The New England Journal of Statistics in Data Science Volume 3, 119–134 (2025)
DOI: https://doi.org/10.51387/24-NEJSDS68

Linear Mixed-effects Models for Censored Data with Serial Correlation Errors Using the Multivariate Student's t-distribution

Kelin ZHONG, Rommy C. OLIVARI, Aldo M. GARAY*, and Víctor H.LACHOS

Abstract

The purpose of this paper is to develop a practical framework for the analysis of the linear mixed-effects models for censored or missing data with serial correlation errors, using the multivariate Student's t-distribution, being a flexible alternative to the use of the corresponding normal distribution. We propose an efficient ECM algorithm for computing the maximum likelihood estimates for these models with standard errors of the fixed effects and likelihood function as a by-product. This algorithm uses closed-form expressions at the E-step, which relies on formulas for the mean and variance of a truncated multivariate Student's t-distribution. In order to illustrate the usefulness of the proposed new methodology, artificial and a real dataset are analyzed. The proposed algorithm and methods are implemented in the `R` package **ARpLMEC**.

keywords and phrases: AR(p) models, Censored data, Damped exponential correlation, EM algorithm, Linear mixed effects models, Student's t-distribution.

1. INTRODUCTION

Linear mixed effects (LME) models 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-LME). 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. For instance, substantial bias in the maximum likelihood (ML) estimates of regression parameters can result when the random-effects distribution is misspecified [11].

Several researchers have investigated alternative distributions for errors in LME models. For example, [28] propose a robust hierarchical linear mixed model in which the random effects and the within-subject errors have a multivariate Student's t-distribution (t-LME). [18] and [16] developed some additional tools for t-LME models from likelihood-based and Bayesian perspectives. However, in longitudinal studies, such as those on environmental pollution and infectious diseases, measurements of some variables may be subjected to certain threshold values below or above which the measurements are not quantifiable. For instance, viral load measures the amount of actively replicating virus and, depending upon the diagnostic assays used, its measurement may be subjected to detection limits, below or above, which are not quantifiable. For this kind of data, [30] proposed the LME model for censored responses (LMEC), and an exact EM-type algorithm is proposed. Recently, [21] and [23] have considered the Student's t-distribution in the context of LMEC models (t-LMEC), including influence diagnostics analyses, under different perturbation schemes. A common feature of these classes of LMECs is that the error terms are unconditionally independent. However, in longitudinal studies, repeated measures are collected over time, and hence, the error term tends to be serially correlated.

There are several proposals in the literature that account for the time dependence in longitudinal data. For instance, [9] studied longitudinal data with random effects and AR(1) errors under the Bayesian paradigm, [4] extended the classical random effects model to deal with time dependence, leaving the random effects distribution unspecified, [13] proposed a robust structure for a censored linear model based on the multivariate Student's t-distribution, considering a damped exponential correlation (DEC) structure. More recently, [26] have proposed a full likelihood-based approach for the Gaussian N-LMEC modeling with autoregressive correlation of order p (AR(p)) errors (AR(p)-LMEC), including the implementation of the EM algorithm for maximum likelihood (ML) estimation.

Even though some proposals have been made to deal with the problem of serial correlation among the observations in LMEC models, to the best of our knowledge, there are no studies of t-LMEC with serially correlated error structures, such as DEC or AR(p). Thus, in this paper we propose the t-LMEC modeling considering some useful correlation structures (CSs). We develop a full likelihood-based treatment, including the implementation of a computationally efficient

*Corresponding author.

estimation method via the EM algorithm with the likelihood function, predictions of unobservable values of the response, and the asymptotic standard errors as byproducts. The model developed here is an extension of those previously presented by [21], [22] and [26] for the analysis of mixed-effects models with censored responses and HIV data.

The proposed algorithm and methods are implemented in the R package **ARpLMEC** [27], available on CRAN repository.

The rest of the paper is organized as follows: Section 2 discusses some preliminary results related to the multivariate Student's t-distribution and its truncated version. Some of its key properties are also presented. In Section 3, the related likelihood-based inference is presented, including estimation of the standard errors for the regression parameters. The application of the proposed method is presented in Sections 4 and 5 through some simulation studies and the analysis of data from a real HIV case study, respectively. Finally, Section 6 concludes with a short discussion of issues raised by this study and some possible directions for future research.

2. PRELIMINARIES

We begin our exposition by defining the notation and presenting some basic concepts which are used throughout the development of our methodology. We denote a random variable by an upper-case letter and its realization by the corresponding lower case and use boldface letters for vectors and matrices. $\mathbf{I}_p$ represents a $p \times p$ identity matrix and $\mathbf{A}^\top$ denotes the transpose of $\mathbf{A}$ matrix. Thus, for $\mathbf{a} = (a_1, \ldots, a_p)^\top$ and $\mathbf{b} = (b_1, \ldots, b_p)^\top$ we have that, if the Borel set in $\mathbb{R}^p$ has the form:

$$\begin{aligned}\mathbb{A} &= \{(y_1, \ldots, y_p) \in \mathbb{R}^p : \ a_1 \le y_1 \le b_1, \ldots, a_p \le y_p \le b_p\} \\ &= \{\mathbf{y} \in \mathbb{R}^p : \mathbf{a} \le \mathbf{y} \le \mathbf{b}\} \end{aligned} \tag{2.1}$$

then, we use the shorthand notations $\{\mathbf{Y} \in \mathbb{A}\} = \{\mathbf{a} \le \mathbf{Y} \le \mathbf{b}\}$ and

$$\int_{\mathbf{a}}^{\mathbf{b}} f(\mathbf{y}) d\mathbf{y} = \int_{a_1}^{b_1} \cdots \int_{a_p}^{b_p} f(y_1, \ldots, y_p) \mathrm{d}y_p \ldots \mathrm{d}y_1.$$

2.1 The Multivariate Student's t-distribution

$\mathbf{Y} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ denotes a random vector $\mathbf{Y}$ following a p-variate Student's-t distribution, with location vector $\boldsymbol{\mu}$, positive-definite scale-covariance matrix $\boldsymbol{\Sigma}$ and degrees of freedom ν and $t_p(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ denotes its probability density function (pdf).

The cumulative distribution function (cdf) of $\mathbf{Y}$ on $[\mathbf{a}, \mathbf{b}]$ is denoted by:

$$L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \int_{\mathbf{a}}^{\mathbf{b}} t_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) d\mathbf{y}. \tag{2.2}$$

An important property of the random vector $\mathbf{Y}$ is that it can be written as a scale mixture of the MVN (multivariate normal) random vector coupled with a positive random variable, i.e.,

$$\mathbf{Y} = \boldsymbol{\mu} + U^{-1/2}\mathbf{Z}, \tag{2.3}$$

where $\mathbf{Z} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$ independent of $U \sim \text{Gamma}(\nu/2, \nu/2)$, where $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\text{Gamma}(a, b)$ denote the p-variate normal with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$ and the gamma distribution with mean a/b, respectively.

2.2 The Multivariate Truncated Student's t-distribution

A p-dimensional random vector $\mathbf{Y}$ is said to follow a doubly truncated Student's t-distribution with location vector $\boldsymbol{\mu}$, scale-covariance matrix $\boldsymbol{\Sigma}$ and degrees of freedom ν over the truncation region $\mathbb{A}$, defined in Eq. (2.1), denoted by $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$, if it has the pdf:

$$Tt_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A}) = \frac{t_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}{L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}, \ \mathbf{a} \le \mathbf{y} \le \mathbf{b},$$

where $L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ as defined in Eq. (2.2).

The cdf of $\mathbf{Y}$ evaluated at the region $\{\mathbf{a} \le \mathbf{y} \le \mathbf{b}\}$ is:

$$\begin{aligned} TT_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A}) &= \frac{1}{L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)} \int_{\mathbf{a}}^{\mathbf{y}} t_p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) d\mathbf{x} \\ &= \frac{L_p(\mathbf{a}, \mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}{L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}. \end{aligned} \tag{2.4}$$

Some propositions and properties related to the p-variate Student's t-distribution and the marginal and conditional moments of the first two moments of TMVT distributions under a double truncation, which are useful for our theoretical developments, can be found in Appendix A and B, respectively.

3. LINEAR MIXED-EFFECTS WITH CENSORED RESPONSE

3.1 Model Specification

We proceed as in [21, 28], by considering a generalization of the classical N-LME model in which the random terms are assumed to follow a Student's-t distribution as follows:

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

$$(\mathbf{b}_i, \boldsymbol{\epsilon}_i)^\top \sim t_{n_i+q}\big(\mathbf{0}, \text{Diag}(\mathbf{D}, \boldsymbol{\Omega}_i), \nu\big). \tag{3.2}$$

The subscript "i" is the subject index; Diag($\mathbf{A}$,$\mathbf{B}$) is a block diagonal matrix whose elements are the matrices $\mathbf{A}$ and $\mathbf{B}$. $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^\top$ is a $n_i \times 1$ vector of observed continuous responses for sample unit "i", $\mathbf{X}_i$ is the $n_i \times s$ design matrix corresponding to the fixed effects $\boldsymbol{\beta}$, which is the $s \times 1$ vector of population-averaged regression coefficients. $\mathbf{Z}_i$ is the $n_i \times q$ design matrix corresponding to the $q \times 1$

vector of random effects $\mathbf{b}_i$. $\boldsymbol{\epsilon}_i$ is the $n_i \times 1$ vector of random errors, the dispersion matrix $\mathbf{D} = \mathbf{D}(\boldsymbol{\alpha})$ depends on unknown and reduced parameters $\boldsymbol{\alpha}$. The correlation structure of the error vector is assumed to be $\boldsymbol{\Omega}_i = \sigma^2\mathbf{R}_i$, where $\mathbf{R}_i = \mathbf{R}_i(\boldsymbol{\phi})$, for $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^\top$ is a $n_i \times n_i$ matrix, that incorporates a time-dependence structure.

Note that $\mathbf{b}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated, once $\mathrm{Cov}(\mathbf{b}_i, \boldsymbol{\epsilon}_i) = \mathbb{E}[\mathbf{b}_i\boldsymbol{\epsilon}_i^\top] = \mathbb{E}[\mathbb{E}(\mathbf{b}_i\boldsymbol{\epsilon}_i^\top|U_i)] = \mathbf{0}$, where U_i is a scalar generated from Gamma$(\nu/2, \nu/2)$. Classical inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top, \boldsymbol{\phi}^\top, \nu)^\top$ is based on the marginal distribution of $\mathbf{Y}_i$, thus $\mathbf{Y}_i \overset{\text{ind.}}{\sim} t_{n_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \nu)$, for $i = 1, \ldots, n$, where $\boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}_i = \sigma^2\mathbf{R}_i + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top$. The estimates from the multivariate t-LME are more robust against outliers than those based on the standard LME; in this sense, [28] showed by simulation study that the t-LME substantially outperforms the normal or standard LME when outliers are present in the data. This issue has also been discussed by [21] in the context of censored mixed-effects models.

In this study, we follow [13, 15, 21, 23, 26] by considering non-informative censoring observations (or censoring at random), i.e. we assume that the response Y_{ij} is not fully observed for all i, j. Thus, let $(\mathbf{V}_i, \mathbf{C}_i)$ be the observed data for the i-th subject, where $\mathbf{V}_i$ represents the vector of uncensored readings ($V_{ij} = V_{0i}$) or censoring interval (V_{1ij}, V_{2ij}) and $\mathbf{C}_i$ is the vector of censoring indicators, such that:

$$C_{ij} = \begin{cases} 1 & \text{if} \quad V_{1ij} \leq Y_{ij} \leq V_{2ij}, \\ 0 & \text{if} \quad Y_{ij} = V_{0i}, \end{cases} \tag{3.3}$$

for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, n_i\}$, i.e., $C_{ij} = 1$ if Y_{ij} is located within a specific interval. Note that for a right-censored observation $V_{2ij} = +\infty$ and for a left-censored observation $V_{1ij} = -\infty$. Moreover, missing at random (MAR) observations can be handled by considering $V_{1ij} = -\infty$ and $V_{2ij} = +\infty$. The model defined in Eqs. (3.1)–(3.3) is henceforth called the t-LMEC model.

3.2 Within-subject Dependence Structures

In order to enable some flexibility when modeling the error covariance, we consider essentially three dependence structures: unconditionally independent (UNC), the continuous-type autoregressive process of order p and the damped exponential correlation, which will be discussed next.

- *Unconditional independence*
  The most common and simplest approach is to assume that the error terms are Unconditionally independent (UNC), i.e. we have $\mathbf{R}_i = \mathbf{I}_{n_i}$, for each $i = 1, \ldots, n$. In practice, the linear mixed models considering UNC errors is very frequent, for example it has been considered by Matos et al. [21]. It will be denoted as UNC-t-LMEC.
  However, in longitudinal studies, 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.

- *Autoregressive dependence of order p*
  In this case, we propose $\mathbf{R}_i$ as a structured AR(p) dependence matrix [8]. Specifically,

$$\mathbf{R}_i = \mathbf{R}_i(\boldsymbol{\phi}) = \frac{1}{1 - \phi_1\rho_1 - \cdots - \phi_p\rho_p}[\rho_{|r-s|}], \tag{3.4}$$

  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 $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^\top$ and satisfy the Yule–Walker equations [7], i.e.,

$$\rho_k = \phi_1\rho_{k-1} + \cdots + \phi_p\rho_{k-p}, \;\; \rho_0 = 1, \;\; k = 1, \ldots, p.$$

  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 [3], 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 [29], see also [17, 16] for details on estimation of LME model with AR(p) dependence.
  The model formulated in Eqs. (3.1)–(3.3) with $\boldsymbol{\Omega}_i = \sigma^2\mathbf{R}_i$, where $\mathbf{R}_i$ is given by Eq. (3.4), $i = 1, \ldots, n$, will be denoted AR(p)-t-LMEC. In order 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.
- *Damped exponential correlation*
  In this case, following [25], we propose to structure $\mathbf{R}_i$ as a damped exponential correlation (DEC) matrix, as follows:

$$\mathbf{R}_i = \mathbf{R}_i(\boldsymbol{\phi}, \mathbf{t}_i) = \left[\phi_1^{|t_{ij} - t_{ik}|^{\phi_2}}\right], \;\; 0 \leq \phi_1 < 1, \;\; \phi_2 \geq 0, \tag{3.5}$$

  where $j, k = 1, \ldots, n_i$, for $i = 1, \ldots, n$ and $\boldsymbol{\phi} = (\phi_1, \phi_2)^\top$. The correlation parameter ϕ_1 describes the autocorrelation between observations separated by the absolute length of two time points, and the damping parameter ϕ_2 allows the acceleration of the exponential decay of the autocorrelation function defining a continuous-time autoregressive (AR) model. It is important to stress that considering the DEC structure it is possible to obtain different correlation structures. For example:

  (i) if $\phi_2 = 0$, then $\mathbf{R}_i$ reduces to the compound symmetry correlation structure (DEC-SYM);

  (ii) if $\phi_2 = 1$, then $\mathbf{R}_i$ reduces to the DEC-AR(1) correlation structure;

(iii) if $0 < \phi_2 < 1$, then $\mathbf{R}_i$ generates a decay rate slower than the DEC-AR(1) structure;

(iv) if $\phi_2 > 1$, then $\mathbf{R}_i$ generates a decay rate faster than the DEC-AR(1) structure; and

(v) if $\phi_2 \to \infty$, then $\mathbf{R}_i$ converges to the correlation matrix of a moving-average of order 1 (DEC-MA(1)).

The model presented in Eqs. (3.1)–(3.3) considering $\boldsymbol{\Omega}_i = \sigma^2\mathbf{R}_i$, where $\mathbf{R}_i$ is given by Eq. (3.5), $i = 1, \ldots, n$, will be denoted DEC-t-LMEC.

3.3 The Likelihood Function

Let $\mathbf{Y}_i = (\mathbf{Y}_i^o, \mathbf{Y}_i^c)^\top$ where $\mathbf{Y}_i^o$ and $\mathbf{Y}_i^c$ represent the n_i^o and n_i^c vectors of observed outcomes and censored observations for subject i, respectively; with $n_i = n_i^o + n_i^c$, such that $C_{ij} = 0$ for all elements in $\mathbf{Y}_i^o$ and 1 for all elements in $\mathbf{Y}_i^c$. After reordering, $\mathbf{Y}_i$, $\mathbf{V}_i$, $\boldsymbol{\mu}_i$, and $\boldsymbol{\Sigma}_i$ can be partitioned as follows:

$$\mathbf{Y}_i = \text{vec}\big(\mathbf{Y}_i^o, \mathbf{Y}_i^c\big),\ \mathbf{V}_i = \text{vec}\big(\mathbf{V}_i^o, \mathbf{V}_i^c\big),\ \boldsymbol{\mu}_i^\top = \big(\boldsymbol{\mu}_i^o, \boldsymbol{\mu}_i^c\big) \text{ and } \boldsymbol{\Sigma}_i = \begin{pmatrix} \boldsymbol{\Sigma}_i^{oo} & \boldsymbol{\Sigma}_i^{oc} \\ \boldsymbol{\Sigma}_i^{co} & \boldsymbol{\Sigma}_i^{cc} \end{pmatrix}, \tag{3.6}$$

where vec(.) denotes the function which stacks vectors or matrices of the same number of columns.

Using Proposition 1 (Appendix A), we have that

$$\mathbf{Y}_i^o \sim t_{n_i^o}\big(\boldsymbol{\mu}_i^o, \boldsymbol{\Sigma}_i^{oo}, \nu\big), \quad \text{and}$$
$$\mathbf{Y}_i^c|\mathbf{Y}_i^o = \mathbf{y}_i^o \sim t_{n_i^c}\big(\boldsymbol{\mu}_i^{co}, \mathbf{S}_i^{co}, \nu + n_i^o\big),$$

where

$$\begin{aligned} \boldsymbol{\mu}_i^o &= \mathbf{X}_i^o\boldsymbol{\beta} \ \text{ and } \ \boldsymbol{\mu}_i^c = \mathbf{X}_i^c\boldsymbol{\beta}, \\ \boldsymbol{\mu}_i^{co} &= \boldsymbol{\mu}_i^c + \boldsymbol{\Sigma}_i^{co}\boldsymbol{\Sigma}_i^{oo-1}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o), \\ \mathbf{S}_i^{co} &= \big(\tfrac{\nu+\delta^2(\mathbf{y}_i^o)}{\nu+n_i^o}\big)\mathbf{S}_i, \\ \mathbf{S}_i &= \boldsymbol{\Sigma}_i^{cc} - \boldsymbol{\Sigma}_i^{co}\boldsymbol{\Sigma}_i^{oo-1}\boldsymbol{\Sigma}_i^{oc} \ \text{ and} \\ \delta^2(\mathbf{y}_i^o) &= (\mathbf{y}_i^o - \boldsymbol{\mu}_i^o)^\top\boldsymbol{\Sigma}_i^{oo-1}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o). \end{aligned}$$

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top, \boldsymbol{\phi}^\top, \nu)^\top$ be the parameters vector, as presented by [21], the likelihood function for subject i is given by:

$$\begin{aligned} L_i(\boldsymbol{\theta}) &= f(\mathbf{y}_i|\boldsymbol{\theta}) = f(\mathbf{V}_i|\mathbf{C}_i, \boldsymbol{\theta}) \\ &= f\big(\mathbf{y}_i^o|\boldsymbol{\theta}\big)P\big(\mathbf{V}_{1i}^c \le \mathbf{Y}_i^c \le \mathbf{V}_{2i}^c|\mathbf{V}_i^o, \boldsymbol{\theta}\big) \\ &= t_{n_i^o}\big(\mathbf{V}_i^o; \boldsymbol{\mu}_i^o, \boldsymbol{\Sigma}_i^{oo}, \boldsymbol{\nu}\big)L_{n_i^c}\big(\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c; \boldsymbol{\mu}_i^{co}, \mathbf{S}_i^{co}, \nu + n_i^o\big), \end{aligned} \tag{3.7}$$

where $L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ is as defined in Eq. (2.2), which can be easily evaluated by using the `R` package **MomTrunc**.

The log-likelihood function for the observed data is given by $\ell(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^n \log L_i(\boldsymbol{\theta})$, and the estimates obtained by maximizing the log-likelihood function $\ell(\boldsymbol{\theta}|\mathbf{y})$ are the ML estimates.

3.4 The EM Algorithm

In order to obtain the ML estimation of the parameters in the t-LMEC model, presented in Section 3.1, we implemented the ECM algorithm, proposed by [24]. This algorithm preserves the properties of simplicity and stability of EM algorithm, however replace the M-step with a sequence of conditional maximization (CM) steps. Based in property of multivariate Student's t-distribution, the t-LMEC model can be expressed in the following hierarchical model:

$$\begin{aligned} \mathbf{Y}i|\mathbf{b}_i, u_i &\overset{\text{ind.}}{\sim} \text{N}_{n_i}\big(\boldsymbol{\mu}_i, u_i^{-1}\boldsymbol{\Omega}_i\big), \\ \mathbf{b}_i|u_i &\overset{\text{ind.}}{\sim} \text{N}_q\big(\mathbf{0}, u_i^{-1}\mathbf{D}\big), \\ u_i &\overset{\text{iid.}}{\sim} Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right). \end{aligned}$$

Thus, it is possible to apply the ECM algorithm by assuming that $\mathbf{Y} = (\mathbf{Y}_1^\top, \ldots, \mathbf{Y}_n^\top)$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)$ and $\mathbf{u} = (u_1, \ldots, u_n)^\top$ are hypothetical missing variables, and augmenting with the observed variables $(\mathbf{V}, \mathbf{C})$ where $\mathbf{V} = \text{vec}(\mathbf{V}_1, \ldots, \mathbf{V}_n)$ and $\mathbf{C} = \text{vec}(\mathbf{C}_1, \ldots, \mathbf{C}_n)$.

Considering $\mathbf{Y}_c = (\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{y}^\top, \mathbf{b}^\top, \mathbf{u}^\top)^\top$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top, \boldsymbol{\phi}^\top, \nu)^\top$, the complete-data log-likelihood, is given by

$$\begin{aligned} \ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = \sum_{i=1}^n \Bigg[&-\frac{n_i}{2}\log\sigma^2 - \frac{1}{2}\log(|\mathbf{R}_i|) \\ &- \frac{u_i}{2\sigma^2}(\mathbf{y}_i - \boldsymbol{\mu}_i - \mathbf{Z}_i\mathbf{b}_i)^\top\mathbf{R}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i - \mathbf{Z}_i\mathbf{b}_i) \\ &- \frac{1}{2}\log|\mathbf{D}| - \frac{u_i}{2}\mathbf{b}_i^\top\mathbf{D}^{-1}\mathbf{b}_i + \log h(u_i|\nu) + C\Bigg], \end{aligned} \tag{3.8}$$

where C is a constant that does not depend on the $\boldsymbol{\theta}$. $h(u_i|\nu)$ represents the *pdf* of a $Gamma(\nu/2, \nu/2)$ distribution.

Given the estimate $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, at iteration "$(k)$" of the algorithm, in the E-step computes the conditional expectation of the complete-data log-likelihood function, given by:

$$\begin{aligned} Q\big(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}\big) &= \mathbb{E}\big[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}\big] \\ &= \sum_{i=1}^n Q_{1i}\big(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}|\widehat{\boldsymbol{\theta}}^{(k)}\big) + \sum_{i=1}^n Q_{2i}\big(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}^{(k)}\big), \end{aligned}$$

where

$$\begin{aligned} &Q_{1i}\big(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}|\widehat{\boldsymbol{\theta}}^{(k)}\big) \\ &\quad= -\frac{n_i}{2}\log\sigma^2 - \frac{1}{2}\log(|\mathbf{R}_i|) \\ &\qquad- \frac{1}{2\sigma^2}\Big[\widehat{a}_i^{(k)} - 2\boldsymbol{\mu}_i^\top\mathbf{R}_i^{-1}\big(\widehat{u_i\mathbf{y}_i}^{(k)} - \mathbf{Z}_i\widehat{u_i\mathbf{b}_i}^{(k)}\big) \\ &\qquad+ \widehat{u}_i^{(k)}\boldsymbol{\mu}_i^\top\mathbf{R}_i^{-1}\boldsymbol{\mu}_i\Big] \quad \text{and} \end{aligned}$$

$$Q_{2i}\big(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}^{(k)}\big) = -\frac{1}{2}\log|\mathbf{D}| - \frac{1}{2}\,\text{tr}\big(\widehat{u_i\mathbf{b}_i\mathbf{b}_i^\top}^{(k)}\mathbf{D}^{-1}\big), \text{ with}$$

$$
\begin{aligned}
\widehat{a}_i^{(k)} &= \operatorname{tr}\big(\widehat{u_i\mathbf{y}_i\mathbf{y}_i^\top}^{(k)}\mathbf{R}_i^{-1} - 2\widehat{u_i\mathbf{y}_i\mathbf{b}_i^\top}^{(k)}\mathbf{Z}_i^\top\mathbf{R}_i^{-1} \\
&\quad + \widehat{u_i\mathbf{b}_i\mathbf{b}_i^\top}^{(k)}\mathbf{Z}_i^\top\mathbf{R}_i^{-1}\mathbf{Z}_i\big), \\
\widehat{u_i\mathbf{b}_i}^{(k)} &= \mathbb{E}\big[u_i\mathbf{b}_i|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\big] = \boldsymbol{\varphi}_i\big(\widehat{u_i\mathbf{y}_i}^{(k)} - \widehat{u}_i^{(k)}\boldsymbol{\mu}_i\big), \\
\widehat{u_i\mathbf{b}_i\mathbf{b}_i^\top}^{(k)} &= \mathbb{E}\big[u_i\mathbf{b}_i\mathbf{b}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\big] \\
&= \boldsymbol{\Lambda}_i + \boldsymbol{\varphi}_i\big(\widehat{u_i\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} - 2\widehat{u_i\mathbf{y}_i}^{(k)}\boldsymbol{\mu}_i \\
&\quad + \widehat{u}_i^{(k)}\boldsymbol{\mu}_i\boldsymbol{\mu}_i^\top\big)\boldsymbol{\varphi}_i^\top, \\
\widehat{u_i\mathbf{y}_i\mathbf{b}_i^\top}^{(k)} &= \mathbb{E}\big[u_i\mathbf{y}_i\mathbf{b}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\big] \\
&= \big(\widehat{u_i\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} - \widehat{u_i\mathbf{y}_i}^{(k)}\boldsymbol{\mu}_i^\top\big)\boldsymbol{\varphi}_i^\top,
\end{aligned}
$$

where $\boldsymbol{\Lambda}_i = (\mathbf{D}^{-1} + \mathbf{Z}_i^\top\mathbf{R}_i^{-1}\mathbf{Z}_i/\sigma^2)^{-1}$ and $\boldsymbol{\varphi}_i = \boldsymbol{\Lambda}_i\mathbf{Z}_i^\top\mathbf{R}_i^{-1}/\sigma^2$.

In conditional maximization (CM) steps, we update $\widehat{\boldsymbol{\theta}}^{(k+1)}$ by maximizing $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ over $\boldsymbol{\theta}$, as follows:

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}^{(k+1)} &= \left(\sum_{i=1}^n \widehat{u}_i^{(k)}\mathbf{X}_i^\top\widehat{\mathbf{R}}_i^{-1(k)}\mathbf{X}_i\right)^{-1}\sum_{i=1}^n \mathbf{X}_i^\top\widehat{\mathbf{R}}_i^{-1(k)} \\
&\quad \times \big(\widehat{u_i\mathbf{y}_i}^{(k)} - \mathbf{Z}_i\widehat{u_i\mathbf{b}_i}^{(k)}\big), \qquad (3.9) \\
\widehat{\sigma^2}^{(k+1)} &= \frac{1}{N}\sum_{i=1}^n \Big[\widehat{a}_i^{(k)} - 2\widehat{\boldsymbol{\mu}}_i^{(k+1)\top}\mathbf{R}_i^{-1} \\
&\quad \times \big(\widehat{u_i\mathbf{y}_i}^{(k)} - \mathbf{Z}_i\widehat{u_i\mathbf{b}_i}^{(k)}\big) \\
&\quad + \widehat{u}_i^{(k)}\widehat{\boldsymbol{\mu}}_i^{(k+1)\top}\mathbf{R}_i^{-1}\widehat{\boldsymbol{\mu}}_i^{(k+1)}\Big], \qquad (3.10) \\
\widehat{\mathbf{D}}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^n \widehat{u_i\mathbf{b}_i\mathbf{b}_i^\top}^{(k)}, \qquad (3.11) \\
\widehat{\phi}^{(k+1)} &= \arg\max_{\phi}\Bigg(-\frac{1}{2}\log(|\mathbf{R}_i|) \\
&\quad - \frac{1}{2\widehat{\sigma^2}^{(k+1)}}\Big[\widehat{a}_i^{(k)} - 2\widehat{\boldsymbol{\mu}}_i^{(k+1)\top}\mathbf{R}_i^{-1} \\
&\quad \times \big(\widehat{u_i\mathbf{y}_i}^{(k)} - \mathbf{Z}_i\widehat{u_i\mathbf{b}_i}^{(k)}\big) \\
&\quad + \widehat{u}_i^{(k)}\widehat{\boldsymbol{\mu}}_i^{(k+1)\top}\mathbf{R}_i^{-1}\widehat{\boldsymbol{\mu}}_i^{(k+1)}\Big]\Bigg), \qquad (3.12) \\
\widehat{\nu}^{(k+1)} &= \arg\max_{\nu}\Bigg\{\sum_{i=1}^n \log \\
&\qquad L_{n_i^c}\big(\mathbf{V}_{1i}^c,\mathbf{V}_{2i}^c;\boldsymbol{\mu}_i^{co^{(k+1)}},\mathbf{S}_i^{co^{(k+1)}},\nu+n_i^o\big) \\
&\quad + \sum_{i=1}^n \log t_{n_i^o}\big(\mathbf{V}_i^o;\boldsymbol{\mu}_i^{o^{(k+1)}},\boldsymbol{\Sigma}_i^{oo^{(k+1)}},\boldsymbol{\nu}\big)\Bigg\}, \qquad (3.13)
\end{aligned}
$$

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

The algorithm is iterated until a suitable convergence rule is satisfied. In this case, the process is iterated until some distance between two successive evaluations of the currently penalized log-likelihood $\ell_p(\boldsymbol{\theta},\lambda)$, such as $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)})/\ell(\widehat{\boldsymbol{\theta}}^{(k)}) - 1|$ becomes small enough. For example, we adopted $\epsilon = 10^{-6}$.

As proposed by [26], a set of starting values may be achieved by computing $\widehat{\boldsymbol{\beta}}^{(0)}$, $\widehat{\sigma^2}^{(0)}$, $\widehat{\mathbf{D}}^{(0)}$ and $\widehat{\boldsymbol{\phi}}^{(0)}$, as the solution of the normal linear mixed-effects model, using the `R` package **nlme**.

On the other hand, it is important to stress that, from Eqs. (3.9)–(3.12), the E-step reduces to the computation of

$$
\begin{aligned}
\widehat{u_i\mathbf{y}_i\mathbf{y}_i^\top} &= \mathbb{E}\big[u_i\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\boldsymbol{\theta}\big], \\
\widehat{u_i\mathbf{y}_i} &= \mathbb{E}[u_i\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i,\boldsymbol{\theta}] \quad \text{and} \\
\widehat{u}_i &= \mathbb{E}[u_i|\mathbf{V}_i,\mathbf{C}_i,\boldsymbol{\theta}].
\end{aligned}
$$

These expected values can be determined, using the marginal and conditional moments of the first two moments of the TMVT distribution under a double truncation, presented in Propositions 3 and 4 (Appendix B). Thus:

1. If the i-th subject has only non-censored components, then:

$$
\widehat{u}_i = \left(\frac{\nu+n_i}{\nu+\delta^2(\mathbf{y}_i)}\right), \quad \widehat{u_i\mathbf{y}_i} = \widehat{u}_i\mathbf{y}_i, \quad \widehat{u_i\mathbf{y}_i\mathbf{y}_i^\top} = \widehat{u}_i\mathbf{y}_i\mathbf{y}_i^\top,
$$

where $\delta^2(\mathbf{y}_i) = (\mathbf{y}_i-\boldsymbol{\mu}_i)^\top\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i-\boldsymbol{\mu}_i)$.

2. If the i-th subject has only censored components then, we have:

$$
\begin{aligned}
\widehat{u}_i &= \frac{L_{n_i}(\mathbf{V}_{1i},\mathbf{V}_{2i};\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i^*,\nu+2)}{L_{n_i}(\mathbf{V}_{1i},\mathbf{V}_{2i};\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i,\nu)}, \\
\widehat{u_i\mathbf{y}_i} &= \widehat{u}_i\mathbb{E}(\mathbf{W}_i), \\
\widehat{u_i\mathbf{y}_i\mathbf{y}_i^\top} &= \widehat{u}_i\mathbb{E}\big(\mathbf{W}_i\mathbf{W}_i^\top\big),
\end{aligned}
$$

where $\mathbf{W} \sim Tt_{n_i}(\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i^*,\nu+2;(\mathbf{V}_{1i},\mathbf{V}_{2i}))$, $\boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_i^* = \frac{\nu}{\nu+2}\boldsymbol{\Sigma}_i$, $\boldsymbol{\Sigma}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top + \boldsymbol{\Omega}_i$.

3. If the i-th subject has censored and uncensored components and given that $\mathbf{Y}_i|\mathbf{V}_i,\mathbf{C}_i$, $\mathbf{Y}_i|\mathbf{V}_i,\mathbf{C}_i,\mathbf{Y}_i^o$ and $\mathbf{Y}_i^c|\mathbf{V}_i$, $\mathbf{C}_i$, $\mathbf{Y}_i^o$ are equivalent process, then, we have:

$$
\begin{aligned}
\widehat{u}_i &= \left(\frac{n_i^o+\nu}{\nu+\delta^2(\mathbf{y}_i^o)}\right) \\
&\quad \times \frac{L_{n_i^c}(\mathbf{V}_{1i}^c,\mathbf{V}_{2i}^c;\boldsymbol{\mu}_i^{co},\widetilde{\mathbf{S}}_i^{co},\nu+n_i^o+2)}{L_{n_i^c}(\mathbf{V}_{1i}^c,\mathbf{V}_{2i}^c;\boldsymbol{\mu}_i^{co},\mathbf{S}_i^{co},\nu+n_i^o)}, \\
\widehat{u_i\mathbf{y}_i} &= \operatorname{vec}\big(\widehat{u}_i\mathbf{y}_i^o,\widehat{u}_i\mathbb{E}[\mathbf{W}_i]\big), \\
\widehat{u_i\mathbf{y}_i\mathbf{y}_i^\top} &= \begin{pmatrix}\widehat{u}_i\mathbf{y}_i^o\mathbf{y}_i^{o\top} & \widehat{u}_i\mathbf{y}_i^o\mathbb{E}^\top[\mathbf{W}_i] \\ \widehat{u}_i\mathbb{E}[\mathbf{W}_i]\mathbf{y}_i^{o^\top} & \widehat{u}_i\mathbb{E}[\mathbf{W}_i\mathbf{W}_i^\top]\end{pmatrix},
\end{aligned}
$$

where $\mathbf{W}_i \sim Tt_{n_i^c}(\boldsymbol{\mu}_i^{co},\widetilde{\mathbf{S}}_i^{co},\nu+n_i^o+2;(\mathbf{V}_{1i}^c,\mathbf{V}_{2i}^c))$, $\widetilde{\mathbf{S}}_i^{co} = (\frac{\nu+\delta^2(\mathbf{y}_i^o)}{\nu+n_i^o+2})\mathbf{S}_i$ and $\mathbf{S}_i$, $\mathbf{S}_i^{co}$ and $\boldsymbol{\mu}_i^{co}$ are defined in Section 3.3.

Formulas for $\mathbb{E}[\mathbf{W}]$ and $\mathbb{E}[\mathbf{W}\mathbf{W}^\top]$, where $\mathbf{W} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$, have been recently developed in the **R** package `relliptical` [31], using a slice sampling algorithm with Gibbs sampler steps.

3.5 The Expected Information Matrix

As developed by [21] and [26], we used the [20]'s technique to obtain an asymptotic approximation for the variances of the fixed effects in our t-MLE model. This approximation is given by:

$$\mathbf{I}_s(\boldsymbol{\beta}|\mathbf{y}) = \mathbf{I}_c(\boldsymbol{\beta}|\mathbf{y}) - \mathbf{I}_m(\boldsymbol{\beta}|\mathbf{y}),$$

where $\mathbf{I}_s(\boldsymbol{\beta}|\mathbf{y})$ represents the information about $\boldsymbol{\beta}$ in the observed data $\mathbf{y}$, $\mathbf{I}_c(\boldsymbol{\beta}|\mathbf{y})$ is the conditional expectation of the complete-data information and $\mathbf{I}_m(\boldsymbol{\beta}|\mathbf{y})$ is the missing information. Therefore, the approximated covariance-matrix of $\widehat{\boldsymbol{\beta}}$ is given by:

$$\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\beta}}) \approx \mathbf{I}_s^{-1}(\boldsymbol{\beta})|_{\widehat{\boldsymbol{\theta}}},$$

where

$$\mathbf{I}_s(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ \left(\frac{\nu + n_i}{\nu + n_i + 2} \right) \mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i - \mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1} \left(\left(\frac{\nu + n_i + 2}{\nu + n_i} \right) \mathrm{E}_2 - \mathrm{E}_1 \right) \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right\},$$

where

$$\mathrm{E}_1 = (\widehat{u_i \mathbf{y}_i} - \widehat{u}_i \boldsymbol{\mu}_i)(\widehat{u_i \mathbf{y}_i} - \widehat{u}_i \boldsymbol{\mu}_i)^\top \quad \text{and}$$
$$\mathrm{E}_2 = \left(\widehat{u_i^2 \mathbf{y}_i \mathbf{y}_i^\top} - \widehat{u_i^2 \mathbf{y}_i} \boldsymbol{\mu}_i^\top - \boldsymbol{\mu}_i \widehat{u_i^2 \mathbf{y}_i^\top} + \widehat{u_i^2} \boldsymbol{\mu}_i \boldsymbol{\mu}_i^\top \right).$$

Note that E_1 depend on the computation of $\widehat{u}_i$, $\widehat{u_i \mathbf{y}_i}$ that can be obtained in Section 3.4 and E_2 depend on the following quantities

$$\widehat{u_i^2} = \mathbb{E}\left[\left(\frac{\nu + n_i}{\nu + \delta^2(\mathbf{y}_i)} \right)^2 | \mathbf{V}_i, \mathbf{C}_i, \boldsymbol{\theta} \right],$$
$$\widehat{u_i^2 \mathbf{y}_i} = \mathbb{E}\left[\left(\frac{\nu + n_i}{\nu + \delta^2(\mathbf{y}_i)} \right)^2 \mathbf{y}_i | \mathbf{V}_i, \mathbf{C}_i, \boldsymbol{\theta} \right] \quad \text{and}$$
$$\widehat{u_i^2 \mathbf{y}_i \mathbf{y}_i^\top} = \mathbb{E}\left[\left(\frac{\nu + n_i}{\nu + \delta^2(\mathbf{y}_i)} \right)^2 \mathbf{y}_i \mathbf{y}_i^\top | \mathbf{V}_i, \mathbf{C}_i, \boldsymbol{\theta} \right].$$

These expected values can be determined in closed form using Propositions 3 and 4, as follows:

1. If the i-th subject has only non-censored components, then,

$$\widehat{u_i^2} = \left(\frac{\nu + n_i}{\nu + \delta^2(\mathbf{y}_i)} \right)^2, \quad \widehat{u_i^2 \mathbf{y}_i} = \widehat{u_i^2} \mathbf{y}_i, \quad \widehat{u_i^2 \mathbf{y}_i \mathbf{y}_i^\top} = \widehat{u_i^2} \mathbf{y}_i \mathbf{y}_i^\top,$$

where $\delta^2(\mathbf{y}_i) = (\mathbf{y}_i - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i)$.

2. If the i-th subject has only censored components then

$$\widehat{u_i^2} = c_p(\nu, 2) \frac{L_{n_i}(\mathbf{V}_{1i}, \mathbf{V}_{2i}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i^*, \nu + 4)}{L_{n_i}(\mathbf{V}_{1i}, \mathbf{V}_{2i}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \nu)},$$
$$\widehat{u_i^2 \mathbf{y}_i} = \widehat{u_i^2} \mathbb{E}[\mathbf{W}_i],$$
$$\widehat{u_i^2 \mathbf{y}_i \mathbf{y}_i^\top} = \widehat{u_i^2} \mathbb{E}\left[\mathbf{W}_i \mathbf{W}_i^\top \right],$$

where $c_p(\nu, 2) = \frac{(n_i + \nu)(\nu + 2)}{\nu(n_i + \nu + 2)}$, $\mathbf{W}_i \sim Tt_{n_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i^*, \nu + 4; (\mathbf{V}_{1i}, \mathbf{V}_{2i}))$, $\boldsymbol{\Sigma}_i^* = \frac{\nu}{\nu+4} \boldsymbol{\Sigma}_i$, $\boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta}$, $\boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top + \boldsymbol{\Omega}_i$.

3. If the i-th subject has censored and uncensored components and given that $\{\mathbf{Y}_i | \mathbf{V}_i, \mathbf{C}_i\}$, $\{\mathbf{Y}_i | \mathbf{V}_i, \mathbf{C}_i, \mathbf{Y}_i^o\}$ and $\{\mathbf{Y}_i^c | \mathbf{V}_i, \mathbf{C}_i, \mathbf{Y}_i^o\}$ are equivalent process, we have:

$$\widehat{u_i^2} = \frac{d_p(n_i^o, \nu, 2)}{(\nu + \delta^2(\mathbf{y}_i^o))^2} \times \frac{L_{n_i^c}(\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c; \boldsymbol{\mu}_i^{co}, \widetilde{\mathbf{S}}_i^{co}, \nu + n_i^o + 4)}{L_{n_i^c}(\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c; \boldsymbol{\mu}_i^{co}, \mathbf{S}_i^{co}, \nu + n_i^o)},$$
$$\widehat{u_i^2 \mathbf{y}_i} = \mathrm{vec}\left(\widehat{u_i^2 \mathbf{y}_i^o}, \widehat{u_i^2} \mathbb{E}[\mathbf{W}_i] \right),$$
$$\widehat{u_i^2 \mathbf{y}_i \mathbf{y}_i^\top} = \begin{pmatrix} \widehat{u_i^2 \mathbf{y}_i^o \mathbf{y}_i^{o\top}} & \widehat{u_i^2 \mathbf{y}_i^o} \mathbb{E}^\top[\mathbf{W}_i] \\ \widehat{u_i^2} \mathbb{E}[\mathbf{W}_i] \mathbf{y}_i^{o\top} & \widehat{u_i^2} \mathbb{E}[\mathbf{W}_i \mathbf{W}_i^\top] \end{pmatrix},$$

where $d_p(n_i^o, \nu, 2) = \frac{(\nu + n_i)(n_i^o + \nu + 2)(n_i^o + \nu)}{n_i + \nu + 2}$, $\mathbf{W}_i \sim Tt_{n_i^c}(\boldsymbol{\mu}_i^{co}, \widetilde{\mathbf{S}}_i^{co}, \nu + n_i^o + 4; (\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c))$, $\widetilde{\mathbf{S}}_i^{co} = (\frac{\nu + \delta^2(\mathbf{y}_i^o)}{\nu + n_i^o + 4}) \mathbf{S}_i$ and $\mathbf{S}_i$, $\mathbf{S}_i^{co}$ and $\boldsymbol{\mu}_i^{co}$ are presented in Section 3.3.

4. SIMULATION STUDIES

In this Section, we conduct two simulation studies in order to analyze the performance of our proposed methods. The first simulation study shows the asymptotic behavior of the ML estimators developed in Section 3.4 and the second one examines the consistency of approximate the standard error (SE) presented in Section 3.5. We consider the AR(p)-LMEC model, defined in Eqs. (3.1)–(3.3) with left censored cases and $n_i = 7$ measurements for each subject. As recommended by [26], for both simulation schemes, the initial set of values to generate the sample were $\beta = (1, 2, 1)^\top$, $\sigma^2 = 0.5$. For the dispersion matrices of the random error and random effects we assume an autoregressive dependence of order $p = 2$ with $\phi = (0.6, -0.2)^\top$ and $D = \begin{pmatrix} 0.490 & 0.001 \\ 0.001 & 0.002 \end{pmatrix}$, respectively. The design matrices for each individual are given by $\mathbf{X}_i = (\mathbf{X}_{i1}, \mathbf{X}_{i2}, \mathbf{X}_{i3})$ and $\mathbf{Z}_i = (\mathbf{1}_7, \mathbf{Z}_{i2})$, with $\mathbf{X}_{ik} = (X_{ik1}, \ldots, X_{ik7})^\top$ and $\mathbf{Z}_{i2} = (Z_{i21}, \ldots, Z_{i27})^\top$ where $\mathbf{1}_m$ represents a $1 \times m$ vector of ones. Each vector, X_{ikj} and Z_{i2j} are generated independently of a uniform distribution, thus $X_{ikj} \sim U(0, 1)$ and $Z_{i2j} \sim U(1, 3)$, $i = 1, \ldots, n$, $k = 2, 3$ and $j = 1, \ldots, 7$. These sets of values are fixed for all the replications. It is important to stress that, for the two simulation, we consider the parameter ν as fixed at 4. Thus, the parameter vector is $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top, \boldsymbol{\phi}^\top)^\top$.

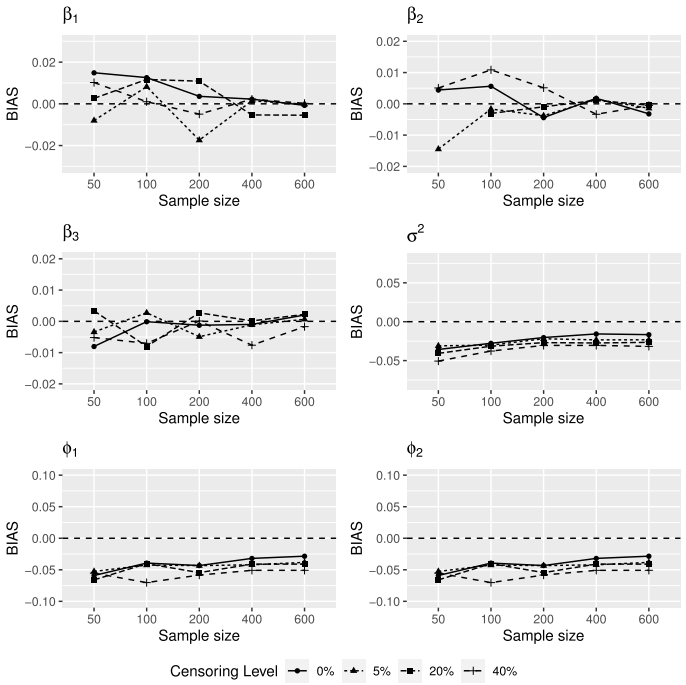


Figure 1: Simulation study 1. Average Bias of parameter estimates under the AR(2)-t-LMEC model.

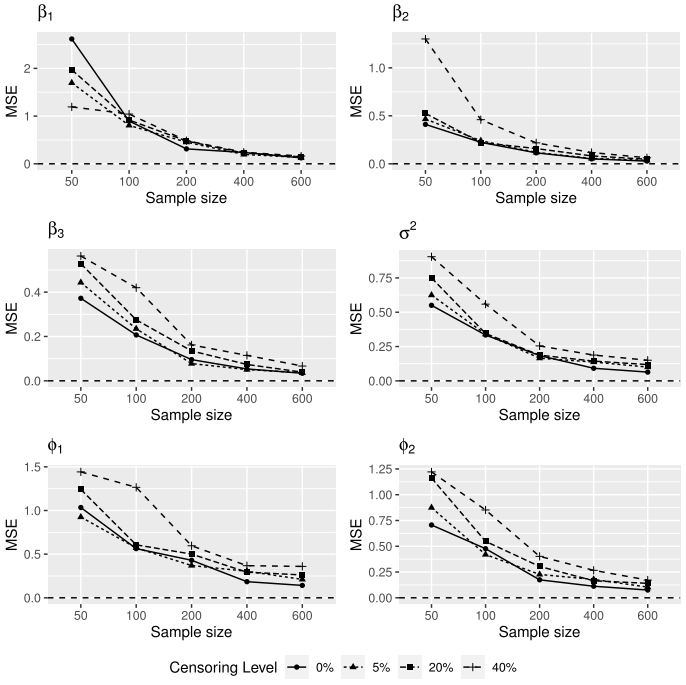


Figure 2: Simulation study 1. Average MSE of parameter estimates under the AR(2)-t-LMEC model.

4.1 Simulation Study 1

We generated $R = 100$ datasets from the AR(2)-t-LMEC model, considering each one of the combinations, of different sample sizes $n \in \{50, 100, 200, 400, 600\}$ and censoring levels $l\% \in \{0\%, 5\%, 20\%, 40\%\}$. In each combination, were fitted the AR(2)-t-LMEC and AR(2)-N-LMEC models. Thus, for both models, we computed the bias ($Bias$) and mean squared error (MSE) of the estimates obtained. These measures are defined by:

$$Bias(\widehat{\theta}_i) = \frac{1}{R}\sum_{j=1}^{R}\left(\widehat{\theta}_i^{(j)} - \theta_i\right) \quad \text{and}$$

$$MSE(\widehat{\theta}_i) = \frac{1}{R}\sum_{j=1}^{R}\left(\widehat{\theta}_i^{(j)} - \theta_i\right)^2,$$

where $\widehat{\theta}_i^{(j)}$ denotes the ML estimate of θ_i, for the j-th replication, $j = 1, \ldots, R$.

From Figures 1–2, we observe as a general rule that, when the distribution's assumption is correct, the $Bias$ and MSE tend to zero when the sample size increases, indicating that the ML estimates based on the proposed EM-type algorithm do provide good asymptotic properties, under the t-LMEC model. However, from Figures 3–4, we noted that when the distribution's assumption is incorrect, the ML estimates do not provide good asymptotic properties; for example, the Bias of $\widehat{\phi}_1$ under the AR(2)-N-LMEC model, departs from zero when the sample size increases.

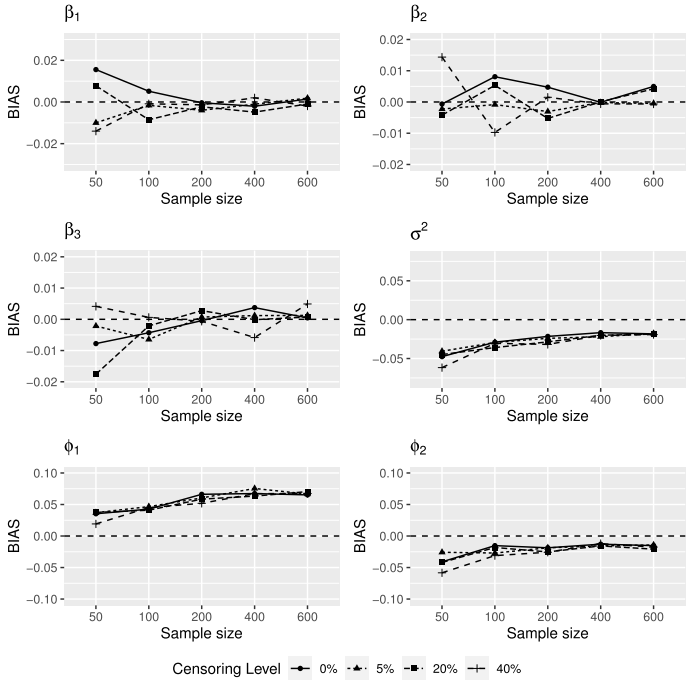


Figure 3: Simulation study 1. Average Bias of parameter estimates under the AR(2)-N-LMEC model.

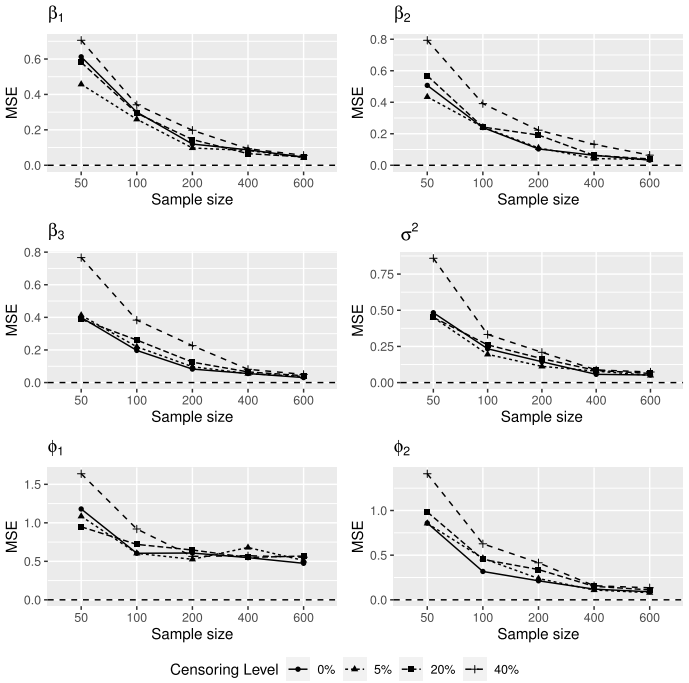


Figure 4: Simulation study 1. Average MSE of parameter estimates under the AR(2)-N-LMEC model.

4.2 Simulation Study 2

In this simulation study, we consider sample sizes $n \in \{40, 100\}$ and generated $R = 100$ Monte Carlo samples from the AR(2)-N-LMEC and AR(2)-t-LMEC models, with different censoring levels $l\% \in \{0\%, 5\%, 15\%, 25\%, 40\%\}$. We analyzed the SE for the $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)$ parameter vector (MC-SE), the averages values of the standard errors computed, using the empirical information matrix (MC-IM-SE) described in Section 3.5, and the percentage of times in the R samples, that the 95% confidence interval contained the true parameter values (MC-COV), assuming asymptotic normality. These measures are defined by:

$$\text{MC-IM-SE}(\widehat{\boldsymbol{\beta}_i}) = \frac{1}{R}\sum_{j=1}^{R} \widehat{SE(\boldsymbol{\beta}_i)}^{(j)}$$

and

$$\text{MC-SE}(\widehat{\boldsymbol{\beta}_i}) = \sqrt{\frac{1}{R-1}\left(\sum_{j=1}^{R}\left(\widehat{\boldsymbol{\beta}_i}^{(j)}\right)^2 - \left(\frac{1}{R}\sum_{j=1}^{R}\widehat{\boldsymbol{\beta}_i}^{(j)}\right)^2\right)},$$

where for $j = 1, \ldots, R$, we have that $\widehat{\boldsymbol{\beta}_i}^{(j)}$ and $\widehat{SE(\boldsymbol{\beta}_i)}^{(j)}$ represent the ML estimates of parameter $\boldsymbol{\beta}_i$ and the SE estimate of $\widehat{\boldsymbol{\beta}_i}^{(j)}$, respectively.

From Tables 1 and 2, we notice that the MC-IM-SE and MC-SE for the $\boldsymbol{\beta}$ parameter vector, from the t-LMEC model are smaller than the corresponding MC-IM-SE and MC-SE, from the N-LMEC model. Besides, the MC-COV under both models are similar.

Table 1. Standard errors of parameter estimates (MC-SE), average values of the standard errors (MC-IM-SE) and MC-COV, under t-MLE model with $n = 40$.

Censoring		AR(2)-N-LMEC		
Levels	Criteria	β_1	β_2	β_3
0%	MC-IM-SE	0.2106	0.2298	0.2160
	MC-SE	0.1648	0.1791	0.1604
	MC-COV	0.98	0.97	0.98
5%	MC-IM-SE	0.2220	0.2151	0.2066
	MC-SE	0.1840	0.1601	0.1770
	MC-COV	0.97	0.96	0.96
15%	MC-IM-SE	0.2136	0.2155	0.2100
	MC-SE	0.1538	0.1661	0.1611
	MC-COV	1	0.96	0.97
25%	MC-IM-SE	0.2177	0.2161	0.2076
	MC-SE	0.2009	0.1708	0.1840
	MC-COV	0.96	0.98	0.93
40%	MC-IM-SE	0.2569	0.2559	0.2551
	MC-SE	0.2050	0.2440	0.2124
	MC-COV	0.96	0.95	0.96
Censoring		AR(2)-t-LMEC		
Levels	Criteria	β_1	β_2	β_3
0%	MC-IM-SE	0.1469	0.1600	0.1467
	MC-SE	0.1366	0.1450	0.1338
	MC-COV	0.95	0.98	0.96
5%	MC-IM-SE	0.1654	0.1575	0.1567
	MC-SE	0.1543	0.1418	0.1492
	MC-COV	0.94	0.95	0.97
15%	MC-IM-SE	0.1834	0.1811	0.1743
	MC-SE	0.1443	0.1471	0.1315
	MC-COV	0.98	0.97	0.98
25%	MC-IM-SE	0.2053	0.1973	0.2004
	MC-SE	0.1636	0.1451	0.1439
	MC-COV	0.98	0.97	0.96
40%	MC-IM-SE	0.2525	0.2545	0.2503
	MC-SE	0.1690	0.1774	0.1781
	MC-COV	0.98	0.94	0.93

5. APPLICATION

In this Section, we study the viral load data of patients infected with the human immunodeficiency virus type 1 (HIV-1), from the clinical trial "A5055", previously analyzed by [32]. This clinical trial consists of 44 HIV-1 infected patients that were treated with one of the two powerful antiretroviral therapies: (i) IDV 800 mg and RTV 200 mg, twice a day every 12 hs. and (ii) IDV 400 mg and RTV 400 mg, twice a day every 12 hs. (see [33] for more details). Were measured the number of RNA copies in the blood plasma (viral load) and the immunological marker of T cells of differentiation groups 4 (CD4), at different times of the treat. It is

Table 2. Standard errors of parameter estimates (MC-SE), average values of the standard errors (MC-IM-SE) and MC-COV, under t-MLE model with $n = 100$.

Censoring		AR(2)-N-LMEC		
Levels	Criteria	β_1	β_2	β_3
0%	MC-IM-SE	0.1267	0.1296	0.1231
	MC-SE	0.1159	0.1240	0.1242
	MC-COV	0.94	0.96	0.95
5%	MC-IM-SE	0.1144	0.1151	0.1241
	MC-SE	0.1040	0.1063	0.1201
	MC-COV	0.95	0.97	0.95
15%	MC-IM-SE	0.1149	0.1240	0.1208
	MC-SE	0.1003	0.1052	0.1062
	MC-COV	0.97	0.97	0.99
25%	MC-IM-SE	0.1259	0.1257	0.1245
	MC-SE	0.1103	0.1315	0.1151
	MC-COV	0.99	0.95	0.96
40%	MC-IM-SE	0.1446	0.1412	0.1371
	MC-SE	0.1144	0.1260	0.1271
	MC-COV	0.97	0.98	0.96
Censoring		AR(2)-t-LMEC		
Levels	Criteria	β_1	β_2	β_3
0%	MC-IM-SE	0.0915	0.0960	0.0913
	MC-SE	0.0938	0.0923	0.0882
	MC-COV	0.97	0.95	0.96
5%	MC-IM-SE	0.1012	0.1006	0.1041
	MC-SE	0.0938	0.0962	0.0939
	MC-COV	0.96	0.97	0.97
15%	MC-IM-SE	0.1096	0.1102	0.1145
	MC-SE	0.1003	0.0978	0.0941
	MC-COV	0.97	0.97	0.97
25%	MC-IM-SE	0.1257	0.1272	0.1249
	MC-SE	0.0932	0.1133	0.1021
	MC-COV	0.99	0.96	0.96
40%	MC-IM-SE	0.1495	0.1478	0.1554
	MC-SE	0.0907	0.0968	0.1085
	MC-COV	0.98	0.99	0.98

important to note that the viral loads were not necessarily measured at equally spaced intervals of days.

Following [33], we investigate the longitudinal evolution of RNA viral load (in log-base-10 scale) of the 44 patients, during antiretroviral treatment. We focus on analyzing only 42 HIV-1 infected patients; two patients were excluded from the analysis because both of them dropped out very early (# 4 and # 8 with 3 and 1 measurements, respectively).

The trajectories of transformed RNA viral load along with the detection limit by plotting dotted red lines are shown in Figure 5. The lower detection limit for RNA viral load is 50 copies/ml, thus 106 out of 312 viral load measurements, below the detection limit, are considered to be censoring (approximately 34%). We start by presenting the

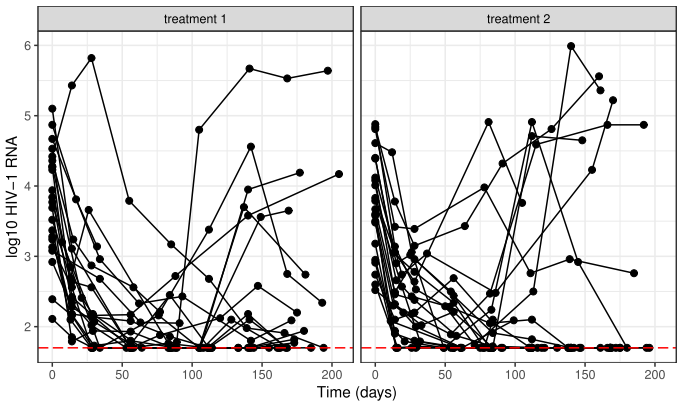


Figure 5: A5055 Data. Trajectories of $\log_{10}(RNA)$ for 42 HIV-1 infected patients who were randomized to one of two treatment regimens.

results obtained by fitting the LMEC model defined in (3.1) and (3.3), as follows:

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

with

$$\begin{aligned}
&\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^\top,\\
&\mathbf{X}_i = (\mathbf{1}_{n_i}^\top, \mathbf{t}_i^\top, \text{Treat}_i \times \mathbf{1}_{n_i}^\top, \text{CD4}^{1/2} \times \mathbf{1}_{n_i}^\top, \text{Treat}_i \times \mathbf{t}_i^\top),\\
&\mathbf{Z}_i = (\mathbf{1}_{n_i}^\top, \mathbf{t}_i^\top),\ \boldsymbol{\beta} = (\beta_0, \ldots, \beta_4)^\top,\ \mathbf{b}_i = (b_{i0}, b_{i1})^\top,\\
&\boldsymbol{\epsilon}_i = (\epsilon_1, \ldots, \epsilon_{n_i})^\top, \text{ for } j = 1, \ldots, n_i \text{ and } i = 1, \ldots, 42,
\end{aligned}$$

where

- $\mathbf{1}_m$ represents a $1 \times m$ vector of ones and $\mathbf{t}_i = (t_{i1}, \ldots, t_{in_i})$.
- Y_{ij} is the is the log10-transformation of the viral load of subject "i" at time t_{ij} (days).
- Treat_i is a treatment indicator (i.e. 0 for the treatments 1 and 1 for the treatments 2).
- $\text{CD4}_{ij}^{1/2}$ represents the square root of CD4 counts of patient i, measured at time t_{ij}.
- The random intercept and random slope are represent by b_{i0} and b_{i1}, respectively.

For comparison purposes, we fitted the N-LMEC and the t-LMEC counterpart with the structure dependency errors discussed in Section 3.2. Thus,

(1) **Assuming the N-LMEC model.**
We made a preliminary analysis using unconditional independence (UNC) in the random effects and random errors, i.e., the model proposed by [30]. The scatter plot of the estimated random intersections (b_0 effect) and random slopes (b_1 effect), presented in Figure 6(a), suggests that there is a possible linear relationship between both of them. We can suspect that there is a dependency structure for the random effect.

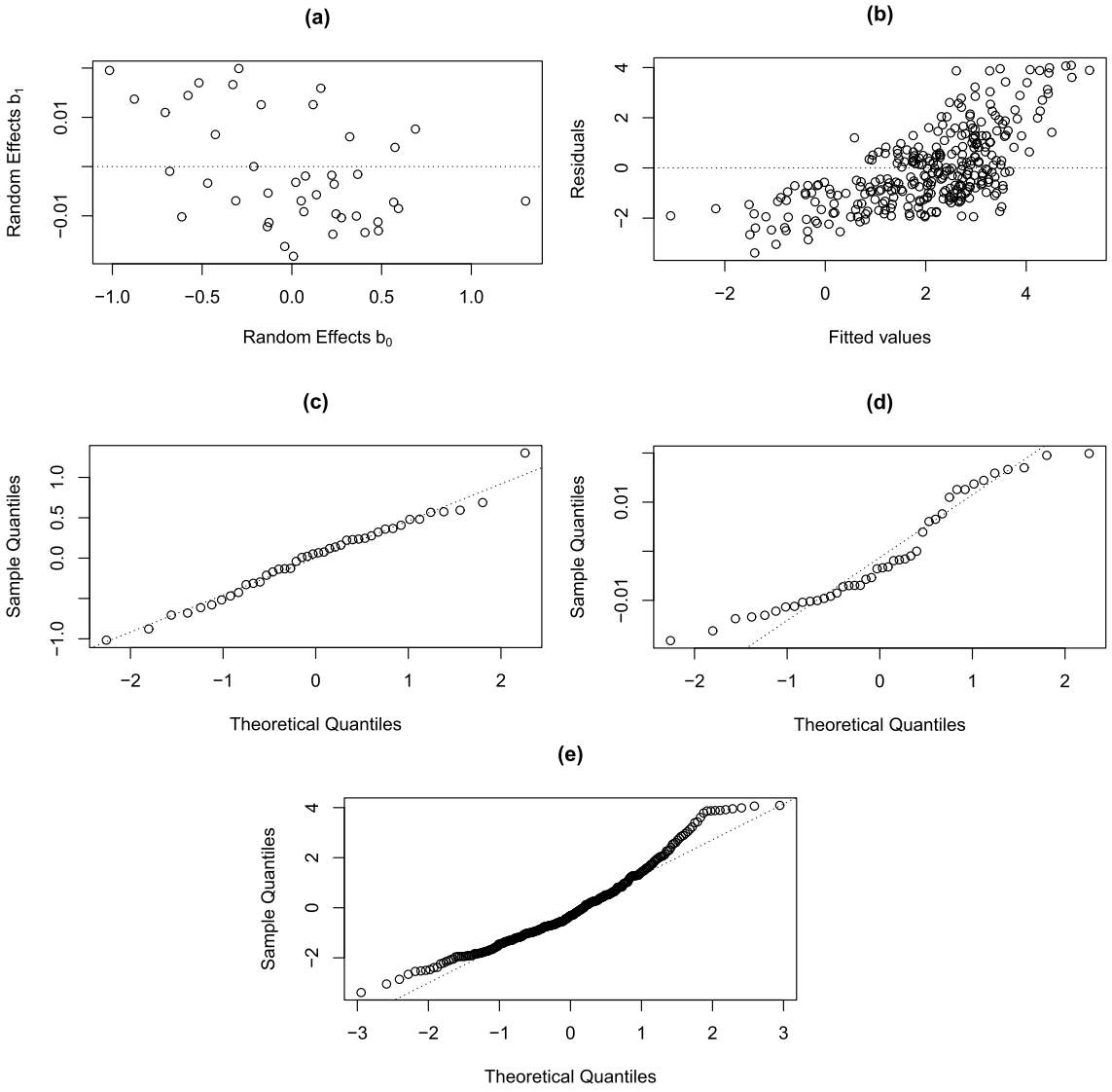


Figure 6: A5055 Data. Scatter diagrams of the random effect and marginal errors and Q-Q plots of estimates of random intercepts and slopes and marginal residuals for log10(RNA) using the N-LMEC model with conditional independence in the random effects and the random errors.

From Figure 6(b), we observe that, apparently, the behavior of the marginal residuals is not totally random, highlighting a trend. Finally, in Figure 6(c)–(e), we show the QQ-plots for the b_0 and b_1 effect and the standard marginal residuals. These figures clearly indicate that the normality assumption for random effects and within-subject errors may be unrealistic. Moreover, the variogram plot of the standard marginal residuals, presented in Figure 7, indicates a long-term autocorrelation, showing a possible intra-individual correlation over time.

Tables 3 presents the results for the "A5055" dataset considering the N-LMEC model, under the different correlation structures. We observed that only the estimation of σ^2 parameters is affected by the choice of the correlation structure.

(2) **Assuming the t-LMEC model.**
In order to address the possible serial correlation among within-subject errors, we fit several models: i) the unconditionally independent (**UNC**), ii) the damped exponential correlation (**DEC**), and iii) the continuous-type autoregressive process (**AR**) of order $p \in \{1, 2, 3\}$ cases. It is important to stress that considering the DEC structure it is possible to obtain different correlation structures, for instance, the symmetry (**SYM**) and **DEC-AR(1)**.

From Table 4, we observed that there are no significant differences in the estimation of the fixed and random

Table 3. A5055 Data. Summary of parameter estimates and the standard errors of fixed effects (in parentheses) under the N-LMEC model, with different correlation structures for the errors.

Parameter estimates	Correlation structures						
		Damped exponential			AR(p)		
	UNC	DEC	DEC-AR(1)	SYM	AR(1)	AR(2)	AR(3)
β_0	4.383	4.379	4.378	4.388	4.384	4.382	4.287
	(0.533)	(0.002)	(0.002)	(0.010)	(0.540)	(0.553)	(0.587)
β_1	−0.004	−0.004	−0.004	−0.004	−0.004	−0.004	−0.003
	(0.009)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)
β_2	0.305	0.304	0.303	0.304	0.307	0.308	0.361
	(0.309)	(0.058)	(0.058)	(0.013)	(0.310)	(0.309)	(0.316)
β_3	−0.111	−0.111	−0.111	−0.112	−0.111	−0.111	−0.113
	(0.031)	(0.010)	(0.010)	(0.008)	(0.032)	(0.032)	(0.034)
β_4	−0.003	−0.003	−0.003	−0.003	−0.003	−0.003	−0.003
	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.006)
σ^2	0.733	0.733	0.732	0.765	0.427	0.417	0.193
	(–)	(–)	(–)	(–)	(–)	(–)	(–)
α_{11}	0.383	0.382	0.381	0.352	0.382	0.384	0.294
	(–)	(–)	(–)	(–)	(–)	(–)	(–)
α_{12}	−0.004	−0.004	−0.004	−0.004	−0.004	−0.004	−0.003
	(–)	(–)	(–)	(–)	(–)	(–)	(–)
α_{22}	>0.001	>0.001	>0.001	>0.001	>0.001	>0.001	>0.001
	(–)	(–)	(–)	(–)	(–)	(–)	(–)
ϕ_1	–	0.998	1.000	0.000	0.652	0.762	0.273
	(–)	(–)	(–)	(–)	(–)	(–)	(–)
ϕ_2	–	0.106	1	0	–	−0.484	0.102
	(–)	(–)	(–)	(–)	(–)	(–)	(–)
ϕ_1	–	–	–	–	–	–	−0.820
	(–)	(–)	(–)	(–)	(–)	(–)	(–)

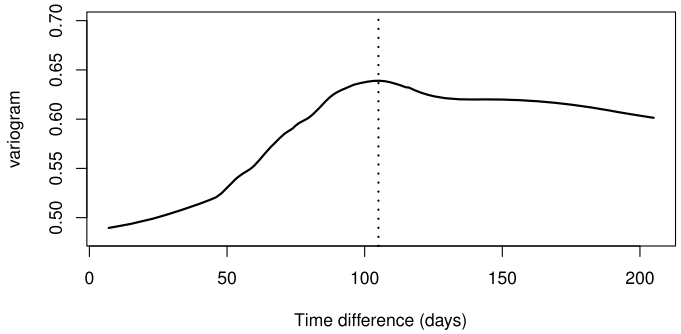


Figure 7: A5055 Data. Variogram from the marginal residuals for log10(RNA) using the N-LMEC model with conditional independence in the random effects and the random errors.

effects parameters under the different correlation structures. On the other hand, the estimation of the ν and σ^2 parameters are affected by the choice of the correlation structure.

Note from Figure 8(a)–(b) that the inclusion of a correlation structure improves the model fit. The QQ plots for the b_0, b_1 effect, and standard marginal residuals, under the t-LMEC model with DEC-AR(1) correlation structure, presented in Figure 8(c)–(e) show a significant improvement, with respect to fit obtained under the N-LMEC model. On the other hand, in Figure 9(a)–(b), we present the simulated envelope plots of Marginal residuals, from fitting the DEC(AR)-N-LMEC and DEC(AR)-t-LMEC models. We observed that some marginal residuals under the DEC(AR)-N-LMEC model are outside the boundary of the envelope, indicating the inadequacy of this model. However, the marginal residuals corresponding to the DEC(AR)-t-LMEC model are within the boundary of the envelope. This shows that although we consider the DEC(AR) correlation structure, a correct specification of the data distribution assumption is essential. Finally, as suggested by [19], we performed the likelihood ratio test to examine the appropriateness of using DEC(AR) dependency structure errors versus UNC errors under the t-LMEC model, obtaining a p-value of 0.0058. As expected, this result indicates the existence of strong dependence among within-patient errors across occasions under the t-LMEC model.

On the other hand, Table 5 shows the summaries of the log-likelihood, AIC and BIC criteria for t-LMEC and N-LMEC models, respectively, under the different autocorrelation structures. We observe that t-LMEC consistently out-

Table 4. A5055 Data. Summary of parameter estimates and the standard errors of fixed effects (in parentheses)under the t-LMEC model, with different correlation structures for the errors.

Parameter estimates	Correlation structures						
		Damped exponential			AR(p)		
	UNC	DEC	DEC-AR(1)	SYM	AR(1)	AR(2)	AR(3)
β_0	3.995	3.998	4.051	4.001	3.988	3.987	4.080
	(0.604)	(0.509)	(0.531)	(0.513)	(0.607)	(0.606)	(0.643)
β_1	−0.005	−0.005	−0.004	−0.005	−0.005	−0.005	−0.004
	(0.008)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.009)
β_2	0.350	0.350	0.385	0.345	0.354	0.355	0.353
	(0.309)	(0.250)	(0.264)	(0.253)	(0.308)	(0.309)	(0.323)
β_3	−0.086	−0.086	−0.092	−0.086	−0.086	−0.086	−0.094
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.021)
β_4	−0.003	−0.003	−0.004	−0.004	−0.004	−0.004	−0.004
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
σ^2	0.473	0.476	0.510	0.507	0.286	0.282	0.175
	(–)	(–)	(–)	(–)	(–)	(–)	(–)
α_{11}	0.220	0.222	0.210	0.194	0.218	0.218	0.222
	(–)	(–)	(–)	(–)	(–)	(–)	(–)
α_{12}	−0.002	−0.002	−0.002	−0.002	−0.002	−0.002	−0.002
	(–)	(–)	(–)	(–)	(–)	(–)	(–)
α_{22}	>0.001	>0.001	>0.001	>0.001	>0.001	>0.001	>0.001
	(–)	(–)	(–)	(–)	(–)	(–)	(–)
ϕ_1	–	0.105	0.899	0.066	0.634	0.731	0.276
	(–)	(–)	(–)	(–)	(–)	(–)	(–)
ϕ_2	–	0.998	1	0	–	−0.421	0.109
	(–)	(–)	(–)	(–)	(–)	(–)	(–)
ϕ_3	–	–	–	–	–	–	−0.789
	(–)	(–)	(–)	(–)	(–)	(–)	(–)
ν	3.382	3.407	3.953	3.304	3.247	3.296	4.915
	(–)	(–)	(–)	(–)	(–)	(–)	(–)

Table 5. A5055 Data. Summary of selection criteria for the t-LMEC and N-LMEC model under different correlation structures.

Model	Criteria	Correlation structures						
			Damped exponential			AR(p)		
		UNC	DEC	DEC-AR(1)	SYM	AR(1)	AR(2)	AR(3)
t-LMEC	loglik	−353.840	−353.877	**−349.323**	−353.696	−353.439	−353.456	−352.021
	AIC	729.681	731.755	**720.646**	729.393	728.878	730.913	730.042
	BIC	770.854	776.671	**761.819**	770.566	770.0517	775.829	778.701
N-LMEC	loglik	−362.571	−362.643	−362.671	−362.569	−362.343	−362.385	**−355.407**
	AIC	743.142	747.285	745.342	745.139	744.686	746.771	**734.815**
	BIC	776.829	788.458	782.773	782.569	782.116	787.944	**779.731**

performs the normal counterpart in all cases. In particular, these criteria indicate a preference of the unspecified correlation structure (DEC(AR)), i.e., the estimated correlation structure of the error is given by:

$$\widehat{\mathbf{\Omega}}_i = \widehat{\sigma^2}\widehat{\mathbf{R}}_i = \widehat{\sigma^2}\big[\widehat{\phi}_1^{|t_{ij}-t_{ik}|}\big] = 0.51\big[0.899^{|t_{ij}-t_{ik}|}\big].$$

Thus, by the previous analysis, we conclude that the t-LMEC model with DEC(AR) correlation structure for the error, is the most appropriate model for these dataset.

All the computational procedures were realized using a 64-bit Windows environment on a notebook machine with a 1.99 GHz Intel Core i7 processor with 12GB of RAM, the EM algorithm took approximately 9.8 min to converge by each model.

6. CONCLUSIONS

In this paper, we proposed a robust linear mixed effect model to analyze longitudinal censored or missing data under the multivariate Student's t-distribution with within-subject correlation, considering some useful dependence structures. In practical implementation, the EM algorithm is used to obtain the ML estimates of the model parameters.

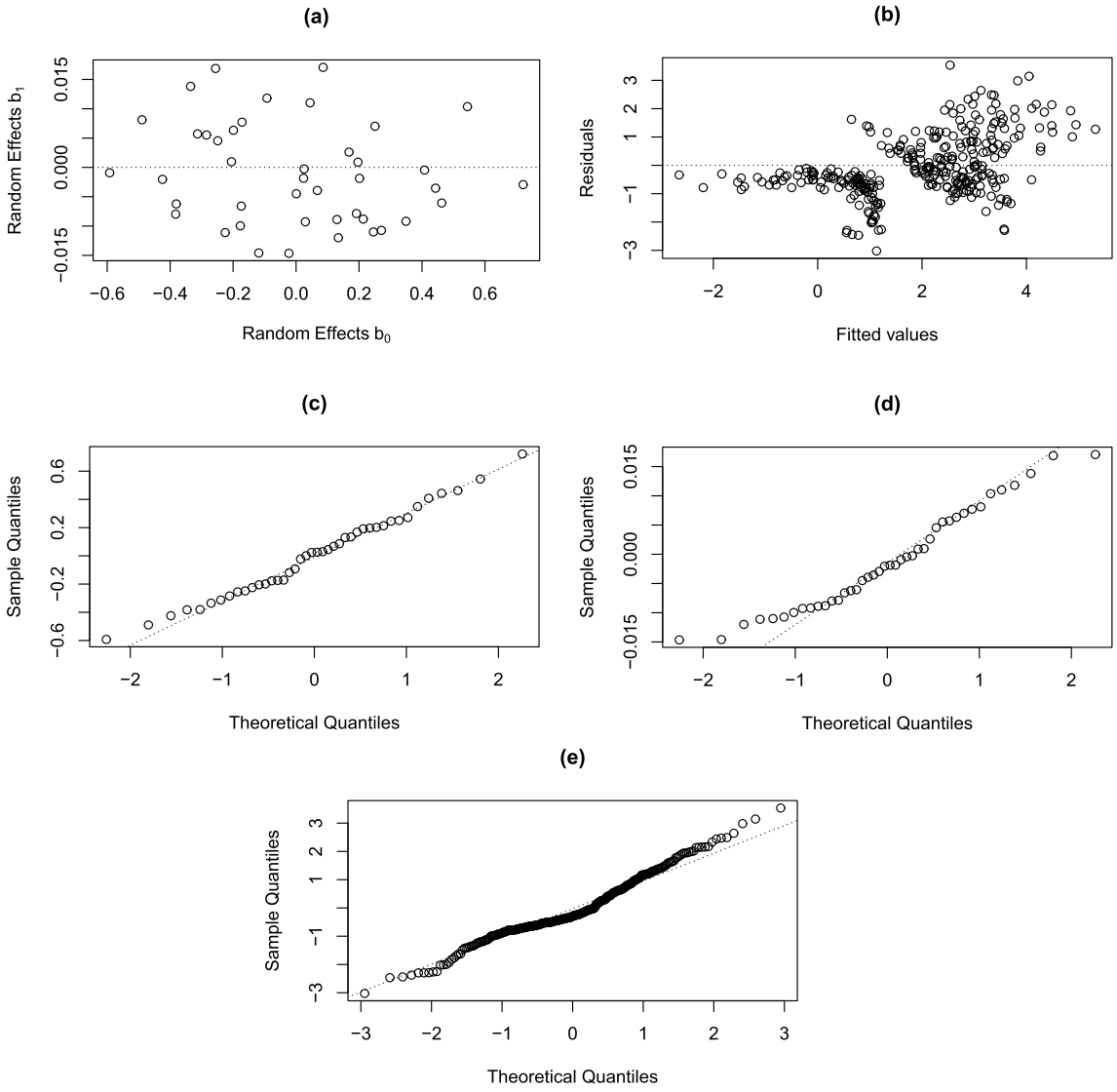


Figure 8: A5055 Data. Scatter diagrams of the random effects and marginal errors and Q-Q plots of estimates of random effects and marginal residuals for log10(RNA) using the DEC(AR)-t-LMEC model.

The model proposed in this manuscript is an extension of [21], [22] and [26] by developing some additional tools for robust inferences in practical data analysis.

Two simulation studies were performed in order to evaluate the proposed model. The simulation studies validate our method's performance and indicate an efficiency gain of the t-LMEC model over the N-LMEC model when data present heavy tails. A real data set, previously analyzed with a conditional independent structure, is analyzed. The results show the t-LMEC model with DEC-AR(1) correlation structure fits the data better. The proposed methods were implemented as part of the new `R` package **ARpLMEC**, which is available for download at the CRAN repository. Moreover, we have prepared a Rmarkdown file which contains the codes to reproduce the results of the application and it is available with this manuscript via GitHub https://github.com/hlachos/tlmec/.

Although the t-LMEC model considered here has shown great flexibility for modeling symmetric data with evidence of heavier tails than the normal distributions, its robustness against outliers can be seriously affected by the presence of skewness. Thus, it is of interest to generalize the t-LMEC model by considering a more flexible family of distributions, such as the scale mixtures of skew-normal (SMSN) distributions [14], to accommodate the censoring, skewness and heavy tails simultaneously. It is also of interest to develop effective algorithms for non-linear censored mixed effect (NLMEC) models or semiparametric structures in order to accommodate substantial non-linearity trends observed in the data.

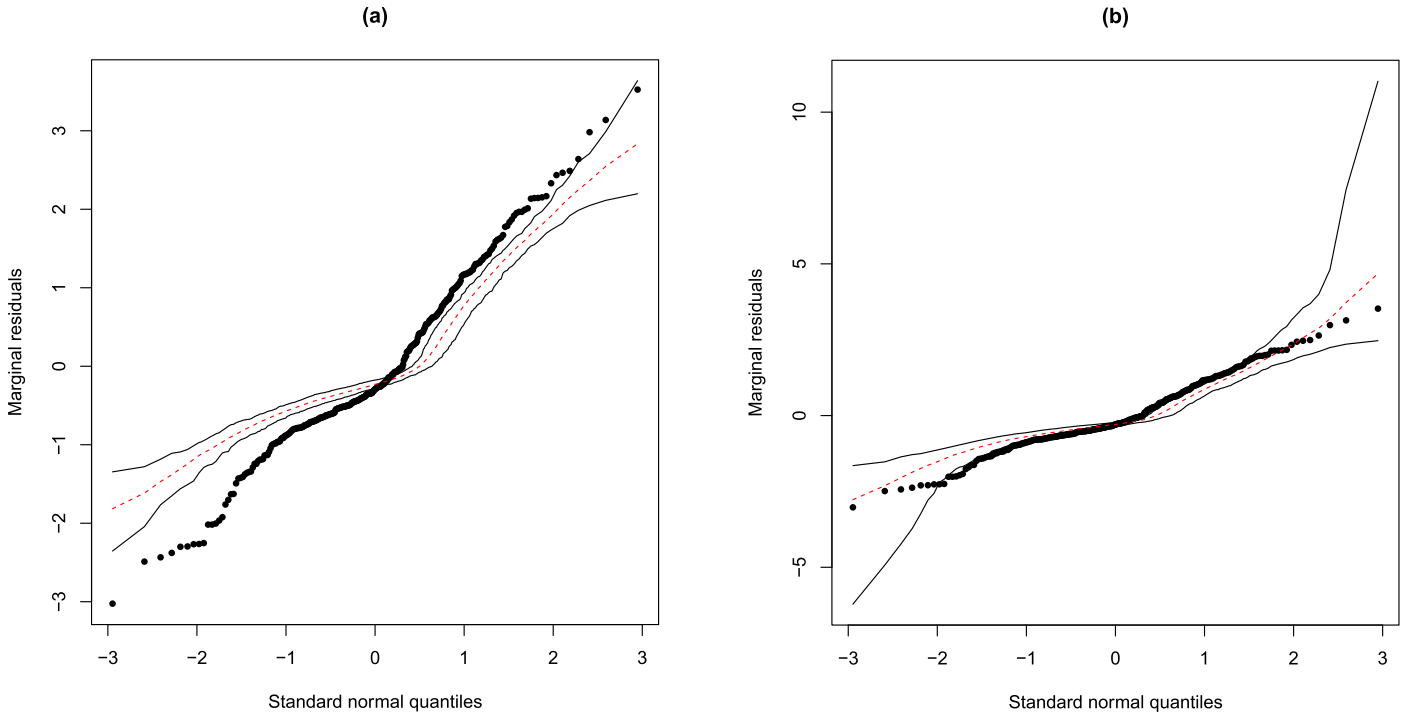


Figure 9: A5055 Data. Simulated envelope of the Marginal residuals from the fitted models: (a) DEC(AR)-N-LMEC model and (b) DEC(AR)-t-LMEC model.

Another possible to explore is to study models that estimate the conditional heteroscedasticity of dataset, as ARCH [12], GARCH [5, 6] models, among others. Finally, another interesting avenue for research is to propose methods within a unified framework in LMEC/NLMEC models for checking random-effects distribution following [1] for generalized linear mixed models and [10] for non-linear mixed models. An in-depth investigation of such extensions is beyond the scope of the present paper, but these are interesting topics for further research.

APPENDIX A. p-VARIATE STUDENT'S-t

The following properties of the p-variate Student's t-distribution are useful for our theoretical developments. We start with the marginal-conditional decomposition of a p-variate Student's t random vector. The proof of the following propositions can be found in [2].

Proposition 1. *Let* $\mathbf{Y} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ *partitioned as* $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ *with* $dim(\mathbf{Y}_1) = p_1$, $dim(\mathbf{Y}_2) = p_2$, *where* $p_1 + p_2 = p$. *Let* $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top$ *and* $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$ *be the corresponding partitions of* $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$, *respectively. Then, we have*

(*i*) $\mathbf{Y}_1 \sim t_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \nu)$; *and*
(*ii*) *The conditional distribution of* $\mathbf{Y}_2 \mid (\mathbf{Y}_1 = \mathbf{y}_1)$ *is given by*

$$\mathbf{Y}_2 \mid (\mathbf{Y}_1 = \mathbf{Y}_1) \sim t_{p_2}(\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}_{22.1}, \nu + p_1),$$

where $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$ *and* $\widetilde{\boldsymbol{\Sigma}}_{22.1} = (\frac{\nu+\delta_1}{\nu+p_1})\boldsymbol{\Sigma}_{22.1}$ *with* $\delta_1 = (\mathbf{y}_1 - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$ *and* $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$.

Proposition 2. *Let* $\mathbf{Y} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. *Then for any fixed vector* $\mathbf{b} \in \mathbb{R}^m$ *and matrix* $\mathbf{A} \in \mathbb{R}^{m \times p}$ *of full rank we get*

$$\mathbf{V} = \mathbf{b} + \mathbf{A}\mathbf{Y} \sim t_m\big(\mathbf{b} + \mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top, \nu\big).$$

APPENDIX B. TMVT DISTRIBUTIONS

The following propositions are related to the marginal and conditional moments of the first two moments of the TMVT distributions under a double truncation. The proof are similar to those given in [21]. In what follows, we shall use the notation $\mathbf{Y}^{(0)} = 1$, $\mathbf{Y}^{(1)} = \mathbf{Y}$, $\mathbf{Y}^{(2)} = \mathbf{Y}\mathbf{Y}^\top$ and $\mathbf{W} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; (\mathbf{a}, \mathbf{b}))$ stands for a p-variate doubly truncated Student's-t distribution on $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^p$.

Proposition 3. *If* $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; (\mathbf{a}, \mathbf{b}))$ *then it follows that*

$$\mathbb{E}\left[\left(\frac{\nu + p}{\nu + \delta(\mathbf{Y})}\right)^r \mathbf{Y}^{(k)}\right] = c_p(\nu, r)\frac{L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \nu + 2r)}{L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}\mathbb{E}\big[\mathbf{W}^{(k)}\big],$$

where

$$c_p(\nu, r) = \left(\frac{\nu+p}{\nu}\right)^r \frac{\Gamma(\frac{p+\nu}{2})\Gamma(\frac{\nu+2r}{2})}{\Gamma(\frac{\nu}{2})\Gamma(\frac{p+\nu+2r}{2})},$$

$\boldsymbol{\Sigma}^* = \frac{\nu}{\nu+2r}\boldsymbol{\Sigma}$ *and* $\nu + 2r > 0$ *with* $\mathbf{W} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \nu + 2r; (\mathbf{a}, \mathbf{b}))$.

Having established the formula on the k-order moment of $\mathbf{Y}$, we provide an explicit formula for the conditional moments with respect to a two-component partition of $\mathbf{Y}$.

Proposition 4. *Let* $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; (\mathbf{a}, \mathbf{b}))$. *Consider the partition* $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ *with* $\dim(\mathbf{Y}_1) = p_1$, $\dim(\mathbf{Y}_2) = p_2$, $p_1 + p_2 = p$, *and the corresponding partitions:* $\mathbf{a} = (\mathbf{a}_1^\top, \mathbf{a}_2^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \mathbf{b}_2^\top)^\top$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top$ *and* $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$. *Then,*

$$\mathbb{E}\left[\left(\frac{\nu+p}{\nu+\delta(\mathbf{Y})}\right)^r \mathbf{Y}_2^{(k)} \mid \mathbf{Y}_1\right] = \frac{d_p(p_1, \nu, r)}{(\nu+\delta(\mathbf{Y}_1))^r} \times \frac{L_{p_2}(\mathbf{a}_2, \mathbf{b}_2; \boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}^*_{22.1}, \nu + p_1 + 2r)}{L_{p_2}(\mathbf{a}_2, \mathbf{b}_2; \boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}_{22.1}, \nu + p_1)} \times \mathbb{E}\left[\mathbf{W}_2^{(k)}\right],$$

for $\nu + p_1 + 2r > 0$, *with* $\delta(\mathbf{Y}_1) = \delta(\mathbf{Y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$,

$$\widetilde{\boldsymbol{\Sigma}}^*_{22.1} = \left(\frac{\nu+\delta_1}{\nu+2r+p_1}\right)\boldsymbol{\Sigma}_{22.1} \quad \textit{and}$$

$$d_p(p_1, \nu, r) = (\nu+p)^r \frac{\Gamma(\frac{p+\nu}{2})\Gamma(\frac{p_1+\nu+2r}{2})}{\Gamma(\frac{p_1+\nu}{2})\Gamma(\frac{p+\nu+2r}{2})},$$

where $\boldsymbol{\Sigma}_{22.1}$ *is defined as in Proposition 1. Moreover,* $\mathbf{W}_2 \sim \mathrm{T}t_{p_2}(\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}^*_{22.1}, \nu + p_1 + 2r; (\mathbf{a}_2, \mathbf{b}_2))$.

ACKNOWLEDGEMENTS

The authors thank the editor and anonymous referees for their valuable comments and suggestions, which significantly contributed to the improvement of this paper.

FUNDING

The research of Aldo M. Garay was supported by Grant APQ-0950-1.02/22 from Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco FACEPE - Brazil and by Grant 441476/2023-6 from National Council for Scientific and Technological Development – CNPq – Brazil. Victor H. Lachos acknowledges the partial financial support from the Office of the Vice President for Research and UConn - CLAS's Summer Research Funding Initiative 2023.

Accepted 6 July 2024

REFERENCES

[1] Abad, A. A., Litière, S. and Molenberghs, G. (2010). Testing for misspecification in generalized linear mixed models. *Biostatistics* **11**(4) 771–786.

[2] Arellano-Valle, R. B. and Bolfarine, H. (1995). On some characterizations of the t-distribution. *Statistics & Probability Letters* **25**(1) 79–85. https://doi.org/10.1016/0167-7152(94)00208-P. MR1364821

[3] Barndorff-Nielsen, O. and Schou, G. (1973). On the parametrization of autoregressive models by partial autocorrelations. *Journal of Multivariate Analysis* **3**(4) 408–419. https://doi.org/10.1016/0047-259X(73)90030-4. MR0343510

[4] Bartolucci, F. and Farcomeni, A. (2009). A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. *Journal of the American Statistical Association* **104**(486) 816–831. https://doi.org/10.1198/jasa.2009.0107. MR2751454

[5] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**(3) 307–327. https://doi.org/10.1016/0304-4076(86)90063-1. MR0853051

[6] Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics* **69**(3) 542–547.

[7] Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco. MR0436499

[8] Box, G. E. P., Jenkins, G. M., Reinsel, G. C. and Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons. MR3379415

[9] Chi, E. M. and Reinsel, G. C. (1989). Models for longitudinal data with random effects and AR(1) errors. *Journal of the American Statistical Association* **84**(406) 452–459. MR1010333

[10] Drikvandi, R. (2020). Nonlinear mixed-effects models with misspecified random-effects distribution. *Pharmaceutical Statistics* **19**(3) 187–201.

[11] Drikvandi, R., Verbeke, G. and Molenberghs, G. (2017). Diagnosing misspecification of the random-effects distribution in mixed models. *Biometrics* **73**(1) 63–71. https://doi.org/10.1111/biom.12551. MR3632352

[12] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**(4) 987–1007. https://doi.org/10.2307/1912773. MR0666121

[13] Garay, A. M., Castro, L. M., Leskow, J. and Lachos, V. H. (2017). Censored linear regression models for irregularly observed longitudinal data using the multivariate-t distribution. *Statistical Methods in Medical Research* **26**(2) 542–566. https://doi.org/10.1177/0962280214551191. MR3635923

[14] Lachos, V. H., Ghosh, P. and Arellano-Valle, R. B. (2010). Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica* **20**(1) 303–322. MR2640696

[15] Lachos, V. H., A. Matos, L., Castro, L. M. and Chen, M.-H. (2019). Flexible longitudinal linear mixed models for multiple censored responses data. *Statistics in Medicine* **38**(6) 1074–1102. https://doi.org/10.1002/sim.8017. MR3916716

[16] Lin, T. I. and Lee, J. C. (2007). Bayesian analysis of hierarchical linear mixed modeling using the multivariate t distribution. *Journal of Statistical Planning and Inference* **137**(2) 484–495. https://doi.org/10.1016/j.jspi.2005.12.010. MR2298952

[17] Lin, T. I. and Lee, J. C. (2003). On modelling data from degradation sample paths over time. *Australian & New Zealand Journal of Statistics* **45**(3) 257–270. https://doi.org/10.1111/1467-842X.00282. MR1999510

[18] Lin, T. I. and Lee, J. C. (2006). A robust approach to t linear mixed models applied to multiple sclerosis data. *Statistics in Medicine* **25**(8) 1397–1412. https://doi.org/10.1002/sim.2384. MR2226794

[19] Lin, T.-I. and Wang, W.-L. (2020). Multivariate-t linear mixed models with censored responses, intermittent missing va-

lues and heavy tails. *Statistical Methods in Medical Research* **29**(5) 1288–1304. https://doi.org/10.1177/0962280219857103. MR4097145

[20] Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* **44**(2) 226–233. MR0676213

[21] Matos, L. A., Prates, M. O., Chen, M. H. and Lachos, V. H. (2013). Likelihood-based inference for mixed-effects models with censored response using the multivariate-t distribution. *Statistica Sinica* **23** 1323–1342. MR3114716

[22] Matos, L. A., Castro, L. M. and Lachos, V. H. (2016). Censored mixed-effects models for irregularly observed repeated measures with applications to HIV viral loads. *Test* **25**(4) 627–653. https://doi.org/10.1007/s11749-016-0486-2. MR3554408

[23] Matos, L. A., Bandyopadhyay, D., Castro, L. M. and Lachos, V. H. (2015). Influence assessment in censored mixed-effects models using the multivariate Student's t distribution. *Journal of Multivariate Analysis* **141** 104–117. https://doi.org/10.1016/j.jmva.2015.06.014. MR3390061

[24] Meng, X. L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **81** 633–648. https://doi.org/10.1093/biomet/80.2.267. MR1243503

[25] Muñoz, A., Carey, V., Schouten, J. P., Segal, M. and Rosner, B. (1992). A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics* **48** 733–742.

[26] Olivari, R. C., Garay, A. M., Lachos, V. H. and Matos, L. A. (2021). Mixed-effects models for censored data with autoregressive errors. *Journal of Biopharmaceutical Statistics* **31**(3) 273–294.

[27] Olivari, R. C., Zhong, K., Garay, A. M. and Lachos, V. H. (2022). ARpLMEC: Fitting autoregressive censored mixed-effects models. R package version 2.3. https://CRAN.R-project.org/package=ARpLMEC.

[28] Pinheiro, J. C., Liu, C. H. and Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using a multivariate t-distribution. *Journal of Computational and Graphical Statistics* **10** 249–276. https://doi.org/10.1198/10618600152628059. MR1939700

[29] Schumacher, F. L., Lachos, V. H. and Dey, D. K. (2017). Censored regression models with autoregressive errors: A likelihood-based perspective. *Canadian Journal of Statistics* **45**(4) 375–392. https://doi.org/10.1002/cjs.11338. MR3729976

[30] Vaida, F. and Liu, L. (2009). Fast implementation for normal mixed effects models with censored response. *Journal of Computational and Graphical Statistics* **18**(4) 797–817. https://doi.org/10.1198/jcgs.2009.07130. MR2750442

[31] Valeriano, K., Matos, L. A. and Morales, C. G. (2022). relliptical: The truncated elliptical family of distributions. R package version 1.1.0. https://CRAN.R-project.org/package=relliptical.

[32] Wang, W.-L. (2013). Multivariate t linear mixed models for irregularly observed multiple repeated measures with missing outcomes. *Biometrical Journal* **55**(4) 554–571. https://doi.org/10.1002/bimj.201200001. MR3079990

[33] Wang, W.-L., Lin, T.-I. and Lachos, V. H. (2018). Extending multivariate-t linear mixed models for multiple longitudinal data with censored responses and heavy tails. *Statistical Methods in Medical Research* **27**(1) 48–64. https://doi.org/10.1177/0962280215620229. MR3745654

Kelin Zhong. Department of Statistics, University of Connecticut, Storrs, USA. E-mail address: kelin.zhong@uconn.edu

Rommy C. Olivari. Department of Statistics, Federal University of Pernambuco, Brazil. E-mail address: rco1@de.ufpe.br

Aldo M. Garay. Department of Statistics, Federal University of Pernambuco, Prof. Moraes Rego, 1235 Cidade Universitaria, Recife, PE, CEP: 50670-901, Brazil. E-mail address: agaray@de.ufpe.br

Víctor H.Lachos. Department of Statistics, University of Connecticut, Storrs, USA. E-mail address: hlachos@uconn.edu