Censored Linear Regression Models for Irregularly Observed Longitudinal Data using the Multivariate-t Distribution

Aldo M. Garay[a], Luis M. Castro[b], Jacek Leskow[c] and Victor H. Lachos[a]

[a] *Departamento de Estatística, Universidade Estadual de Campinas, Brazil*

[b] *Departamento de Estatística, Universidad de Concepción, Chile*

[c] *Institute of Mathematics, Cracow Technical University, Cracow, Poland*

Abstract: In AIDS studies it is quite common to observe viral load measurements collected irregularly over time. Moreover, these measurements can be subjected to some upper and/or lower detection limits depending on the quantification assays. A complication arises when these continuous repeated measures have a heavy-tailed behavior. For such data structures, we propose a robust structure for a censored linear model based on the multivariate Student-t distribution. To compensate for the autocorrelation existing among irregularly observed measures, a damped exponential correlation structure is employed. An efficient EM-type algorithm is developed for computing the maximum likelihood estimates, obtaining as a by-product the standard errors of the fixed effects and the log-likelihood function. The proposed algorithm uses closed-form expressions at the E-step, that rely on formulas for the mean and variance of a truncated multivariate Student-t distribution. The methodology is illustrated through an application to an HIV-AIDS study and several simulation studies.

Key words and phrases: Censored data, ECM Algorithm, longitudinal data, HIV viral load, outliers.

1 Introduction

In many biomedical studies and clinical trials, the use of longitudinal models has shown a significant growth in recent years, becoming a powerful tool for modeling correlated outcomes. In clinical trials of anti-retroviral therapy in AIDS studies, HIV-1 RNA (viral load) measures are collected over a period of treatment to determine rates of changes in the amount of actively replicating virus. These measures are used as a key primary endpoint because the viral load monitoring during the therapy is mostly available, a failure in the treatment can be defined virologically, and a new regimen of therapy is recommended as soon as virological rebound occurs[1]. Since for each patient the viral load measures are collected over time, the correlation structure among responses must be taken into account.

Longitudinal models allow us to estimate viral load trajectories as well as to quantify the correlation structure between viral load measurements[2,3]. However, in practice, the statistical modeling of viral loads can be challenging due to the following features. First, the measurements can be subject to a upper/lower limit of quantification. As a result, the viral load responses are either left or right censored depending upon the diagnostics assay used. In general, the range of limit detection varies from 400 copies/ml for the earlier assays to 40 copies/ml for the more sophisticated assays of the recent times. Second, the viral load are usually recorded at irregular occasions because the timings of measurements often vary from one patient to another, and typically measurement times are associated with the course of the disease. Third, the viral load measurements often contain influential and/or outlying observations, which can cause misleading results in parameter estimates as well as their standard errors when a Gaussian assumption is considered. Therefore, one of the greatest challenges related to longitudinal data modeling in AIDS research is to consider the inherent features of viral load measurements simultaneously.

In statistical and biomedical literature, linear and nonlinear mixed effects models based on Gaussian assumptions are routinely used to model longitudinal data[4–6]. Nevertheless, such an assumption could be not realistic because of the presence of atypical observations (outliers). To deal with this weakness, some alternatives based on heavy-tailed distributions have been proposed. For example, Pinheiro et al.[7] proposed a Student-t linear mixed model demonstrating its robustness against outliers. Other authors focused their research interest in developing strategies for fitting both linear and nonlinear mixed effects models under heavy-tailed distributions such as the Student-t, slash and contaminated normal distributions[8–12]. The regression and mixed effects models for censored responses under heavy-tailed distributions have been studied in detail in the last years.[13–16]

Recently, a variety of heavy-tailed statistical models have been proposed for longitudinal data considering not only the correlation structure induced by the random effects term but also by other types of correlation in the error term. For example, Wang[17] studied the multivariate Student-t linear mixed model (t-LMM) for outcome variables recorded at irregular occasions by using a parsimonious damping exponential correlation (DEC) structure. This type of correlation structure, proposed by Muñoz et al.[18], takes into account the autocorrelation generated by the within-subject dependence among irregular occasions. Wang and Fan[19] considered the multivariate Student-t linear mixed with AR(p) dependence structure (a particular case of DEC structure) for the within-subject errors in the case of multiple outcomes. However, and as was mentioned by Goldstein et al.[20] and Browne and Goldstein[21], in cases when the repeated measures are collected close in time or correlations among the measures do not decay quickly, random effects modes may not adequately account for the dependency and a more complex correlation structure must be specified.

Following Wang[17], the aim of this paper is to consider the DEC structure for the across-occasion covariance matrix of the random errors under censored responses. As a consequence, the robust so called Student-t multivariate linear censored (t-MLC) model with DEC structure is defined and a fully likelihood-based approach is conducted, including the implementation of an exact expectation conditional maximization (ECM) algorithm for the maximum likelihood (ML) estimation. As in Matos et al.[22], we show that the E-step reduces to computing the first two moments of certain truncated multivariate Student-t distribution, with the computation of the likelihood function and the asymptotic standard errors as a by-product of the E-step. General formulas for these moments were derived by using the results given in Ho et al.[23]. The likelihood function is used for monitoring convergence and the model selection is conducted through the Akaike information criterion (AIC) and Bayesian information criterion (BIC).

The rest of the paper is organized as follows. Section 2 describes a motivating real life data set of HIV-AIDS infected patients. In Section 3 we introduce some notation related to the truncated Student-t distribution. Then, the t-MLC is presented. In Section 4, the related likelihood-based inference are presented including the imputation procedure of censored cases. The method for the prediction of future observations is presented in Section 5. The advantage of the proposed method is presented through the analysis of a case studies of HIV viral load in Section 6. Section 7 presents two simulation studies for comparing the performance of our method with other normality-based one. Section 8 concludes with a short discussion of issues raised by our study and some possible directions for a future research.

2 Motivating example: UTI data

In this section, we present a longitudinal dataset corresponding to UTI - unstructured antiretroviral therapy treatment interruption in HIV-infected adolescents in four institutions in the US. In this case, the HIV-1 RNA measures are subject to censoring below the lower limit of detection of the assay (50 copies/mL) and the presence of outlying and influential observations is noted.

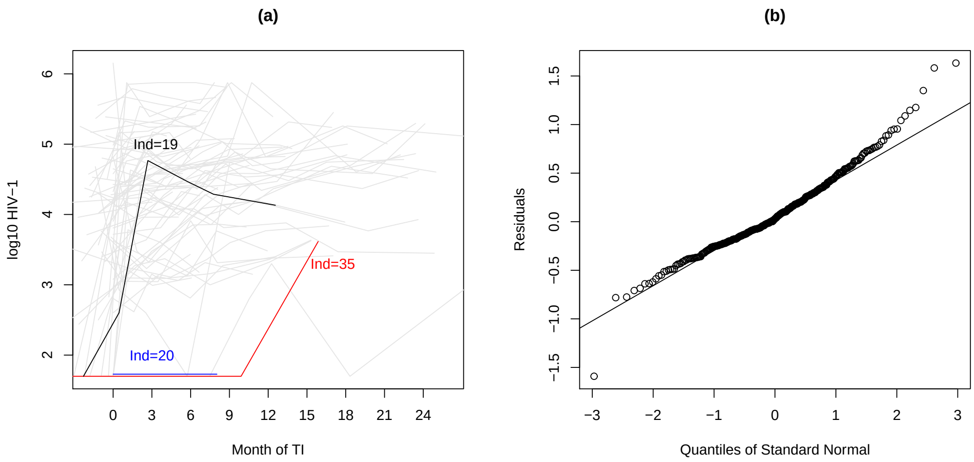

Figure 1: UTI data. (a) Individual profiles (in $\log_{10}$ scale) for HIV viral load at different follow-up times. The trajectories for three censored individuals are numbered. (b) Normal Q–Q plot for model residuals obtained by using the *lmec* package of R.

This dataset consist on the measurements of HIV-1 RNA measures after unstructured treatment interruption (UTI) in 72 adolescents from US. UTI was defined as discontinuation of all antiretroviral drugs for any period of time, after which treatment was resumed. The reasons for interruption might have been diverse. For example, Saitoh et al.[24] mentioned (a) the medication fatigue, (b) patients who were unable to take antiretroviral medications, (c) toxicity associated with the use of

		$\log_{10}$RNA							
		month 0	month 1	month 3	month 6	month 9	month 12	month 18	month 24
$\log_{10}$RNA	month 0		0.4877	0.4100	0.4052	0.4820	0.4435	0.3441	0.6529
	month 1	0.4877		0.9145	0.8551	0.8455	0.6978	0.7090	0.6140
	month 3	0.4100	0.9145		0.9255	0.8638	0.7209	0.7601	0.6301
	month 6	0.4052	0.8551	0.9255		0.8238	0.6490	0.6548	0.5314
	month 9	0.4820	0.8455	0.8638	0.8238		0.9185	0.7642	0.8061
	month 12	0.4435	0.6978	0.7209	0.6490	0.9185		0.6646	0.6897
	month 18	0.3441	0.7090	0.7601	0.6548	0.7642	0.6646		0.8947
	month 24	0.6529	0.6140	0.6301	0.5314	0.8061	0.6897	0.8947	

Table 1: Observed correlation of $\log_{10}$RNA for a single response over different times.

antiretroviral medications, (d) adverse effects.

The dataset present about 7% of observations below the detection limits of assay quantifications (left-censored). The viral load were monitored from the closest time points at $0, 1, 3, 6, 9, 12, 18$ and 24 months after the treatment interruption, that is, they were irregularly collected over time. A more detailed explanation of the data can be founded in Saitoh et al.[24] and Vaida and Liu[25]. The individual profiles of viral load at different followup times after UTI appears in Figure 1 (panel a). This figure also depicts the normal quantile–quantile (Q-Q) plot for the residuals (panel b) obtained by fitting a censored (Gaussian) mixed effect model proposed by Vaida and Liu[25] using the *lmec* package of R. The Q–Q plot exhibits a heavy-tailed behavior, suggesting that the normality assumption for the within-subject errors might be inappropriate. The non-normality of the distribution gives an indication that some atypical observations or outliers might exist in the data. Outliers may cause misleading results in parameter estimates as well as their standard errors and they could have an enormous impact on statistical inferences. In addition, Table 1 shows the observed correlation for the responses in different time points. Clearly, the data do not appear to satisfy an uncorrelated structure of zero covariances or a compound symmetry assumption of equal variances and covariances across time.

3 Model Specification

3.1 Preliminaries

In this section, we present some useful results associated to the p-variate Student-t distribution that will need for implementing the EM algorithm. We start with the density function (*pdf*) of a Student-t random vector $\mathbf{Y} \in \mathbb{R}^p$ with location vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$ and ν degrees of freedom. Its *pdf* is given by

$$t_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \frac{\Gamma(\frac{p+\nu}{2})}{\Gamma(\frac{\nu}{2})\pi^{p/2}}\nu^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\left(1 + \frac{\delta}{\nu}\right)^{-(p+\nu)/2},$$

where $\Gamma(\cdot)$ is the standard gamma function and $\delta = (\mathbf{y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})$ is the Mahalanobis distance. The notation adopted for the Student-t *pdf* is $t_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$.

The cumulative distribution function (*cdf*) is denoted by $T_p(\cdot|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$. It is important to stress that if $\nu > 1$, the mean of $\mathbf{y}$ is $\boldsymbol{\mu}$ and if $\nu > 2$, the covariance matrix is given by $\nu(\nu-2)^{-1}\boldsymbol{\Sigma}$. Moreover, as ν tends to infinity, $\mathbf{Y}$ is converges in distribution to a multivariate normal with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

An important property of the random vector $\mathbf{Y}$ is that it can be written as a mixture of a normal random vector and a positive random variable, *i.e*,

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

where $\mathbf{Z}$ is a normal random vector, with zero-mean vector and covariance $\boldsymbol{\Sigma}$, independent of U, that is a positive random variable with a gamma distribution $Gamma(\nu/2,\nu/2)$.

The distribution of $\mathbf{Y}$ constrained to lie within the right-truncated hyperplane

$$\mathbb{A} = \{\mathbf{y} \in \mathbb{R}^p | \mathbf{y} \le \mathbf{a}\}, \tag{1}$$

where $\mathbf{y} = (y_1,\ldots,y_p)^\top$ and $\mathbf{a} = (a_1,\ldots,a_p)^\top$, is a truncated Student-t distribution, denoted by $Tt_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu;\mathbb{A})$, with *pdf* given by $f(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu;\mathbb{A}) = \dfrac{t_p(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}{T_p(\mathbf{a}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}\mathbb{I}_{\mathbb{A}}(\mathbf{y})$, where $\mathbb{I}_{\mathbb{A}}(\mathbf{y})$ is the indicator function of $\mathbb{A}$.

As was mentioned at the beginning of this section, the next properties of the multivariate Student-t and truncated Student-t distributions are useful for the implementation of the EM-algorithm. We start with the marginal-conditional decomposition of a Student-t random vector. Details of the proofs are provided in Arellano-Valle and Bolfarine[26].

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

(i) $\mathbf{Y}_1 \sim t_{p_1}(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11},\nu)$*; and*

(ii) The conditional cdf of $\mathbf{Y}_2|\mathbf{Y}_1 = \mathbf{y}_1$ is given by

$$P(\mathbf{Y}_2 \le \mathbf{y}_2|\mathbf{Y}_1 = \mathbf{y}_1) = T_{p_2}\left(\mathbf{y}_2|\boldsymbol{\mu}_{2.1},\widetilde{\boldsymbol{\Sigma}}_{22.1},\nu+p_1\right),$$

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

The following results provide the truncated moments of a Student-t random vector. The proofs

Gamma(a,b) denotes a gamma distribution with a/b mean.

of Proposition 2 and 3 are given in Matos et al.[22]. The proof of Proposition 4 is given in Hot et al.[23].

Proposition 2. *If* $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$ *with* $\mathbb{A}$ *as (1), then the* k*-th moment of* $\mathbf{Y}$*,* $k = 0, 1, 2,$ *is*

$$E\left\{\left(\frac{\nu+p}{\nu+\delta}\right)^r \mathbf{Y}^{(k)}\right\} = c_p(\nu, r)\frac{T_p(\mathbf{a}|\boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \nu+2r)}{T_p(\mathbf{a}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)} E_{\mathbf{W}}\{\mathbf{W}^{(k)}\},\ \mathbf{W} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \nu+2r; \mathbb{A}),$$

where $c_p(\nu, r) = \left(\frac{\nu+p}{\nu}\right)^r \left(\frac{\Gamma((p+\nu)/2)\Gamma((\nu+2r)/2)}{\Gamma(\nu/2)\Gamma((p+\nu+2r)/2)}\right)$, $\delta = (\mathbf{Y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})$, $\mathbf{a} = (a_1, \ldots, a_p)^\top$, $\boldsymbol{\Sigma}^* = \frac{\nu}{\nu+2r}\boldsymbol{\Sigma}$, $\mathbf{Y}^{(0)} = 1$, $\mathbf{Y}^{(1)} = \mathbf{Y}$, $\mathbf{Y}^{(2)} = \mathbf{Y}\mathbf{Y}^\top$*, and* $\nu + 2r > 0$.

Having established a formula for the k-order moments of $\mathbf{Y}$, we now present a result on conditional moments of the partition of $\mathbf{Y}$.

Proposition 3. *Let* $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$ *with* $\mathbb{A}$ *as (1). Consider the partition* $\mathbf{Y}^\top = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)$ *with* $dim(\mathbf{Y}_1) = p_1$*,* $dim(\mathbf{Y}_2) = p_2$*,* $p_1 + p_2 = p$*, and the corresponding partition of the parameters* $\boldsymbol{\mu}$*,* $\boldsymbol{\Sigma}$*,* $\mathbf{a}$ *(*$\mathbf{a}^{y_1}, \mathbf{a}^{y_2}$*) and* $\mathbb{A}$ *(*$\mathbb{A}^{y_1}, \mathbb{A}^{y_2}$*). Then, under the notation of Proposition 1, the conditional* k*-th moment of* $\mathbf{Y}_2$ *is*

$$E\left\{\left(\frac{\nu+p}{\nu+\delta}\right)^r \mathbf{Y}_2^{(k)}|\mathbf{Y}_1\right\} = \frac{d_p(p_1, \nu, r)}{(\nu+\delta_1)^r}\frac{T_{p_2}(\mathbf{a}^{y_2}|\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}^*_{22.1}, \nu+p_1+2r)}{T_{p_2}(\mathbf{a}^{y_2}|\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}_{22.1}, \nu+p_1)} E_{\mathbf{W}}\{\mathbf{W}^{(k)}\},$$

where $\mathbf{W} \sim Tt_{p_2}(\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}^*_{22.1}, \nu+p_1+2r; \mathbb{A}^{y_2})$, $\delta = (\mathbf{Y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})$, $\delta_1 = (\mathbf{Y}_1-\boldsymbol{\mu}_1)^\top\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{Y}_1-\boldsymbol{\mu}_1)$, $\mathbf{a}^{y_2} = (a_1, \ldots, a_{p_2})^\top$, $\widetilde{\boldsymbol{\Sigma}}^*_{22.1} = \left(\frac{\nu+\delta_1}{\nu+2r+p_1}\right)\boldsymbol{\Sigma}_{22.1}$, $\nu + p_1 + 2r > 0$ *and* $d_p(p_1, \nu, r) = (\nu+p)^r\left(\frac{\Gamma((p+\nu)/2)\Gamma((p_1+\nu+2r)/2)}{\Gamma((p_1+\nu)/2)\Gamma((p+\nu+2r)/2)}\right)$.

In the following Proposition, we establish relationships between the expectation and covariance of $\mathbf{Y}$ and $\mathbf{W}$.

Proposition 4. *Let* $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu}, \Sigma, \nu; \mathbb{A}^*)$*, with* $\mathbb{A}^* = \{\mathbf{y} \in \mathbb{R}^p|\mathbf{a}^* < \mathbf{y} \leq \mathbf{b}^*\}$*, where* $\mathbf{a}^* = (a_1^*, \ldots, a_p^*)^\top$*,* $\mathbf{b}^* = (b_1^*, \ldots, b_p^*)^\top$*,* $\Sigma = \Lambda\mathbf{R}\Lambda$ *and* $\Lambda = Diag(\sigma_{11}, \ldots, \sigma_{pp})$ *is a* $p \times p$ *diagonal matrix with each element being positive. We have that* $\mathbf{W} = \Lambda^{-1}(\mathbf{Y}-\boldsymbol{\mu}) \sim Tt_p(\mathbf{0}, \mathbf{R}, \nu; \mathbb{A})$*, where* $\mathbf{a} = \Lambda^{-1}(\mathbf{a}^* - \boldsymbol{\mu})$ *and* $\mathbf{b} = \Lambda^{-1}(\mathbf{b}^* - \boldsymbol{\mu})$*. Therefore,*

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \Lambda E[\mathbf{W}]$$

$$E[\mathbf{Y}\mathbf{Y}^\top] = \boldsymbol{\mu}\boldsymbol{\mu}^\top + \Lambda E[\mathbf{W}]\boldsymbol{\mu}^\top + \mu E[\mathbf{W}^\top]\Lambda + \Lambda E[\mathbf{W}\mathbf{W}^\top]\Lambda^\top,$$

where $E[\mathbf{W}]$ *and* $E[\mathbf{W}\mathbf{W}^\top]$ *are given in Ho et al.*[23]*.*

3.2 The statistical model

Our multivariate Student-t linear (t-ML) model, for longitudinal responses, is defined by:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \tag{2}$$

with $\boldsymbol{\epsilon}_i \sim t_{n_i}\{\mathbf{0}, \boldsymbol{\Sigma}_i, \nu\}$, where $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^\top$ is a $n_i \times 1$ vector of continuous responses for sample unit i measured at particular time points $\mathbf{t}_i = (t_{i1}, \ldots, t_{n_i})^\top$, $\mathbf{X}_i$ is the $n_i \times p$ design matrix corresponding to the $p \times 1$ vector of fixed effects $\boldsymbol{\beta}$ and $\boldsymbol{\epsilon}_i$ is the $n_i \times 1$ vector of random errors. As was noted in Section 2, the measurements of the viral load for each subject present evidence of serial correlation. Therefore, to obtain accurate parameter estimates, we consider a parsimonious structure on the dispersion matrix $\boldsymbol{\Sigma}_i = \sigma^2\mathbf{E}_i$, where the matrix $\mathbf{E}_i$ incorporates a time-dependence structure. Thus, we adopt the DEC structure for $\boldsymbol{\Sigma}_i$, as proposed by Muñoz et al.[18]. This correlation structure allows us to deal with unequally spaced and unbalanced observations and is defined as

$$\boldsymbol{\Sigma}_i = \sigma^2\mathbf{E}_i = \sigma^2\mathbf{E}_i(\boldsymbol{\phi}, \mathbf{t}_i) = \sigma^2\left[\phi_1^{|t_{ij}-t_{ik}|^{\phi_2}}\right], \tag{3}$$

where $\mathbf{E}_i(\boldsymbol{\phi}, \mathbf{t}_i) = \sigma^2\left[\phi_1^{|t_{ij}-t_{ik}|^{\phi_2}}\right]$, $i = 1, \ldots, n,\ \ j,k = 1, \ldots, n_i,$. 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, for a given nonnegative ϕ_1:

- If $\phi_2 = 0$, then the $\mathbf{E}_i$ is the compound symmetry structure.
- If $0 < \phi_2 < 1$, then we have $\mathbf{E}_i$ with a decay rate between that of compound symmetry and the first-order AR (AR(1)) model.
- If $\phi_2 = 1$, then the $\mathbf{E}_i$ is an AR(1) model.
- If $\phi_2 > 1$, then we have $\mathbf{E}_i$ with a decay rate faster than AR(1) model.
- If $\phi_2 \to \infty$,then the $\mathbf{E}_i$ is the first-order moving average (MA(1)) model.

For a more detailed discussion about the DEC structure, we refer to Muñoz et al.[18] and Wang[17].

From a practical point of view and in order to avoid computational problems, the parameter space of ϕ_1 and ϕ_2 is confined within $\{(\phi_1, \phi_1) : 0 < \phi_1 < 1, \phi_2 > 0\}$, that is, we restrict our attention to nonnegative values of ϕ_1 and ϕ_2. Such assumption is quite common in most biomedical and epidemiological applications. Finally, we consider the approach proposed by Vaida and Liu[25] and Matos et al.[22] to modeling the censored responses. Thus, the observed data for the ith subject is given by $(\mathbf{V}_i, \mathbf{C}_i)$, where $\mathbf{V}_i$ represents the vector of uncensored reading or censoring level and

$\mathbf{C}_i$ the vector of censoring indicators. In other words,

$$y_{ij} \leq V_{ij} \quad \text{if} \;\; C_{ij} = 1, \;\text{and}\;\; y_{ij} = V_{ij} \quad \text{if} \;\; C_{ij} = 0, \tag{4}$$

so that, considering (4) along with (2)-(3) defined the multivariate Student-t linear censored (t-MLC) model. In our theoretical development, we will use a left censoring pattern, but the extensions to right censored data are immediate. The right censored problem can be represented by a left censored problem by simultaneously transforming the response y_{ij} and censoring level V_{ij} to $-y_{ij}$ and $-V_{ij}$.

3.3 The likelihood function

To obtain the likelihood function of the t-MLC model, first we treat separately the observed and censored components of $\mathbf{y}_i$, *i.e.*, $\mathbf{y}_i = (\mathbf{y}_i^{o\top}, \mathbf{y}_i^{c\top})^\top$, with $C_{ij} = 0$ for all elements in $\mathbf{y}_i^o$, and $C_{ij} = 1$ for all elements in $\mathbf{y}_i^c$. Analogous, we write $\mathbf{V}_i = vec(\mathbf{V}_i^o, \mathbf{V}_i^c)$, where $vec(\cdot)$ denotes the function which stacks vectors or matrices of the same number of columns, with $\mathbf{\Sigma}_i = \begin{pmatrix} \mathbf{\Sigma}_i^{oo} & \mathbf{\Sigma}_i^{oc} \\ \mathbf{\Sigma}_i^{co} & \mathbf{\Sigma}_i^{cc} \end{pmatrix}$. Then, using Proposition 1, we have that $\mathbf{y}_i^o \sim t_{n_i^o}(\mathbf{X}_i^o\boldsymbol{\beta}, \mathbf{\Sigma}_i^{oo}, \nu)$ and $\mathbf{y}_i^c | \mathbf{y}_i^o, \sim t_{n_i^c}(\boldsymbol{\mu}_i^{co}, \mathbf{S}_i^{co}, \nu + n_i^o)$, where

$$\boldsymbol{\mu}_i^{co} = \mathbf{X}_i^c\boldsymbol{\beta} + \mathbf{\Sigma}_i^{co}\mathbf{\Sigma}_i^{oo-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\boldsymbol{\beta}), \;\; \mathbf{S}_i^{co} = \left(\frac{\nu + Q(\mathbf{y}_i^o)}{\nu + n_i^o}\right)\mathbf{\Sigma}_i^{cc.o}, \tag{5}$$

with $\mathbf{\Sigma}_i^{cc.o} = \mathbf{\Sigma}_i^{cc} - \mathbf{\Sigma}_i^{co}\mathbf{\Sigma}_i^{oo-1}\mathbf{\Sigma}_i^{oc}$ and $Q(\mathbf{y}_i^o) = (\mathbf{y}_i^o - \mathbf{X}_i^o\boldsymbol{\beta})^\top\mathbf{\Sigma}_i^{oo-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\boldsymbol{\beta})$. Therefore, the likelihood function for subject i is

$$\begin{aligned} L_i(\boldsymbol{\theta}|\mathbf{y}) &= f(\mathbf{V}_i|\mathbf{C}_i, \boldsymbol{\theta}) = f(\mathbf{y}_i^c \leq \mathbf{V}_i^c | \mathbf{y}_i^o = \mathbf{V}_i^o, \boldsymbol{\theta}) f(\mathbf{y}_i^o = \mathbf{V}_i^o|\boldsymbol{\theta}), \\ &= T_{n_i^c}(\mathbf{V}_i^c|\boldsymbol{\mu}_i^{co}, \mathbf{S}_i^{co}, \nu + n_i^o) t_{n_i^o}(\mathbf{V}_i^o|\mathbf{X}_i^o\boldsymbol{\beta}, \mathbf{\Sigma}_i^{oo}, \nu) = L_i. \end{aligned}$$

Straightforwardly, the log-likelihood function for the observed data is given by $\ell(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^n \log L_i$. It is important to note that this function can be computed at each step of the EM-type algorithm without additional computational burden since the L_i's have already been computed at the E-step. We assume that the degrees of freedom parameter of the Student-t distribution is fixed. For choosing the most appropriate value of this parameter, we will use the log-likelihood profile[10,27]. This assumption is based on the work by Lucas[28], in which the author showed that the protection against outliers is preserved only if the degrees of freedom parameter is fixed. Consequently, the parameter vector for the t-MLC model is $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \phi_1, \phi_2)^\top$.

4 The EM algorithm

We describe in detail how to carry out ML estimation for the proposed t-MLC model. The EM algorithm, originally proposed by Dempster et al.[29], is a very popular iterative optimization strategy commonly used to obtain ML estimates for incomplete data problems. This algorithm has many attractive features such as the numerical stability and the simplicity of implementation and its memory requirements are quite reasonable[30]. However, ML estimation for the t-MLC model is complicated because of the censoring and the DEC structure, and the EM algorithm is less advisable due to the computational difficulty at the M-step. To overcome this problem, we used an extension of the EM algorithm, called the ECM algorithm[31]. A key feature of this algorithm is that it preserves the stability of the EM and has a typically faster convergence rate than the original EM.

In order to propose the ECM algorithm for our t-MLC model, firstly we define $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{u} = (u_1, \ldots, u_n)^\top$, $\mathbf{V} = vec(\mathbf{V}_1, \ldots, \mathbf{V}_n)$, and $\mathbf{C} = vec(\mathbf{C}_1, \ldots, \mathbf{C}_n)$ such that we observe $(\mathbf{V}_i, \mathbf{C}_i)$ for the i-th subject. Now, we treat $\mathbf{u}$ and $\mathbf{y}$ as hypothetical missing data, and augmenting with the observed data $\mathbf{V}, \mathbf{C}$ corresponding to the censoring mechanism. Consequently, we set the complete-data vector as $\mathbf{y}_c = (\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{y}^\top, \mathbf{u}^\top)^\top$. As is well known, the ECM algorithm must be applied to the complete data log-likelihood function given by

$$\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}|\mathbf{y}_c),$$

where

$$\ell_i(\boldsymbol{\theta}|\mathbf{y}_c) = -\frac{1}{2}\left[n_i \log\sigma^2 + \log|\mathbf{E}_i| + \frac{u_i}{\sigma^2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top \mathbf{E}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right] + h(u_i|\nu) + c,$$

c is a constant that does not depend on $\boldsymbol{\theta}$ and $h(u_i|\nu)$ is the Gamma$(\nu/2, \nu/2)$ *pdf.* Finally, the ECM algorithm for the t-MLC model can be summarized through the following two steps.

E-step:

Given the current value $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, the E-step obtains the conditional expectation of the complete data log-likelihood function

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}), \tag{6}$$

where

$$Q_i(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}|\widehat{\boldsymbol{\theta}}^{(k)}) = -\frac{n_i}{2}\log\sigma^2 - \frac{1}{2}\log|\mathbf{E}_i| - \frac{1}{2\sigma^2}A_i^{(k)}(\boldsymbol{\beta}, \boldsymbol{\phi}),$$

with

$$A_i^{(k)}(\boldsymbol{\beta}, \boldsymbol{\phi}) = \left[\mathrm{tr}\left(\widehat{u\mathbf{y}_i^2}^{(k)}\mathbf{E}_i^{-1}\right) - 2\boldsymbol{\beta}\mathbf{X}_i^\top\mathbf{E}_i^{-1}\widehat{u\mathbf{y}}_i^{(k)} + \widehat{u}_i^{(k)}\boldsymbol{\beta}\mathbf{X}_i^\top\mathbf{E}_i^{-1}\mathbf{X}_i\boldsymbol{\beta}\right].$$

Note that, since ν is fixed, we do not need to obtain $E[h(u_i|\nu)|\mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}]$.

CM-step:

In this step, $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ is conditionally maximized with respect to $\boldsymbol{\theta}$ and a new estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$ is obtained. Specifically, we have that

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}^{(k+1)} &= \left(\sum_{i=1}^{n} \widehat{u}_i^{(k)} \mathbf{X}_i^\top \left(\widehat{\mathbf{E}}_i^{(k)}\right)^{-1} \mathbf{X}_i\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i^\top \left(\widehat{\mathbf{E}}_i^{(k)}\right)^{-1} \widehat{u\mathbf{y}}_i^{(k)}, && (7)\\
\widehat{\sigma^2}^{(k+1)} &= \frac{1}{N}\sum_{i=1}^{n} A_i^{(k)}(\widehat{\boldsymbol{\beta}}^{(k+1)}, \widehat{\boldsymbol{\phi}}^{(k)}) && (8)\\
\widehat{\boldsymbol{\phi}}^{(k+1)} &= argmax_{\boldsymbol{\phi}} \left\{ -\frac{1}{2}\sum_{i=1}^{n} [\log(|\mathbf{E}_i|) + A_i^{(k)}(\widehat{\boldsymbol{\beta}}^{(k+1)}, \boldsymbol{\phi})] \right\}, && (9)
\end{aligned}
$$

where $N = \sum_{i=1}^{n} n_i$. The *optim* routine in `R` package[32] can be used to perform a two-dimensional search of (ϕ_1, ϕ_2) in (9) subject to box constraints: $\phi_1 \in [0, 1)$ and $\phi_2 \in (0, \infty)$.

The algorithm is iterated until a suitable convergence rule is satisfied. In this case, we adopt the distance involving two successive evaluations of the log-likelihood $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)})/\ell(\widehat{\boldsymbol{\theta}}^{(k)}) - 1|$ as a convergence criterion. It is important to stress that from equations (7)-(9), the E-step reduces to the computation of $\widehat{u\mathbf{y}_i^2}$, $\widehat{u\mathbf{y}}_i$, and $\widehat{u}_i$. These expected values can be determined in closed form, using Propositions 1-4, as follows:

1. If the subject i has only censored components, from Proposition 2

$$
\begin{aligned}
\widehat{u\mathbf{y}_i^2} &= E\{u_i \mathbf{y}_i \mathbf{y}_i^\top | \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\} = \frac{T_{n_i}(\mathbf{V}_i|\widehat{\boldsymbol{\mu}}_i^{(k)}, \widehat{\boldsymbol{\Sigma}}_i^{*(k)}, \nu+2)}{T_{n_i}(\mathbf{V}_i|\widehat{\boldsymbol{\mu}}_i^{(k)}, \widehat{\boldsymbol{\Sigma}}_i^{(k)}, \nu)} E\{\mathbf{W}_i \mathbf{W}_i^\top\},\\
\widehat{u\mathbf{y}}_i &= E\{u_i \mathbf{y}_i | \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\} = \frac{T_{n_i}(\mathbf{V}_i|\widehat{\boldsymbol{\mu}}_i^{(k)}, \widehat{\boldsymbol{\Sigma}}_i^{*(k)}, \nu+2)}{T_{n_i}(\mathbf{V}_i|\widehat{\boldsymbol{\mu}}_i^{(k)}, \widehat{\boldsymbol{\Sigma}}_i^{(k)}, \nu)} E\{\mathbf{W}_i\},\\
\widehat{u}_i &= E\{u_i | \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\} = \frac{T_{n_i}(\mathbf{V}_i|\widehat{\boldsymbol{\mu}}_i^{(k)}, \widehat{\boldsymbol{\Sigma}}_i^{*(k)}, \nu+2)}{T_{n_i}(\mathbf{V}_i|\widehat{\boldsymbol{\mu}}_i^{(k)}, \widehat{\boldsymbol{\Sigma}}_i^{(k)}, \nu)},
\end{aligned}
$$

   where $\mathbf{W}_i \sim Tt_{n_i}(\widehat{\boldsymbol{\mu}}_i^{(k)}, \widehat{\boldsymbol{\Sigma}}_i^{*(k)}, \nu+2; \mathbb{A}_i)$, $\widehat{\boldsymbol{\mu}}_i^{(k)} = \mathbf{X}_i \widehat{\boldsymbol{\beta}}^{(k)}$, $\widehat{\boldsymbol{\Sigma}}_i^{*(k)} = \dfrac{\nu}{\nu+2}\widehat{\boldsymbol{\Sigma}}_i^{(k)}$, $\widehat{\boldsymbol{\Sigma}}_i^{(k)} = \widehat{\sigma^{2(k)}}\widehat{\mathbf{E}}_i^{(k)}$ and $\mathbb{A}_i = \{\mathbf{W}_i \in \mathbb{R}^{n_i} | \mathbf{w}_i \leq \mathbf{V}_i\}$ where $\mathbf{w}_i = (w_{i1}, \ldots, w_{in_i})^\top$ and $\mathbf{V}_i = (V_{i1}, \ldots, V_{in_i})^\top$.

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

$$
\widehat{u\mathbf{y}_i^2} = \frac{\nu + n_i}{\nu + Q(\mathbf{y}_i)} \mathbf{y}_i \mathbf{y}_i^\top, \; \widehat{u\mathbf{y}}_i = \frac{\nu + n_i}{\nu + Q(\mathbf{y}_i)} \mathbf{y}_i, \; \widehat{u}_i = \frac{\nu + n_i}{\nu + Q(\mathbf{y}_i)},
$$

   where $Q(\mathbf{y}_i) = (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$.

3. If the subject i has censored and uncensored components, then from Proposition 3 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 processes, we have that

$$\begin{aligned}
\widehat{u\mathbf{y}_i^2} &= E\{u_i\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{y}_i^o,\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\} = \begin{pmatrix} \mathbf{y}_i^o\mathbf{y}_i^{o\top}\widehat{u}_i & \widehat{u}_i\mathbf{y}_i^o\widehat{\mathbf{w}}_i^{c\top} \\ \widehat{u}_i\widehat{\mathbf{w}}_i^c\mathbf{y}_i^{o\top} & \widehat{u}_i\widehat{\mathbf{w}^2}_i^c \end{pmatrix}, \\
\widehat{u\mathbf{y}}_i &= E\{u_i\mathbf{y}_i|\mathbf{y}_i^o,\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\} = vec(y_i^o\widehat{u}_i,\widehat{\mathbf{w}}_i^c), \\
\widehat{u}_i &= E\{u_i|\mathbf{y}_i^o,\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\} = \left(\frac{n_i^o+\nu}{\nu+Q(\mathbf{y}_i^o)}\right)\frac{T_{n_i}(\mathbf{V}_i|\boldsymbol{\mu}_i^{co},\widetilde{\mathbf{S}}^{co},\nu+n_i^o+2)}{T_{n_i}(\mathbf{V}_i|\boldsymbol{\mu}_i^{co},\mathbf{S}^{co},\nu+n_i^o)},
\end{aligned}$$

where $\widetilde{\mathbf{S}}^{co} = \left(\frac{\nu+Q(\mathbf{y}_i^o)}{\nu+2+n_i^o}\right)\boldsymbol{\Sigma}_i^{cc.o}$, $\widehat{\mathbf{w}}_i^c = E\{\mathbf{W}_i\}$, and $\widehat{\mathbf{w}^2}_i^c = E\{\mathbf{W}_i\mathbf{W}_i^\top\}$, with $\mathbf{W}_i \sim Tt_{n_i^c}(\boldsymbol{\mu}_i^{co},\widetilde{\mathbf{S}}^{co},\nu+n_i^o+2;\mathbb{A}_i^c)$ and $\boldsymbol{\Sigma}_i^{cc.o}$, $\boldsymbol{\mu}_i^{co}$, and $\mathbf{S}^{co}$ are as in (5).

As was mentioned in Proposition 4, formulas for $E[\mathbf{W}]$ and $E[\mathbf{W}\mathbf{W}^\top]$, where $\mathbf{W}\sim Tt_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu;\mathbb{A})$, can be found in Ho et al.[23]. For the the computation of multivariate Student-t *cdf* we used the *pmvt* function of the *mvtnorm* package[33] from R software. Additional details about the ECM algorithm for our proposed t-MLC model can be found in Appendix.

4.1 The expected information matrix

Louis[34] proposed a technique for computing the observed information matrix within the EM algorithm framework. Using this method, and from the results given by Lange et al.[27], we can find an asymptotic approximation for the variances of the fixed effects in the t-MLC model. This approximation is given by

$$\mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\beta}} = Var(\widehat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^n \frac{\nu+n_i}{\nu+n_i+2}\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i - \sum_{i=1}^n \mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{B}_i\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\right)^{-1}, \tag{10}$$

where $\mathbf{B}_i = Var\left\{\frac{\nu+n_i}{\nu+Q(\mathbf{y}_i)}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})|\mathbf{V}_i,\mathbf{C}_i\right\}$, with $\mathbf{y}_i\sim Tt_{n_i}(\mathbf{X}_i\boldsymbol{\beta},\boldsymbol{\Sigma}_i,\nu;\mathbb{A}_i)$ where $\mathbb{A}_i=\{\mathbf{y}_i\in\mathbb{R}^{n_i}|\mathbf{y}_i\leq\mathbf{a}_i\}$ with $\mathbf{y}_i=(y_{i1},\ldots,y_{in_i})^\top$ and $\mathbf{a}_i=(a_{i1},\ldots,a_{in_i})^\top$.

Note that clearly $\mathbf{B}_i$ depends on the following quantities:

$$\widehat{u\mathbf{y}_i^2}^* = E\left[\left(\frac{\nu+n_i}{\nu+Q(\mathbf{y}_i)}\right)^2\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}\right], \quad \widehat{u\mathbf{y}_i^1}^* = E\left[\left(\frac{\nu+n_i}{\nu+Q(\mathbf{y}_i)}\right)^2\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}\right] \text{ and}$$

$$\widehat{u\mathbf{y}_i^0}^* = E\left[\left(\frac{\nu+n_i}{\nu+Q(\mathbf{y}_i)}\right)^2|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}\right].$$

After some algebraic manipulations, we have three possible scenarios for the calculation of these quantities:

- If individual i has only non-censored components, then

$$\widehat{u\mathbf{y}_i^2}^* = \left(\frac{\nu+n_i}{\nu+Q(\mathbf{y}_i)}\right)^2\mathbf{y}_i\mathbf{y}_i^\top,\ \widehat{u\mathbf{y}_i^1}^* = \left(\frac{\nu+n_i}{\nu+Q(\mathbf{y}_i)}\right)^2\mathbf{y}_i,\ \widehat{u\mathbf{y}_i^0}^* = \left(\frac{\nu+n_i}{\nu+Q(\mathbf{y}_i)}\right)^2,$$

where $Q(\mathbf{y}_i) = (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$.

- If individual i has only censored components then from Proposition 2

$$\widehat{u\mathbf{y}_i^2}^* = c_{n_i}(\nu,2)\frac{T_{n_i}(\mathbf{V}_i|\widehat{\boldsymbol{\mu}}_i, \widehat{\boldsymbol{\Sigma}}_i^*, \nu+4)}{T_{n_i}(\mathbf{V}_i|\widehat{\boldsymbol{\mu}}_i, \widehat{\boldsymbol{\Sigma}}_i, \nu)} E\left[\mathbf{W}_i\mathbf{W}_i^\top\right],$$

$$\widehat{u\mathbf{y}_i^1}^* = c_{n_i}(\nu,2)\frac{T_{n_i}(\mathbf{V}_i|\widehat{\boldsymbol{\mu}}_i, \widehat{\boldsymbol{\Sigma}}_i^*, \nu+4)}{T_{n_i}(\mathbf{V}_i|\widehat{\boldsymbol{\mu}}_i, \widehat{\boldsymbol{\Sigma}}_i, \nu)} E\left[\mathbf{W}_i\right], \quad \widehat{u\mathbf{y}_i^0}^* = c_{n_i}(\nu,2)\frac{T_{n_i}(\mathbf{V}_i|\widehat{\boldsymbol{\mu}}_i, \widehat{\boldsymbol{\Sigma}}_i^*, \nu+4)}{T_{n_i}(\mathbf{V}_i|\widehat{\boldsymbol{\mu}}_i, \widehat{\boldsymbol{\Sigma}}_i, \nu)},$$

where $\mathbf{W}_i \sim Tt_{n_i}(\widehat{\boldsymbol{\mu}}_i, \widehat{\boldsymbol{\Sigma}}_i, \nu+4, \mathbb{A}_i)$, $\widehat{\boldsymbol{\mu}_i} = \mathbf{X}_i\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\Sigma}}_i^* = \dfrac{\nu}{\nu+4}\widehat{\boldsymbol{\Sigma}}_i$ and

$$c_{n_i}(\nu,2) = \left(\frac{\nu+n_i}{\nu}\right)^2 \left[\frac{\Gamma\left(\frac{\nu+n_i}{2}\right)\Gamma\left(\frac{\nu+4}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{\nu+n_i+4}{2}\right)}\right].$$

- If individual i has censored and uncensored components, then from Proposition 3 and the fact 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 processes, we have:

$$\widehat{u\mathbf{y}_i^2}^* = \begin{pmatrix} \widehat{u\mathbf{y}_i^0}^* \mathbf{y}_i^o\mathbf{y}_i^{o\top} & \widehat{u\mathbf{y}_i^0}^* \mathbf{y}_i^o\widehat{\mathbf{w}_i^c}^\top \\ \widehat{u\mathbf{y}_i^0}^* \widehat{\mathbf{w}_i^c}\mathbf{y}_i^{o\top} & \widehat{u\mathbf{y}_i^0}^* \widehat{\mathbf{w}_i^c\mathbf{w}_i^{c\top}} \end{pmatrix}, \quad \widehat{u\mathbf{y}_i^1}^* = vec(\widehat{u\mathbf{y}_i^0}\,\mathbf{y}_i^o, \widehat{\mathbf{w}_i^c}),$$

$$\widehat{u\mathbf{y}_i^0}^* = \left(\frac{d_{n_i}}{(\nu+\delta_i^o)^2}\right)\frac{T_{n_i^c}(\mathbf{V}_i|\boldsymbol{\mu}_i^{co}, \tilde{\mathbf{S}}_i^{co}, \nu+n_i^o+4)}{T_{n_i^c}(\mathbf{V}_i|\boldsymbol{\mu}_i^{co}, \boldsymbol{\Sigma}_i^{co}, \nu+n_i^o)},$$

where

$$d_{n_i} = (\nu+n_i)^2\left(\frac{\Gamma\left(\frac{n_i+\nu}{2}\right)\Gamma\left(\frac{n_i^o+\nu+4}{2}\right)}{\Gamma\left(\frac{n_i^o+\nu}{2}\right)\Gamma\left(\frac{n_i+\nu+4}{2}\right)}\right),$$

$\tilde{\mathbf{S}}_i^{co} = \left(\frac{\nu+\delta_i^o}{\nu+4+n_i^o}\right)\boldsymbol{\Sigma}_i^{cc.o}$, $\widehat{\mathbf{w}_i^c} = E[\mathbf{W}_i]$ and $\widehat{\mathbf{w}_i^c\mathbf{w}_i^{c\top}} = E[\mathbf{W}_i\mathbf{W}_i^\top]$, with $\mathbf{W}_i \sim Tt_{n_i^c}(\boldsymbol{\mu}_i^{co}, \tilde{\mathbf{S}}_i^{co}, \nu + n_i^o+2, \mathbb{A}_i^c)$ and $\boldsymbol{\mu}_i^{co}$,$\boldsymbol{\Sigma}_i^{cc.o}$ and $\mathbf{S}_i^{co}$ as defined in Subsection 3.3.

Asymptotic confidence intervals for the fixed effects and hypothesis tests as well are obtained assuming that the ML estimates $\widehat{\boldsymbol{\beta}}$ has approximately a $N_p(\boldsymbol{\beta}, \mathbf{J}_{\boldsymbol{\beta\beta}}^{-1})$ distribution. In practice, $\mathbf{J}_{\boldsymbol{\beta\beta}}$ is usually unknown and it needs to be replaced by its ML estimates $\mathbf{J}_{\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}}$.

4.2 Imputation of censored components

Let $\mathbf{y}_i^{(c)}$ the true unobserved response vector for the censored components of the ith subject. Now, as a byproduct of the EM algorithm we can obtain the predictor of the censored components, denoted by $\widetilde{\mathbf{y}}_i^{(c)}$, as follows

$$\widetilde{\mathbf{y}}_i^{(c)} = E\{\mathbf{y}_i|\mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\}, \tag{11}$$

which is obtained considering two possible cases:

1. If subject i has only censored components

$$\widetilde{\mathbf{y}}_i^{(c)} = E\{\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\},$$

where $\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}} \sim Tt_{n_i}(\mathbf{X}_i\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}_i, \nu; \mathbb{A}_i)$, with $\mathbb{A}_i = \{\mathbf{y}_i \in \mathbb{R}^{n_i}|\mathbf{y}_i \leq \mathbf{a}_i\}$, $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$ and $\mathbf{a}_i = (a_{i1}, \ldots, a_{in_i})^\top$. This expression is obtained using Proposition 4.

2. If subject i has censored and uncensored components, then from Proposition 3 with $r = 0$ and $k = 1$, we have that

$$\widetilde{\mathbf{y}}_i^{(c)} = E\{\mathbf{y}_i^c|\mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\},$$

with $\mathbf{y}_i^c|\mathbf{y}_i^o, \sim Tt_{n_i^c}(\widehat{\boldsymbol{\mu}}_i^{co}, \widehat{\mathbf{S}}_i^{co}, \nu+n_i^o; \mathbb{A}_i)$ where $\mathbb{A}_i = \{\mathbf{y}_i \in \mathbb{R}^{n_i}|\mathbf{y}_i \leq \mathbf{a}_i\}$ with $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$, $\mathbf{a}_i = (a_{i1}, \ldots, a_{in_i})^\top$ and

$$\widehat{\boldsymbol{\mu}}_i^{co} = \mathbf{X}_i^c\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\Sigma}}_i^{co}\widehat{\boldsymbol{\Sigma}}_i^{oo-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\widehat{\boldsymbol{\beta}}), \;\; \widehat{\mathbf{S}}_i^{co} = \left(\frac{\nu + Q(\mathbf{y}_i^o)}{\nu + n_i^o}\right)\widehat{\boldsymbol{\Sigma}}_i^{cc.o},$$

with $\widehat{\boldsymbol{\Sigma}}_i^{cc.o} = \widehat{\boldsymbol{\Sigma}}_i^{cc} - \widehat{\boldsymbol{\Sigma}}_i^{co}\widehat{\boldsymbol{\Sigma}}_i^{oo-1}\widehat{\boldsymbol{\Sigma}}_i^{oc}$ and $Q(\mathbf{y}_i^o) = (\mathbf{y}_i^o - \mathbf{X}_i^o\widehat{\boldsymbol{\beta}})^\top\widehat{\boldsymbol{\Sigma}}_i^{oo-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\widehat{\boldsymbol{\beta}})$.

5 Prediction of future values

The problem related to the prediction of future values has a great impact in many practical applications. Rao[35] pointed out that the predictive accuracy of future observations can be taken as an alternative measure of "goodness-of-fit". In order to propose an strategy for generating predicted values from our t-MLC model, we used the approach proposed by Wang[17]. Thus, let $\mathbf{y}_{i,obs}$ be an observed response vector of dimension $n_{i,obs} \times 1$ for a new subject i over the first portion of time and $\mathbf{y}_{i,pred}$ the corresponding $n_{i,pred} \times 1$ response vector over the future portion of time. Let $\bar{\mathbf{X}}_i = (\mathbf{X}_{i,obs}, \mathbf{X}_{i,pred})$ be the $(n_{i,obs}+n_{i,pred}) \times p$ design matrix corresponding to $\bar{\mathbf{y}}_i = (\mathbf{y}_{i,obs}^\top, \mathbf{y}_{i,pred}^\top)$.

To deal with the censored values existing in $\mathbf{y}_{i,obs}$, we used the imputation procedure presented in the Subsection 4.2. Therefore, when the censored values are imputed, a complete data set, denoted by $\mathbf{y}_{i,obs^*}$, is obtained. The reason to use the imputation procedure is that we avoid to compute truncated conditional expectations of multivariate Student-t distribution originated by the censoring scheme. Hence, we have that

$$\bar{\mathbf{y}}_i^* = \left(\mathbf{y}_{i,obs^*}^\top, \mathbf{y}_{i,pred}^\top\right)^\top \sim t_{n_{i,obs}+n_{i,pred}}\left(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i, \nu\right),$$

where the matrix $\boldsymbol{\Sigma}_i$ defined in (3), can be represented by $\boldsymbol{\Sigma}_i = \begin{pmatrix} \boldsymbol{\Sigma}_i^{obs^*,obs^*} & \boldsymbol{\Sigma}_i^{obs^*,pred} \\ \boldsymbol{\Sigma}_i^{pred,obs^*} & \boldsymbol{\Sigma}_i^{pred,pred} \end{pmatrix}$. As was mentioned in Wang[17] e Rao[36], the best linear predictor of $\mathbf{y}_{i,pred}$ with respect to the minimum mean squared error (MSE) criterion is the conditional expectation of $\mathbf{y}_{i,pred}$ given $\mathbf{y}_{i,obs^*}$, which, from Proposition 1, is given by

$$\widehat{\mathbf{y}}_{i,pred}(\boldsymbol{\theta}) = \mathbf{X}_{i,pred}\boldsymbol{\beta} + \boldsymbol{\Sigma}_i^{pred,obs^*}\boldsymbol{\Sigma}_i^{obs^*,obs^*-1}\left(\mathbf{y}_{i,obs^*} - \mathbf{X}_{i,obs^*}\boldsymbol{\beta}\right). \tag{12}$$

Therefore, $\mathbf{y}_{i,pred}$ can be estimated directly by substituting $\widehat{\boldsymbol{\theta}}$ into (12), leading to $\widehat{\widehat{\mathbf{y}}_{i,pred}} = \widehat{\mathbf{y}}_{i,pred}(\widehat{\boldsymbol{\theta}})$.

6 Simulation studies

In order to study the performance of our proposed method, we present two simulations studies. The first one shows the performance of t-MLC with DEC structure on the imputation procedure. The second one shows the asymptotic behavior of the ML estimates for our proposed model.

For both simulation schemes, we considered the t-MLC model defined in the Subsection 3.2, with parameters setting at $\beta_1 = 2.5$, $\beta_2 = 4$, $\sigma^2 = 4$, $\phi_1 = 0.8$ and $\phi_2 = 1$. In addition, the time points are set as $\mathbf{t}_i = (1, 3, 5, 7, 10, 14)^\top$, for $i = 1, \ldots, n$.

6.1 Imputation performance

As was mentioned above, the goal of this simulation study is to compare the performance of the t-MLC models with DEC structure under two scenarios: a) when the parameters ϕ_1 and ϕ_2 are unknown and estimated from the data, called unspecified (U) structure, and b) when $\mathbf{E}_i = \mathbf{I}_{n_i}$, that is, when an uncorrelated structure (UNC) is considered. For these purposes, we proceeded as follows:

1. We generated $M = 100$ data sets of size $n = 300$ from the t-MLC model with a DEC structure $\mathbf{E}_i = 0.8^{|t_{ij}-t_{ik}|}$, under four different settings of censoring proportions say, $\gamma = 5\%, 15\%, 25\%$ and 35%. It is important to note that, the goal here is to study the effect of the level of censoring in the estimation under misspecification of the correlation structure.

2. All the censored observations were imputed using the mechanism described in Subsection 4.2 and considering both, U and UNC structure.

In order to compare performance of the U and UNC structures through the EM-imputation defined in (11), we utilized two empirical discrepancy measures called the mean absolute error (MAE) and mean square error (MSE)[17,37]. They are defined by

$$\text{MAE} = \frac{1}{k}\sum_{i,j}|y_{ij} - \widetilde{y}_{ij}| \quad \text{and} \quad \text{MSE} = \frac{1}{k}\sum_{i,j}\left(y_{ij} - \widetilde{y}_{ij}\right)^2, \tag{13}$$

Censoring level	Correlation structure: Unspecified - U MAE	Unspecified - U MSE	Uncorrelated - UNC MAE	Uncorrelated - UNC MSE
5%	1.120052	2.744973	1.199131	3.075949
15%	1.293753	3.106423	1.563213	4.340442
25%	1.409025	3.902546	1.684068	5.475168
35%	1.568360	4.647703	1.830202	6.170776

Table 2: **Simulated data**. Arithmetics means of the MAE and MSE over $M = 100$ datasets.

where y_{ij} is the original simulated value (before being considered as a censored observation) and $\widetilde{y}_{ij}$ is the EM-imputation, for $i = 1, \ldots, 300$ and $j = 1, \ldots, 6$. Note that for $\gamma = 5\%$ we have that $k = 90$, for $\gamma = 15\%$, $k = 270$, for $\gamma = 25\%$ $k = 270$ and for $\gamma = 35\%$, $k = 630$.

Arithmetic means of MSE and MAE over the 100 datasets are displayed in Table 2 and Figure 2. We can see that in all cases, the U structure presents the smallest MSE and MAE than the UNC one, as expected.

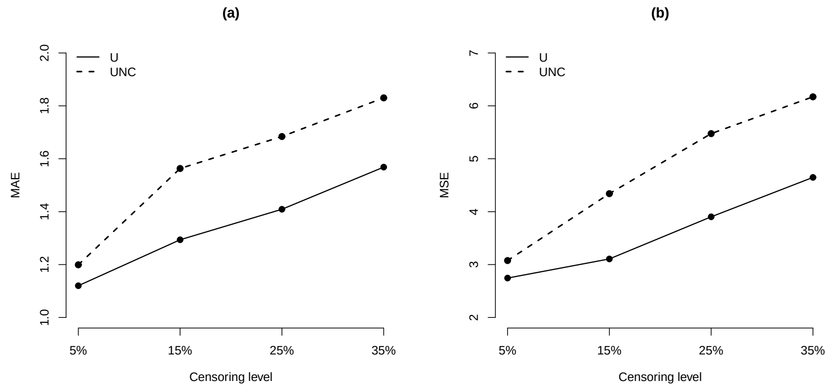

Figure 2: **Simulated data**. Arithmetics means of (a) MAE and (b) MSE over $M = 100$ datasets under the t-MLC model with U and UNC structure.

6.2 Asymptotic Properties

In this simulation study, we analyze the absolute bias (Bias) and mean square error (MSE) of the regression coefficient estimates obtained from the t-MLC model for six different sample sizes ($n = 50, 100, 200, 300, 400$ and 600). These measures are defined by

$$\text{Bias}\,(\theta_i) = \frac{1}{M}\sum_{j=1}^{M}\left|\widehat{\theta}_i^{(j)} - \theta_i\right| \quad \text{and} \quad \text{MSE}\,(\theta_i) = \frac{1}{M}\sum_{j=1}^{M}\left(\widehat{\theta}_i^{(j)} - \theta_i\right)^2,$$

where $\widehat{\theta}_i^{(j)}$ is the ML estimate of the parameter θ_i for the jth sample.

The idea of this simulation is to provide empirical evidence about consistency of the ML estimators under the t-MLC model with DEC structure. For each sample size, we generate $M = 100$ dataset with 5% of censoring proportion. In this simulation scheme the parameter ϕ_2 is fixed at 1 and thus a continuous-time AR(1) model is considered. Using the ECM algorithm, the absolute bias and mean squared error for each parameter over the 100 datasets were computed.

From Figure 3 we can see that the absolute bias and the MSE tend to zero as the sample size increase. As is expected, under the t-MLC model the EM algorithm provide estimates with good asymptotic properties.

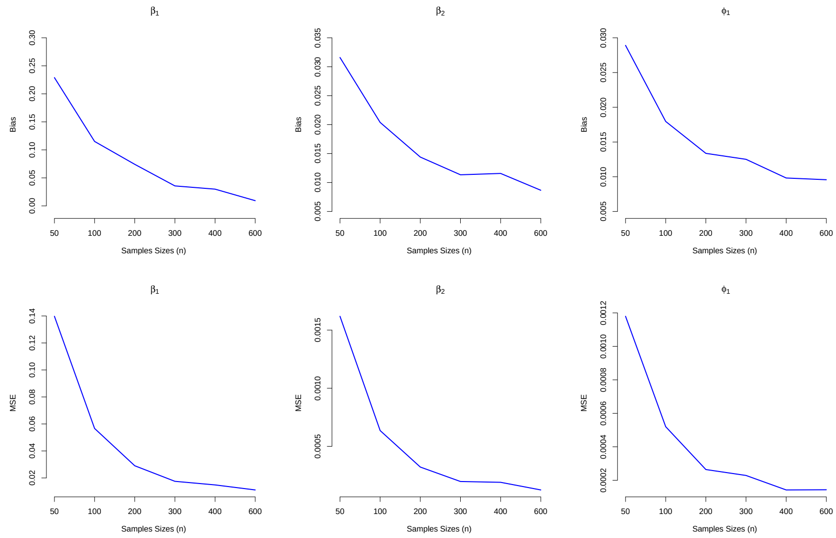

Figure 3: **Simulated data**. Bias (first row) and MSE (second row) of parameter estimates in the t-MLC model under 5% of censoring.

7 Application

We applied our method to the UTI data described in Section 2. This dataset consist of 362 observations, 26 were below the detection limits (50 or 400 copies/mL). The UTI data was analyzed previously by Lachos et al.[13], where it was observed that inferences based on Gaussian assumptions are questionable. Consequently, we revisited this dataset with the aim of carrying out robust inference by considering the Student-t model. We consider our proposed t-MLC model with the DEC correlation structure $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{E}_i$ defined in Subsection 3.2 and for the sake of model comparison, we also fit the normal MLC (N-MLC) counterparts, which can be treated as the reduced t-MLC as ν tends to infinity. Here we have that y_{ij} is the $\log_{10}$ HIV-1 RNA for subject i at time t_j, with $t_1 = 0$, $t_2 = 1$, $t_3 = 3$, $t_4 = 6$, $t_5 = 9$, $t_6 = 12$, $t_7 = 18$, and $t_8 = 24$.

We consider four cases of correlation structure induced by the specification of the matrix $\mathbf{E}_i$, namely, (a) the uncorrelated UNC structure, (b) the continuous-time AR(1) structure, (c) the MA(1) structure and (d) the unknown U structure (when ϕ_1 and ϕ_2 are unknown).

The degrees of freedom parameter ν was fixed at the value that maximizes the t-MLC likelihood function. Figure 4 shows that the likelihood function reaches the maximum at $\nu = 10$, indicating the lack of adequacy of the normal assumption for the UTI data. The ML estimates of the other parameters were obtained using the ECM algorithm described in Section 4, with starting values obtained through the library *lmec*[25].

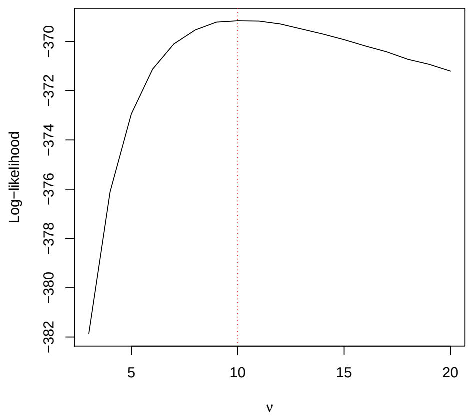

Figure 4: **UTI data**. Plot of the profile log-likelihood of the degrees of freedom ν.

Table 3 presents the ML estimates and standard errors of the regression parameters $\boldsymbol{\beta}$ for the t-MLC and N-MLC models. Although the estimates are quite similar in both cases, the standard errors are in general smaller under the Student-t, indicating that our proposed censored model (t-LMC) produces more precise estimates.

Note that we fitted eight models, resulting from the combinations of the four correlations struc-

Parameters	N-MLC							
	UNC		AR(1)		MA(1)		U	
	Est	SE	Est	SE	Est	SE	Est	SE
β_1	3.6160	0.0153	3.6334	0.0162	3.6194	0.0150	3.6196	0.0156
β_2	4.1527	0.0172	4.2095	0.0168	4.1825	0.0166	4.1834	0.0164
β_3	4.2381	0.0184	4.2502	0.0182	4.2384	0.0181	4.2568	0.0169
β_4	4.3727	0.0187	4.3224	0.0189	4.3729	0.0184	4.3738	0.0170
β_5	4.3650	0.0248	4.4680	0.0237	4.3652	0.0245	4.5791	0.0195
β_6	4.2326	0.0313	4.3781	0.0303	4.2327	0.0309	4.5819	0.0221
β_7	4.3258	0.0444	4.3749	0.0463	4.3260	0.0438	4.6879	0.0275
β_8	4.5620	0.0818	4.5762	0.0842	4.5620	0.0807	4.8061	0.0418
σ^2	1.0631		1.1498		1.0486		1.1053	
ϕ_1	–		0.8251		0.4068		0.7027	
ϕ_2	–		1.00		∞		0.0286	

Parameters	t-MLC							
	UNC		AR(1)		MA(1)		U	
	Est	SE	Est	SE	Est	SE	Est	SE
β_1	3.6511	0.0120	3.6410	0.0155	3.6578	0.0120	3.6330	0.0153
β_2	4.2386	0.0146	4.3022	0.0172	4.2706	0.0144	4.2697	0.0171
β_3	4.3149	0.0156	4.3312	0.0187	4.3246	0.0156	4.3290	0.0177
β_4	4.4715	0.0159	4.4297	0.0195	4.4792	0.0159	4.4715	0.0178
β_5	4.5268	0.0210	4.5476	0.0248	4.5293	0.0209	4.6359	0.0206
β_6	4.3923	0.0267	4.4435	0.0317	4.3963	0.0266	4.6238	0.0235
β_7	4.5012	0.0373	4.4660	0.0475	4.5092	0.0377	4.7082	0.0295
β_8	4.6896	0.0692	4.6481	0.0863	4.5092	0.0687	4.7998	0.0455
σ^2	0.8092		1.0272		0.8003		1.0103	
ϕ_1	–		0.7754		0.2752		0.6629	
ϕ_2	–		1.00		∞		0.0222	
ν	10.00		10.00		10.00		10.00	–

Table 3: **UTI data**. ML estimation and standard errors for the regression coefficients under the normal and Student-t MLC models with different DEC structures.

tures, say UNC, AR(1), MA(1) and U, and two distributional assumptions (normal and Student-t). The value of the log-likelihood function and the AIC and BIC criteria for these 8 models are presented in Table 4, where we can see that the t-MLC outperform consistently the normal counterpart in all cases. In particular, these criteria indicate a preference of the unspecified correlation structure (U structure), that is, when the parameters ϕ_1 and ϕ_2 of the matrix $\mathbf{E}_i$ are estimated from the data.

The regression coefficients β_j, for $j = 1, \ldots, 8$, increase gradually under the two models. This evidence the negative effect of the interruption of the antiretroviral therapy on the viral load levels. In other words, the viral load increments consistently along the time when the antiretroviral therapy begins to be interrupted. For the best model (t-MLC), the coefficients increase from 3.63 at the beginning of the study to 4.79 at the end of this. Note that considering an asymptotic 95% confidence interval, the estimates of all regression coefficients are significant. The estimate of the within-subject (σ^2) scale parameter (in $\log_{10}$ scale) is 1.01.

Outlying observations may affect the estimation of the parameters under assumptions of normality. Our t-MLC model with DEC structure accommodates these discrepant observations attributing to them small weights in the estimation procedure. The estimated weights ($\widehat{u}_i, i = 1, \ldots, 72$) for

	N-MLC				t-MLC			
Criteria	**UNC**	**AR(1)**	**MA(1)**	**U**	**UNC**	**AR(1)**	**MA(1)**	**U**
log-likelihood	-524.166	-463.043	-516.507	-411.926	-484.165	-421.249	-476.647	**-369.129**
AIC	1066.333	946.087	1053.014	845.852	986.331	862.498	973.295	**760.259**
BIC	1101.357	985.004	1091.931	888.660	1021.357	901.415	1012.212	**803.067**
AIC_{corr}	1066.844	946.714	1053.641	846.607	986.843	863.125	973.922	**761.014**

Table 4: **UTI data**. Comparison between the normal and Student-t MLC models using different model selection criteria.

the t-MLC model with U structure are presented in Figure 5. In this Figure, we observe that observations ♯20, ♯35, ♯41 and ♯42 present smaller weights, verifying the robust aspects of the ML estimation under the Student-t distribution. These results agree with those obtained by Lachos et al.[13] under the Bayesian paradigm.

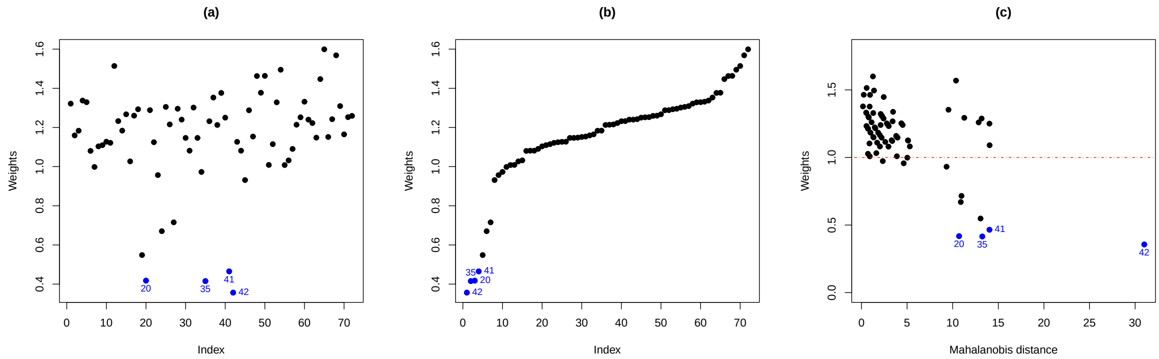

Figure 5: **UTI data**. Estimated weights $\widehat{u}_i$ for the t-MLC model with unspecific U structure.

Now we turn our attention to the one-step-ahead and two-step-ahead forecast of future observations using the approach proposed in Section 5 for the UTI data. As a simple illustration, we considered in the analysis the cases who were measured on at least six occasions (29 individuals in total) and we predicted the last two measures. As in the simulation scheme presented in Subsection 6.1, we considered the MAE and MSE measures for comparing the performance of the prediction under different DEC structures. Table 5 shows the comparison between the predicted values (one-step-ahead and two-step-ahead) with the real ones, under the t-MLC model considering three different DEC structures, say, AR(1), MA(1) and U. Figure 6 shows the comparison between the real data with the predicted values, two-step-ahead, under three different DEC structures: AR(1), MA(1) and U for the individuals ♯4, ♯15, ♯61. We can see from these results once again how the U structure outperforms the other correlation structures from a predictive point of view, *i.e.* the U structure generates predictive values close to the real ones.

	t-MLC					
	U		**AR(1)**		**MA(1)**	
Forecast	MAE	MSE	MAE	MSE	MAE	MSE
one step	0.3308357	0.1912845	0.4388853	0.2702304	0.623717	0.5197352
two step	0.3721411	0.2159791	0.5222417	0.5049302	0.6417997	0.702774

Table 5: **UTI data**. Evaluation of the prediction accuracy for the t-MLC model with different DEC structures.

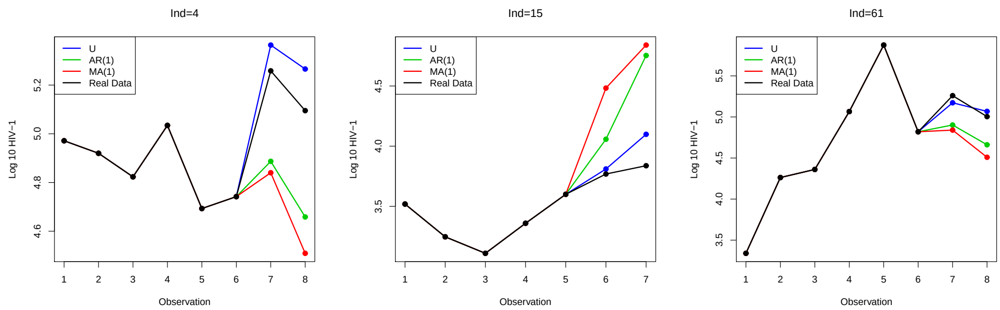

Figure 6: **UTI data**. Evaluation of the prediction performance for three random subjects.

8 Conclusions

We have proposed a robust approach to linear regression models with censored observations based on the multivariate Student-t distribution, called the t-MLC model. This offers a flexible alternative for dealing with longitudinal censored data in the presence of outliers and/or influential observations. For modeling the autocorrelation existing among irregularly observed measures, a damped exponential correlation structure was adopted as proposed by Muñoz et al.[18]. A novel ECM algorithm to obtain the ML estimates is developed by exploring the statistical properties of the multivariate truncated Student-t distribution. Our proposed algorithm has a closed-form expression for the E-step, based on formulas for the mean and variance of the truncated Student-t distribution. We applied our methods to a recent AIDS study (freely downloadable from R), concluding that when the antiretroviral therapy is interrupted, the HIV-1 RNA levels in blood increase consistently along the period of evaluation. We also perform two simulation studies, showing the superiority of t-MLC model on the provision of more adequate results when the available data has censored components. Furthermore, the simulation results demonstrate that our method gives very competitive performance in terms of imputation when a DEC structure is considered. From these results it is encouraging that the use of the t-MLC model with DEC structure offer a better fit, protection against outliers and more precise inferences.

An established way to validate the model (2) is to use the bootstrap technique (see e.g. Efron and Tibshirani[38] or Chernick[39]). Here, the appropriate way to use bootstrap will be to bootstrap observations $(Y_i, X_i),\ i = 1, \dots, n$. This means drawing randomly with replacement from the set of indices $\{1, \dots, n\}$ and obtaining a "new" sample of the same size n as the original sample. Since the bootstrap sampling is with replacement, it is quite likely here to draw twice or more the observation with the same index i, where $1 \le i \le n$. For each of the new sample we can recalculate the estimates of the parameter β in the model. After creating B samples we can have a data-driven approximation of the sampling distribution of the estimate $\hat{\beta}$ and a better idea on variability of $\hat{\beta}$. This now can be compared with the chosen p dimensional Student-t distribution to check whether this parametric model is reflecting data-driven sampling distribution of $\hat{\beta}$ and its variance.

Although the t-MLC considered here has shown great flexibility for modeling symmetric data, its robustness against outliers can be seriously affected by the presence of skewness. Recently, Lachos et al.[40] (see also Bandyopadhyay et al.[14]) proposed a remedy to accommodate skewness and heavy-tailedness simultaneously, using scale mixtures of skew-normal (SMSN) distributions. We conjecture that our methods can be used under MLC models, and should yield satisfactory results at the expense of additional complexity in the implementation. An in-depth investigation of such extensions is beyond the scope of the present paper, but it is an interesting topic for further research.

Acknowledgment

Victor H. Lachos and Aldo M. Garay would like to acknowledge the support of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil) and the Fundação de Amparo à Pesquisa do Estado de São Paulo (Grant 2013/21468-0 from FAPESP-Brazil). Luis M. Castro acknowledges funding support by Grant FONDECYT 1130233 from the Chilean government and Grant 2012/19445-0 from FAPESP-Brazil.

Appendix: Details of the ECM algorithm

In this Appendix, we derive the ECM algorithm equations (7)–(9) for the t-MLC model. Let $\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top)^\top$, $\mathbf{u} = (u_1, \dots, u_n)^\top$, $\mathbf{V} = vec(\mathbf{V}_1, \dots, \mathbf{V}_n)$, and $\mathbf{C} = vec(\mathbf{C}_1, \dots, \mathbf{C}_n)$ such that we observe $(\mathbf{V}_i, \mathbf{C}_i)$ for the i-th subject. Treating $\mathbf{u}$ and $\mathbf{y}$ as hypothetical missing data, and augmenting with the observed data $\mathbf{V}, \mathbf{C}$, we set $\mathbf{y}_c = (\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{y}^\top, \mathbf{u}^\top)^\top$.
Denoting the complete-data likelihood by $L(\cdot|\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{y}^\top, \mathbf{u}^\top)$ and pdf's in general by $f(\cdot)$, we have that for $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \phi_1, \phi_2)^\top$

$$
\begin{aligned}
L\left(\boldsymbol{\theta}|\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{y}^\top, \mathbf{u}^\top\right) &= f\left(\mathbf{y}|\mathbf{V}, \mathbf{C}, \mathbf{u}\right) f\left(\mathbf{u}\right) \\
&= f(\mathbf{y}|\mathbf{u}) f(\mathbf{u}) = \prod_{i=1}^{n} f(\mathbf{y}_i|u_i) h(u_i|\boldsymbol{\nu}).
\end{aligned}
$$

Dropping unimportant constants, the complete-data log-likelihood function is given by

$$
\begin{aligned}
\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) &= \log\{L[\boldsymbol{\theta}|\mathbf{y}_c]\} = \log\left\{\prod_{i=1}^{n} f(\mathbf{y}_i|u_i)h(u_i|\nu)\right\} \\
&= \sum_{i=1}^{n}\log\left\{(2\pi)^{-p/2}u_i^{1/2}|\Sigma_i|^{-1/2}\exp\left(\frac{u_i}{2}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})^\top\Sigma_i^{-1}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})\right)\right\} \\
&+ \sum_{i=1}^{n}\log\{h(u_i|\boldsymbol{\nu})\}, \\
&= -\frac{1}{2}\sum_{i=1}^{n}\left[n_i\log\sigma^2+\log|\mathbf{E}_i|+\frac{u_i}{\sigma^2}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})^\top\mathbf{E}_i^{-1}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})\right] \\
&+ \sum_{i=1}^{n}\log\{h(u_i|\nu)\}+c,
\end{aligned}
$$

where c is a constant that is independent of the parameter vector $\boldsymbol{\theta}$ and $h(u_i|\nu)$ is the gamma density (Gamma$(\nu/2,\nu/2)$). Our EM-type algorithm (ECM) for the t-MLC model can be summarized in the following way

E-step:
Given the current value $\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(k)}$, the E-step calculates the conditional expectation of the complete data log-likelihood function

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) &= \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}), \\
&= \sum_{i=1}^{n}\left\{-\frac{n_i}{2}\log\sigma^2-\frac{1}{2}\log|\mathbf{E}_i|-\frac{1}{2\sigma^2}A_i^{(k)}(\boldsymbol{\beta},\boldsymbol{\phi})\right\},
\end{aligned}
$$

with

$$
A_i^{(k)}(\boldsymbol{\beta},\boldsymbol{\phi}) = \left[\mathrm{tr}\left(\widehat{u\mathbf{y}_i^2}^{(k)}\mathbf{E}_i^{-1}\right)-2\boldsymbol{\beta}\mathbf{X}_i^\top\mathbf{E}_i^{-1}\widehat{u\mathbf{y}}_i^{(k)}+\widehat{u}_i^{(k)}\boldsymbol{\beta}\mathbf{X}_i^\top\mathbf{E}_i^{-1}\mathbf{X}_i\boldsymbol{\beta}\right].
$$

Note that in this case we do not consider the computation of $E[h(u_i|\nu)|\mathbf{V},\mathbf{C},\widehat{\boldsymbol{\theta}}^{(k)}]$ because ν is fixed.

CM-step:
The conditional maximization (CM) step then conditionally maximizes $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}=(\boldsymbol{\beta}^\top,\sigma^2,\phi_1,\phi_2)^\top$ and obtains a new estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$

$$
\begin{aligned}
\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})}{\partial\boldsymbol{\beta}} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left[\mathbf{X}_i^\top\left(\widehat{\mathbf{E}}_i^{(k)}\right)^{-1}\widehat{u\mathbf{y}}_i^{(k)}-\left(\widehat{u}_i^{(k)}\mathbf{X}_i^\top\left(\widehat{\mathbf{E}}_i^{(k)}\right)^{-1}\mathbf{X}_i\right)\boldsymbol{\beta}\right]; \\
\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})}{\partial\sigma^2} &= -\frac{N}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}A_i^{(k)}(\widehat{\boldsymbol{\beta}}^{(k+1)},\widehat{\boldsymbol{\phi}}^{(k)}).
\end{aligned}
$$

Thus, the solutions of $\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\beta}} = 0$ and $\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})}{\partial \sigma^2} = 0$ are

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^{(k+1)} &= \left(\sum_{i=1}^{n} \widehat{u}_i^{(k)} \mathbf{X}_i^\top \left(\widehat{\mathbf{E}}_i^{(k)}\right)^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i^\top \left(\widehat{\mathbf{E}}_i^{(k)}\right)^{-1} \widehat{u\mathbf{y}}_i^{(k)}, \\
\widehat{\sigma^2}^{(k+1)} &= \frac{1}{N} \sum_{i=1}^{n} A_i^{(k)}(\widehat{\boldsymbol{\beta}}^{(k+1)}, \widehat{\boldsymbol{\phi}}^{(k)}),
\end{aligned}$$

where $N = \sum_{i=1}^{n} n_i$. We estimate $\boldsymbol{\phi}$ by maximizing the marginal log-likelihood, circumventing the (in general) complicated task of computing $\frac{\partial \mathbf{E}_i}{\partial \boldsymbol{\phi}}$. This strategy were used, for instance, by Wang and Fan[37] and Wang[17]. Then,

$$\widehat{\boldsymbol{\phi}}^{(k+1)} = argmax_{\boldsymbol{\phi}} \left\{ -\frac{1}{2} \sum_{i=1}^{n} [\log(|\mathbf{E}_i|) + A_i^{(k)}(\widehat{\boldsymbol{\beta}}^{(k+1)}, \boldsymbol{\phi})] \right\},$$

The algorithm is iterated until the distance involving two successive evaluations of the log-likelihood, $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)})/\ell(\widehat{\boldsymbol{\theta}}^{(k)}) - 1|$, is sufficiently small.

References

[1] Ndembi N, Goodall R, Dunn D, McCormick A, Burke A, Lyagoba F, et al. Viral Rebound and Emergence of Drug Resistance in the Absence of Viral Load Testing: A Randomized Comparison between Zidovudine-Lamivudine plus Nevirapine and Zidovudine-Lamivudine plus Abacavir. Journal of Infectious Diseases. 2010;201(1):106–113.

[2] Wu L, Liu W, Hu X. Joint inference on HIV viral dynamics and immune suppression in presence of measurement errors. Biometrics. 2010;66:327–335.

[3] Qiu W, Wu L. HIV viral dynamic models with censoring and informative dropouts. Statistics in Biopharmaceutical Research. 2010;2(2):220–228.

[4] Laird N, Ware J. Random effects models for longitudinal data. Biometrics. 1982;38:963–974.

[5] Pinheiro JC, Bates M Douglas. Mixed-Effects Models in S and S-PLUS. New York, NY: Springer; 2000.

[6] Davidian M, Giltinan D. Nonlinear models for repeated measurements: An overview and update. Journal of Agricultural, Biological, and Environmental Statstics. 2003;8:387–419.

[7] Pinheiro JC, Liu CH, Wu YN. Efficient algorithms for robust estimation in linear mixed-effects models using a multivariate t-distribution. Journal of Computational and Graphical Statistics. 2001;10:249–276.

[8] Rosa GJM, Padovani CR, Gianola D. Robust linear mixed models with Normal/Independent distributions and Bayesian MCMC implementation. Biometrical Journal. 2003;45:573–590.

[9] Lin TI, Lee JC. Bayesian analysis of hierarchical linear mixed modeling using the multivariate t distribution. Journal of Statistical Planning and Inference. 2007;137(2):484–495.

[10] Meza C, Osorio F, De la Cruz R. Estimation in nonlinear mixed-effects models using heavy-tailed distributions. Statistics and Computing. 2011;22:1–19.

[11] Lachos V, Castro L, Dey D. Bayesian inference in nonlinear mixed-effects models using normal independent distributions. Computational Statistical & Data Analysis. 2013;64:237–252.

[12] Castro LM, Lachos VH, Galvis D, Bandyopadhyay D. Bayesian semiparametric longitudinal data modeling using normal/independent densities. In: Current Trends in Bayesian Methodology with Applications. Chapman & Hall/CRC; 2014. .

[13] Lachos VH, Bandyopadhyay D, Dey DK. Linear and nonlinear mixed-effects models for censored HIV viral loads using normal/independent distributions. Biometrics. 2011;67:1594–1604.

[14] Bandyopadhyay D, Lachos V, Castro L, Dey D. Skew-normal/independent linear mixed models for censored responses with applications to HIV viral loads. Biometrical Journal. 2012;54:405–425.

[15] Arellano-Valle RB, Castro LM, González-Farías G, Muños Gajardo K. Student-t censored regression model: properties and inference. Statistical Methods and Applications. 2012;21(4):453–473.

[16] Castro LM, Lachos VH, Arellano-Valle RB. Partially linear censored regression models using heavy-tailed distributions: a Bayesian approach. Statistical Methodology. 2014;18:14–31.

[17] Wang WL. Multivariate t linear mixed models for irregularly observed multiple repeated measures with missing outcomes. Biometrical Journal. 2013;55(4):554–571.

[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] Wang WL, Fan TH. Estimation in multivariate t linear mixed models for multivariate longitudinal data. Statistica Sinica. 2011;21:1857–1880.

[20] Goldstein H, Healy M, Rasbash J. Multilevel time series models with applications to repeated measures data. Statistics in Medicine. 1994;13(16):1643–1655.

[21] Browne W, Goldstein H. MCMC sampling for a multilevel model with nonindependent residuals within and between cluster units. Journal of Educational and Behavioral Statistics. 2010;35(4):453–473.

[22] Matos LA, Prates MO, Chen MH, Lachos VH. Likelihood-based inference for mixed-effects models with censored response using the multivariate-t distribution. Statistica Sinica. 2013;23:1323–1342.

[23] Ho HJ, Lin TI, Chen HY, Wang WL. Some results on the truncated multivariate t distribution. Journal of Statistical Planning and Inference. 2012;142:25–40.

[24] Saitoh A, Foca M, Viani RM, Heffernan-Vacca S, Vaida F, Lujan-Zilbermann J, et al. Clinical outcomes after an unstructured treatment interruption in children and adolescents with perinatally acquired HIV infection. Pediatrics. 2008;121(3):e513–e521.

[25] Vaida F, Liu L. Fast implementation for normal mixed effects models with censored response. Journal of Computational and Graphical Statistics. 2009;18(4):797–817.

[26] Arellano-Valle RB, Bolfarine H. On some characterizations of the t-distribution. Statistics & Probability Letters. 1995;25:79–85.

[27] Lange KL, Little RJA, Taylor JMG. Robust statistical modeling using t distribution. Journal of the American Statistical Association. 1989;84:881–896.

[28] Lucas A. Robustness of the student t based M-estimator. Communications in Statistics-Theory and Methods. 1997;26(5):1165–1182.

[29] Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B,. 1977;39:1–38.

[30] Couvreur C. The EM algorithm: a guided tour. In: Proceedings of the 2d IIEE European Workshop on Computationaly Intensive Methos in Control and Signal Processing; 1996. .

[31] Meng X, Rubin DB. Maximum likelihood estimation via the ECM algorithm: A general framework. Biometrika. 1993;81:633–648.

[32] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2014. Available from: `http://www.R-project.org/`.

[33] Genz A, Bretz F, Hothorn T, Miwa T, Mi X, Leisch F, et al. mvtnorm: Multivariate Normal and t Distribution. R package version 09-2, URL http://CRANR-projectorg/package=mvtnorm. 2008;.

[34] Louis TA. Finding the observed information matrix when using the EM algorithm. Journal of the Royal Statistical Society, Series B. 1982;44:226–233.

[35] Rao CR. Prediction of future observations in growth curve models. Statistical Science. 1987;2:434–447.

[36] Rao CR. Linear Statistical Inference and its Applications, Second Edition,. John Wiley & Sons; 1973.

[37] Wang WL, Fan TH. ECM based maximum likelihood inference for multivariate linear mixed models with autoregressive errors. Computational Statistics & Data Analysis. 2010;54(5):1328 – 1341.

[38] Efron B, Tibshirani R. An Introduction to Bootstrap. Chapman & Hall/CRC; 1993.

[39] Chernick MR. Bootstrap Methods: A Guide for Practitioners and Researchers. Wiley; 2008.

[40] Lachos VH, Ghosh P, Arellano-Valle RB. Likelihood based inference for skew–normal independent Linear mixed models. Statistica Sinica. 2010;20:303–322.