Scale mixture of skew-normal linear mixed models with within-subject serial dependence

Fernanda L. Schumacher1 | Victor H. Lachos2 | Larissa A. Matos1

1Department of Statistics, Universidade Estadual de Campinas, São Paulo, Brazil
2Department of Statistics, University of Connecticut, Storrs, Connecticut, USA

Correspondence
Larissa A. Matos, Department of Statistics, Universidade Estadual de Campinas, São Paulo, Brazil.
Email: larissam@unicamp.br

Funding information
Conselho Nacional de Desenvolvimento Científico e Tecnológico; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Grant/Award Number: 001

In longitudinal studies, repeated measures are collected over time and hence they tend to be serially correlated. These studies are commonly analyzed using linear mixed models (LMMs), and in this article we consider an extension of the skew-normal/independent LMM, where the error term has a dependence structure, such as damped exponential correlation or autoregressive correlation of order p. The proposed model provides flexibility in capturing the effects of skewness and heavy tails simultaneously when continuous repeated measures are serially correlated. For this robust model, we present an efficient EM-type algorithm for parameters estimation via maximum likelihood and the observed information matrix is derived analytically to account for standard errors. The methodology is illustrated through an application to schizophrenia data and some simulation studies. The proposed algorithm and methods are implemented in the new R package skewlmm.

KEYWORDS
autoregressive AR(p), damped exponential correlation, EM-algorithm, irregularly observed longitudinal data, linear mixed models, scale mixtures of skew-normal distributions

1 INTRODUCTION

Linear mixed models (LMM) are frequently used to analyze repeated measures data, because they model flexibly the within-subject correlation often present in this type of data. Usually for mathematical convenience, it is assumed that both random effect and error term follow normal distributions (N-LMM). These restrictive assumptions, however, may result in a lack of robustness against departures from the normal distribution and invalid statistical inferences, especially when the data show heavy tails and skewness. For instance, substantial bias in the maximum likelihood (ML) estimates of regression parameters can result when the random-effects distribution is misspecified.1,2

To deal with this problem, some proposals have been made in the literature by replacing the assumption of normality by a more flexible class of distributions. For instance, Pinheiro et al.3 proposed a multivariate t linear mixed model (T-LMM) and showed that it performed well in the presence of outliers, and Lin and Lee4 developed additional tools for the T-LMM and discussed its application to multiple sclerosis data. Rosa et al.5 adopted a Bayesian framework to carry out posterior analysis in LMM with the thick-tailed class of normal/independent distributions. Arellano-Valle et al.6 proposed a skew-normal linear mixed model (SN-LMM) based on the skew-normal (SN) distribution introduced by Azzalini and Dalla Valle,7 and Lin and Lee8 developed additional tools for the SN-LMM. Ho and Lin9 proposed a skew-t linear mixed model (ST-LMM) based on the skew-t (ST) distribution introduced by Azzalini and Capitanio.10

From a wider perspective, Lachos et al.11 proposed a parametric robust modeling of LMM based on skew-normal/independent (SNI) distributions, where random effects follow a SNI distribution and within-subject errors
follow a NI distribution, so that observed responses follow a SNI distribution, and they define what they call the 

skew-normal/independent linear mixed model. They presented an efficient EM-type algorithm for the computation of 

ML estimates of parameters on the basis of the hierarchical formulation of the SNI class. It is important to note that 

the SNI class is a subclass of the scale mixture of skew-normal (SMSN) class introduced by Branco and Dey,12 which will be 

considered in this article. More recently, Pereira and Russo13 developed asymmetric nonlinear regression models with 

mixed-effects by assuming that the random components of the model follow distributions from the SMSN class. From a 

Bayesian perspective, Maleki et al.14 considered a linear mixed effect model assuming that the random terms follow an 

unrestricted SN generalized-hyperbolic distribution, which provide flexibility for modeling complex data.

A common feature of these classes of LMMs is that the error terms are conditionally independent. However, in longitudi-

dinal studies, repeated measures are collected over time and hence the error term tends to be serially correlated. There 

are some recent proposes in the literature that account for the time dependence in longitudinal data. For instance, Chang 

and Zimmerman15 proposed to use SN antedependence models for modeling skewed longitudinal data exhibiting serial 

correlation, Asaret et al.16 proposed a methodology using multivariate normal variance-mean mixtures to fit linear mixed 

effects models for non-Gaussian continuous repeated measurement data, and Lachos et al.17 considered a robust general-

ization of the multivariate censored LMM based on the scale mixtures of normal (SMN) distributions, with a damped 
exponential correlation (DEC) structure to take into account the autocorrelation among measurements.

Nevertheless, to the best of our knowledge, there are no studies in the SMSN-LMM with serially correlated error 
structures, such as DEC,18 or autoregressive correlation of order $p$ [(AR($p$)].19 Therefore, the aim of this article is to 
develop a full likelihood approach to SMSN-LMM with serially correlated errors, considering some useful correlation 
structures (CSs). Our proposal intends to develop additional tools not considered in Lachos et al.,11 such as modeling a 
possible within-subject serial dependence and some measures for model evaluation, and to apply these techniques for 
making robust inferences in practical longitudinal data analysis. Moreover, the proposed method has been coded and imple-
mented in the $R$ package skewlmm.20 A great advantage of this package is that it offers an automatic fit of all the 
SMSN-LMM taken into consideration.

The rest of the article is organized as follows. Section 2 gives a brief introduction to SMSN class, further we define the 
SMSN-LMM and present some important dependence structures. A likelihood approach for parameter estimation is given in 
Section 3, including the estimation of random effects and standard errors (SEs). Section 4 presents some simulation 
studies that were conducted to evaluate the empirical performance of the proposed model under several scenarios and in 
Section 5 we fit the SMSN-LMM to a schizophrenia dataset. Finally, Section 6 presents some concluding remarks.

## 2 | MODEL FORMULATION

### 2.1 | Scale mixture of SN distributions

Let $Y$ be a $p \times 1$ random vector, $\mu$ a $p \times 1$ location vector, $\Sigma$ a $p \times p$ positive definite dispersion matrix, $\lambda$ a $p \times 1$ skewness parameter, and let $U$ be a positive random variable with a cumulative distribution function (cdf) $H(u; \nu)$, where $\nu$ is a scalar or parameter vector indexing the distribution of $U$. The multivariate SMSN class of distributions, denoted by 

$\text{SMSN}_p(\mu, \Sigma, \lambda; H)$, can be defined through the following probability density function (pdf):

$$
 f(y) = 2 \int_0^{\infty} \phi_p(\gamma; \mu, \lambda^T \Sigma^{-1/2} (y - \mu)) H(u; \nu), \quad y \in \mathbb{R}^p,
$$

(1)

for some positive weight function $\kappa(u)$, where $\phi_p(\cdot; \mu, \Sigma)$ denotes the pdf of the $p$-variate normal distribution with a mean vector $\mu$ and a covariance matrix $\Sigma$, $\Sigma^{-1/2}$ is such that $\Sigma^{-1/2} \Sigma^{-1/2} = \Sigma^{-1}$, and $\Phi(\cdot)$ denotes the cdf of the standard normal distribution.

A special case of the SMSN class is the skew normal distribution,7 denoted by $\text{SN}_p(\mu, \Sigma, \lambda)$, for which $H$ is degenerate at 1 (i.e., $U = 1$ with probability 1), leading to the usual pdf

$$
 f(y) = 2 \phi_p(y; \mu, \Sigma) \Phi(A), \quad y \in \mathbb{R}^p,
$$

where $A = \lambda^T \Sigma^{-1/2} (y - \mu)$. Another special case is obtained when the skewness parameter $\lambda = 0$, then the SMSN distribution in (1) reduces to the SMN distribution ($Y \sim \text{SMN}_p(\mu, \Sigma; H)$), discussed earlier by Lange and Sinsheimer.21
An important feature of this class, that can be used to derive many of its properties, is its stochastic representation. Let \( \mathbf{Y} \) be a \( p \)-dimensional random vector with pdf as in (1), then \( \mathbf{Y} \) can be represented in a stochastic way as follows:

\[
\mathbf{Y} \overset{d}{=} \mathbf{\mu} + \mathbf{\kappa}(\mathbf{U})^{1/2}\mathbf{\Sigma}^{1/2}(\mathbf{\delta}|\mathbf{T}_0| + (\mathbf{I}_p - \mathbf{\delta}\mathbf{T})^{1/2}\mathbf{T}_1), \quad \text{with} \quad \mathbf{\delta} = \frac{\lambda}{\sqrt{1 + \lambda^2}},
\]

(2)

where "\( \overset{d}{=} \)" means "equal in distribution," \( |\mathbf{T}_0| \) denotes the absolute value of \( \mathbf{T}_0 \), \( \mathbf{U} \sim \mathcal{N}(\mathbf{0}, 1) \), and \( \mathbf{T}_1 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p) \) are all independent variables, with \( \mathbf{I}_n \) being the \( n \times n \) identity matrix. The representation in (2) facilitates the implementation of EM-type algorithm. In this representation, it is straightforward that \( \mathbf{Y} \mid \mathbf{U} = \mathbf{u} \sim \mathcal{SN}_p(\mathbf{\mu}, \mathbf{\kappa}(\mathbf{\mu})\mathbf{\Sigma}) \).

Another useful feature of this class is that if \( \mathbf{Y} \sim \mathcal{SMSN}_p(\mathbf{\mu}, \mathbf{\Sigma}, \lambda, \nu) \) and \( \mathbf{X} \sim \mathcal{SN}_p(\mathbf{\mu}, \mathbf{\Sigma}, \nu) \), then for any even function \( g \), \( g(\mathbf{Y} - \mathbf{\mu}) \) has the same distribution as \( g(\mathbf{X} - \mathbf{\mu}) \) \(^{12,23} \). As a consequence, the Mahalanobis distance from the asymmetrical class \( d = (\mathbf{Y} - \mathbf{\mu})^\top \mathbf{\Sigma}^{-1}(\mathbf{Y} - \mathbf{\mu}) \) has the same distribution as the one from the symmetrical class \( (\mathbf{X} - \mathbf{\mu})^\top \mathbf{\Sigma}^{-1}(\mathbf{X} - \mathbf{\mu}) \).

For simplicity and following Lachos et al. \(^{11} \) in the remaining of this work we restrict to the case where \( \mathbf{\kappa}(\mathbf{u}) = \mathbf{u}^{-1} \). The asymmetrical class of SMSN distributions includes many distributions as special cases, and we consider explicitly the following distributions:

- **The multivariate ST distribution** with \( \nu \) degrees of freedom, \( \mathcal{ST}_p(\mathbf{\mu}, \mathbf{\Sigma}, \lambda, \nu) \), \(^{12,23} \) which can be derived from the mixture model (1) by taking \( \mathbf{U} \sim \text{Gamma}(\nu/2, \nu/2) \), with \( \nu > 0 \), and whose pdf can be written as

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

where \( t_p(\cdot; \mathbf{\mu}, \mathbf{\Sigma}, \nu) \) and \( T_p(\cdot; \nu) \) denote, respectively, the pdf of the \( p \)-variate Student-t distribution and the cdf of the standard univariate Student-t distribution with \( \nu \) degrees of freedom, and \( d = (\mathbf{Y} - \mathbf{\mu})^\top \mathbf{\Sigma}^{-1}(\mathbf{Y} - \mathbf{\mu}) \) is the Mahalanobis distance. In this case, it can be shown that \( d = (\mathbf{Y} - \mathbf{\mu})^\top \mathbf{\Sigma}^{-1}(\mathbf{Y} - \mathbf{\mu}) \sim pF_p(\mathbf{\nu}, \mathbf{\nu}) \), where \( F(a, b) \) denotes the Snedecor’s F distribution with parameters \( a \) and \( b \).

- **The multivariate skew-slash distribution**, \( \mathcal{SSL}_p(\mathbf{\mu}, \mathbf{\Sigma}, \lambda, \nu) \), that arises by taking \( \mathbf{U} \sim \text{Beta}(\nu, 1) \), with \( \nu \in [0, 1) \) and \( \nu > 0 \), and whose pdf function takes the form

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

For this distribution, the cdf of the Mahalanobis distance is \( P(d \leq r) = P\left( \chi^2_p \leq r \right) - \frac{\chi^2(p/2+1)}{\Gamma(p/2)} P\left( \chi^2_{p+2\nu} \leq r \right) \).

- **The multivariate skew-contaminated normal distribution**, \( \mathcal{SCN}_p(\mathbf{\mu}, \mathbf{\Sigma}, \lambda_1, \lambda_2) \), where \( \lambda_1, \lambda_2 \in (0, 1) \) which arises when the mixing scale factor \( \mathbf{U} \) is a discrete random variable taking one of two values and with probability function given by \( h(u|\nu) = \nu_1 I_{\{1\}}(u) + (1 - \nu_1)I_{\{1\}}(u) \), where \( \nu = (\nu_1, \nu_2) \) and \( I_{\{1\}}(u) \) is the indicator function of the set \( \nu \) whose value equals one if \( u \in \nu \) and zero elsewhere. In this case, the pdf becomes

\[
f(\mathbf{y}) = 2 \left\{ \nu_1 \phi_p(\mathbf{y}; \mathbf{\mu}, \nu_2^{-1}\mathbf{\Sigma}) \Phi(\nu_2^{1/2}A) + (1 - \nu_1) \phi_p(\mathbf{y}; \mathbf{\mu}, \mathbf{\Sigma}) \Phi(\mathbf{A}) \right\}, \quad \mathbf{y} \in \mathbb{R}^p.
\]

It is easy to see that the cdf of the Mahalanobis distance in this case is given by \( P(d \leq r) = \nu_1 P\left( \chi^2_p \leq \nu_2 r \right) + (1 - \nu_1) P\left( \chi^2_p \leq r \right) \).

We refer to Lachos et al. \(^{11} \) and Lachos et al. \(^{22} \) for details and additional properties related to this class of distributions.

### 2.2 The SMSN-LMM

Suppose that a variable of interest together with several covariates is repeatedly measured for each of \( n \) subjects at certain occasions over a period of time. For the \( i \)-th subject, \( i = 1, \ldots, n \), let \( \mathbf{Y}_i \) be a \( n_i \times 1 \) vector of observed continuous responses. In general, a normal linear mixed effects model is defined as
where $X_i$ of dimension $n_i \times l$ is the design matrix corresponding to the fixed effects, $\beta$ of dimension $l \times 1$ is a vector of population-averaged regression coefficients called fixed effects, $Z_i$ of dimension $n_i \times q$ is the design matrix corresponding to the $q \times 1$ random effects vector $b_i$, and $e_i$ of dimension $n_i \times 1$ is the vector of random errors. It is assumed that the random effects $b_i$ and the residual components $e_i$ are independent with $b_i \overset{\text{iid}}{\sim} N_q(0, D)$ and $e_i \overset{\text{iid}}{\sim} N_n(0, \Sigma_i)$. The $q \times q$ random effects covariance matrix $D$ may be unstructured or structured, and the $n_i \times n_i$ error covariance matrix $\Sigma_i$ is commonly written as $\sigma_i^2 R_i$, where $R_i$ can be a known matrix or a structured matrix depending on a vector of parameter, say $\phi$.

Likewise, the SMSN-LMM can be defined by considering

$$
\begin{pmatrix} b_i \\ e_i \end{pmatrix} \overset{\text{iid}}{\sim} \text{SMNS}_{q+n_i}(\mathbf{cA}, \mathbf{D}, \lambda; H), \quad i = 1, \ldots, n,
$$

where $c = c(\upsilon) = -\sqrt{\frac{\upsilon}{\pi}} k_1$, with $k_1 = E[U^{-1/2}]$, $\Delta = D^{1/2}\delta$, $D = D(\alpha)$ depends on unknown and reduced parameter vector $\alpha$, and we consider $\Sigma_i = \sigma_i^2 R_i$, with $R_i = R_i(\phi)$, $\phi = (\phi_1, \ldots, \phi_p)^T$, being a structured matrix. Calculating $k_1$ is straightforward and the results for the distributions discussed in Subsection 2.1 are presented in Table 1 in the Supplementary Material.

Some remarks about the model formulated in (3) and (4) are worth noting:

(i) From Lemma 1 in Appendix A of Lachos et al.\textsuperscript{11} it follows that, marginally,

$$
\begin{pmatrix} b_i \\ e_i \end{pmatrix} \overset{\text{iid}}{\sim} \text{SMNS}_{q}(\mathbf{cA}, \mathbf{D}, \lambda; H) \quad \text{and} \quad e_i \overset{\text{iid}}{\sim} \text{SMNS}_{n_i}(0, \sigma_i^2 R_i; H), \quad i = 1, \ldots, n.
$$

Thus the skewness parameter $\lambda$ incorporates asymmetry only in the distribution of the random effects (and consequently in the marginal distribution of $Y$, which is given below). In addition, as long as $k_1 < \infty$ the chosen location parameter ensures that $E[b_i] = E[e_i] = 0$, so that $E[Y_i] = X_i \beta$, for each $i = 1, \ldots, n$, and the regression parameter are all comparable. This is important since centering $b_i$‘s distribution in 0, and consequently having $E[b_i] \neq 0$, might lead to biased estimates of $\beta$, as illustrated in Section 3 in the Supplementary Material.

(ii) Even though for each $i = 1, \ldots, n$, $b_i$ and $e_i$ are indexed by the same scale mixing factor $U_i$—and hence they are not independent in general, conditional on $U_i$, we have that $b_i$ and $e_i$ are independent, what can be written as $b_i|U_i \perp e_i|U_i$. Since $\text{Cov}(b_i, e_i) = E[b_i e_i^T] = E[U_i|E[b_i e_i^T|U_i]} = 0$, $b_i$ and $e_i$ are uncorrelated (UNC). Consequently, if $k_2 = E(U^{-1}) < \infty$ we have

$$
\text{Var}(Y_i) = \text{Var}(e_i) + Z_i \text{Var}(b_i)Z_i^T = k_2 (\Sigma_i + Z_i DZ_i^T) - c^2 Z_i \Delta \Delta^T Z_i^T = Y_i.
$$

(iii) Under the SMSN-LMM at (3) and (4), for $i = 1, \ldots, n$, we have marginally

$$
Y_i \overset{\text{iid}}{\sim} \text{SMNS}_{n_i}(X_i \beta + Z_i c \Delta, \Psi_i, \lambda_i; H),
$$

where $\Psi_i = \Sigma_i + Z_i DZ_i^T$, $\lambda_i = \frac{\Psi^{-1/2} Z_i \zeta}{\sqrt{1 + \zeta^2}}$, with $\zeta = D^{-1/2}\lambda$ and $\Lambda_i = (D^{-1} + Z_i^T \Sigma_i^{-1} Z_i)^{-1}$. Hence, the marginal pdf of $Y_i$ is

$$
\begin{aligned}
f(y_i; \theta) &= 2 \int_0^{\infty} \phi_n(y_i; X_i \beta + Z_i c \Delta, u^{-1} \Psi_i) \Phi \left( u^{1/2} \lambda_i^T \Psi_i^{-1/2} (y_i - X_i \beta - Z_i c \Delta) \right) dH(u; \nu).
\end{aligned}
$$

This result can be shown using arguments from conditional probability and the moment generating function of the multivariate SN distribution, which is given in Section 2 in the Supplementary Material.

(iv) The SMSN-LMM can be written hierarchically as follows:

$$
Y_i | b_i, U_i = u_i \overset{\text{iid}}{\sim} N_{n_i}(X_i \beta + Z_i b_i, u_i^{-1} \sigma_i^2 R_i),
$$

where $\chi$ is the vector of random errors. It is assumed that the random effects $b_i$ and the residual components $e_i$ are independent with $b_i \overset{\text{iid}}{\sim} N_q(0, D)$ and $e_i \overset{\text{iid}}{\sim} N_n(0, \Sigma_i)$.
\[ \mathbf{b}_i | T_i = t_i, U_i = u_i^{\text{ind}} \sim N_q \left( \Delta t_i, u_i^{-1} \mathbf{I} \right), \]  
\[ T_i | U_i = u_i^{\text{ind}} \sim \text{TN} \left( c, u_i^{-1}, (c, \infty) \right), \]  
\[ U_i^{\text{ind}} \sim H(\cdot; \lambda), \quad i = 1, \ldots, n, \] (10)

which are all independent, where \( \Delta = \mathbf{D}^{1/2} \delta \), \( \Gamma = \mathbf{D} - \Delta \mathbf{A}^T \), with \( \delta = \sqrt{1 + \lambda^T \lambda} \) and \( \mathbf{D}^{1/2} \) is the square root of \( \mathbf{D} \), such that \( \mathbf{D}^{1/2} \mathbf{D}^{1/2} = \mathbf{D} \), containing \( q(q + 1)/2 \) distinct elements, and \( \text{TN}(\mu, \tau, (a, b)) \) denotes the univariate normal distribution \( (N(\mu, \tau)) \) truncated on the interval \((a, b)\). The hierarchical representation given in (9)–(12) is useful for the implementation of the EM algorithm as will be seen in the following section.

### 2.3 Within-subject dependence structures

In order to enable some flexibility when modeling the error covariance, we consider essentially three dependence structures: UNC, AR\((p)\) and DEC, which will be discussed next.

#### 2.3.1 Uncorrelated

The most common and simplest approach is to assume that the error terms are conditionally UNC. Under this assumption, for each \( i = 1, \ldots, n \), we have \( \mathbf{R}_i = \mathbf{I}_n \). The use of LMMs with conditionally UNC errors is very frequent in practice, for example, it has been considered by Lachos et al.\(^{11}\) in their applications, and it will be denoted as UNC-SMSN-LMM.

In longitudinal studies, however, repeated measures are collected over time and hence the error term might be serially correlated. In order to account for the within-subject serial correlation, we consider other two general structures.

#### 2.3.2 Autoregressive dependence of order \( p \)

Consider at first the case where for each subject \( i = 1, \ldots, n \) a variable of interest is observed regularly over discrete time, \( n_i \) times. Then, we propose to model \( \mathbf{R}_i \) as a structured AR\((p)\) dependence matrix.\(^{19}\) Specifically,

\[ \mathbf{R}_i = \mathbf{R}_i(\phi) = \frac{1}{1 - \phi_1 \rho_1 - \cdots - \phi_p \rho_p} [\rho_{|r-s|}], \] (13)

where \( r, s = 1, \ldots, n_i \) and \( \rho_1, \ldots, \rho_p \) are the theoretical autocorrelations of the process, and thereby they are functions of autoregressive parameters \( \phi = (\phi_1, \ldots, \phi_p)^T \), and satisfy the Yule–Walker equations,\(^{19}\) that is,

\[ \rho_k = \phi_1 \rho_{k-1} + \cdots + \phi_p \rho_{k-p}, \quad \rho_0 = 1, \quad k = 1, \ldots, p.\]

In addition, the roots of \( 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p = 0 \) must lie outside the unit circle to ensure stationarity of the AR\((p)\) model. Following Barndorff-Nielsen and Schou,\(^{24}\) the autoregressive process can be reparameterized using a one-to-one, continuous and differentiable transformation in order to simplify the conditions for stationarity. For details on the estimation of the autoregressive coefficients, we refer to Schumacher et al.\(^{25}\) and see also Lin and Lee\(^{26,27}\) for details on estimation of LMM with AR\((p)\) dependence.

The model formulated in (3) and (4) with error covariance \( \mathbf{\Sigma}_i = \sigma^2_i \mathbf{R}_i \), where \( \mathbf{R}_i \) is given by (13), \( i = 1, \ldots, n \), will be denoted AR\((p)\)-SMSN-LMM. To accommodate situations in which measurements are taken irregularly over discrete time, we modify \( \mathbf{R}_i \) by computing it for a regular range of time and then suppressing the line and column regarding the position from the missing measurements.
2.3.3 Damped exponential correlation

More generally, consider now that for each subject \(i = 1, \ldots , n\) a variable of interest is observed at times \(t_i = (t_{i1}, t_{i2}, \ldots , t_{in})\). Following Muñoz et al.,\cite{muñoz2019} we propose to structure \(R_i\) as a DEC matrix, as follows:

\[
R_i = R_i(\phi, t_i) = \begin{bmatrix} \phi_1^{t_i-t_{ik}} \\
\vdots \\
\phi_n^{t_i-t_{ik}} \end{bmatrix}, \quad 0 \leq \phi_1 < 1, \quad \phi_n \geq 0.
\]

(14)

where \(j, k = 1, \ldots , n_i\), for \(i = 1, \ldots , n\), and \(\phi = (\phi_1, \phi_2)^T\). Note that for \(\phi_2 = 1\), \(R_i\) reduces to the correlation matrix of a continuous-time autoregressive processes of order 1 (CAR(1)), hence \(\phi_2\) enables attenuation or acceleration of the exponential decay from a CAR(1) autocorrelation function (ACF), depending on its value. Moreover, \(\phi_1\) describes the autocorrelation between observations such that \(|t_{ij} - t_{ik}| = 1\). More details on DEC structure can be found in Muñoz et al.\cite{muñoz2019}

The DEC structure is rather flexible, and some particular cases are worth pointing out:

1. if \(\phi_2 = 0\), then \(R_i\) reduces to the compound symmetry CS;
2. if \(\phi_2 = 1\), then \(R_i\) reduces to the CAR(1) CS;
3. if \(0 < \phi_2 < 1\), then \(R_i\) generates a decay rate slower than the CAR(1) structure;
4. if \(\phi_2 > 1\), then \(R_i\) generates a decay rate faster than the CAR(1) structure; and
5. if \(\phi_2 \to \infty\), then \(R_i\) converges to the correlation matrix of a moving-average of order 1 (MA(1)).

The model formulated in (3) and (4) with error covariance \(\Sigma_i = \sigma^2_i R_i\), where \(R_i\) is given by (14), \(i = 1, \ldots , n\), will be denoted DEC-SMSN-LMM.

2.4 The likelihood function

The marginal pdf of \(Y_i\), \(i = 1, \ldots , n\), is given in (8), with \(R_i\) depending on the chosen CS, as described in Subsection 2.3. Hence, the log-likelihood function for \(\theta\) based on the observed sample \(y = (y_1^T, \ldots , y_n^T)^T\) is given by

\[
\ell(\theta|y) = \sum_{i=1}^n \ell_i(\theta|y_i) = \sum_{i=1}^n \log(f(y_i|\theta)),
\]

where \(\theta = (\beta^T, \sigma^2, \phi^T, \alpha^T, \lambda^T, \nu^T)^T\). As the observed log-likelihood function involves complex expressions, it is very difficult to work directly with \(\ell(\theta|y)\) to find the ML estimates of \(\theta\). Thus, in this work we propose to use an EM-type algorithm\cite{schumacher2019} for parameter estimation via ML.

3 ML ESTIMATION

3.1 The EM algorithm

A convenient feature of the SMSN-LMM is its hierarchical representation, as given in (9)–(12). Following Lachos et al.,\cite{lachos2019} \(b, u\) and \(t\) can be treated as hypothetical missing data and therefore we propose to use the ECME algorithm\cite{schumacher2019} for parameter estimation.

Let the augmented dataset be \(\hat{y}_i = (y_i^T, b_i^T, u_i^T, t_i^T)^T\), where \(y = (y_1^T, \ldots , y_n^T)^T, b = (b_1^T, \ldots , b_n^T)^T, u = (u_1, \ldots , u_n)^T\) and \(t = (t_1, \ldots , t_n)^T\). Hence, an EM-type algorithm can be applied to the complete-data log-likelihood function \(\ell_c(\theta|\hat{y}_i) = \sum_{i=1}^n \ell_i(\theta|\hat{y}_i)\), given by

\[
\ell_c(\theta|\hat{y}_i) = \sum_{i=1}^n \left[ -\frac{1}{2} \log |\Sigma_i| - \frac{\mu_i}{2\sigma^2_i} (y_i - X_i\beta - Z_i b_i)^T R_i^{-1} (y_i - X_i\beta - Z_i b_i) \\
- \frac{1}{2} \log |\Gamma| - \frac{\mu_i}{2} (b_i - \Delta_i)^T \Gamma^{-1} (b_i - \Delta_i) + K(\nu) + C. \right]
\]
where $C$ is a constant that is independent of the parameter vector $\theta$ and $K(\nu)$ is a function that depends on $\theta$ only through $\nu$.

For the current value $\theta = \hat{\theta}^{(k)}$, the E-step of the EM-type algorithm requires the evaluation of $Q^{(k)}(\theta) = E \left\{ \ell(\theta|y_i) | y, \hat{\theta}^{(k)} \right\} = \sum_{i=1}^{n} \hat{Q}^{(k)}(\theta)$, where the expectation is taken with respect to the joint conditional distribution of $b, u,$ and $t$, given $y$ and $\hat{\theta}$. Thus, we have

$$
\hat{Q}^{(k)}_i(\theta) = \hat{Q}^{(k)}_{i1}(\beta, \sigma^2_e, \phi) + \hat{Q}^{(k)}_{i2}(\alpha, \lambda) + \hat{Q}^{(k)}_{i3}(\nu),
$$

where

$$
\hat{Q}^{(k)}_{i1}(\beta, \sigma^2_e, \phi) = -\frac{n_i}{2} \log (\sigma^2_e) - \frac{1}{2} \log |R_i| - \frac{\hat{u}^{(k)}_i}{2\sigma^2_e} (y_i - X_i\beta)^\top R_i^{-1} (y_i - X_i\beta) + \frac{1}{\sigma^2_e} (y_i - X_i\beta)^\top R_i^{-1} Z_i \hat{u}^{(k)}_i - \frac{1}{2\sigma^2_e} \tr \left( R_i^{-1} Z_i \hat{u}^{(k)}_i Z_i^\top \right),
$$

$$
\hat{Q}^{(k)}_{i2}(\alpha, \lambda) = -\frac{1}{2} \log |\Gamma| - \frac{1}{2} \tr \left( \Gamma^{-1} \hat{u}^{(k)}_i \hat{u}^{(k)}_i - \frac{\hat{u}^{(k)}_i}{2} \Delta \Gamma^{-1} \hat{u}^{(k)}_i - \frac{\hat{u}^{(k)}_i}{2} \Delta \Gamma^{-1},
$$

with $\tr(A)$ and $|A|$ indicating trace and determinant of matrix $A$, respectively, $\hat{u}^{(k)}_i = E\{U_i|\hat{\theta}^{(k)}, y_i\}$, $\hat{u}^{(k)}_i = E\{U_i|\hat{\theta}^{(k)}, y_i\}$, $\hat{u}^{(k)}_i = E\{U_i|\hat{\theta}^{(k)}, y_i\}$, $\hat{u}^{(k)}_i = E\{U_i|\hat{\theta}^{(k)}, y_i\}$, and $\hat{u}^{(k)}_i = E\{U_i|\hat{\theta}^{(k)}, y_i\}$, $i = 1, \ldots, n$.

These expressions can be readily evaluated once we have the following conditional distributions, which can be derived using arguments from conditional probability:

$$
b_i|t_i, u_i, y_i, \theta \sim N_q(s_i t_i + r_i, u_i^{-1} B_i),
$$

$$
T_i|u_i, y_i, \theta \sim \text{TN} \left( \sigma + \mu_i, u_i^{-1} M_i^2; (c, \infty) \right),
$$

$$
Y_i|\theta \sim \text{SMSN}_{\eta_i}(X_i\beta + c Z_i\Delta, \Psi_i, \tilde{\lambda}_i; H),
$$

where $M_i = (1 + \Delta^\top Z_i^\top Q_i^{-1} Z_i \Delta)^{-1/2}$, $\mu_i = M_i \Delta^\top Z_i^\top Q_i^{-1} (y_i - X_i\beta - c Z_i \Delta)$, $B_i = (\Gamma^{-1} + Z_i^\top Q_i^{-1} Z_i)^{-1}$, $s_i = (I_q - B_i Z_i^\top Q_i^{-1} Z_i) \Delta$, $r_i = B_i Z_i^\top Q_i^{-1} (y_i - X_i\beta)$, $\Omega_i = \Sigma_i^\top + \Delta^\top \Sigma_i \Delta$, for $i = 1, \ldots, n$.

Thence, after some algebra and omitting the supra-index $(k)$, we get the following expressions:

$$
\hat{u}^{(k)}_i = (\hat{\mu}^{(k)} + \hat{c}^{(k)}) \hat{u}_i + \hat{M}_i \hat{r}_i, \quad \hat{u}^{(k)}_i = \hat{M}_i^2 + (\hat{\mu}^{(k)} + \hat{c}^{(k)})^2 \hat{u}_i + \hat{M}_i (\hat{u}_i + 2\hat{c} \hat{r}_i),
$$

$$
\hat{u}^{(k)}_i = \hat{r}_i \hat{u}_i + \hat{s}_i \hat{u}_i, \quad \hat{u}^{(k)}_i = \hat{r}_i \hat{u}_i + \hat{s}_i \hat{u}^{(k)}_i,
$$

$$
\hat{u}^{(k)}_i = \hat{r}_i \hat{u}_i + \hat{s}_i \hat{u}^{(k)}_i + \hat{u}^{(k)}_i (\hat{s}_i \hat{r}_i + \hat{r}_i \hat{s}_i) + \hat{u}^{(k)}_i \hat{u}^{(k)}_i, \quad (16)
$$

where $\hat{c} = c^{(k)}$, and the expressions for $\hat{u}_i$ and $\hat{r}_i$ is $E \left\{ U_i^{1/2} W \left( U_i^{1/2} A_i \right) | \hat{\theta}, y_i \right\}$, for $W(x) = \phi(x) / \Phi(x)$, $x \in \mathbb{R}$ and $A_i = \mu_i / M_i = \Delta^\top \Psi_i^{1/2} (y_i - X_i\beta - c Z_i \Delta)$, can be found in Section 2 from Lachos et al.\cite{11} which can be easily implemented for the ST and SCN distributions, but involve numerical integration for the SSL case.

The M-step requires the maximization of $\hat{Q}^{(k)}(\theta)$ with respect to $\theta$. The motivation for employing an EM-type algorithm is that it can be utilized efficiently to obtain closed-form equations for the M-step. The conditional maximization step conditionally maximize $\hat{Q}^{(k)}(\theta)$ with respect to $\theta$, obtaining a new estimate $\hat{\theta}^{(k+1)}$, as follows:

1. Update $\beta^{(k)}, \sigma^2_e^{(k)}, \phi^{(k)}, \Delta^{(k)},$ and $\hat{\Gamma}^{(k)}$ using the following expressions:

$$
\hat{\beta}^{(k+1)} = \left( \sum_{i=1}^{n} \hat{\beta}_i^{(k)} X_i^\top \Sigma_i^{-1}(X_i) \right)^{-1} \sum_{i=1}^{n} X_i^\top \Sigma_i^{-1}(X_i) \left( \hat{u}_i^{(k)} y_i - Z_i \hat{u}^{(k)}_i \right),
$$

$$
\hat{\sigma}_e^{(k+1)} = \frac{1}{N} \sum_{i=1}^{n} \left( \hat{u}_i^{(k)} (y_i - X_i \hat{\beta}^{(k+1)}_i)^\top R_i^{-1} \left( \phi^{(k)}(y_i - X_i \hat{\beta}^{(k+1)}_i) \right) \right),
$$

$$
\hat{\phi}^{(k+1)} = \frac{1}{N} \sum_{i=1}^{n} \left( \hat{u}_i^{(k)} (y_i - X_i \hat{\beta}^{(k+1)}_i)^\top R_i^{-1} \left( \phi^{(k)}(y_i - X_i \hat{\beta}^{(k+1)}_i) \right) \right),
$$

$$
\Delta^{(k+1)} = \frac{1}{N} \sum_{i=1}^{n} \left( \hat{u}_i^{(k)} (y_i - X_i \hat{\beta}^{(k+1)}_i)^\top R_i^{-1} \left( \phi^{(k)}(y_i - X_i \hat{\beta}^{(k+1)}_i) \right) \right),
$$

$$
\hat{\Gamma}^{(k+1)} = \left( \frac{1}{N} \sum_{i=1}^{n} \left( \hat{u}_i^{(k)} (y_i - X_i \hat{\beta}^{(k+1)}_i)^\top R_i^{-1} \left( \phi^{(k)}(y_i - X_i \hat{\beta}^{(k+1)}_i) \right) \right) \right)^{-1},
$$

$$
\hat{\Gamma}^{(k+1)} = \left( \frac{1}{N} \sum_{i=1}^{n} \left( \hat{u}_i^{(k)} (y_i - X_i \hat{\beta}^{(k+1)}_i)^\top R_i^{-1} \left( \phi^{(k)}(y_i - X_i \hat{\beta}^{(k+1)}_i) \right) \right) \right)^{-1}.
Estimation of random effects and prediction

In practice, the estimator of \( \tau \) is updated iteratively until a predefined criterion is reached, such as when the number of estimated parameters and the skewness vector and the parameters from the scale matrix of random effects are small enough.

Furthermore, in practical applications it is usual to include the interest in predicting \( \mathbf{y}_i \) by substituting the ML estimate \( \hat{\mathbf{b}} \) into (8). Explicit expression for \( \tau_{-i|j} = \mathbb{E}(U_{-i|j}^{1/2}W\phi(U_{-i|j}^{1/2}A_i|y_j)) \), where \( \lambda_i, \zeta \) and \( \bar{\lambda}_i \) are as in (8). The algorithm is iterated until a predefined criteria is reached, such as when \( \left| \ell'(\hat{\theta}^{(k+1)}|\mathbf{y})/\ell'(\hat{\theta}^{(k)}|\mathbf{y}) - 1 \right| \) becomes small enough.

In practice, to select between various SMSN-LMM distributions we can consider the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), given by AIC = \(-2\ell'(\hat{\theta}) + 2m\) and BIC = \(-2\ell'(\hat{\theta}) + m \log(N)\), where \( m \) is the number of estimated parameters and \( \hat{\theta} \) is the ML estimate of \( \theta \).

### 3.2 Estimation of random effects and prediction

The minimum mean-squared error (MSE) estimator of \( \mathbf{b}_i \) is obtained by the conditional mean of \( \mathbf{b}_i \) given \( \mathbf{y}_i = \mathbf{y}_i \), that can be shown to be

\[
\hat{\mathbf{b}}_i(\theta) = \mathbb{E}(\mathbf{b}_i|\mathbf{y}_i = \mathbf{y}_i, \theta) = \mathbb{E}_{U_i} [\mathbb{E}(\mathbf{b}_i|U_i, \mathbf{y}_i = \mathbf{y}_i, \theta)|\mathbf{Y}_i = \mathbf{y}_i, \theta] = \mu_{\mathbf{b}_i} + \frac{\tau_{-i|j}}{\sqrt{1 + \zeta^T \Lambda_i \zeta}} \Lambda_i \zeta.
\] (17)

where \( \mu_{\mathbf{b}_i} = c\Delta + D\Psi_i^{-1/2}y_{0i} \), with \( y_{0i} = \Psi_i^{-1/2}(\mathbf{y}_i - \mathbf{X}_i\beta - c\zeta\Delta) \), and \( \Lambda_i, \zeta \) and \( \bar{\lambda}_i \) are as in (8). In practice, the estimator of \( \mathbf{b}_i \) (also known as empirical Bayes estimator), \( \hat{\mathbf{b}}_i \), can be obtained by substituting the ML estimate \( \hat{\theta} \) into (17).

Furthermore, in practical applications it is usual to include the interest in predicting \( \mathbf{Y}_i^+ \), a future \( n_{\text{pred}} \times 1 \) vector of measurement of \( \mathbf{Y}_i \), given the observed measurement \( \mathbf{y} = (\mathbf{y}_{(i)}, \mathbf{y}_{i})^T \), where \( \mathbf{y}_{(i)} = (\mathbf{y}_{1}^T, \ldots, \mathbf{y}_{i-1}^T, \mathbf{y}_{i+1}^T, \ldots, \mathbf{y}_n^T) \). If \( \mathbf{X}_i^+ \)
and $Z_i^+$ denote $n_{\text{pred}} \times p$ and $n_{\text{pred}} \times q$ matrices of prediction regression variables corresponding to $Y_i^+$, we assume that

$$
\begin{bmatrix}
Y_i^+ \\
\Upsilon_i^+
\end{bmatrix} \sim \text{SMSN}_{n_i+n_{\text{pred}}}(X_i^+ \beta, \Psi_i^+, \lambda_i; H),
$$

where $X_i^+ = (X_i^T, X_i^+ \Psi_i^+)^T$, $Z_i^+ = (Z_i^T, Z_i^+ \Psi_i^+)^T$, $\Psi_i^+ = \Sigma_i^+ + Z_i^+ D Z_i^+ \Psi_i^+$, $\Lambda_i^+ = (D^{-1} + Z_i^+ \Sigma_i^+)^{-1}$, $\lambda_i = \Psi_i^{-1/2} \Psi_i^{1/2}$, and

$$
\Psi_i^+ = (\Psi_i^{+,11}, \Psi_i^{+,12}, \Psi_i^{+,21}, \Psi_i^{+,22}).
$$

From Lachos et al., it follows that the minimum MSE predictor of future measurements of $Y_i$ is the conditional expectation of $y_i^+$ given $Y_i = y_i$, that is,

$$
\hat{Y}_i^+(\theta) = E\{Y_i^+ | Y_i = y_i, \theta\} = \mu_{i,21} + \frac{\tilde{v}_i \Psi_{i,22,1}^{1/2} \psi_i^{1/2}}{\sqrt{1 + \psi_i^{1/2} \psi_i^{1/2}}}.
$$

where

$$
\mu_{i,21} = X_i^+ \beta + cZ_i^+ \Delta + \Psi_{i,11}^{-1}(Y_i - X_i \beta - cZ_i \Delta), \Psi_{i,22,1}^{-1} = \Psi_{i,22} - \Psi_{i,12} \Psi_{i,11}^{-1} \Psi_{i,12}, \quad \psi_i = \Psi_i^{-1/2} \lambda_i = (v_i^{(1)} , v_i^{(2)})^T,
$$

and

$$
\tilde{v}_i = E\{U_i^{-1/2} W_\Phi \left[ U_i^{-1/2} \psi_i^{1/2} (Y_i - X_i \beta - cZ_i \Delta) \right] Y_i = y_i \}, \quad \psi_i = (\psi_{i,11}^{1/2}, \psi_{i,22,1}^{1/2}).
$$

In practice, the prediction of $Y_i^+$ can be obtained by substituting the ML estimate $\hat{\theta}$ into (18), so $\hat{Y}_i^+ = \hat{Y}_i^+(\hat{\theta})$.

### 3.3 Estimation of SEs

In this section, we derive the observed information matrix from the score vector with respect to $\theta^* = \theta \setminus \nu$. First, we reparameterize $D = F^2$ for ease of computation and theoretical derivation, where $F$ is the square root of $D$ containing $q(q+1)/2$ distinct elements $\alpha_y = (a_1, \ldots, a_{q(q+1)/2})^T$. Given the observed sample $y = (y_1, \ldots, y_n)^T$ and $\nu$, the log-likelihood function for $\theta^* = (\theta_1^T, \theta_2^T)^T$, with $\theta_1 = (\beta^T, \sigma_\nu^2, \phi^T)^T$ and $\theta_2 = (\alpha_y^T, \lambda^T)^T$ is given by $\ell(\theta^*) = \sum_{i=1}^n \ell_i(\theta^*; \nu)$, where

$$
\ell_i(\theta^*; \nu) = \log 2 - \frac{n_i}{2} \log 2 \pi - \frac{1}{2} \log |\Psi_i| + \log K_i,
$$

with

$$
K_i(\theta^*; \nu) = \int_0^\infty u_i^{n_i/2} \exp \left\{ -\frac{1}{2} u_i d_i \right\} \Phi \left( u_i^{1/2} A_i \right) dH(u_i; \nu),
$$

where $d_i = (y_i - X_i \beta - cZ_i \Delta)^T \Psi_i^{-1}(y_i - X_i \beta - cZ_i \Delta)$ and $A_i = \frac{\lambda F Z_i \Psi_i^{-1} (y_i - X_i \beta - cZ_i \Delta)}{(1 + \lambda F^{-1} A_i F^{-1} \lambda)^{1/2}}$. Thus, we have after some algebraic manipulations that the score vector is given by

$$
s = \sum_{i=1}^n s_i = \sum_{i=1}^n \frac{\partial \ell_i(\theta^*; \nu)}{\partial \theta^*} = -\frac{1}{2} \sum_{i=1}^n \frac{\partial \log |\Psi_i|}{\partial \theta^*} + \sum_{i=1}^n \frac{1}{K_i} \frac{\partial K_i}{\partial \theta^*},
$$

where

$$
\frac{\partial K_i}{\partial \theta^*} = I_i^\theta \left( \frac{n_i+1}{2} \right) \frac{\partial A_i}{\partial \theta^*} - \frac{1}{2} I_i^\theta \left( \frac{n_i+1}{2} \right) \frac{\partial d_i}{\partial \theta^*},
$$

with

$$
I_i^\theta(w) = \int_0^\infty u_i^w \exp \left\{ -\frac{1}{2} u_i d_i \right\} \Phi \left( u_i^{1/2} A_i \right) dH(u_i; \nu), \quad \text{and} \quad I_i^\nu(w) = \int_0^\infty u_i^w \exp \left\{ -\frac{1}{2} u_i d_i \right\} \phi_1 \left( u_i^{1/2} A_i \right) dH(u_i; \nu).
$$

and $K_i = I_i^\theta \left( \frac{n_i+1}{2} \right)$. The results from substituting $H$ for each distribution considered are presented in Section 1 in the Supplementary Material, along with the derivatives of $\log |\Psi_i|$, $d_i$, and $A_i$, which involve tedious but not complicated algebraic manipulations.
Under some regularity conditions, asymptotic covariance of ML estimates can be estimated by the inverse of the observed information matrix, \( I(\hat{\theta}) = \sum_{i=1}^{n} \mathbf{s}_i \mathbf{s}_i^T \), where \( \mathbf{s}_i = s_i(\hat{\theta}) \) is the score vector in (20) evaluated at \( \theta^* = \hat{\theta} \).

### 3.4 Likelihood ratio test

Considering the usual interest in testing if a restricted model represents the data well enough, in this section we present a likelihood ratio test. An important particular case is testing the hypothesis that an asymmetrical model is not necessary, that could be written as \( H_0: \lambda = 0 \).

Let \( H_0: \tau = 0, \tau = (\tau_1, \ldots, \tau_r)^T \), be a hypothesis of interest, \( \Theta \) be the \( k \)-dimensional parameter space of the unrestricted model, and \( \Theta_0 \) be the parameter space under \( H_0 \), for \( 1 \leq r < k \). With the interest of measuring the impact of \( H_0 \) in the maximum of the likelihood function, consider the statistic \( \Lambda_n = 2 (\ell(\hat{\theta}) - \ell(\hat{\Theta}_0)) \), where \( \hat{\Theta}_0 \) is the ML estimate of \( \theta \) under the restriction in \( H_0 \). Then, under \( H_0 \), \( \Lambda_n \) is asymptotic distributed as a chi-square random variable with \( r \) degrees of freedom \( (\chi^2)^r \).32

### 3.5 Additional tools for model evaluation

Evaluating the suitability of a fitted model to a real dataset is an important step in data analysis and there are several methods that can be used to this purpose. When dealing with heavy-tailed data, the Mahalanobis distance is a convenient measure which can be used to identify potential outlying observations and to assess the validity of the underlying distributional assumption of the response variable, once if the fitted model is appropriate the distribution of the Mahalanobis distance is known and presented in Subsection 2.1.

Following Ho and Lin,9 to assess the goodness of fit of SMSN-LMM one can construct a Healy type plot33 by plotting the empirical ACF.19 In the context of mixed models, Pinheiro and Bates35 proposes to use the empirical ACF for within-subject errors. In the context of time series data, a commonly used tool for investigation serial correlation is the correlation matrix. If the fitted model is appropriate, the plot should resemble a straight line through the origin with unit slope.

In addition, based on Zeller et al.,34 the observed Mahalanobis distance can be decomposed as follows:

\[
d_i(\hat{\theta}) = (\mathbf{y}_i - \mathbf{X}_i \hat{\beta} - \hat{c} \mathbf{Z}_i \hat{\Delta})^T \hat{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\beta} - \hat{c} \mathbf{Z}_i \hat{\Delta}) = e_i^T \hat{\Sigma}_i^{-1} e_i + (\hat{\mu}_b - \hat{c} \hat{\Delta})^T \hat{D}^{-1} (\hat{\mu}_b - \hat{c} \hat{\Delta}) = d_{e_i}(\hat{\theta}) + d_{b_i}(\hat{\theta}),
\]  

where \( e_i = \mathbf{y}_i - \mathbf{X}_i \hat{\beta} - \hat{\mu}_b \) and \( \hat{\mu}_b \) is as in (17). This decomposition gives some insight on how the estimated random effects \( \hat{\mathbf{b}}_i \) and the estimated residuals \( e_i \) affect the overall distance.

Another important assumption that should be taken into account is the dependence structure assumed to the within-subject errors. In the context of time series data, a commonly used tool for investigation serial correlation is the empirical ACF.19 In the context of mixed models, Pinheiro and Bates35 proposes to use the empirical ACF for the residuals of a fitted LMM. Based on this approach and restricting to the case that the data is observed at discrete times, let \( \mathbf{r}_i = \hat{Y}_i^{-1/2} (\mathbf{y}_i - \mathbf{X}_i \hat{\beta}) \) be the standardized marginal residual vector for subject \( i \), where \( \mathbf{Y}_i \) is given in (6) and \( \mathbf{r}_i^T = (r_{it_1}, \ldots, r_{it_n}) \). The empirical autocorrelation at lag \( l \) can be defined as

\[
\hat{\rho}(l) = \frac{\sum_{i=1}^{n} \sum_{(j,k): j-l=k} r_{it_j} r_{it_k} / N(l)}{\sum_{i=1}^{n} \sum_{j=1}^{n} r_{it_j}^2 / N(0)},
\]

where \( N(\cdot) \) is the number of pairs used in the respective numerator summation. If the within-subject dependence structure is correct, \( \hat{\rho}(l) \) is expected to be close to zero.

Since \( \mathbf{r}_i \)'s distribution is not symmetrical, the interval estimates of \( \rho(\cdot) \) that are commonly used in time series models are not appropriate. Alternatively, we consider a Monte Carlo estimate for an UNC model, by generating \( M \) samples from a UNC-SMSN-LMM similar to the fitted model, calculating the standardized marginal residuals and \( \hat{\rho}(l) \) for each sample, and using empirical 100(\(\alpha/2\))th and 100(\(1 - \alpha/2\))th
percentiles as interval estimates of level $1 - a$. If the considered dependence structure is appropriate, we would expect approximately $100(1 - a)\%$ of the empirical autocorrelations to belong to the UNC interval.

## 4 | SIMULATION STUDIES

In the interest of investigating empirical properties of the proposed model four simulation studies were performed, and their results are presented in this section. In all simulation studies we initialized $\nu$ as follows: 10 for the ST distribution, 5 for the SSL distribution, and $(0.05, 0.8)$ for the SCN distribution. Besides, for AR($p$) dependence $\phi$ was initialized as its estimate from fitting an AR($p$)-LMM using lme function from nlme package in R, while for DEC dependence $\phi$ was initialized by finding the maximum marginal log-likelihood function as in (8) on a grid of $\phi$ and for other parameters fixed. Finally, $\beta, \sigma^2$, $D$, and $\lambda$ were initialized at the true value plus a small random error. For all studies, the bias and relative bias for estimating a parameter $\theta$ based on the $k$th sample is calculated by $\hat{\theta}(k) - \theta$ and $(\hat{\theta}(k) - \theta)/\theta$, respectively. The computational procedures were implemented using the R software, through the package skewlmm.

### 4.1 | First study

This simulation study aims to investigate asymptotic properties of the proposed model. Thence, we generated and estimated 500 Monte Carlo samples from the model

$$Y_i = (\beta_0 + b_i)\mathbf{1}_{10} + \beta_1 x_{ij} + \epsilon_i, \quad i = 1, \ldots, n,$$

where $\beta_0 = 1, \beta_1 = 2$, $\mathbf{1}_k$ is the all-ones vector of length $k$ and $x_{ij} = (x_{i1}, \ldots, x_{i10})^T$, with $x_{ij}$ being generated from the $U(0, 2)$ distribution, $i = 1, \ldots, n$ and $j = 1, \ldots, 10$, with $n$ taking values 50, 100, 200, and 350. Let $R$ be a $10 \times 10$ AR(2) dependence matrix, as given in Subsection 2.3, with $\phi_1 = 0.6$ and $\phi_2 = -0.2$. Four scenarios were considered:

(a) $b_i \sim \text{idn}(-1.0705, 2, 3)$ and $\epsilon_i \sim \text{iidn}(0, 0.25R_i);$ 
(b) $b_i \sim \text{iidn}(-1.2324, 2, 3; 6)$ and $\epsilon_i \sim \text{iidn}(0, 0.25R_i; 6);$ 
(c) $b_i \sim \text{iidn}(-1.5165, 2, 3; 1.7)$ and $\epsilon_i \sim \text{iidn}(0, 0.25R_i; 1.7);$ and 
(d) $b_i \sim \text{iidn}(-1.2915, 2, 3; 0.25, 0.3)$ and $\epsilon_i \sim \text{iidn}(0, 0.25R_i; 0.25, 0.3).$

The ML estimates and their associated SEs were recorded. In order to examine the consistency of the approximated method to get SEs described in Subsection 3.3, we computed the SD of the ML estimates obtained from the 500 Monte Carlo samples (denoted by MC-SD) and compared it with the average of the SE estimates obtained as described in Subsection 3.3 (denoted by ML-SE), for each scenario. Likewise, the average of the ML estimates will be denoted by MC-AV.

Table 1 presents results for $n = 100$ and $n = 350$. In general, the estimation method of the SEs provide results close to the empirical ones, and the closeness improves as the number of subjects increases. However, the SE approximation for the skewness parameter seems to be poor. In addition, there is a bias in the estimates of $D^{1/2}, \lambda,$ and $\nu$, but it gets smaller as $n$ increases. Furthermore, Table 2 presents coverage probability of the 95% confidence intervals for $\beta$, which are in general close to the nominal value.

The bias trend can be seem more clearly in Figures 1 and 2, which present the mean bias and $\pm 1$ SD for each distribution and parameter, by number of subjects. As the number of subjects increases, the bias (when it exists) gets closer to zero and its SD gets smaller, indicating consistency of the estimators.

### 4.2 | Second study

In order to evaluate the performance of the proposed model and the impact in estimating with the wrong distribution, we generated 500 Monte Carlo datasets from the following model:

$$Y_i = (\beta_0 + b_{i0})\mathbf{1}_{10} + (\beta_1 + b_{i11})x_{ij} + \beta_2 x_{ij} + \epsilon_i, \quad i = 1, \ldots, 100,$$
where $\beta_0 = 1, \beta_1 = 2, \beta_2 = 1.5, \mathbf{1}_k$ is the all-ones vector of length $k$, $\mathbf{x}_{1i} = (1, 2, \ldots, 10)^T$ and $\mathbf{x}_{2i} = \mathbf{w}_i \mathbf{1}_{10}$ is a group indicator taken as $\mathbf{w}_i = 0$ if $i \leq 50$ and $\mathbf{w}_i = 1$ if $i > 50$. Let $\mathbf{R}_i$ be the AR(2) dependence matrix, as given in Subsection 2.3, with $\phi_1 = 0.6$ and $\phi_2 = -0.2$ and let $\mathbf{b}_i = (b_{0i}, b_{1i})^T$. For data generation, two scenarios were considered:

(a) $\mathbf{b}_i^{\text{ind}} \sim \text{SN}_2 \left( \begin{bmatrix} -0.4718 \\ -0.7500 \end{bmatrix}, \begin{bmatrix} 2.0 & 0.2 \\ 0.2 & 1.0 \end{bmatrix} \right)$ and $\epsilon_i^{\text{ind}} \sim N_{10}(0, 0.25 \mathbf{R}_i)$; and

(b) $\mathbf{b}_i^{\text{ind}} \sim \text{ST}_2 \left( \begin{bmatrix} -0.5432 \\ -0.8635 \end{bmatrix}, \begin{bmatrix} 2.0 & 0.2 \\ 0.2 & 1.0 \end{bmatrix} \right)$ and $\epsilon_i^{\text{ind}} \sim t_{10}(0, 0.25 \mathbf{R}_i; 6)$.

Figure 3 presents the mean relative bias of $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$ for both scenarios of data generation when estimating with all the asymmetrical distributions considered in this article and the normal distribution. When the generating model is SN, all distributions except the normal seem to fit the data equally well. On the other hand, when the generating model is ST, estimating with wrong distributions seems to slightly increase variance of the estimator.

Furthermore, to evaluate the capability of the proposed selection criteria in selecting the appropriate distribution, we computed the AIC and BIC for each model and for each sample. Table 3 presents the number of times that each model was selected for both scenarios considered and the average value for each criterion. One can see that in general the criteria can classify the correct model well, and they seem to be specially good at distinguishing between the normal/SN distribution and the heavier-tailed distributions.
TABLE 2  Simulation study 1

| n  | \( \beta_0 \) | SN | ST | SSL | SCN | SN | ST | SSL | SCN |
|----|---------------|----|----|-----|-----|----|----|-----|-----|
| 50 | 96.4          | 96.4 | 96.4 | 97.2   | 95.4 | 94.0 | 96.2 | 96.8 |
| 100| 97.6          | 96.8 | 97.8 | 97.2   | 95.4 | 95.8 | 95.0 | 96.2 |
| 200| 97.4          | 98.2 | 96.0 | 97.2   | 95.8 | 93.8 | 94.4 | 94.2 |
| 350| 97.4          | 98.2 | 99.0 | 98.2   | 96.0 | 96.0 | 95.8 | 94.4 |

Note: Coverage probability (in percentages) of the 95% confidence intervals based on 500 Monte Carlo samples with different number of subjects (n).
Abbreviations: SN, skew-normal; ST, skew-t.

FIGURE 1  Simulation study 1. Mean bias and ±1 SD of estimates, according to distribution and parameter, by number of subjects (n), based on 500 Monte Carlo samples.

FIGURE 2  Simulation study 1. Mean bias and ±1 SD for estimates of \( \nu \), according to distribution, by number of subjects (n), based on 500 Monte Carlo samples.
Figure 3  Simulation study 2. Mean relative bias and ±1 SD for estimates of β₀, β₁, and β₂ when generating data from both scenarios considered—(A) SN and (B) ST—and estimating the models N, SN, ST, SSL, and SCN. SN, skew-normal; ST, skew-t.

Table 3  Simulation study 2

| Measure        | Scenario (A) - SN | Scenario (B) - ST |
|----------------|-------------------|-------------------|
|                | N     | SN     | ST    | SSL   | SCN   | N     | SN     | ST    | SSL   | SCN   |
| Selected times | 7     | 469    | 15    | 6     | 3     | 0     | 0     | 0     | 0     | 373   | 70    | 57    |
| AIC            | 176   | 323    | 1     | 0     | 0     | 0     | 0     | 0     | 0     | 411   | 79    | 10    |
| BIC            | 2256.0| 2242.5 | 2245.2| 2244.4| 2246.1| 2667.3| 2638.0| 2524.9| 2529.0| 2537.0|
| Average AIC    | 2300.0| 2296.5 | 2304.1| 2303.3| 2309.9| 2711.4| 2692.0| 2583.8| 2588.0| 2601.0|
| Average BIC    | 2304.1| 2303.3| 2309.9| 2711.4| 2692.0| 2583.8| 2588.0| 2601.0|

Note: Number of times that each distribution was selected based on each selection criterion and average value of each criterion, under both scenarios considered.
Abbreviations: AIC, Akaike information criterion; BIC, Bayesian information criterion; SN, skew-normal; ST, skew-t.

4.3  Third study

In the interest of analyzing the impact of specifying the wrong dependence structure on parameter estimates, we generated 500 Monte Carlo samples from the following LMM:

\[ Y_i = (\beta_0 + b_i)1_{n_i} + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, 100, \]

where \( \beta_0 = 1, \beta_1 = 2, 1_{n_i} \) is the all-ones vector of length \( n_i \) and \( x_i = (x_{i1}, \ldots, x_{in_i})^T \), with \( x_{ij} \) being generated from the \( U(0, 2) \) distribution, for \( i = 1, \ldots, 100, j = 1, \ldots, n_i, \) and \( n_1 = \ldots = n_{100} \) taking values 5, 10, and 15. Let \( R_i \) be an \( n_i \times n_i \) AR(2) dependence matrix, as given in Subsection 2.3, with \( \phi_1 = 0.6 \) and \( \phi_2 = -0.2 \). For this study, we considered \( b_i \) iid \( ST_1(-1.2324, 2, 3; 6) \) and \( \epsilon_i \) iid \( t_6(0, 0.25 \mathbf{R}_i; 6) \). Then, we estimated the ST-LMM by considering four covariance structures: UNC, AR(1), AR(2), and DEC.

Figure 4 presents violin plots of the relative bias of \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \) for all dependence structures considered, showing a rotated kernel density plot and some summary statistics of the relative bias. When the number of observations per subject is small (\( n_i = 5 \)), the estimates are similar for all covariance structures. As the number of observations per subject increases, the impact in considering the UNC model becomes more evident, although for the other dependence structures considered, the wrong choice of the covariance function does not seem to cause much effect on \( \hat{\beta} \). For the UNC model, the relative bias of \( \hat{\beta}_0 \) has high density above 0, indicating a bias on the intercept estimate that increases with \( n_i \), and the relative bias of \( \hat{\beta}_1 \) presents more variability for all sample sizes.

Table 4 presents the number of times that each model was selected based on AIC and BIC criteria and their average value. For \( n_i = 5 \), the BIC criterion—which gives more penalty for the number of parameters—selects the model with AR(1) dependence most of the times, but it does not select the UNC model any time. The AIC criterion, by the other hand, select
the correct model most of times for all sample sizes. In addition, as \( n_l \) increases the selection criteria can distinguish better between the covariance structures, as expected.

Once the parameter estimates of \( \phi \) and \( \sigma^2_e \) are not directly comparable between different covariance structures, we present in Table 5 the first row of the estimated within-subject variance matrix based on the average of parameter estimates obtained for all dependence structures.

## 5 Application: Schizophrenia Data

Schizophrenia is a severe psychiatric disorder characterized by delusions, hallucinations, persistent delusions and sometimes disorganized behavior and speech. In order to study the equivalency of a new antipsychotic drug for schizophrenia, Lapierre et al.\(^{37}\) presented a double-blinded clinical trial with randomization among four treatments: three doses (low, medium, and high) of a new therapy (NT) against a standard therapy (ST), for 245 patients with acute schizophrenia. Initial studies prior to this double-blinded study suggested that the experimental drug had equivalent antipsychotic activity, with less side effects. The study was conducted at 13 clinical centers, and the primary response variable was assessed using the Brief Psychiatric Rating Scale (BPRS) at baseline (week 0), and at weeks 1, 2, 3, 4, and 6 of treatment. This scale measures the extent of a total of 18 features, and rates each one on a seven point scale, with a higher number reflecting a worse evaluation. The total BPRS score is the sum of the scores on the 18 items.
| lag | 0   | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| True| 0.347| 0.174| 0.035| -0.014| -0.015| -0.006| -0.001| 0.001| 0.001| 0.000|
| UNC | 0.291| 0.000| 0.000| 0.000| 0.000| 0.000| 0.000| 0.000| 0.000| 0.000|
| AR(1)| 0.393| 0.220| 0.124| 0.069| 0.039| 0.022| 0.012| 0.007| 0.004| 0.002|
| AR(2)| 0.348| 0.174| 0.035| -0.013| -0.015| -0.006| -0.001| 0.001| 0.001| 0.000|
| DEC | 0.362| 0.189| 0.053| 0.010| 0.001| 0.000| 0.000| 0.000| 0.000| 0.000|

Note: First row of the estimated within-subject variance matrix, based on average estimate from 500 Monte Carlo samples, for different dependence structures. Abbreviations: DEC, damped exponential correlation; UNC, uncorrelated.

Figure 5 displays individual BPRS trajectories evolved over six visits along with their mean profiles for the ST and the high dose of the NT. For simplicity, we will consider only this two treatments (118 patients), but an extension for modeling the four treatments is straightforward. Ho and Lin\(^9\) showed that both subject-specific intercepts and slopes are positively skewed and that there are outliers at the level of the random effects, indicating the need of a robust model that accommodates the random effect skewness. Ho and Lin\(^9\) suggests a robust LMM using the skew t distribution, however, the article does not take into account a possible serial correlation, despite the fact that the repeated measures of each subject were collected over time.

Thus, it is of practical interest to develop a statistical model with considerable flexibility in the distributional assumptions of the random effects as well as the error terms, and that can accommodate some possible within-subject serial correlation. Following Ho and Lin\(^9\) and based on the trajectories presented in Figure 5, we propose to fit the model

\[
Y_i = (\beta_0 + b_{0i})1_{n_i} + (\beta_1 + b_{1i})x_{i1} + \beta_2 x_{i2}^2 + \beta_3 NT_i + \epsilon_i, \quad i = 1, \ldots, 118,
\]

where \(Y_i\) is the BPRS score vector divided by 10 for the \(i\)th participant, \(1_{n_i}\) is the all-ones vector of length \(n_i\), \(x_i = (x_{i1}, \ldots, x_{in_i})^\top\), with \(x_{ij}\) taken as \((\text{time} - 3)/10\) with time being measured in weeks from the baseline, \(x_{i2} = (x_{i21}, \ldots, x_{in_i})^\top\), and \(NT_i\) is an all-ones vector if the \(i\)th subject received the NT and an all-zeros vector otherwise. We fit the SMSN-LMM considering the covariance structures presented in Subsection 2.3 and preserving the last three observations from subjects with identification (ID) numbers 204 and 1608 for prediction evaluation purposes. Moreover, initial values for \(\nu\) and \(\phi\) were obtained as described in Section 4; for \(\beta, \sigma^2, D\) they were obtained using the classical LMM through the \text{lme} function from \text{nlme} package in \text{R}; and finally for \(\lambda\) they were chosen as \(3 \times \text{sign}(\rho)\), where \(\rho\) is the sample skewness coefficient from the random effect estimated from the classical LMM.

Table 6 presents AIC and BIC criteria for all distributions and covariance structures considered. The lowest value for both criteria is the one from the AR(1)-ST-LMM and therefore this model is selected for further analyses. Table 7 summarizes the results from ML estimation of the AR(1)-ST-LMM. Moreover, 95% confidence intervals were calculated for \(\beta\) by considering the asymptotic normal approximation for the distribution of ML estimators. We conclude that all parameters are significantly different from 0, except for \(\beta_3\), the estimate of NT effect. This result corroborates the equivalent hypothesis of the new antipsychotic drug.
### TABLE 6  Selection criteria for fitting the SMSN-LMM to the schizophrenia dataset

| Distribution | AIC | BIC |
|--------------|-----|-----|
| N            | 1592.3 | 1559.6 | 1561.1 | 1561.8 | 1599.0 | 1604.9 | 1610.0 | 1605.2 |
| SN           | 1566.4 | 1542.1 | 1544.1 | 1542.7 | 1544.1 | 1590.3 | 1596.6 | 1599.7 | 1596.6 |
| ST           | 1489.6 | **1456.1** | 1458.1 | 1457.2 | 1457.9 | 1537.8 | **1508.7** | 1515.1 | 1518.5 | 1514.9 |
| SSL          | 1499.0 | 1464.6 | 1466.5 | 1465.4 | 1466.3 | 1547.2 | 1517.2 | 1523.5 | 1526.7 | 1523.3 |
| SCN          | 1502.7 | 1470.2 | 1472.1 | 1470.8 | 1471.9 | 1555.3 | 1527.2 | 1533.5 | 1536.5 | 1533.3 |

Note: Bold values indicate the smallest value from each criterion.

Abbreviations: AIC, Akaike information criterion; BIC, Bayesian information criterion; DEC, damped exponential correlation; SN, skew-normal; ST, skew-t; UNC, uncorrelated.

### TABLE 7  ML results from fitting the AR(1)-ST-LMM to the schizophrenia dataset, where $F = D^{1/2}$

| Parameter | Estimate | SE  | 95% CI  |
|-----------|----------|-----|---------|
| $\beta_0$ | 2.58     | 0.18 | 2.22    | 2.93    |
| $\beta_1$ | $-1.57$  | 0.43 | $-2.40$ | $-0.73$ |
| $\beta_2$ | 5.97     | 0.76 | 4.48    | 7.46    |
| $\beta_3$ | $-0.17$  | 0.16 | $-0.47$ | 0.14    |
| $\sigma^2_e$ | 0.27  | 0.05 |         |         |
| $\phi$    | 0.60     | 0.16 |         |         |
| $F_{11}$  | 1.13     | 0.41 |         |         |
| $F_{12}$  | 1.24     | 0.30 |         |         |
| $F_{22}$  | 2.32     | 0.68 |         |         |
| $\lambda_1$ | 10.53  | –   |         |         |
| $\lambda_2$ | 13.15  | –   |         |         |
| $\nu$     | 4.11     | –   |         |         |

Abbreviations: ML, maximum likelihood; SE, standard error; ST-LMM, skew-t linear mixed model.

Since simulations studies showed that the estimated SE of the skewness parameter is not reliable, instead of presenting its estimate we performed a likelihood ratio test for testing the hypothesis $H_0 : \lambda = 0$, as described in Subsection 3.4. Since there are two restrictions under $H_0$, the asymptotic distribution of $\Lambda_n$ is $\chi^2_2$. We have $\Lambda_n = 35.832$, hence the $p$-value of the likelihood ratio test is $P(\chi^2_2 > 35.832) \approx 0$ and we conclude that the asymmetric model is necessary for modeling the schizophrenia dataset, corroborating previous studies.

Figure 6 presents trajectories of six random subjects along with their fitted and predicted values, indicating adequacy of the proposed model. Moreover, to assess the goodness of fit of the selected model, we construct a Healy type plot, as described in Subsection 3.5, for the AR(1)-N-LMM, AR(1)-SN-LMM, AR(1)-t-LMM, and AR(1)-ST-LMM. Since Healy’s plot requires all subject to have the same number of observations, for this analysis the nonobserved measurements were imputed by the prediction procedure described in Subsection 3.2. From Figure 7 one can see that the ST model (d) is closer to the identity line than the competitive models, corroborating with the likelihood ratio test result for the asymmetry parameter and illustrating the gain in considering heavy-tailed distributions.

In order to evaluate the adequacy of the selected dependence structure, we compute the empirical autocorrelations as in (23) for UNC-ST-LMM and AR(1)-ST-LMM. The results are presented in Figure 8 and show that the empirical autocorrelation at lag 3 is significantly different from 0 for the UNC model, indicating that there are some dependence not accommodated by the UNC-ST-LMM, whereas for the autoregressive model...
FIGURE 6 Evaluation of fit and prediction for six random subjects of the schizophrenia dataset, who are identified by their ID on title.

FIGURE 7 Healy’s plot for assessing the goodness-of-fit of fitted models.

no significant autocorrelations are observed, indicating that the correlation structured from the AR(1)-ST-LMM is appropriate.

In the interest of detecting outlying observations, Figure 9(A) presents the Mahalanobis distance and the 99% quantile from its theoretical distributions (as discussed in Subsection 3.5), by number of observations, once not all patients concluded the study. From this analysis, only subject with ID 348 is classified as a possible outlier. Furthermore, from Figure 9(B) one can see that the weights ($\hat{u}_i$) are close to zero for the subjects with large Mahalanobis distance, illustrating the distribution’s capability of accommodating discrepant observations, in spite of the SN distribution (in which all observations have the same weight). Finally, plots (C) and (D) show the decomposition of the Mahalanobis distance as in (22), suggesting outlying observations only at the within-subject level.
6 | CONCLUDING REMARKS

We proposed a likelihood approach for estimation via an EM-type algorithm, model evaluation and inference of LMMs under scale mixture of SN distributions with within-subject correlation, considering some useful dependence structures. This work generalizes the results of Lachos et al.\textsuperscript{11} by developing some additional tools and making robust inferences in practical data analysis. Several simulation studies were performed in order to evaluate the proposed model. The proposed methods were implemented as part of the new \texttt{R} package \texttt{skewlmm},\textsuperscript{20} which is available for download at the CRAN repository.\textsuperscript{38}

It is worth emphasizing that even though the model formulation ensures that the random effect and the error terms are UNC, they are not independent in general. In this regard, an interesting extension would be to consider different mixing variables for the random effect and for the error, as in Asar et al.\textsuperscript{16} but in this case the likelihood function has no closed form and therefore the use of approximated approaches, such as a Monte Carlo EM algorithm, is necessary. Another promising avenue for future research is to consider the class of generalized hyperbolic (GH) distributions\textsuperscript{39} which is generated by a variance-mean mixture of a multivariate Gaussian with a generalized inverse Gaussian distribution. This rich family of GH distributions include some well-known heavy-tailed and symmetric multivariate distributions including the Normal Inverse Gaussian and some members of the family of scale-mixture of SN distributions. Finally, a possible extension of the proposed model to account for multiple responses and censored data, as proposed by Lin and Wang\textsuperscript{40,41} and Wang et al.,\textsuperscript{42} would be a promising path for future works.
ACKNOWLEDGMENTS

We are grateful to two anonymous referees, the editor and the associate editor for very useful comments and suggestions, which greatly improved this article. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and by Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil (CNPq).

ORCID

Fernanda L. Schumacher  https://orcid.org/0000-0002-5724-8918
Victor H. Lachos  https://orcid.org/0000-0002-7239-2459
Larissa A. Matos  https://orcid.org/0000-0002-2635-0901

REFERENCES

1. Drikvandi R, Verbeke G, Molenberghs G. Diagnosing misspecification of the random-effects distribution in mixed models. *Biometrics*. 2017;73(1):63-71.
2. Drikvandi R. Nonlinear mixed-effects models with misspecified random-effects distribution. *Pharm Stat*. 2019;19(3):187-201.
3. Pinheiro JC, Liu CH, Wu YN. Efficient algorithms for robust estimation in linear mixed-effects models using a multivariate t-distribution. *J Comput Graph Stat*. 2001;10:249-276.
4. Lin TI, Lee JC. A robust approach to t linear mixed models applied to multiple sclerosis data. *Stat Med*. 2006;25(8):1397-1412.
5. Rosa G, Padovani C, Gianola D. Robust linear mixed models with normal/independent distributions and Bayesian MCMC implementation. *Biom J*. 2003;45(5):573-590.
6. Arellano-Valle RB, Bolfarine H, Lachos V. Skew-normal linear mixed models. *J Data Sci*. 2005;3:415-438.
7. Azzalini A, Dalla Valle A. The multivariate skew-normal distribution. *Biometrika*. 1996;83(4):715-726.
8. Lin TI, Lee JC. Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. *Stat Med*. 2008;27(9):1490-1507.
9. Ho HJ, Lin TI. Robust linear mixed models using the skew t distribution with application to schizophrenia data. *Biom J*. 2010;52(4):449-469.
10. Azzalini A, Capitanio A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *J Royal Stat Soc Ser B (Stat Methodol)*. 2003;65(2):367-389.
11. Lachos VH, Ghosh P, Arellano-Valle RB. Likelihood based inference for skew–normal independent linear mixed models. *Stat Sin*. 2010;20:303-322.
12. Branco MD, Dey DK. A general class of multivariate skew-elliptical distributions. *J Multivar Anal*. 2001;79:99-113.
13. Pereira MAA, Russo CM. Nonlinear mixed-effects models with scale mixture of skew-normal distributions. *J Appl Stat*. 2019;46(9):1602-1620.
14. Maleki M, Wraith D, Arellano-Valle RB. A flexible class of parametric distributions for Bayesian linear mixed models. *Test*. 2019;28(2):543-564.
15. Chang SC, Zimmerman DL. Skew-normal antedependence models for skewed longitudinal data. *Biometrika*. 2016;103(2):363-376.
16. Asar Ö, Bolin D, Diggle PJ, Wallin J. Linear mixed effects models for non-Gaussian continuous repeated measurement data. *J Royal Stat Soc C (Appl Stat)*. 2020;69(5):1015-1065. https://doi.org/10.1111/rssc.12405.
17. Lachos VH, Matos LA, Castro LM, Chen MH. Flexible longitudinal linear mixed models for multiple censored responses data. *Stat Med*. 2019;38(10):1074-1102.
18. Muñoz A, Carey V, Schouten JP, Segal M, Rosner B. A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics*. 1992;48:733-742.
19. Box GEP, Jenkins GM. *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day; 1976.
20. Schumacher FL, Matos LA, Lachos VH. SKEWLMM: Scale Mixture of Skew-Normal Linear Mixed Models. R Package Version 0.2.0; 2020.
21. Lange KL, Sinsheimer JS. Normal/independent distributions and their applications in robust regression. *J Comput Graph Stat*. 1993;2:175-198.
22. Lachos VH, Cabral CRB, Zeller CB. *Finite Mixture of Skewed Distributions*. New York, NY: Springer; 2018.
23. Azzalini A, Genton M. Robust likelihood methods based on the skew-t and related distributions. *Int Stat Rev*. 2008;76:1490-1507.
24. Barndorff-Nielsen O, Schou G. On the parametrization of autoregressive models by partial autocorrelations. *J Multivar Anal*. 1973;3(4):408-419.
25. Schumacher FL, Lachos VH, Dey DK. Censored regression models with autoregressive errors: a likelihood-based perspective. *Can J Stat*. 2017;45(4):375-392.
26. Lin TI, Lee JC. On modelling data from degradation sample paths over time. *Aust N Z J Stat*. 2003;45(3):257-270.
27. Lin TI, Lee JC. Bayesian analysis of hierarchical linear mixed modeling using the multivariate t distribution. *J Stat Plan Infer*. 2007;137(2):484-495.
28. Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the EM algorithm. *J Royal Stat Soc Ser B (Methodol)*. 1977;39(1):1-38.
29. Liu C, Rubin DB. The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. *Biometrika*. 1994;81(4):633-648.
30. Wit E, Heuvel EVD, Romeijn JW. "All models are wrong...": an introduction to model uncertainty. Stat Neerl. 2012;66(3):217-236.
31. Azzalini A, Capitanio A. Statistical applications of the multivariate skew normal distribution. J Royal Stat Soc Ser B (Stat Methodol). 1999;61(3):579-602.
32. Mood AM. Introduction to the Theory of Statistics. New York, NY: McGraw-hill; 1950.
33. Healy MJR. Multivariate normal plotting. J Royal Stat Soc Ser C (Appl Stat). 1968;17(2):157-161.
34. Zeller CB, Labra FV, Lachos VH, Balakrishnan N. Influence analyses of skew-normal/independent linear mixed models. Comput Stat Data Anal. 2010;54(5):1266-1280.
35. Pinheiro JC, Bates DM. Mixed-Effects Models in S and S-PLUS. New York, NY: Springer; 2000.
36. Pinheiro J, Bates D, DeBroy S, Sarkar D, R Core Team nlme: Linear and Nonlinear Mixed Effects Models.R package version 3.1-140; 2019.
37. Lapierre Y, Nair N, Chouinard G, et al. A controlled dose-ranging study of remoxipride and haloperidol in schizophrenia-a Canadian multicentre trial. Acta Psychiatr Scand. 1990;82(S358):72-77.
38. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.
39. Browne RP, McNicholas PD. A mixture of generalized hyperbolic distributions. Can J Stat. 2015;43(2):176-198.
40. Lin TI, Wang WL. Multivariate-t nonlinear mixed models with application to censored multi-outcome AIDS studies. Biostatistics. 2017;18(4):666-681.
41. Lin TI, Wang WL. Multivariate-t linear mixed models with censored responses, intermittent missing values and heavy tails. Stat Methods Med Res. 2020;29(5):1288-1304.
42. Wang WL, Lin TI, Lachos VH. Extending multivariate-t linear mixed models for multiple longitudinal data with censored responses and heavy tails. Stat Methods Med Res. 2018;27(1):48-64.

**SUPPORTING INFORMATION**
Additional supporting information may be found online in the Supporting Information section at the end of this article.

---

**How to cite this article:** Schumacher FL, Lachos VH, Matos LA. Scale mixture of skew-normal linear mixed models with within-subject serial dependence. *Statistics in Medicine*. 2021;40:1790–1810. https://doi.org/10.1002/sim.8870