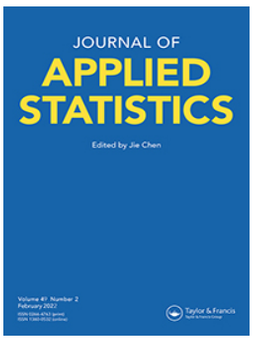

Journal of Applied Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

Fast inference for robust nonlinear mixed-effects models

José Clelto Barros Gomes, Reiko Aoki, Victor Hugo Lachos, Gilberto Alvarenga Paula & Cibele Maria Russo

To cite this article: José Clelto Barros Gomes, Reiko Aoki, Victor Hugo Lachos, Gilberto Alvarenga Paula & Cibele Maria Russo (2022): Fast inference for robust nonlinear mixed-effects models, Journal of Applied Statistics, DOI: 10.1080/02664763.2022.2034141

To link to this article: https://doi.org/10.1080/02664763.2022.2034141

JOURNAL OF APPLIED STATISTICS
https://doi.org/10.1080/02664763.2022.2034141

Taylor & Francis
Taylor & Francis Group

Fast inference for robust nonlinear mixed-effects models

José Clelto Barros Gomes [a], Reiko Aoki [b], Victor Hugo Lachos [c], Gilberto Alvarenga Paula [d] and Cibele Maria Russo [b]

[a]Department of Statistics, Federal University of Amazonas, Manaus, Brazil; [b]Department of Applied Mathematics and Statistics, Institute of Mathematics and Computer Science, University of São Paulo, São Carlos, Brazil; [c]Department of Statistics, University of Connecticut, Storrs, CT, USA; [d]Department of Statistics, Institute of Mathematics and Statistics, University of São Paulo, São Paulo, Brazil

ABSTRACT

The interest for nonlinear mixed-effects models comes from application areas as pharmacokinetics, growth curves and HIV viral dynamics. However, the modeling procedure usually leads to many difficulties, as the inclusion of random effects, the estimation process and the model sensitivity to atypical or nonnormal data. The scale mixture of normal distributions include heavy-tailed models, as the Student-*t*, slash and contaminated normal distributions, and provide competitive alternatives to the usual models, enabling the obtention of robust estimates against outlying observations. Our proposal is to compare two estimation methods in nonlinear mixed-effects models where the random components follow a multivariate scale mixture of normal distributions. For this purpose, a Monte Carlo expectation-maximization algorithm (MCEM) and an efficient likelihood-based approximate method are developed. Results show that the approximate method is much faster and enables a fairly efficient likelihood maximization, although a slightly larger bias may be produced, especially for the fixed-effects parameters. A discussion on the robustness aspects of the proposed models are also provided. Two real nonlinear applications are discussed and a brief simulation study is presented.

ARTICLE HISTORY
Received 8 December 2018
Accepted 19 January 2022

KEYWORDS
Nonlinear mixed-effects models; scale mixture of normal distributions; Monte Carlo EM; likelihood-based approximation; computational efficiency

1. Introduction

Nonlinear mixed-effects models (NLMEMs) provide a number of possibilities to dealing with correlated data, such as growth curves or pharmacokinetic data, for example. The most common assumption for the distribution of errors and random effects in NLMEMs is the normality (see, for instance, [8,12,30]), which may not be the most appropriate choice in cases of heavy-tailed data or in the presence of outliers. Moreover, it is extensively discussed in the literature that, even under the normality assumption, the estimation process in NLMEM may require additional tools as numerical integration, Monte Carlo methods or linear approximations based on Taylor expansion or Laplace approximations (see, for instance, [30]).

CONTACT Cibele Maria Russo cibele@icmc.usp.br

Recently, heavy-tailed distributions have been used to avoid the disproportional influence of outlying observations or heavy-tailed data, and some results in linear mixed-effects models are presented, for example, in Savalli *et al.* [23] and Osorio *et al.* [17]. However, it is not straightforward to extend the developed theory for nonlinear mixed-effects models, mainly because it requires numerical integration to obtain the marginal model. One possibility to avoid this problem is to add the random effects to the nonlinear model in an adequate linear form, which allows for obtaining the marginal model directly by applying properties from elliptical distributions (see, for example, [22]).

An alternative is to consider a subclass of the elliptical distributions, known as the scale mixture of normal distributions, which covers important heavy-tailed elliptical families, such as the multivariate Student-*t* (MSt), multivariate slash (MSl), multivariate contaminated normal (MCN), among others. A likelihood-based approach was recently presented for linear and nonlinear mixed-effects models for censored data using Student-*t* distributions (see [15]). A relevant discussion is presented in Meza *et al.* [16], with a proposal of using a stochastic approximation of the EM algorithm in nonlinear mixed-effects models. However, an unanswered question still remains about how to compare the estimation under exact- or approximate-based likelihood, since the approximate method has been applied in many papers (see, for instance, [9]). Another important point refers to the interpretation of the random effects in nonlinear models, rarely discussed in the literature, but which is crucial in these models.

Aiming to elucidate these points, we propose the comparison of two estimation methods in nonlinear mixed-effects models with a scale mixture of normal distributions supposed for the random effects and errors. The first method is a Monte Carlo EM method (MCEM), based on the exact likelihood, and the second one is an approximate method based on iterative approximations to a linear mixed-effects model. This comparison was considered under normality by Wu [30] and approximate inferences for nonlinear mixed-effects models with scale mixtures of skew-normal distributions were addressed recently by Schumacher *et al.* [24]. We provide a discussion on the inclusion and interpretation of random effects in nonlinear mixed-effects models and discuss the results from a simulation study, which showed that the approximate method may be a fairly efficient alternative, though it adds a larger bias in some situations.

The article is organized as follows. Two real-data motivating examples are described in Section 2. Nonlinear mixed-effects models with a scale mixture of normal distributions are discussed in Section 3 and the two estimation methods are proposed in Section 4. In Section 5 the methodology is illustrated with the two motivating examples, and a brief analysis about the robustness aspects of the proposed models is presented. In Section 6 we present a compact Monte Carlo simulation study to compare the two methodologies. Finally, the results are discussed in Section 7.

2. Motivation

In this section, two motivating examples of nonlinear correlated data are presented, as well as a brief discussion about the inclusion of random effects in nonlinear problems.

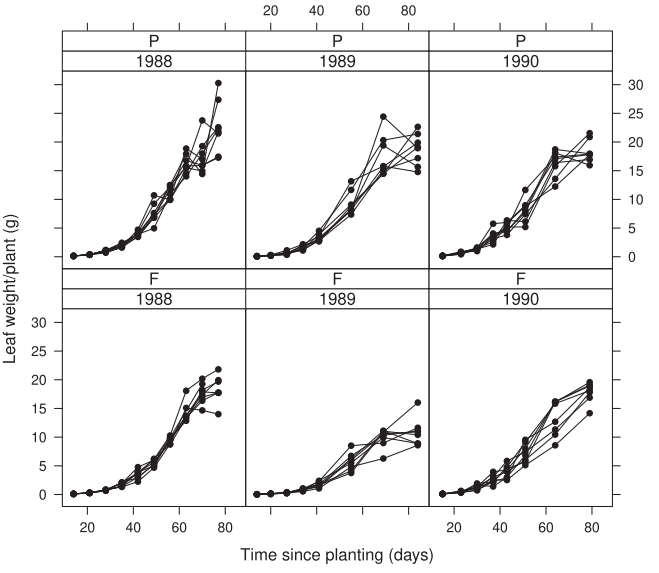

Figure 1. Leaf weight versus the time for the soybean plants data set.

2.1. Growth curve problem

The three-parameter logistic model is frequently used to model growth curve data where the mean of a response variate Y is related to a covariate T (frequently the time) according to the nonlinear function g as follows:

$$\mathrm{E}(Y) = g(\beta_1, \beta_2, \beta_3, T) = \frac{\beta_1}{1 + \exp\{-[T - \beta_2]/\beta_3\}}.$$

One example is the growth of soybean plants (see, for instance, [19]), where Y is the average leaf weight per plant (in g) and T is the time after planting (in days). The parameters β_1, β_2 and β_3 have physical interpretations according to the response variable, in the example where Y represents the leaf weight, the parameters β_1, β_2 and β_3 represent the asymptotic leaf weight, the time at which the leaf reaches half of its asymptotic weight and the time elapsed between the leaf reaching half and $1/(1 + e^{-1}) \approx 3/4$ of its asymptotic weight, respectively. The observed data set is presented in Figure 1, where the points indicate the measurements and line segments illustrate subsequent measurements taken in the same plot. This data set is available in the computational package R under the name Soybean {nlme}. Further details on the data set will be discussed in Subsection 5.1.

2.2. Pharmacokinetic problem

Another nonlinear problem considered here is the pharmacokinetics of a substance in the body. This type of problem involves the absorption and elimination of a substance, and it is usual to model the mean concentration of the substance, Y, by using the nonlinear

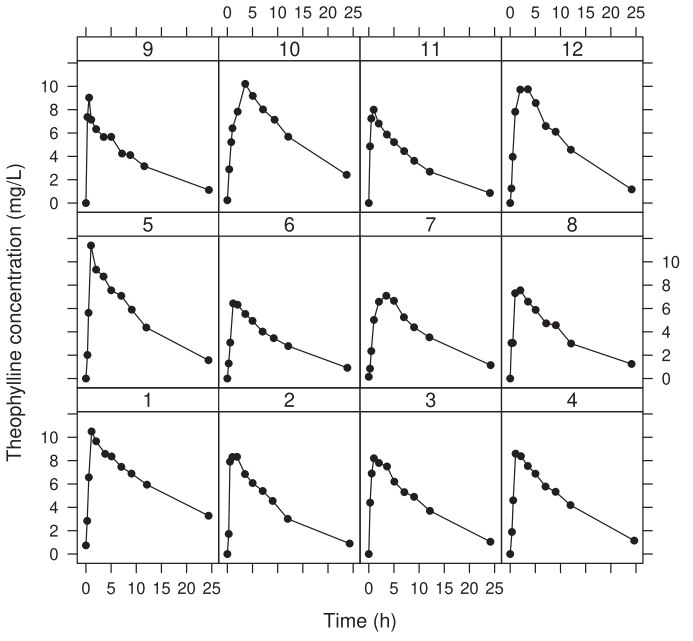

Figure 2. Theophylline dataset.

function g of time, T, and dose, D, as follows

$$\mathrm{E}(Y) = g(lK_\mathrm{e}, lK_\mathrm{a}, lC_\mathrm{l}, T, D) = D\exp(lK_\mathrm{e} + lK_\mathrm{a} - lC_\mathrm{l})\frac{[\exp(-\mathrm{e}^{lK_\mathrm{e}}T) - \exp(-\mathrm{e}^{lK_\mathrm{a}}T)]}{\mathrm{e}^{lK_\mathrm{a}} - \mathrm{e}^{lK_\mathrm{e}}}.$$

An example is described by Pinheiro and Bates [19] on the anti-asthmatic agent theophylline, where the serum concentration of the substance, Y (in mg/L), was measured at eleven times (in h) after administering D dose (in mg/kg) in each of the 12 patients. The nonlinearity of the data can be observed in Figure 2. This dataset is available in R under the name *Theoph* {*datasets*}.

It is usual to call this nonlinear model a first-order compartment model, with the following interpretation for the parameters: lK_a represents the logarithm of the substance absorption rate, lK_e is the logarithm of the substance elimination rate and lC_l represents the logarithm of plasma clearance. For more details, refer to Section 5.2.

2.3. Random effects interpretation in nonlinear mixed-effects models

One of the purposes of including random effects to a regression model is to enable different fitted models to distinct experimental units. In linear mixed-effects models, random effects are usually related to the intercept and slope of the regression lines. Furthermore, the inclusion of random effects in nonlinear models provide markedly different fitted curves, according to the distinct interpretation of each parameter. Moreover, it is worth analyzing the result of each random effect included in the model, since including a large number of random effects may increase significantly the computational cost without bringing much benefits.

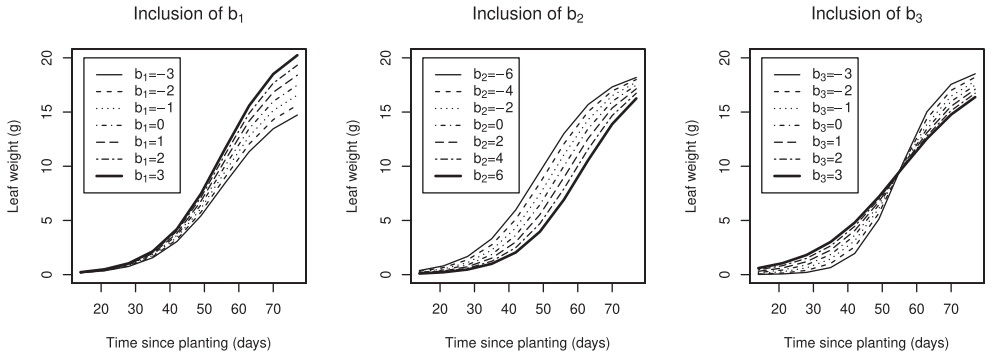

Figure 3. Result of the random effects in soybean plants hypothetical example.

2.3.1. Three parameters logistic model

For an illustration, let us suppose that different random effects would be included to a three-parameter logistic model for modeling the leaf weight (g) with time after planting (days). Let us consider a theoretical model with three random effects, given by

$$Y = \frac{\beta_1 + b_1}{1 + \exp\{-[T - (\beta_2 + b_2)]/(\beta_3 + b_3)\}}$$

and assume the values $\boldsymbol{\beta} = (19, 55, 9)^\top$, $\boldsymbol{T} = (14, 21, 28, 35, 42, 49, 56, 63, 70, 77)^\top$. An illustration of the effect of individually including the random effects b_1, b_2 and b_3 can be observed in Figure 3.

2.3.2. First-order one compartment model

For the theoretical first-order compartment model, consider the inclusion of different random effects as it follows

$$Y = D\exp[(lK_e + b_1) + (lK_a + b_2) - (lC_l + b_3)]\frac{\{\exp(-e^{(lK_e+b_1)}T) - \exp[-e^{(lK_a+b_2)}T]\}}{e^{(lK_a+b_2)} - e^{(lK_e+b_1)}}.$$

and assume the values $(lK_e, lK_a, lC_l)' = (-2.4, 0.4, -3)^\top$, $\boldsymbol{T} = (0, 1, 2, \ldots, 25)^\top$. An illustration of the effect of individually including the random effects b_1, b_2 and b_3 can be observed in Figure 4. As it can be seen in the graph, including random effects in each fixed-effect parameter enables the model to take into account different variability patterns for the absorption and elimination of the substance in the body.

3. Nonlinear mixed-effects models

In this section, we discuss briefly some properties of the scale mixture of normal distributions and then propose the nonlinear mixed-effects models with the appropriate assumptions.

3.1. Scale mixture of normal distributions (SMN)

The symmetrical class of SMN distributions is known by their robustness properties and it includes important distributions as Student-*t*, slash and contaminated normal (see, for

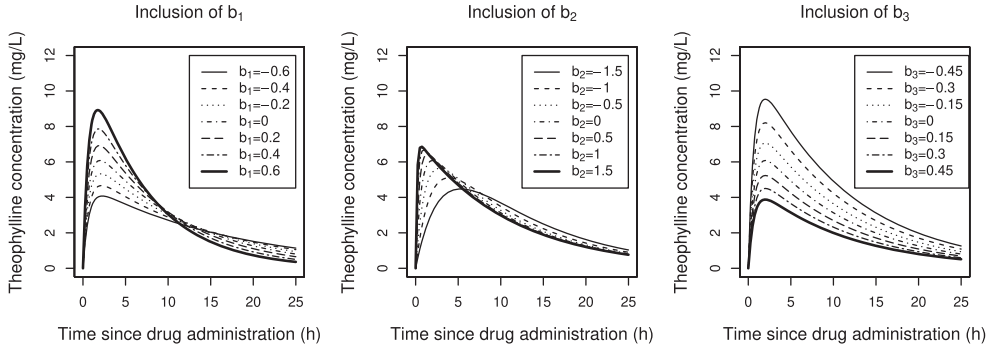

Figure 4. Result of the random effects in first order, one compartment model.

example, [1,10]). It is frequently advantageous to represent an $m-$dimensional random vector $\mathbf{Y}$ following a SMN distribution in its stochastic form

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

where $\boldsymbol{\mu}$ is the location vector, U is a positive random variable with cumulative distribution function (cdf) $H(\boldsymbol{\nu})$ and probability density function (pdf) $h(\boldsymbol{\nu})$, where $\boldsymbol{\nu}$ is a scalar or vector parameter indexing the distribution of U, $\kappa(U)$ is the weight function, $\mathbf{Z} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ with $\mathbf{Z}$ and U independent. Given $U = u$, $\mathbf{Y}$ follows a multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance-covariance $\kappa(u)\boldsymbol{\Sigma}$. In other words, the SMN distribution is a scale mixture of normal distributions, where the distribution of the scale factor U is the mixing distribution. The marginal pdf of $\mathbf{Y}$ may be written as

$$f(\mathbf{y}) = \int_0^\infty \phi_m(\mathbf{y} \mid \boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma})\, \mathrm{d}H(\boldsymbol{\nu}), \tag{1}$$

where $\phi_m(\cdot \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for the probability density function of the m-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. We will use the notation $\mathbf{Y} \sim \mathrm{SMN}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$.

3.2. Robust nonlinear mixed-effects models

Suppose that $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$ is a vector of observed continuous multivariate responses with $\mathbf{y}_i$ a $(n_i \times 1)$ vector containing the observations for the experimental unit i, $i = 1, \ldots, n$, such that

$$\begin{aligned} \mathbf{y}_i &= \mathbf{g}(\boldsymbol{\varphi}_i, \mathbf{X}_i) + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n, \\ \boldsymbol{\varphi}_i &= \mathbf{A}_i\boldsymbol{\beta} + \mathbf{b}_i, \end{aligned} \tag{2}$$

in which $\mathbf{X}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{in_i})^\top$ is a matrix of explanatory variables for the ith unit, $\mathbf{b}_i$ is a $(q \times 1)$ vector of random effects, $\boldsymbol{\epsilon}_i$ is an $(n_i \times 1)$ vector of random errors for $i = 1, \ldots, n$, $\boldsymbol{\beta}$ is a $(p \times 1)$ location vector and $\mathbf{A}_i$ is a full rank $(q \times p)$ matrix of known constants. This nonlinear model was considered by Lee and Xu [11], for instance, under normality. In this

Table 1. Characterization of some SMN distributions.

Distribution	$\kappa(u)$	U	Density function $f(\mathbf{y})$
$\mathrm{MSt}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$	$\frac{1}{u}$	Gamma $\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$,	$\frac{\Gamma(\frac{m+\nu}{2})}{\Gamma(\frac{\nu}{2})\pi^{m/2}} \nu^{-m/2} \lvert\Sigma\rvert^{-1/2} \left(1+\frac{d}{\nu}\right)^{-\frac{m+\nu}{2}}$
		$u > 0, \nu > 0$	
$\mathrm{MSl}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$	$\frac{1}{u}$	Beta$(\nu, 1)$,	$\nu \int_0^1 u^{\nu-1} \phi_m(\mathbf{y} \mid \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})\, \mathrm{d}u$
		$0 < u < 1, \nu > 0$	

with $d = (\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})$.

paper, we will assume that

$$\begin{pmatrix} \boldsymbol{\epsilon}_i \\ \mathbf{b}_i \end{pmatrix} \overset{\text{ind.}}{\sim} \mathrm{SMN}_{n_i+q}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{pmatrix}; H\right), \tag{3}$$

where $\mathbf{D}$ and $\boldsymbol{\Sigma}_i$ are positive-definite dispersion matrices. We assume that $\mathbf{D} = \mathbf{D}(\boldsymbol{\tau}) = \mathrm{diag}(\boldsymbol{\tau})$ is a diagonal matrix and denote its elements by $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_q)^\top$. The matrix $\boldsymbol{\Sigma}_i$ with dimension $(n_i \times n_i)$ is typically dependent upon i through its dimension, and it will be considered, for example, $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}_{n_i}$ for $i = 1, \ldots, n$ and $\sigma > 0$ a scalar. Since $\mathbf{A}_i$ and $\mathbf{X}_i$ are known matrices, we will simplify the notation by writing $\mathbf{g}(\boldsymbol{\beta}, \mathbf{b}_i)$ to represent $\mathbf{g}(\boldsymbol{\varphi}_i, \mathbf{X}_i) = \mathbf{g}(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{b}_i, \mathbf{X}_i)$. Finally, as it was indicated in the previous section, $H = H(\boldsymbol{\nu})$ is the cdf generator that determines the specific SMN model that was assumed.

Remarks: (i) From (3), it follows that marginally

$$\boldsymbol{\epsilon}_i \overset{\text{ind.}}{\sim} \mathrm{SMN}_{n_i}(\mathbf{0}, \boldsymbol{\Sigma}_i; H) \text{ and } \mathbf{b}_i \overset{\text{iid.}}{\sim} \mathrm{SMN}_q(\mathbf{0}, \mathbf{D}; H) \quad i = 1, \ldots, n. \tag{4}$$

(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. The independence corresponds to the case where $U_i = 1$, $i = 1, \ldots, n$, so that the SMN–NLMEM reduces to the normal NLMEM as defined in Walker [27]. 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 not correlated, once $\mathrm{Cov}(\mathbf{b}_i, \boldsymbol{\epsilon}_i) = \mathrm{E}[\mathbf{b}_i\boldsymbol{\epsilon}_i^\top] = \mathrm{E}_{U_i}[\mathrm{E}[\mathbf{b}_i\boldsymbol{\epsilon}_i^\top \mid U_i]] = \mathbf{0}$. Therefore, an attractive and convenient way to specify (2) and (3) is the following hierarchical representation:

$$\mathbf{b}_i \mid U_i = u_i \overset{\text{ind.}}{\sim} \mathrm{N}_q(\mathbf{0}, \kappa(u_i)\mathbf{D}) \text{ and } \boldsymbol{\epsilon}_i \mid U_i = u_i \overset{\text{ind.}}{\sim} \mathrm{N}_{n_i}(\mathbf{0}, \kappa(u_i)\boldsymbol{\Sigma}_i), \quad i = 1, \ldots, n,$$

$\mathbf{b}_i \mid U_i = u_i$ and $\boldsymbol{\epsilon}_i \mid U_i = u_i$ are independent, where $U_i \overset{\text{iid}}{\sim} h(\boldsymbol{\nu})$, and $\kappa(\cdot)$ is the weight function, $i = 1, \ldots, n$.

(iii) Aiming to choose between the different fitted models, we use the Akaike information criterion (AIC), which also provide an alternative to select the parameter $\boldsymbol{\nu}$ from the scale mixture of normal distributions. To obtain an approximation for the log-likelihood, the importance sampling method is considered, following the suggestion of Meza *et al.* [16]. Further discussion about fixing or estimating the extra parameters may be found in the literature (for more details, see [14]), taking into account

the possible sensitivity added by unbounded behavior of the influence and change-of-variance functions of the location parameter. However, the methodology proposed here could be easily adapted for the case in which $\boldsymbol{\nu}$ is estimated.

(iv) The hierarchical representation (three-stage) to the NLMEM defined in (2) and (3) is given by

$$\mathbf{y}_i \mid \mathbf{b}_i, U_i = u_i \overset{\text{ind.}}{\sim} \mathrm{N}_{n_i}(\mathbf{g}(\boldsymbol{\beta}, \mathbf{b}_i), \kappa(u_i)\boldsymbol{\Sigma}_i), \tag{5}$$

$$\mathbf{b}_i \mid U_i = u_i \overset{\text{ind.}}{\sim} \mathrm{N}_q(\mathbf{0}, \kappa(u_i)\mathbf{D}), \text{and} \tag{6}$$

$$U_i \overset{\text{iid.}}{\sim} h(\boldsymbol{\nu}). \tag{7}$$

Classical inference of the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\tau}^\top, \sigma^2)^\top$ is based on the marginal distribution of $\mathbf{y}_i$, particularly by the frequentist approach, and the maximum likelihood estimates (MLEs) of the parameters can be obtained from the joint distribution

$$f(\mathbf{y}_1, \ldots, \mathbf{y}_n) = \prod_{i=1}^{n} \int\int \phi_{n_i}(\mathbf{y}_i \mid \mathbf{g}(\boldsymbol{\beta}, \mathbf{b}_i), \kappa(u_i)\boldsymbol{\Sigma}_i)\phi_q(\mathbf{b}_i \mid \mathbf{0}, \kappa(u_i)\mathbf{D})\, \mathrm{d}\mathbf{b}_i\, \mathrm{d}H(\boldsymbol{\nu}), \tag{8}$$

which generally does not have a closed form expression because the model function is not linear in the random effect. In the next section, we propose a Monte Carlo EM algorithm that facilitates the likelihood inference and also an approximate method based on iterative approximations to the linear mixed-effects model.

4. Maximum likelihood estimation

4.1. Monte Carlo EM method

Initially proposed by Dempster *et al.* [3], the EM algorithm represents an efficient tool to obtain the maximum likelihood estimates in problems with incomplete data, for instance. By augmenting the observed data with a 'missing' quantity, it is worth using this iterative procedure when the maximization of the complete data likelihood is easier than the original data. Basically, the process consists of repeating the steps of expectation and maximization of the complete data likelihood until the convergence is achieved. These steps are known in the literature as the E-step and M-step, respectively. Since it is not always straightforward to obtain the E-step expressions explicitly, additional tools may be required to estimate the expected values numerically. In particular, the Monte Carlo EM algorithm (MCEM) has been used for mixed-effects models by treating the random effects as latent data (see, for instance, [27,28]) and it will be used here to obtain the maximum likelihood estimates for the NLMEM parameters. Modified versions of the EM algorithm for NLMEM are presented in the literature, as for instance the SEM or SAEM algorithms [8]. However, previous papers, for instance [16], have shown that the results considering these different methods are very close to each other, therefore in this paper we only consider the MCEM method.

Considering the model defined in (2)–(3), the unobserved random effect $\mathbf{b}_i$ and the scale factor U_i are considered as missing data, so that the 'complete data' is given by

$\{(\mathbf{y}_i, \mathbf{b}_i, U_i), i = 1, \ldots, n\}$. Therefore, the complete–data log-likelihood for all individuals can be written as

$$\begin{aligned}\ell_c(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \ell(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{b}_i, u_i) \\ &= \sum_{i=1}^{n} \{\log f(\mathbf{y}_i|\mathbf{b}_i, u_i, \boldsymbol{\beta}, \sigma^2) + \log f(\mathbf{b}_i|u_i, \boldsymbol{\tau}) + \log h(\boldsymbol{\nu})\},\end{aligned}$$

where $\boldsymbol{\tau}$ is the parameter vector of the scale matrix $\mathbf{D}$ and $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}_{n_i}$. Let $\boldsymbol{\theta}^{(t)}$ be the parameter estimates from the tth EM iteration. The Expectation step ('E step') from individual i at the $(t + 1)$th iteration can be written as

$$\begin{aligned}Q_i(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}) &= E[\ell(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{b}_i, u_i) \mid \mathbf{y}_i, \boldsymbol{\theta}^{(t)}] \\ &= \iint \{\log \phi_{n_i}(\mathbf{y}_i \mid \mathbf{g}(\boldsymbol{\beta}, \mathbf{b}_i), \kappa(u_i)\sigma^2 \mathbf{I}_{n_i}) + \log \phi_q(\mathbf{b}_i \mid \mathbf{0}, \kappa(u_i)\mathbf{D}) + \log h(\boldsymbol{\nu})\} \\ &\quad \times f(u_i, \mathbf{b}_i \mid \mathbf{y}_i, \boldsymbol{\theta}^{(t)})\, \mathrm{d}u_i\, \mathrm{d}\mathbf{b}_i. \qquad (9)\end{aligned}$$

It is well known that the foregoing integral does not have a closed form in general and the evaluation of the integral by numerical quadrature is in general unfeasible except for simple cases. However, note that expression (9) is an expectation with respect to $f(u_i, \mathbf{b}_i \mid \mathbf{y}_i, \boldsymbol{\theta}^{(t)})$, so it may be evaluated by using the Monte Carlo EM algorithm by Wei and Tanner [28], as discussed by Wu [30]. Specifically, we may use the Gibbs sampler with a Metropolis-Hastings step (see, for instance, [7]) to generate samples from $[U_i, \mathbf{b}_i \mid \mathbf{y}_i, \boldsymbol{\theta}^{(t)}]$ by sampling from the full conditionals $[U_i \mid \mathbf{b}_i, \mathbf{y}_i, \boldsymbol{\theta}^{(t)}]$ and $[\mathbf{b}_i \mid u_i, \mathbf{y}_i, \boldsymbol{\theta}^{(t)}]$. Note that

$$\begin{aligned}f(u_i \mid \mathbf{b}_i, \mathbf{y}_i, \boldsymbol{\theta}^{(t)}) &\propto h(\boldsymbol{\nu})\phi_{n_i}(\mathbf{y}_i \mid \mathbf{g}(\boldsymbol{\beta}, \mathbf{b}_i), \kappa(u_i)\sigma^2 \mathbf{I}_{n_i})\phi_q(\mathbf{b}_i \mid \mathbf{0}, \kappa(u_i)\mathbf{D}) \\ &\propto h(\boldsymbol{\nu})\kappa^{-(n_i+q)/2}(u_i) \exp\left(-\frac{1}{2}\kappa^{-1}(u_i)\left[\mathbf{b}_i^\top \mathbf{D}^{-1}\mathbf{b}_i + \frac{\|\mathbf{y}_i - \mathbf{g}(\boldsymbol{\beta}, \mathbf{b}_i)\|^2}{\sigma^2}\right]\right)\end{aligned}$$

and

$$f(\mathbf{b}_i \mid u_i, \mathbf{y}_i, \boldsymbol{\theta}^{(t)}) \propto \phi_{n_i}(\mathbf{y}_i \mid \mathbf{g}(\boldsymbol{\beta}, \mathbf{b}_i), \kappa(u_i)\sigma^2 \mathbf{I}_{n_i})\phi_q(\mathbf{b}_i \mid \mathbf{0}, \kappa(u_i)\mathbf{D}). \qquad (10)$$

Monte Carlo samples can be generated from these full conditionals by using rejection sampling methods, in this case we will consider the Gibbs sampler (see, for instance, [5]). It can be seen that the full conditional distribution of U_i under the Student-t model follows a gamma distribution and under the slash case it follows a truncated gamma (TGamma) distribution (see Table 2), in which the notation tgamma(a, b, t) represents a random variable with Gamma(a, b) distribution with right truncation at value t. Meanwhile, the full conditional distribution of $\mathbf{b}_i$ does not have a closed form as it appears inside the nonlinear function $\mathbf{g}(\boldsymbol{\beta}, \mathbf{b}_i)$, and one alternative would be to implement the Metropolis-Hastings algorithm (see, for instance, [7]) to obtain samples from $\mathbf{b}_i$. Notice that if $\mathbf{g}(\boldsymbol{\beta}, \mathbf{b}_i)$ is linear with respect to $\mathbf{b}_i$, for instance $\mathbf{g}(\boldsymbol{\beta}, \mathbf{b}_i) = \boldsymbol{\varphi}_i = \mathbf{A}_i\boldsymbol{\beta} + \mathbf{b}_i$, then depending on the distribution of U_i the full conditional distribution of $\mathbf{b}_i$ could also be written in a closed form expression.

Table 2. Full conditional distributions of $(U_i \mid \mathbf{b}_i, \mathbf{y}_i, \boldsymbol{\theta})$.

Distribution	$f(u_i \mid \mathbf{b}_i, \mathbf{y}_i, \boldsymbol{\theta})$
$\text{MSt}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$	$\text{Gamma}\left(\frac{\nu+n_i+q}{2}, \frac{\nu+G_i}{2}\right)$, with $u_i > 0$ and $\nu > 0$
$\text{MSl}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$	$\text{TGamma}\left(\nu + \frac{n_i+q}{2}, \frac{G_i}{2}, 1\right)$, with $0 < u_i < 1$ and $\nu > 0$

with $G_i = [\mathbf{b}_i^\top \mathbf{D}^{-1}\mathbf{b}_i + \frac{\|\mathbf{y}_i - \mathbf{g}(\boldsymbol{\beta},\mathbf{b}_i)\|^2}{\sigma^2}]$.

For individual i, let $\{(\mathbf{b}_i^{(1)}, U_i^{(1)}), \ldots, (\mathbf{b}_i^{(M)}, U_i^{(M)})\}$ denote a random sample of size M generated from $[U_i, \mathbf{b}_i \mid \mathbf{y}_i, \boldsymbol{\theta}^{(t)}]$ then the E step at the $(t+1)$th EM iteration can be written as

$$
\begin{aligned}
Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}) &= \sum_{i=1}^{n} Q_i(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}) = \sum_{i=1}^{n}\left[\frac{1}{M}\sum_{j=1}^{M} \ell(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{b}_i^{(j)}, u_i^{(j)})\right] \\
&\propto \sum_{i=1}^{n}\sum_{j=1}^{M}\frac{1}{M}\left[-\frac{n_i}{2}\log\sigma^2 - \frac{\kappa^{-1}(u_i^{(j)})}{2\sigma^2}\|\mathbf{y}_i - \mathbf{g}(\boldsymbol{\beta}, \mathbf{b}_i^{(j)})\|^2\right] \\
&\quad + \sum_{i=1}^{n}\sum_{j=1}^{M}\frac{1}{M}\left[-\frac{1}{2}\log|\mathbf{D}| - \frac{\kappa^{-1}(u_i^{(j)})}{2}\mathbf{b}_i^{(j)\top}\mathbf{D}^{-1}\mathbf{b}_i^{(j)}\right].
\end{aligned}
$$

The Maximization step ('M step') of the Monte Carlo EM algorithm (MCEM) maximizes $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)})$ to produce an updated estimate $\boldsymbol{\theta}^{(t+1)}$, and therefore it is like a complete-data maximization. From $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)})$ it can be easily seen that the unique solution is given by

$$
\hat{\sigma}^{2(t+1)} = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{M}\frac{1}{M}[\kappa^{-1}(u_i^{(j)})\|\mathbf{y}_i - \mathbf{g}(\boldsymbol{\beta}^{(t)}, \mathbf{b}_i^{(j)})\|^2], \quad N = \sum_{i=1}^{n} n_i,
$$

$$
\hat{\mathbf{D}}^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{M}\frac{1}{M}\{\kappa^{-1}(u_i^{(j)})\text{diag}(\mathbf{b}_i^{(j)}\mathbf{b}_i^{(j)\top})\}.
$$

The estimation of $\boldsymbol{\beta}$ is obtained via a Newton-Raphson type algorithm inside the M step. First, the Q function is derived with respect to $\boldsymbol{\beta}$,

$$
U(\boldsymbol{\beta}) = \frac{\partial Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)})}{\partial \boldsymbol{\beta}} = \sigma^{-2}\sum_{i=1}^{n}\sum_{j=1}^{M}\frac{1}{M}\kappa^{-1}(u_i^{(j)})\mathbf{J}_i^{(t)\top}[\mathbf{y}_i - \mathbf{g}(\boldsymbol{\beta}^{(t)}, \mathbf{b}_i^{(j)})],
$$

with $\mathbf{J}_i = \partial\mathbf{g}(\boldsymbol{\beta}, \mathbf{b}_i)/\partial\boldsymbol{\beta}^\top$. Then the (negative) derivative of $U(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}^\top$ is obtained as

$$
H(\boldsymbol{\beta}) = -\frac{\partial U(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top} = \sigma^{-2}\sum_{i=1}^{n}\sum_{j=1}^{M}\frac{1}{M}\kappa^{-1}(u_i^{(j)})\{\mathbf{J}_i^{(t)\top}\mathbf{J}_i^{(t)} - \dot{\mathbf{J}}_i^{(t)}[\mathbf{y}_i - \mathbf{g}(\boldsymbol{\beta}^{(t)}, \mathbf{b}_i^{(j)})]\},
$$

where $\dot{\mathbf{J}}_i = \partial\mathbf{J}_i^\top/\partial\boldsymbol{\beta}^\top = [\partial\mathbf{J}_i^\top/\partial\beta_1; \partial\mathbf{J}_i^\top/\partial\beta_2; \ldots; \partial\mathbf{J}_i^\top/\partial\beta_p]$.

So the $\boldsymbol{\beta}$ estimates are obtained by iterating the steps

$$\hat{\boldsymbol{\beta}}^{(t+1)} = \hat{\boldsymbol{\beta}}^{(t)} + [H(\hat{\boldsymbol{\beta}}^{(t)})]^{-1} U(\hat{\boldsymbol{\beta}}^{(t)}).$$

4.2. An approximate method to MLE

In this section, we discuss using an approximate method proposed by Wu [30] in the normal case, which represents an alternative to the MCEM method and avoid some challenges found in the MCEM approach, such as the difficulty of convergence, for instance. As it is stated by Wu [30], the approximate method may involve less computational effort than the MCEM method, specially when the dimension of the random effects vector is high. The main advantages of the approximate method are that sampling the random effects in the E step is not necessary and the explicit M step expressions may be obtained trivially. On the other hand, the approximation of the model may also provide additional errors to the problem.

The most frequently approximate methods found in the literature are based on Taylor expansions of the nonlinear function or Laplace approximations. Here we consider a method which is similar to the one used by Wolfinger [29] and consists of iteratively solving the linear mixed-effects (LME) model and proceeds in the standard way to estimate the parameters.

First, we rewrite the SMN–NLMEM (2) and (3) as a single equation by combining the first two stages

$$y_{ij} = g_{ij}(\boldsymbol{\beta}, \mathbf{b}_i) + \epsilon_{ij} \quad \text{for } i = 1, \ldots, n \text{ and } j = 1, \ldots, n_i.$$

Denote the current estimates of $(\boldsymbol{\beta}, \mathbf{b}_i)$ by $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}_i)$. Taking the first-order Taylor expansion of g_{ij} around the current parameter estimate $\hat{\boldsymbol{\beta}}$ and the random effect estimates $\hat{\mathbf{b}}_i$, the approximate method consists of iteratively solving the LME response model

$$\tilde{\mathbf{y}}_i = \mathbf{W}_i \boldsymbol{\beta} + \mathbf{T}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i, \tag{11}$$

where $\tilde{\mathbf{y}}_i = \mathbf{y}_i - \mathbf{g}_i(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}_i) + \mathbf{W}_i \hat{\boldsymbol{\beta}} + \mathbf{T}_i \hat{\mathbf{b}}_i$ with $\mathbf{g}_i = (g_{i1}, \ldots, g_{in_i})^\top$, $\mathbf{W}_i = (\mathbf{W}_{i1}^\top, \ldots, \mathbf{W}_{in_i}^\top)^\top$, $\mathbf{T}_i = (\mathbf{T}_{i1}^\top, \ldots, \mathbf{T}_{in_i}^\top)^\top$ and $\tilde{\mathbf{y}}_i = (\tilde{y}_{i1}, \ldots, \tilde{y}_{in_i})^\top$, in which

$$\mathbf{W}_{ij} = \left.\frac{\partial g_{ij}(\boldsymbol{\beta}, \hat{\mathbf{b}}_i)}{\partial \boldsymbol{\beta}^\top}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \quad \text{and} \quad \mathbf{T}_{ij} = \left.\frac{\partial g_{ij}(\hat{\boldsymbol{\beta}}, \mathbf{b}_i)}{\partial \mathbf{b}_i^\top}\right|_{\mathbf{b}_i=\hat{\mathbf{b}}_i}$$
$$\text{for } i = 1, \ldots, n, \text{ and } j = 1, \ldots, n_i.$$

Note that the dimensions of $\mathbf{W}_i$ and $\mathbf{T}_i$ are $(n_i \times p)$ and $(n_i \times q)$, respectively.

Now we combine the LME response model (11) with (3). Therefore, by standard arguments of scale mixture of normal distributions and matrix algebra, it is not difficult to see that

$$\mathbf{b}_i \,|\, \tilde{\mathbf{y}}_i, \hat{\boldsymbol{\theta}}, u_i \sim \mathrm{N}_q(\tilde{\mathbf{b}}_i, \kappa(u_i)\tilde{\boldsymbol{\Sigma}}_i), \tag{12}$$

with $\tilde{\boldsymbol{\Sigma}}_i = (\hat{\mathbf{D}}^{-1} + \hat{\sigma}^{-2}\mathbf{T}_i^\top \mathbf{T}_i)^{-1}$ and $\tilde{\mathbf{b}}_i = \hat{\sigma}^{-2}\tilde{\boldsymbol{\Sigma}}_i \mathbf{T}_i^\top (\tilde{\mathbf{y}}_i - \mathbf{W}_i \hat{\boldsymbol{\beta}})$. Observe that $\tilde{\mathbf{b}}_i$ is obtained by the empirical Bayes method (see [26, Chap 7]).

Table 3. Characterization of $\tilde{u}_i = E[\kappa^{-1}(U_i) \mid \tilde{\mathbf{y}}_i, \boldsymbol{\theta}^{(t)}]$ for some distributions.

Distribution	$\tilde{u}_i$
$\text{MSt}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$	$\dfrac{\nu + n_i}{\nu + d_i}$
$\text{MSl}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$	$\dfrac{2\nu + n_i}{d_i} \dfrac{P_1(n_i/2 + \nu + 1, d_i/2)}{P_1(n_i/2 + \nu, d_i/2)}$

with $d_i = (\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})$ and $P_x(a, b) = \frac{b^a}{\Gamma(a)} \int_0^x s^{a-1} \mathrm{e}^{-bs}\, \mathrm{d}s$ is the cdf of a random variable with distribution Gamma(a, b).

After some algebra, we can then integrate out $\mathbf{b}_i$ and U_i from (9) and obtain the following E step

$$Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}) = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}),$$

where

$$\begin{aligned} Q_i(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}) = E[\ell(\boldsymbol{\theta}; \tilde{\mathbf{y}}_i, \mathbf{b}_i, u_i) \mid \tilde{\mathbf{y}}_i, \boldsymbol{\theta}^{(t)}] &\propto -\frac{n_i}{2} \log \sigma^2 \\ &- \frac{1}{2\sigma^2}[\tilde{u}_i(\tilde{\mathbf{y}}_i - \mathbf{W}_i\boldsymbol{\beta} - \mathbf{T}_i\tilde{\mathbf{b}}_i)^\top(\tilde{\mathbf{y}}_i - \mathbf{W}_i\boldsymbol{\beta} - \mathbf{T}_i\tilde{\mathbf{b}}_i) + \operatorname{tr}(\tilde{\boldsymbol{\Sigma}}_i \mathbf{T}_i^\top \mathbf{T}_i)] \\ &- \frac{1}{2}\log|\mathbf{D}| - \frac{1}{2}\operatorname{tr}((\tilde{\boldsymbol{\Sigma}}_i + \tilde{u}_i\tilde{\mathbf{b}}_i\tilde{\mathbf{b}}_i^\top)\mathbf{D}^{-1}), \end{aligned}$$

with $\tilde{u}_i = E[\kappa^{-1}(U_i) \mid \tilde{\mathbf{y}}_i, \boldsymbol{\theta}^{(t)}]$.

Finally, we can update the parameter estimates as follows:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left(\sum_{i=1}^{n} \tilde{u}_i \mathbf{W}_i^\top \mathbf{W}_i\right)^{-1} \left[\sum_{i=1}^{n} \tilde{u}_i \mathbf{W}_i^\top (\tilde{\mathbf{y}}_i - \mathbf{T}_i\tilde{\mathbf{b}}_i)\right], \\ \hat{\sigma}^2 &= \frac{1}{N}\sum_{i=1}^{n}[\tilde{u}_i(\tilde{\mathbf{y}}_i - \mathbf{W}_i\hat{\boldsymbol{\beta}} - \mathbf{T}_i\tilde{\mathbf{b}}_i)^\top(\tilde{\mathbf{y}}_i - \mathbf{W}_i\hat{\boldsymbol{\beta}} - \mathbf{T}_i\tilde{\mathbf{b}}_i) + \operatorname{tr}(\tilde{\boldsymbol{\Sigma}}_i \mathbf{T}_i^\top \mathbf{T}_i)], \\ \hat{\mathbf{D}} &= \frac{1}{n}\sum_{i=1}^{n}(\tilde{\boldsymbol{\Sigma}}_i + \tilde{u}_i\tilde{\mathbf{b}}_i\tilde{\mathbf{b}}_i^\top). \end{aligned}$$

It can be shown that $\tilde{\mathbf{y}}_i \sim \text{SMN}_{n_i}(\mathbf{W}_i\boldsymbol{\beta}, \mathbf{T}_i\mathbf{D}\mathbf{T}_i^\top + \sigma^2\mathbf{I}_{n_i}; H)$ and the values of $\tilde{u}_i = E[\kappa^{-1}(U_i) \mid \tilde{\mathbf{y}}_i, \boldsymbol{\theta}^{(t)}]$ for some distributions are given in Table 3. It is important to note in the same table that the maximum likelihood estimates bring a type of robustness to the model as the iterative processes encompass terms to control the influence of large Mahalanobis distances. Further discussion about the robustness in heavy-tailed mixed-effects models can be found in Paula *et al.* [18], Osorio *et al.* [17] and Russo *et al.* [21] for example.

4.2.1. Standard error estimates

Following the expressions developed in Louis [13] and used by Tan *et al.* [25], the observed information matrix may be written as

$$-\mathrm{E}\left\{\frac{\partial^2 \ell(\boldsymbol{\theta};\mathbf{y},\mathbf{b},\mathbf{u})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right\}\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} - \mathrm{Var}\left\{\frac{\partial \ell(\boldsymbol{\theta};\mathbf{y},\mathbf{b},\mathbf{u})}{\partial\boldsymbol{\theta}}\right\}\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \tag{13}$$

in which $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$ and $\mathbf{u} = (u_1, \ldots, u_n)^\top$ and the expectation and variance are computed with respect to $f(u_i, \mathbf{b}_i \,|\, \mathbf{y}_i, \boldsymbol{\theta}^{(t)})$. Since the expressions in Equation (13) may not be easily evaluated analytically, one alternative is to obtain estimates of these quantities by using the samples generated in the Monte Carlo method, and the standard errors are obtained from the square roots of the diagonal elements of the inverse of the estimated information matrix.

5. Application

Two real applications are discussed in this section; the first one about a growth curve problem and the second one about a pharmacokinetic problem.

5.1. Growth curve data

Considering the growth soybean data set analyzed by Pinheiro and Bates [19, Chap. 6] and Davidian and Giltinan [2, Chap. 1], it is usual to consider a mixed-effects model with the random effect in the three fixed effects parameters, which leads to the model

$$y_{ij} = \frac{\varphi_{1i}}{1 + \exp\{-[x_{ij} - \varphi_{2i}]/\varphi_{3i}\}} + \epsilon_{ij}, \quad j = 1, \ldots, n_i,\ i = 1, \ldots, n, \tag{14}$$

where $\varphi_{1i} = \beta_1 + b_{1i}$, $\varphi_{2i} = \beta_2 + b_{2i}$, $\varphi_{3i} = \beta_3 + b_{3i}$ and n_i assumes the values 8, 9 or 10 depending on the value of $i \in \{1, \ldots, n = 48\}$.

The measurements of leaf weights were taken within approximately weekly intervals after planting, over three years, 1988, 1989 and 1990, and two genotypes, P (*plant introduction*) and F (*forrest*). The observed value y_{ij} represents the jth mean weight (in g) of leafs from soybean plants in the ith plot, after t days of being planted, where for each of the 6 year-genotype combination there were 8 plots. In this case, β_1, β_2 and β_3 represent the asymptotic leaf weight, the time at which the leaf reaches half of its asymptotic weight and the time elapsed between the leaf reaching half and $1/(1 + \mathrm{e}^{-1})$ of its asymptotic weight, respectively.

The maximum likelihood estimates of the parameters obtained by the MCEM and the approximate method with standard errors considering the normal, Student-t and slash distribution are given in Table 4. Parameter ν was selected from a range of integer values, according to the lower AIC achieved. The parameter estimates and the values of the asymptotic standard errors of the parameters are close considering these two methodologies, but it is worth noting that the approximate method is much faster than the MCEM method. A brief study on the computational performance can be seen in Section 6. The fitted profiles under the Student-t model with 4 degrees of freedom for one plant randomly chosen for each combination of variety and year are presented in Figure 5. The model seems to deliver an adequate fit to the data set.

Table 4. Maximum likelihood estimates of the parameters for the soybean plants growth curve problem using the MCEM and the approximate method.

	Normal				Student-t_4				Slash_4			
	MCEM		Approximate		MCEM		Approximate		MCEM		Approximate	
	Estim.	SE	Estim.	SE	Estim.	SE	Estim.	SE	Estim.	SE	Estim.	SE
β_1	19.2160	0.2621	18.9530	0.2311	19.5070	0.1810	19.3440	0.1635	18.9880	0.2078	18.8290	0.1850
β_2	55.516	0.3805	55.1360	0.3392	55.9590	0.2625	55.6150	0.2391	55.3890	0.3083	55.0190	0.2771
β_3	8.7767	0.2420	8.5291	0.2247	8.8708	0.1699	8.6275	0.1569	8.7239	0.1930	8.4812	0.1849
σ^2	1.4055	0.1009	1.3985	0.0974	0.6360	0.0485	0.6504	0.0454	0.7004	0.0523	0.6991	0.0488
τ_1	17.2610	3.5755	16.7190	2.5303	15.2780	3.2994	15.0170	2.0979	12.9550	2.6967	12.4940	1.7537
τ_2	6.6926	1.3692	6.6606	0.5515	8.0256	1.6650	8.0636	0.7483	6.5294	1.3396	6.5989	0.6087
τ_3	0.1248	0.0255	0.1892	0.0025	0.4426	0.0904	0.4661	0.0095	0.3122	0.0637	0.3686	0.0074
Log-lik	−755.381		−755.619		−707.923		−708.698		−725.562		−726.589	
AIC	1524.763		1525.237		1429.845		1431.396		1465.124		1467.1708	

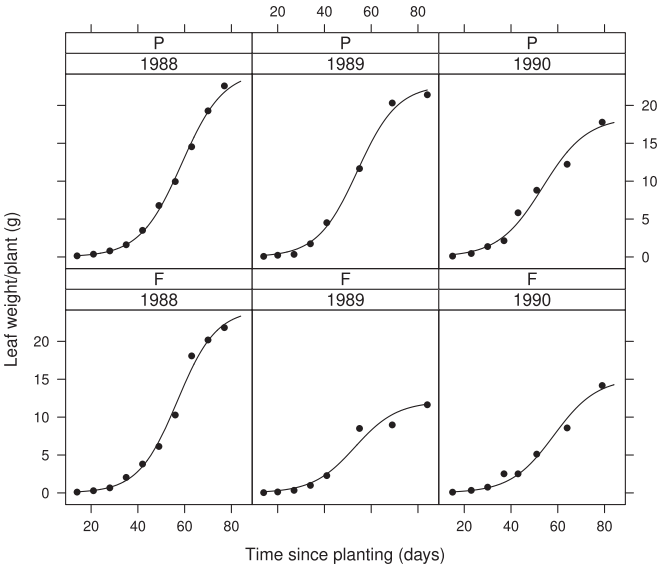

Figure 5. Fitted profiles for randomly chosen plants under the Student-*t* model with 4 degrees of freedom for the growth curves problem.

5.2. Pharmacokinetic data

In the experiment described by Pinheiro and Bates [19] on the agent theophylline, serum concentration (in mg/L) of the substance was measured at eleven times (in h) after administering D dose (in mg/kg) in each of the twelve patients. First, nonlinear mixed-effects models with three random effects were considered, namely

$$y_{ij} = D\exp(\varphi_{1i} + \varphi_{2i} - \varphi_{3i})\frac{[\exp(-e^{\varphi_{1i}}T) - \exp(-e^{\varphi_{2i}}T)]}{e^{\varphi_{2i}} - e^{\varphi_{1i}}} + \epsilon_{ij},$$

where $\varphi_{1i} = lK_e + b_{1i}$, $\varphi_{2i} = lK_a + b_{2i}$ and $\varphi_{3i} = lC_l + b_{3i}$, where lK_e, lK_a and lC_l are the fixed-effects and b_{1i}, b_{2i} and b_{3i} are the random effects.

The estimates and standard errors for the parameters in the normal selected models, Student-*t* with $\nu = 4$ and the slash with $\nu = 4$ are presented in Table 5 and the fitted profiles for the chosen model (Student-*t* with 4 degrees of freedom) are illustrated in Figure 6. Note that the model delivers an adequate fit for most of the individuals. A further investigation could be developed on the significance of the variance components, which is not straightforward in nonlinear mixed-effects models, as discussed by Russo *et al.* [21].

5.2.1. Computational aspects

For the estimation process based on MCEM, samples of size $M \geq 10{,}000$ of the full conditionals of $[U_i \mid \mathbf{b}_i, \mathbf{y}_i, \boldsymbol{\theta}^{(t)}]$ and $[\mathbf{b}_i \mid U_i, \mathbf{y}_i, \boldsymbol{\theta}^{(t)}]$ were generated by using the Monte Carlo EM method, with Metropolis Hastings within Gibbs algorithm for the E-step. Initially we made $M = 10{,}000$ and for the following iterations, we increment M with more 1000 until reaching convergence. Four parallel runs were generated in each case; the first 80% were discarded and, with a spacing of size 100, four samples with the remaining elements were used. The convergence was monitored by using the ANOVA diagnostic method proposed

Table 5. Maximum likelihood estimates of the parameters for the theophylline data set using the MCEM and approximate method.

	Normal				Student-t_4				Slash$_4$			
	MCEM		Approximate		MCEM		Approximate		MCEM		Approximate	
	Estim.	SE	Estim.	SE	Estim.	SE	Estim.	SE	Estim.	SE	Estim.	SE
lK_e	−2.4590	0.0465	−2.4546	0.0497	−2.4379	0.0364	−2.4319	0.0393	−2.4413	0.0431	−2.4383	0.0462
lK_a	0.4608	0.0534	0.4541	0.0505	0.4617	0.0444	0.4202	0.0414	0.4571	0.0519	0.4301	0.0475
lC_l	−3.2283	0.0328	−3.2246	0.0336	−3.1689	0.0263	−3.1617	0.0264	−3.1977	0.0313	−3.2000	0.0311
σ^2	0.4998	0.0632	0.5022	0.0618	0.2968	0.0400	0.2992	0.0368	0.3245	0.0432	0.3244	0.0399
τ_1	0.0004	0.0001	0.0005	< 0.0001	0.0002	< 0.0001	0.0003	< 0.0001	0.0003	0.0001	0.0003	< 0.0001
τ_2	0.4361	0.0115	0.4136	0.1582	0.4919	0.0090	0.4686	0.1827	0.3667	0.0082	0.3515	0.1359
τ_3	0.0283	0.0115	0.0281	0.0021	0.0222	0.0090	0.0223	0.0014	0.0201	0.0082	0.0200	0.0014
Log-lik	−177.859		−177.778		−170.027		−170.079		−174.739		−174.687	
AIC	369.718		369.555		354.055		354.158		363.478		363.374	

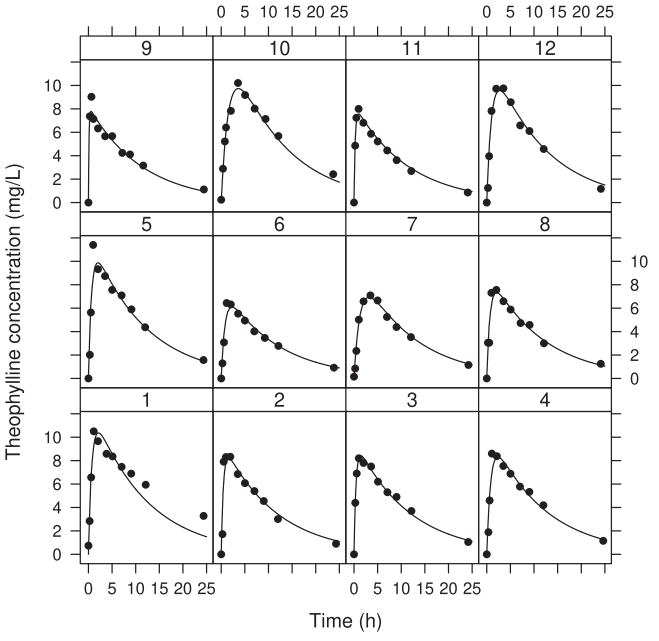

Figure 6. Fitted profiles under Student-*t* model with 4 degrees of freedom for the pharmacokinetic problem.

by Gelman and Rubin [6], observing that the estimated potential scale reduction factor $\hat{R}$ was smaller than 1.1 in all the cases.

An important discussion concerns the computational aspects of the two estimation methods. Although the Monte Carlo EM may deliver slightly better estimates to the parameters, it is significantly more expensive than the approximate method, as it requires sampling from the distribution of $[U_i, \mathbf{b}_i \mid \mathbf{y}_i, \boldsymbol{\theta}^{(t)}]$. This computational effort increases with the inclusion of more random effects, the number of individuals and depends on initial values. On the other hand, the approximate method provides sufficiently good estimates for the parameters in few seconds. For the pharmacokinetic application, for instance, the MCEM method takes around one hour (3600s) and the approximate method takes around 2s to reach convergence. Moreover, it is worth noting that the approximate method can be easily implemented. The routines were implemented in Ox [4] and run in a DELL PowerEdge 1950 server, with 2 Xeon 5430 with 2.66 GHz and 16 GB of RAM. The figures were produced using R [20].

5.2.2. Robustness aspects

It is widely discussed in the literature that heavy-tailed distributions may deliver robust estimates for the parameters (see, for instance, [17,18,21]). For models with scale-mixture of normal distributions, this robustness is due to U_i, included in the model according to Equations (5)–(7) through the function $\kappa(u_i)$. It is worth noting that $\kappa(u_i)$ appears as weights on the expressions of the maximization step for the Monte Carlo EM estimation. For the approximate method, the weights are defined by $\tilde{u}_i$ in the iterative procedure

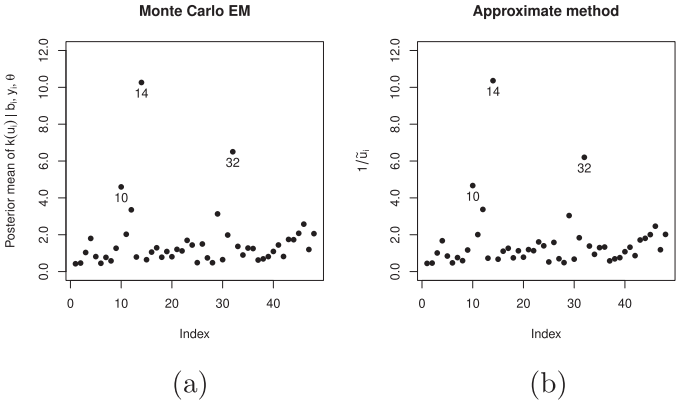

Figure 7. Posterior means of $\kappa(U_i)$ given $\mathbf{b}_i, \mathbf{y}_i, \boldsymbol{\theta}$ (MCEM) and $1/\tilde{u}_i$ (approximate method) under Student-*t* model with 4 degrees of freedom for the soybean application.

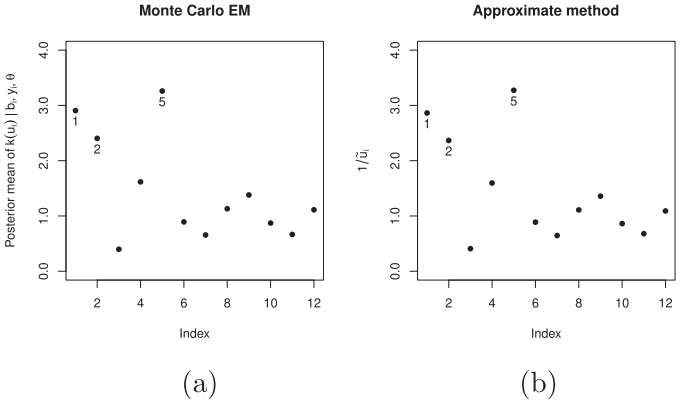

Figure 8. Posterior means of $\kappa(U_i)$ given $\mathbf{b}_i, \mathbf{y}_i, \boldsymbol{\theta}$ (MCEM) and $1/\tilde{u}_i$ (approximate method) under Student-*t* model with 4 degrees of freedom for the theophylline application.

expressions, and the robustness may be achieved due to the characterization of this quantity, according to Table 3.

For the presented models, the observations with bigger Mahalanobis distance receive the lowest weights in the iterative procedure. Thus, a possible procedure to identify outlying observations would be the graphs of the posterior mean of $\kappa(U_i) \mid \mathbf{b}_i, \mathbf{y}_i, \boldsymbol{\theta}$ for the Monte Carlo EM method and $\tilde{u}_i$ for the approximate method, as shown in Figures 7 and 8 for the growth curves and pharmacokinetic applications considering the Student-*t* with 4 degrees of freedom. Both estimation methods lead to the identification of the same observations. For the growth curves problem, observations 10, 14 and 32 are identified as outliers and for the pharmacokinetic application, individuals 1, 2 and 5 are pointed out as outlying observations.

In the growth curve application, the three experimental units identified (Figure 7) present a non-expected behavior, as it can be observed in Figure 9. Although the leaf weight

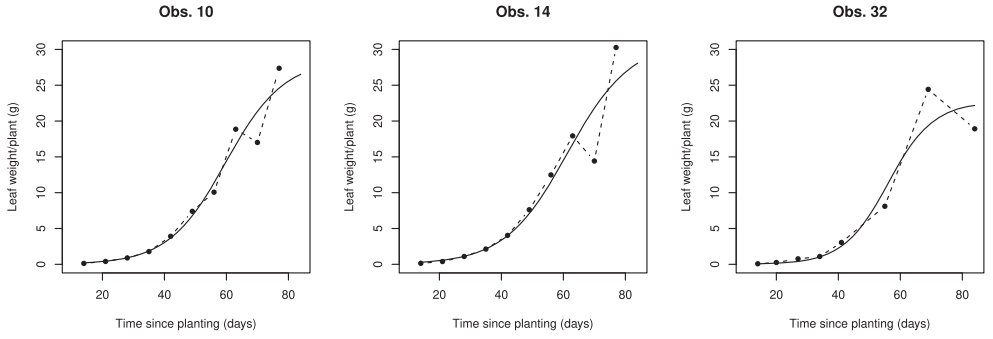

Figure 9. Leaf weight measurements and fitted profiles for the observations identified in the robustness analysis.

measurements are expected to grow between two subsequent observing times, the opposite situation occurs for some experimental units including the three identified, namely the observations 10, 14 and 32. When compared to the other data collected in 1988 from variety P, observation 10 has the smallest of the measurements and the second-largest measurement of that group. Observation 14 has the biggest of the measurements when compared to the data collected in 1988 from variety P and that measurement has a large distance of the previous point. Observation 32 has a high growth curve when compared with the data collected in 1989 from variety P. For that observation, the 7th and 8th measurements are the biggest and the smallest of the group, respectively, indicating an unexpected decrease in the mean leaf weight.

For the theophylline application, patient 1 presented a slower substance elimination than predicted by the model (see Figure 6). Moreover, the highest dose of theophylline was administered to individual 5, who also presented the highest substance concentration among all the individuals. It is important to observe that the substance doses administered were 4.02, 4.40, 4.53, 4.40, 5.86, 4.00, 4.95, 4.53, 3.10, 5.50, 4.92 and 5.30 for the twelve subjects.

The case deletion diagnostics was performed and indicated a larger variation in the estimates under a normal model than under Student-t and slash models for the fixed-effects parameters, which confirms the robustness of heavy-tailed models. For the variance components, in some cases, the Student-t and slash distributions led to larger variations.

In the next section, we present a simulation study to compare the two methodologies.

6. Simulation study

A Monte Carlo simulation study was conducted to compare the MCEM and approximate methods considering the normal, Student-t and slash distributions. In each of the scenarios considered, 2000 samples were generated according to the growth curves model or to the pharmacokinetic model and the bias and the mean squared error (MSE) were computed. For the fixed-effects parameters, theoretical values were fixed close to the maximum likelihood estimates since they provide an interpretation to the physical phenomenon in the two datasets. For the variance components, we have considered some variations, as follows.

Table 6. Simulation results for the growth curves model with theoretical fixed effects parameters $(\beta_1, \beta_2, \beta_3)^\top = (19, 55, 9)^\top$, theoretical variance components $(\boldsymbol{\tau}^\top, \sigma^2)_1^\top = (16, 6, 0.1, 1)^\top$ and Monte Carlo standard deviation (MC-SD), Monte Carlo mean of the approximate standard error obtained through the information-based method described in Section 4.2.1 (IM-SE).

		MCEM				Approximate			
		Bias	MSE	MC-SD	IM-SE	Bias	MSE	MC-SD	IM-SE
Normal	β_1	0.504	0.443	0.434	0.219	0.536	0.592	0.552	0.199
	β_2	0.012	0.271	0.520	0.328	−0.340	0.384	0.518	0.301
	β_3	0.579	0.394	0.244	0.208	0.381	0.201	0.235	0.197
	τ_1	0.055	13.368	3.657	3.322	−0.371	12.921	3.576	2.425
	τ_2	−0.037	3.764	1.940	1.222	−0.212	3.570	1.878	0.484
	τ_3	0.119	0.095	0.285	0.045	0.113	0.079	0.258	0.005
	σ^2	−0.141	0.025	0.072	0.062	−0.140	0.025	0.072	0.060
Student-t_4	β_1	0.613	0.602	0.476	0.207	0.619	0.731	0.590	0.196
	β_2	0.031	0.322	0.566	0.301	−0.384	0.468	0.567	0.297
	β_3	0.643	0.488	0.275	0.194	0.408	0.232	0.256	0.194
	τ_1	0.058	18.444	4.295	3.409	−0.479	17.229	4.124	2.339
	τ_2	0.761	6.278	2.388	1.397	0.605	5.640	2.297	0.584
	τ_3	0.187	0.167	0.364	0.059	0.129	0.097	0.284	0.005
	σ^2	−0.157	0.039	0.121	0.065	−0.155	0.039	0.122	0.059
Slash$_4$	β_1	0.694	0.747	0.516	0.242	0.729	0.934	0.635	0.218
	β_2	0.020	0.331	0.575	0.358	−0.445	0.515	0.563	0.327
	β_3	0.733	0.617	0.280	0.227	0.474	0.295	0.263	0.215
	τ_1	−0.062	13.736	3.707	3.325	−0.586	12.882	3.542	2.373
	τ_2	0.123	4.118	2.026	1.257	−0.081	3.713	1.926	0.502
	τ_3	0.150	0.124	0.319	0.051	0.122	0.087	0.269	0.005
	σ^2	−0.154	0.030	0.082	0.062	−0.152	0.030	0.081	0.059

- Growth curves model: the theoretical values for the variance components were $(\boldsymbol{\tau}^\top, \sigma^2)_1^\top = (16, 6, 0.1, 1)^\top$, $(\boldsymbol{\tau}^\top, \sigma^2)_2^\top = (10, 10, 1, 1)^\top$ and $(\boldsymbol{\tau}^\top, \sigma^2)_3^\top = (16, 6, 0.1, 5)^\top$. The theoretical fixed-effects parameters were taken as $(\beta_1, \beta_2, \beta_3)^\top = (19, 55, 9)^\top$ for the three scenarios.
- Pharmacokinetic model: the theoretical values for the variance components were $(\boldsymbol{\tau}^\top, \sigma^2)_1^\top = (0.02, 0.05, 0.5, 0.5)^\top$, $(\boldsymbol{\tau}^\top, \sigma^2)_2^\top = (0.05, 0.05, 0.05, 3)^\top$ and $(\boldsymbol{\tau}^\top, \sigma^2)_3^\top = (0.02, 0.05, 0.5, 1)^\top$. For the fixed-effects parameters, theoretical values were taken as $(lK_e, lK_a, lC_l)^\top = (-2.5, 0.5, -3)^\top$ for the three scenarios.

The results presented in Table 6 were obtained considering the growth curves model in the first scenarios. In all the scenarios, $n = 48$ and the sizes of the groups were the same as in soybean application. The normal distributions, Student-t with $\nu \in \{4, 5, 6, 7, 8\}$ and slash with $\nu \in \{4, 5, 6, 7, 8\}$ were considered and we present here the results for normal, Student-t with $\nu = 4$ and slash with $\nu = 4$. No significant differences were observed for the other cases. The obtained results in the second and third scenario, where $(\boldsymbol{\tau}^\top, \sigma^2)_2^\top = (10, 10, 1, 1)^\top$ and $(\boldsymbol{\tau}^\top, \sigma^2)_3^\top = (16, 6, 0.1, 5)^\top$, respectively, were similar to the other one and will be omitted here. In general, the methods are equivalent, but in some cases the estimates presented a slightly larger bias under the MCEM method. Regarding the MSE, in most cases the MCEM method also performed worse than the approximate method.

The results presented in Table 7 were obtained considering the pharmacokinetic model and the theoretical values $(\boldsymbol{\tau}^\top, \sigma^2)_1^\top = (0.02, 0.05, 0.5, 0.5)^\top$. In the three scenarios,

Table 7. Simulation results for the pharmacokinetic model with theoretical fixed-effects parameters $(lK_e, lK_a, lC_l)^\top = (-2.5, 0.5, -3)^\top$, theoretical variance components $(\boldsymbol{\tau}^\top, \sigma^2)_1^\top = (0.02, 0.05, 0.5, 0.5)^\top$ and Monte Carlo standard deviation (MC-SD), Monte Carlo mean of the approximate standard error obtained through the information-based method described in Section 4.2.1 (IM-SE).

		MCEM				Approximate			
		Bias	MSE	MC-SD	IM-SE	Bias	MSE	MC-SD	IM-SE
Normal	lK_e	−0.016	0.006	0.074	0.048	−0.010	0.006	0.074	0.047
	lK_a	0.003	0.009	0.093	0.054	−0.006	0.009	0.094	0.052
	lC_l	−0.012	0.003	0.056	0.033	−0.012	0.006	0.077	0.032
	τ_1	−0.009	0.000	0.019	0.005	−0.010	0.000	0.018	0.001
	τ_2	0.007	0.001	0.037	0.022	0.006	0.001	0.036	0.008
	τ_3	−0.180	0.048	0.125	0.022	−0.191	0.051	0.123	0.108
	σ^2	−0.020	0.005	0.066	0.061	−0.019	0.005	0.067	0.059
Student-t_4	lK_e	−0.027	0.007	0.082	0.047	−0.020	0.007	0.082	0.049
	lK_a	0.009	0.010	0.100	0.053	−0.002	0.010	0.101	0.054
	lC_l	−0.024	0.005	0.066	0.032	−0.026	0.008	0.088	0.034
	τ_1	−0.007	0.001	0.023	0.006	−0.008	0.001	0.022	0.001
	τ_2	0.013	0.002	0.046	0.024	0.010	0.002	0.043	0.009
	τ_3	−0.225	0.066	0.125	0.025	−0.236	0.070	0.121	0.089
	σ^2	0.010	0.019	0.138	0.070	0.011	0.020	0.140	0.063
Slash$_4$	lK_e	−0.022	0.008	0.085	0.051	−0.014	0.007	0.085	0.051
	lK_a	0.006	0.011	0.103	0.058	−0.006	0.011	0.103	0.056
	lC_l	−0.021	0.005	0.066	0.035	−0.022	0.008	0.088	0.035
	τ_1	−0.007	0.000	0.022	0.006	−0.009	0.000	0.021	0.001
	τ_2	0.010	0.002	0.039	0.023	0.008	0.002	0.038	0.009
	τ_3	−0.226	0.063	0.109	0.025	−0.236	0.067	0.106	0.089
	σ^2	−0.017	0.006	0.076	0.063	−0.015	0.006	0.076	0.060

$n = 12$ and the sizes of the groups are $n_i = 11, i = 1, \ldots, 12$, the same as in the theophylline application. The obtained results for the other cases were similar to the results in Table 7 and they will be omitted here.

As the results of the simulation study, we highlight that the MCEM method produced slightly larger bias in some of the cases but as a general result, this method does not seem to produce important deviations from the results obtained in the approximate method.

Tables 6 and 7 also show the Monte Carlo standard deviation (MC-SD) and the Information-based standard errors (IM-SE) of the estimates considering the growth curve and pharmacokinetics models, respectively, with MCEM and approximated methods. The IM-SE are mean standard errors based on the observed information matrix obtained in Section 4.2.1. For the considered cases, the MC-SD is, in general, greater than the IM-SE, and both MC-SD and IM-SE are greater for the MCEM method in comparison to the approximate method. These results are expected since MC-SD depicts the randomness involved in the Monte Carlo simulation, and a similar conclusion can be obtained when comparing MCEM to the approximate method.

In order to evaluate the computational performance of both proposed methods, a simulation study was conducted with 100 artificial replicates for the first scenario of the dataset, the growth curve problem, namely theoretical parameters of fixed effects $(\beta_1, \beta_2, \beta_3)^\top = (19, 55, 9)^\top$ and theoretical parameters of variance components $(\boldsymbol{\tau}^\top, \sigma^2)_1^\top = (16, 6, 0.1, 1)^\top$. A second scenario was considered for the pharmacokinetics model with theoretical values for fixed-effects $(lK_e, lK_a, lC_l)^\top = (-2.5, 0.5, -3)^\top$ and theoretical variance components $(\boldsymbol{\tau}^\top, \sigma^2)_1^\top = (0.02, 0.05, 0.5, 0.5)^\top$. The results are presented

Table 8. Computational average cost (in one 100th of a second) of the two methods based on 100 artificial datasets from the Theoph and Soybean scenarios.

	Theoph		Soybean	
	MCEM	Approximate	MCEM	Approximate
Normal	18572.75	0.113	22813.76	0.201
Student_4	23981.36	0.121	26844.01	0.235
Slash_4	23918.42	0.118	29625.71	0.228

in Table 8. The simulation study was conducted using Ox [4] and run in a two CPUs Linux system with 24 GB RAM each.

7. Discussion

The assumption of the scale mixture of normal distributions for the joint distribution of the random effects and errors in nonlinear mixed-effects models represent an important tool to fit nonlinear correlated data as it may provide robust estimates to the involved parameters. In this paper, we compare two approaches to obtain the maximum likelihood estimates in these models, and perform a simulation study to compare the MCEM and approximate method. It was observed that in general the approximate method may perform well when compared to the MCEM method and it is computationally efficient. Although there are no important differences in the bias of the estimates of the parameters related to the fixed effects and to the variance of the random errors, there is a significant gain in the computational time when the approximate method is applied. In conclusion, we recommend using the approximate method to reach reasonably good estimates for the parameters in nonlinear mixed-effects models, and if the researcher aims to obtain more accurate estimates, the approximate method can also provide fairly good initial values to the MCEM algorithm.

Acknowledgments

We thank Emmanuel Lesaffre for the suggestions on the manuscript. The researchers used HPC resources from the Information Technology Superintendence of the University of São Paulo to run the computational routines related to the simulation study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The authors gratefully acknowledge Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP [grant number 12/22696-4], Fundação de Amparo à Pesquisa do Estado do Amazonas - FAPEAM [grant number Edital 001/2014] and Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq, Brazil, which supported this research.

ORCID

José Clelto Barros Gomes http://orcid.org/0000-0003-4681-0116
Reiko Aoki http://orcid.org/0000-0002-4280-1258

Victor Hugo Lachos http://orcid.org/0000-0002-7239-2459
Gilberto Alvarenga Paula http://orcid.org/0000-0002-9906-9942
Cibele Maria Russo http://orcid.org/0000-0003-1356-0245

References

[1] D.R. Andrews and C.L. Mallows, *Scale mixtures of normal distributions*, J. R. Stat. Soc. 36 (1974), pp. 99–102.
[2] M. Davidian and D.M. Giltinan, *Nonlinear Models for Repeated Measurement Data*, Chapman and Hall, London, 1995.
[3] A.P. Dempster, N.M. Laird, and D.B. Rubin, *Maximun likelihood from incomplete data via the em algorithm (with discussion)*, J. R. Stat. Soc. B 39 (1977), pp. 1–38.
[4] J.A. Doornik, *An object-oriented matrix programming language ox 6*, 2009.
[5] A. Gelfand, S. Hills, A. Racine-Poon, and A. Smith, *Illustration of Bayesian inference in normal data models using Gibbs sampling*, J. Am. Stat. Assoc. 85 (1990), pp. 972–985.
[6] A. Gelman and D.B. Rubin, *Inference from iterative simulation using multiple sequences*, Stat. Sci. 7 (1992), pp. 457–472.
[7] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, *Markov Chain Monte Carlo In Practice*, Chapman and Hall, London, 1996.
[8] E. Kuhn and M. Lavielle, *Maximum likelihood estimation in nonlinear mixed effects models*, Comput. Stat. Data Anal. 49 (2005), pp. 1020–1038.
[9] V.H. Lachos, D. Bandyopadhyay, and D.K. Dey, *Linear and nonlinear mixed-effects models for censored HIV viral loads using normal/independent distributions*, Biometrics 67 (2011), pp. 1594–1604.
[10] K. Lange and J.S. Sinsheimer, *Normal/independent distributions and their applications in robust regression*, J. Comput. Graph. Stat. 2 (1993), pp. 175–198.
[11] S.Y. Lee and L. Xu, *Influence analyses of nonlinear mixed-effects models*, Comput. Stat. Data Anal. 45 (2004), pp. 321–341.
[12] M.J. Lindstrom and D.M. Bates, *Nonlinear mixed effects models for repeated measures data*, Biometrics 46 (1990), pp. 673–687.
[13] T.A. Louis, *Finding the observed information matrix when using the em algorithm*, J. R. Stat. Soc B (Methodol.) 441982), pp. 226–233.
[14] A. Lucas, *Robustness of the student-t based m-estimator*, Commun. Stat. - Theory Methods 26 (1997), pp. 1165–1182.
[15] L.A. Matos, M.O. Prates, M.H. Chen, and V.H. Lachos, *Likelihood-based inference for mixed-effects models with censored response using the multivariate-t distribution*, Stat. Sin. 23 (2013), pp. 1323–1342.
[16] C. Meza, F. Osorio, and R. De la Cruz, *Estimation in nonlinear mixed-effects models using heavy-tailed distributions*, Stat. Comput. 22 (2012), pp. 121–139.
[17] F. Osorio, G.A. Paula, and M. Galea, *Assessment of local influence in elliptical linear models with longitudinal structure*, Comput. Stat. Data Anal. 51 (2007), pp. 4354–4368.
[18] G.A. Paula, M. Medeiros, and F.E. Vilca-Labra, *Influence diagnostics for linear models with first-order autoregressive elliptical errors*, Stat. Probab. Lett. 79 (2009), pp. 339–346.
[19] J.C. Pinheiro and D.M. Bates, *Mixed-Effects Models in S and S-Plus*, Springer, 2000.
[20] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014. Available at http://www.R-project.org/.
[21] C.M. Russo, R. Aoki, and G.A. Paula, *Assessment of variance components in nonlinear mixed-effects elliptical models*, TEST (Madrid) 21 (2012), pp. 519–545.
[22] C.M. Russo, G.A. Paula, and R. Aoki, *Influence diagnostics in nonlinear mixed-effects elliptical models*, Comput. Stat. Data Anal. 53 (2009), pp. 4143–4156.
[23] C. Savalli, G.A. Paula, and F.J.A. Cysneiros, *Assessment of variance components in elliptical linear mixed models*, Stat. Modell. 6 (2006), pp. 59–76.

[24] F.L. Schumacher, D.K. Dey, and V.H. Lachos, *Approximate inferences for nonlinear mixed effects models with scale mixtures of skew-normal distributions*, J. Stat. Theory Pract. 15 (2021), pp. 1–26. doi:10.1007/s42519-021-00172-5
[25] M. Tan, G.L. Tian, and H.B. Fang, *An efficient MCEM algorithm for fitting generalized linear mixed models for correlated binary data*, J. Stat. Comput. Simul. 77 (2007), pp. 929–943.
[26] G. Verbeke and G. Molenberghs, *Linear Mixed Models for Longitudinal Data*, Springer, 2000.
[27] S. Walker, *An EM algorithm for nonlinear random effects models*, Biometrics 52 (1996), pp. 934–944.
[28] G.C.G. Wei and M.A. Tanner, *A monte carlo implementation of the em algorithm and the poor man's data augmentation algorithm*, J. Am. Stat. Assoc. 85 (1990), pp. 699–704.
[29] R. Wolfinger, *Laplace's approximation for nonlinear mixed models*, Biometrika 80 (1993), pp. 791–795.
[30] L. Wu, *Exact and approximate inferences for nonlinear mixed-effects models with missing covariates*, J. Am. Stat. Assoc. 99 (2004), pp. 700–709.