}$	−0.008	0.188	0.016	0.194	−0.012	0.187
$\beta_2^{(2)}$	0.009	0.139	−0.003	0.143	−0.008	0.131
$\beta_3^{(2)}$	0.029	0.243	−0.003	0.242	0.007	0.260
$\mathrm{SD}(b_{i1}^{(1)})$	−0.066	0.112	−0.092	0.142	−0.161	0.327
$\mathrm{SD}(b_{i2}^{(1)})$	−0.073	0.112	−0.054	0.136	−0.096	0.530
$\mathrm{SD}(b_{i1}^{(2)})$	−0.004	0.098	−0.010	0.112	−0.080	0.302
$\mathrm{SD}(b_{i2}^{(2)})$	−0.017	0.105	0.000	0.106	−0.101	0.341
LMM with MST random effects						
$\beta_1^{(1)}$	−0.014	0.190	−0.013	0.195	0.003	0.176
$\beta_2^{(1)}$	−0.003	0.145	0.005	0.156	0.006	0.137
$\beta_3^{(1)}$	0.024	0.250	−0.007	0.228	0.012	0.208
$\beta_1^{(2)}$	−0.005	0.193	−0.016	0.197	−0.012	0.162
$\beta_2^{(2)}$	0.012	0.145	−0.004	0.149	−0.005	0.125
$\beta_3^{(2)}$	0.027	0.253	−0.008	0.225	0.009	0.202
$\mathrm{SD}(b_{i1}^{(1)})$	−0.001	0.109	−0.024	0.138	−0.091	0.320
$\mathrm{SD}(b_{i2}^{(1)})$	−0.008	0.109	0.013	0.134	−0.028	0.522
$\mathrm{SD}(b_{i1}^{(2)})$	0.002	0.097	−0.003	0.112	−0.073	0.301
$\mathrm{SD}(b_{i2}^{(2)})$	−0.011	0.105	0.007	0.106	−0.095	0.340

† ESD is defined as the standard deviation of the posterior estimates from the 250 simulations.

random intercepts and random slopes are used for two independent outcomes. We see similar results as shown in Table 1, except that the ESDs for parameters $\beta_2^{(1)}$ and $\beta_2^{(2)}$ from Table 2 are larger. This is because random slopes are added in Scenario 2. Based on the simulation, we have the following findings:

- The estimates of regression coefficients $\boldsymbol{\beta}^{(k)}$ from the LMM-NOR are unbiased and robust to departure of normality of random effects. As a generalization, the estimates of $\boldsymbol{\beta}^{(k)}$ from the LMM-MST are also unbiased.
- The estimates of the SD of the random effects tend to have negative bias. Gelman (2006) found that using a uniform prior for the variance parameter lead to positive biases. The negative bias of the SD parameter in the simulation is probably due to the inverted Gamma prior specification of the variance parameter. However, the underlying relationship between the direction of bias and prior specifications is not clear and requires further investigation. Again, we see that the estimates from the LMM-MST have smaller negative bias than those from the LMM-NOR.
- When the underlying true distribution of random-effects is indeed normal, using the LMM-MST may result in a slight efficiency loss. When the normality

assumption is violated, for example, the random effects are skewed or have heavy tails, the LMM-NOR may lead to inefficient inference of the fixed parameters. The loss of efficiency is more pronounced for the regression coefficient associated with the individual-level covariate.
- The LMM-MST lead to more accurate estimates of the SD of the random effects, even when the true underlying distribution is normal.

Indeed, impact of misspecification of random effects may vary, depending on both the underlying distribution and model structure (Neuhaus, McCulloch and Boylan, 2011). However, the LMM-MST maintain high efficiency and are more robust to distributions with heavy skewness and thick tails.

4. APPLICATIONS

4.1 Bivariate modeling for AIDS marker data

Longitudinal modeling for bivariate AIDS markers have been considered by Boscardin, Taylor and Law (1998), Thiebaut et al. (2002) and Thiebaut et al. (2005). Most LMM have used normal random effects. Ghosh and Hanson (2010) proposed a semiparametric Bayesian approach to multivariate longitudinal data when the random effects were assumed to follow a mixture of Polya trees prior distribution. They considered two markers of AIDS, i.e., CD4 and CD8 cell counts, from a large clinical trial on HIV-AIDS, ACTG 398. There were two treatment options for patients in the trial, that was double protease inhibitor (PI) arms or single-PI arm. The subjects in the trial were either NNRTI-naive, that is, they had not previously received drugs from the NNRTI class, or NNRTI-experienced, that is, they had previously received NNRTIs as part of their treatment. NNRTI-naive subjects had a better virologic response in the trial. The primary objective of the study was to compare the proportion of subjects who had virologic failure after 24 weeks on study between the double-PI arms and the single-PI arm. The CD4 and CD8 counts were measured at weeks 0 (baseline), 2, 4, 8, 16, 24, 32, 40, and 48 for all subjects. Some subjects had repeated observation at the same week. We considered two outcomes, i.e., CD4 and CD8 cell counts. In the analyzed dataset, there were 479 subjects, 3,664 CD4 and CD8 observations, 2–13 per subject. In the analysis, we found that both markers are highly skewed. Thus, we considered the two transformed outcomes $y^{(1)} = \log(1 + \mathrm{CD4})$ and $y^{(2)} = \log(1 + \mathrm{CD8})$. The covariates of interest were time after baseline (weeks), treatment (binary covariate, the three dual PI arms combined versus placebo), NNRTI experience, and baseline log10 viral load (VL).

The bivariate LMM was specified as

$$y_{ij}^{(k)} = \beta_1^{(k)} + \beta_2^{(k)}\texttt{week} + \beta_3^{(k)}\texttt{NNRTI} + \beta_4^{(k)}\texttt{treatment}$$
$$+ \beta_5^{(k)}\log_{10}(\texttt{VL}) + b_{i1}^{(k)} + b_{i2}^{(k)}\texttt{week} + e_{ij}^{(k)},$$
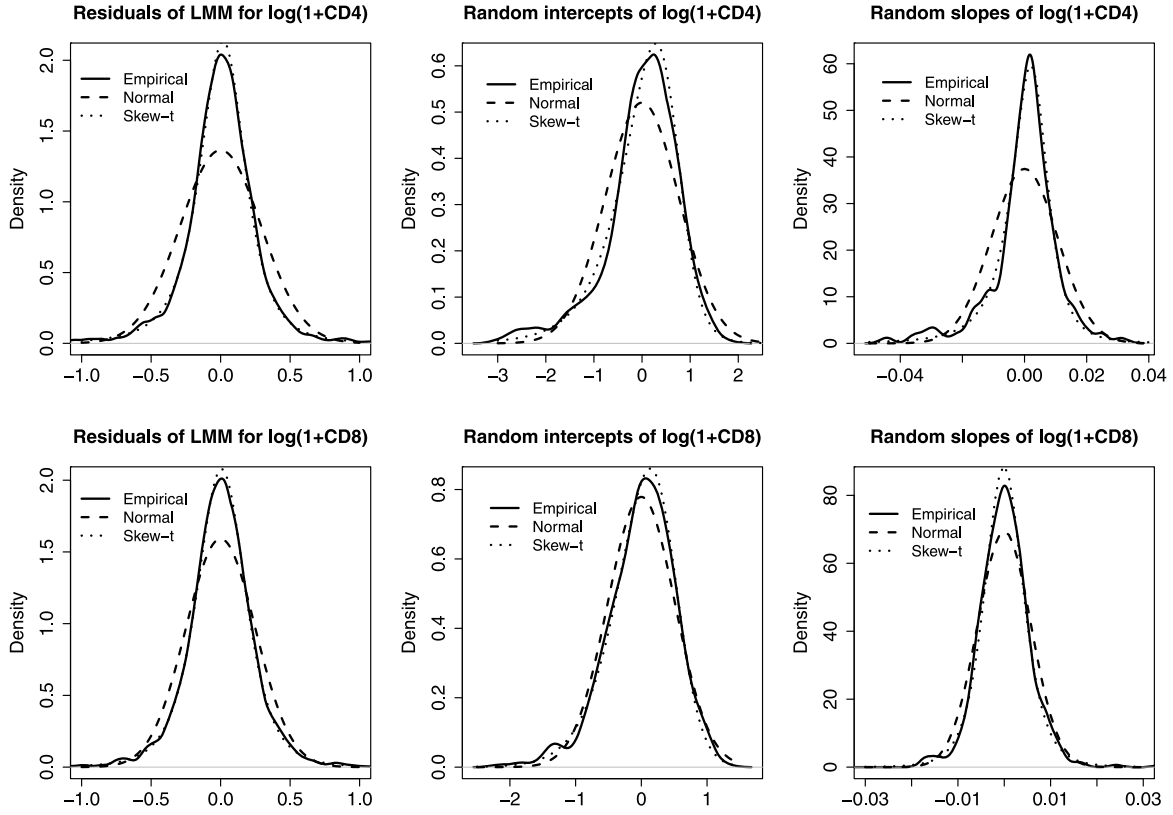
Figure 2. Distributions of the random effects and residuals from the normal LMM for the HIV study.

for $k = 1, 2$; $i = 1, \ldots, 479$; $j = 1, \ldots, n_i$, and $y_{ij}^{(k)}$ the value of the kth marker for the ith subject at the jth time point t_{ij} (in weeks). Figure 2 shows the empirical density functions of the residuals, random intercepts and random slopes for $Y^{(1)}$ and $Y^{(2)}$ from the normal LMM. The densities of the fitted normal and skew-t distributions are also presented. We can see that the residuals and random effects still present certain skewness and heavy tails, even after the log-transformation. The skew-t distributions show remarkably better fit than the normal distribution, which motivated us to use the MST distributions for random intercepts and slopes. As the random effects for the two outcomes are obviously correlated and may have different DFs, we considered the extended MST for the random effects $\boldsymbol{b}_i = (b_{i1}^{(1)}, b_{i2}^{(1)}, b_{i1}^{(2)}, b_{i2}^{(2)})$ and used univariate skew-t distributions for the error terms for each outcome. Prior values were fixed as follows: the regression coefficients $\beta^{(k)}$ were zero vectors; the skewness parameters were zeros; the scale parameters were generated from a gamma distribution with mean 1 and standard deviation 100; the DFs were generated from a uniform distribution $U(3, 100)$ and the inverse of covariance matrices $(\boldsymbol{\Sigma}_e^{(k)})^{-1}$ and $(\boldsymbol{\Sigma}_b)^{-1}$ were Diag$(1, 1)$. Reasonable MCMC mixing occurred well after 100,000 iterations. Based on the criterion by Gelman and Rubin (1992), the convergence of the MCMC sample was achieved. The posterior estimates were based on 10,000 samples from the additional 100,000 iterations with thinning interval of 10.

The parameter estimates with standard deviation (SD) and 95% credible intervals (CIs) of the LMM-MST are shown in Table 3. The estimates of regression coefficients are presented in the top section. The results based on the regression coefficients are similar to those by Ghosh and Hanson (2010). For example, as seen by the negative coefficients on the week term, the number of CD4 and CD8 decrease insignificantly with respect to time after baseline. The estimates of treatment effects and NNRTI experience are insignificant for both markers. The baseline HIV-RNA viral load is a significant covariate for both CD4 and CD8. By using the LMM with extended MST, we are able to examine the distributions of the residuals and random effects. The residuals from the LMM are strongly skewed and have heavy tails. For example, the skewness and DF for the residuals for $Y^{(1)}$ are -0.236 with 95% CI $(-0.261, -0.212)$ and 3.047 with 95% CI $(3.001, 3.170)$. The random intercepts for both markers also present certain skewness. The skewness parameters for the random intercepts are $\nu_{b1}^{(1)} = -0.587$ with 95% CI $(-0.750, -0.411)$ for log(1+CD4) and $\nu_{b1}^{(1)} = 0.128$ with 95% CI $(-0.341, 0.149)$ for log(1+CD8), respectively. The distributions of random slopes are approximately normal with skewness parameters close to 0 and DFs greater than 35.

Table 3. Parameter estimates of the LMM with extended MST for the HIV biomarker study

Variable		$Y^{(1)} = \log(1 + \text{CD4})$			$Y^{(2)} = \log(1 + \text{CD8})$		
		Mean	SD	(95% CI)	Mean	SD	(95% CI)
Intercept	$\beta_1^{(k)}$	7.060	0.194	(6.685, 7.455)	7.152	0.154	(6.857, 7.451)
Week	$\beta_2^{(k)}$	-0.001	0.002	$(-0.006, 0.004)$	-0.004	0.002	$(-0.008, 0.001)$
$\log_{10}(\text{Viral load})$	$\beta_3^{(k)}$	-0.359	0.041	$(-0.441, -0.282)$	-0.093	0.032	$(-0.156, -0.031)$
Treatment	$\beta_4^{(k)}$	-0.064	0.061	$(-0.194, 0.052)$	0.015	0.049	$(-0.080, 0.112)$
NNRTI	$\beta_5^{(k)}$	-0.101	0.058	$(-0.212, 0.017)$	-0.041	0.048	$(-0.134, 0.054)$
Parameters of skew-t distributions for random errors							
Skewness	$\lambda_e^{(k)}$	-0.236	0.012	$(-0.261, -0.212)$	-0.192	0.014	$(-0.220, -0.164)$
DF	$\nu_e^{(1)}$	3.047	0.046	(3.001, 3.170)	3.243	0.174	(3.011, 3.654)
Scale	$\sigma_e^{(k)}$	0.099	0.005	(0.089, 0.110)	0.128	0.006	(0.116, 0.141)
Parameters of extended MST for random effects (intercepts and slopes)							
Skewness (intercept)	$\lambda_{b1}^{(k)}$	-0.587	0.086	$(-0.750, -0.411)$	-0.112	0.128	$(-0.341, 0.149)$
DF (intercept)	$\nu_{b1}^{(k)}$	7.607	3.004	(4.179, 15.190)	16.740	7.977	(6.988, 38.610)
Scale (intercept)	$\sigma_{b1}^{(k)}$	0.431	0.035	(0.366, 0.502)	0.451	0.026	(0.398, 0.500)
Skewness (slope)	$\lambda_{b2}^{(k)}$	-0.000	0.004	$(-0.008, 0.008)$	0.000	0.004	$(-0.008, 0.008)$
DF (slope)	$\nu_{b2}^{(k)}$	35.30	8.365	(19.06, 49.06)	35.20	8.407	(18.91, 49.07)
Scale (slope)	$\sigma_{b2}^{(k)}$	0.051	0.002	(0.048, 0.055)	0.049	0.002	(0.046, 0.053)

Table 4. Correlations between the random effects

Correlation	Mean	SD	(95% CI)
$\text{Corr}(b_{i1}^{(1)}, b_{i2}^{(1)})$	-0.031	0.050	$(-0.128, 0.068)$
$\text{Corr}(b_{i1}^{(2)}, b_{i2}^{(2)})$	-0.073	0.049	$(-0.168, 0.024)$
$\text{Corr}(b_{i1}^{(1)}, b_{i1}^{(2)})$	0.707	0.049	(0.604, 0.797)
$\text{Corr}(b_{i2}^{(1)}, b_{i2}^{(2)})$	0.057	0.047	$(-0.034, 0.148)$

In addition, we present the correlations between the random effects in Table 4. The first two rows are the correlations between the random intercept and random slope for the two outcomes. We see that for both CD4 and CD8, the random intercepts and random slopes are negatively correlated, meaning that the subjects with higher cell counts at baseline tend to have a faster decreasing rate of cell counts, although the effect is not significant at the 0.05 level. The third row shows the correlation of random intercepts for the two outcomes, which is highly significant with 95% CI (0.604, 0.797). This means that the number of baseline CD4 and CD8 cell counts are strongly correlated. The last row shows the correlation of the two random slopes, which is not significant. This indicates that the decreasing rates for CD4 and CD8 cell counts are not highly correlated.

4.2 Behavioral factors related to sexually transmitted infections

Understanding human sexual behaviors is essential for the effective prevention of sexually transmitted infections (STI). Young women were recruited for participation in a behavioral epidemiological study from three urban primary care clinics. The overall objective of the study was to examine the behavioral factors related to STIs. Eligibility criteria included that the women aged between 14 and 17 years, be able to understand English, have no serious psychiatric disturbances or mental handicaps, and attend one of the three recruiting clinics. All participating subjects had quarterly clinic visits for the duration of the study period and they also completed daily behavioral diaries, which provided detailed records of the subject's sexual behaviors in their original time sequence. Specifically, the diary was a structured minisurvey in which the subject reported sexual intercourse, condom protection, STI symptoms, and daily mood and sexual interest. For details of the study, see Ghosh and Tu (2009).

In this analysis, we used a subset of 282 subjects who had been enrolled into the study for at least 6 months (24 weeks). We considered a bivariate outcome $Y_{ij} = (\text{MOOD}_{ij}, \text{SI}_{ij})^T$, where MOOD_{ij} and SI_{ij} were, respectively, the weekly average mood and sexual interest (SI) scores reported by the ith subject in the jth week. The subject characteristics that we considered in this analysis included age, lifetime number of partners, and history of STI, all measured at enrollment. The STI history is thought to be a marker of the more risky sexual behaviors in young women, and the lifetime number of partners to be a marker of a subject's sexual experience and partner availability.

As an initial analysis, we fit a normal LMM to the two outcomes using covariates, time, number of partners in life (NP), age, status of having sexually transmitted disease (STD) and previous outcomes. Thus, the model for mood and sexual interest is:

$$y_{ij}^{(k)} = \beta_1^{(k)} + \beta_2^{(k)}\texttt{time} + \beta_3^{(k)}\texttt{NP} + \beta_4^{(k)}\texttt{Age} + \beta_5^{(k)}\texttt{STD}$$
$$+ \beta_6^{(k)}y_{i,j-1}^{(k)} + b_i^{(k)} + e_{ij}^{(k)}, k = 1, 2.$$
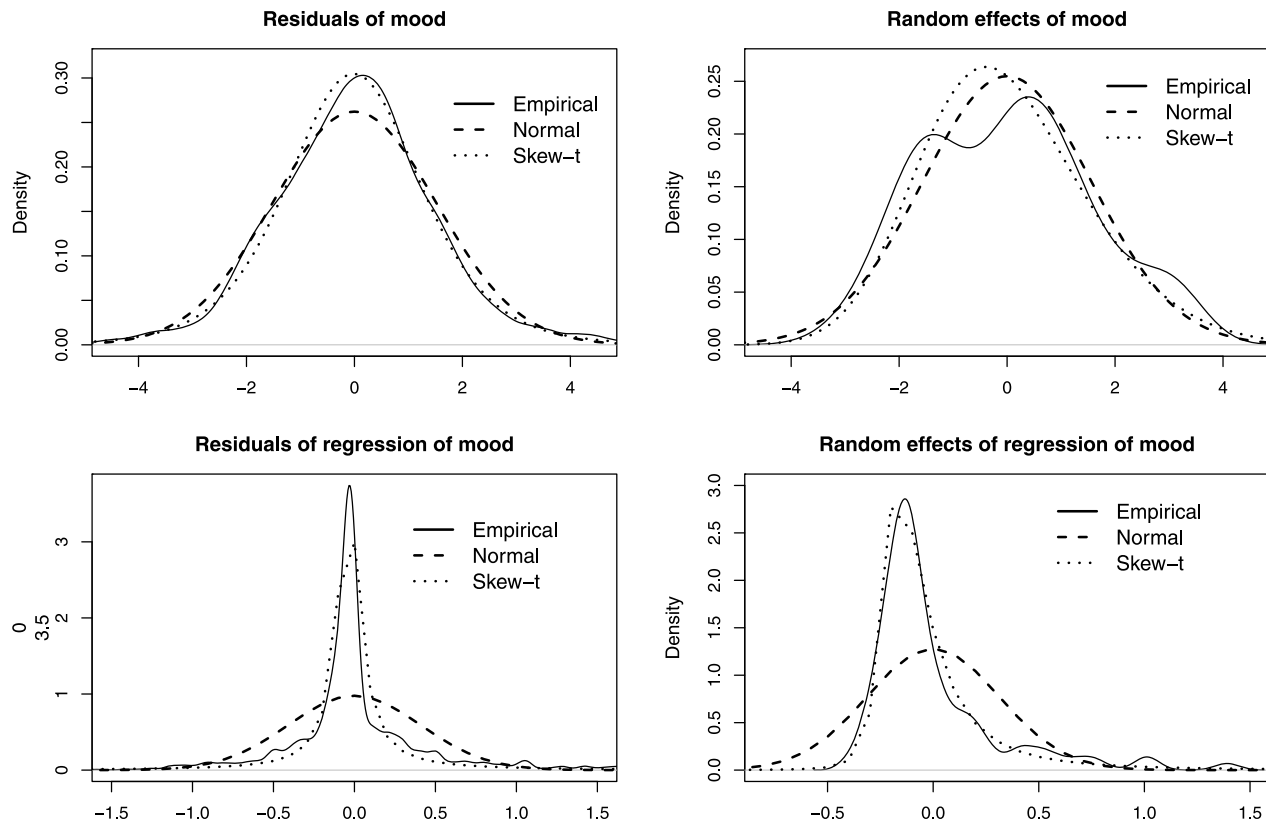
Figure 3. Distributions of the random effects and residuals from the normal LMM for the STI study.

The autoregressive structures embedded in the model allowed us to examine the strength of the autocorrelation within the outcomes. This was not only of scientific interest to the investigation, but also helpful for the exploration of the modeling structure. For example, a very strong autocorrelation in mood would not only counter the speculation of a mood swing in adolescents, but also render it unnecessary to collect mood measurements so frequently. The specification implied a first order autoregressive model. We chose the AR(1) model in part due to its simplicity. In general, the order of an autoregressive model can be selected using the AIC or BIC criteria (Shibata, 1976). Under the Bayesian framework, one can use the Deviance information criterion.

Figure 3 shows the empirical distributions of the random effects and residuals errors for sexual interest and mood from the normal LMM. The densities of the fitted normal and skew-*t* distributions are also presented. We see clearly that the fit from the skew-*t* distributions is much better than that from the normal distributions, indicating that use of a skewed distribution is warranted.

The priors for the parameters are specified as in the application for HIV markers. The parameter estimates of the LMM using extended MST are shown in Table 5. Again, the estimates of regression coefficients are presented in the top section. We see that follow-up time and the number of partners are positively associated with mood. Age and having an STD are negatively associated with mood. But none of the associations are statistically significant. The most significant factor that is associated with current mood is the mood in previous interview. For sexual interest, we see a similar pattern. Previous observation of sexual interest is the most significant predictor of current sexual interest. The other four variables, i.e., time, number of partners, age and having an STD, are not statistically significant. The parameters of the MST distributions for random errors and random effects are shown in the bottom. We see that both the random errors and random effects present various degrees of skewness. For example, the skewness parameter of the random effect for mood is 2.824 with 95% CI (1.781, 3.748). In addition, both random errors and random effects show heavier tails than the normal distribution. The correlation parameter between mood and sexual interest is positive but not significant.

5. DISCUSSION

We proposed an LMM with extended MST distributions, which is more computationally intensive compared to the normal LMM. However, the proposed model enhances modeling flexibility and allows applied researchers to analyze correlated outcomes with a wide range of choices. It is a useful tool to handle longitudinal or clustered outcomes with asymmetric or heavy-tailed distributions. In this paper, we

Variable		$Y^{(1)} =$ Mood			$Y^{(2)} =$ Sexual Interest		
		Mean	SD	(95% CI)	Mean	SD	(95% CI)
Intercept	$\beta_1^{(k)}$	6.044	0.500	(5.092, 7.049)	1.059	0.043	(0.975, 1.146)
Time	$\beta_2^{(k)}$	0.019	0.014	(−0.008, 0.046)	−0.000	0.001	(−0.002, 0.001)
Number of partners	$\beta_3^{(k)}$	0.130	0.114	(−0.087, 0.357)	0.015	0.013	(−0.011, 0.041)
Age	$\beta_4^{(k)}$	−0.021	0.224	(−0.468, 0.417)	0.001	0.023	(−0.044, 0.047)
STD	$\beta_5^{(k)}$	−0.017	0.506	(−1.033, 0.960)	−0.021	0.052	(−0.119, 0.084)
Previous Obs.	$\beta_6^{(k)}$	0.305	0.038	(0.230, 0.382)	0.055	0.017	(0.024, 0.089)
Parameters of skew-t distributions for random errors							
Skewness	$\lambda_e^{(k)}$	1.364	0.145	(1.079, 1.645)	0.115	0.009	(0.099, 0.133)
DF	$\nu_e^{(k)}$	3.620	0.480	(3.026, 4.802)	3.016	0.016	(3.000, 3.060)
Scale	$\sigma_e^{(k)}$	0.735	0.078	(0.590, 0.896)	0.037	0.002	(0.033, 0.041)
Parameters of extended MST for random intercepts							
Skewness	$\lambda_b^{(k)}$	2.824	0.514	(1.781, 3.748)	0.102	0.045	(0.019, 0.194)
DF	$\nu_b^{(k)}$	17.31	9.705	(4.644, 41.82)	3.363	0.373	(3.009, 4.374)
Scale	$\sigma_b^{(k)}$	0.759	0.263	(0.378, 1.402)	0.169	0.018	(0.138, 0.209)
Correlation	ρ	0.047	0.153	(−0.256, 0.341)			

only consider the extended MST distribution for the random effects. In the same spirit, we can use extended MST distribution for the error terms. This will increase the complexity of the analysis, however. In the situations where the multiple outcomes have more than one mode, a finite mixture of univariate and multivariate skew-t distributions may be used for multivariate outcomes.

In the analysis, we considered up to four random effects for two correlated outcomes. Conceptually, the proposed method can be applied to more than four random effect. However, the number of parameters of the unstructured correlation matrix for q random effects is roughly $q^2/2$. The computational burden will increase substantially for high-dimensional random effects. In practice, we suggest not to use a large number, say 10, of random effects.

ACKNOWLEDGEMENTS

REFERENCES

ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88** 669–679. MR1224394

AZZALINI, A. and CAPITANIO, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **61** 579–602. MR1707862

AZZALINI, A. and CAPITANIO, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **65** 367–389. MR1983753

AZZALINI, A. and GENTON, M. G. (2008). Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review* **76** 106–129.

BESAG, J., GREEN, P., HIGDON, D. and MENGERSEN, K. (1995). Bayesian computation and stochastic systems. *Statistical Science* **10** 3–66. MR1349818

BOSCARDIN, W. J., TAYLOR, J. M. G. and LAW, N. (1998). Longitudinal models for AIDS marker data. *Statistical Methods in Medical Research* **7** 13–27.

BRANCO, M. D. and DEY, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis* **79** 99–113. MR1867257

CHIB, S. and GREENBERG, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician* **49** 327–335.

FRUHWIRTH-SCHNATTER, S. and PYNE, S. (2010). Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions. *Biostatistics* **11** 317–336.

GELMAN, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1** 1–19. MR2221284

GELMAN, A. and HILL, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*, 1st ed. Cambridge University Press.

GELMAN, A. and RUBIN, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7** 457–472.

GEWEKE, J. (1993). Bayesian treatment of the independent Student-t linear model. *Journal of Applied Econometrics* **8** S19–S40.

GHOSH, P., BRANCO, M. D. and CHAKRABORTY, H. (2006). Bivariate random effect model using skew-normal distribution with application to HIV-RNA. *Statistics in Medicine* **26** 1255–1267. MR2345719

GHOSH, P. and HANSON, T. (2010). A semiparametric Bayesian approach to multivariate longitudinal data. *Australian and New Zealand Journal of Statistics* **52** 275–288. MR2744574

GHOSH, P. and TU, W. (2009). Assessing sexual attitudes and behaviors of young women: A joint model with nonlinear time effects, time varying covariates, and dropouts. *Journal of the American Statistical Association* **104** 474–485. MR2751432

GURKA, M. J., EDWARDS, L. J., MULLER, K. E. and KUPPER, L. L. (2006). Extending the Box-Cox transformation to the linear mixed model. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **169** 273–288. MR2225543

HO, H. J. and LIN, T. I. (2010). Robust linear mixed models using the skew t distribution with application to schizophrenia data. *Biometrical Journal* **52** 449–469. MR2751913

JACQMIN-GADDA, H., SIBILLOT, S., PROUST, C., MOLINA, J. M. and THIABAUT, R. (2007). Robustness of the linear mixed model to misspecified error distribution. *Computational Statistics and Data Analysis* **51** 5142–5154. MR2370713

JARA, A., QUINTANA, F. and SAN MARTIN, E. (2008). Linear mixed models with skew-elliptical distributions: A Bayesian approach. *Computational Statistics and Data Analysis* **52** 5033–5045. MR2526212

JONES, M. C. (2002). A dependent bivariate t distribution with marginals on different degrees of freedom. *Statistics and Probability Letters* **56** 163–170. MR1881170

LACHOS, V. H., GHOSH, P. and ARELLANO-VALLE, R. B. (2010). Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica* **20** 303–322. MR2640696

LAIRD, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association* **73** 805–811. MR0521328

LIN, T. I. (2010). Robust mixture modeling using multivariate skew t distributions. *Statistics and Computing* **20** 343–356. MR2725392

LIN, T. I. and LEE, J. C. (2008). Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. *Statistics in medicine* **27** 1490–1507. MR2420253

LITIERE, S., ALONSO, A. and MOLENBERGHS, G. (2007). Type I and type II error under random-effects misspecification in generalized linear mixed models. *Biometrics* **63** 1038–1044. MR2414580

LITIERE, S., ALONSO, A. and MOLENBERGHS, G. (2011). Rejoinder to a note on type II error under random effects misspecification in generalized linear mixed models. *Biometrics* **67** 656–660.

LUNN, D. J., SPIEGELHALTER, D., THOMAS, A. and BEST, N. (2009). The BUGS project: Evolution, critique and future directions (with discussion). *Statistics in Medicine* **28** 3049–3082. MR2750401

MÜLLER, P. and ROSNER, G. L. (1997). A Bayesian population model with hierarchical mixture priors applied to blood count data. *Journal of the American Statistical Association* **92** 1279–1292.

NEUHAUS, J. M., MCCULLOCH, C. E. and BOYLAN, R. (2011). A note on type II error under random effects misspecification in generalized linear mixed models. *Biometrics* **67** 654–656. MR2829095

ODELL, P. L. and FEIVESON, A. H. (1966). A numerical procedure to generate a sample covariance matrix. *Journal of the American Statistical Association* **61** 199–203. MR0192635

O'MALLEY, A. J. and ZOU, K. H. (2006). Bayesian multivariate hierarchical transformation models for ROC analysis. *Statistics in Medicine* **25** 459–479. MR2225177

PANAGIOTELIS, A. and SMITH, M. (2008). Bayesian density forecasting of intraday electricity prices using multivariate skew t distributions. *International Journal of Forecasting* **24** 710–727. Energy Forecasting.

PANTAZIS, N., TOULOUMI, G., WALKER, A. S. and BABIKER, A. G. (2005). Bivariate modelling of longitudinal measurements of two human immunodeficiency type 1 disease progression markers in the presence of informative drop-outs. *Journal of the Royal Statistical Society. Series C: Applied Statistics* **54** 405–423. MR2135882

SAMMEL, M., LIN, X. and RYAN, L. (1999). Multivariate linear mixed models for multiple outcomes. *Statistics in medicine* **18** 2479–2492.

SHAH, A., LAIRD, N. and SCHOENFELD, D. (1997). A random-effects model for multiple characteristics with possibly missing data. *Journal of the American Statistical Association* **92** 775–779. MR1467867

SHAW, W. T. and LEE, K. T. A. (2008). Bivariate Student t distributions with variable marginal degrees of freedom and independence. *Journal of Multivariate Analysis* **99** 1276–1287. MR2419347

SHIBATA, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* **63** 117–126. MR0403130

SPIEGELHALTER, D., THOMAS, A., BEST, N. and LUNN, D. (2003). WinBUGS user manual. Version 1.4. Cambridge: MRC Biostatistics Unit.

THIEBAUT, R., JACQMIN-GADDA, H., CHENE, G., LEPORT, C. and COMMENGES, D. (2002). Bivariate linear mixed models using SAS proc MIXED. *Computer Methods and Programs in Biomedicine* **69** 249–256.

THIEBAUT, R., JACQMIN-GADDA, H., BABIKER, A. and COMMENGES, D. (2005). Joint modelling of bivariate longitudinal data with informative dropout and left-censoring, with application to the evolution of CD4+ cell count and HIV RNA viral load in response to treatment of HIV infection. *Statistics in Medicine* **24** 65–82. MR2134496

VERBEKE, G. and LESAFFRE, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association* **91** 217–221.

VERBEKE, G. and MOLENBERGHS, G. (2001). *Linear Mixed Models for Longitudinal Data.* Springer. MR1880596

WANG, W. L. and FAN, T. H. (2010). ECM-based maximum likelihood inference for multivariate linear mixed models with autoregressive errors. *Computational Statistics and Data Analysis* **54** 1328–1341. MR2600835

WANG, W. L. and FAN, T. H. (2011). Estimation in multivariate t linear mixed models for multiple longitudinal data. *Statistica Sinica* **21** 1857–1880. MR2896002

WANG, W. L. and FAN, T. H. (2012). Bayesian analysis of multivariate t linear mixed models using a combination of IBF and Gibbs samplers. *Journal of Multivariate Analysis* **105** 300–310. MR2877519

WANG, K., NG, S.-K. and MCLACHLAN, G. J. (2009). Multivariate skew t mixture models: Applications to fluorescence-activated cell sorting data. *Digital Image Computing: Techniques and Applications* **0** 526–531.

ZHANG, D. and DAVIDIAN, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics* **57** 795–802. MR1859815

Binbing Yu

7201 Wisconsin Ave. Suite 3C309

Laboratory of Epidemiology and Population Sciences

National Institute on Aging

Bethesda, MD 20904

USA

E-mail address: whybb@yahoo.com

A. James O'Malley

Harvard Medical School

Department of Health Care Policy

180 Longwood Avenue

Boston, MA 02115

USA

E-mail address: omalley@hcp.med.harvard.edu

Pulak Ghosh

Department of Quantitative Methods
  and Information Sciences

Indian Institute of Management, Bangalore

Bannerghatta Road, 560076, Bangalore

India

E-mail address: pulakghoshc@gmail.com