Statistica Sinica **20** (2010), 303-322

LIKELIHOOD BASED INFERENCE FOR SKEW-NORMAL INDEPENDENT LINEAR MIXED MODELS

Victor H. Lachos, Pulak Ghosh and Reinaldo B. Arellano-Valle

Universidade Estadual de Campinas, Indian Institute of Management, Bangalore and Pontificia Universidad Católica de Chile

Abstract: Linear mixed models with normally distributed response are routinely used in longitudinal data. However, the accuracy of the assumed normal distribution is crucial for valid inference of the parameters. We present a new class of asymmetric linear mixed models that provides for an efficient estimation of the parameters in the analysis of longitudinal data. We assume that, marginally, the random effects follow a multivariate skew-normal/independent distribution (Branco and Dey (2001)) and that the random errors follow a symmetric normal/independent distribution (Lange and Sinsheimer (1993)), providing an appealing robust alternative to the usual symmetric normal distribution in linear mixed models. Specific distributions examined include the skew-normal, the skew-t, the skew-slash, and the skew-contaminated normal distribution. We present an efficient EM-type algorithm for the computation of maximum likelihood estimation of parameters. The technique for the prediction of future responses under this class of distributions is also investigated. The methodology is illustrated through an application to Framingham cholesterol data and a simulation study.

Key words and phrases: EM-algorithm, linear mixed models, skew-normal/independent distributions, skewness.

1. Introduction

Linear mixed models (LMM; Laird and Ware (1982)) are the most frequently used tool for longitudinal data analysis with continuous repeated measures. In a LMM framework it is routinely assumed that the random effects and the within-subject measurement errors have a normal distribution. While this assumption makes the model easy to apply in widely used software such as SAS, its accuracy is difficult to check and the routine use of normality has been questioned by many authors (Verbeke and Lesaffre (1997), Pinheiro, Liu and Wu (2001), Zhang and Davidian (2001), Ghidey, Lesaffre and Eilers (2004), and Lin and Lee (2008)). The normality assumption suffers from a lack of robustness against departures from the normal, particularly when data show multimodality and skewness, and may not provide an accurate estimation of between-subject variation. For example, Zhang and Davidian (2001) showed that the estimated subject-specific intercept from the Framingham heart study data was not normally distributed and that the use of the normal distribution in this scenario may obscure important features of between-subject variation. Thus it is of practical interest to develop statistical model with considerable flexibility in the distributional assumptions of the random effects, as well as the error terms.

There has been considerable work in this direction. Verbeke and Lesaffre (1996) introduce a heterogeneous linear mixed model where the random effects distribution is relaxed using normal mixtures. Pinheiro, Liu and Wu (2001) proposed a multivariate t linear mixed (TLMM) model and showed that it performed well in the presence of outliers. Lin and Lee (2006, 2007) developed some additional tools for TLMM from the likelihood-based and Bayesian perspective. Zhang and Davidian (2001) proposed a LMM in which the random effects follow the so-called semi-nonparametric (SNP) distribution. Rosa, Padovani and Gianola (2003) adopted a Bayesian framework to carry out posterior analysis in LMM with the thick-tailed class of normal/independent (NI) distributions (Lange and Sinsheimer (1993)). Ghidey et al. (2004) developed a LMM with a smooth random effects density. Ma, Genton and Davidian (2004) considered a generalized flexible skew-elliptical distribution for the random effects density and proposed somewhat complicated algorithms for maximum likelihood (ML) estimation and Bayesian inference via Markov Chain Monte Carlo (MCMC). Recently, Arellano-Valle, Bolfarine and Lachos (2005a), Lin and Lee (2008), and Lachos, Bolfarine, Arellano-Valle and Montenegro (2007) proposed a skew-normal linear mixed model (SN-LMM) based on multivariate skew-normal (SN) distribution introduced by Azzalini and Dalla-Valle (1996). They also developed an EM-type algorithm for maximum likelihood estimation (MLE). A common feature of these classes of LMMs is that the normal linear mixed model (N-LMM) is a member of the class.

In this paper we propose a parametric robust modeling of LMM based on skew-normal/independent (SNI) distributions. In particular, we assume a SNI distribution for the random effects, and a NI distribution for the within-subject errors. Together, the observed responses follow a SNI distribution and define what we call a skew-normal/independent linear mixed model (SNI-LMM). Particularly, the SNI distributions provide a group of skew-thick-tailed distributions that are useful for robust inference and that contain as proper elements the skew-normal (SN), the skew-t (ST), the skew-slash (SSL), and the skew-contaminated normal (SCN) distributions. The marginal density of the observed quantities are obtained analytically by integrating out the random effects, leading to an observed (marginal) likelihood function that can be maximized directly by using existing statistical software such as Ox, R or Matlab. The hierarchical representation of the proposed model makes possible the implementation of an EM-type algorithm, which for special cases and common situations yields "closed form" expressions for the E and M-steps. We further analyze the longitudinal Framingham cholesterol data whose random effects distribution has been found to be non-normal and positively skewed by Zhang and Davidian (2001), Ghidey et al. (2004), and Lin and Lee (2008).

The rest of the article is organized as follows. After a brief introduction to SNI distributions in Section 2, the SNI-LMM is presented in Section 3; a likelihood-based methodology is used for estimation and inference, including the estimation of the random effects and the prediction of future values. In Section 4, a simulation study is conducted to examine the performance of the estimation for subject-specific random effects and for prediction of futures values. The advantage of the proposed methodology is illustrated through the Framingham cholesterol data in Section 5, and some concluding remarks are presented in Section 6.

2. Skew-Normal/Independent Distributions

To better motivate our proposed methodology, we give a brief introduction of SNI distributions, starting with a definition of the SN-distribution. We say that a $p \times 1$ random vector $\mathbf{Y}$ follows a SN-distribution with $p \times 1$ location vector $\boldsymbol{\mu}$, $p \times p$ positive definite dispersion matrix $\boldsymbol{\Sigma}$, and $p \times 1$ skewness parameter vector $\boldsymbol{\lambda}$, and write $\mathbf{Y} \sim SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, if its probability density function (pdf) is

$$f(\mathbf{y}) = 2\phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi(\boldsymbol{\lambda}^\top \mathbf{y}_0), \tag{2.1}$$

where $\mathbf{y}_0 = \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$, $\phi_p(.; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for the pdf of the p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariate matrix $\boldsymbol{\Sigma}$, $N_{\mathrm{p}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ say, and $\Phi(.)$ is the cumulative distribution function (cdf) of the standard univariate normal. Note for $\boldsymbol{\lambda} = \mathbf{0}$ that (2.1) reduces to the symmetric $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$-pdf, while for non-zero values of $\boldsymbol{\lambda}$, it produces a perturbed (asymmetric) family of $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$-pdf's. Except for a straightforward difference in the parametrization considered in (2.1), this model corresponds to that introduced by Azzalini and Dalla-Valle (1996), with properties extensively studied in Azzalini and Capitanio (1999), and in Arellano-Valle and Genton (2005b). Let $\mathbf{Z} = \mathbf{Y} - \boldsymbol{\mu}$. Since $a\mathbf{Z} \sim SN_p(\mathbf{0}, a^2\boldsymbol{\Sigma}, \boldsymbol{\lambda})$, for all $a > 0$, the SNI family can be defined as follows: a SNI distribution is that of a $p-$dimensional random vector

$$\mathbf{Y} = \boldsymbol{\mu} + U^{-1/2}\mathbf{Z}, \tag{2.2}$$

where U is a positive random variable with the cdf $H(u;\boldsymbol{\nu})$ and pdf $h(u;\boldsymbol{\nu})$, and independent of the $SN_p(\mathbf{0},\boldsymbol{\Sigma},\boldsymbol{\lambda})$-random vector $\mathbf{Z}$. Here $\boldsymbol{\nu}$ is a scalar or vector parameter indexing the distribution of the mixing scale factor U. Given $U=u$, $\mathbf{Y}$ follows a multivariate skew-normal distribution with location vector $\boldsymbol{\mu}$, scale matrix $u^{-1}\boldsymbol{\Sigma}$, and skewness parameter vector $\boldsymbol{\lambda}$, i.e., $\mathbf{Y}|U=u\sim \mathrm{SN}_p(\boldsymbol{\mu},u^{-1}\boldsymbol{\Sigma},\boldsymbol{\lambda})$. In other words, the SNI distributions are scale mixtures of the skew-normal distribution, where the distribution of the scale factor U is the mixing distribution. Thus, by (2.1), the marginal pdf of $\mathbf{Y}$ is

$$f(\mathbf{y})=2\int_0^\infty \phi_p(\mathbf{y};\boldsymbol{\mu},u^{-1}\boldsymbol{\Sigma})\Phi(u^{1/2}\boldsymbol{\lambda}^\top\mathbf{y}_0)dH(u;\boldsymbol{\nu}), \tag{2.3}$$

where $\mathbf{y}_0=\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})$. The notation $\mathbf{Y}\sim\mathrm{SNI}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},H)$ will be used when $\mathbf{Y}$ has pdf (2.3). When $\boldsymbol{\lambda}=\mathbf{0}$, the SNI distributions reduces to the normal-independent (NI) class, i.e., the class of scale-mixtures of the normal distribution represented by the pdf $f_0(\mathbf{y})=\int_0^\infty\phi_p(\mathbf{y};\boldsymbol{\mu},u^{-1}\boldsymbol{\Sigma})dH(u;\boldsymbol{\nu})$. We use the notation $\mathbf{Y}\sim\mathrm{NI}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},H)$ when $\mathbf{Y}$ has distribution in the NI class.

The asymmetrical class of SNI distributions includes the skew-t, the skew-slash, and the skew-contaminated normal. All these distributions have heavier tails than the skew-normal and can be used for robust inferences. Some of these distributions are described subsequently. For each element of this class, we also compute the conditional moments $u_r=\mathrm{E}\{U^r|\mathbf{y}\}$ and $\tau_r=\mathrm{E}\{U^{r/2}W_\Phi(U^{1/2}\mathrm{A})|\mathbf{y}\}$, where $\mathrm{A}=\boldsymbol{\lambda}^\top\mathbf{y}_0$ and $W_\Phi(x)=\phi_1(x)/\Phi(x)$, $x\in\mathbb{R}$; these are useful in the implementation of the EM-algorithm, the estimation of the random effects and the prediction of futures values. The proof of the following results given in Appendix A of the supplementary material.

Proposition 1. *Let* $\mathbf{Y}\sim SNI_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},H)$ *and let* $U\sim H$ *be the mixing random scale factor. Then*

$$u_r=\frac{2f_0(\mathbf{y})}{f(\mathbf{y})}\mathrm{E}\{U_{\mathbf{y}}^r\Phi(U_{\mathbf{y}}^{1/2}\mathrm{A})\}\quad\text{and}\quad \tau_r=\frac{2f_0(\mathbf{y})}{f(\mathbf{y})}\mathrm{E}\{U_{\mathbf{y}}^{r/2}\phi_1(U_{\mathbf{y}}^{1/2}\mathrm{A})\},$$

where $A=\boldsymbol{\lambda}^\top\mathbf{y}_0$ *with* $\mathbf{y}_0=\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})$, f_0 *is the pdf of* $\mathbf{Y}_0\sim NI_p(\boldsymbol{\mu},\boldsymbol{\Sigma},H)$, *and* $U_{\mathbf{y}}\overset{d}{=}U|\mathbf{Y}_0=\mathbf{y}$.

2.1. Multivariate skew-t distribution

The multivariate skew-t distribution (Branco and Dey (2001), and Azzalini and Capitanio (2003)) with ν degrees of freedom, $\mathrm{ST_p}(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)$ say, can be derived from the mixture model (2.3), by taking $U\sim Gamma(\nu/2,\nu/2)$, $\nu>0$.

The pdf of $\mathbf{Y}$ is

$$f(\mathbf{y}) = 2t_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) T\left(\sqrt{\frac{\nu+p}{\nu+d}}\mathrm{A}; \nu+p\right), \quad \mathbf{y} \in \mathbb{R}^p, \tag{2.4}$$

where $t_p(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ and $T(\cdot; \nu)$ denote, respectively, the pdf of the p-variate Student-t distribution, namely $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ and the cdf of the standard univariate t-distribution, and $d = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ is the Mahalanobis distance. A particular case of the skew-t distribution is the skew-Cauchy distribution with $\nu = 1$. Also, as $\nu \uparrow \infty$, we get the skew-normal distribution as the limiting case. Applications of the skew-t distribution to robust estimation can be found in Lin, Lee and Hsieh (2007) and Azzalini and Genton (2007). Moreover, for this model we have in Proposition 1 that $\mathbf{Y}_0 \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, i.e., $\mathbf{Y}_0 | U = u \sim N_p(\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})$ and $U \sim Gamma(\nu/2, \nu/2)$. From the fact that $U_{\mathbf{y}} \overset{d}{=} U|\mathbf{Y}_0 = \mathbf{y} \sim Gamma((\nu + p)/2, (\nu + \mathrm{d})/2)$, we find the conditional expectations of u_r and τ_r.

Corollary 1. *Suppose* $\mathbf{Y} \sim ST_p(\mu, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$. *Then*

$$u_r = \frac{f_0(\mathbf{y})}{f(\mathbf{y})} \frac{2^{r+1}\Gamma((\nu+p+2r)/2)(\nu+\mathrm{d})^{-r}}{\Gamma((\nu+p)/2)} T\left(\sqrt{\frac{\nu+p+2r}{\nu+\mathrm{d}}}\mathrm{A}; \nu+p+2r\right), \ \textit{and}$$

$$\tau_r = \frac{f_0(\mathbf{y})}{f(\mathbf{y})} \frac{2^{(r+1)/2}\Gamma((\nu+p+r)/2)}{\pi^{1/2}\Gamma((\nu+p)/2)} \frac{(\nu+\mathrm{d})^{(\nu+p)/2}}{(\nu+\mathrm{d}+\mathrm{A}^2)^{(\nu+p+r)/2}}.$$

2.2. Multivariate skew-slash distribution

Another SNI distribution, termed as the multivariate skew-slash distribution and denoted by $\mathrm{SSL}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$, arises when the distribution of U is $Beta(\nu, 1)$, $\nu > 0$. Its pdf is given by

$$f(\mathbf{y}) = 2\nu \int_0^1 u^{\nu-1}\phi_p(\mathbf{y}; \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})\Phi(u^{1/2}\mathrm{A})du, \quad \mathbf{y} \in \mathbb{R}^p. \tag{2.5}$$

The skew-slash distribution reduces to the skew-normal distribution as $\nu \uparrow \infty$. The conditional moments u_r and τ_r for the skew-slash distribution (2.5) follow by considering in Proposition 1 that $U_{\mathbf{y}} \sim Gamma((2\nu + p + 2r)/2, d/2)\mathbb{I}_{(0,1)}$. Applications of the skew-slash distribution can be found in Wang and Genton (2006).

Corollary 2. *Suppose* $\mathbf{Y} \sim SSL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$. *Then*

$$u_r = \frac{f_0(\mathbf{y})}{f(\mathbf{y})} \frac{2\Gamma((2\nu+p+2r)/2)}{\Gamma((2\nu+p)/2)} \left(\frac{2}{\mathrm{d}}\right)^r \frac{P_1((2\nu+p+2r)/2, \mathrm{d}/2)}{P_1((p+2\nu)/2, d/2)} \mathrm{E}\left\{\Phi(S^{1/2}\mathrm{A})\right\}, \ \textit{and}$$

$$\tau_r = \frac{f_0(\mathbf{y})}{f(\mathbf{y})} \frac{2^{r/2+1/2}\Gamma((2\nu+p+r)/2)}{\Gamma((2\nu+p)/2)\pi^{1/2}} \frac{\mathrm{d}^{(2\nu+p)/2}}{(\mathrm{d}+\mathrm{A}^2)^{(2\nu+p+r)/2}} \frac{P_1((2\nu+p+r)/2,(\mathrm{d}+\mathrm{A}^2)/2)}{P_1((p+2\nu)/2,d/2)},$$

where $P_x(a,b)$ denotes the cdf of the Gamma(a,b) distribution evaluated at x and $S \sim Gamma((2\nu+p+2r)/2, d/2)\mathbb{I}_{(0,1)}$.

We note in Corollary 2 that $\mathrm{E}\{\Phi(S^{1/2}\mathrm{A})\}$ can be computed as $\mathrm{E}\{\Phi(S_0^{1/2}\mathrm{A}) |0 < S_0 < 1\}$, where $S_0 \sim Gamma((2\nu+p+2r)/2, d/2)$, and can be approximated by Monte Carlo integration as follows: generate L samples $S_1,\ldots,S_L$ from $0 < S_0 < 1$, then approximate $\mathrm{E}\{\Phi(S^{1/2}\mathrm{A})\}$ directly by $(1/L)\sum_{i=1}^{L}\Phi(S_iA)$. In Appendix B, we give an algorithm to generate a truncated gamma random variable.

2.3. Multivariate skew-contaminated normal distribution

The multivariate skew-contaminated normal distribution arises when the mixing scale factor U is a discrete random variable taking one of two values. The pdf of U, given a parameter vector $\boldsymbol{\nu} = (\nu_1, \nu_2)^\top$, is

$$nh(u;\boldsymbol{\nu}) = \nu_1\mathbb{I}_{(u=\nu_2)} + (1-\nu_1)\mathbb{I}_{(u=1)}, \quad 0 < \nu_1 < 1,\ 0 < \nu_2 \le 1. \tag{2.6}$$

It follows

$$f(\mathbf{y}) = 2\Big\{\nu_1\phi_p(\mathbf{y};\boldsymbol{\mu},\nu_2^{-1}\boldsymbol{\Sigma})\Phi(\nu_2^{1/2}\mathrm{A}) + (1-\nu_1)\phi_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma})\Phi(\mathrm{A})\Big\}.$$

This distribution is denoted by $SCN_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu_1,\nu_2)$, $0 < \nu_1 < 1$, $0 < \nu_2 \le 1$. Parameter ν_1 can be interpreted as the proportion of outliers, while ν_2 may be interpreted as a scale factor. The skew-contaminated normal distribution reduces to the skew-normal distribution when $\nu_1 = \nu_2 = 1$. In this case, considering that $U_{\mathbf{y}}$ is a discrete random variable with conditional probability function $h_0(u|\mathbf{y}) = (1/f_0(\mathbf{y}))\{\nu_1\phi_p(\mathbf{y};\boldsymbol{\mu},\nu_2^{-1}\boldsymbol{\Sigma})\mathbb{I}_{(u=\nu_2)} + (1-\nu_1)\phi_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma})\mathbb{I}_{(u=1)}\}$, Proposition 1 yields the following.

Corollary 3. *Suppose $\mathbf{Y} \sim SCN_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu_1,\nu_2)$. Then*

$$u_r = \frac{2}{f(\mathbf{y})}\Big\{\nu_1\nu_2^r\phi_p(\mathbf{y};\boldsymbol{\mu},\nu_2^{-1}\boldsymbol{\Sigma})\Phi(\nu_2^{1/2}\mathrm{A}) + (1-\nu_1)\phi_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma})\Phi(\mathrm{A})\Big\}, \quad \textit{and}$$
$$\tau_r = \frac{2}{f(\mathbf{y})}\Big\{\nu_1\nu_2^{r/2}\phi_p(\mathbf{y};\boldsymbol{\mu},\nu_2^{-1}\boldsymbol{\Sigma})\phi_1(\nu_2^{1/2}\mathrm{A}) + (1-\nu_1)\phi_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma})\phi_1(\mathrm{A})\Big\}.$$

3. The Skew-Normal/Independent Linear Mixed Model

We consider a generalization of N-LMM in which the within-subject errors are assumed to follow a NI distribution and the random effects are assumed to have a multivariate SNI distribution within the class (2.3). Simultaneously, the model can be written as

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \tag{3.1}$$

with the assumption that

$$\begin{pmatrix} \mathbf{b}_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \overset{ind.}{\sim} \mathrm{SNI}_{n_i+q}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_i \end{pmatrix}, \begin{pmatrix} \boldsymbol{\lambda} \\ \mathbf{0} \end{pmatrix}, H\right), i = 1, \ldots, n, \tag{3.2}$$

where the subscript i is the subject index, $\mathbf{Y}_i$ is a $n_i \times 1$ vector of observed continuous responses for subject i, $\mathbf{X}_i$ is the $n_i \times p$ design matrix corresponding to the fixed effects, $\boldsymbol{\beta}$ is a $p \times 1$ vector of population-averaged regression coefficients called fixed effects, $\mathbf{Z}_i$ is the $n_i \times q$ design matrix corresponding to the $q \times 1$ vector of random effects $\mathbf{b}_i$, and $\boldsymbol{\epsilon}_i$ is the $n_i \times 1$ vector of random errors. The matrices $\mathbf{D} = \mathbf{D}(\boldsymbol{\alpha})$ and $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\boldsymbol{\gamma})$, $i = 1, \ldots, n$, are dispersion matrices corresponding to the between and within subjects variability, and depend on unknown and reduced parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, respectively. Finally, as was indicated in the previous section, $H = H(\cdot; \boldsymbol{\nu})$ is the cdf-generator that determines the specific SNI model that we assume.

Remarks.

(i) From Lemma 1 in Appendix A it follows that, marginally,

$$\mathbf{b}_i \overset{iid}{\sim} \mathrm{SNI}_q(\mathbf{0}, \mathbf{D}, \boldsymbol{\lambda}, H) \text{ and } \boldsymbol{\epsilon}_i \overset{ind.}{\sim} \mathrm{NI}_{n_i}(\mathbf{0}, \boldsymbol{\Sigma}_i, H), \quad i = 1, \ldots, n. \tag{3.3}$$

Thus this model considers that the $\boldsymbol{\epsilon}_i$'s, related to within-subject errors are symmetrically distributed, while the distribution of random effects is assumed to be asymmetric. That is, the skewness parameter $\boldsymbol{\lambda}$ incorporates asymmetry in the distribution of the random effects only, and in the vector of observed responses $\mathbf{Y}_i$, $i = 1, \ldots, n$, which will be shown to have, marginally, a multivariate SNI distribution.

(ii) Since for each $i = 1, \ldots, n$, $\mathbf{b}_i$ and $\boldsymbol{\epsilon}_i$ are indexed by the same scale mixing factor U_i, they are not independent in general. Independence corresponds to the case $U_i = 1$ $(i = 1, \ldots, n)$, so that the SNI-LMM reduces to the SN-LMM as defined in Arellano-Valle et al. (2005a) and Lin and Lee (2008). However, conditional on U_i, $\mathbf{b}_i$ and $\boldsymbol{\epsilon}_i$ are independent for each $i = 1, \ldots, n$, which implies that $\mathbf{b}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated, since $Cov(\mathbf{b}_i, \boldsymbol{\epsilon}_i) = \mathrm{E}\{\mathbf{b}_i\boldsymbol{\epsilon}_i^\top\} = \mathrm{E}\{\mathrm{E}\{\mathbf{b}_i\boldsymbol{\epsilon}_i^\top | U_i\}\} = \mathbf{0}$. Thus, an attractive and convenient way to specify (3.2) is the following:

$$\mathbf{b}_i | U_i = u_i \overset{ind.}{\sim} \mathrm{SN}_q(\mathbf{0}, u_i^{-1}\mathbf{D}, \boldsymbol{\lambda}),\ \boldsymbol{\epsilon}_i | U_i = u_i \overset{ind.}{\sim} N_{n_i}(\mathbf{0}, u_i^{-1}\boldsymbol{\Sigma}_i), \quad i = 1, \ldots, n,$$

and they are independent, where $U_i \stackrel{iid}{\sim} H$, $i = 1, \ldots, n$.

(iii) When $\boldsymbol{\lambda} = 0$ and the $U_i \stackrel{iid}{\sim} Gamma(\nu/2, \nu/2)$, the SNI-LMM reduces to the (hierarchical) Student-t LMM proposed by Pinheiro et al. (2001), and when $\boldsymbol{\lambda} = 0$ the SNI-LMM reduces to the NI-LMM defined by Osorio (2006). Moreover, when the $\mathbf{b}_i$'s are normally distributed, the SNI-LMM reduces to the robust LMM defined by the Rosa et al. (2003).

Classical inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\alpha}^\top, \boldsymbol{\lambda}^\top, \boldsymbol{\nu}^\top)^\top$ is based on the marginal distribution for $\mathbf{Y}_i$ (Verbeke and Molenberghs (2000)), which we present below. The proof follows directly from Corollary 2 in Arellano-Valle et al. (2005a), replacing the scale matrix $\boldsymbol{\Psi}_i$ by $u_i^{-1}\boldsymbol{\Psi}_i$ there.

Proposition 2. *Under the SNI-LMM at* (3.1)$-$(3.2), *the marginal distribution of* $\mathbf{Y}_i$ *is*

$$f(\mathbf{y}_i;\boldsymbol{\theta}) = 2\int_0^\infty \phi_{n_i}(\mathbf{y}_i;\mathbf{X}_i\boldsymbol{\beta}, u_i^{-1}\boldsymbol{\Psi}_i)\Phi\Big(u_i^{1/2}\bar{\boldsymbol{\lambda}}_i^\top\boldsymbol{\Psi}_i^{-1/2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\Big)dH(u_i;\boldsymbol{\nu}). \tag{3.4}$$

Thus $\mathbf{Y}_i \stackrel{\text{ind.}}{\sim} SNI_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Psi}_i, \bar{\boldsymbol{\lambda}}_i, H)$, $i = 1, \ldots, n$, *where* $\boldsymbol{\Psi}_i = \boldsymbol{\Sigma}_i + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top$,

$$\boldsymbol{\Lambda}_i = (\mathbf{D}^{-1} + \mathbf{Z}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{Z}_i)^{-1} \ \ and \ \ \bar{\boldsymbol{\lambda}}_i = \frac{\boldsymbol{\Psi}_i^{-1/2}\mathbf{Z}_i\mathbf{D}\boldsymbol{\zeta}}{\sqrt{1+\boldsymbol{\zeta}^\top\boldsymbol{\Lambda}_i\boldsymbol{\zeta}}}, \ \ with \ \ \boldsymbol{\zeta} = \mathbf{D}^{-1/2}\boldsymbol{\lambda}.$$

The result presented in Proposition 2 facilitates implementation of inferences with standard optimization routines and existing statistical software such as the *optim* routine in platform R. In this paper we use the EM algorithm (Dempster, Laird and Rubin (1977)) for parameter estimation via two simple modifications, including the ECM algorithm (Meng and Rubin (1993)) and the ECME algorithm (Liu and Rubin (1994), and Meng and Van Dyk (1997)).

3.1. Maximum likelihood estimation

In this section, we demonstrate how to use the EM-type algorithm for ML estimation of the SNI-LMM. A key feature of this model is that it can be formulated in a flexible hierarchical representation that is useful for theoretical derivations. From (2.2) and the marginal stochastic representation of a SN random vector (see Lachos et al. (2007)), it follows that

$$\mathbf{Y}_i|\mathbf{b}_i, U_i = u_i \stackrel{\text{ind.}}{\sim} N_{n_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, u_i^{-1}\boldsymbol{\Sigma}_i); \ \mathbf{b}_i|T_i = t_i, U_i = u_i \stackrel{\text{ind.}}{\sim} N_q(\boldsymbol{\Delta}t_i, u_i^{-1}\boldsymbol{\Gamma});$$
$$T_i|U_i = u_i \stackrel{\text{ind.}}{\sim} HN_1(0, u_i^{-1}); \ \ U_i \stackrel{\text{i.i.d.}}{\sim} H(u_i;\boldsymbol{\nu}), \tag{3.5}$$

$i = 1, \ldots, n$, where $HN_1(0, \sigma^2)$ is the half-$N_1(0, \sigma^2)$ distribution, $\boldsymbol{\Delta} = \mathbf{D}^{1/2}\boldsymbol{\delta}$, and $\boldsymbol{\Gamma} = \mathbf{D} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top$, with $\boldsymbol{\delta} = \boldsymbol{\lambda}/(1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}$ and $\mathbf{D}^{1/2}$ being the square root of $\mathbf{D}$ containing $q(q+1)/2$ distinct elements. Let $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{b}^\top, \mathbf{t}^\top)^\top$, with $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$, $\mathbf{u} = (u_1, \ldots, u_n)^\top$, $\mathbf{t} = (t_1, \ldots, t_n)^\top$, and let $\boldsymbol{\theta}^{(k)} = (\boldsymbol{\beta}^{(k)\top}, \boldsymbol{\gamma}^{(k)\top}, \boldsymbol{\alpha}^{(k)\top}, \boldsymbol{\lambda}^{(k)\top}, \widehat{\boldsymbol{\nu}}^{(k)})^\top$, denote the estimate of $\boldsymbol{\theta}$ at the kth iteration. It follows from (3.5) that the complete-data log-likelihood function is of the form

$$\ell_c(\boldsymbol{\theta}; \mathbf{y}_c) = \sum_{i=1}^{n} \Big[-\frac{1}{2}\log|\boldsymbol{\Sigma}_i| - \frac{u_i}{2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)$$
$$-\frac{1}{2}\log|\boldsymbol{\Gamma}| - \frac{u_i}{2}(\mathbf{b}_i - \boldsymbol{\Delta}t_i)^\top \boldsymbol{\Gamma}^{-1}(\mathbf{b}_i - \boldsymbol{\Delta}t_i) + \log h(u_i; \boldsymbol{\nu}) \Big] + C,$$

where C is a constant that is independent of the parameter vector $\boldsymbol{\theta}$. Given the current estimate $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, the E-step calculates

$$\begin{aligned} Q(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}^{(k)}) &= \mathrm{E}\{\ell_c(\boldsymbol{\theta}; \mathbf{y}_c); \widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}\} \\ &= \sum_{i=1}^{n} Q_{1i}(\boldsymbol{\theta}_1; \widehat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^{n} Q_{2i}(\boldsymbol{\theta}_2; \widehat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^{n} Q_{3i}(\boldsymbol{\nu}; \widehat{\boldsymbol{\theta}}^{(k)}), \end{aligned}$$

where $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$, $\boldsymbol{\theta}_2 = (\boldsymbol{\alpha}^\top, \boldsymbol{\lambda}^\top)^\top$, and $Q_{3i}(\boldsymbol{\nu}; \widehat{\boldsymbol{\theta}}^{(k)}) = E\{\log h(U_i; \boldsymbol{\nu}); \widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}\}$,

$$\begin{aligned} Q_{1i}(\boldsymbol{\theta}_1; \widehat{\boldsymbol{\theta}}^{(k)}) &= -\frac{1}{2}\log|\widehat{\boldsymbol{\Sigma}}_i^{(k)}| - \frac{1}{2}\widehat{u}_i^{(k)}(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)})^\top[\widehat{\boldsymbol{\Sigma}}_i^{(k)}]^{-1}(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}) \\ &\quad + (\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)})^\top[\widehat{\boldsymbol{\Sigma}}_i^{(k)}]^{-1}\mathbf{Z}_i\widehat{(u\mathbf{b})}_i^{(k)} \\ &\quad - \frac{1}{2}\mathrm{tr}\Big\{[\widehat{\boldsymbol{\Sigma}}_i^{(k)}]^{-1}\mathbf{Z}_i\widehat{(u\mathbf{b}\mathbf{b}^\top)}_i^{(k)}\mathbf{Z}_i^\top\Big\}, \\ Q_{2i}(\boldsymbol{\theta}_2; \widehat{\boldsymbol{\theta}}^{(k)}) &= -\frac{1}{2}\log|\widehat{\boldsymbol{\Gamma}}^{(k)}| - \frac{1}{2}\mathrm{tr}\Big\{[\widehat{\boldsymbol{\Gamma}}^{(k)}]^{-1}\Big(\widehat{(u\mathbf{b}\mathbf{b}^\top)}_i^{(k)} - \widehat{(ut\mathbf{b})}_i^{\top(k)}\widehat{\boldsymbol{\Delta}}^{(k)} \\ &\quad - \widehat{\boldsymbol{\Delta}}^{\top(k)}\widehat{(ut\mathbf{b})}_i^{(k)} + \widehat{(ut_2)}_i^{(k)}\widehat{\boldsymbol{\Delta}}^{(k)}\widehat{\boldsymbol{\Delta}}^{\top(k)}\Big)\Big\}, \end{aligned}$$

where $\mathrm{tr}\{\mathbf{Z}\}$ indicates the trace of matrix $\mathbf{Z}$. The calculation of these functions require expressions for $\widehat{u}_i^{(k)} = \mathrm{E}\{U_i|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i\}$, $\widehat{(u\mathbf{b})}_i^{(k)} = \mathrm{E}\{U_i\mathbf{b}_i|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i\}$, $\widehat{(u\mathbf{b}\mathbf{b}^\top)}_i^{(k)} = \mathrm{E}\{U_i\mathbf{b}_i\mathbf{b}_i^\top|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i\}$, $\widehat{(ut)}_i^{(k)} = \mathrm{E}\{U_iT_i|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i\}$, $\widehat{(ut_2)}_i^{(k)} = \mathrm{E}\{UT_i^2 |\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i\}$, and $\widehat{(ut\mathbf{b})}_i^{(k)} = \mathrm{E}\{U_iT_i\mathbf{b}_i|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i\}$. These can be readily evaluated as (see Appendix B)

$$\widehat{(ut)}_i^{(k)} = \widehat{u}_i^{(k)}\widehat{\mu}_i^{(k)} + \widehat{M}_i^{(k)}\widehat{\tau}_{1i}^{(k)}, \quad \widehat{(ut_2)}_i^{(k)} = \widehat{u}_i^{(k)}[\widehat{\mu}_i^{(k)}]^2 + [\widehat{M}_i^{(k)}]^2 + \widehat{M}_i^{(k)}\widehat{\mu}_i^{(k)}\widehat{\tau}_{1i}^{(k)}, \quad (3.6)$$

$$\widehat{(u\mathbf{b})}_i^{(k)} = \widehat{u}_i^{(k)}\widehat{\mathbf{r}}_i^{(k)} + \widehat{\mathbf{s}}_i^{(k)}\widehat{(ut)}_i^{(k)}, \quad \widehat{(ut\mathbf{b})}_i^{(k)} = \widehat{\mathbf{r}}_i^{(k)}\widehat{(ut)}_i^{(k)} + \widehat{\mathbf{s}}_i^{(k)}\widehat{(ut_2)}_i^{(k)},$$
$$\widehat{(u\mathbf{b}\mathbf{b}^\top)}_i^{(k)} = \widehat{\mathbf{B}}_i^{(k)} + \widehat{u}_i^{(k)}\widehat{\mathbf{r}}_i^{(k)}\widehat{\mathbf{r}}_i^{\top(k)} + \widehat{\mathbf{r}}_i^{(k)}\widehat{\mathbf{s}}_i^{\top(k)}\widehat{(ut)}_i^{(k)} + \widehat{\mathbf{s}}_i^{(k)}\widehat{\mathbf{r}}_i^{(k)\top}\widehat{(ut)}_i + \widehat{\mathbf{s}}_i^{(k)}\widehat{\mathbf{s}}_i^{\top(k)}\widehat{(ut_2)}_i^{(k)},$$

where, omitting the supraindex (k), $\widehat{M}_i = [1 + \widehat{\boldsymbol{\Delta}}^\top \mathbf{Z}_i^\top \widehat{\boldsymbol{\Omega}}_i^{-1}\mathbf{Z}_i\widehat{\boldsymbol{\Delta}}]^{-1/2}$, $\widehat{\mu}_i = [\widehat{M}_i]^2\widehat{\boldsymbol{\Delta}}^\top\mathbf{Z}_i^\top\widehat{\boldsymbol{\Omega}}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}})$, $\widehat{\mathbf{B}}_i = [\widehat{\boldsymbol{\Gamma}}^{-1} + \mathbf{Z}_i^\top\widehat{\boldsymbol{\Sigma}}_i^{-1}\mathbf{Z}_i]^{-1}$, $\widehat{\mathbf{r}}_i = \widehat{\mathbf{B}}_i\mathbf{Z}_i^\top\widehat{\boldsymbol{\Sigma}}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}})$, $\widehat{\mathbf{s}}_i = (\mathbf{I}_q - \widehat{\mathbf{B}}_i\mathbf{Z}_i^\top\widehat{\boldsymbol{\Sigma}}_i^{-1}\mathbf{Z}_i)\widehat{\boldsymbol{\Delta}}$, $\widehat{\boldsymbol{\Omega}}_i = \widehat{\boldsymbol{\Sigma}}_i + \mathbf{Z}_i\widehat{\boldsymbol{\Gamma}}\mathbf{Z}_i^\top$, and, as were defined in Section 2, $\hat{u}_i^{(k)} = \hat{u}_{1i}^{(k)} = \mathrm{E}\{U_i|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i\}$, and $\widehat{\tau}_{1i} = \mathrm{E}\{U_i^{1/2}W_\Phi[(U_i^{1/2}\widehat{\mu}_i^{(k)})/(\widehat{M}_i^{(k)})]|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i\}$, with $W_\Phi(x) = \phi_1(x)/\Phi(x),\ x \in \mathbb{R}$.

Since $\mathbf{Y}_i \overset{\text{ind.}}{\sim} \mathrm{SNI}_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Psi}_i, \bar{\boldsymbol{\lambda}}_i, H)$ and $\mathrm{A}_i = (\mu_{T_i})/(M_{T_i}) = \bar{\boldsymbol{\lambda}}_i^\top\mathbf{y}_{0i}$, with $\mathbf{y}_{0i} = \boldsymbol{\Psi}_i^{-1/2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$, $i = 1, \ldots, n$, in each step, the conditional expectations $\widehat{u}_i$ and $\widehat{\tau}_{1i}$ can be easily derived from the result given in Section 2. For the skew-t and skew-contaminated normal distributions of the SNI class we have computationally attractive expressions that can be easily implemented. However, for the skew-slash case, Monte Carlo integration may be employed, which yields the so-called MC-EM algorithm. Once at the kth iteration, the conditional moments $\widehat{u}_i^{(k)}$ and $\widehat{\tau}_{1i}^{(k)}$ need to be approximated by Monte Carlo integration (see Corollary 2).

The CM steps then conditionally maximize $Q(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$, obtaining a new estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$, as follows.

CM-step 1: Fix $\widehat{\boldsymbol{\gamma}}^{(k)}$ and update $\widehat{\boldsymbol{\beta}}^{(k)}$ as

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = (\sum_{i=1}^n \widehat{u}_i^{(k)}\mathbf{X}_i^\top\widehat{\boldsymbol{\Sigma}}_i^{(k)-1}\mathbf{X}_i)^{-1}\sum_{i=1}^n \mathbf{X}_i^\top\widehat{\boldsymbol{\Sigma}}_i^{(k)-1}(\widehat{u}_i^{(k)}\mathbf{y}_i - \mathbf{Z}_i\widehat{(u\mathbf{b})}_i^{(k)}). \quad (3.7)$$

CM-step 2: Fix $\widehat{\boldsymbol{\beta}}^{(k+1)}$ and update $\widehat{\boldsymbol{\gamma}}^{(k)}$ as $\widehat{\boldsymbol{\gamma}}^{(k+1)} = \mathrm{argmax}_{\boldsymbol{\gamma}}\{Q_{1i}(\widehat{\boldsymbol{\beta}}^{(k+1)}, \boldsymbol{\gamma}; \widehat{\boldsymbol{\theta}}^{(k)})\}$.

CM-step 3: Update $\widehat{\boldsymbol{\Delta}}^{(k)}$ as $\widehat{\boldsymbol{\Delta}}^{(k+1)} = (\sum_{i=1}^n \widehat{(ut\mathbf{b})}_i^{(k)})/(\sum_{i=1}^n \widehat{(ut_2)}_i^{(k)})$.

CM-step 4: Fix $\widehat{\boldsymbol{\Delta}}^{(k+1)}$ and update $\widehat{\boldsymbol{\Gamma}}^{(k)}$ as

$$\widehat{\boldsymbol{\Gamma}}^{(k+1)} = \frac{1}{n}\sum_{i=1}^n \Big(\widehat{(u\mathbf{b}\mathbf{b}^\top)}_i^{(k)} - \widehat{(ut\mathbf{b})}_i^{(k)}[\widehat{\boldsymbol{\Delta}}^{(k+1)}]^\top - \widehat{\boldsymbol{\Delta}}^{(k+1)}[\widehat{(ut\mathbf{b})}_i^{(k)}]^\top + \widehat{(ut_2)}_i^{(k)}\widehat{\boldsymbol{\Delta}}^{(k+1)}[\boldsymbol{\Delta}^{(k+1)}]^\top\Big).$$

CM-step 5: Update $\widehat{\boldsymbol{\nu}}^{(k)}$ by optimizing the constrained actual marginal log-likelihood function $\widehat{\boldsymbol{\nu}}^{(k+1)} = \text{argmax}_{\boldsymbol{\nu}}\{f(\mathbf{y};\widehat{\boldsymbol{\theta}}_1^{(k+1)},\widehat{\boldsymbol{\theta}}_2^{(k+1)},\boldsymbol{\nu})\}$, where $f(\mathbf{y};\boldsymbol{\theta})$ is as in Proposition 2.

The more efficient CM-step 5 follows Liu and Rubin (1994) (ECME, see also Meng and Van Dyk (1997)). It is referred to as conditional marginal likelihood step (CML-step), where we replace the usual M-step by a step that maximizes the restricted actual log-likelihood function. Further, this step along with the CM-step 2 can be easily accomplished by using, for instance, the "optim" routine in R software. Another strategy for speeding up the convergence rate is to use the PX-EM algorithm of Liu, Rubin and Wu (1998); however, its application is not straightforward for SNI-LMM, and requires further exploration. The skewness parameter vector, and the parameters of the scale matrix of the random effects $\mathbf{b}$, can be estimated by noting that $\widehat{\mathbf{D}}^{(k)} = \widehat{\boldsymbol{\Gamma}}^{(k)} + \widehat{\boldsymbol{\Delta}}^{(k)}[\boldsymbol{\Delta}^{(k)}]^\top$ and $\widehat{\boldsymbol{\lambda}}^{(k)} = [\widehat{\mathbf{D}}^{(k)}]^{-1/2}\widehat{\boldsymbol{\Delta}}^{(k)}/(1 - [\boldsymbol{\Delta}^{(k)}]^\top[\widehat{\mathbf{D}}^{(k)}]^{-1}\widehat{\boldsymbol{\Delta}}^{(k)})^{1/2}$. For the special (and common) situation in which $\boldsymbol{\Sigma}_i = \sigma_e^2\mathbf{R}_i$, where $\mathbf{R}_i$ is a known matrix of dimension $(n_i \times n_i)$ and $\boldsymbol{\gamma} = \sigma_e^2$, CM-step 2 reduces to the closed form

$$\begin{aligned}\widehat{\sigma_e}^{2(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\Big[&\widehat{u}_i^{(k)}(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k+1)})^\top\mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k+1)})\\ &-(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k+1)})^\top\mathbf{R}_i^{-1}\mathbf{Z}_i\widehat{(u\mathbf{b})}_i^{(k)} - [\widehat{(u\mathbf{b})}_i^{(k)}]^\top\mathbf{Z}_i^\top\mathbf{R}_i^{-1}\\ &\times(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k+1)}) + \mathbf{tr}(\mathbf{R}_i^{-1}\mathbf{Z}_i\widehat{(u\mathbf{b}\mathbf{b}^\top)}_i^{(k)}\mathbf{Z}_i^\top)\Big].\end{aligned}$$

In Appendix B we give the EM algorithm for restricted estimation of the parameters, that can be used to construct, for instance, the likelihood ratio statistics. A common problem, with any iterative optimization procedures, is that one needs appropriate initial values to avoid divergence or time-consuming computations. A simple way of selecting useful starting values is to use those obtained under skew-normal assumption (Lachos et al. (2007) and Lin and Lee (2008)) and repeat the iterations until the difference between two successive log-likelihood evaluations (3.4) is small enough to achieve convergence. Information criteria such as AIC and BIC (Lachos et al. (2007)), can be used in practice to select between various SNI-LMM distributions.

Assuming the regularity conditions in Zacks (1971, Chap. 5) asymptotic covariance of the ML estimates can be estimated by the inverse of the observed information matrix, $\mathbf{L}(\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n}\widehat{\mathbf{s}}_i\widehat{\mathbf{s}}_i^\top$, where $\widehat{\mathbf{s}}_i = [(\partial\log f(\mathbf{y}_i;\boldsymbol{\theta}))/(\partial\boldsymbol{\theta})]|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ is the score vector corresponding corresponding to the observation $\mathbf{y}_i$ evaluated at

$\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$. Expressions for the elements of the score vector with respect to $\boldsymbol{\theta}$, and for each element of the SNI class, are given in Appendix C.

3.2. Estimation of random effects and prediction

In this section, we consider an empirical Bayes inference for the random effects that is useful for evaluating subject-specific quantities such as individual intercepts and slopes. From (3.5), the conditional distribution of the $\mathbf{b}_i$ given $(\mathbf{Y}_i, U_i) = (\mathbf{y}_i, u_i)$ belong to the extended skew-normal (EST) family of distributions (Azzalini and Capitanio (1999)), and its pdf is

$$f(\mathbf{b}_i|\mathbf{y}_i, u_i, \boldsymbol{\theta}) = \frac{1}{\Phi(u_i^{1/2}\mathrm{A}_i)}\ \phi_q(\mathbf{b}_i; \boldsymbol{\mu}_{bi}, u_i^{-1}\boldsymbol{\Lambda}_i)\Phi(u_i^{1/2}\boldsymbol{\zeta}^\top\mathbf{b}_i),$$

where $\boldsymbol{\mu}_{bi} = \mathbf{D}\mathbf{Z}_i^\top\boldsymbol{\Psi}_i^{-1/2}\mathbf{y}_{0i}$ and $\mathrm{A}_i = \bar{\boldsymbol{\lambda}}_i^\top\mathbf{y}_{0i}$, with $\mathbf{y}_{0i} = \boldsymbol{\Psi}_i^{-1/2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$, and $\boldsymbol{\Lambda}_i$, $\boldsymbol{\zeta}$ and $\bar{\boldsymbol{\lambda}}_i$ as in Proposition 2. Thus, from Lemma 2 in Appendix A, it follows that

$$\mathrm{E}\left\{\mathbf{b}_i|\mathbf{Y}_i = \mathbf{y}_i, U_i = u_i, \boldsymbol{\theta}\right\} = \boldsymbol{\mu}_{bi} + \frac{u_i^{-1/2}W_\Phi(u_i^{1/2}\mathrm{A}_i)}{\sqrt{1+\boldsymbol{\zeta}^\top\boldsymbol{\Lambda}_i\boldsymbol{\zeta}}}\boldsymbol{\Lambda}_i\boldsymbol{\zeta}.$$

The minimum mean-squared error (MSE) estimator of $\mathbf{b}_i$ obtained by the conditional mean of $\mathbf{b}_i$ given $\mathbf{Y}_i = \mathbf{y}_i$ is

$$\begin{aligned}\widehat{\mathbf{b}}_i(\boldsymbol{\theta}) &= \mathrm{E}\left\{\mathbf{b}_i|\mathbf{Y}_i = \mathbf{y}_i, \boldsymbol{\theta}\right\} = \mathrm{E}\left\{\mathrm{E}\left\{\mathbf{b}_i|U_i, \mathbf{Y}_i = \mathbf{y}_i, \boldsymbol{\theta}\right\}|\mathbf{Y}_i = \mathbf{y}_i, \boldsymbol{\theta}\right\},\\ &= \boldsymbol{\mu}_{bi} + \frac{\tau_{-1i}}{\sqrt{1+\boldsymbol{\zeta}^\top\boldsymbol{\Lambda}_i\boldsymbol{\zeta}}}\boldsymbol{\Lambda}_i\boldsymbol{\zeta}, \end{aligned} \tag{3.8}$$

where, $\tau_{-1i} = \mathrm{E}\{U_i^{-1/2}W_\Phi(U_i^{1/2}\mathrm{A}_i)|\mathbf{y}_i\}$. In practice, the empirical Bayes estimators of $\mathbf{b}_i$, $\widehat{\mathbf{b}}_i$, can be obtained by substituting the ML estimate $\widehat{\boldsymbol{\theta}}$ into (3.8). Furthermore, we are interested in the prediction of ${\mathbf{y}^+}_i$, a future $\upsilon \times 1$ vector of measurement of $\mathbf{Y}_i$, given the observed measurement $\mathbf{Y} = (\mathbf{Y}_{(i)}^\top, \mathbf{Y}_i^\top)^\top$, where $\mathbf{Y}_{(i)} = (\mathbf{Y}_1^\top, \ldots, \mathbf{Y}_{i-1}^\top, \mathbf{Y}_{i+1}^\top, \ldots, \mathbf{Y}_n^\top)$. If $\mathbf{x}_i^+$ and $\mathbf{z}_i^+$ denote $\upsilon \times p$ and $\upsilon \times q$ matrices of prediction regressors corresponding to $\mathbf{y}_i^+$, we assume that

$$\begin{bmatrix}\mathbf{Y}_i\\ {\mathbf{y}^+}_i\end{bmatrix} \sim \mathrm{SNI}_{n_i+\nu}(\mathbf{X}_i^*\boldsymbol{\beta}, \boldsymbol{\Psi}_i^*, \bar{\boldsymbol{\lambda}}_i^*; H),$$

where $\mathbf{X}_i^* = (\mathbf{X}_i^\top, \mathbf{x}_i^{+\top})^\top$, $\mathbf{Z}_i^* = (\mathbf{Z}_i^\top, \mathbf{z}_i^{+\top})^\top$, $\boldsymbol{\Psi}_i^* = \boldsymbol{\Sigma}_i^* + \mathbf{Z}_i^*\mathbf{D}\mathbf{Z}_i^{*\top}$, $\boldsymbol{\Lambda}_i^* = (\mathbf{D}^{-1} + {\mathbf{Z}^*}_i^\top\boldsymbol{\Sigma}_i^{*-1}{\mathbf{Z}^*}_i)^{-1}$, ${\bar{\boldsymbol{\lambda}}^*}_i = (\boldsymbol{\Psi}_i^{*-1/2}{\mathbf{Z}^*}_i\mathbf{D}\boldsymbol{\zeta})/(\sqrt{1+\boldsymbol{\zeta}^\top\boldsymbol{\Lambda}_i^*\boldsymbol{\zeta}})$. From Lemma 1 in Appendix A, jointly with equation (2.2), it follows that $\mathbf{Y}_i \sim \mathrm{SNI}_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Psi}_i, \boldsymbol{\Psi}_i^{1/2}\widetilde{\boldsymbol{\upsilon}}; H)$, and

$$\mathrm{E}\left\{\mathbf{y}_i^+|\mathbf{Y}_i, u_i, \boldsymbol{\theta}\right\} = \boldsymbol{\mu}_{2.1} + u_i^{-1/2}\frac{\phi(u_i^{1/2}\widetilde{\boldsymbol{\upsilon}}_i^\top(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}))}{\Phi_i(u_i^{1/2}\widetilde{\boldsymbol{\upsilon}}_i^\top(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}))}\frac{\boldsymbol{\Psi}_{22.1}^*\boldsymbol{\upsilon}_i^{(2)}}{\sqrt{1+\boldsymbol{\upsilon}_i^{(2)\top}\boldsymbol{\Psi}_{22.1}^*\boldsymbol{\upsilon}_i^{(2)}}},$$

where $\boldsymbol{\mu}_{2.1} = \mathbf{x}_i^{+}\boldsymbol{\beta} + \boldsymbol{\Psi}_{21}^{*}\boldsymbol{\Psi}_{11}^{*-1}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})$, $\boldsymbol{\Psi}_{22.1}^{*} = \boldsymbol{\Psi}_{22}^{*} - \boldsymbol{\Psi}_{21}^{*}\boldsymbol{\Psi}_{11}^{*-1}\boldsymbol{\Psi}_{12}^{*}$, $\widetilde{\boldsymbol{\upsilon}}_i = (\boldsymbol{\upsilon}_i^{(1)} + \boldsymbol{\Psi}_{11}^{*-1}\boldsymbol{\Psi}_{12}^{*}\boldsymbol{\upsilon}_i^{(2)})/(\sqrt{1 + \boldsymbol{\upsilon}_i^{(2)\top}\boldsymbol{\Psi}_{22.1}^{*}\boldsymbol{\upsilon}_i^{(2)}})$, with $\boldsymbol{\upsilon}_i = \boldsymbol{\Psi}_i^{*-1/2}\bar{\boldsymbol{\lambda}}_i^{*} = (\boldsymbol{\upsilon}_i^{(1)\top}, \boldsymbol{\upsilon}_i^{(2)\top})^{\top}$ and $\boldsymbol{\Psi}_{11}^{*} = \boldsymbol{\Psi}_i$ and $\boldsymbol{\Psi}_{12}^{*} = \boldsymbol{\Psi}_{21}^{*}$. The minimum MSE predictor of $\mathbf{y}_i$ is the conditional expectation of $\mathbf{y}_i$ given $\mathbf{Y}_i$, i.e.,

$$\begin{aligned}\widehat{\mathbf{y}}_i^{+}(\boldsymbol{\theta}) &= \mathrm{E}\{\mathbf{y}_i^{+}|\mathbf{Y}_i, \boldsymbol{\theta}\} = \mathrm{E}\{\mathrm{E}\{\mathbf{y}_i^{+}|U_i, \mathbf{Y}_i\}|\mathbf{Y}_i, \boldsymbol{\theta}\} \\ &= \boldsymbol{\mu}_{2.1} + \frac{\tau_{-1i}\boldsymbol{\Psi}_{22.1}^{*}\boldsymbol{\upsilon}_i^{(2)}}{\sqrt{1 + \boldsymbol{\upsilon}_i^{(2)\top}\boldsymbol{\Psi}_{22.1}^{*}\boldsymbol{\upsilon}_i^{(2)}}}. \end{aligned} \tag{3.9}$$

The prediction of $\mathbf{y}_i^{+}$ can be obtained by substituting the ML estimate $\widehat{\boldsymbol{\theta}}$ into (3.9), $\widehat{\mathbf{y}}_i^{+} = \widehat{\mathbf{y}}_i^{+}(\widehat{\boldsymbol{\theta}})$.

4. Simulation Study

In this section we present a simulation study to evaluate the performance of the proposed method and of the conditional mean prediction of the subject-specific effects proposed in Section 3. In particular, we want to asses the robustness or bias incurred when one assumes a normal or skew-normal distribution for random effects when the actual distribution is ST.

For the simulation, we generated 1,000 Monte Carlo data set from the model

$$Y_{ij} = \beta_0 + t_{ij}\beta_1 + \omega_i\beta_2 + b_i + e_{ij}, \tag{4.1}$$

where, for $j = 1, \ldots, 5$, $t_{ij} = j - 3$, $\beta_1 = 2$, $\beta_2 = 1$, $e_{ij} \sim t_1(0, 0.5^2, 4)$, and additional specifications to be described below. To show the advantage of the skew-t distribution, we further generated the $\beta_0 + b_i$ according to a $ST_1(1, 2, 3, 4)$ distribution, yielding a highly skewed and heavy tailed distribution, as suggested by the solid line in Figure 4.1(a).

Note that t_{ij} represents a covariate with values changing within individuals and the same for all individuals, while ω_i is the individual level-covariate, e.g., a treatment indicator. We took $n = 100$ with $\omega_i = 1$ if $i \leq 50$ and $\omega_i = 0$ if $i > 50$. For each of $1,000$ simulated data sets, model (4.1) was fit three times under the assumption that the density of b_i was (i) the ST, (ii) SN, and (iii) Normal (N) distribution. We used, the Akaike information criterion (AIC) to select the model that better fit the data.

When the data was actually generated from the ST case, four of AIC values selected the SN-LMM specification for the 1,000 data sets, and none of AIC values selected the normal specification. Table 4.1 gives the numerical results when the original data was generated from the ST and estimates obtained under skew-t LMM (ST-LMM), normal LMM (N-LMM), and skew-normal LMM (SN-LMM).

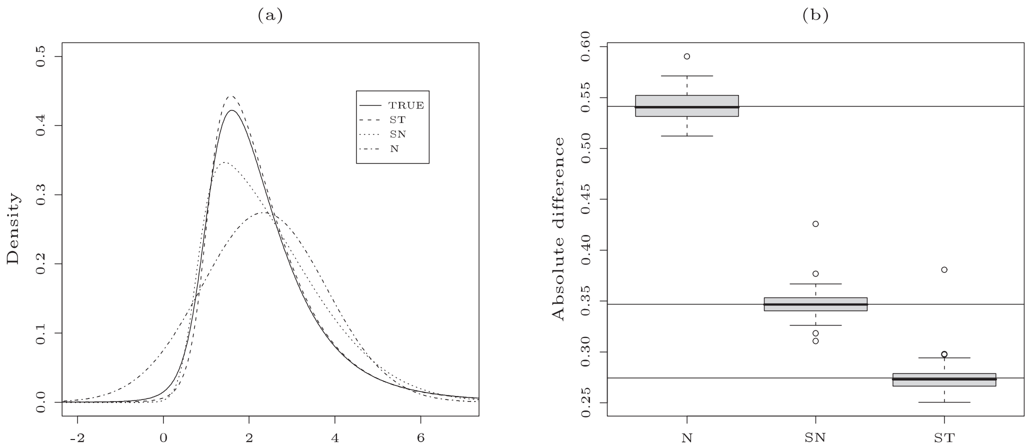

Figure 4.1. Simulation study based on 1,000 data sets of ST-LMM. (a) box-plot of the mean absolute difference of the estimated and simulated random effects for the 100 individuals. (b) True density of the random effects (solid line) and Monte Carlo average estimated densities for 1,000 data set: using N-LMM (dashed-dotted), SN-LMM (dotted) and ST-LMM (dotted-line) fitted. The solid lines are the respective means.

Table 4.1. Monte Carlo results based on 1,000 data sets, true $ST_1(0, 2, 3, 4)$ distribution for the random effects and $t_1(0, 0.25, 4)$ for the random errors. MEAN and SD are average and standard deviation of the estimates, AVE SE is average of estimated standard errors. True values of parameters are in parentheses.

Parameter	MEAN	SD	AVE SE	MEAN	SD	AVE SE	MEAN	SD	AVE SE
	(i) ST-LMM			(ii) SN-LMM			(iii) N-LMM		
β_0 (1)	1.0174	0.2037	0.1871	0.7666	0.2116	0.1982	2.3404	0.2258	0.2168
β_1 (2)	1.9988	0.0179	0.0177	1.9988	0.0228	0.0260	1.9988	0.0228	0.0265
β_2 (1)	0.9953	0.1920	0.1846	0.9807	0.2241	0.2232	1.0237	0.2994	0.3056
σ_e^2 (0.25)	0.2527	0.0307	0.0288	0.4904	0.1031	0.0378	0.4911	0.1035	0.0394
σ_b^2 (2)	2.0410	0.5388	0.5112	4.6433	1.5472	0.7758	2.1170	1.0270	0.3233
λ (3)	3.8644	2.2167	2.2016	6.8108	2.8361	3.3430	-	-	-
ν (4)	4.3234	1.1389	0.9708	-	-	-	-	-	-

In the ST-LMM, the average of estimates of standard errors agreed well with the Monte Carlo standard deviations. We can notice that the slope estimates were similar among the three fitted models, however the standard errors of the models seemed to produce more accurate maximum likelihood estimates. The inferences for the variance components are different for the three fitted models, but the estimates are not comparable since they are in different scales. As found by other authors (Arellano-Valle et al. (2005a)), efficiency of estimation of β_2, associated with the individual-level covariate ω_i, was degraded when normality is assumed relative to allowing a more flexible representation via the SN distribution. Additionally, this estimate was also degraded when the skew-normal distribution was assumed. Since the main focus of such analysis may be the evaluation of treatment effects, this suggests that adopting normality (or skew-normality) assumptions routinely may lead to inefficient inferences on fixed effects of primary interest when the actual distribution is not normal.

To investigate the performance of the empirical Bayes estimates of the subject-specific effects, in Figure 4.1(b), we depicted the conditional mean predictor obtained under N-LMM, SN-LMM and ST-LMM. Accuracy is evaluated by the absolute difference between the simulated and estimated random effects for each individual. The mean values plotted in this figure, clearly indicate that the ST-LMM outperformed the N-LMM and SN-LMM regarding the prediction of random effects. A similar simulation study has also been considered to evaluate the performance of the SSL-LMM and SCN-LMM, with results given in Appendix D.

5. An Illustrative Example

The Framingham heart study examined the role of serum cholesterol as a risk factor for the evolution of cardiovascular disease. Arellano-Valle et al. (2005a) and Lachos et al. (2007), analyzed the same data set by fitting a SN-LMM. In this section, we revisit the Framingham cholesterol data with the aim of providing additional inferences by using SNI distributions. Assuming a linear growth-curve model with subject-specific random intercepts and slopes, we fit a LMM model to the data, as specified by Zhang and Davidian (2001),

$$Y_{ij} = \beta_o + \beta_1 \mathrm{sex}_i + \beta_2 \mathrm{age}_i + \beta_3 t_{ij} + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij}, \tag{5.1}$$

where Y_{ij} is the cholesterol level, divided by 100, at the jth time for subject i, t_{ij} is (time $-$ 5)/10, with time measured in years from the start of the study, and age_i is age at the start of the study; and sex_i is the gender indicator (0 = female, 1 = male). Thus, $\mathbf{x}_{ij} = (1, \mathrm{sex}_i, \mathrm{age}_i, t_{ij})^\top$, $\mathbf{b}_i = (b_{0i}, b_{1i})^\top$, and $\mathbf{Z}_{ij} = (1, t_{ij})^\top$, $i = 1, \ldots, 200$. The histogram of the cholesterol levels (not shown here) clearly indicates an underlying asymmetric distribution, and thus it would seem appropriate to fit a SNI-LMM to the data. To verify the existence of skewness in the random effects, we started by fitting an ordinary N-LMM. Figure 5.2 depicts histograms and corresponding envelopes of the empirical Bayes estimates of $\mathbf{b}_i$, $\widehat{\mathbf{b}}_i = \widehat{\boldsymbol{\mu}}_{b_i}$ and shows that there are no apparent non-normal patterns for subject-specific slopes. However, the subject-specific intercept are positively skewed, and

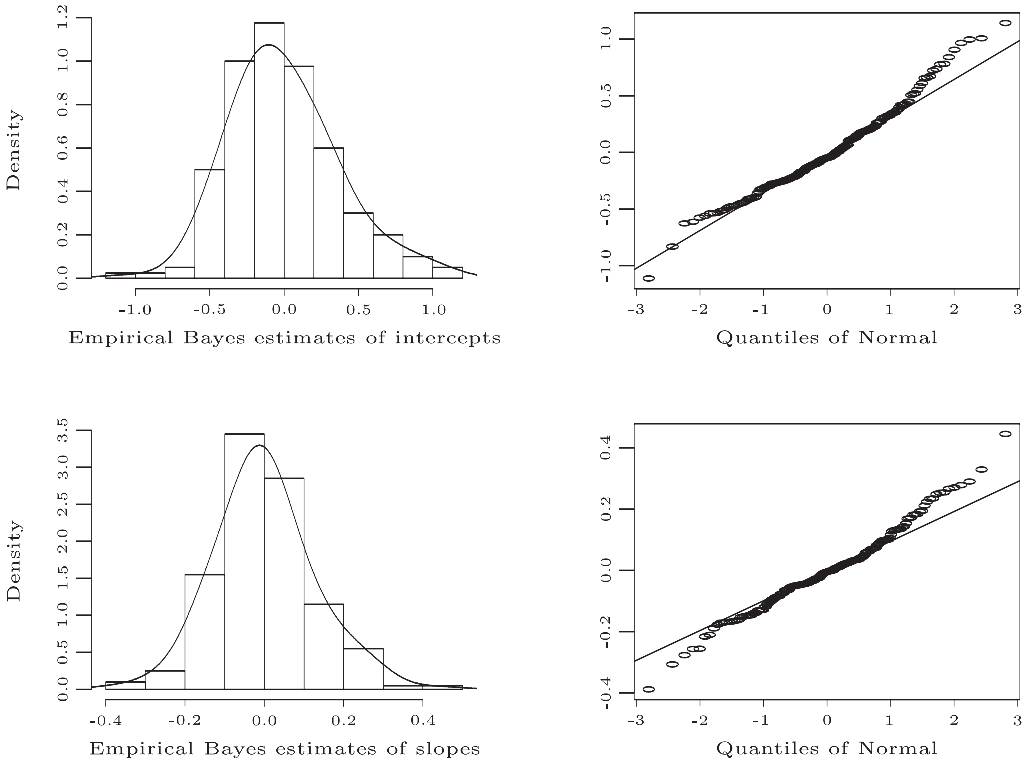

Figure 5.2. Histogram and normal Q-Q plots of empirical bayes estimates of: subject-specific intercepts (first row) and subject-specific slopes (second row).

therefore the suggested Gaussian model did not fit well. Moreover, the QQ - plots clearly support the use of thick-tailed distributions.

Based on the above observations, we now consider a SNI distribution for $\boldsymbol{b}_i$ and NI distribution for $\boldsymbol{e}_i$ with heavy tails. In our analysis we assume SN, ST, SCN and SSL distributions from the SNI class for comparative purposes.

Table 5.2 contains the ML estimates for the parameters of the four models, viz, SN-LMM, ST-LMM, SCN-LMM and SSL-LMM, together with their corresponding standard errors calculated via the approximate observed information matrix given in Appendix C. The AIC criterion indicates that the SNI distributions with heavy tails presents a better fit than the SN-LMM model, due to the departure of the data from normality. We also note from Table 5.2 that the intercept and slope estimates are similar among the four fitted models, however the standard errors of the ST-LMM , SCN-LMM and SSL-LMM are smaller than those in the SN model, indicating that the three models with longer tails than SN produce more accurate maximum likelihood estimates. The estimates for the

Table 5.2. Results from fitting the three models to the Framingham cholesterol data set. (d_{11}, d_{12}, d_{22}) are the distinct elements of the matrix $\mathbf{D}^{1/2}$. The SE values are estimated asymptotic standard errors based on the observed information matrix given in Appendix C. Here AIC denotes the Akaike Information criterion.

	SN-LMM		ST-LMM		SCN-LMM		SSL-LMM	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_o	1.3520	0.1502	1.3888	0.1311	1.4045	0.1396	1.4089	0.1433
β_1	-0.0488	0.0509	-0.0548	0.0447	-0.0461	0.0468	-0.0430	0.0482
β_2	0.0152	0.0033	0.0149	0.0029	0.0144	0.0030	0.0140	0.0031
β_3	0.3562	0.0667	0.3641	0.0611	0.4006	0.0630	0.3998	0.0638
σ_e^2	0.0430	0.0017	0.0325	0.0025	0.0264	0.0028	0.0228	0.0025
d_{11}	0.5261	0.0474	0.4417	0.0477	0.4079	0.0541	0.3918	0.0472
d_{12}	0.0018	0.0302	-0.0030	0.0305	-0.0246	0.0290	-0.0232	0.0277
d_{22}	0.2166	0.0330	0.2035	0.0370	0.2099	0.0386	0.1953	0.0353
λ_1	13.8050	4.2423	13.7822	4.4242	13.4875	4.6855	14.1171	4.7110
λ_2	-6.3654	4.3984	-8.0691	3.9867	-8.7607	4.0621	-8.4215	4.2099
ν	-	-	8.1799	2.1980	0.2981	0.0865	2.0898	0.4669
γ	-	-	-	-	0.3345	0.0425	-	-
$\ell(\widehat{\boldsymbol{\theta}})$	-152.0090		-127.4155		-125.9182		-130.3672	
AIC	0.1552		0.1326		0.1321		0.1354	

variance components are not comparable since they are on a different scale.

To asses the predictive performance of the SN-LMM and SNI-LMM with heavy tails we drop out the last three measurement y_{i4}, y_{i5}, y_{i6} from individual 133, then compute the ML estimates using the remaining data. The prediction of $\mathbf{y}_i = (y_{i4}, y_{i5}, y_{i6})^\top$, denoted by $\widehat{\mathbf{y}}_i = (\widehat{y}_{i4}, \widehat{y}_{i5}, \widehat{y}_{i6})^\top$, is made using formula (3.9). As a measure of precision we use the MARD, the mean of absolute relative deviation $|(y_{ip} - \widehat{y}_{ip})/y_{ip}|$, where p is the time point being forecast. The comparison of the predictors based on the different models is given in Table 5.3. As expected, the result indicated that the SNI distribution yields better predictions than the SN and the Normal (see Lin and Lee (2008)) predictors. Thus, the SNI-LMM with heavy tails not only provide better model fitting, they also yield smaller prediction errors for the cholesterol data.

6. Concluding Remarks

In this paper, we have proposed the application of a new class of asymmetric distributions, called the SNI distribution, to LMMs. This facilitates the fit of a linear mixed model even when the data deviates from the usual normal distribution assumption. A closed-form expression is obtained for the likelihood

Table 5.3. Comparison of forecast accuracy in term of MARD.

points being forecast	N-LMM	SN-LMM	ST-LMM	SCN-LMM	SSL-LMM
y_4	0.0801	0.0791	0.0780	0.0781	0.0780
y_5	0.0941	0.0916	0.0879	0.0879	0.0881
y_6	0.0959	0.0909	0.0877	0.0870	0.0870
Average (%)	9.00	8.72	8.45	8.43	8.44

function of the observed data that can be maximized by using existing statistical software. An EM-type algorithm is developed by exploring the statistical properties of the SNI class. The observed information matrix is derived analytically and allows direct implementation of inference on this class of models. A small simulation study is presented, showing the potential to gain efficiency in parameter estimation when the normality assumption is violated. We believe that the approaches proposed here can also be used to study other asymmetric multivariate models. For the Cholesterol Framingham data, the SNI distributions with heavy tails give a better fit. We found some difficulties in the EM implementation of the slash distribution, since it involves an integral in the marginal likelihood (M-step) and in the computation of the conditional quantities τ_r and u_r (E-step), though numerical integration can be used. Matlab and R programs are available from the first author‘s homepage at website address `http://www.ime.unicamp.br/~hlachos/ListaPub.html`.

7. Supplementary Materials

The web Appendices referenced in the paper are available under the Paper Information link at the Statistica Sinica website `http://www.stat.sinica.edu.tw/statistica`.

Acknowledgement

The authors thank two referees and an associate editor for constructive comments. We also thank PhD. Student C. Zeller from Campinas State University for her help on the preprint version of the paper. The first author acknowledges partial financial support from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). The research of R. B. Arellano-Valle was supported in part by FONDECYT (Chile), 1085241.

References

Arellano-Valle, R. B. and Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scand. J. Statist.* **33**, 561-574.

Arellano-Valle, R. B., Bolfarine, H. and Lachos, V. H. (2005a). Skew-normal linear mixed models. *Journal of Data Science.* **3**, 415-438.

Arellano-Valle, R. B. and Genton, M. G. (2005b). On fundamental skew distributions. *J. Multivariate Anal.* **96**, 93-116.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *J. Roy. Statist. Soc. Ser. B* **61**, 579-602.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on the multivariate skew-t distribution. *J. Roy. Statist. Soc. Ser. B* **65**, 367-389.

Azzalini, A. and Dalla-Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715-726.

Azzalini, A. and Genton, M. (2007). On Gauss's characterization of the normal distribution. *Bernoulli* **13**, 169-174.

Branco, M. and Dey, D. (2001). A general class of multivariate skew-elliptical distribution. *J. Multivariate Anal.* **79**, 93-113.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39**, 1-22.

Ghidey, W., Lesaffre, E. and Eilers, P. (2004). Smooth random effects distribution in a linear mixed model. *Biometrics* **60**, 945-953.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions,* Vol. 1. Wiley, New York.

Lachos, V. H., Bolfarine, H., Arellano-Valle, R. B. and Montenegro, L. C. (2007). Likelihood based inference for multivariate skew-normal regression models. *Comm. Statist. Theory Methods* **36**, 1769-1786.

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

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

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

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

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

Lin, T. I., Lee, J. C. and Hsieh, W. J. (2007). Roubust mixture modeling using the skew t distribution. *Statist. Comput.* **17**, 81-92.

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

Liu, C. H., Rubin, D. B. and Wu, Y. (1998). Parameter expansion to accelerate EM: the PX-EM algorithm. *Biometrika* 85, 755-770.

Ma, Y., Genton, M. G. and Davidian, M. (2004). Linear mixed models with flexible generalized skew-elliptical random effects. In *Skew-Elliptical Distributions and their Applications: A Journey Beyond Normality,* (Edited by M. G. Genton), 339-358. Chapman & Hall/CRC, Boca Raton, FL.

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

Meng, X. L. and Van Dyk, D.(1997). The EM algorithm n old folk song sung to a fast new tune (with discussion). *J. Roy. Statist. Soc. Ser. B* **59**, 511-567.

Osorio, F. S. (2006). Diagnóstico de Influência em Modelos Elípticos com Efeitos Mistos. Ph.D. dissertation, Department of Statistics, IME-USP. São Paulo.

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

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

Verbeke, G. and Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *J. Amer. Statist. Assoc.* **91**, 217-221.

Verbeke, G. and Lesaffre, E. (1997). The effect of misspecifying the random effects distribution in linear mixed models for longitudinal data. *Comput. Statist. Data Anal.* **23**, 541-556.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York.

Wang, J. and Genton, M. (2006). The multivariate skew-slash distribution. *J. Statist. Plann. Inference* **136**, 209-220.

Zacks, S. (1971). *The Theory of Statistical Inference*. Wiley, New York.

Zhang, D. and Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics* **57**, 795-802.

Departamento de Estatística, Universidade Estadual de Campinas, Sao Paulo, Brazil.

E-mail: hlachos@ime.unicamp.br

Department of Quantitative Methods and Information Sciences, Indian Institute of Management, Bannerghatta Road, Bangalore-566076, India.

E-mail: pulakghosh@gmail.com

Departamento de Estadística, Pontificia Universidad Católica de Chile, Santiago, Chile.

E-mail: reivalle@mat.puc.cl

(Received April 2008; accepted December 2008)