Extending multivariate-t semiparametric mixed models for longitudinal data with censored responses and heavy tails

Thalita B. Mattos[1a], Larissa A. Matos[a], and Victor H. Lachos[b*]
[a] *Departamento de Estatística, Universidade Estadual de Campinas, Brazil*
[b] *Department of Statistics, University of Connecticut, USA*

Abstract

In this paper we extended the semiparametric mixed model for longitudinal censored data with normal errors to Student-t erros. This models allows flexible functional dependence of an outcome variable on covariates by using nonparametric regression, while accounting for correlation between observations by using random effects. Penalized likelihood equations are applied to derive the maximum likelihood estimates which appear to be robust against outlying observations in the sense of the Mahalanobis distance. We estimate nonparametric functions by using smoothing splines jointly estimate smoothing parameter by the EM algorithm. Finally, the performance of the proposed approach is evaluated through extensive simulation studies as well as application to dataset from AIDS study.

Keywords: Censored data; EM algorithm; HIV viral load; Linear mixed-effects; Semiparametric models; multivariate-t distribution.

1. Introduction

Longitudinal data analysis has attracted considerable research interest and a large number of statistical modeling and analysis methods have been suggested to analyze such data with various features. Linear and nonlinear mixed effects (LME and NLME, respectively) are parametric models for longitudinal data that have been extensively studied in the last few decades; see Davidian & Giltinan (1995); Diggle (2002); Pinheiro & Bates (2006) among others, for more ideas and methodologies for longitudinal data analysis using parametric modeling. These models are very useful for longitudinal data analysis, as they provide a parsimonious description of the relationship between the response and its covariance. However, parametric models are efficient when they are correctly specified, the model misspecification can result in biased estimation. To relax the assumptions on parametric forms, an attractive approach is the semiparametric mixed model, which retains the flexibility of the nonparametric model while preserving good properties such as easy implementation and good interpretability of parametric models.

Semiparametric mixed models have received great attention in the literature with approaches based on kernel smoothing (Zeger & Diggle, 1994), or, more often, on smoothing spline (Zhang

* *Address for correspondence*: Víctor Hugo Lachos Dávila, Department of Statistics, University of Connecticut, Storrs, CT 06269, U.S.A. E-mail: hlachos@uconn.edu

et al., 1998). However, these models (LME/NLME and semiparametric) are in general made on the assumption of Gaussian errors. Some studies have investigated alternative distributions for errors in LME/NLME, for example, Pinheiro *et al.* (2001) propose a robust hierarchical linear mixed model in which the random effects and the within-subject errors have a multivariate t-distribution. Moreover, Meza *et al.* (2012) presented an extension of a Gaussian nonlinear mixed effects model considering a class of heavy tailed multivariate distributions for both random effects and residual errors. In the semiparametric context, Ibacache-Pulgar *et al.* (2012) extended semiparametric mixed linear models with normal errors to elliptical errors in order to permit distributions with heavier and lighter tails than the normal ones.

At the same time, longitudinal data can be complicated when the response is censored for some of the observations due to an assay detection limit used to quantify the marker. For example, this can occur when measuring the chemical content of a collection of samples (Palarea-Albaladejo & Martin-Fernandez, 2013), when measuring the concentration of some pollutants in environmental data (Helsel, 2011) or measuring Human Immunodeficiency Virus viral load in blood compartment (HIV RNA) (Hughes, 1999). Several methods have been proposed to deal with such limits of detection, censored mixed-effects models are frequently used in the analysis of longitudinal AIDS data. Lachos *et al.* (2011) considered a Bayesian treatment of the linear mixed model with censored responses (LMEC) and the nonlinear mixed model with censored responses (NLMEC) models based on the normal/independent distributions. Further, Matos *et al.* (2013b) developed a likelihood-based inference for LMEC and NLMEC based on the multivariate-t distribution, named as tLMEC and tNLMEC.

The aim of this paper is to consider the study of censored mixed-effects models using, simultaneously, semiparametric techniques such as smoothing splines and the distribution-t multivariate, due to its capability of down-weighting out lying observations. This paper is organized as follows. Section 2 describes the multivariate-t distribution and some of its properties. In Section 3, the Student-t semiparametric censored mixed-effects model is defined, where the estimation and inference procedures of the regression coefficients, nonparametric function, and scale parameter are presented. Some inferential results and discussions of estimation of the smoothing parameter are given in Section 4. Moreover, in Section 5, the goodness of fit and model selection procedures are proposed to check the quality of fit. Some simulation results are presented in Section 6 and an application to the data set of HIV viral loads is presented in Section 7. Finally, in Section 8 some concluding remarks are given with some future research directions.

2. The multivariate t distribution

In this section we present the p-variate Student's t-distribution and some of its useful properties. The following properties are useful for the implementation of the expectation maximization (EM) algorithm.

A random variable $\mathbf{Y}$ having a p-variate t distribution with location vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$ (positive definite) and degrees of freedom ν ($\nu > 0$) denoted by $\mathbf{Y} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, has the probability density fucntion (pdf):

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

where $\Gamma(\cdot)$ is the standard gamma function and $\delta^2(\mathbf{y}) = (\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})$ is the Mahalanobis distance. The cumulative distribution fucntion (cdf) of $\mathbf{Y}$ is denoted by

$$\mathrm{T}_p(\mathbf{b}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) = \int_{-\infty}^{\mathbf{b}} t_p(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)d\mathbf{y}.$$

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.

$$\mathbf{Y} = \boldsymbol{\mu} + U^{-1/2}\mathbf{Z}, \quad \mathbf{Z} \sim \mathrm{N}_p(\mathbf{0},\boldsymbol{\Sigma}), \quad U \sim \mathrm{Gamma}(\nu/2,\nu/2),$$

where $\mathbf{Z}$ and U are independent and $\mathrm{Gamma}(\alpha,\beta)$ stands for a gamma distribution with mean α/β, and density denoted by $\mathrm{G}(\cdot|\alpha,\beta)$. It is important to stress that if $\nu > 1$, $\boldsymbol{\mu}$ is the mean of $\mathbf{Y}$, and if $\nu > 2$, $(\nu/(\nu-2))\boldsymbol{\Sigma}$ is its covariance matrix. As $\nu \to \infty$, U converges to one with probability one, and so $\mathbf{Y}$ becomes marginally multivariate normal with mean μ and covariance matrix $\boldsymbol{\Sigma}$, denoted by $\mathrm{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$.

In order to introduce some notation, for the multivariate t-distribution, the following property is useful for our theoretical developments. We start with the marginal-conditional decomposition of a Student's t random vector. Details of the proofs are provided in Arellano-Valle & Bolfarine (1995).

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* $p = p_1 + p_2$. *Let* $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top,\boldsymbol{\mu}_2^\top)^\top$ *and* $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$ *be the corresponding partions of* $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$. *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*

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

where $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1-\boldsymbol{\mu}_1)$ *and* $\widetilde{\boldsymbol{\Sigma}}_{22.1} = \left(\dfrac{\nu+\delta^2(\mathbf{y}_1)}{\nu+p_1}\right)\boldsymbol{\Sigma}_{22.1}$ *with* $\delta^2(\mathbf{y}_1) = (\mathbf{y}_1 - \boldsymbol{\mu}_1)^\top\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1-\boldsymbol{\mu}_1)$ *and* $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$.

A p-dimensional random vector $\mathbf{Y}$ is said to follow a truncated Student's t distribution with location $\boldsymbol{\mu}$, scale-covariance matrix $\boldsymbol{\Sigma}$ and degrees of freedom ν over the truncation region $\mathbb{A} = \{(y_1,\ldots,y_p) \in \mathbb{R}^p : a_1 \le y_1 \le b_1,\ldots,a_p \le y_p \le b_p\} = \{\mathbf{y} \in \mathbb{R}^p : \mathbf{a} \le \mathbf{y} \le \mathbf{b}\}$, denoted by $\mathbf{Y} \sim \mathrm{Tt}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu;\mathbb{A})$, if its density is given by:

$$f(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu;\mathbb{A}) = \frac{t_p(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}{\mathrm{T}_p(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}, \quad \mathbf{a} \le \mathbf{y} \le \mathbf{b}.$$

The following results provide the truncated moments of a Student's t random vector. The proofs of Proposition 2 and 3 are given in Matos *et al.* (2013b).

Proposition 2. *If* $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu;(\mathbf{a},\mathbf{b}))$ *then it holds that*

$$\mathbb{E}\left[\left(\frac{\nu+p}{\nu+\delta^2(\mathbf{Y})}\right)^r \mathbf{Y}^{(k)}\right] = c_p(\nu,r)\frac{T_p(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu+2r)}{T_p(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}\mathbb{E}[\mathbf{W}^{(k)}], \quad k = 0,1,2,$$

where $c_p(\nu,r) = \left(\dfrac{\nu+p}{\nu}\right)^r\left(\dfrac{\Gamma((p+\nu)/2)\Gamma((\nu+2r)/2)}{\Gamma(\nu/2)\Gamma((p+\nu+2r)/2)}\right)$, $\boldsymbol{\Sigma}^* = \dfrac{\nu}{\nu+2r}\boldsymbol{\Sigma}$,
$\mathbf{W} \sim Tt_p(\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu+2r;(\mathbf{a},\mathbf{b}))$, $\mathbf{W}^{(0)} = 1$, $\mathbf{W}^{(1)} = \mathbf{W}$, $\mathbf{W}^{(2)} = \mathbf{W}\mathbf{W}^\top$ *and* $\nu+2r > 0$.

Proposition 3. *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 of* $\mathbf{a}$, $\mathbf{b}$, $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$. *Then, under the notation of Proposition 1, the conditional k-th moment of* $\mathbf{Y}_2$ *is*

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

where $d_p(p_1,\nu,r) = (\nu+p)^r\left(\frac{\Gamma((p+\nu)/2)\Gamma((p_1+\nu+2r)/2)}{\Gamma((p_1+\nu)/2)\Gamma((p+\nu+2r)/2)}\right)$,
$\widetilde{\boldsymbol{\Sigma}}^*_{22.1} = \left(\frac{\nu+\delta^2(\mathbf{y}_1)}{\nu+2r+p_1}\right)\boldsymbol{\Sigma}_{22.1}$, $\mathbf{W} \sim Tt_{p_2}(\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}^*_{22.1}, \nu+p_1+2r; (\mathbf{a}_2,\mathbf{b}_2))$, $\mathbf{W}^{(0)} = 1$, $\mathbf{W}^{(1)} = \mathbf{W}$, $\mathbf{W}^{(2)} = \mathbf{W}\mathbf{W}^\top$ *and* $\nu + p_1 + 2r > 0$, $k = 0, 1, 2$.

3. The Student-t semiparametric mixed effects model with censored responses

3.1. The model specification

Let the sample consist of n subjects, with the ith subject having n_i observations over time. Let y_{ij} denote the measurement of the ith subject at time t_{ij}, then the semiparametric mixed model for outcome y_{ij} is given by

$$y_{ij} = \mathbf{x}_{ij}^\top\boldsymbol{\beta} + f(t_{ij}) + \mathbf{z}_{ij}^\top\mathbf{b}_i + \epsilon_{ij}, \quad i = 1,\dots,n, \quad j = 1,\dots,n_i, \tag{1}$$

where $\boldsymbol{\beta}$ is the $p \times 1$ vector of regression coefficients associated with covariates $\mathbf{x}_{ij}$ $(p \times 1)$, $f(\cdot)$ is a twice-differentiable smooth function of time, the $\mathbf{b}_i$ are independent $q \times 1$ vectors of random effects associated with covariates $\mathbf{z}_{ij}$ $(q \times 1)$, and the ϵ_{ij} are independent measurement errors.

In order to write model (1) computationally more advantageous, we can express in a matrix form as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{N}_i\mathbf{f} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \tag{2}$$

where $\mathbf{y}_i = (y_{i1},\dots,y_{in_i})^\top$ is a $(n_i \times 1)$ random vector of observed responses from the ith subject, $\mathbf{X}_i$ is an $n_i \times p$ design matrix with rows $\mathbf{x}_{ij}^\top$, $\mathbf{N}_i$ is an $n_i \times r$ incidence matrix for the ith subject connecting $\mathbf{t}_i$ and $\mathbf{t}^0$ such that the (j,s)th element of $\mathbf{N}_i$ equals the indicator function $\mathbb{I}(t_{ij} = t_s^0)$ for $j = 1,\dots,n_i$ and $s = 1,\dots,r$, $\mathbf{f} = (f(t_1^0),\dots,f(t_r^0))^\top$ with $t_1^0,\dots,t_r^0$ being the distinct and ordered values of t_{ij}, $\mathbf{Z}_i$ is the $n_i \times q$ design matrix of the random effects with $\mathbf{z}_{ij}^\top$ and $\boldsymbol{\epsilon}_i$ is an $n_i \times 1$ vector of within-subjects errors.

In this work, we assume that the random effects and the errors follow a Student-t distribution:

$$\begin{pmatrix}\mathbf{b}_i\\ \boldsymbol{\epsilon}_i\end{pmatrix} \overset{\text{ind.}}{\sim} t_{q+n_i}\left(\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\mathbf{D} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Omega}_i\end{pmatrix}, \nu\right), i = 1,\dots,n, \tag{3}$$

where ν represents the multivariate t-distribution degrees-of-freedom (df), $\mathbf{D}$ is a $q \times q$ symmetric positive-definite covariance matrix of the random effects ($\mathbf{b}_i$) that depends upon a set of unknown parameter vector $\boldsymbol{\alpha}$ and $\boldsymbol{\Omega}_i = \sigma^2\mathbf{E}_i$ represents the within-subject variance-covariance matrix for subject i, σ^2 is the scalar within-subject variance parameter and $\mathbf{E}_i$ is a $n_i \times n_i$ matrix that incorparate a time-dependence structure. Note that $\mathbf{b}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated, but not necessarily independent.

Munoz *et al.* (1992) proposed a family of correlation structures, damped exponential correlation (DEC) structure, which allows to deal with unequally spaced and unbalanced observations. We adopt the DEC structure for $\mathbf{E}_i$, defined as

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

where ϕ_1 is the correlation between observations separated by one t-unit in time and ϕ_2 is the "scale parameter", which permits attenuation or acceleration of the exponential decay of the autocorrelation function, defining a continuous-time autoregressive model. Examples of particular cases in this family of correlation structures include the compound symmetry (CS), AR(1), and MA(1) - moving average of order 1, correlation structures when ϕ_2 takes the values 0,1, and ∞, respectively. A more detailed discussion of the DEC structure can be found in Munoz *et al.* (1992).

It follows that the semiparametric mixed model with t-distribution assumes the following joint distribution:

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{b}_i \end{pmatrix} \overset{\text{ind.}}{\sim} t_{n_i+q}\left(\begin{pmatrix} \mathbf{X}_i\boldsymbol{\beta}+\mathbf{N}_i\mathbf{f} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top+\boldsymbol{\Omega}_i & \mathbf{Z}_i\mathbf{D} \\ \mathbf{D}\mathbf{Z}_i^\top & \mathbf{D} \end{pmatrix}, \nu\right). \tag{4}$$

Thus, the $\mathbf{y}_i$ are independent and marginally distributed as

$$\mathbf{y}_i \overset{\text{ind.}}{\sim} t_{n_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \nu),$$

where $\boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta}+\mathbf{N}_i\mathbf{f}$, $\boldsymbol{\Sigma}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top+\boldsymbol{\Omega}_i$, for $i = 1,\ldots,n$.

As mentioned earlier, the proposed model also considers censored observations, i.e., we assume that the response y_{ij} is not fully observed for all i,j. Let the observed data for the i-th subject be $(\mathbf{V}_i, \mathbf{C}_i)$, 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:

$$C_{ij} = \begin{cases} 1 & \text{if} \quad V_{1ij} \leq y_{ij} \leq V_{2ij}, \\ 0 & \text{if} \quad Y_{ij} = V_{0i}, \end{cases} \tag{5}$$

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 $V_{1ij} = -\infty$. The model defined in (1)-(5) is henceforth called the DEC-tSMEC model.

For responses with censoring pattern as in (5), we have that marginally

$$\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i \sim \mathrm{Tt}_{n_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \nu; \mathbb{A}),$$

where $\mathrm{Tt}_{n_i}(.;\mathbb{A})$ denotes the truncated Student-t distribution on the interval $\mathbb{A}$, $\mathbb{A}_i = A_{i1} \times \ldots \times A_{in_i}$, with A_{ij} being the interval $(-\infty,\infty)$ if $C_{ij} = 0$ and the interval $(V_{1ij}, V_{2ij}]$ if $C_{ij} = 1$.

3.2. The likelihood function

We are interested in maximum likelihood estimation of model (1) when $\mathbf{y}_i$ has a censored response. To compute the likelihood function associated with the model defined by (1)-(5), the first step is to treat separately the observed and censored components of $\mathbf{y}_i$. Let $\mathbf{y}_i^o$ be the n_i^o-vector of observed outcomes and $\mathbf{y}_i^c$ be the n_i^c-vector of censored observations for subject i 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:

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

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

Using properties of multivariate Student's-t distribution (see Arellano-Valle & Bolfarine, 1995), we have that

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

where

$$\boldsymbol{\mu}_i^o = \mathbf{X}_i^o\boldsymbol{\beta} + \mathbf{N}_i^o\mathbf{f}, \quad \boldsymbol{\mu}_i^c = \mathbf{X}_i^c\boldsymbol{\beta} + \mathbf{N}_i^c\mathbf{f}, \quad \boldsymbol{\mu}_i^{co} = \boldsymbol{\mu}_i^c + \boldsymbol{\Sigma}_i^{co}\boldsymbol{\Sigma}_i^{oo-1}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o),$$
$$\mathbf{S}_i^{co} = \left(\frac{\nu + \delta^2(\mathbf{y}_i^o)}{\nu + n_i^o}\right)\mathbf{S}_i, \quad \mathbf{S}_i = \boldsymbol{\Sigma}_i^{cc} - \boldsymbol{\Sigma}_i^{co}\boldsymbol{\Sigma}_i^{oo-1}\boldsymbol{\Sigma}_i^{oc} \quad \text{and}$$
$$\delta^2(\mathbf{y}_i^o) = (\mathbf{y}_i^o - \boldsymbol{\mu}_i^o)^\top\boldsymbol{\Sigma}_i^{oo-1}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o).$$

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \mathbf{f}^\top, \sigma^2, \boldsymbol{\alpha}^\top, \boldsymbol{\phi}, \nu)^\top$ be the parameters vector. From Matos *et al.* (2013a), the likelihood for subject i is given by

$$\begin{aligned} L_i(\boldsymbol{\theta}) = f(\mathbf{y}_i|\boldsymbol{\theta}) &= f(\mathbf{V}_i|\mathbf{C}_i, \boldsymbol{\theta}) \\ &= f(\mathbf{y}_i^o|\boldsymbol{\theta})P(\mathbf{V}_{1i}^c \le \mathbf{y}_i^c \le \mathbf{V}_{2i}^c|\mathbf{V}_i^o, \boldsymbol{\theta}) \\ &= t_{n_i^o}(\mathbf{V}_i^o; \boldsymbol{\mu}_i^o, \boldsymbol{\Sigma}_i^{oo}, \boldsymbol{\nu})\mathrm{T}_{n_i^c}(\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c; \boldsymbol{\mu}_i^{co}, \mathbf{S}_i^{co}, \nu + n_i^o) = L_i, \end{aligned} \tag{6}$$

where $T_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ denotes the cumulative distribution function (cdf) of the multivariate Students-t distribution with parameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ and ν.

The log-likelihood function for the observed data is given by $\ell(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^n \log L_i$, and the estimates obtained by maximizing the log-likelihood function $\ell(\boldsymbol{\theta}|\mathbf{y})$ are the maximum likelihood estimates (MLEs). For the reason that $f(\cdot)$ is an infinite-dimensional parameter, the direct maximization of (6) without imposing restrictions over the function $f(\cdot)$ may cause overfitting and non-identifiability of $\boldsymbol{\beta}$ (see Green, 1987). A well-know procedure based on the idea of log-likelihood penalization and consists of incorporating a penalty function in the log-likelihood, such that

$$\ell_p(\boldsymbol{\theta}|\mathbf{y}) = \ell(\boldsymbol{\theta}|\mathbf{y}) - \frac{\lambda}{2}J(\mathbf{f}), \tag{7}$$

where $J(\mathbf{f})$ denotes the penalty function over $\mathbf{f}$ and $\lambda \ge 0$ is a smoothing parameter which controls the tradeoff between goodness of fit and the smoothness estimated function. By maximizing (7), one obtains the MPL estimate.

Similarly to Ibacache-Pulgar *et al.* (2013), we will consider the following penalty function:

$$J(\mathbf{f}) = \int_a^b [f''(t)]^2 dt = \mathbf{f}^\top\mathbf{K}\mathbf{f},$$

where $[f''(t)]$ denotes the second derivative of $f(t)$ with $[a, b]$ containing the values t_j^0, of $j = 1, \dots, r$ and $\mathbf{K}$ is the nonnegative definite smoothing matrix that depends only on the knots defined in Green & Silverman (1994). In this case, the estimation of $\mathbf{f}$ leads to a smooth cubic spline with knots at the points t_j^0.

3.3. The EM algorithm for MPL estimation

In this subsection, we discuss the estimation of $\boldsymbol{\theta}$ based on penalized log-likelihood.

The EM algorithm (Dempster *et al.*, 1977) is a popular iterative algorithm for ML estimation of models with incomplete data and has several appealing features such as stability of monotone convergence and simplicity of implementation. We adopt a variant of the the EM-type algorithm, called the ECME algorithm, for computing MPL estimates of model parameters. Liu & Rubin (1994) showed that ECME typically shares with EM the simplicity and stability, but has a faster rate of convergence, especially for multivariate t-distribution with unknown degrees-of-freedom.

Based on the essential property of multivariate t-distribution, the model (4) can be expressed in the following hierarchical model:

$$\begin{aligned} \mathbf{y}_i|\mathbf{b}_i, u_i &\overset{\text{ind.}}{\sim} \mathrm{N}_{n_i}(\boldsymbol{\mu}_i, u_i^{-1}\boldsymbol{\Omega}_i), \\ \mathbf{b}_i|u_i &\overset{\text{ind.}}{\sim} \mathrm{N}_q(\mathbf{0}, u_i^{-1}\mathbf{D}), \\ u_i &\overset{\text{ind.}}{\sim} \mathrm{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \end{aligned} \tag{8}$$

where Gamma(a, b) denotes the gamma distribution with mean a/b and variance a/b^2. Thus, it is possible to apply the penalized EM algorithm (Green, 1990) 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 $\mathbf{V} = vec(\mathbf{V}_1, \ldots, \mathbf{V}_n)$, and $\mathbf{C} = vec(\mathbf{C}_1, \ldots, \mathbf{C}_n)$. Hence, the penalized log-likelihood function for the model based on complete data $\mathbf{y}_c = (\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{y}^\top, \mathbf{b}^\top, \mathbf{u}^\top)^\top$ is given by

$$\ell_{pc}(\boldsymbol{\theta}|\mathbf{y}_c) = \ell_c(\boldsymbol{\theta}|\mathbf{y}_c) - \frac{\lambda}{2}\mathbf{f}^\top\mathbf{K}\mathbf{f}, \tag{9}$$

with

$$\begin{aligned} \ell_c(\boldsymbol{\theta}|\mathbf{y}_c) &= \sum_{i=1}^{n}\left[-\frac{n_i}{2}\log\sigma^2 - \frac{1}{2}\log(|\mathbf{E}_i|) - \frac{u_i}{2\sigma^2}(\mathbf{y}_i - \boldsymbol{\mu}_i - \mathbf{Z}_i\mathbf{b}_i)^\top\mathbf{E}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i - \mathbf{Z}_i\mathbf{b}_i)\right. \\ &- \left.\frac{1}{2}\log|\mathbf{D}| - \frac{u_i}{2}\mathbf{b}_i^\top\mathbf{D}^{-1}\mathbf{b}_i + \log h(u_i|\nu) + C\right], \end{aligned} \tag{10}$$

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

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

$$\begin{aligned} Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) &= \mathbb{E}\left[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)\Big|\mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}\right] - \frac{\lambda}{2}\mathbf{f}^\top\mathbf{K}\mathbf{f}, \\ &= \sum_{i=1}^{n} Q_{1i}(\boldsymbol{\beta}, \mathbf{f}, \sigma^2, \boldsymbol{\phi}|\widehat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^{n} Q_{2i}(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}^{(k)}), \end{aligned}$$

where

$$\begin{aligned} Q_{1i}(\boldsymbol{\beta}, \mathbf{f}, \sigma^2, \boldsymbol{\phi}|\widehat{\boldsymbol{\theta}}^{(k)}) &= -\frac{n_i}{2}\log\sigma^2 - \frac{1}{2}\log(|\mathbf{E}_i|) \\ &- \frac{1}{2\sigma^2}\left[\widehat{a}_i^{(k)} - 2\boldsymbol{\mu}_i^\top\mathbf{E}_i^{-1}\left(\widehat{u_i\mathbf{y}_i}^{(k)} - \mathbf{Z}_i\widehat{u_i\mathbf{b}_i}^{(k)}\right)\widehat{u}_i^{(k)}\boldsymbol{\mu}_i^\top\mathbf{E}_i^{-1}\boldsymbol{\mu}_i\right] \\ &- \frac{\lambda}{2n}\mathbf{f}^\top\mathbf{K}\mathbf{f} \end{aligned}$$

and

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

with

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

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

The conditional maximization (CM) steps then conditionally maximizes $Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$ and obtains a new estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$, as follows:

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \left(\sum_{i=1}^m \widehat{u_i}^{(k)}\mathbf{X}_i^\top\widehat{\mathbf{E}}_i^{-1(k)}\mathbf{X}_i\right)^{-1}\sum_{i=1}^m \mathbf{X}_i^\top\widehat{\mathbf{E}}_i^{-1(k)}\left(\widehat{u_i\mathbf{y}_i}^{(k)} - \widehat{u_i}^{(k)}\mathbf{N}_i\widehat{\mathbf{f}}^{(k)} - \mathbf{Z}_i\widehat{u_i\mathbf{b}_i}^{(k)}\right) \quad (11)$$

$$\begin{aligned}
\widehat{\mathbf{f}}^{(k+1)} &= \left(\sum_{i=1}^m \widehat{u_i}^{(k)}\mathbf{N}_i^\top\widehat{\mathbf{E}}_i^{-1(k)}\mathbf{N}_i + \widehat{\sigma^2}^{(k)}\lambda\mathbf{K}\right)^{-1}\sum_{i=1}^m \mathbf{N}_i^\top\widehat{\mathbf{E}}_i^{-1(k)}\left(\widehat{u_i\mathbf{y}_i}^{(k)} - \widehat{u_i}^{(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k+1)}\right.\\
&- \left.\mathbf{Z}_i\widehat{u_i\mathbf{b}}^{(k)}\right) \quad (12)
\end{aligned}$$

$$\widehat{\sigma^2}^{(k+1)} = \frac{1}{N}\sum_{i=1}^m\left[\widehat{a}_i^{(k)} - 2\widehat{\boldsymbol{\mu}}_i^{(k+1)\top}\mathbf{E}_i^{-1}\left(\widehat{u_i\mathbf{y}_i}^{(k)} - \mathbf{Z}_i\widehat{u_i\mathbf{b}_i}_i^{(k)}\right) + \widehat{u_i}^{(k)}\widehat{\boldsymbol{\mu}}_i^{(k+1)\top}\mathbf{E}_i^{-1}\widehat{\boldsymbol{\mu}}_i^{(k+1)}\right] \quad (13)$$

$$\widehat{\mathbf{D}}^{(k+1)} = \frac{1}{m}\sum_{i=1}^m \widehat{u_i\mathbf{b}_i\mathbf{b}_i^\top}^{(k)} \quad (14)$$

$$\begin{aligned}
\widehat{\phi}^{(k+1)} &= \underset{\phi\in(0,1)\times\mathcal{R}^+}{\arg\max}\left(-\frac{1}{2}\log(|\mathbf{E}_i|) - \frac{1}{2\widehat{\sigma^2}^{(k+1)}}\left[\widehat{a}_i^{(k)} - 2\widehat{\boldsymbol{\mu}}_i^{(k+1)\top}\mathbf{E}_i^{-1}\left(\widehat{u_i\mathbf{y}_i}^{(k)} - \mathbf{Z}_i\widehat{u_i\mathbf{b}_i}_i^{(k)}\right)\right.\right.\\
&+ \left.\left.\widehat{u_i}^{(k)}\widehat{\boldsymbol{\mu}}_i^{(k+1)\top}\mathbf{E}_i^{-1}\widehat{\boldsymbol{\mu}}_i^{(k+1)}\right]\right) \quad (15)
\end{aligned}$$

$$\begin{aligned}
\widehat{\nu}^{(k+1)} &= \underset{\nu}{\arg\max}\left\{\sum_{i=1}^m \log \mathrm{T}_{n_i^c}\left(\mathbf{V}_{1i}^c,\mathbf{V}_{2i}^c;\boldsymbol{\mu}_i^{co(k+1)},\mathbf{S}_i^{co(k+1)},\nu+n_i^o\right)\right.\\
&+ \left.\sum_{i=1}^m \log t_{n_i^o}\left(\mathbf{V}_i^o;\boldsymbol{\mu}_i^{o(k+1)},\boldsymbol{\Sigma}_i^{oo(k+1)},\boldsymbol{\nu}\right)\right\}, \quad (16)
\end{aligned}$$

where $N = \sum_{i=1}^m n_i$. The algorithm is iterated until a suitable convergence rule is satisfied, in this case, we adopt the distance involving two successive evaluations of the actual penalized log-likelihood. So, this process is iterated until some distance between two successive evaluations of the actual penalized log-likelihood $\ell_p(\boldsymbol{\theta},\lambda)$ in Section 3.2, such as $|\ell_p(\widehat{\boldsymbol{\theta}}^{(k+1)}) - \ell_p(\widehat{\boldsymbol{\theta}}^{(k)})|$ or $|\ell_p(\widehat{\boldsymbol{\theta}}^{(k+1)})/\ell_p(\widehat{\boldsymbol{\theta}}^{(k)}) - 1|$, becomes small enough, for example, $\epsilon = 10^-6$.

It is important to stress that from equations (11) to (15), the E-step reduces to the computation of

$$\widehat{u_i\mathbf{y}_i\mathbf{y}_i^\top} = \mathbb{E}\left[u_i\mathbf{y}_i\mathbf{y}_i^\top\middle|\mathbf{V}_i,\mathbf{C}_i,\boldsymbol{\theta}\right], \quad \widehat{u_i\mathbf{y}_i} = \mathbb{E}\left[u_i\mathbf{y}_i\middle|\mathbf{V}_i,\mathbf{C}_i,\boldsymbol{\theta}\right], \quad \text{and} \quad \widehat{u_i} = \mathbb{E}\left[u_i\middle|\mathbf{V}_i,\mathbf{C}_i,\boldsymbol{\theta}\right],$$

that is, the first and second moments of a truncated multivariate-t distribution. These expected values can be determined in closed form, using Propositions 2-3, as follows:

1. If the ith 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 ith subject has only censored components then from Proposition 2, we have:

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

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

3. If the ith subject has censored and uncensored components and given that $(\mathbf{Y}_i\big|\mathbf{V}_i, \mathbf{C}_i)$, $(\mathbf{Y}_i\big|\mathbf{V}_i, \mathbf{C}_i, \mathbf{Y}_i^o)$, and $(\mathbf{Y}_i^c\big|\mathbf{V}_i, \mathbf{C}_i, \mathbf{Y}_i^o)$ are equivalent process, then from Proposition 3, we have

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

   where $\mathbf{W}_i \sim \mathrm{Tt}_{n_i^c}(\boldsymbol{\mu}_i^{co}, \widetilde{\mathbf{S}}_i^{co}, \nu + n_i^o + 2, (\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c))$, $\widetilde{\mathbf{S}}_i^{co} = \left(\dfrac{\nu + \delta^2(\mathbf{y}_i^o)}{\nu + n_i^o + 2}\right)\mathbf{S}_i$ and $\mathbf{S}_i$, $\mathbf{S}_i^{co}$ and $\boldsymbol{\mu}_i^{co}$ are as in Section 3.2.

Formulas for $\mathbb{E}[\mathbf{W}]$ and $\mathbb{E}[\mathbf{W}\mathbf{W}^\top]$, where $\mathbf{W} \sim \mathrm{Tt}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$, have been recently developed using recurrence relations involving the density of multivariate t-distribution. These can be obtained in the R package `MomTrunc` (Galarza *et al.*, 2020).

3.4. Estimation of the random effects

In this section, we are interested in the estimation of random effects, which is useful for evaluating subject-specific quantities of interest such as individually changed intercepts and slopes. To estimate the random effects, we consider the conditional mean of $\mathbf{b}_i$ given the observed data $\mathbf{V}_i$, and $\mathbf{C}_i$, that is, $\mathbb{E}[\mathbf{b}_i|\mathbf{V}_i, \mathbf{C}_i]$, empirical Bayes approach. Thus, when the parameter values of $\boldsymbol{\theta}$ are known, the conditional mean of $\mathbf{b}_i$ given $\mathbf{C}_i$, $\mathbf{V}_i$ is

$$\widehat{\mathbf{b}}_i(\boldsymbol{\theta}) = \mathbb{E}\left[\mathbf{b}_i|\mathbf{V}_i, \mathbf{C}_i\right] = \boldsymbol{\varphi}_i(\widehat{\mathbf{y}}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{N}_i\mathbf{f}), \tag{17}$$

where $\boldsymbol{\varphi}_i$ is defined in Subsection 3.3 and $\widehat{\mathbf{y}}_i = \mathbb{E}[\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i]$ is the first moment of the truncated t-distribution.

The empirical Bayes estimates of random effects are obtained by substituting the MPL estimates $\widehat{\boldsymbol{\theta}}$ into $\mathbf{b}_i(\boldsymbol{\theta})$, leading to $\widehat{\mathbf{b}}_i = \mathbf{b}_i(\widehat{\boldsymbol{\theta}})$. In addition, the fitted values of responses can be estimated directly by $\widehat{\mathbf{y}}_i = \mathbf{X}_i\widehat{\boldsymbol{\beta}} + \mathbf{N}_i\widehat{\mathbf{f}} + \mathbf{Z}_i\widehat{\mathbf{b}}_i$.

3.5. The expected information matrix

In the context of nonparametric regression, the covariance matrix of the MPL estimates can be evaluated by inverting the observed information matrix obtained by treating the penalized likelihood as a usual likelihood (Segal *et al.*, 1994). Louis (1982) proposed a technique for computing the observed matrix within the EM algorithm framework, this method adjust the variance of the estimated fixed effects for the information lost owing to censoring. Using this method, and from the results given by Lange *et al.* (1989), the information matrix for $(\boldsymbol{\beta}, \mathbf{f})$ can be approximated by

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

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

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

where the penalized expected information matrix $\mathbf{I}_p(\boldsymbol{\beta}, \mathbf{f})$ takes the form:

$$\mathbf{I}_p(\boldsymbol{\beta}, \mathbf{f}) = \begin{pmatrix} \mathbf{I}_p(\boldsymbol{\beta}, \boldsymbol{\beta}) & \mathbf{I}_p(\boldsymbol{\beta}, \mathbf{f}) \\ \mathbf{I}_p^\top(\boldsymbol{\beta}, \mathbf{f}) & \mathbf{I}_p(\mathbf{f}, \mathbf{f}) \end{pmatrix},$$

where

$$\begin{aligned}
\mathbf{I}_p(\boldsymbol{\beta}, \boldsymbol{\beta}) &= \sum_{i=1}^{n} \left\{ \left(\frac{\nu + n_i}{\nu + n_i + 2} \right) \mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i - \mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1} \left[\left(\frac{\nu + n_i + 2}{\nu + n_i} \right) \mathbf{E}_2 - \mathbf{E}_1 \right] \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right\}, \\
\mathbf{I}_p(\boldsymbol{\beta}, \mathbf{f}) &= \sum_{i=1}^{n} \left\{ \left(\frac{\nu + n_i}{\nu + n_i + 2} \right) \mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{N}_i - \mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1} \left[\left(\frac{\nu + n_i + 2}{\nu + n_i} \right) \mathbf{E}_2 - \mathbf{E}_1 \right] \boldsymbol{\Sigma}_i^{-1} \mathbf{N}_i \right\}, \\
\mathbf{I}_p(\mathbf{f}, \mathbf{f}) &= \sum_{i=1}^{n} \left\{ \left(\frac{\nu + n_i}{\nu + n_i + 2} \right) \mathbf{N}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{N}_i - \mathbf{N}_i^\top \boldsymbol{\Sigma}_i^{-1} \left[\left(\frac{\nu + n_i + 2}{\nu + n_i} \right) \mathbf{E}_2 - \mathbf{E}_1 \right] \boldsymbol{\Sigma}_i^{-1} \mathbf{N}_i \right\} \\
&+ \lambda^2 \mathbf{K} \mathbf{f} \mathbf{f}^\top \mathbf{K},
\end{aligned}$$

where $\mathbf{E}_1 = (\widehat{u_i\mathbf{y}_i} - \widehat{u}_i\boldsymbol{\mu}_i)(\widehat{u_i\mathbf{y}_i} - \widehat{u}_i\boldsymbol{\mu}_i)^\top$ and $\mathbf{E}_2 = (\widehat{u_i^2\mathbf{y}_i\mathbf{y}_i^\top} - \widehat{u_i^2\mathbf{y}_i}\boldsymbol{\mu}_i^\top - \boldsymbol{\mu}_i\widehat{u_i^2\mathbf{y}_i^\top} + \widehat{u_i^2}\boldsymbol{\mu}_i\boldsymbol{\mu}_i^\top)$. Note that $\mathbf{E}_1$ depend on the computation of $\widehat{u}_i$, $\widehat{u_i\mathbf{y}_i}$ that can be obtained in Subsection 3.3 and $\mathbf{E}_2$ depend on the following quantities

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

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

These expected values can be determined in closed form using Proposition 2 and 3, as follows

1. If the ith 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 ith subject has only censored components then

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

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

3. If the ith 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

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

where $d_p(n_i^o,\nu,2) = \dfrac{(\nu+n_i)(n_i^o+\nu+2)(n_i^o+\nu)}{n_i+\nu+2}$, $\mathbf{W}_i \sim Tt_{n_i^c}(\boldsymbol{\mu}_i^{co},\widetilde{\mathbf{S}}_i^{co},\nu+n_i^o+4;(\mathbf{V}_{1i}^c,\mathbf{V}_{2i}^c))$, $\widetilde{\mathbf{S}}_i^{co} = \left(\dfrac{\nu+\delta^2(\mathbf{y}_i^o)}{\nu+n_i^o+4}\right)\mathbf{S}_i$ and $\mathbf{S}_i$, $\mathbf{S}_i^{co}$ and $\boldsymbol{\mu}_i^{co}$ are as in Subsection 3.2.

It can be noted that here we also need the first and second moments of truncated-t distribution. And, as mentioned before, these moments can be obtained in the R package `MomTrunc` (Galarza *et al.*, 2020).

4. Estimation of the smoothing parameter

In the previous sections we considered the smoothing parameter λ fixed to make inference for the nonparametric function $\mathbf{f}$. However, in practice, this parameter need to be estimated from the data. Many authors have pointed out that the proper selection of smoothing parameters is essential for good a performance of the spline estimates (Green & Silverman, 1994). Wahba & Wold (1975) examine how much the smoothing should be because if λ is too small, the spline is too wiggly and picks up too much noise (overfit), and if λ is too large, the spline is too smooth and the signal is lost (underfit). A classical data-driven approach to selecting the smoothing parameter is cross validation (CV), which leaves out one subject's entire data at a time, but this approach is often computationally expensive (Zeger & Diggle, 1994).

Several authors have shown the connection between a smoothing spline and a linear mixed effects model for analysis of longitudinal data (see, Wang, 1998; Kohn *et al.*, 1991, for instance,). Zhang *et al.* (1998) treated the smoothing parameter as an additional variance component and estimated it with other variance components simultaneously using REML. According to Green (1987); Zhang *et al.* (1998), we can write $\mathbf{f}$ via a one-to-one linear transformation as:

$$
\mathbf{f} = \mathbf{T}\boldsymbol{\delta}+\mathbf{B}\mathbf{d}, \tag{18}
$$

where $\boldsymbol{\delta}$ and $\mathbf{d}$ are vectors with dimensions 2 and $r-2$, $\mathbf{B}=\mathbf{L}(\mathbf{L}^\top\mathbf{L})^{-1}$ and $\mathbf{L}$ is an $r\times(r-2)$ full-rank matrix satisfying $\mathbf{K}=\mathbf{L}\mathbf{L}^\top$ and $\mathbf{L}^\top\mathbf{T}=0$. Given (18), Equation (2) can be reformulated as:

$$\mathbf{y}_i=\mathbf{X}_i^*\boldsymbol{\beta}^*+\mathbf{Z}_i^*\mathbf{b}_i^*+\boldsymbol{\epsilon}_i,$$

where $\mathbf{X}_i^*=[\mathbf{X}_i,\mathbf{N}_i\mathbf{T}]$, $\mathbf{Z}_i=[\mathbf{N}_i\mathbf{B},\mathbf{Z}_i]$, $\boldsymbol{\beta}^*=(\boldsymbol{\beta}^\top,\boldsymbol{\delta}^\top)^\top$ are the regression coefficients and $\mathbf{b}^*=(\mathbf{d}^\top,\mathbf{b}_i^\top)^\top$ are mutually independent random effects with $\mathbf{d}\sim t_{r-2}(\mathbf{0},\frac{\sigma^2}{\lambda}\mathbf{I}_{r-2})$ and $\mathbf{b}_i$ and $\boldsymbol{\epsilon}_i$ have the same distributions as those given in Section 3.1.

Motivated by Zhang *et al.* (1998)'s results and using the connection between the smoothing spline and the linear mixed models, we propose to estimate λ using the EM algorithm, due to its simplicity of implementation and stable monotone convergence. This novel procedure is described as follows. Consider the following model:

$$\begin{aligned}\mathbf{y}_i|\mathbf{b}_i^*,u_i &\sim \mathrm{N}_{n_i}(\mathbf{X}_i^*\boldsymbol{\beta}^*+\mathbf{Z}_i^*\mathbf{b}_i^*,u_i^{-1}\boldsymbol{\Omega}_i)\\ \mathbf{b}_i^*|u_i &\sim \mathrm{N}_{r-2+q}(\mathbf{0},u_i^{-1}\boldsymbol{\Psi}),\\ u_i &\sim \mathrm{Gamma}(\nu/2,\nu/2),\end{aligned}$$

where

$$\boldsymbol{\Psi}=\begin{pmatrix}\frac{\sigma^2}{\lambda}\mathbf{I}_{r-2} & \mathbf{0}\\ \mathbf{0} & \mathbf{D}\end{pmatrix}.$$

Let $\mathbf{y}_i$ denote the observed data and $(\mathbf{b}_i^*,u_i)$ denote the missing data. Then, we consider the augmented data vector $\mathbf{y}_{ic}^*=(\mathbf{y}_i^\top,\mathbf{b}_i^{*\top},u_i^\top)$. In this case, the log-likelihood function for the augmented data model, dropping all the terms that are not functions of λ, takes the form:

$$\ell(\lambda;\mathbf{y}_c^*)\propto\sum_{i=1}^n\left\{-\frac{1}{2}\log|u_i^{-1}\boldsymbol{\Psi}_i|-\frac{1}{2}u_i\mathbf{b}_i^*\boldsymbol{\Psi}_i^{-1}\mathbf{b}_i^{*\top}\right\}.$$

The solution $\widehat{\lambda}$ can be obtained via the following joint iterative process:

Step 1: Obtain $\widehat{\boldsymbol{\theta}}^{(k+1)}$, as described in Subsection 3.3;

Step 2: (E-step) Given the observed data, evaluate the expectation of $\ell(\lambda;\mathbf{y}_c^*)$ and estimate in the kth iteration :

$$Q(\lambda|\widehat{\lambda}^{(k)}) = \mathbb{E}\left[\ell(\lambda;\mathbf{y}_c^*)|\mathbf{y},\widehat{\lambda}^{(k)}\right]=-\frac{1}{2}\sum_{i=1}^n\log|\boldsymbol{\Psi}_i|-\frac{1}{2}\sum_{i=1}^n\mathrm{tr}(\boldsymbol{\Psi}_i^{-1}\widehat{u_i\mathbf{b}_i^*\mathbf{b}_i^{*\top}}^{(k)}),$$

with $\widehat{u_i\mathbf{b}_i^*\mathbf{b}_i^{*\top}}^{(k)}=\mathbb{E}\left[u_i\mathbf{b}_i^*\mathbf{b}_i^{*\top}|\mathbf{y},\widehat{\lambda}^{(k)}\right]=\boldsymbol{\Lambda}_i^*+\dfrac{\nu+n_i}{\nu+Q(\mathbf{y}_i)}\boldsymbol{\Lambda}_i^*\mathbf{Z}_i^{*\top}\boldsymbol{\Omega}_i^{-1}(\mathbf{y}_i-\mathbf{X}_i^*\boldsymbol{\beta}^*)(\mathbf{y}_i-\mathbf{X}_i^*\boldsymbol{\beta}^*)^\top\boldsymbol{\Omega}_i^{-1}\mathbf{Z}_i^*\boldsymbol{\Lambda}_i^*$, $\boldsymbol{\Lambda}_i^*=(\boldsymbol{\Psi}^{-1}+\mathbf{Z}_i^{*\top}\boldsymbol{\Omega}_i^{-1}\mathbf{Z}_i^*)^{-1}$.

Step 3: (M-step) Uptade λ by

$$\widehat{\lambda}^{(k+1)}=-\frac{n(r-2)}{\sum_{i=1}^n\mathrm{tr}\left(\boldsymbol{\Psi}^{-1}\frac{\partial\boldsymbol{\Psi}}{\partial\lambda}\boldsymbol{\Psi}^{-1}\widehat{u_i\mathbf{b}_i^*\mathbf{b}_i^{*\top}}^{(k)}\right)}.$$

Thus, by repeating Step 1, Step 2 and Step 3, this iterative process leads to the MPL estimates of $\boldsymbol{\theta}$ and the smoothing parameter λ.

5. Goodness of fit and model selection

In this section, we consider diagnoses to assess the adequacy of the fit of the proposed model and detect influential observations.

Under condition that $\mathbf{y}_i \overset{\text{ind.}}{\sim} t_{n_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \nu)$, the Mahalanobis distance, $\delta_i^2(\boldsymbol{\theta}) = (\mathbf{y}_i - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i)$, has been considered by several authors to detect outliers in multivariate t models. To deal with the censored values existing in $\mathbf{y}_i$, we used the imputation procedure, that is, for censored values $\widehat{\mathbf{y}}_i = \mathbb{E}[\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i]$. According to Lange *et al.* (1989), under t model $F_i = \delta_i^2(\boldsymbol{\theta})/n_i$ is F-distributed with n_i and ν degrees of freedom, where n_i corresponds to the number of measurements associated with the ith subject. In addition, $\widehat{F}_i = \delta_i^2(\widehat{\boldsymbol{\theta}})/n_i$ has asymptotically the same distribution as F_i, $i = 1, \ldots, n$. Therefore, using the Wilson-Hilferty approximation (Johnson *et al.*, 1994; Galea-Rojas, 1995), we have that the transformed distance is

$$F_i^{[z]} = \frac{\left(1 - \dfrac{2}{9\nu}\right) F_i^{1/3} - \left(1 - \dfrac{2}{9n_i}\right)}{\left[\left(\dfrac{2}{9\nu}\right) F_i^{2/3} + \left(\dfrac{2}{9n_i}\right)\right]^{1/2}}, \quad i = 1, \ldots, m,$$

and follows approximately a standard normal distribution. Thus, a Q-Q plot of the transformed distances, $F_i^{[z]}$, can be used to assess the fit of the multivariate t distribution.

For a model selection criterion, we adopt the Akaike Information Criterion (AIC) (Akaike, 1974) and the Bayesian information criterion (BIC) (Schwarz *et al.*, 1978, so BIC is also known as SIC) which have been extended for standard LME and NLME models (Davidian & Giltinan, 1995). For t-SMEC model, we can define the AIC and BIC as follows:

$$\begin{aligned} AIC(\widehat{\boldsymbol{\theta}}) &= -2\ell_p(\widehat{\boldsymbol{\theta}}) + 2p^*, \\ BIC(\widehat{\boldsymbol{\theta}}) &= -2\ell_p(\widehat{\boldsymbol{\theta}}) + p^* \log N, \end{aligned}$$

where $\ell_p(\widehat{\boldsymbol{\theta}})$ corresponds to the logarithm of the penalized likelihood function, defined in Equation (7), p^* is the total number of parameters in the model, and N denotes the size of the sample.

6. Simulation studies

In order to examine the performance of our proposed models and algorithm, we present two simulation studies. The first one examines the finite sample properties of the estimators. The second study compares the performance of the estimates of the t-SMEC model and the N-SMEC model. For both simulation schemes, we simulate longitudinal data from the following model:

$$y_{ij} = \beta_1 x_{1ij} + \beta_2 x_{2_{ij}} + f(t_{ij}) + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij}, \quad i = 1, \ldots, n, \quad , j = 1, \ldots, n_i. \tag{19}$$

The parameters were set at $\boldsymbol{\beta}^\top = (\beta_1, \beta_2) = (2, -1.5)$, $\sigma^2 = 0.13$, and $\mathbf{D}$ with elements $\alpha_{11} = 0.25$, $\alpha_{12} = 0.01$, $\alpha_{22} = 0.1$. We chose a smooth function $f(t_{ij}) = \exp(\sin(0.3t_{ij})\cos(0.6t_{ij}))$, where $t_{ij} = (1, 2, 3, 4, 5, 6, 7)$. The values $\mathbf{x}_i^\top = (x_1, x_2)$ were generated independently from a uniform distribution in the intervals (0,1) and (-1,1), respectively, and those values were kept constant throughout the experiment.

All computational procedures were implemented using the R software (R Core Team, 2020), which is available from us upon request.

6.1. Asymptotic properties

In this simulation study, the main focus is to evaluate the finite-sample performance of the parameter estimates. Another goal is to examine the consistency of the standard errors for the MPL estimates of $\boldsymbol{\beta}$ and $\mathbf{f}$. To do, so we generated samples from the t-SMEC model, with $(b_{0i}, b_{1i}) \overset{\text{ind.}}{\sim} t_2(\mathbf{0}, \mathbf{D}, \nu)$ and $\boldsymbol{\epsilon}_i \sim t_{n_i}(\mathbf{0}, \boldsymbol{\Omega}_i, \nu)$, where $\boldsymbol{\Omega}_i = \sigma^2 \mathbf{E}_i$, with a correlation structure AR(1) for $\mathbf{E}_i$ considering $\phi_1 = 0.8$ and $\nu = 5$. Moreover, to study the effect on the level of censoring and sample sizes, we consider two censoring proportions (10% and 20%) and sample sizes fixed at $m = 50, 100$ and 300. For each combination of sample size and censoring level, we generated 200 simulated datasets.

To evaluate the computational accuracy and to examine the consistency of the estimates of the standard erros suggested in Subsection 3.4, we computed the following measures:

- The arithmetic average of estimates:

$$\text{MC Mean}(\widehat{\theta}_i) = \frac{1}{200} \sum_{j=1}^{200} \widehat{\theta}_i^{(j)}$$

- The average values of the estimates of the standard erros obtained through the method described in Subsection 3.4 using the expected information matrix (MC IM).

- The Monte Carlo standard deviation of $\boldsymbol{\beta}$ and $\mathbf{f}$ (MC SD).

Table 1 summarize the simulation results based on 200 Monte Carlo data sets for the model parameters $(\boldsymbol{\beta}, \mathbf{f})$. It can be observed that the MC Mean approaches the true value for fixed components and when the sample size increases the value of MC SD decreases. It can also be seen that the approximate standard errors (MC IM) obtained in Subsection 3.4 and the standard deviation estimates (MC SD) closely agree with each other, suggesting that the derived standard errors work well. From Figure 1 it can be observed that the variability among the estimates of the nonparametric function declines as the sample size increases, and the censorship does not influence the estimation of the nonparametric part. Therefore, we can conclude that the t-SMEC model provides estimates with good asymptotic properties for the fixed components and the nonparametric part is able to capture the true unknown function.

6.2. Robustness of the estimates

The purpose of this simulation study is to compare the fits of the t-SMEC and N-SMEC models when we assume the normal distribution for the errors and random effects. Also, we are interested in comparing the fits when the usual assumption of normality is violated. Then, in this case, we replace the multivariate normal distribution by the multivariate contaminated normal, which is a particular case of the SMN distributions.

First, for the normal distribution, we consider $(b_{0i}, b_{1i}) \overset{\text{ind.}}{\sim} \text{N}_2(\mathbf{0}, \mathbf{D})$ and $\boldsymbol{\epsilon}_i \sim \text{N}_{n_i}(\mathbf{0}, \boldsymbol{\Omega}_i)$, where $\boldsymbol{\Omega}_i = \sigma^2 \mathbf{E}_i$ with a correlation structure AR(1) for $\mathbf{E}_i$ and $\phi_1 = 0.4$. For the contamined normal, we consider $(b_{0i}, b_{1i}) \overset{\text{ind.}}{\sim} \text{N}_2(\mathbf{0}, u_i^{-1}\mathbf{D})$, $\boldsymbol{\epsilon}_i \sim \text{N}_{n_i}(\mathbf{0}, u_i^{-1}\boldsymbol{\Omega}_i)$ and

$$U_i = \begin{cases} 0.3 & \text{with probability} \quad 0.3, \\ 1 & \text{with probability} \quad 0.7, \end{cases}$$

Table 1: **Simulation study - Asymptotic properties**. Results based on 200 simulated samples. MC IM, MC SD are the respective average of the approximate standard errors obtained using the expected information matrix, and the average of the approximate standard deviations from fitting t-SMEC model.

Cens. level	Parameter	m=50			m=100			m=300		
		MC Mean	MC IM	MC SD	MC Mean	MC IM	MC SD	MC Mean	MC IM	MC SD
10%	β_1	1.9993	0.0362	0.0389	2.0040	0.0263	0.0289	1.9996	0.0148	0.0145
	β_2	-1.4989	0.0186	0.0199	-1.5020	0.0131	0.0141	-1.4992	0.0073	0.0076
	$f(1) = 1.2762$	1.2759	0.1091	0.1128	1.2777	0.0787	0.0747	1.2713	0.0453	0.0462
	$f(2) = 1.2270$	1.2297	0.1384	0.1397	1.2301	0.1005	0.0933	1.2198	0.0576	0.0566
	$f(3) = 0.8370$	0.8470	0.1759	0.1820	0.8369	0.1281	0.1197	0.8270	0.0733	0.0713
	$f(4) = 0.5029$	0.5120	0.2176	0.2244	0.5020	0.1585	0.1507	0.4915	0.0906	0.0868
	$f(5) = 0.3725$	0.3866	0.2613	0.2675	0.3659	0.1904	0.1818	0.3568	0.1086	0.1059
	$f(6) = 0.4176$	0.4335	0.3061	0.3145	0.4077	0.2230	0.2111	0.3989	0.1272	0.1264
	$f(7) = 0.6549$	0.6785	0.3519	0.3616	0.6456	0.2561	0.2433	0.6349	0.1461	0.1433
20%	β_1	1.9981	0.0292	0.0466	2.0037	0.0209	0.0316	1.9987	0.0117	0.0165
	β_2	-1.4971	0.0150	0.0222	-1.5011	0.0104	0.0166	-1.4986	0.0058	0.0090
	$f(1) = 1.2762$	1.2666	0.0934	0.1129	1.2672	0.0654	0.0783	1.2582	0.0376	0.0477
	$f(2) = 1.2270$	1.2163	0.1184	0.1385	1.2156	0.0833	0.0957	1.1995	0.0478	0.0577
	$f(3) = 0.8370$	0.8301	0.1501	0.1815	0.8174	0.1061	0.1227	0.7999	0.0607	0.0722
	$f(4) = 0.5029$	0.4909	0.1853	0.2254	0.4777	0.1313	0.1543	0.4583	0.0749	0.0874
	$f(5) = 0.3725$	0.3629	0.2221	0.2698	0.3372	0.1577	0.1865	0.3163	0.0898	0.1067
	$f(6) = 0.4176$	0.4055	0.2599	0.3174	0.3731	0.1846	0.2166	0.3522	0.1051	0.1273
	$f(7) = 0.6549$	0.6476	0.2985	0.3649	0.6056	0.2120	0.2483	0.5820	0.1207	0.1443

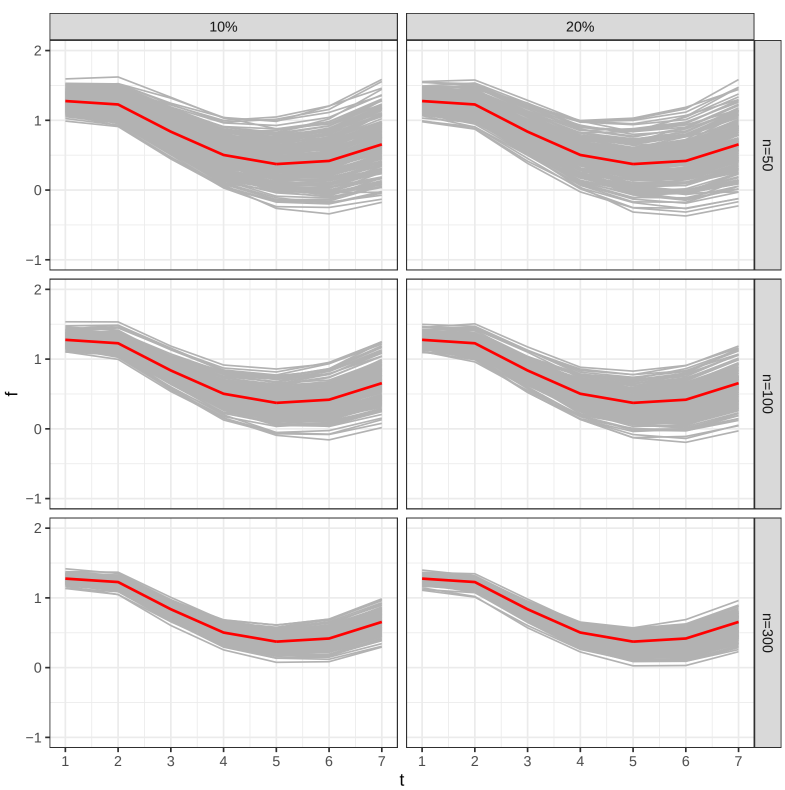


Figure 1: **Simulation study - Asymptotic properties**. Graphs of the nonparametric components with 200 replications. Adjusted curves (gray lines) and true curves (red lines) for all scenarios.

where $\mathbf{\Omega}_i = \sigma^2 \mathbf{E}_i$, with a correlation structure AR(1) for $\mathbf{E}_i$ and $\phi_1 = 0.4$. We generated $M = 200$ datasets of size $m = 150$ with censoring proportion 15%. Once the simulated data were generated, we fit the N-SMEC model and t-SMEC model to each simulated dataset.

The model selection criterion as well as the estimates of the model parameters were recorded for each simulation. The detailed numerical results under the scenarios considered, including the average BIC values and the MPL estimates are summarized in Table 2. From Table 2, can be noted that when the data generated follow the normal distribution the performances of the N-SMEC model and t-SMEC model are similar, indicating that the t-SMEC model gives reliable estimates. Also, to evaluate the use of the BIC criterion, the N-SMEC model was chosen by the criterion in 79.5% (159/200) of the samples generated as the best model. When the data generated follow the contamined normal, the t-SMEC model has better estimates and the standard errors are less than that of the N-SMEC model. Evaluating the BIC criterion, the N-SMEC model was chosen in 38.5% (77/200) of the samples.

Another important feature in our model is the ability to detect whether the distribution has heavily tails or not. It can be seen from Table 2 that when we fit the t-SMEC model to normal data, the estimate of ν on average is high, that is, the data does not have heavy-tails behavior. Now, when the data generated is Contaminated normal, the estimated ν on average is small since we are dealing with a distribution with heavier tails that the normal distribution. Therefore, we can be observed that the t-SMEC model fits better than the N-SMEC model counterpart when the data have tails heavier behavior.

7. Application

In this section, we apply our method to analyze a longitudinal dataset corresponding to the interruption of treatment of UTI (unstructured antiretroviral therapy) in HIV-infected adolescents at four institutions in the USA.

The UTI data is referred to a study of 72 perinatally HIV-infected children (Saitoh *et al.*, 2008). This dataset is available in the R package **lmec** (Vaida & Liu, 2012). Primarily due to treatment fatigue, unstructured treatment interruptions (UTI) are common in this population. Suboptimal adherence can lead to antiretroviral (ARV) resistance and diminished treatment options in the future. The aim of this study was to monitor the HIV-1 viral laod (RNA) after unstructured treatment interruption. The subjects in the study had taken ARV therapy for at least 6 months before UTI, and the medication was discontinued for more than 3 months. The HIV viral load were studied from the closest time points at $0, 1, 3, 6, 9, 12, 18, 24$ months after UTI. The number of observations from baseline (month 0) to month 24 are $71, 62, 58, 57, 43, 34, 24$, and 13, respectively. Out of 362 observations, 26(7%) observations were below the detection limits (50 or 400 copies/mL) and were left-censored at these values. The individual profiles are shown in Figure 2. This dataset was analyzed by Vaida & Liu (2009) and Matos *et al.* (2013b) using the N-LMEC and t-LMEC models, respectively.

Here, we revisit the UTI data assuming that the functional form of the HIV RNA levels over time is not known. We considered the following model:

$$y_{ij} = f(t_{ij}) + b_i + \epsilon_{ij}, \tag{20}$$

where y_{ij} is the $\log_{10}$HIV RNA for subject i at time t_{ij} ($i = 1, 2, \ldots, 72$; $j = 1, 2, \ldots, n_i$), $f(t_{ij})$ is an arbitrary smoothing function, b_i is the random intercept for the i-th subject, and ϵ_{ij} are

Table 2: **Simulation study - Robustness of the estimates**. Summary statistics based on 200 simulated AR(1) samples for the estimates parameters.

Distribution	Parameter	Fit					
		Normal			Student-t		
		MC Mean	MC IM	MC SD	MC Mean	MC IM	MC SD
Normal	β_1 (2)	2.0065	0.0475	0.0406	1.9968	0.0349	0.0413
	β_2 (-1.5)	-1.5071	0.0233	0.0197	-1.5011	0.0172	0.0197
	$f(1) = 1.2762$	1.2798	0.0640	0.0583	1.2869	0.0601	0.0611
	$f(2) = 1.2270$	1.2287	0.0780	0.0719	1.2387	0.0751	0.0767
	$f(3) = 0.8370$	0.8428	0.0980	0.0994	0.8526	0.0948	0.1048
	$f(4) = 0.5029$	0.5130	0.1208	0.1203	0.5236	0.1168	0.1274
	$f(5) = 0.3725$	0.3889	0.1435	0.1414	0.3981	0.1394	0.1486
	$f(6) = 0.4176$	0.4434	0.1673	0.1670	0.4487	0.1628	0.1733
	$f(7) = 0.6549$	0.6867	0.1918	0.1958	0.6881	0.1868	0.2014
	σ^2 (0.13)	0.1396			0.0993		
	α_{11} (0.25)	0.2606			0.2294		
	α_{12} (0.01)	0.0124			0.0108		
	α_{22} (0.1)	0.0967			0.0845		
	ϕ_1 (0.4)	0.3212			0.3929		
	ν	-			24.5473		
	λ	3.0217			1.3226		
	BIC	1490.536			1500.18		
Contamined Normal	β_1 (2)	1.9966	0.0552	0.0458	1.9950	0.0373	0.0474
	β_2 (-1.5)	-1.4998	0.0271	0.0257	-1.4992	0.0182	0.0244
	$f(1) = 1.2762$	1.2910	0.0806	0.0761	1.2874	0.0652	0.0696
	$f(2) = 1.2270$	1.2463	0.0991	0.0931	1.2263	0.0814	0.0899
	$f(3) = 0.8370$	0.8706	0.1247	0.1168	0.8582	0.1037	0.1121
	$f(4) = 0.5029$	0.5494	0.1539	0.1548	0.5279	0.1279	0.1457
	$f(5) = 0.3725$	0.4314	0.1835	0.1863	0.3984	0.1529	0.1735
	$f(6) = 0.4176$	0.4930	0.2142	0.2225	0.4552	0.1785	0.2044
	$f(7) = 0.6549$	0.7400	0.2458	0.2523	0.6884	0.2049	0.2288
	σ^2 (0.13)	0.2190			0.0932		
	α_{11} (0.25)	0.4311			0.2135		
	α_{12} (0.01)	0.0215			0.0120		
	α_{22} (0.1)	0.1603			0.0810		
	ϕ_1 (0.4)	0.3593			0.3925		
	ν	-			3.8593		
	λ	4.7094			24.0673		
	BIC	1925.797			1920.187		

random errors. The model (20) can be express in matrix form as:

$$\mathbf{y}_i = \mathbf{N}_i\mathbf{f} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \tag{21}$$

where $\mathbf{y}_i$ is a $(n_i \times 1)$ vector of responses for the i-th children, $\mathbf{N}_i$ is the incidence matrix, $\mathbf{f}$ is a (8×1) vector whose components are function $f(\cdot)$ evaluated at the times in the set $\mathbf{t}^0 = (t_1^0 = 0, t_2^0 = 1, t_3^0 = 3, \ldots, t_8^0 = 24)$, $\mathbf{Z}_i = \mathbf{1}_{n_i}$, with $\mathbf{1}_{n_i}$ a $(n_i \times 1)$ vector of ones and $\mathbf{t}_i = [t_{i1}, \ldots, t_{in_i}]^\top$, $\mathbf{b}_i = b_i$ the random intercept and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})^\top$ represents the within-subject random error.

We acknowledge four cases of correlation structure to specify the matrix $\mathbf{E}_i$: the continuous-time AR(1) structure, the compound symmetry (CS), the damped exponential (DEC) and the uncor-

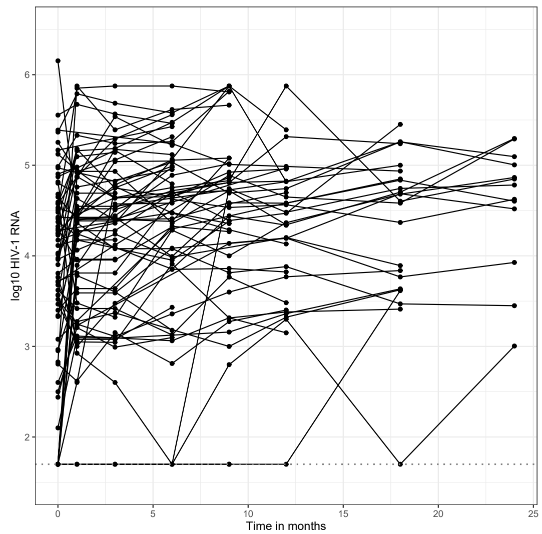


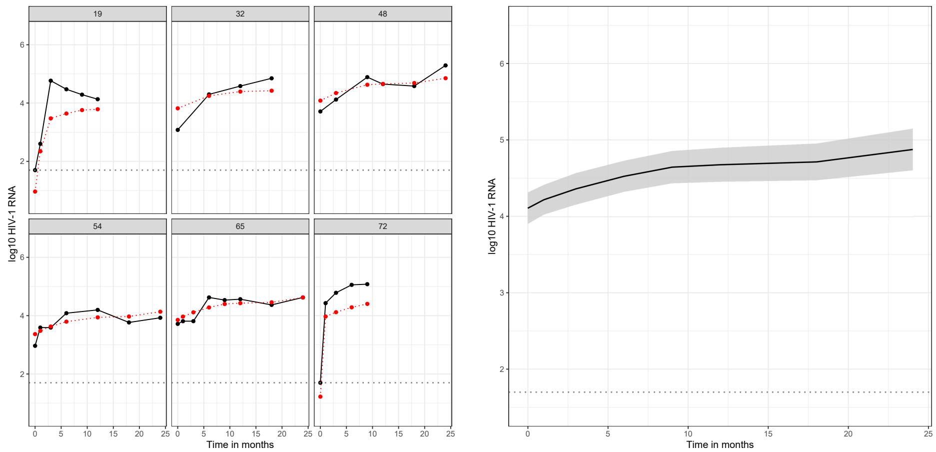
Figure 2: **UTI data**. Individual profiles (in log10 scale) for HIV viral load at different follow-up times.

related (UNC). Table 3 represents the MPL estimates of $\boldsymbol{\theta} = (\mathbf{f}^\top, \sigma^2, \alpha, \boldsymbol{\phi}^\top, \nu)^\top$, the smoothing parameter estimate (λ), the corresponding penalized log-likelihood function evaluated at $\widehat{\boldsymbol{\theta}}$ in the fitted models, and the values of AIC and BIC. These results reveal that the model with an UNC structure has lower AIC and BIC compared to the other structures, that is, the measures over the time of the same subject are not correlated. From the fit of (20), estimates of individual profiles are shown for six subjects in Figure 3a, it can be seen that the model seems to provide a good fit.

Table 3: **UTI dataset**. Parameter estimates of the t-SMEC model (20) for the UTI dataset. SE indicates the standard errors.

	AR(1)		CS		DEC		UNC	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
f_1	4.1106	0.0948	4.0863	0.1169	4.1276	0.0998	4.0929	0.1082
f_2	4.2219	0.0921	4.1853	0.1137	4.2146	0.0975	4.2116	0.1039
f_3	4.3647	0.0991	4.3462	0.1243	4.3563	0.1046	4.3672	0.1126
f_4	4.5284	0.0975	4.5213	0.1218	4.5226	0.1029	4.5323	0.1103
f_5	4.6484	0.1013	4.6294	0.1277	4.6324	0.1066	4.6478	0.1153
f_6	4.6795	0.1048	4.6704	0.1331	4.6729	0.1108	4.6786	0.1195
f_7	4.7080	0.1129	4.7161	0.1446	4.7132	0.1200	4.7090	0.1300
f_8	4.8736	0.1309	4.8499	0.1691	4.8542	0.1389	4.8697	0.1516
σ^2	0.0777		0.4668		0.3407		0.1036	
α	0.3171		0.0615		0.0559		0.3741	
ϕ_1	0.0007		0.7369		0.7777			
ϕ_2	1		0		0.0317			
ν	3.0344		3.0998		3.036		3.0978	
λ	699.1542		2095.618		2137.032		695.6855	
loglikp	-341.2195		-339.7186		-340.5147		-339.4198	
AIC	706.439		703.4372		707.0294		**700.8397**	
BIC	753.1388		750.1369		757.6208		**743.6477**	

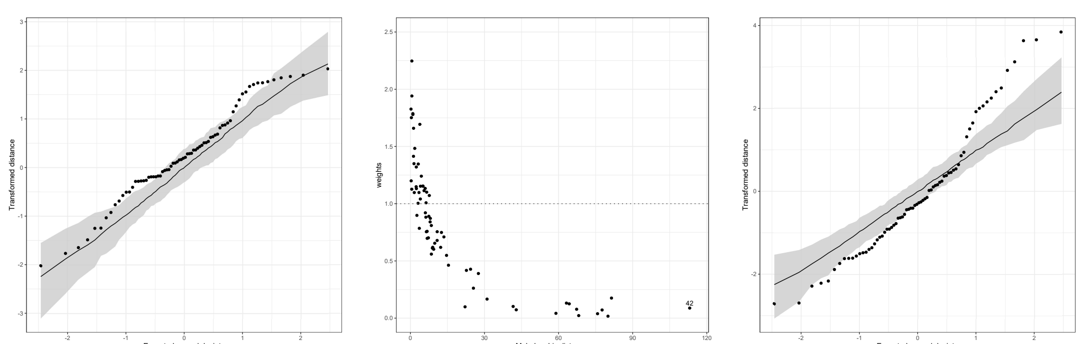
For the t-SMEC model under UNC correlation structure (our best model), we present in Figure 3b the curve of the estimated nonparametric function and the corresponding confidence bands. It can be noted that the estimated nonparametric function increase gradually. This is the evidence of the negative effect of the antiretroviral therapy interruption on the viral load levels. It means, the viral load increments consistently along the time when the antiretroviral therapy begins to be interrupted. For the fit model, the mean viral load ($\mathbb{E}(y_{ij}) = f(t_{ij})$) increases from 4.09 at the time of UTI to 4.87 at 24 months.



(a) (b)

Figure 3: **UTI dataset**. (a) Viral loads in $\log_{10}$ scale (solid line) for 6 randomly chosen subjects and estimated trajectories (red, dotted line) for the t-SMEC model in the UNC structure. (b) Fitted curve of nonparametric part. The shaded regions denote the 95% confidence intervals obtained by $\widehat{\mathbf{f}} \pm 1.96\sqrt{\widehat{\mathrm{Var}}(\widehat{\mathbf{f}})}$.

Figure 4 displays the transformed distance plots, for the Student-t (Figure 4a) and the normal (Figure 4c) models. The transformed distance under the Student-t model seems to be closer to normality than under the normal model. Therefore, it can be seen that the adjusted model t-SMEC with the UNC correlation structure seems to present an adequate adjustment. Identification of outlying observations under the t-SMEC model may be performed, for instance, by the scatter plot between the estimated weight and the estimated Mahalanobis distance, Figure 4b. As can be seen, the subject 42 receive a smaller weight and the higher Mahalanobis distance. Besides, in this Figure, it can be observed that many observations present smaller weights, verifying the robust aspects of the MPL estimation under the Student's t-distribution.



(a) (b) (c)

Figure 4: **UTI dataset**. (a) Normal probability plot for the transformed distance under the t-SMEC model with UNC structure. (b) Estimated weights ($\widehat{u}_i$) for the estimated t-SMEC model with UNC structure. (c) Normal probability plot for the transformed distance under the N-SMEC model with UNC structure.

8. Conclusion

In this paper, we proposed a semiparametric mixed model for the analysis of longitudinal censored data, assuming that the within-individual measurement errors and the random effects were distributed with t-multivariate distribution. This work can be considered as an extension of Matos *et al.* (2013b), where a linear/nonlinear mixed effects model was considered for censored data with t-distribution.

In practical implementation, the EM algorithm is used to obtain MPL estimates of the regression coefficients of the parametric part and to estimate the nonparametric component as a natural cubic spline. We proposed the EM algorithm to estimate the smoothing parameter using a modification of the mixed model proposed by Green (1987). The first simulation study validates the performance of our method and the second study indicates that there is an efficiency gain of the t-SMEC model in relation to the N-SMEC model for data with tails heavier than normal. A real data set previously analyzed under N-LMEC and t-LMEC models is reanalyzed under the semiparametric mixed model, showing the flexibility of the t-SMEC model to fit the data set in which we do not know the functional form that relates the variables. The codes in R (R Core Team, 2020) used in the application can be obtained from the authors upon request.

In this work, we have discussed the estimation of a single nonparametric function. But the methods can be generalized to additive mixed models in the presence of multiple nonparametric additive covariate effects and non-Gaussian outcomes (Ibacache-Pulgar *et al.*, 2013). Although the t-SMEC model considered here has shown great flexibility for modeling symmetric data with indications of lighter or 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-SMEC model by considering a more flexible family of distributions, such as the scale mixtures of skew-normal (SMSN) distribution class, to accommodate the censoring, skewness, and heaviness in tails simultaneously.

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Cont.*, **19**, 716–723.

Arellano-Valle, R. B. & Bolfarine, H. (1995). On some characterizations of the t-distribution. *Statistics & Probability Letters*, **25**(1), 79–85.

Davidian, M. & Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data*. Routledge.

Dempster, A., Laird, N. & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B,*, **39**, 1–38.

Diggle, P. (2002). *Analysis of longitudinal data*. Oxford University Press, USA. ISBN 0198524846.

Galarza, C. E., Kan, R. & Lachos, V. H. (2020). *MomTrunc: Moments of Folded and Doubly Truncated Multivariate Distributions*. R package version 5.69.

Galea-Rojas, M. (1995). *Calibraçao comparativa estrutural e funcional*. Ph.D. thesis, Tese de Doutorado em Estatıstica, Universidade de Sao Paulo, Sao Paulo.

Green, P. J. (1987). Penalized likelihood for general semi-parametric regression models. *International Statistical Review/Revue Internationale de Statistique*, **55**, 245–259.

Green, P. J. (1990). On use of the em for penalized likelihood estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, **52**(3), 443–452.

Green, P. J. & Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: a roughness penalty approach*. CRC Press.

Helsel, D. R. (2011). *Statistics for censored environmental data using Minitab and R*, volume 77. John Wiley & Sons.

Hughes, J. (1999). Mixed effects models with censored data with application to HIV RNA levels. *Biometrics*, **55**, 625–629.

Ibacache-Pulgar, G., Paula, G. A. & Galea, M. (2012). Influence diagnostics for elliptical semi-parametric mixed models. *Statistical Modelling*, **12**(2), 165–193.

Ibacache-Pulgar, G., Paula, G. A. & Cysneiros, F. J. A. (2013). Semiparametric additive models under symmetric distributions. *TEST*, **22**(1), 103–121.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994). *Continuous univariate distributions*. Wiley New York.

Kohn, R., Ansley, C. F. & Tharm, D. (1991). The performance of cross-validation and maximum likelihood estimators of spline smoothing parameters. *Journal of the American Statistical Association*, **86**(416), 1042–1050.

Lachos, V., Bandyopadhyay, D. & Dey, D. (2011). Linear and nonlinear mixed-effects models for censored HIV viral loads using normal/independent distributions. *Biometrics*, **55**, 1304–1318.

Lange, K. L., Little, R. & Taylor, J. (1989). Robust statistical modeling using t distribution. *Journal of the American Statistical Association*, **84**, 881–896.

Liu, C. & Rubin, D. B. (1994). The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika*, **80**, 267–278.

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.

Matos, L. A., Lachos, V. H., Balakrishnan, N. & Labra, F. V. (2013a). Influence diagnostics in linear and nonlinear mixed-effects models with censored data. *Computational Statistics & Data Analysis*, **57**(1), 450 – 464.

Matos, L. A., Prates, M. O., Chen, M.-H. & Lachos, V. H. (2013b). Likelihood-based inference for mixed-effects models with censored response using the multivariate-t distribution. *Statistica Sinica*, **23**, 1323–1345.

Meza, C., Osorio, F. & De la Cruz, R. (2012). Estimation in nonlinear mixed-effects models using heavy-tailed distributions. *Statistics and Computing*, **22**(1), 121–139.

Munoz, A., Carey, V., Schouten, J. P., Segal, M. & Rosner, B. (1992). A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics*, **48**(3), 733–742.

Palarea-Albaladejo, J. & Martin-Fernandez, J. (2013). Values below detection limit in compositional chemical data. *Analytica chimica acta*, **764**, 32–43.

Pinheiro, J. & Bates, D. (2006). *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media.

Pinheiro, J. C., Liu, C. & Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics*, **10**(2), 249–276.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Saitoh, A., Foca, M., Viani, R., Heffernan-Vacca, S., Vaida, F., Lujan-Zilbermann, J., Emmanuel, P., Deville, J. & Spector, S. (2008). Clinical outcomes after an unstructured treatment interruption in children and adolescents with perinatally acquired HIV infection. *Pediatrics*, **121**, 513–521.

Schwarz, G. *et al.* (1978). Estimating the dimension of a model. *The annals of statistics*, **6**(2), 461–464.

Segal, M. R., Bacchetti, P. & Jewell, N. P. (1994). Variances for maximum penalized likelihood estimates obtained via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, **56**(2), 345–352.

Vaida, F. & Liu, L. (2009). Fast implementation for normal mixed effects models with censored response. *Journal of Computational and Graphical Statistics*, **18**, 797–817.

Vaida, F. & Liu, L. (2012). *lmec: Linear Mixed-Effects Models with Censored Responses*. R package version 1.0.

Wahba, G. & Wold, S. (1975). A completely automatic french curve: fitting spline functions by cross validation. *Communications in Statistics*, **4**(1), 1–17.

Wang, Y. (1998). Smoothing spline models with correlated random errors. *Journal of the American Statistical Association*, **93**(441), 341–348.

Zeger, S. L. & Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics*, **50**, 689–699.

Zhang, D., Lin, X., Raz, J. & Sowers, M. (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association*, **93**(442), 710–719.