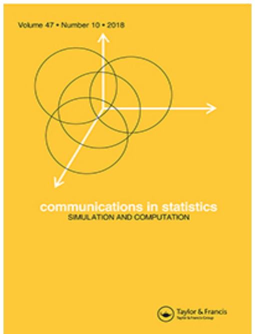

Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Simulation and Computation

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lssp20

Linear mixed models based on skew scale mixtures of normal distributions

Clécio S. Ferreira , Heleno Bolfarine & Víctor H. Lachos

To cite this article: Clécio S. Ferreira , Heleno Bolfarine & Víctor H. Lachos (2020): Linear mixed models based on skew scale mixtures of normal distributions, Communications in Statistics - Simulation and Computation, DOI: 10.1080/03610918.2020.1827265

To link to this article: https://doi.org/10.1080/03610918.2020.1827265

COMMUNICATIONS IN STATISTICS - SIMULATION AND COMPUTATION®
https://doi.org/10.1080/03610918.2020.1827265

Linear mixed models based on skew scale mixtures of normal distributions

Clécio S. Ferreira[a], Heleno Bolfarine[b], and Víctor H. Lachos[c]

[a]Department of Statistics, Federal University of Juiz de Fora, Juiz de Fora, Brazil; [b]Department of Statistics, University of São Paulo, São Paulo, Brazil; [c]Department of Statistics, University of Connecticut, Storrs, Connecticut, USA

ABSTRACT

Scale mixtures of normal distributions are useful for statistical procedures involving symmetric and heavy-tailed data. Ferreira, Lachos, and Bolfarine (2016) defined a multivariate skewed version of these distributions that offers much-needed flexibility by combining both skewness and heavy tails. In this work, we develop a linear mixed model based on skew scale mixtures of normal distributions, with emphasis on the skew Student-*t* normal, skew–slash and skew–contaminated normal distributions. Using the hierarchical structure of the model, we develop maximum likelihood estimation of the model parameters via the expectation-maximization (EM) algorithm. In addition, the standard errors are obtained via the approximate information matrix and the local influence analysis is explored under some perturbation schemes. To examine the performance and the usefulness of the proposed method, we present simulation studies and analyze a real dataset.

ARTICLE HISTORY
Received 24 April 2020
Accepted 18 September 2020

KEYWORDS
EM algorithm; Linear mixed models; Local influence analysis; Skew scale mixtures of normal distributions

1. Introduction

The scale mixtures of normal distributions (Andrews and Mallows 1974) are a group of thick–tailed distributions that are often used for robust inference of symmetrical data. Moreover, this class includes distributions such as the Student-*t*, slash and contaminated normal, among others. However, this class is inappropriate for dataset that are skewed and present heavy-tails, such as, data on family income (Azzalini, Capello, and Kotz 2003) or on substance concentration (Lachos and Bolfarine 2007). Thus, appropriate distributions to fit these skewed and heavy tailed data are needed.

Azzalini (1985) proposed the univariate skew-normal distribution and it was generalized to the multivariate case by Azzalini and Dalla-Valle (1996) and Azzalini and Capitanio (1999). The multivariate skew-normal densities extend the multivariate normal model by allowing a shape parameter to account for skewness. The probability density function (PDF) of the generic element of a multivariate skew-normal distribution is given by:

$$f(\mathbf{y}) = 2\phi_n(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})), \qquad \mathbf{y} \in \mathbb{R}^n, \tag{1}$$

where $\phi_n(.|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for the PDF of the n-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariate matrix $\boldsymbol{\Sigma}, \Phi_1(.)$ represents the cumulative distribution function (CDF) of the standard normal distribution. When $\boldsymbol{\lambda} = 0$, the skew-normal distribution reduces to the normal distribution ($\mathbf{Y} \sim \mathrm{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$). An n-dimensional random vector $\mathbf{Y}$ with PDF as in (1) will be denoted by $\mathrm{SN}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. Its marginal stochastic representation is given by:

CONTACT Clécio da Silva Ferreira clecio.ferreira@ufjf.edu.br DeDepartamento de Estatística, Universidade Federal de Juiz de Fora, Rua José Lourenço Kelmer, s/n - Campus Universitário, Bairro São Pedro, CEP 36036-900, Juiz de Fora, Minas Gerais, Brazil.

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}(\boldsymbol{\delta}|T_0| + (\mathbf{I}_n - \boldsymbol{\delta}\boldsymbol{\delta}^\top)^{1/2}\mathbf{T}_1), \quad \text{with} \quad \boldsymbol{\delta} = \frac{\boldsymbol{\lambda}}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}}, \tag{2}$$

where $|T_0|$ denotes the absolute value of T_0, $T_0 \sim N_1(0,1)$ and $\mathbf{T}_1 \sim N_n(0,\mathbf{I}_n)$ are independent, "$\stackrel{d}{=}$" means "distributed as" and $\mathbf{I}_n$ denotes the identity matrix of order n. From (2) it follows that the expectation and variance of $\mathbf{Y}$ are given, respectively, by:

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta},$$
$$\text{Var}[\mathbf{Y}] = \boldsymbol{\Sigma} - \frac{2}{\pi}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^{1/2}.$$

Linear mixed models (LMMs) have become important tools for practicing statisticians, commonly utilized to analyze continuous repeated measures, grouped and longitudinal data, among others. They are used in several areas, like agricultural, environmental, biomedical, economic and social science applications (see, e.g. Diggle, Liang, and Zeger 1996). The increasing popularity of these models is due to the flexibility they offer to model the correlation between and within samples, frequently present in longitudinal data (Laird and Ware 1982) and grouped data (Henderson 1984), as well as the capacity to model balanced and unbalanced data.

Repeated measures data are typically generated by observing a number of subjects repeatedly under different experimental conditions. Observations of the same subject are usually made at different times, as in longitudinal studies. Mixed-effects models assume that the intra-subject model relating the response variables to time is the same for all subjects, but the model parameters may vary between subjects. Despite their appealing statistical properties, a standard but possibly restrictive assumption in LMMs is that the random effects and the residual components follow a normal distribution. Hence, considerable interest has been focused on relaxing the normality (symmetry) assumption and jointly estimating the random effects and model parameters.

Some recent works have incorporated asymmetric distributions to the random effects in LMMs. For instance, Ma, Genton, and Davidian (2004) considered a generalized flexible skew-elliptical distribution to represent the density of the random effects and proposed complicated algorithms for maximum likelihood (ML) and Bayesian inference using MCMC methods. Arellano-Valle, Bolfarine, and Lachos (2005) defined a skew-normal linear mixed model which assumes that the random effects follow a skew-normal distribution and presented an expectation-maximization (EM)-type algorithm to perform ML estimation. Lachos, Ghosh, and Arellano-Valle (2010) proposed the skew-normal/independent linear mixed model, based on the multivariate scale mixtures of skew-normal (SMSN) family of distributions, where an EM-type algorithm for ML estimation is also presented. Yu, O'Malley, and Ghosh (2014) introduced a new class of extended multivariate skew- t distributions, which allows different degrees of freedom to accommodate heterogeneity in tail-heaviness across outcomes. More recently, Kahrari, Ferreira, and Arellano-Valle (2019) introduced a flexible class of linear mixed models by assuming that the random effects and model errors follow a multivariate skew-normal-Cauchy distribution.

In the context of asymmetric distributions, Ferreira, Bolfarine, and Lachos (2011) proposed a new family of asymmetric univariate distributions called skew scale mixtures of normal distributions (SSMN), generated by the normal kernel (as the skewing function), using otherwise symmetric distributions of the class of scale mixtures of normal distributions (Andrews and Mallows 1974; Lange and Sinsheimer 1993). Ferreira, Lachos, and Bolfarine (2016) developed a multivariate version of the skew scale mixtures of normal distributions, providing an EM algorithm and the observed information matrix for multivariate responses. Following Lachos, Ghosh, and Arellano-Valle (2010), in this article, we develop a new family of linear mixed models, where the random effect follows a SSMN distribution and the error term follows a multivariate scale mixtures of normal (SMN) distribution.

Investigation of the influence is an important step in data analysis after the parameter estimation. The identification of problems caused by influential aspects may give ideas for improving the model assumptions and/or input data to establish a better model. Local influence by minor perturbations of a statistical model is a useful tool for sensitivity analysis (Cook 1986). For example, Lesaffre and Verbeke (1998) developed local influence for LMMs under the normal distribution. This measure involves the first and second partial derivatives of the log-likelihood function. So, when this function involves intractable integrals, direct application of Cook's (1986) approach may require hard calculation. A second approach for local influence, based on conditional expectation of the complete-data log-likelihood at the E-step of the EM algorithm, was developed by Zhu and Lee (2001). For some applications using their approach, see for example, Zeller et al. (2010), Zeller, Lachos, and Labra (2011), Ferreira and Paula (2017) and Ferreira and Arellano-Valle (2018).

Section 2 presents the multivariate skew scale mixtures of normal distributions and their particular cases. Section 3 proposes the SSMN linear mixed models, the maximum likelihood estimation using the EM algorithm (Dempster, Laird, and Rubin 1977) and a simulation study investigate the functionality of the proposed model. In Sec. 4, we develop local influence analysis by Zhu and Lee (2001)'s approach. Finally, Sec. 5 presents an application of the model to the data collected as part of the famous Framingham heart study.

2. Multivariate skew scale mixtures of normal distributions

Andrews and Mallows (1974) presented necessary and sufficient conditions under which a continuous symmetric random variable Y may be generated as the ratio Z/U where Z and U are independent, Z has a standard normal distribution and U is a positive random variable. This symmetric family, named scale mixtures of normal (SMN) distributions, includes distributions such as the Student-t, slash and contaminated-normal distributions. All these distributions have heavier tails than the normal.

We say that an n-dimensional vector $\mathbf{Y}$ has a SMN distribution (Lange and Sinsheimer 1993) with location parameter $\boldsymbol{\mu} \in \mathbb{R}^n$ and a positive definite scale matrix $\boldsymbol{\Sigma}$ if its density function assumes the form:

$$\begin{aligned} f_0(\mathbf{y}) &= \int_0^{\infty} \phi_n(\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}) dH(u, \boldsymbol{\tau}) \\ &= |2\pi\boldsymbol{\Sigma}|^{-1/2} \int_0^{\infty} u^{n/2} \exp\left[-\frac{1}{2} u(\mathbf{y}-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right] dH(u, \boldsymbol{\tau}), \ \ \mathbf{y} \in \mathbb{R}^n, \end{aligned} \tag{3}$$

where $H(u, \boldsymbol{\tau})$ is the CDF of a one-dimensional positive random variable U indexed by the parameter vector $\boldsymbol{\tau}$. For a random vector with a PDF as in (3), we use the notation $\mathbf{Y} \sim \text{SMN}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$, while when $\boldsymbol{\mu} = 0$ and $\boldsymbol{\Sigma} = \mathbf{I}_n$, we use the notation $\mathbf{Y} \sim \text{SMN}_n(H)$.

Its stochastic representation is given by:

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

where $\mathbf{Z} \sim \text{N}_n(0, \boldsymbol{\Sigma})$ and U is a positive random variable with CDF H independent of $\mathbf{Z}$.

Definition 1. A random n-dimensional vector $\mathbf{Y}$ follows a skew scale mixtures of normal distribution with location parameter $\boldsymbol{\mu} \in \mathbb{R}^n$, a positive definite scale matrix $\boldsymbol{\Sigma}$ and skewness parameter $\boldsymbol{\lambda} \in \mathbb{R}^n$ (Ferreira, Lachos, and Bolfarine 2016), if its density function is given by:

$$f(\mathbf{y}) = 2f_0(\mathbf{y})\Phi_1\left(\boldsymbol{\lambda}^{\top}\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})\right), \ \ \mathbf{y} \in \mathbb{R}^n, \tag{4}$$

where $f_0(\cdot)$ is defined as in (3). For a random vector with a PDF as in (4), we use the notation $\mathbf{Y} \sim SSMN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$.

Remark. Conditioned on U, we have that:

$$f_{\mathbf{Y}|U=u}(\mathbf{y}) = 2\phi_n(\mathbf{y}, \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})\Phi_1\left(\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})\right)$$
$$= 2\phi_n(\mathbf{y}, \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})\Phi_1\left(u^{-1/2}\boldsymbol{\lambda}^\top\left[(u^{-1}\boldsymbol{\Sigma})^{-1/2}\right](\mathbf{y}-\boldsymbol{\mu})\right).$$

Therefore, $\mathbf{Y}|U=u \sim SN_n(\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}, u^{-1/2}\boldsymbol{\lambda})$.

If $n=1$, then, we have the univariate SSMN distribution developed in Ferreira, Bolfarine, and Lachos (2011). In the next section, we describe some distributions of the SSMN class.

2.1. Examples of SSMN distributions

• *The skew Student-t normal (StN) distribution with $\nu > 0$ degrees of freedom, denoted by $StN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; \nu)$.*

The use of the t-distribution as an alternative to the normal distribution has been frequently suggested in the literature. For instance, Little (1988) and Lange, Little, and Taylor (1989) recommended using the Student-t distribution for robust modeling.

Considering $U \sim \text{Gamma}(\nu/2, \nu/2)$, the PDF of $\mathbf{Y}$ takes the form:

$$f(\mathbf{y}) = 2\frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\pi^{n/2}}\nu^{-n/2}|\boldsymbol{\Sigma}|^{-1/2}\left(1+\frac{d}{\nu}\right)^{-\left(\frac{\nu+n}{2}\right)}\Phi_1\left(\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})\right), \quad \mathbf{y} \in \mathbb{R}^n,$$

where $d = (\mathbf{y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})$. In this case, $U|\mathbf{Y}=\mathbf{y} \sim \text{Gamma}((\nu+n)/2, (\nu+d)/2)$.

The skew Student-t normal distribution was first developed by Gómez, Venegas, and Bolfarine (2007). In that paper, the authors showed that the StN distribution can present a much wider asymmetry range than the one presented by the ordinary skew-normal distribution (Azzalini 1985). When $\nu \uparrow \infty$, one gets the skew-normal distribution as the limiting case.

• *The skew-slash (SSL) distribution, with shape parameter $\nu > 0$, denoted by $SSL_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; \nu)$.*

The distribution of U is $Beta(\nu, 1)$, $0 < u < 1$ and $\nu > 0$. Its PDF is given by:

$$f(\mathbf{y}) = 2\nu\int_0^1 u^{\nu-1}\phi_n\left(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}\right)du\Phi_1\left(\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})\right), \quad \mathbf{y} \in \mathbb{R}^n.$$

When $\nu \uparrow \infty$, one gets the skew-normal distribution as the limiting case. For this distribution, we have that $U|\mathbf{Y}=\mathbf{y} \sim TG(\nu+n/2, d/2, 1)$, where $TG(a, b, t)$ is the right truncated gamma distribution, with PDF $f(x|a, b, t) = \frac{b^a}{\gamma(a, bt)}x^{a-1}\exp(-bx)\mathbb{I}_{(0,t)}(x)$, $\gamma(a, b) = \int_0^b u^{a-1}e^{-u}du$ is the incomplete gamma function and $\mathbb{I}_A(\cdot)$ denotes the indicator function.

• *The skew-contaminated normal (SCN) distribution, denoted by $SCN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; \nu, \gamma), 0 \le \nu \le 1,\ 0 < \gamma \le 1$.*

Here, U is a discrete random variable taking one of two states. The probability density function of U, given the parameter vector $\boldsymbol{\tau} = (\nu, \gamma)^\top$, is denoted by $h(u; \boldsymbol{\tau}) = \nu\mathbb{I}_{(u=\gamma)} + (1-\nu)\mathbb{I}_{(u=1)}$. It, thus, follows that:

$$f(\mathbf{y}) = 2\{\nu\phi_n\left(\mathbf{y}|\boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Sigma}\right) + (1-\nu)\phi_n(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\}\Phi_1\left(\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})\right), \quad \mathbf{y} \in \mathbb{R}^n.$$

The conditional distribution $U|\mathbf{Y}=\mathbf{Y}$ is also a degenerated function, given by $f(u|\mathbf{Y}=\mathbf{y}) = \frac{1}{f_0(\mathbf{y})}\{\nu\phi_n(\mathbf{y}, \boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Sigma})\mathbb{I}_{(u=\gamma)} + (1-\nu)\phi_n(\mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\Sigma})\mathbb{I}_{(u=1)}\}$, where $f_0(\mathbf{y}) = \nu\phi_n(\mathbf{y}, \boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Sigma}) + (1-\nu)\phi_n(\mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

The skew-normal distribution is a particular case of the skew-contaminated normal distribution when $\gamma = 1$.

3. The SSMN linear mixed model

The mixed linear model for continuous responses, proposed by Laird and Ware (1982), is expressed by:

$$\mathbf{Y}_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, ..., m, \tag{5}$$

where $\mathbf{Y}_i : n_i \times 1$ is a vector of responses of the m individuals, $\mathbf{x}_i^\top : n_i \times p$ is the model matrix corresponding to the fixed effects, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters describing the population mean, named fixed effects, $\mathbf{Z}_i : n_i \times q$ is the model matrix corresponding to the vector of random effects $\mathbf{b}_i : q \times 1$ and $\boldsymbol{\epsilon}_i : n_i \times 1$ is the vector of the errors. Typically it is assumed that the random effects $\mathbf{b}_i$ and the residual components $\boldsymbol{\epsilon}_i$ are independents with $\mathbf{b}_i \overset{iid}{\sim} N_q(0, \mathbf{B})$ and $\boldsymbol{\epsilon}_i \overset{ind}{\sim} N_{n_i}(0, \boldsymbol{\psi}_i)$, for $i = 1, ..., m$. $\mathbf{B} = \mathbf{B}(\boldsymbol{\alpha})$ and $\boldsymbol{\psi}_i = \psi_i(\boldsymbol{\gamma})$ are dispersion matrices, usually associated with inter- and intra-unit, depending on a number of unknown parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, respectively.

Arellano-Valle, Bolfarine, and Lachos (2005) (see also Lin and Lee 2008) proposed a skew-normal linear mixed model (SN-LMM) based on multivariate skew-normal distributions (Azzalini and Dalla-Valle 1996; Azzalini and Capitanio 1999) defined as:

$$\begin{aligned} &\mathbf{Y}_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, ..., m, \\ &\mathbf{b}_i \overset{iid}{\sim} SN_q(0, \mathbf{B}, \boldsymbol{\lambda}), \quad \boldsymbol{\epsilon}_i \overset{ind}{\sim} N_{n_i}(0, \boldsymbol{\psi}_i), \end{aligned} \tag{6}$$

where $\mathbf{b}_i$ and $\boldsymbol{\epsilon}_i$ are independent and $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_q)^\top$ is the vector of asymmetry of $\mathbf{b}_i$, $i = 1, ..., m$.

The linear mixed model developed in this work extends the SN-LMM defined in Eq. (6), by considering that b_i follows a distribution in the SSMN class, as follows:

$$\begin{aligned} &\mathbf{Y}_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{Z}_i b_i + \boldsymbol{\epsilon}_i, \\ &\begin{pmatrix} b_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \sim \text{SSMN}_{n_i+1}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_b^2 & 0 \\ 0 & \sigma_e^2 \mathbf{I}_{n_i} \end{pmatrix}, \begin{pmatrix} \lambda \\ 0 \end{pmatrix}; H \right), \quad i = 1, ..., m. \end{aligned}$$

Although, we cannot find a marginal distribution for $\mathbf{Y}_i$, we have an attractive hierarchical representation, which enables estimating the model parameters via the EM algorithm. Next, we prove that, although, the terms b_i and $\boldsymbol{\epsilon}_i$ are not independent, they are not correlated.

The joint conditional distribution of $(b_i, \boldsymbol{\epsilon}_i)|U_i$ is given by:

$$\begin{pmatrix} b_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \Bigg| U_i = u_i \sim SN_{n_i+1}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, u_i^{-1} \begin{pmatrix} \sigma_b^2 & 0 \\ 0 & \sigma_e^2 \mathbf{I}_{n_i} \end{pmatrix}, u_i^{-1/2} \begin{pmatrix} \lambda \\ 0 \end{pmatrix} \right), \quad i = 1, ..., m. \tag{7}$$

Using the proposition 2.18 given in Lachos (2004) and considering the conditional distribution in (7), with $h = 2$, $\mathbf{C}_1 = (1 \;\; 0)$ $(1 \times (n_i + 1))$ and $\mathbf{C}_2 = (0 \;\; \mathbf{I}_{n_i})$ $(n_i \times (n_i + 1))$, we have that $\mathbf{C}_1^\top \mathbf{Y}_i = b_i$ and $\mathbf{C}_2^\top \mathbf{Y}_i = \boldsymbol{\epsilon}_i$. So, $b_i|U_i$ and $\boldsymbol{\epsilon}_i|U_i$ are conditionally independent.

Let $\boldsymbol{\delta}_i = \frac{u_i^{-1/2}}{\sqrt{1 + u_i^{-1}\lambda^2}} \begin{pmatrix} \lambda \\ 0 \end{pmatrix}$ and $\mathbf{C} = (1 \;\; 0)$ $(1 \times (n_i + 1))$. Then, using the proposition 2.9 given in Lachos (2004), we have that $\mathbf{C}^\top \mathbf{Y}_i = b_i$, $\mathbf{C}^\top \boldsymbol{\Sigma} \mathbf{C} = \sigma_b^2 / u_i$, $\mathbf{C}^\top \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta}_i = \frac{\lambda \sigma_b}{u_i \sqrt{1 + u_i^{-1} \lambda^2}}$ and $\delta^* = \frac{\lambda}{\sqrt{u_i + \lambda^2}}$, $\lambda^* = \frac{\lambda}{\sqrt{u_i}}$. In this case,

$$b_i|U_i = u_i \sim SN_1\left(0, \frac{\sigma_b^2}{u_i}, \frac{\lambda}{\sqrt{u_i}}\right). \quad (8)$$

Therefore,

$$b_i \sim \text{SSMN}_1\left(0, \sigma_b^2, \lambda; H\right). \quad (9)$$

Let $\mathbf{C} = (0 \ \mathbf{I}_{n_1}) \ (n_i \times (n_i + 1))$. It follows that $\mathbf{C}^\top \mathbf{Y}_i = \boldsymbol{\epsilon}_i$, $\mathbf{C}^\top \boldsymbol{\Sigma} \mathbf{C} = \frac{\sigma_e^2}{u_i} \mathbf{I}_{n_i}$, $\mathbf{C}^\top \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta} = 0$, $\delta^* = 0$, $\lambda^* = 0$. Hence,

$$\boldsymbol{\epsilon}_i|U_i = u_i \sim N_{n_i}\left(\mathbf{0}, \frac{\sigma_e^2}{u_i}\mathbf{I}_{n_i}\right)$$

and so

$$\boldsymbol{\epsilon}_i \sim \text{SMN}_{n_i}\left(0, \sigma_e^2 \mathbf{I}_{n_i}; H\right). \quad (10)$$

As a result, $E[\boldsymbol{\epsilon}_i] = 0$ and $Cov(b_i, \boldsymbol{\epsilon}_i) = E[b_i \boldsymbol{\epsilon}_i] - E[b_i]E[\boldsymbol{\epsilon}_i] = E_{U_i}[E[b_i|U_i]E[\boldsymbol{\epsilon}_i|U_i]] = 0$.

By the results given in Eqs. (9) and (10), the SSMN-LMM can be represented as:

$$\begin{aligned} &\mathbf{Y}_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{Z}_i b_i + \boldsymbol{\epsilon}_i, \\ &b_i \overset{iid}{\sim} \text{SSMN}_1\left(0, \sigma_b^2, \lambda; H\right), \\ &\boldsymbol{\epsilon}_i \overset{ind}{\sim} \text{SMN}_{n_i}\left(0, \sigma_e^2 \mathbf{I}_{n_i}; H\right), \quad i = 1, \ldots, m. \end{aligned} \quad (11)$$

Using the conditional distribution of $b_i|U_i = u_i$ in Eq. (8), the model can be represented hierarchically as follows:

$$\begin{aligned} \mathbf{Y}_i|b_i, u_i &\overset{\text{ind}}{\sim} N_{n_i}\left(\mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{Z}_i b_i, u_i^{-1}\sigma_e^2 \mathbf{I}_{n_i}\right), \\ b_i|u_i &\overset{\text{ind}}{\sim} SN_1\left(0, u_i^{-1}\sigma_b^2, u_i^{-1/2}\lambda\right) \text{e} \\ U_i &\overset{\text{iid}}{\sim} H(\boldsymbol{\tau}). \end{aligned} \quad (12)$$

Lee and Nelder (2004) provide a discussion of the differences between conditional and marginal models in models in the presence of random effects, discussing the advantages of the conditional over marginal models and considering the first as fundamental, for which marginal predictions can be made.

If $\boldsymbol{\epsilon}_i \sim \text{SMN}_q(0, \boldsymbol{\Sigma}; H)$, we have that $E[\boldsymbol{\epsilon}_i] = 0$. When $b_i \sim \text{SSMN}_1\left(0, \sigma_b^2, \lambda; H\right)$, then, $E[b_i] = c\sigma_b \lambda E_{U_i}\left[\frac{U_i^{-1}}{\sqrt{1+U_i^{-1}\lambda^2}}\right]$, with $c = \sqrt{\frac{2}{\pi}}$ (see Ferreira, Bolfarine, and Lachos 2011). So,

$$E[\mathbf{Y}_i] = \mathbf{x}_i \boldsymbol{\beta} + c\sigma_b \lambda E_{U_i}\left[\frac{U_i^{-1}}{\sqrt{1 + U_i^{-1}\lambda^2}}\right]\mathbf{Z}_i.$$

The PDF corresponding to (12) is given by:

$$f(\mathbf{y}_i) = 2\int_{-\infty}^{+\infty}\int_0^{+\infty} \phi_{n_i}\left(\mathbf{y}_i|\mathbf{x}_i^\top \boldsymbol{\beta} + b_i \mathbf{Z}_i, u_i^{-1}\sigma_e^2 \mathbf{I}_{n_i}\right)\phi_1\left(b_i|0, u_i^{-1}\sigma_b^2\right)\Phi_1\left(\frac{\lambda b_i}{\sigma_b}\right)h(u_i; \boldsymbol{\tau})du_i db_i. \quad (13)$$

The model in (12) can be rewritten as

$$\begin{aligned}
\mathbf{Y}_i|b_i, u_i, t_i &\overset{\text{ind}}{\sim} N_{n_i}\left(\mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{Z}_i b_i, u_i^{-1}\sigma_e^2 \mathbf{I}_{n_i}\right),\\
b_i|u_i, t_i &\overset{\text{ind}}{\sim} N_1\left(\frac{\sigma_b \lambda}{u_i\sqrt{1+\lambda^2/u_i}} t_i, \frac{\sigma_b^2}{u_i+\lambda^2}\right),\\
U_i &\overset{\text{iid}}{\sim} H(\boldsymbol{\tau}) \text{ and}\\
T_i &\overset{\text{iid}}{\sim} \text{TN}(0,1;(0,+\infty)),\ i=1,\dots,m,
\end{aligned}$$

where $\text{TN}\left(\mu, \sigma^2; (a,b)\right)$ represents the univariate truncated-normal distribution of $N_1\left(\mu, \sigma^2\right)$ lying within the interval (a, b) (Johnson, Kotz, and Balakrishnan 1994). The PDF in (13) can be expressed as:

$$f(\mathbf{y}_i) = 2\int_{-\infty}^{+\infty}\int_0^{+\infty}\int_0^{+\infty} \phi_{n_i}\left(\mathbf{y}_i|\mathbf{x}_i^\top\boldsymbol{\beta} + b_i\mathbf{Z}_i, \sigma_e^2/u_i\mathbf{I}_{n_i}\right)\phi_1\left(b_i|0,\sigma_b^2/u_i\right) \times\phi_1\left(t_i|\lambda b_i, \sigma_b^2\right)h(u_i;\boldsymbol{\tau})dt_i du_i db_i. \tag{14}$$

If $\lambda = 0$ and $U = 1$, the model (11) becomes the classical mixed effects model. If $\lambda \neq 0$, but $U = 1$, it becomes the skew-normal linear mixed model introduced by Arellano-Valle, Bolfarine, and Lachos (2005), with b_i a scalar. If $b_i = 0$, it has the features of skew scale mixtures of normal models discussed by Ferreira, Bolfarine, and Lachos (2011). The mixed model presented here, in addition to the generalizing class of asymmetric normal mixed effects models, enables a process of estimation through the EM algorithm with analytical expression in the M-step. In addition, we propose an approximation to the observed Fisher information matrix as in Lin (2010).

3.1. Maximum likelihood estimation via the EM algorithm

Finding the ML estimate of the parameter vector $\boldsymbol{\theta}$ by direct maximization of the log-likelihood (14) can be a hard task because of the intractable integrals. Thus, we prefer to implement the EM algorithm introduced by Dempster, Laird, and Rubin (1977), which has several appealing features, such as stability of monotone convergence with each iteration increasing the likelihood, and simplicity of implementation. One of the major reasons for the popularity of the EM algorithm is that the M-step involves only complete data ML estimation, which is often computationally simple. We refer to McLachlan and Krishnan (2008) for details and many applications of the EM algorithm.

Let $\mathbf{y} = \left(\mathbf{y}_1^\top, \dots, \mathbf{y}_m^\top\right)^\top$, $\mathbf{b} = (b_1, \dots, b_m)^\top, \mathbf{u} = (u_1, \dots, u_m)^\top$ and $\mathbf{t} = (t_1, \dots, t_m)^\top$. Then, treating $\mathbf{u}$, $\mathbf{b}$ and $\mathbf{t}$ as missing data, it follows that the complete log-likelihood function associated with $\mathbf{y}_c = \left(\mathbf{y}^\top, \mathbf{u}^\top, \mathbf{b}^\top, \mathbf{t}^\top\right)^\top$ is given by

$$\begin{aligned}
\ell_c\left(\boldsymbol{\theta}|\mathbf{y}_c\right) \propto& -\frac{N}{2}\log\sigma_e^2 - \frac{1}{2\sigma_e^2}\sum_{i=1}^m u_i\left(\mathbf{y}_i - \mathbf{x}_i^\top\boldsymbol{\beta} - b_i\mathbf{Z}_i\right)^\top\left(\mathbf{y}_i - \mathbf{x}_i^\top\boldsymbol{\beta} - b_i\mathbf{Z}_i\right)\\
&- m\log\sigma_b^2 - \frac{1}{2\sigma_b^2}\sum_{i=1}^m u_i b_i^2 - \frac{1}{2\sigma_b^2}\sum_{i=1}^m\left(t_i^2 - 2\lambda t_i b_i + \lambda^2 b_i^2\right),
\end{aligned}$$

where $N = \sum_{i=1}^m n_i$.

It follows that the expectancy with respect to $\mathbf{t}$, $\mathbf{u}$ and $\mathbf{b}$, conditional on $\mathbf{y}$ of the complete log-likelihood function (E-step), is given by:

$$
\begin{aligned}
Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{k}\right) &= \mathrm{E}\left[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y},\hat{\boldsymbol{\theta}}^{(k)}\right] \\
&= -\frac{1}{2\sigma_e^2}\sum_{i=1}^{m}\left(\hat{u}_i^{(k)}(\mathbf{y}_i-\mathbf{x}_i^\top\boldsymbol{\beta})^\top(\mathbf{y}_i-\mathbf{x}_i^\top\boldsymbol{\beta}) - 2\widehat{ub}_i^{(k)}\mathbf{Z}_i^\top(\mathbf{y}_i-\mathbf{x}_i^\top\boldsymbol{\beta}) + \widehat{ub^2}_i^{(k)}\mathbf{Z}_i^\top\mathbf{Z}_i\right) \\
&\quad -\frac{N}{2}\log\sigma_e^2 - m\log\sigma_b^2 - \frac{1}{2\sigma_b^2}\sum_{i=1}^{m}\left(\widehat{ub^2}_i^{(k)} + \widehat{t^2}_i^{(k)} - 2\lambda\widehat{tb}_i^{(k)} + \lambda^2\widehat{b}_i^{2^{(k)}}\right),
\end{aligned}
\tag{15}
$$

where $\widehat{t^d u^e b^f}_i^{(k)} = E\left[T_i^d U_i^e b_i^f|\mathbf{y}_i,\boldsymbol{\theta}^{(k)}\right],\ d,e,f>0.$

The M-step maximizes $Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)$ with respect to $\boldsymbol{\theta}$, obtaining a new estimate $\hat{\boldsymbol{\theta}}^{(k+1)}$. The EM algorithm of the model is described below:

E-step: Given $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{(k)}$, calculate for $i=1,\ldots,n$, $\hat{u}_i^{(k)}$, $\widehat{ub}_i^{(k)}$, $\widehat{ub^2}_i^{(k)}$, $\widehat{b^2}_i^{(k)}$, $\widehat{tb}_i^{(k)}$ and $\widehat{t^2}_i^{(k)}$ using (A5)–(A10) (Appendix A).

M-step: Update $\hat{\boldsymbol{\theta}}^{(k+1)}$ maximizing $Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right)$ about $\boldsymbol{\theta}$, obtaining the following analytic expressions

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}^{(k+1)} &= \left(\mathbf{X}^\top\mathbf{D}\left(\mathbf{U}^{*^{(k)}}\right)\mathbf{X}\right)^{-1}\mathbf{X}^\top\left(\mathbf{D}\left(\mathbf{U}^{*^{(k)}}\right)\mathbf{Y}-\mathbf{D}\left(\mathbf{Ub}^{*^{(k)}}\right)\mathbf{Z}\right), \\
\hat{\lambda}^{(k+1)} &= \frac{\sum_{i=1}^{m}\widehat{tb}_i^{(k)}}{\sum_{i=1}^{m}\widehat{b^2}_i^{(k)}}, \\
\widehat{\sigma_e^2}^{(k+1)} &= \frac{1}{N}\sum_{i=1}^{m}\left(\hat{u}_i^{(k)}\left(\mathbf{y}_i-\mathbf{x}_i^\top\boldsymbol{\beta}^{(k)}\right)^\top\left(\mathbf{y}_i-\mathbf{x}_i^\top\boldsymbol{\beta}^{(k)}\right) - 2\widehat{ub}_i^{(k)}\mathbf{Z}_i^\top\left(\mathbf{y}_i-\mathbf{x}_i^\top\boldsymbol{\beta}^{(k)}\right) + \widehat{ub^2}_i^{(k)}\mathbf{Z}_i^\top\mathbf{Z}_i\right) \text{ and} \\
\widehat{\sigma_b^2}^{(k+1)} &= \frac{1}{2m}\sum_{i=1}^{m}\left(\widehat{ub^2}_i^{(k)} + \widehat{t^2}_i^{(k)} - 2\lambda^{(k)}\widehat{tb}_i^{(k)} + \lambda^{(k)^2}\widehat{b^2}_i^{(k)}\right),
\end{aligned}
\tag{16}
$$

where $\mathbf{X}=\left(\mathbf{x}_1^\top,\ldots,\mathbf{x}_m^\top\right)^\top$ of dimension $N\times p$, $\mathbf{Z}=\left(\mathbf{Z}_1^\top,\ldots,\mathbf{Z}_m^\top\right)^\top$ dimension $N\times 1$, $\hat{\mathbf{U}}^{*^{(k)}}=\left[\hat{u}_1^{(k)}1_{n_1}^\top,\ldots,\hat{u}_m^{(k)}1_{n_m}^\top\right]^\top$, $\widehat{\mathbf{Ub}}^{*^{(k)}}=\left[\widehat{ub}_1^{(k)}1_{n_1}^\top,\ldots,\widehat{ub}_m^{(k)}1_{n_m}^\top\right]^\top$ and 1_k is a ones vector of length k, $k>0$. As recommended by Lange, Little, and Taylor (1989), Lucas (1997) and Berkane, Kano, and Bentler (1994), who pointed out difficulties in estimating $\boldsymbol{\tau}$ due to problems of unbounded and local maxima in the likelihood function, we consider the value of $\boldsymbol{\tau}$ to be known. In the application, we use the Akaike information criterion (AIC) (Akaike 1974) to select an appropriate value of the parameter $\boldsymbol{\tau}$.

The iterations of the above algorithms are repeated until a suitable convergence rule is satisfied, e.g. $||\boldsymbol{\theta}^{(k+1)}-\boldsymbol{\theta}^{(k)}||$ or $|\ell\left(\boldsymbol{\theta}^{(k+1)}\right)-\ell\left(\boldsymbol{\theta}^{(k)}\right)|$ is sufficiently small, say 10^{-6}. A set of reasonable initial values can be obtained by computing $\hat{\boldsymbol{\beta}}^{(0)}$, where $\hat{\sigma}_e^{2(0)}$ and $\hat{\sigma}_b^{2(0)}$ are the estimates of the normal linear mixed model. The estimate $\hat{\lambda}^{(0)}$ can be the skewness of $\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}^{(0)}$.

3.1.1. Simulation study

The objective of this study is to investigate the functionality of the proposed model and to check the standard errors provided by the approximation of the observed Fisher information matrix (see Appendix B). Similarly to Arellano-Valle, Bolfarine, and Lachos (2005), we formulate the following SSMN-LMM model:

Table 1. Mean, empirical standard error (*ESE*) and the standard deviation (*SD*) for estimates based on 500 samples from SSMN-LMM.

SN model									
Parameter (True value)	$m=200$			$m=500$			$m=1000$		
	Mean	*SD*	*ESE*	Mean	*SD*	*ESE*	Mean	*SD*	*ESE*
$\beta_0(-1)$	−0.972	0.131	0.494	−0.995	0.052	0.054	−0.999	0.038	0.037
$\beta_1(2)$	2.000	0.011	0.011	2.000	0.007	0.007	2.000	0.005	0.005
$\beta_2(1)$	0.996	0.070	0.070	1.002	0.043	0.044	1.001	0.030	0.031
$\sigma_e^2(0.25)$	0.249	0.013	0.013	0.250	0.008	0.008	0.250	0.006	0.006
$\sigma_b^2(0.5)$	0.486	0.107	0.129	0.498	0.058	0.063	0.499	0.043	0.044
$\lambda(3)$	3.379	1.548	3.518	3.131	0.836	1.004	3.099	0.605	0.657
StN model ($\nu=3$)									
Parameter (True value)	$m=200$			$m=500$			$m=1000$		
	Mean	*SD*	*ESE*	Mean	*SD*	*ESE*	Mean	*SD*	*ESE*
$\beta_0(-1)$	−0.978	0.089	0.092	−0.993	0.053	0.055	−0.995	0.035	0.039
$\beta_1(2)$	2.001	0.012	0.013	2.000	0.008	0.008	2.000	0.006	0.006
$\beta_2(1)$	0.997	0.074	0.078	1.000	0.048	0.048	0.999	0.034	0.034
$\sigma_e^2(0.25)$	0.250	0.023	0.021	0.252	0.016	0.013	0.251	0.011	0.010
$\sigma_b^2(0.5)$	0.484	0.099	0.109	0.498	0.065	0.069	0.499	0.049	0.048
$\lambda(3)$	3.054	1.093	1.896	3.056	0.688	0.994	2.946	0.478	0.646
SSL model ($\nu=3$)									
Parameter (True value)	$m=200$			$m=500$			$m=1000$		
	Mean	*SD*	*ESE*	Mean	*SD*	*ESE*	Mean	*SD*	*ESE*
$\beta_0(-1)$	−0.976	0.103	0.101	−0.992	0.055	0.059	−0.996	0.038	0.041
$\beta_1(2)$	2.000	0.014	0.013	2.000	0.008	0.008	2.000	0.006	0.006
$\beta_2(1)$	0.997	0.080	0.081	1.001	0.048	0.050	1.001	0.034	0.035
$\sigma_e^2(0.25)$	0.249	0.014	0.015	0.250	0.009	0.009	0.250	0.006	0.006
$\sigma_b^2(0.5)$	0.487	0.101	0.104	0.493	0.059	0.063	0.495	0.041	0.044
$\lambda(3)$	3.186	1.240	2.342	3.124	0.813	1.183	3.036	0.563	0.737
SCN model ($\nu=0.3$ and $\gamma=0.8$)									
Parameter (True value)	$m=200$			$m=500$			$m=1000$		
	Mean	*SD*	*ESE*	Mean	*SD*	*ESE*	Mean	*SD*	*ESE*
$\beta_0(-1)$	−0.977	0.118	0.115	−0.995	0.051	0.055	−0.999	0.035	0.037
$\beta_1(2)$	1.999	0.011	0.012	2.000	0.008	0.007	2.000	0.005	0.006
$\beta_2(1)$	1.004	0.074	0.072	1.005	0.044	0.045	0.999	0.033	0.032
$\sigma_e^2(0.25)$	0.249	0.013	0.013	0.250	0.008	0.008	0.250	0.006	0.006
$\sigma_b^2(0.5)$	0.486	0.094	0.105	0.492	0.057	0.061	0.499	0.039	0.043
$\lambda(3)$	3.267	1.305	2.156	3.131	0.844	1.045	3.102	0.582	0.670

$$
\begin{aligned}
Y_{ij} &= \beta_0 + \beta_1 t_{ij} + \beta_2 w_i + b_i + \varepsilon_{ij},\\
b_i &\sim \mathrm{SSMN}_1\left(0, \sigma_b^2, \lambda; H\right),\\
\varepsilon_{ij} &\sim \mathrm{SMN}_{n_i}\left(0_{n_i}, \sigma_e^2 \mathbf{I}_{n_i}; H\right),\quad i = 1, \ldots, m,
\end{aligned}
\tag{17}
$$

where $t_{ij}=j-3,\ j=1,\ldots,5,\ w_i=1$ for $i=1,\ldots,m/2$ and 0 elsewhere and $m=200$, 500 and 1000. The true values of the parameters are: $\beta_0=-1,\ \beta_1=2,\ \beta_2=1,\ \sigma_e^2=0.25,\ \sigma_b^2=0.5,\ \lambda=3$. For StN and SSL we use $\nu=3$ and for SCN, $\nu=0.3$ and $\gamma=0.8$ (which are considered fixed in the simulation study). We simulate 500 random samples of each SSMN-LMM. In each replication, we obtain the parameter estimates based on EM algorithm and the standard error estimate, that is, the mean square root of the diagonal elements of the inverse of the approximated observed information matrix (Appendix B). With 500 estimates of the parameters and the empirical standard errors, we present in the Table 1 the mean of the parameter estimates (Mean), the empirical standard errors (*ESE*) measured by the average values of the standard error estimates and the Monte Carlo standard deviation of the parameters (*SD*). This table shows that the bias related to all parameters tends to zero and the estimation method of the standard errors

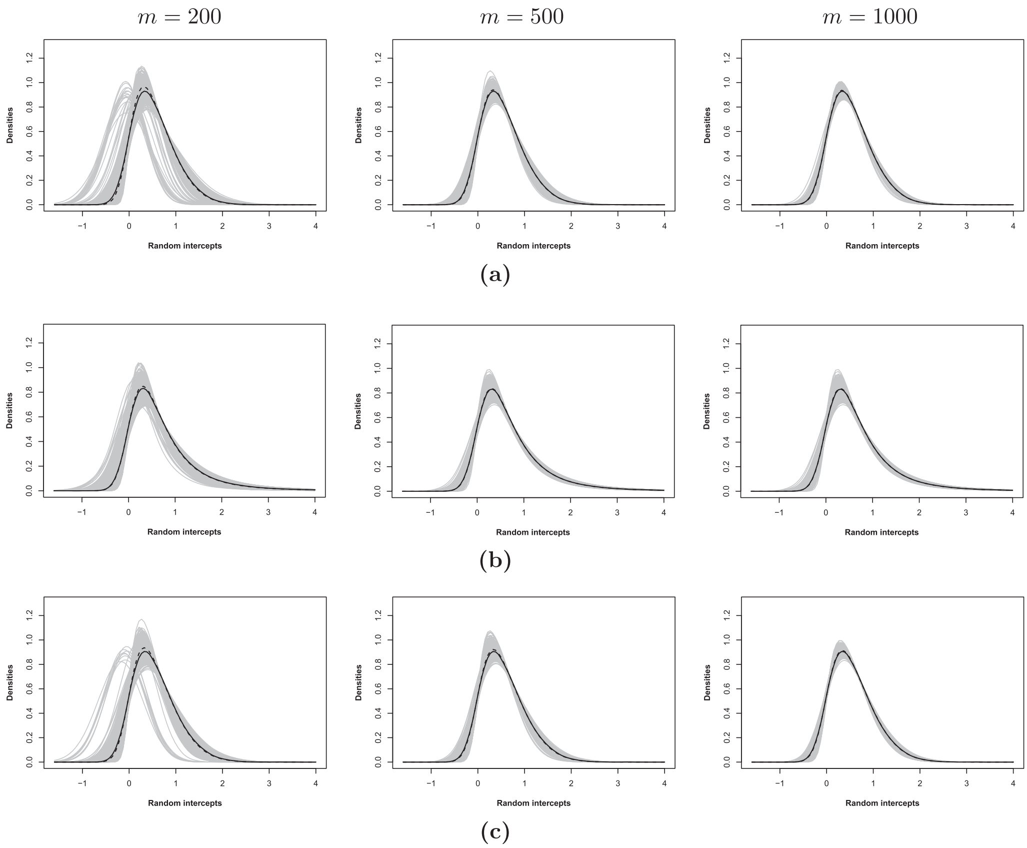

Figure 1. Simulation results based on 500 datasets. True density (solid line), density with mean of the estimated parameters (dashed line) and estimated densities of 500 random samples (grey lines). Random samples with $m = 200$ (left panel), $m = 500$ (center panel) and $m = 1000$ (right panel): (a) skew-normal; (b) skew Student-t normal; (c) skew-contaminated normal.

provides relatively close results when the sample size increases in all SSMN-LMM models. So, it seems that the ML estimates using the EM-type algorithm present good large sample properties and calculations of the observed information matrix are correct and reliable.

Also, for each parameter estimate of the random sample, we construct the density estimates of the random effects of the SSMN-LMM. Figure 1 shows the density estimates over the 500 datasets with length $m = 200$, $m = 500$ and $m = 1000$, the Monte Carlo average of this density estimates and the true densities, for SN, StN, SSL and SCN models. It seems that the additional flexibility afforded by the SSMN representation is sufficient to capture quite accurately the true underlying features of the random effects.

4. Local influence analysis

We use the approach of Zhu and Lee (2001) in the incomplete data context. Consider a perturbation vector $\boldsymbol{\omega}$ varying in an open region $\boldsymbol{\Omega} \in \mathbb{R}^q$. Let $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{y}_c)$, $\boldsymbol{\theta} \in \mathbb{R}^p$ be the complete-data log-likelihood of the perturbed model. We assume there is a $\boldsymbol{\omega}_0$ such that $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}_0|\mathbf{y}_c) = \ell_c(\boldsymbol{\theta}|\mathbf{y}_c)$ for all $\boldsymbol{\theta}$. Let $\hat{\boldsymbol{\theta}}_\omega$ be the maximizer of the function $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \mathrm{E}[\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{y}_c)|\mathbf{y}, \hat{\boldsymbol{\theta}}]$. Then, the influence graph is defined as $\boldsymbol{\alpha}(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, f_Q(\boldsymbol{\omega}))^\top$, where $f_Q(\boldsymbol{\omega})$ is the Q-displacement function defined as:

$$f_Q(\boldsymbol{\omega}) = 2\left[Q\left(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}\right) - Q\left(\hat{\boldsymbol{\theta}}_\omega|\hat{\boldsymbol{\theta}}\right)\right].$$

The normal curvature $C_{f_Q,\mathbf{d}}$, of $\boldsymbol{\alpha}(\boldsymbol{\omega})$ at $\boldsymbol{\omega}_0$ in the direction of a unit vector $\mathbf{d}$, which is used to summarize the local behavior of the Q-displacement function, is given by

$$C_{f_Q,\mathbf{d}}(\boldsymbol{\theta}) = -2\mathbf{d}^\top \ddot{Q}_{\boldsymbol{\omega}_0}\mathbf{d} = 2\mathbf{d}^\top \Delta_{\boldsymbol{\omega}_0}^\top \left\{-\ddot{Q}\left(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}\right)\right\}^{-1} \Delta_{\boldsymbol{\omega}_0}\mathbf{d}^\top$$

where $\ddot{Q}\left(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}\right) = \partial^2 Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}\right)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ and $\Delta_{\boldsymbol{\omega}} = \partial^2 Q\left(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}}\right)/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}}$. The expression $-\ddot{Q}_{\boldsymbol{\omega}_0} = \Delta_{\boldsymbol{\omega}_0}^\top \left\{-\ddot{Q}\left(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}\right)\right\}^{-1} \Delta_{\boldsymbol{\omega}_0}$ is used to detect influential observations.

We use the spectral decomposition of the $\ddot{Q}_{\boldsymbol{\omega}_o}$ to construct the measure $M(m_0)$ and its graphics. We have that

$$-2\ddot{Q}_{\boldsymbol{\omega}_o} = \sum_{k=1}^{n} \lambda_k \mathbf{e}_k \mathbf{e}_k^\top,$$

where $(\lambda_1, \mathbf{e}_1), \ldots, (\lambda_n, \mathbf{e}_n)$ are the eigenvalue-eigenvector pairs of the matrix $-2\ddot{Q}_{\boldsymbol{\omega}_o}$, with $\lambda_1 \geq \ldots \geq \lambda_q > \lambda_{q+1} = \ldots = \lambda_n = 0$ and $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are elements of the associated orthonormal basis. As in Zhu and Lee (2001), we use all eigenvectors corresponding to nonzero eigenvalues to detect influential observations. Let $\bar{\lambda}_k = \lambda_k/\left(\lambda_1 + \ldots + \lambda_q\right)$, $\mathbf{e}_k^2 = \left(e_{k1}^2, \ldots, e_{kn}^2\right)$ and

$$M(0) = \sum_{k=1}^{q} \bar{\lambda}_k \mathbf{e}_k^2.$$

Following Lee and Xu (2004), we use $1/n + c^* SM(0)$ as a benchmark to regard the lth case as influential, where c^* is an arbitrary constant (depending on the real application) and $SM(0)$ is the standard deviation of $\{M(0)_l, l = 1, \ldots, n\}$.

4.1. The Hessian matrix

The matrix $\ddot{Q}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})$ has elements given by:

$$\frac{\partial^2 Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}\right)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} = -\frac{1}{\sigma_e^2}\mathbf{X}^\top \mathbf{D}(\hat{\mathbf{U}}^*)\mathbf{X},$$

$$\frac{\partial^2 Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}\right)}{\partial\boldsymbol{\beta}\partial\sigma_e^2} = -\frac{1}{\sigma_e^4}\mathbf{X}^\top \left[\mathbf{D}(\hat{\mathbf{U}}^*)(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) - \mathbf{D}(\widehat{\mathbf{Ub}}^*)\mathbf{Z}\right],$$

$$\frac{\partial^2 Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}\right)}{\partial\sigma_e^2\partial\sigma_e^2} = \frac{N}{2\sigma_e^4} - \frac{1}{\sigma_e^6}\left[Q_U(\boldsymbol{\beta}) - 2(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^\top \mathbf{D}(\widehat{\mathbf{Ub}}^*)\mathbf{Z} + \mathbf{Z}^\top \mathbf{D}\left(\widehat{\mathbf{Ub}^2}^*\right)\mathbf{Z}\right],$$

$$\frac{\partial^2 Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}\right)}{\partial\sigma_b^2\partial\sigma_b^2} = \frac{m}{\sigma_b^4} - \frac{1}{\sigma_b^6}\left[\widehat{\mathbf{Ub}^2} + \widehat{\mathbf{t}^2} - 2\lambda\widehat{\mathbf{tb}} + \lambda^2\widehat{\mathbf{b}^2}\right]^\top \mathbf{1}_m,$$

$$\frac{\partial^2 Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}\right)}{\partial\sigma_b^2\partial\lambda} = \frac{1}{\sigma_b^4}\left[\lambda\widehat{\mathbf{b}^2} - \widehat{\mathbf{tb}}\right]^\top \mathbf{1}_m, \quad \frac{\partial^2 Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}\right)}{\partial\lambda\partial\lambda} = -\frac{1}{\sigma_b^2}\widehat{\mathbf{b}^2}^\top \mathbf{1}_m,$$

$$\frac{\partial^2 Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}\right)}{\partial\boldsymbol{\beta}\partial\sigma_b^2} = \frac{\partial^2 Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}\right)}{\partial\boldsymbol{\beta}\partial\lambda} = \mathbf{0} : p \times 1 \text{ and}$$

$$\frac{\partial^2 Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}\right)}{\partial\sigma_e^2\partial\sigma_b^2} = \frac{\partial^2 Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}\right)}{\partial\sigma_e^2\partial\lambda} = 0,$$

where $Q_U(\boldsymbol{\beta}) = (\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^\top \mathbf{D}(\hat{\mathbf{U}}^*)(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})$, $\widehat{\mathbf{t}^2} = [\widehat{t^2}_1, \ldots, \widehat{t^2}_m]^\top$, $\hat{\mathbf{U}}^* = [\hat{u}_1 1_{n_1}^\top, \ldots, \hat{u}_m 1_{n_m}^\top]^\top$, $\widehat{\mathbf{Ub}}^* = [\widehat{ub}_1 1_{n_1}^\top, \ldots, \widehat{ub}_m 1_{n_m}^\top]^\top$, $\widehat{\mathbf{Ub}^2}^* = [\widehat{ub^2}_1 1_{n_1}^\top, \ldots, \widehat{ub^2}_m 1_{n_m}^\top]^\top$, $\mathbf{D}$ is a diagonal matrix and $\mathbf{1}_m : m \times 1$ is a column vector of 1's.

4.2. Perturbation schemes

In this section, we consider six different perturbation schemes for SSMN-LMM.

4.2.1. Case weights perturbation

Let $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)^\top$ a vector $m \times 1$ with $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$. Then, the perturbed Q-function is given by

$$Q\left(\boldsymbol{\theta}, \boldsymbol{\omega} | \hat{\boldsymbol{\theta}}\right) = \sum_{i=1}^m \omega_i Q_i\left(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}\right),$$

where $Q_i\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}\right)$ is as in (15). In this case, the matrix $\boldsymbol{\Delta}_{\omega_0} = \left.\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0,} = \left(\boldsymbol{\Delta}_{\boldsymbol{\beta}}^\top, \boldsymbol{\Delta}_{\sigma_e^2}^\top, \boldsymbol{\Delta}_{\sigma_b^2}^\top, \boldsymbol{\Delta}_{\lambda}^\top\right)^\top$ has elements given by

$$\begin{aligned}
\boldsymbol{\Delta}_{\boldsymbol{\beta}} &= \frac{1}{\sigma_e^2} \mathbf{X}^\top \left[\mathbf{D}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})\mathbf{D}^*(\hat{\mathbf{U}}^*) - \mathbf{D}(\mathbf{Z})\mathbf{D}^*\left(\widehat{\mathbf{Ub}}^*\right)\right], \\
\boldsymbol{\Delta}_{\sigma_e^2} &= \frac{1}{2\sigma_e^4}\left[(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^\top \mathbf{D}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})\mathbf{D}^*(\hat{\mathbf{U}}^*) + \mathbf{Z}^\top \mathbf{D}(\mathbf{Z})\mathbf{D}^*\left(\widehat{\mathbf{Ub}^2}^*\right)\right. \\
&\quad \left. -2\mathbf{Z}^\top \mathbf{D}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})\mathbf{D}^*\left(\widehat{\mathbf{Ub}}^*\right)\right] - \frac{1}{2\sigma_e^2}[n_1 \ n_2 \ \ldots \ n_m], \\
\boldsymbol{\Delta}_{\sigma_b^2} &= -\frac{1}{\sigma_b^2} 1_m^\top + \frac{1}{2\sigma_b^4}\left[\widehat{\mathbf{Ub}^2}^\top + \widehat{\mathbf{t}^2} - 2\lambda \widehat{\mathbf{tb}} + \lambda^2 \widehat{\mathbf{b}^2}\right]^\top \text{ and} \\
\boldsymbol{\Delta}_{\lambda} &= \frac{1}{\sigma_b^2}\left[\widehat{\mathbf{tb}} - \lambda \widehat{\mathbf{b}^2}\right]^\top,
\end{aligned}$$

where

$$\mathbf{D}^*(\hat{\mathbf{a}}^*) = \begin{pmatrix} \hat{a}_1 \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \cdots & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \hat{a}_2 \mathbf{1}_{n_2} & \cdots & \mathbf{0}_{n_2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n_m} & \mathbf{0}_{n_m} & \cdots & \hat{a}_m \mathbf{1}_{n_m} \end{pmatrix}$$

is a matrix of order $N \times m$, $\mathbf{0}_{n_i}$ and $\mathbf{1}_{n_i}$ are vectors of zeros and ones, respectively, of length n_i.

4.2.2. Perturbation on the scale parameter σ_e^2

This perturbation scheme is introduced in the form $\sigma_{e_{\omega_i}}^2 = \sigma_e^2/\omega_i$, $\omega_i > 0$, $i = 1, \ldots, m$. So, the perturbed Q-function is given by

$$\begin{aligned}
Q\left(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}\right) &= -\frac{1}{2}\sum_{i=1}^m n_i \log \frac{\sigma_e^2}{\omega_i} - \frac{1}{2\sigma_e^2}\sum_{i=1}^m \omega_i \left(\hat{U}_i \mathbf{Q}_i(\boldsymbol{\beta}) - 2\widehat{Ub}_i \mathbf{Z}_i^\top (\mathbf{y}_i - \mathbf{x}_i^\top \boldsymbol{\beta}) + \widehat{Ub^2}_i \mathbf{Z}_i^\top \mathbf{Z}_i\right) \\
&\quad - m \log \sigma_b^2 - \frac{1}{2\sigma_b^2}\sum_{i=1}^m \widehat{ub^2}_i - \frac{1}{2\sigma_b^2}\sum_{i=1}^m\left(\widehat{t^2}_i - 2\lambda \widehat{tb}_i + \lambda^2 \widehat{b^2}_i\right).
\end{aligned}$$

The vector of non-perturbation is $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$ and the matrix $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$ has elements

$$\begin{aligned}
\boldsymbol{\Delta}_{\boldsymbol{\beta}} &= \frac{1}{\sigma_e^2}\mathbf{X}^\top\left[\mathbf{D}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})\mathbf{D}^*(\hat{\mathbf{U}}^*) - \mathbf{D}(\mathbf{Z})\mathbf{D}^*\left(\widehat{\mathbf{Ub}}^*\right)\right],\\
\boldsymbol{\Delta}_{\sigma_e^2} &= \frac{1}{2\sigma_e^4}\left[(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^\top\mathbf{D}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})\mathbf{D}^*(\hat{\mathbf{U}}^*) + \mathbf{Z}^\top\mathbf{D}(\mathbf{Z})\mathbf{D}^*\left(\widehat{\mathbf{Ub}^2}^*\right)\right.\\
&\qquad \left. -2\mathbf{Z}^\top\mathbf{D}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})\mathbf{D}^*\left(\widehat{\mathbf{Ub}}^*\right)\right],\\
\boldsymbol{\Delta}_{\sigma_b^2} &= \boldsymbol{\Delta}_\lambda = \mathbf{0}_m,
\end{aligned}$$

with $\mathbf{0}_m$ is a vector of zeros of length m.

4.2.3. Perturbation on the scale parameter σ_b^2

This scheme is obtained perturbing the parameter $\sigma^2_{b_{\omega_i}} = \sigma_b^2/\omega_i, \;\; \omega_i > 0, \;\; i = 1, \ldots, m$. The perturbed Q function has the form

$$\begin{aligned}
Q\left(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}\right) &= -\frac{N}{2}\log\sigma_e^2 - \frac{1}{2\sigma_e^2}\sum_{i=1}^m\left(\hat{U}_i\mathbf{Q}_i(\boldsymbol{\beta}) - 2\widehat{Ub}_i\mathbf{Z}_i^\top\left(\mathbf{y}_i - \mathbf{x}_i^\top\boldsymbol{\beta}\right) + \widehat{Ub^2}_i\mathbf{Z}_i^\top\mathbf{Z}_i\right)\\
&\quad -\sum_{i=1}^m\log\frac{\sigma_b^2}{\omega_i} - \frac{1}{2\sigma_b^2}\sum_{i=1}^m\omega_i\widehat{Ub^2}_i - \frac{1}{2\sigma_b^2}\sum_{i=1}^m\omega_i\left(\widehat{t^2}_i - 2\lambda\widehat{tb}_i + \lambda^2\widehat{b^2}_i\right).
\end{aligned}$$

Under this perturbation, the vector $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$ and the matrix $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$ has elements

$$\begin{aligned}
&\boldsymbol{\Delta}_{\boldsymbol{\beta}} = \mathbf{0}_{m,p} \text{ and } \boldsymbol{\Delta}_{\sigma_e^2} = \mathbf{0}_m,\\
&\boldsymbol{\Delta}_{\sigma_b^2} = \frac{1}{2\sigma_b^4}\left[\widehat{\mathbf{Ub}^2} + \widehat{\mathbf{t}^2} - 2\lambda\widehat{\mathbf{tb}} + \lambda^2\widehat{\mathbf{b}^2}\right]^\top \text{ and } \boldsymbol{\Delta}_\lambda = \frac{1}{\sigma_b^2}\left[\widehat{\mathbf{tb}} - \lambda\widehat{\mathbf{b}^2}\right]^\top,
\end{aligned}$$

where $\mathbf{0}_{m,p}$ is a matrix of zeros of order $m \times p$.

4.2.4. Perturbation on the asymmetric parameter λ

This perturbation scheme is introduced considering $\lambda_{\omega_i} = \omega_i\lambda$. Letting $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ and $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$, the perturbed Q-function is

$$\begin{aligned}
Q\left(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}\right) &= -\frac{N}{2}\log\sigma_e^2 - \frac{1}{2\sigma_e^2}\sum_{i=1}^m\left(\hat{U}_i\mathbf{Q}_i(\boldsymbol{\beta}) - 2\widehat{Ub}_i\mathbf{Z}_i^\top\left(\mathbf{y}_i - \mathbf{x}_i^\top\boldsymbol{\beta}\right) + \widehat{Ub^2}_i\mathbf{Z}_i^\top\mathbf{Z}_i\right)\\
&\quad - m\log\sigma_b^2 - \frac{1}{2\sigma_b^2}\sum_{i=1}^m\widehat{Ub^2}_i - \frac{1}{2\sigma_b^2}\sum_{i=1}^m\left(\widehat{t^2}_i - 2\lambda\omega_i\widehat{tb}_i + \lambda^2\omega_i^2\widehat{b^2}_i\right).
\end{aligned}$$

The matrix $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$ has elements

$$\begin{aligned}
&\boldsymbol{\Delta}_{\boldsymbol{\beta}} = \mathbf{0}_{m,p} \text{ and } \boldsymbol{\Delta}_{\sigma_e^2} = \mathbf{0}_m,\\
&\boldsymbol{\Delta}_{\sigma_b^2} = \frac{\lambda}{\sigma_b^4}\left[\lambda\widehat{\mathbf{b}^2} - \widehat{\mathbf{tb}}\right]^\top \text{ and } \boldsymbol{\Delta}_\lambda = \frac{1}{\sigma_b^2}\left[\widehat{\mathbf{tb}} - 2\lambda\widehat{\mathbf{b}^2}\right]^\top.
\end{aligned}$$

4.2.5. Explanatory variable perturbation

The interest here is to perturb a particular explanatory variable allowing, for example, to detect ill-conditioning in the matrix $\mathbf{X}$ (Belsley 1991). With this condition, the perturbation scheme of variable X_v has the representation:

$$\mathbf{x}_{iv_{\omega_i}}^{\top} = \mathbf{x}_i^{\top} + S_v \omega_i \mathbf{A}_{n_i,p}, \qquad v \in \{1, ..., p\},$$

where S_v is the standard deviation of the variable $\mathbf{X}_v$ and $\mathbf{A}_{n_i,p}$ is a matrix of zeros of order $n_i \times p$, with the vth column of 1s. In this case, $\boldsymbol{\omega}_0 = \mathbf{0}_m$ and the perturbed Q-function is given by:

$$\begin{aligned} Q\left(\boldsymbol{\theta}, \boldsymbol{\omega} | \hat{\boldsymbol{\theta}}\right) = & -\frac{1}{2\sigma_e^2} \sum_{i=1}^{m} \left[\hat{U}_i \left(\mathbf{Q}_i(\boldsymbol{\beta}) - 2\omega_i S_v \beta_v 1_{n_i}^{\top} (\mathbf{y}_i - \mathbf{x}_i^{\top} \boldsymbol{\beta}) + n_i \omega_i^2 S_v^2 \beta_v^2 \right) \right. \\ & \left. -2\widehat{Ub}_i \mathbf{Z}_i^{\top} (\mathbf{y}_i - \mathbf{x}_i^{\top} \boldsymbol{\beta} - S_v \omega_i \beta_v \mathbf{1}_{n_i}) + \widehat{Ub^2}_i \mathbf{Z}_i^{\top} \mathbf{Z}_i \right] \\ & -\frac{N}{2} \log \sigma_e^2 - m \log \sigma_b^2 - \frac{1}{2\sigma_b^2} \sum_{i=1}^{m} \widehat{ub^2}_i - \frac{1}{2\sigma_b^2} \sum_{i=1}^{m} \left(\widehat{t^2}_i - 2\lambda \widehat{tb}_i + \lambda^2 \widehat{b^2}_i \right). \end{aligned}$$

It follows that the matrix $\boldsymbol{\Delta}_{\omega_0}$ is given by:

$$\begin{aligned} \boldsymbol{\Delta}_{\boldsymbol{\beta}} &= \frac{S_v}{\sigma_e^2} \left[\mathbf{A}_{v1}^{\top} \mathbf{D}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \mathbf{D}^*(\hat{\mathbf{U}}^*) - \beta_v \mathbf{X}^{\top} \mathbf{D}^*(\hat{\mathbf{U}}^*) - \mathbf{A}_{v1}^{\top} \mathbf{D}^*(\mathbf{Z}) \mathbf{D}^*\left(\widehat{\mathbf{Ub}}^*\right) \right], \\ \boldsymbol{\Delta}_{\sigma_e^2} &= \frac{S_v \beta_v}{\sigma_e^4} \left[-(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^{\top} \mathbf{D}^*(\hat{\mathbf{U}}^*) + \mathbf{Z}^{\top} \mathbf{D}^*\left(\widehat{\mathbf{Ub}}^*\right) \right] \text{ and} \\ \boldsymbol{\Delta}_{\sigma_b^2} &= \boldsymbol{\Delta}_{\lambda} = \mathbf{0}_m. \end{aligned}$$

4.2.6. Response variable perturbation

The perturbation of the response variable Y is in the form $\mathbf{y}_{i_{\omega_i}} = \mathbf{y}_i + S_y \omega_i \mathbf{1}_{n_i}$, where S_y is the standard deviation of the vector of the observed values $\mathbf{Y}$. In this case, $\boldsymbol{\omega}_0 = \mathbf{0}_m$ and

$$\begin{aligned} Q\left(\boldsymbol{\theta}, \boldsymbol{\omega} | \hat{\boldsymbol{\theta}}\right) = & -\frac{1}{2\sigma_e^2} \sum_{i=1}^{m} \left[\hat{U}_i \left(\mathbf{Q}_i(\boldsymbol{\beta}) + 2\omega_i S_y 1_{n_i}^{\top} (\mathbf{y}_i - \mathbf{x}_i^{\top} \boldsymbol{\beta}) + n_i \omega_i^2 S_y^2 \right) \right. \\ & \left. -2\widehat{Ub}_i \mathbf{Z}_i^{\top} \left(\mathbf{y}_i - \mathbf{x}_i^{\top} \boldsymbol{\beta} + S_y \omega_i \mathbf{1}_{n_i} \right) + \widehat{Ub^2}_i \mathbf{Z}_i^{\top} \mathbf{Z}_i \right] \\ & -\frac{N}{2} \log \sigma_e^2 - m \log \sigma_b^2 - \frac{1}{2\sigma_b^2} \sum_{i=1}^{m} \widehat{ub^2}_i - \frac{1}{2\sigma_b^2} \sum_{i=1}^{m} \left(\widehat{t^2}_i - 2\lambda \widehat{tb}_i + \lambda^2 \widehat{b^2}_i \right). \end{aligned}$$

It follows that the matrix $\boldsymbol{\Delta}_{\omega_0}$ has elements

$$\begin{aligned} \boldsymbol{\Delta}_{\boldsymbol{\beta}} &= \frac{S_y}{\sigma_e^2} \mathbf{X}^{\top} \mathbf{D}^*(\hat{\mathbf{U}}^*), \\ \boldsymbol{\Delta}_{\sigma_e^2} &= \frac{S_y}{\sigma_e^4} \left[(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^{\top} \mathbf{D}^*(\hat{\mathbf{U}}^*) - \mathbf{Z}^{\top} \mathbf{D}^*\left(\widehat{\mathbf{Ub}}^*\right) \right] \text{ and} \\ \boldsymbol{\Delta}_{\sigma_b^2} &= \boldsymbol{\Delta}_{\lambda} = \mathbf{0}_m. \end{aligned}$$

5. Application

The mixed models developed in this article are applied to a conjoint longitudinal dataset collected as part of the famed Framingham heart study, analyzed previously by Lachos et al. (2007). The conjoint dataset includes the levels of blood cholesterol over time according to age and sex of $m = 200$ individuals selected randomly. Thus, we can formulate the following SSMN-LMM model:

$$\begin{aligned} Y_{ij} &= \beta_0 + \beta_1 \text{sex}_i + \beta_2 \text{age}_i + \beta_3 t_{ij} + b_i + \varepsilon_{ij}, \\ b_i &\sim \text{SSMN}_1\left(0, \sigma_b^2, \lambda; H\right), \\ \varepsilon_{ij} &\sim \text{SMN}_{n_i}\left(\mathbf{0}_{n_i}, \sigma_e^2 \mathbf{I}_{n_i}; H\right), \ i = 1, ..., 200, \end{aligned} \tag{18}$$

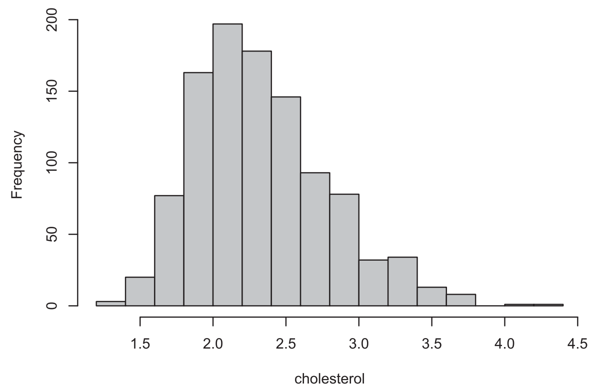

Figure 2. Histogram of the cholesterol (Framingham data).

where Y_{ij} is the level of cholesterol (mg/DL) divided by 100 at the jth sampling time for subject i and t_{ij} is (time − 5)/10, with time measured in years from baseline; age$_i$ is age at baseline; and sex$_i$ is the gender indicator (0 = female, 1 = male). The number of repetitions j of each individual varies from 1 to 6. The histogram of the dependent variable is presented in Figure 2, which indicates large skewness (sample skewness of 0.69). Table 2 presents the descriptive statistics for the levels of cholesterol and age of the individuals.

Table 3 presents the estimates of the parameters of the Normal-LMM and SSMN-LMM, together with the standard errors of the estimates. We use the AIC to choose the best values for ν and γ. Notice that the estimates of regression coefficients are similar in all models, but the scale and asymmetry parameters are slightly different, since the skew-normal model has a higher level of asymmetry than the others. Note also that in all models, only the regression coefficients of the sex variable are not significant. The AIC value, based on the likelihood function, indicates that the data are best fitted utilizing the skew contaminated-normal mixed model, although, the skew Student-t normal mixed model presents a close value.

With the aim to compare the SSMN-LMM with a competitor, we fitted the in Eq. (18) using the scale mixtures of skew-normal (SMSN) distributions (Branco and Dey 2001), using the package "skewlmm" (Schumacher, Matos, and Lachos 2020). This class of distributions incorporates asymmetry and heavy tails but presenting different coefficients of asymmetry and kurtosis than SSMN. The particular cases of this class are the "skew-t," "skew-slash" and "skew contaminated-normal" distributions. Besides, both SSMN and SMSN belong to a broad and flexible class of distributions obtained by both scale and shape mixtures of skew-normal distributions (Arellano-Valle, Ferreira, and Genton 2018). According to Table 4, we obtained similar log-likelihoods and AIC values of the competitors SMSN, in particular StN and SCN, but the SSMN model was slightly better than SMSN.

5.1. Diagnostic analysis

The previous results indicate that the SSMN-LMM models fit that dataset better, with the SNC distribution significantly better. Thus, now we realize diagnostic analysis for the SNC-LMM model in the perturbations schemes discussed in Sec. 4.

The graphs of diagnostics ($M(0)$) are shown in the Figure 3, with cutoff points in the form $\overline{M(0)}) + c^* SD(M(0))$, with c^* equals 2 or 3, where $SD(\cdot)$ is the standard deviation.

Individual 39 exerts a large influence on the case weight and in the scale parameter σ_b^2 perturbations, still exerting influence on the perturbation in the parameter of scale σ_e^2. This individual,

Table 2. Framingham cholesterol data: Descriptive statistics for the levels of cholesterol and age of the individuals.

	Minimum	Mean	Maximum	Standard deviation
Level of cholesterol (mg/DL)/100	1.29	2.34	4.30	0.46
Age (years)	31.0	42.47	62.0	7.89

Table 3. Estimates of the parameters of the SSMN-LMM models for cholesterol data. Estimates of the asymptotic standard errors are given in parentheses.

	Models				
Parameters	Normal	SN	SSL	StN	SCN
β_0	1.715(0.152)	1.476(0.134)	1.511(0.132)	1.526(0.130)	1.518(0.131)
β_1	−0.013(0.055)	−0.026(0.049)	−0.024(0.048)	−0.031(0.047)	−0.029(0.048)
β_2	0.015(0.004)	0.011(0.003)	0.011(0.003)	0.011(0.003)	0.012(0.003)
β_3	0.283(0.020)	0.282(0.020)	0.279(0.019)	0.271(0.019)	0.274(0.019)
σ_e^2	0.049(0.002)	0.049(0.001)	0.026(0.001)	0.036(0.001)	0.029(0.001)
σ_b^2	0.141(0.015)	0.317(0.044)	0.137(0.020)	0.186(0.021)	0.134(0.015)
λ	–	2.892(0.297)	1.515(0.131)	1.730(0.129)	1.302(0.106)
ν	–	–	3	7	0.3
γ	–	–	–	–	0.3
$\ell(\hat{\boldsymbol{\theta}})$	– 186.081	−167.632	−154.129	−141.606	**−140.344**
AIC	384.162	349.264	324.258	299.212	**298.688**

Table 4. Log-likelihoods and AIC values of the SMSN-LMM.

Distribution	Skew-slash	Skew-*t*	Skew contaminated-normal
$\ell(\hat{\boldsymbol{\theta}})$	−145.237	−142.693	−140.369
AIC	306.474	301.386	298.738

a woman, has low cholesterol levels (mean of 1.5) and advanced age of 59 years in relation to the others.

The perturbation analysis in the parameter of asymmetry λ indicates several points possibly exerting influence (2, 7, 26,131, 160, 172 and 174). These individuals have high levels of cholesterol, without specific characterization of age and sex. These individuals possibly affect, in addition to the estimation of the λ, the estimated mean and variance of the model.

The perturbation analysis in the response variable and in the explanatory variable age present individuals 43 (male) and 156 (female) as influential. They are characterized by having advanced age, 62 and 57 years, respectively.

6. Conclusions

The article presented a new class of linear mixed models, through of the class of asymmetric scale of mixtures of normal distributions developed by Ferreira, Bolfarine, and Lachos (2011) and Ferreira, Lachos, and Bolfarine (2016). An EM algorithm was developed, which presented analytic expressions for the estimators at the M-step. The approximated observed information matrix was computed by the product of the score vector based on the Q-function. Diagnostic analysis was developed using the approach of Zhu and Lee (2001). A simulation study was developed to evaluate the performance of the maximum likelihood estimators, the accuracy of the standard errors and the estimates of the random effects. Moreover, we showed that the skew contaminated-normal and skew Student-t models generate better results than the normal and skew-normal ones, in the context of linear mixed models for the Framingham cholesterol dataset. Finally, the local influence analysis was applied to the best model (SCN-LMM).

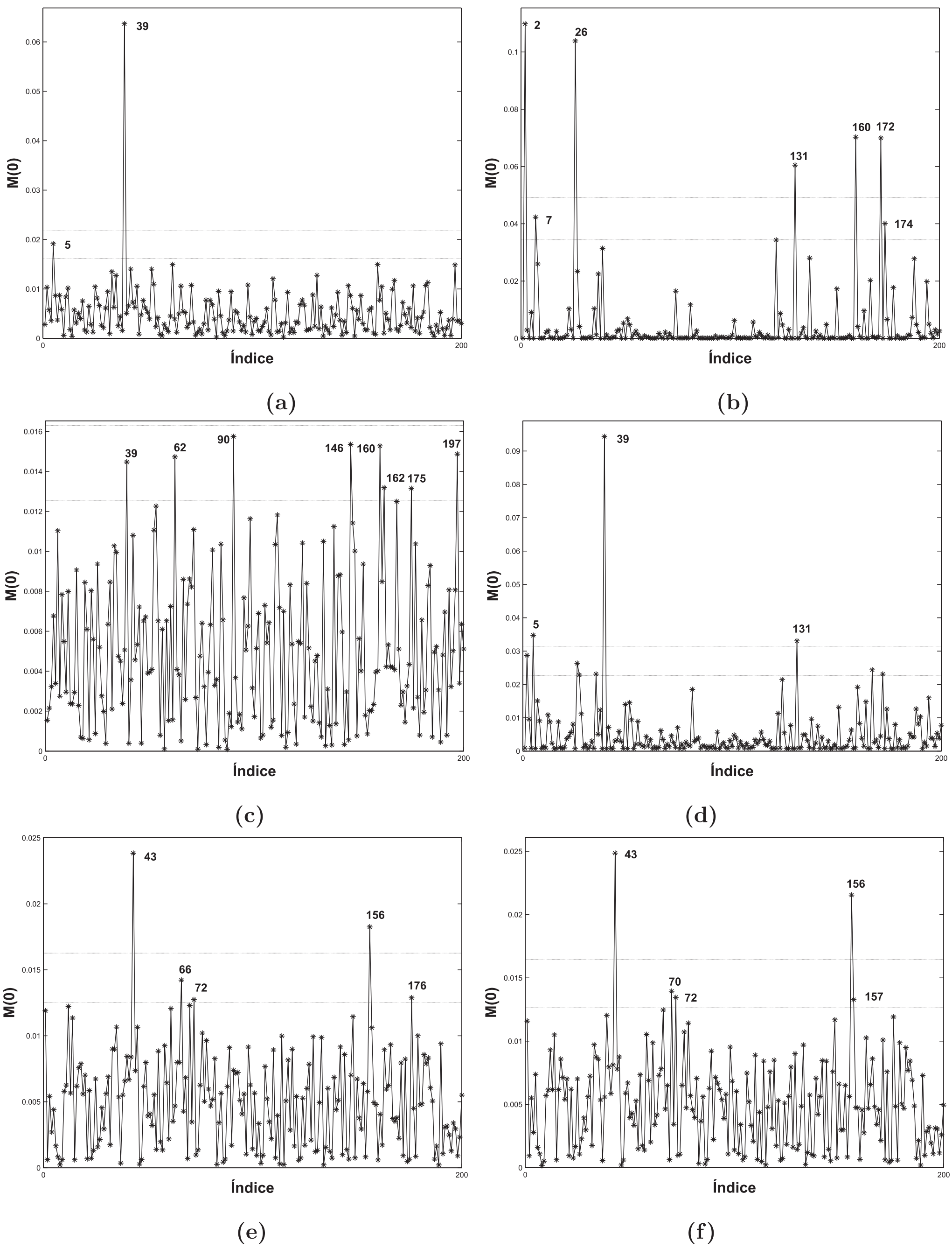

Figure 3. Diagnostic analysis to cholesterol data: (a) Case-weight perturbation, (b) Perturbation on the asymmetric parameter λ, (c) Perturbation on the scale parameter σ_e^2, (d) Perturbation on the scale parameter σ_b^2, (e) Perturbation on the response variable and (f) Perturbation on the explanatory variable age.

Of course, further extensions of the current work are possible. For example, the proposed method can be naturally extended by considering a multidimensional random effect. Other extensions include a Bayesian treatment via Markov chain Monte Carlo (MCMC) sampling methods in the context of SSMN-LMM.

Appendix A

Elements of the E-step of the EM algorithm

We have that:

$$f(u_i, b_i, t_i|\mathbf{y}_i, \boldsymbol{\theta}) = f(b_i|\mathbf{y}_i, u_i, t_i, \boldsymbol{\theta})f(t_i|\mathbf{y}_i, u_i, \boldsymbol{\theta})f(u_i|\mathbf{y}_i, \boldsymbol{\theta}).$$

Hence,

$$\widehat{ub}_i = E_{U_i}\left[U_i E_{t_i}\left[E_{b_i}\left[b_i|\mathbf{y}_i, u_i, t_i\right]|\mathbf{y}_i, u_i\right]|\mathbf{y}_i, \boldsymbol{\theta}\right].$$

From Eq. (14), we have that:

$$f(\mathbf{y}_i|\boldsymbol{\theta}) = \int_{-\infty}^{+\infty}\int_0^{+\infty}\int_0^{+\infty} f(\mathbf{y}_i|b_i, u_i, \boldsymbol{\theta})f(b_i|u_i, \boldsymbol{\theta})f(t_i|b_i, \boldsymbol{\theta})f(u_i, \boldsymbol{\theta})dt_i du_i db_i.$$

From lemma 2.2 of Lachos (2004), since $k(u) = 1/u$, we have that:

$$\begin{aligned} f(\mathbf{y}_i|b_i, u_i, \boldsymbol{\theta})f(b_i|u_i, \boldsymbol{\theta}) &= \phi_{n_i}\left(\mathbf{y}_i|\mathbf{x}_i^\top\boldsymbol{\beta} + b_i\mathbf{Z}_i, \sigma_e^2/u_i\mathbf{I}_{n_i}\right)\phi_1\left(b_i|0, \sigma_b^2/u_i\right) \\ &= \phi_{n_i}\left(\mathbf{y}_i|\mathbf{x}_i^\top\boldsymbol{\beta}, u_i^{-1}\left[\sigma_e^2\mathbf{I}_{n_i} + \sigma_b^2\mathbf{Z}_i\mathbf{Z}_i^\top\right]\right)\phi_1\left(b_i|\Lambda_i\sigma_e^{-2}u_i\mathbf{Z}_i^\top\left(\mathbf{y}_i - \mathbf{x}_i^\top\boldsymbol{\beta}\right), \Lambda_i\right) = f(\mathbf{y}_i|u_i, \boldsymbol{\theta})f(b_i|\mathbf{y}_i, u_i, \boldsymbol{\theta}), \end{aligned} \tag{A1}$$

where $\Lambda_i = u_i^{-1}\left(\sigma_b^{-2} + \sigma_e^{-2}\mathbf{Z}_i^\top\mathbf{Z}_i\right)^{-1}$. Thus, the joint density can be expressed as $f(\mathbf{y}_i, u_i, b_i, t_i|\boldsymbol{\theta}) = f(\mathbf{y}_i|u_i, \boldsymbol{\theta})f(b_i|\mathbf{y}_i, u_i, \boldsymbol{\theta})f(t_i|b_i, \boldsymbol{\theta})f(u_i, \boldsymbol{\theta})$. Using the same Lemma, we have that

$$\begin{aligned} f(b_i|\mathbf{y}_i, u_i, \boldsymbol{\theta})f(t_i|b_i, \boldsymbol{\theta}) &= \phi_1\left(b_i|\Lambda_i^*\sigma_e^{-2}\mathbf{Z}_i^\top\left(\mathbf{y}_i - \mathbf{x}_i^\top\boldsymbol{\beta}\right), \Lambda_i\right)\phi_1\left(t_i|\lambda b_i, \sigma_b^2\right)\mathbb{I}(t_i > 0) \\ &= \phi_1\left(b_i|\eta_i + \lambda\sigma_b^{-2}\delta_i(t_i - \lambda\eta_i), \delta_i\right)\phi_1\left(t_i|\lambda\eta_i, \sigma_b^2 + \lambda^2\Lambda_i\right)\mathbb{I}(t_i > 0) \\ &= f(b_i|\mathbf{y}_i, u_i, t_i, \boldsymbol{\theta})f(t_i|\mathbf{y}_i, u_i, \boldsymbol{\theta}), \end{aligned}$$

where $\eta_i = \Lambda_i^*\sigma_e^{-2}\mathbf{Z}_i^\top\left(\mathbf{y}_i - \mathbf{x}_i^\top\boldsymbol{\beta}\right)$, $\Lambda_i^* = \left(\sigma_b^{-2} + \sigma_e^{-2}\mathbf{Z}_i^\top\mathbf{Z}_i\right)^{-1}$ and $\delta_i = \left(\Lambda_i^{-1} + \lambda^2\sigma_b^{-2}\right)^{-1}$. So, we have the following relation:

$$\begin{aligned} f(u_i, b_i, t_i|\mathbf{y}_i, \boldsymbol{\theta}) &= \phi_{n_i}\left(\mathbf{y}_i|\mathbf{x}_i^\top\boldsymbol{\beta}, u_i^{-1}\left[\sigma_e^2\mathbf{I}_{n_i} + \sigma_b^2\mathbf{Z}_i\mathbf{Z}_i^\top\right]\right) \\ &\times \phi_1\left(b_i|\eta_i + \lambda\sigma_b^{-2}\delta_i(t_i - \lambda\eta_i), \delta_i\right)\phi_1\left(t_i|\lambda\eta_i, \sigma_b^2 + \lambda^2\Lambda_i\right)\mathbb{I}(t_i > 0). \end{aligned} \tag{A2}$$

Let $Z_i = \frac{\lambda\eta_i}{\sqrt{\sigma_b^2 + \lambda^2\Lambda_i}}$. Since $T_i|\mathbf{y}_i, u_i, \boldsymbol{\theta} \sim TN\left(\lambda\eta_i, \sigma_b^2 + \lambda^2\Lambda_i, (0, +\infty)\right)$, then, we have that

$$E\left[T_i|\mathbf{y}_i, u_i, \boldsymbol{\theta}\right] = \lambda\eta_i + \sqrt{\sigma_b^2 + \lambda^2\Lambda_i}W_\Phi(Z_i) \text{ and} \tag{A3}$$

$$E\left[T_i^2|\mathbf{y}_i, u_i, \boldsymbol{\theta}\right] = \lambda^2\eta_i^2 + \sigma_b^2 + \lambda^2\Lambda_i + \lambda\eta_i\sqrt{\sigma_b^2 + \lambda^2\Lambda_i}W_\Phi(Z_i). \tag{A4}$$

From (A2), we have that $b_i|\mathbf{y}_i, u_i, t_i, \boldsymbol{\theta} \sim N_1\left(\eta_i + \frac{\lambda}{\sigma_2^b}\delta_i(t_i - \lambda\eta_i), \delta_i\right)$, and then, the necessary useful quantities for the implementation of the EM algorithm are given in the following equations:

$$\widehat{ub}_i = \eta_i\hat{U}_i + \frac{\lambda}{\sigma_b^2}E_{U_i}\left[U_i\delta_i\sqrt{\sigma_b^2 + \lambda^2\Lambda_i}W_\Phi(Z_i)|\mathbf{y}_i, \boldsymbol{\theta}\right], \tag{A5}$$

$$\begin{aligned} \widehat{ub^2}_i = \eta_i^2\hat{U}_i &+ \frac{\lambda^2}{\sigma_b^2}E_{U_i}\left(U_i\delta_i^2|\mathbf{y}_i, \boldsymbol{\theta}\right) + \frac{2\lambda\eta_i}{\sigma_b^2}E_{U_i}\left[U_i\delta_i\sqrt{\sigma_b^2 + \lambda^2\Lambda_i}W_\Phi(Z_i)|\mathbf{y}_i, \boldsymbol{\theta}\right] \\ &+ E_{U_i}\left[U_i\delta_i|\mathbf{y}_i, \boldsymbol{\theta}\right] + \frac{\lambda^4\Lambda_i^*}{\sigma_b^4}E_{U_i}\left(\delta_i^2|\mathbf{y}_i, \boldsymbol{\theta}\right) \\ &- \frac{\lambda^3\eta_i}{\sigma_b^4}E_{U_i}\left[U_i\delta_i^2\sqrt{\sigma_b^2 + \lambda^2\Lambda_i}W_\Phi(Z_i)|\mathbf{y}_i, \boldsymbol{\theta}\right], \end{aligned} \tag{A6}$$

$$\begin{aligned} \widehat{b^2}_i = \eta_i^2 &+ \frac{\lambda^2}{\sigma_b^2}E_{U_i}\left(\delta_i^2|\mathbf{y}_i, \boldsymbol{\theta}\right) + \frac{2\lambda\eta_i}{\sigma_b^2}E_{U_i}\left[\delta_i\sqrt{\sigma_b^2 + \lambda^2\Lambda_i}W_\Phi(Z_i)|\mathbf{y}_i, \boldsymbol{\theta}\right] \\ &+ E_{U_i}\left[\delta_i|\mathbf{y}_i, \boldsymbol{\theta}\right] + \frac{\lambda^4}{\sigma_b^4}E_{U_i}\left(\Lambda_i\delta_i^2|\mathbf{y}_i, \boldsymbol{\theta}\right) - \frac{\lambda^3\eta_i}{\sigma_b^4}E_{U_i}\left[\delta_i^2\sqrt{\sigma_b^2 + \lambda^2\Lambda_i}W_\Phi(Z_i)|\mathbf{y}_i, \boldsymbol{\theta}\right], \end{aligned} \tag{A7}$$

$$\widehat{tb}_i = \lambda\eta_i^2 + \lambda E_{U_i}\left(\delta_i|\mathbf{y}_i,\boldsymbol{\theta}\right) + \frac{\lambda^3}{\sigma_b^2}E_{U_i}\left(\Lambda_i\delta_i|\mathbf{y}_i,\boldsymbol{\theta}\right) + \eta_i E_{U_i}\left[\sqrt{\sigma_b^2+\lambda^2\Lambda_i}\,W_\Phi(Z_i)\big|\mathbf{y}_i,\boldsymbol{\theta}\right], \tag{A8}$$

$$\hat{t}_i = \lambda\eta_i + E_{U_i}\left[\sqrt{\sigma_b^2+\lambda^2\Lambda_i}\,W_\Phi(Z_i)\big|\mathbf{y}_i,\boldsymbol{\theta}\right] \text{ and} \tag{A9}$$

$$\widehat{t^2}_i = \lambda^2\eta_i^2 + \sigma_b^2 + \lambda^2 E_{U_i}\left(\Lambda_i|\mathbf{y}_i,\boldsymbol{\theta}\right) + \lambda\eta_i E_{U_i}\left[\sqrt{\sigma_b^2+\lambda^2\Lambda_i}\,W_\Phi(Z_i)\big|\mathbf{y}_i,\boldsymbol{\theta}\right]. \tag{A10}$$

Calculation of $E_{U_i}\left[g(U_i)|\mathbf{y}_i,\boldsymbol{\theta}\right]$, where g is an integrable real function, as presented in Eqs. (A5)–(A10).

From Eq. (A1), we have that

$$\begin{aligned} f\left(\mathbf{y}_i|\boldsymbol{\theta}\right) &= 2\int_0^{+\infty}\int_{-\infty}^{+\infty}\phi_{n_i}\left(\mathbf{y}_i|\mathbf{x}_i^\top\boldsymbol{\beta}, u_i^{-1}\left[\sigma_e^2\mathbb{I}_{n_i}+\sigma_b^2\mathbf{Z}_i\mathbf{Z}_i^\top\right]\right)\phi\left(b_i|\sigma_e^{-2}\Lambda_i^*\mathbf{Z}_i^\top\left(\mathbf{y}_i-\mathbf{x}_i^\top\boldsymbol{\beta}\right),\Lambda_i\right) \\ &\quad\times\Phi\left(\frac{\lambda b_i}{\sigma_b}\right)h(u_i;\boldsymbol{\tau})db_i du_i \\ &= 2\int_0^{+\infty}\phi_{n_i}\left(\mathbf{y}_i|\mathbf{x}_i^\top\boldsymbol{\beta}, u_i^{-1}\left[\sigma_e^2\mathbb{I}_{n_i}+\sigma_b^2\mathbf{Z}_i\mathbf{Z}_i^\top\right]\right)E_{b_i|\mathbf{y}_i,u_i}\left[\Phi\left(\frac{\lambda b_i}{\sigma_b}\right)\right]h(u_i;\boldsymbol{\tau})du_i. \end{aligned} \tag{A11}$$

By lemma 1 given in Arellano-Valle, Bolfarine, and Lachos (2005), for $\mu = \sigma_e^{-2}\Lambda_i^*\mathbf{Z}_i^\top\left(\mathbf{y}_i-\mathbf{x}_i^\top\boldsymbol{\beta}\right)=\eta_i$, $\Sigma=\Lambda_i$, $a=0$, $B=\lambda/\sigma_b$, $\eta=0$ and $\Omega=1$, we have that:

$$E_{b_i}\left[\Phi\left(\frac{\lambda b_i}{\sigma_b}\right)\Big|\mathbf{y}_i,u_i,\boldsymbol{\theta}\right] = \Phi\left(\frac{\lambda\eta_i}{\sqrt{\sigma_b^2+\lambda^2\Lambda_i}}\right).$$

Therefore,

$$f\left(\mathbf{y}_i|\boldsymbol{\theta}\right) = 2\int_0^{+\infty}\phi_{n_i}\left(\mathbf{y}_i|\mathbf{x}_i^\top\boldsymbol{\beta}, u_i^{-1}\left[\sigma_e^2\mathbb{I}_{n_i}+\sigma_b^2\mathbf{Z}_i\mathbf{Z}_i^\top\right]\right)\Phi\left(\frac{\lambda\eta_i}{\sqrt{\sigma_b^2+\lambda^2\Lambda_i}}\right)h(u_i;\boldsymbol{\tau})du_i. \tag{A12}$$

The conditional expectation $E_{U_i}\left[g(U_i)|\mathbf{y}_i,\boldsymbol{\theta}\right]$ can be rewritten as:

$$E_{U_i|\mathbf{y}_i}\left[g(U_i)\right] = \int g(u_i)f\left(u_i|\mathbf{y}_i,\boldsymbol{\theta}\right)du_i = \frac{\int g(u_i)f\left(\mathbf{y}_i|u_i,\boldsymbol{\theta}\right)h(u_i;\boldsymbol{\tau})du_i}{\int f\left(\mathbf{y}_i|u_i,\boldsymbol{\theta}\right)h(u_i;\boldsymbol{\tau})du_i}.$$

According to (A12), $f\left(\mathbf{y}_i|u_i,\boldsymbol{\theta}\right)$ is given by

$$f\left(\mathbf{y}_i|u_i,\boldsymbol{\theta}\right) = 2\phi_{n_i}\left(\mathbf{y}_i|\mathbf{x}_i^\top\boldsymbol{\beta}, u_i^{-1}\left[\sigma_e^2\mathbb{I}_{n_i}+\sigma_b^2\mathbf{Z}_i\mathbf{Z}_i^\top\right]\right)\Phi\left(\frac{\lambda\eta_i}{\sqrt{\sigma_b^2+\lambda^2\Lambda_i}}\right),$$

that is, $\mathbf{Y}_i|u_i,\boldsymbol{\theta} \sim SN_{n_i}(\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i,\boldsymbol{\lambda}_i)$, with $\boldsymbol{\mu}_i=\mathbf{x}_i^\top\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_i = u_i^{-1}\left[\sigma_e^2\mathbb{I}_{n_i}+\sigma_b^2\mathbf{Z}_i\mathbf{Z}_i^\top\right]$ and

$$\boldsymbol{\lambda}_i = \frac{\lambda u_i^{-1/2}\Lambda_i^*}{\sigma_e^2\sqrt{\sigma_b^2+\lambda^2\Lambda_i}}\left[\sigma_e^2\mathbb{I}_{n_i}+\sigma_b^2\mathbf{Z}_i\mathbf{Z}_i^\top\right]^{1/2}\mathbf{Z}_i.$$

So, the expected values above are obtained via numerical integration based on the conditional distribution of $U_i|\mathbf{y}_i$.

Appendix B

Approximated information matrix

The PDF of $\mathbf{y}_i$ in (13) involves an integral, which makes it hard to compute the expected information matrix ($\mathbf{I}(\boldsymbol{\theta})$). Thus, we consider the same strategy used by Lin (2010) by obtaining an approximation to $\mathbf{I}(\boldsymbol{\theta})$ through the following expression:

$$\mathbf{I}_e\left(\hat{\boldsymbol{\theta}}|\mathbf{y}\right) = \sum_{i=1}^{n} \hat{\mathbf{s}}_i \hat{\mathbf{s}}_i^\top \tag{B1}$$

where $\mathbf{s}_i = E\left[\frac{\partial \ell_c(\boldsymbol{\theta}|\mathbf{y}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]$ is the individual score vector, $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^\top, \sigma_e^2, \sigma_b^2, \lambda\right)^\top$. The elements of $\mathbf{s}_i = \left(\mathbf{s}_i(\boldsymbol{\beta})^\top, \mathbf{s}_i(\sigma_e^2), \mathbf{s}_i(\sigma_b^2), \mathbf{s}_i(\lambda)\right)^\top$ are given by,

- $\mathbf{s}_i(\boldsymbol{\beta}) = \frac{1}{2\sigma_e^2}\mathbf{x}_i^\top\left[\hat{u}_i\left(\mathbf{y_i} - \mathbf{x}_i^\top\boldsymbol{\beta}\right) - \widehat{ub}_i\mathbf{Z}_i\right]$
- $\mathbf{s}_i(\sigma_e^2) = \frac{1}{2\sigma_e^2}\left[-n_i + \hat{u}_i\left(\mathbf{y_i} - \mathbf{x}_i^\top\boldsymbol{\beta}\right)^\top\left(\mathbf{y_i} - \mathbf{x}_i^\top\boldsymbol{\beta}\right) - 2\widehat{ub}_i\mathbf{Z}_i^\top\left(\mathbf{y_i} - \mathbf{x}_i^\top\boldsymbol{\beta}\right) + \widehat{ub^2}_i\mathbf{Z}_i^\top\mathbf{Z}_i\right]$
- $\mathbf{s}_i(\sigma_b^2) = \frac{1}{2\sigma_b^2}\left[\widehat{ub^2}_i + \widehat{t^2}_i - 2\lambda\widehat{tb}_i + \lambda^2\widehat{b^2}_i - 2\right]$
- $\mathbf{s}_i(\lambda) = \frac{1}{2\sigma_b^2}\left[\widehat{tb}_i - \lambda\widehat{b^2}_i\right]$

ORCID

Clécio S. Ferreira http://orcid.org/0000-0002-7820-1350

References

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19 (6):716–23. doi:10.1109/TAC.1974.1100705.

Andrews, D. F., and C. L. Mallows. 1974. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society: Series B (Methodological)* 36 (1):99–102. doi:10.1111/j.2517-6161.1974.tb00989.x.

Arellano-Valle, R. B., H. Bolfarine, and V. H. Lachos. 2005. Skew-normal linear mixed models. *Journal of Data Science* 3:415–38.

Arellano-Valle, R. B., C. S. Ferreira, and M. G. Genton. 2018. Scale and shape mixtures of multivariate skew-normal distributions. *Journal of Multivariate Analysis* 166:98–110. doi:10.1016/j.jmva.2018.02.007.

Azzalini, A. 1985. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 12:171–78.

Azzalini, A., T. D. Capello, and S. Kotz. 2003. Log-skew-normal and log-skew-t distributions as models for family income data. *Journal of Income Distribution* 11:13–21.

Azzalini, A., and A. Capitanio. 1999. Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61 (3):579–602. doi:10.1111/1467-9868.00194.

Azzalini, A., and A. Dalla-Valle. 1996. The multivariate skew-normal distribution. *Biometrika* 83 (4):715–26. doi:10.1093/biomet/83.4.715.

Belsley, D. A. 1991. *Conditioning diagnostics: Collinearity and weak data in regression*. New York: Wiley.

Berkane, M., Y. Kano, and P. M. Bentler. 1994. Pseudo maximum likelihood estimation in elliptical theory: Effects of misspecification. *Computational Statistics & Data Analysis* 18 (2):255–67. doi:10.1016/0167-9473(94)90175-9.

Branco, M. D., and D. K. Dey. 2001. A general class of multivariate skew elliptical distributions. *Journal of Multivariate Analysis* 79 (1):99–113. doi:10.1006/jmva.2000.1960.

Cook, R. D. 1986. Assessment of local influence. *Journal of the Royal Statistical Society: Series B (Methodological)* 48 (2):133–69. doi:10.1111/j.2517-6161.1986.tb01398.x.

Dempster, A., N. Laird, and D. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* 39 (1):1–38. doi:10.1111/j.2517-6161.1977.tb01600.x.

Diggle, P. J., K. Y. Liang, and S. L. Zeger. 1996. *Analysis of longitudinal data*. Oxford: Clarendon Press.

Ferreira, C. S., and R. B. Arellano-Valle. 2018. Estimation and diagnostic analysis in skew-generalized-normal regression models. *Journal of Statistical Computation and Simulation* 88 (6):1039–59. doi:10.1080/00949655.2017.1419351.

Ferreira, C. S., H. Bolfarine, and V. H. Lachos. 2011. Skew scale mixtures of normal distributions: Properties and estimation. *Statistical Methodology* 8 (2):154–71. doi:10.1016/j.stamet.2010.09.001.

Ferreira, C. S., V. H. Lachos, and H. Bolfarine. 2016. Likelihood-based inference for multivariate skew scale mixtures of normal distributions. *AStA Advances in Statistical Analysis* 100 (4):421–41. doi:10.1007/s10182-016-0266-z.

Ferreira, C. S., and G. A. Paula. 2017. Estimation and diagnostic for skew-normal partially linear models. *Journal of Applied Statistics* 44 (16):3033–53. doi:10.1080/02664763.2016.1267124.

Gómez, H. W., O. Venegas, and H. Bolfarine. 2007. Skew-symmetric distributions generated by the normal distribution function. *Environmetrics* 18 (4):395–407. doi:10.1002/env.817.

Henderson, C. R. 1984. *Applications of linear models in animal breeding*. Guelph: University of Guelph Press.

Johnson, N. L., S. Kotz, and N. Balakrishnan. 1994. *Continuous univariate distributions*. Vol. 1. New York: Wiley.

Kahrari, F., C. S. Ferreira, and R. B. Arellano-Valle. 2019. Skew-normal-cauchy linear mixed models. *Sankhya B* 81 (2):185–202. doi:10.1007/s13571-018-0173-2.

Lachos, V. H. 2004. Modelos lineares mistos assimétricos. PhD thesis, Departamento de Estatística, IME-USP.

Lachos, V. H., and H. Bolfarine. 2007. Skew probit error-in-variables models. *Statistical Methodology* 3:1–12.

Lachos, V. H., H. Bolfarine, R. B. Arellano-Valle, and L. C. Montenegro. 2007. Likelihood based inference for multivariate skew-normal regression models. *Communications in Statistics - Theory and Methods* 36 (9): 1769–86. doi:10.1080/03610920601126241.

Lachos, V. H., P. Ghosh, and R. B. Arellano-Valle. 2010. Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica* 20:303–22.

Laird, N. M., and J. H. Ware. 1982. Random effects models for longitudinal data. Biometrics 38 (4):963–74. doi:10.2307/2529876.

Lange, K. L., R. Little, and J. Taylor. 1989. Robust statistical modeling using t distribution. *Journal of the American Statistical Association* 84 (408):881–96. doi:10.2307/2290063.

Lange, K. L., and J. S. Sinsheimer. 1993. Normal/independent distributions and their applications in robust regression. Journal of Computational and Graphical Statistics 2 (2):175–98. doi:10.2307/1390698.

Lee, S. Y., and L. Xu. 2004. Influence analysis of nonlinear mixed-effects models. *Computational Statistics & Data Analysis* 45 (2):321–41. doi:10.1016/S0167-9473(02)00303-1.

Lee, Y., and J. A. Nelder. 2004. Conditional and marginal model: Another view. *Statistical Science* 19 (2):219–38. doi:10.1214/088342304000000305.

Lesaffre, E., and G. Verbeke. 1998. Local influence in linear mixed models. *Biometrics* 54 (2):570–82. doi:10.2307/3109764.

Lin, T. I. 2010. Robust mixture modeling using multivariate skew t distributions. *Statistics and Computing* 20 (3): 343–56. doi:10.1007/s11222-009-9128-9.

Lin, T. I., and J. C. Lee. 2008. Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. *Statistics in Medicine* 27 (9):1490–507. doi:10.1002/sim.3026.

Little, R. J. A. 1988. Robust estimation of the mean and covariance matrix from data with missing values. *Applied Statistics* 37 (1):23–38. doi:10.2307/2347491.

Lucas, A. 1997. Robustness of the student t based m-estimator. *Communications in Statistics - Theory and Methods* 26 (5):1165–82. doi:10.1080/03610929708831974.

Ma, Y., M. G. Genton, and M. Davidian. 2004. Linear mixed effects models with flexible generalized skew-elliptical random effects. In *Skew-elliptical distributions and their applications: A journey beyond normality*, ed. M. G. Genton, 339–58. Boca Raton: Chapman & Hall.

McLachlan, G. J., and T. Krishnan. 2008. *The EM algorithm and extensions*. 2nd ed. Hoboken: Wiley.

Schumacher, F. L., L. A. Matos, and V. H. Lachos. 2020. Skewlmm: Scale Mixture of Skew-Normal Linear Mixed Models. R package version 0.2.2. https://CRAN.R-project.org/package=skewlmm.

Yu, B., A. J. O'Malley, and P. Ghosh. 2014. Linear mixed models for multiple outcomes using extended multivariate skew-t distributions. *Statistics and Its Interface* 7 (1):101–11. doi:10.4310/SII.2014.v7.n1.a11.

Zeller, C. B., F. V. Labra, V. H. Lachos, and N. Balakrishnan. 2010. Influence analyses of skew-normal/independent linear mixed models. *Computational Statistics & Data Analysis* 54 (5):1266–80. doi:10.1016/j.csda.2009.11.008.

Zeller, C. B., V. H. Lachos, and F. V. Labra. 2011. Local influence analysis for regression models with scale mixtures of skew-normal distributions. *Journal of Applied Statistics* 38 (2):343–68. doi:10.1080/02664760903406504.

Zhu, H., and S. Lee. 2001. Local influence for incomplete-data models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63 (1):111–26. doi:10.1111/1467-9868.00279.