

Heavy-tailed longitudinal regression models for censored data: A likelihood based perspective

Larissa A. Matos^a, Victor H. Lachos^a, Tsung-I Lin^b and Luis M. Castro^{c1}

^a Department of Statistics, Campinas State University, Campinas, São Paulo, Brazil

^b Institute of Statistics, National Chung Hsing University, Taichung, Taiwan

^c Department of Statistics and CI²MA, Universidad de Concepción, Concepción, Chile

Abstract

HIV RNA viral load measures are often subjected to some upper and lower detection limits depending on the quantification assays. Hence, the responses are either left or right censored. Moreover, it is quite common to observe viral load measurements collected irregularly over time. A complication arises when these continuous repeated measures have a heavy-tailed behaviour. For such data structures, we propose a robust censored linear model based on the scale mixtures of normal distributions (SMN family). To take into account the autocorrelation existing among irregularly observed measures, a damped exponential correlation structure is considered. A stochastic approximation of the EM algorithm (SAEM algorithm) is developed to obtain the maximum likelihood estimates of the model parameters. The main advantage of this new procedure allows us to estimate the parameters of interest and evaluate the log-likelihood function in an easy and fast way. Furthermore, the standard errors of the fixed effects and predictions of unobservable values of the response can be obtained as a by-product. The practical utility of the proposed methodology is exemplified using both simulated and real data.

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

1 Introduction

The study of models in which the variable of interest is subjected to certain threshold values below or above which the measurements are not quantifiable has been the scope of the biomedical and biostatistical literature in recent years. Particularly, this situation occurs commonly in the study of the human immunodeficiency virus (HIV) behaviour, where the quantification of HIV-1 RNA viral load is done using assays with different detection limits for monitoring the copy number of virus per millilitre of plasma. Lower detection limits ranging from 400 to 500 RNA copies/mL are considered for standard assays such as Amplicor HIV-1 monitor test 1.5 and Nuclisens HIV-1 QT assay (Antunes et al. 2003), while the range is 50 to 100 RNA copies/mL for ultra-sensitive assays such as the TaqMan assay, version 1 and 2 (Swenson et al. 2014).

In practice, longitudinal data coming from follow-up studies (*e.g.* acquired immune deficiency syndrome - AIDS - studies) can be modelled using censored linear and nonlinear mixed effects models (see for example Wu 2010, and references therein) and also regression models with a specific correlation structures on the error term (Garay et al. 2014). Although it is quite common to consider a Gaussian assumption for the random components of the model due mainly to the computational flexibility for parameter estimation (see Vaida and Liu 2009). From a practical point of view, such an assumption may not be realistic. In this context, some recent works in censored models (Garay et al. 2014, 2015) have indicated that likelihood-based inference can be seriously affected by the presence of atypical observations and/or the misspecification of the parametric distributions for both random effects and errors.

¹*Address for correspondence:* Luis M. Castro, Departamento de Estadística and CI²MA, Universidad de Concepción, Av. Esteban Iturra s/n, Barrio Universitario, Concepción, Chile. E-mail: luiscastroc@udec.cl.

The research herein was performed in part while Victor H. Lachos and Luis M. Castro were visiting the Department of Science at Pontificia Universidad Católica del Perú.

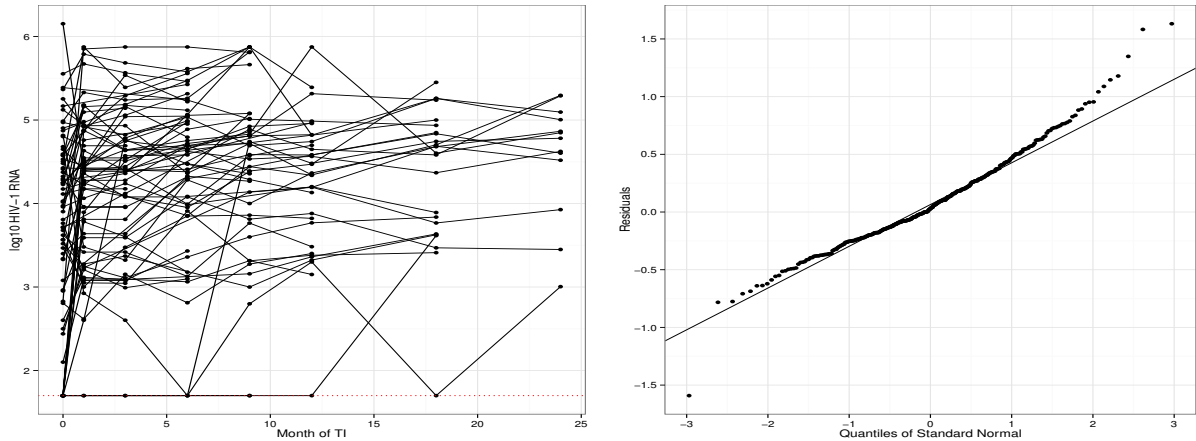


Figure 1: UTI data. Individual profiles (in \log_{10} scale) for HIV viral load at different follow-up times (Left panel) . Normal Q–Q plot for model residuals obtained by fitting a censored (Gaussian) mixed effect model (Right panel).

Consequently, in situations where the inferential results are sensitive to the assumed distributions for the random components of the model, it may be desirable to consider more flexible distributional assumptions, specifically, a heavy-tailed class of distributions.

For example, Pinheiro et al. (2001) proposed the a multivariate Student’s- t linear mixed model (t -LME). Lin and Lee (2006) and Lin and Lee (2007) developed some additional tools for the t -LME from likelihood-based and Bayesian perspectives, respectively. Is important to stress that, from the Bayesian point of view, Rosa et al. (2003) proposed the linear mixed model considering the normal/independent (NI) class of distributions (NI-LME). In the case of univariate censored responses, Arellano-Valle et al. (2012) and Massuia et al. (2015) proposed an extension of the normal censored regression (N-CR or Tobit) model to the case where the error terms follow a univariate Student’s- t distribution. Lachos et al. (2011) considered the use of the NI class in mixed effects models for longitudinal data with censored responses and adopted a Bayesian treatment to carry out posterior inference, extending, in some sense, the proposals of Samson et al. (2006), Vaida et al. (2007) and Vaida and Liu (2009).

From a likelihood-based perspective, a few alternatives have been proposed for longitudinal models under censored responses and considering heavy-tailed distributions. Recently, Garay et al. (2014) and Matos et al. (2013) utilized the Student’s- t distribution in the context of censored regression (t -CR) and mixed effects (t -CME) models respectively. They considered exact EM algorithms for maximum likelihood (ML) estimation, based on the mean and variance of a truncated multivariate Student’s- t distribution.

However, and from the frequentist point of view, the use of others heavy-tailed distributions has not been explored in the context of censored longitudinal models. For that reason, the aim of this paper is to consider the multivariate scale mixture of normal (SMN) distributions as the distribution of the random error in the context of the linear regression (SMN-CR) model for censored and longitudinal data. Our contribution extends the recent works of Garay et al. (2014) and Garay et al. (2015) since they only used the Student’s- t distribution which is a member of the SMN class. It is important to stress that, for the estimation of the model parameters, we consider an stochastic approximation of the EM algorithm, the so-called SAEM algorithm. This algorithm introduced by Delyon et al. (1999), seems to be more efficient than the EM (Dempster et al. 1977) and Monte Carlo EM (MCEM) (Wei and Tanner 1990) algorithms because it does not need the computation of the two first moments of the truncated multivariate SMN distributions, which requires high-dimensional numerical integration instead of a very intensive computation step of Monte Carlo simulation to evaluate those complex integrals. Moreover, Jank (2006) showed that the computational burden of SAEM is much smaller and reach the convergence

in just a fraction of the simulation size when compared to MCEM. This is because the memory effect persists in the SAEM method, in which the previous simulations are considered in the computation of the posterior ones. Note that, in the case of mixed effects models, Kuhn and Lavielle (2005), Meza et al. (2012) and Lavielle and Mbogning (2014) showed a good efficiency of the SAEM algorithm for ML estimation when the hypothesis of normality of the random components of the model is not considered.

In order to evaluate the performance of our proposal, we consider a dataset which contains 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 be diverse, *e.g.*, medication fatigue, patients inability to take antiretroviral medications, toxicity associated with the use of antiretroviral medications, adverse effects among many others (for more details see Saitoh et al. 2008). The dataset presents about 7% of observations below the detection limits of assay quantifications (left censored) and the viral loads were monitored from the closest time points at 0, 1, 3, 6, 9, 12, 18, and 24 months after the interruption. The individual profiles of viral load at different follow-up times after UTI are displayed in Figure 1 (left panel). This figure also presents the normal quantile-quantile (QQ) plot for the residuals (right panel) obtained by fitting a normal censored mixed effect (N-CME) model. Note that, this plot exhibits a heavy-tailed behavior, suggesting that the normality assumption for the within-subject errors might be inappropriate and maybe that some atypical observations or outliers might exist in the data. Since the outcome variables were recorded at irregular occasions, we consider a parsimonious damping exponential correlation (DEC) structure to address the within-subject autocorrelation. This type of correlation structure, proposed by Muñoz et al. (1992), takes into account the autocorrelation generated by the dependence among irregular occasions, see Figure 1 (left panel).

The paper is organized as follows. Section 2 provides some preliminaries of the SMN and truncated-SMN distributions and a brief review of the SAEM algorithms. Section 3 proposes the SMN-CR model and shows how to compute the ML estimates through the SAEM algorithm. In Section 4, we formulate analytically the empirical information matrix of model parameters. The issue concerning the prediction of future observations is also discussed. In Section 5, our proposed techniques are compared with the normality-based approach using simulated data and illustrated with the analysis of the UTI case study. Section 6 concludes with a short discussion of issues raised by our methods and some possible directions for a future research.

2 Preliminaries

2.1 Scale mixture of normal distributions (SMN)

An element of the symmetric class of scale mixture of multivariate normal distributions (Andrews and Mallows 1974; Lange and Sinsheimer 1993) is defined as the distribution of the p -variate random vector

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

where $\boldsymbol{\mu}$ is a location vector, \mathbf{Z} is a normal random vector with mean vector $\mathbf{0}$, variance-covariance matrix $\boldsymbol{\Sigma}$, U is a positive random variable with cumulative distribution function (*cdf*) $H(u | \boldsymbol{\nu})$ and probability density function (*pdf*) $h(u | \boldsymbol{\nu})$, independent of \mathbf{Z} , where $\boldsymbol{\nu}$ is a scalar or parameter vector indexing the distribution of U and $\kappa(U)$ is the weight function. Note that given $U = u$, \mathbf{y} follows a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\kappa(u)\boldsymbol{\Sigma}$. Hence, the *pdf* of \mathbf{y} is given by

$$\text{SMN}_p(\mathbf{y} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) = \int_0^\infty \phi_p(\mathbf{y}; \boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma})dH(u | \boldsymbol{\nu}),$$

where $\phi_p(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for the *pdf* of the p -variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariate matrix $\boldsymbol{\Sigma}$. By convention, we shall write $\mathbf{y} \sim \text{SMN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu})$. Three members of the scale mixture of normal class of distributions are commonly used for robust estimation:

- The multivariate Student's-t distribution, $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, where ν is called the degrees of freedom, can be derived from the mixture model (1), where U is distributed as $Gamma(\nu/2, \nu/2)$ and $\kappa(u) = 1/u$, with $\nu > 0$. The pdf of \mathbf{y} takes the form of

$$T_p(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \frac{\Gamma(\frac{p+\nu}{2})}{\Gamma(\frac{\nu}{2})\pi^{p/2}} \nu^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \left(1 + \frac{d}{\nu}\right)^{-(p+\nu)/2}, \quad \mathbf{y} \in \mathbb{R}^p,$$

where $\Gamma(\cdot)$ is the standard gamma function and $d = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})$ is the Mahalanobis distance.

- The multivariate slash distribution, $SL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, arises when $\kappa(u) = 1/u$ and the distribution of U is $Beta(\nu, 1)$, with $u \in (0, 1)$ and $\nu > 0$. Its pdf is given by

$$SL_p(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \nu \int_0^1 u^{\nu-1} \phi_p(\mathbf{y}; \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}) du, \quad \mathbf{y} \in \mathbb{R}^p.$$

- The multivariate contaminated normal distribution, $CN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu, \gamma)$, where $\nu, \gamma \in (0, 1)$. Here, $\kappa(u) = 1/u$ and U is a discrete random variable taking one of two states and with probability function given by

$$h(u \mid \boldsymbol{\nu}) = \nu \mathbb{I}_{\{\gamma\}}(u) + (1 - \nu) \mathbb{I}_{\{1\}}(u),$$

where $\boldsymbol{\nu} = (\nu, \gamma)$ and $\mathbb{I}_{\{\tau\}}(u)$ is the indicator function of the set τ whose value equals one if $u \in \tau$ and zero elsewhere. The associated density is

$$CN_p(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) = \nu \phi_p(\mathbf{y}; \boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Sigma}) + (1 - \nu) \phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

The parameter ν can be interpreted as the proportion of outliers while γ may be interpreted as a scale factor.

Now, let $TSMN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$ represent a p -variate truncated SMN distribution for $SMN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ lying within a right-truncated hyperplane $\mathbb{A} = \{\mathbf{x} = (x_1, \dots, x_p)^\top \mid x_1 \leq a_1, \dots, x_p \leq a_p\}$. We say that the p -dimensional vector $\mathbf{X} \sim TSMN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$, if its density is given by:

$$TSMN_p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A}) = \frac{SMN_p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}{\prod_{r=1}^p \int_{-\infty}^{a_r} SMN_p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) d\mathbf{x}} \mathbb{I}_{\{\mathbb{A}\}}(\mathbf{x}) \quad (2)$$

where the notation $\prod_{r=1}^p \int_{-\infty}^{a_r} = \int_{-\infty}^{a_1} \dots \int_{-\infty}^{a_p}$ stands for the abbreviation of multiple integrals.

2.2 The SAEM algorithm

The EM algorithm, introduced by (Dempster et al. 1977), is powerful frequentist approach to estimate parameters via ML when the data has missing/censored observations and/or latent variables. The main features of EM algorithm is the ease of implementation and the stability of monotone convergence. Let $\boldsymbol{\theta}$ be the parameter vector and $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{q}^\top)$ be the vector of complete data, *i.e.*, the observed data \mathbf{y} and the missing/censored data (or the latent variables, depending on the situation) \mathbf{q} . The EM algorithm consists basically of two steps: the expectation (E-step) and the maximization (M-step). These steps are performed iteratively in the complete likelihood function, $l_c(\boldsymbol{\theta} \mid \mathbf{y}_c)$, until it reaches the convergence. Each iteration is performed as follows:

E-Step: Calculate the conditional expectation $Q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}^{(k)}) = E \left[l_c(\boldsymbol{\theta} \mid \mathbf{y}_c) \mid \mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)} \right]$, where $\hat{\boldsymbol{\theta}}^{(k)}$ is the estimate of $\boldsymbol{\theta}$ at the k -th iteration.

M-Step: update $\boldsymbol{\theta}^{(k)}$ according to $\hat{\boldsymbol{\theta}}^{(k+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} Q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}^{(k)})$.

Although the EM algorithm is a powerful tool when the analytical expressions required by the E-steps have a closed form, it becomes a problem when the analytical expressions cannot be evaluated. To alleviate this difficulty, Wei and Tanner (1990) proposed the MCEM algorithm, where the E-step is replaced by a Monte Carlo approximation based on a large number of independent simulations of the latent variables. However, a large number of simulations are required, making the MCEM algorithm computationally expensive.

As an alternative, Delyon et al. (1999) presented a stochastic approximation of the EM algorithm, called SAEM. In this algorithm, at each iteration, the latent variables are successively simulated by the conditional distribution and the unknown parameters are updated. According to Meza et al. (2012), the SAEM method at iteration k proceeds as follows:

E-Step:

1. Simulation-step:

- (a) draw $\mathbf{q}^{(k,l)}$ ($l = 1, \dots, m$) from the conditional distribution $f(\mathbf{q} | \mathbf{y}, \hat{\boldsymbol{\theta}}^{(k-1)})$, or
- (b) MCMC procedure: when random samples cannot be simulated directly from the conditional distribution, draw $\mathbf{q}^{(k,l)}$ ($l = 1, \dots, m$) instead from the transition probability $\Pi_{\hat{\boldsymbol{\theta}}^{(k)}}(\mathbf{q}^{(k-1)}, \cdot)$, the sequence $\mathbf{q}^{(k)}$ is a Markov Chain with transition kernels $\Pi_{\hat{\boldsymbol{\theta}}^{(k)}}$.

2. Stochastic approximation: update $Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(k)})$ according to

$$Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(k)}) = Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(k-1)}) + \delta_k \left[\frac{1}{m} \sum_{l=1}^m \ell_c(\boldsymbol{\theta} | \mathbf{q}^{(k,l)}, \mathbf{y}) - Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(k-1)}) \right], \quad (3)$$

where $\ell_c(\boldsymbol{\theta} | \mathbf{y}_c) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta} | \mathbf{y}_c)$ is the complete log-likelihood function and δ_k is a smoothness parameter, *i.e.*, a decreasing sequence of positive numbers such that $\sum_{k=1}^{\infty} \delta_k = \infty$ and $\sum_{k=1}^{\infty} \delta_k^2 < \infty$.

M-Step:

1. Maximization: update $\boldsymbol{\theta}^{(k)}$ according to

$$\hat{\boldsymbol{\theta}}^{(k+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(k)}).$$

When we need to perform (b) in the E-Step of the SAEM algorithm, this algorithm is called MCMC-SAEM and was proposed by Kuhn and Lavielle (2004). In this paper we will refer indistinctly as SAEM algorithm. As proposed by Galarza et al. (2015) we will consider the following smoothing parameter

$$\delta_k = \begin{cases} 1, & \text{if } 1 \leq k \leq cW; \\ \frac{1}{k-cW}, & \text{if } cW + 1 \leq k \leq W, \end{cases} \quad (4)$$

where W is the maximum number of iterations and c a cut point ($0 \leq c \leq 1$) which determines the percentage of the initial iterations. By Equation (3), we have that if the smoothing parameter δ_k is equal to 1 for all k , the SAEM algorithm has “no memory” and it coincides with the MCEM algorithm. While the SAEM has no memory, the algorithm will converge quickly (convergence in distribution) to a solution neighborhood. However when the algorithm has memory it will converge slowly (almost sure convergence) to the ML solution.

Note that, for the SAEM algorithm, the E-Step coincides with the MCEM algorithm, however a small number of simulations m (suggested to be $m \leq 20$) is necessary. This is possible because unlike the traditional EM algorithm and its variants, the SAEM algorithm uses not only the current simulation of the missing/censored/latent data at the iteration k denoted by $(\mathbf{q}^{(k,l)})$, $l = 1, \dots, m$ but some or all previous simulations, where this “memory” property is set by the smoothing parameter δ_k .

3 Linear regression models for irregularly observed longitudinal data

3.1 The statistical model

Let $\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top)^\top$ denote the vector of observed continuous multivariate responses. Herein, \mathbf{y}_i is a $n_i \times 1$ vector containing the observations for subject i measured at particular time points $\mathbf{t}_i = (t_{i1}, \dots, t_{in_i})$. Formally, the regression model is given by

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \quad (5)$$

where $\boldsymbol{\epsilon}_i$ is the vector of random errors of dimension $(n_i \times 1)$ with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_i$; and \mathbf{X}_i is an $n_i \times p$ design matrix corresponding to the fixed effects, $\boldsymbol{\beta}$, of dimension $p \times 1$. Instead of the usual assumption of normality, we replace the multivariate normal distribution by the scale mixture of multivariate normal distributions. Therefore, it follows that

$$\boldsymbol{\epsilon}_i \stackrel{\text{ind.}}{\sim} \text{SMN}_{n_i}(\mathbf{0}, \boldsymbol{\Omega}_i, \boldsymbol{\nu}), \quad i = 1, \dots, n. \quad (6)$$

The correlation structure of the error vector is assumed to be $\boldsymbol{\Omega}_i = \sigma^2 \mathbf{E}_i$, where the $n_i \times n_i$ matrix \mathbf{E}_i incorporates a time-dependence structure. Consequently, to capture the serial correlation among irregularly observed longitudinal data, it is necessary to consider a parsimonious parameterization of the matrix \mathbf{E}_i . Following Muñoz et al. (1992), we adopt a DEC (damped exponential correlation) structure for \mathbf{E}_i , which is defined as:

$$\mathbf{E}_i = \mathbf{E}_i(\boldsymbol{\phi}, \mathbf{t}_i) = \left[\phi_1^{|t_{ij} - t_{ik}| \phi_2} \right], \quad i = 1, \dots, n, \quad j, k = 1, \dots, n_i, \quad (7)$$

where $\boldsymbol{\phi} = (\phi_1, \phi_2)^\top$, the parameter ϕ_1 describes the autocorrelation between observations separated by the absolute length of two time points, and the parameter ϕ_2 permits acceleration of the exponential decay of the autocorrelation function, defining a continuous-time autoregressive model.

For practical reasons, the parameter space of ϕ_1 and ϕ_2 is confined within $\boldsymbol{\Phi} = \{(\phi_1, \phi_2) : 0 < \phi_1 < 1, \phi_2 > 0\}$. It is important to stress that different values of the damping parameter ϕ_2 produce a variety of correlation structures for a given value of $\phi_1 > 0$, as follows:

- (a) if $\phi_2 = 0$, then \mathbf{E}_i generates the compound symmetry correlation structure;
- (b) when $0 < \phi_2 < 1$, then \mathbf{E}_i presents a decay rate between the compound symmetry structure and the first-order AR (AR (1)) model;
- (c) if $\phi_2 = 1$, then \mathbf{E}_i generates an AR(1) structure;
- (d) when $\phi_2 > 1$, \mathbf{E}_i presents a decay rate faster than the AR(1) structure; and
- (e) if $\phi_2 \rightarrow \infty$, then \mathbf{E}_i represents the first-order moving average model, MA(1).

A more detailed discussion of the DEC structure presenting more complex scenarios of the parameter space $\boldsymbol{\Phi}$ can be found in Muñoz et al. (1992).

Using the stochastic representation (1), the hierarchical representation (two-stages) of the linear regression model defined in (5) - (7) is given by

$$\begin{aligned} \mathbf{y}_i | U_i = u_i &\stackrel{\text{ind.}}{\sim} \mathbf{N}_{n_i}(\mathbf{X}_i \boldsymbol{\beta}, \kappa(u_i) \boldsymbol{\Omega}_i), \\ U_i &\stackrel{\text{iid.}}{\sim} h(u_i | \boldsymbol{\nu}). \end{aligned} \quad (8)$$

Recall that we are interested in the case where left-censored observations can occur. That is, the observations are of the form

$$\begin{aligned} y_{ij} &\leq V_{ij} \quad \text{if } C_{ij} = 1, \\ y_{ij} &= V_{ij} \quad \text{if } C_{ij} = 0, \end{aligned} \quad (9)$$

where V_{ij} represents the uncensored observation or limit of quantification and C_{ij} is the censoring indicator whose value equals one if censored observation and zero if uncensored observation. So, the observed data for the i -th subject is represented by $(\mathbf{V}_i, \mathbf{C}_i)$. We have chosen to work with the left censored case, but the results are easily extendable to other censoring types. Thus, the model defined in (5) – (9) will be called the SMN-CR model.

3.2 The likelihood function

Frequentist inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\phi}^\top, \boldsymbol{\nu}^\top)^\top$ is based on the marginal distribution for $\mathbf{y}_i, i = 1, \dots, n$. For the SMN-CR model with complete data, we have that, marginally,

$$\mathbf{y}_i \stackrel{\text{ind.}}{\sim} \text{SMN}_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Omega}_i, \boldsymbol{\nu}), \quad i = 1, \dots, n. \quad (10)$$

For computing the marginal likelihood, the first step is to treat separately the observed and censored components of \mathbf{y}_i . This procedure is described in the next Proposition.

Proposition 1. *Let \mathbf{y} be partitioned as $\mathbf{y}_i = \text{vec}(\mathbf{y}_i^o, \mathbf{y}_i^c)$ with $\dim(\mathbf{y}_i^o) = n_i^o$, $\dim(\mathbf{y}_i^c) = n_i^c$ and $n_i^o + n_i^c = n_i$, where $\text{vec}(\cdot)$ denotes the operator which stacks vectors or matrices of the same number of columns and $C_{ij} = 0$ for all elements in \mathbf{y}_i^o , and 1 for all elements in \mathbf{y}_i^c . Let $\mathbf{V}_i, \mathbf{X}_i$, and $\boldsymbol{\Omega}_i$ also be partitioned as follows: $\mathbf{V}_i = \text{vec}(\mathbf{V}_i^o, \mathbf{V}_i^c)$, $\mathbf{X}_i^\top = (\mathbf{X}_i^o, \mathbf{X}_i^c)$, and $\boldsymbol{\Omega}_i = \begin{pmatrix} \boldsymbol{\Omega}_i^{oo} & \boldsymbol{\Omega}_i^{oc} \\ \boldsymbol{\Omega}_i^{co} & \boldsymbol{\Omega}_i^{cc} \end{pmatrix}$. Then, we have $\mathbf{y}_i \mid u_i \sim N_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \kappa(u_i)\boldsymbol{\Omega}_i)$, where*

$$\mathbf{y}_i^o \mid u_i \sim N_{n_i^o}(\mathbf{X}_i^o\boldsymbol{\beta}, \kappa(u_i)\boldsymbol{\Omega}_i^{oo}) \quad \text{and} \quad \mathbf{y}_i^c \mid \mathbf{y}_i^o, u_i \sim N_{n_i^c}(\boldsymbol{\mu}_i, \kappa(u_i)\mathbf{S}_i), \quad (11)$$

with $\boldsymbol{\mu}_i = \mathbf{X}_i^c\boldsymbol{\beta} + \boldsymbol{\Omega}_i^{co}(\boldsymbol{\Omega}_i^{oo})^{-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\boldsymbol{\beta})$ and $\mathbf{S}_i = \boldsymbol{\Omega}_i^{cc} - \boldsymbol{\Omega}_i^{co}(\boldsymbol{\Omega}_i^{oo})^{-1}\boldsymbol{\Omega}_i^{oc}$.

Proof 1. *The proof is straightforward from the conditional properties of a normal random vector.*

Following Vaida and Liu (2009), we have the following proposition to calculate the likelihood function.

Proposition 2. *Let $\Phi_{n_i}(\mathbf{u}; \mathbf{a}, \mathbf{A})$ and $\phi_{n_i}(\mathbf{u}; \mathbf{a}, \mathbf{A})$ be the cdf (left tail) and pdf, respectively, of $N_{n_i}(\mathbf{a}, \mathbf{A})$ computed at \mathbf{u} . The likelihood function for the i -th subject is given by*

$$\begin{aligned} L_i(\boldsymbol{\theta}) &= f(\mathbf{y}_i^o \mid \boldsymbol{\theta})f(\mathbf{y}_i^c \leq \mathbf{V}_i^c \mid \mathbf{y}_i^o, \boldsymbol{\theta}) \\ &= \int_0^\infty f(\mathbf{y}_i^o \mid u_i, \boldsymbol{\theta})f(\mathbf{y}_i^c \leq \mathbf{V}_i^c \mid u_i, \mathbf{y}_i^o, \boldsymbol{\theta})dH(u_i) \\ &= \int_0^\infty \phi_{n_i^o}(\mathbf{y}_i^o; \mathbf{X}_i^o\boldsymbol{\beta}, \kappa(u_i)\boldsymbol{\Omega}_i^{oo})\Phi_{n_i^c}(\mathbf{V}_i^c; \boldsymbol{\mu}_i, \kappa(u_i)\mathbf{S}_i)dH(u_i). \end{aligned} \quad (12)$$

Proof 2. *The proof follows directly from the conditional probability arguments sketched in Jacqmin-Gadda et al. (2000) and Vaida and Liu (2009).*

The log-likelihood function for the observed data is given by $\ell(\boldsymbol{\theta} \mid \mathbf{y}) = \sum_{i=1}^n \{\log L_i\}$ and can be used to monitor the convergence of the SAEM algorithm. The likelihood function for particular cases of the SMN-CR model are given in following Corollary. The proof is given in Appendix A.

Corollary 1. *The likelihood function for special elements of the SMN class are given by.*

(a) (normal) *If U is degenerate in 1, i.e., $P(U = 1) = 1$, then*

$$L_i(\boldsymbol{\theta}) = \phi_{n_i^o}(\mathbf{y}_i^o; \mathbf{X}_i^o\boldsymbol{\beta}, \kappa(u_i)\boldsymbol{\Omega}_i^{oo})\Phi_{n_i^c}(\mathbf{V}_i^c; \boldsymbol{\mu}_i, \mathbf{S}_i).$$

(b) (Student's-t) If $\kappa(u) = 1/u$ and U is distributed as $\text{Gamma}(\nu/2, \nu/2)$, with $\nu > 0$, then

$$L_i(\boldsymbol{\theta}) = t_{n_i^o}(\mathbf{y}_i^o; \mathbf{X}_i^o \boldsymbol{\beta}, \boldsymbol{\Omega}_i^{oo}, \nu) T_{n_i^c} \left(\mathbf{V}_i^c; \boldsymbol{\mu}_i, \left(\frac{\nu + \boldsymbol{\delta}}{\nu + n_i^o} \right) \mathbf{S}_i, \nu + n_i^o \right),$$

$$\text{where } \boldsymbol{\delta} = (\mathbf{y}_i^o - \mathbf{X}_i^o \boldsymbol{\beta})^\top (\boldsymbol{\Omega}_i^{oo})^{-1} (\mathbf{y}_i^o - \mathbf{X}_i^o \boldsymbol{\beta}).$$

(c) (contaminated normal) If $\kappa(u) = 1/u$ and \mathbf{U} is a discrete random variable taking one of two states and with probability function given by $h(u|\boldsymbol{\nu}) = \nu \mathbb{I}_{\{\gamma\}}(u) + (1 - \nu) \mathbb{I}_{\{1\}}(u)$, then

$$\begin{aligned} L_i(\boldsymbol{\theta}) &= \nu [\phi_{n_i^o}(\mathbf{y}_i^o; \mathbf{X}_i^o \boldsymbol{\beta}, \gamma^{-1} \boldsymbol{\Omega}_i^{oo}) \Phi_{n_i^c}(\mathbf{V}_i^c; \boldsymbol{\mu}_i, \gamma^{-1} \mathbf{S}_i)] \\ &+ (1 - \nu) [\phi_{n_i^o}(\mathbf{y}_i^o; \mathbf{X}_i^o \boldsymbol{\beta}, \boldsymbol{\Omega}_i^{oo}) \Phi_{n_i^c}(\mathbf{V}_i^c; \boldsymbol{\mu}_i, \mathbf{S}_i)]. \end{aligned}$$

Lucas (1997) carried out an interesting study on the robust aspects of the Student's-t M-estimator in the univariate case using influence functions. He showed that the protection against outliers is preserved only if the degrees of freedom parameter are fixed. In this paper, we assume that the parameter $\boldsymbol{\nu}$ is fixed and choose the most appropriate value of $\boldsymbol{\nu}$ (see Lange et al. 1989; Meza et al. 2012) based on AIC or BIC. The entire parameter vector is $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\phi}^\top)^\top$ hereafter.

3.3 Maximum likelihood estimation

In this subsection, we develop the MCMC-SAEM (hereafter SAEM) algorithm for ML estimation of the parameters in the SMN-CR model defined previously. Consider the model defined in (5) – (7), $\mathbf{u} = (u_1, \dots, u_n)^\top$, $\mathbf{V} = \text{vec}(\mathbf{V}_1, \dots, \mathbf{V}_n)$, and $\mathbf{C} = \text{vec}(\mathbf{C}_1, \dots, \mathbf{C}_n)$ 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 $\mathbf{y}_c = (\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{y}^\top, \mathbf{u}^\top)^\top$ as the complete data. Therefore, the complete data log-likelihood function for all individuals can be written, using the representation defined in (8), as $\ell_c(\boldsymbol{\theta} | \mathbf{y}_c) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta} | \mathbf{y}_c)$,

$$\begin{aligned} \ell_c(\boldsymbol{\theta} | \mathbf{y}_c) &= \sum_{i=1}^n \{\log f(\mathbf{y}_i | u_i) + \log h(u_i | \boldsymbol{\nu})\} \\ &= -\frac{N}{2} \log \sigma^2 - \sum_{i=1}^n \frac{1}{2} \log |\mathbf{E}_i| - \sum_{i=1}^n \frac{\kappa^{-1}(u_i)}{2\sigma^2} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^\top \mathbf{E}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \\ &+ \sum_{i=1}^n \log h(u_i | \boldsymbol{\nu}) + C, \end{aligned}$$

with C being a constant that does not depend on the parameter vector $\boldsymbol{\theta}$ and $\sum_{i=1}^n n_i = N$. Given the current estimate (at the k -th iteration) $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(k)}$, the conditional expectation of the complete data log-likelihood function is given by:

$$Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(k)}) = E \left[\ell_c(\boldsymbol{\theta} | \mathbf{y}_c) | \mathbf{V}, \mathbf{C}, \hat{\boldsymbol{\theta}}^{(k)} \right] = \sum_{i=1}^n Q_i(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(k)}),$$

where

$$\begin{aligned} Q_i(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(k)}) &= -\frac{n_i}{2} \log \widehat{\sigma}^{2(k)} - \frac{1}{2} \log |\widehat{\mathbf{E}}_i^{(k)}| \\ &- \frac{1}{2\widehat{\sigma}^{2(k)}} E \left[\kappa^{-1}(u_i) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^\top \mathbf{E}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) | \mathbf{V}, \mathbf{C}, \hat{\boldsymbol{\theta}}^{(k)} \right] \\ &= -\frac{n_i}{2} \log \widehat{\sigma}^{2(k)} - \frac{1}{2} \log |\widehat{\mathbf{E}}_i^{(k)}| - \frac{1}{2\widehat{\sigma}^{2(k)}} \left[\text{tr} \left(\widehat{\boldsymbol{\kappa}}_i^{2(k)} \widehat{\mathbf{E}}_i^{-1(k)} \right) \right. \\ &\left. - 2\widehat{\boldsymbol{\beta}}^{(k)\top} \mathbf{X}_i^\top \widehat{\mathbf{E}}_i^{-1(k)} \widehat{\boldsymbol{\kappa}}_i \mathbf{y}_i^{(k)} + \widehat{\boldsymbol{\kappa}}_i^{(k)} \widehat{\boldsymbol{\beta}}^{(k)\top} \mathbf{X}_i^\top \widehat{\mathbf{E}}_i^{-1(k)} \mathbf{X}_i \widehat{\boldsymbol{\beta}}^{(k)} \right], \end{aligned}$$

with

$$\widehat{\kappa \mathbf{y}_i^2}^{(k)} = E \left[\kappa^{-1}(u_i) \mathbf{y}_i \mathbf{y}_i^\top \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)} \right], \quad (13)$$

$$\widehat{\kappa \mathbf{y}_i}^{(k)} = E \left[\kappa^{-1}(u_i) \mathbf{y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)} \right], \quad (14)$$

$$\widehat{\kappa}_i^{(k)} = E \left[\kappa^{-1}(u_i) \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)} \right]. \quad (15)$$

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

In the traditional EM algorithm, we evaluate the conditional expectations given in Equations (13) – (15). As there are no closed-form expressions for them, two intermediate steps are introduced, including the simulation and approximation steps. In the simulation, for the i -th subject, we generate samples from the full conditional distributions of the latent variables (u_i, \mathbf{y}_i) through the Gibbs sampler algorithm according to the following scheme (at the k -th iteration):

Step 1:

Sample $\mathbf{y}_i^{c(k,l)}$ from $f(\mathbf{y}_i^c \mid \mathbf{V}_i^c, \mathbf{y}_i^o, u_i, \widehat{\boldsymbol{\theta}}^{(k)})$, which is a truncated normal distribution. By Proposition 1, we obtain $\mathbf{y}_i^c \mid \mathbf{V}_i^c, \mathbf{y}_i^o, u_i, \boldsymbol{\theta} \sim \text{TN}_{n_i^c}(\boldsymbol{\mu}_i, \kappa(u_i) \mathbf{S}_i; \mathbb{A}_i)$, with $\mathbb{A}_i = \{\mathbf{y}_i^c = (y_{i1}^c, \dots, y_{in_i^c}^c)^\top \mid y_{i1}^c \leq V_{i1}^c, \dots, y_{in_i^c}^c \leq V_{in_i^c}^c\}$, $\boldsymbol{\mu}_i = \mathbf{X}_i^c \boldsymbol{\beta} + \boldsymbol{\Omega}_i^{co} (\boldsymbol{\Omega}_i^{oo})^{-1} (\mathbf{y}_i^o - \mathbf{X}_i^o \boldsymbol{\beta})$ and $\mathbf{S}_i = \boldsymbol{\Omega}_i^{cc} - \boldsymbol{\Omega}_i^{co} (\boldsymbol{\Omega}_i^{oo})^{-1} \boldsymbol{\Omega}_i^{oc}$.

Then, the new observation $\mathbf{y}_i^{(k,l)} = (y_{i1}^{c(k,l)}, \dots, y_{in_i^c}^{c(k,l)}, y_{n_i^c+1}, \dots, y_{n_i})$ is a sample generated for the n_i^c censored cases and the observed values (uncensored cases).

Step 2:

Sample $u_i^{(k,l)}$ from $f(u_i \mid \mathbf{y}_i^{(k,l)}, \widehat{\boldsymbol{\theta}}^{(k)})$. This gives rise to

(a) Student's- t

$$u_i \mid \mathbf{y}_i, \boldsymbol{\theta} \sim \text{Gamma} \left(\frac{\nu + n_i}{2}, \frac{\nu + (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^\top \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})}{2} \right);$$

(b) Slash

$$u_i \mid \mathbf{y}_i, \boldsymbol{\theta} \sim \text{TGamma} \left(\nu + \frac{n_i}{2}, \frac{(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^\top \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})}{2}; (0, 1) \right),$$

which follows a truncated gamma distribution lying in the interval $(0, 1)$;

(c) Contaminated normal

$f(u_i \mid \mathbf{y}_i, \boldsymbol{\theta})$, is a discrete distribution taking values γ with probability $\frac{p_1}{p_1 + p_2}$ and 1 with probability $\frac{p_2}{p_1 + p_2}$, where

$$p_1 = \nu \gamma^{\frac{n_i}{2}} \exp \left(-\frac{\gamma}{2} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^\top \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right),$$

$$p_2 = (1 - \nu) \exp \left(-\frac{1}{2} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^\top \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right).$$

The next step is the **Stochastic Approximation**. Since the sequence $(\mathbf{y}_i^{(k,l)}, u_i^{(k,l)})$ for $l = 1, \dots, m$ is collected at the k -th iteration, we replace the conditional expectations given in (13) – (15) by the following stochastic approximations:

$$\widehat{\kappa\mathbf{y}_i^2}^{(k)} = \widehat{\kappa\mathbf{y}_i^2}^{(k-1)} + \delta_k \left[\frac{1}{m} \sum_{l=1}^m \kappa^{-1}(u_i^{(k,l)}) \mathbf{y}_i^{(k,l)} \mathbf{y}_i^{(k,l)\top} - \widehat{\kappa\mathbf{y}_i^2}^{(k-1)} \right], \quad (16)$$

$$\widehat{\kappa\mathbf{y}_i}^{(k)} = \widehat{\kappa\mathbf{y}_i}^{(k-1)} + \delta_k \left[\frac{1}{m} \sum_{l=1}^m \kappa^{-1}(u_i^{(k,l)}) \mathbf{y}_i^{(k,l)} - \widehat{\kappa\mathbf{y}_i}^{(k-1)} \right], \quad (17)$$

$$\widehat{\kappa}_i^{(k)} = \widehat{\kappa}_i^{(k-1)} + \delta_k \left[\frac{1}{m} \sum_{l=1}^m \kappa^{-1}(u_i^{(k,l)}) - \widehat{\kappa}_i^{(k-1)} \right]. \quad (18)$$

An advantage of the SAEM algorithm is that, even though it performs a MCMC E-step, it requires a small and fixed sample size, making it much faster than MCEM. Some authors claim that $m \leq 10$ is large enough, but to be more conservative, we chose $m = 20$. As a consequence, the MCMC samples are incorporated in a smooth way with the previous step of the algorithm.

Finally, the conditional maximization step is carried out and $\widehat{\boldsymbol{\theta}}^{(k)}$ is updated by maximizing $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ over $\widehat{\boldsymbol{\theta}}^{(k)}$, which leads to the following expressions:

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \left(\sum_{i=1}^n \widehat{\kappa}_i^{(k)} \mathbf{X}_i^\top \widehat{\mathbf{E}}_i^{-1(k)} \mathbf{X}_i \right)^{-1} \sum_{i=1}^n \mathbf{X}_i^\top \widehat{\mathbf{E}}_i^{-1(k)} \widehat{\kappa\mathbf{y}_i}^{(k)}, \quad (19)$$

$$\begin{aligned} \widehat{\sigma}^2^{(k+1)} &= \frac{1}{N} \sum_{i=1}^n \left[\text{tr} \left(\widehat{\kappa\mathbf{y}_i^2}^{(k)} \widehat{\mathbf{E}}_i^{-1(k)} \right) - 2\widehat{\boldsymbol{\beta}}^{(k)\top} \mathbf{X}_i^\top \widehat{\mathbf{E}}_i^{-1(k)} \widehat{\kappa\mathbf{y}_i}^{(k)} \right. \\ &\quad \left. + \widehat{\kappa}_i^{(k)} \widehat{\boldsymbol{\beta}}^{(k)\top} \mathbf{X}_i^\top \widehat{\mathbf{E}}_i^{-1(k)} \mathbf{X}_i \widehat{\boldsymbol{\beta}}^{(k)} \right], \end{aligned} \quad (20)$$

$$\begin{aligned} \widehat{\phi}^{(k+1)} &= \underset{\boldsymbol{\phi} \in (0,1) \times \mathbb{R}^+}{\text{argmax}} \left(-\frac{1}{2\widehat{\sigma}^2^{(k)}} \sum_{i=1}^n \left[\text{tr} \left(\widehat{\kappa\mathbf{y}_i^2}^{(k)} \mathbf{E}_i^{-1} \right) - 2\widehat{\boldsymbol{\beta}}^{(k)\top} \mathbf{X}_i^\top \mathbf{E}_i^{-1} \widehat{\kappa\mathbf{y}_i}^{(k)} \right. \right. \\ &\quad \left. \left. + \widehat{\kappa}_i^{(k)} \widehat{\boldsymbol{\beta}}^{(k)\top} \mathbf{X}_i^\top \mathbf{E}_i^{-1} \mathbf{X}_i \widehat{\boldsymbol{\beta}}^{(k)} \right] - \frac{1}{2} \sum_{i=1}^n \log(|\mathbf{E}_i^{-1}|) \right). \end{aligned} \quad (21)$$

It is important to stress that $\widehat{\kappa\mathbf{y}_i^2}^{(k)}$, $\widehat{\kappa\mathbf{y}_i}^{(k)}$ and $\widehat{\kappa}_i^{(k)}$ rely on minimal sufficient statistics. Since the complete likelihood function does belong to the exponential family, the parameters estimates of this SAEM algorithm converges. Under several conditions, Kuhn and Lavielle (2005) and Samson et al. (2006) have verified that the estimate sequence produced by the SAEM algorithm converges towards a (local) maximum of the likelihood function.

3.4 Imputation of censored components

We are also interested in the prediction of the censored components of the i -th subject. Let \mathbf{y}_i^c be the true unobserved response vector for the censored components. In the implementation of the SAEM algorithm, the predictions of the censored components, denoted by $\widehat{\mathbf{y}}_i^{c(k)}$, are calculated as

$$\widehat{\mathbf{y}}_i^{c(k)} = E\{\mathbf{y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\}, \quad i = 1, \dots, n,$$

where

$$\widehat{\mathbf{y}}_i^{c(k)} = \widehat{\mathbf{y}}_i^{c(k-1)} + \delta_k \left[\frac{1}{m} \sum_{l=1}^m \mathbf{y}_i^{c(k,l)} - \widehat{\mathbf{y}}_i^{c(k)} \right] \quad (22)$$

and the $\mathbf{y}_i^{c(k,l)}$'s are obtained without computational effort from the **Step 1** of the proposed SAEM algorithm.

4 Standard errors and prediction of future observations

4.1 Empirical information matrix

Following Lin (2010), the asymptotic covariance of the ML estimates can be approximated by

$$\mathbf{I}_e(\boldsymbol{\theta} | \mathbf{y}) = \sum_{i=1}^n \mathbf{s}(\mathbf{y}_i | \boldsymbol{\theta}) \mathbf{s}^\top(\mathbf{y}_i | \boldsymbol{\theta}) - \frac{1}{n} \mathbf{S}(\mathbf{y}_i | \boldsymbol{\theta}) \mathbf{S}^\top(\mathbf{y}_i | \boldsymbol{\theta}), \quad (23)$$

where $\mathbf{S}(\mathbf{y}_i | \boldsymbol{\theta}) = \sum_{i=1}^n \mathbf{s}(\mathbf{y}_i | \boldsymbol{\theta})$ and $\mathbf{s}(\mathbf{y}_i | \boldsymbol{\theta})$ is the empirical score function for the i -th subject. According to Louis (1982), the individual score is determined as

$$\mathbf{s}(\mathbf{y}_i | \boldsymbol{\theta}) = \frac{\partial \log f(\mathbf{y}_i | \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = E \left(\frac{\partial \ell_{ic}(\boldsymbol{\theta} | \mathbf{y}_{ci})}{\partial \boldsymbol{\theta}} | \mathbf{V}_i, \mathbf{C}_i, \boldsymbol{\theta} \right), \quad (24)$$

where $\ell_{ic}(\boldsymbol{\theta} | \mathbf{y}_{ci})$ is the complete data log-likelihood formed from the complete observation \mathbf{y}_{ci} . Substituting the ML estimate of $\boldsymbol{\theta}$ in (24), it leads to $\mathbf{s}(\mathbf{y}_i | \hat{\boldsymbol{\theta}}) = 0$. As a result, the empirical information matrix $\mathbf{I}_e(\hat{\boldsymbol{\theta}} | \mathbf{y})$ is reduced to

$$\mathbf{I}_e(\hat{\boldsymbol{\theta}} | \mathbf{y}) = \sum_{i=1}^n \hat{\mathbf{s}}_i \hat{\mathbf{s}}_i^\top = \begin{pmatrix} \hat{\mathbf{s}}_{i,\beta} \\ \hat{\mathbf{s}}_{i,\sigma^2} \\ \hat{\mathbf{s}}_{i,\phi} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{s}}_{i,\beta} & \hat{\mathbf{s}}_{i,\sigma^2} & \hat{\mathbf{s}}_{i,\phi} \end{pmatrix}, \quad (25)$$

where

$$\begin{aligned} \hat{\mathbf{s}}_{i,\beta} &= (\hat{\mathbf{s}}_{i,\beta_1}, \dots, \hat{\mathbf{s}}_{i,\beta_p})^\top = \frac{1}{\sigma^2} \left(\mathbf{X}_i^\top \hat{\mathbf{E}}_i^{-1} \widehat{\boldsymbol{\kappa}} \mathbf{y}_i - \hat{\kappa}_i \mathbf{X}_i^\top \hat{\mathbf{E}}_i^{-1} \mathbf{X}_i \hat{\boldsymbol{\beta}} \right), \\ \hat{\mathbf{s}}_{i,\sigma^2} &= -\frac{n_i}{2\sigma^2} + \frac{1}{2\sigma^4} \left[\text{tr} \left(\widehat{\boldsymbol{\kappa}} \mathbf{y}_i^\top \hat{\mathbf{E}}_i^{-1} \right) - 2\hat{\boldsymbol{\beta}}^\top \mathbf{X}_i^\top \hat{\mathbf{E}}_i^{-1} \widehat{\boldsymbol{\kappa}} \mathbf{y}_i + \hat{\kappa}_i \hat{\boldsymbol{\beta}}^\top \mathbf{X}_i^\top \hat{\mathbf{E}}_i^{-1} \mathbf{X}_i \hat{\boldsymbol{\beta}} \right], \\ \hat{\mathbf{s}}_{i,\phi} &= (\hat{\mathbf{s}}_{i,\phi_1}, \hat{\mathbf{s}}_{i,\phi_2})^\top \end{aligned}$$

with

$$\begin{aligned} \hat{\mathbf{s}}_{i,\phi_s} &= \frac{1}{2\sigma^2} \left[\text{tr} \left(\widehat{\boldsymbol{\kappa}} \mathbf{y}_i^\top \hat{\mathbf{E}}_i^{-1} \dot{\mathbf{E}}_i(s) \hat{\mathbf{E}}_i^{-1} \right) - 2\hat{\boldsymbol{\beta}}^\top \mathbf{X}_i^\top \hat{\mathbf{E}}_i^{-1} \dot{\mathbf{E}}_i(s) \hat{\mathbf{E}}_i^{-1} \widehat{\boldsymbol{\kappa}} \mathbf{y}_i + \hat{\kappa}_i \hat{\boldsymbol{\beta}}^\top \mathbf{X}_i^\top \hat{\mathbf{E}}_i^{-1} \dot{\mathbf{E}}_i(s) \hat{\mathbf{E}}_i^{-1} \mathbf{X}_i \hat{\boldsymbol{\beta}} \right], \\ &\quad - \frac{1}{2} \text{tr} \left(\hat{\mathbf{E}}_i^{-1} \dot{\mathbf{E}}_i(s) \right), \end{aligned}$$

and $\dot{\mathbf{E}}_i(s) = \frac{\partial \mathbf{E}_i}{\partial \phi_s} |_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}}}$ for $s = 1, 2$. For the DEC structure, we have the following partial derivatives

$$\begin{aligned} \frac{\partial \mathbf{E}_i}{\partial \phi_1} &= |t_{ij} - t_{ik}|^{\phi_2} \phi_1^{|t_{ij} - t_{ik}|^{\phi_2 - 1}}, \\ \frac{\partial \mathbf{E}_i}{\partial \phi_2} &= |t_{ij} - t_{ik}|^{\phi_2} \log(|t_{ij} - t_{ik}|) \log(\phi_1) \phi_1^{|t_{ij} - t_{ik}|^{\phi_2}}. \end{aligned}$$

4.2 Prediction

For generating predicted values from the SMN-CR model, we follow the scheme adopted by Wang (2013) and Garay et al. (2014). 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 $\bar{\mathbf{y}}_i = (\mathbf{y}_{i,obs}^\top, \mathbf{y}_{i,pred}^\top)$.

The censored values existing in $\mathbf{y}_{i,obs}$ are imputed by (22). Therefore, after this imputation step, a complete data set, \mathbf{y}_{i,obs^*} , is obtained. We obtain

$$\bar{\mathbf{y}}_i^* = \left(\mathbf{y}_{i,obs^*}^\top, \mathbf{y}_{i,pred}^\top \right)^\top \sim SMN_{n_{i,obs}+n_{i,pred}}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Omega}_i; \mathbf{H}),$$

where $\boldsymbol{\Omega}_i = \begin{pmatrix} \boldsymbol{\Omega}_i^{obs^*,obs^*} & \boldsymbol{\Omega}_i^{obs^*,pred} \\ \boldsymbol{\Omega}_i^{pred,obs^*} & \boldsymbol{\Omega}_i^{pred,pred} \end{pmatrix}$. The best linear predictor of $\mathbf{y}_{i,pred}$ (with respect to the minimum mean squared error) is the conditional expectation of $\mathbf{y}_{i,pred}$ given \mathbf{y}_{i,obs^*} , namely

$$\hat{\mathbf{y}}_{i,pred}(\boldsymbol{\theta}) = \mathbf{X}_{i,pred}\boldsymbol{\beta} + \boldsymbol{\Omega}_i^{pred,obs^*} \boldsymbol{\Omega}_i^{obs^*,obs^*}{}^{-1} (\mathbf{y}_{i,obs^*} - \mathbf{X}_{i,obs^*}\boldsymbol{\beta}). \quad (26)$$

Consequently, $\mathbf{y}_{i,pred}$ can be estimated directly by substituting $\hat{\boldsymbol{\theta}}$ into (26).

5 Application

In this section, we illustrate the performance of the proposed techniques through simulated datasets. Afterward, we apply the methods to the analysis of a HIV dataset previously analyzed by Garay et al. (2014) under the Student's- t distribution.

5.1 Simulation study

The main goal of this simulation study is to investigate the effects on the parameter inference when the traditional normality assumption is violated. We examine the behavior of the models under different proportions of censoring and sample sizes.

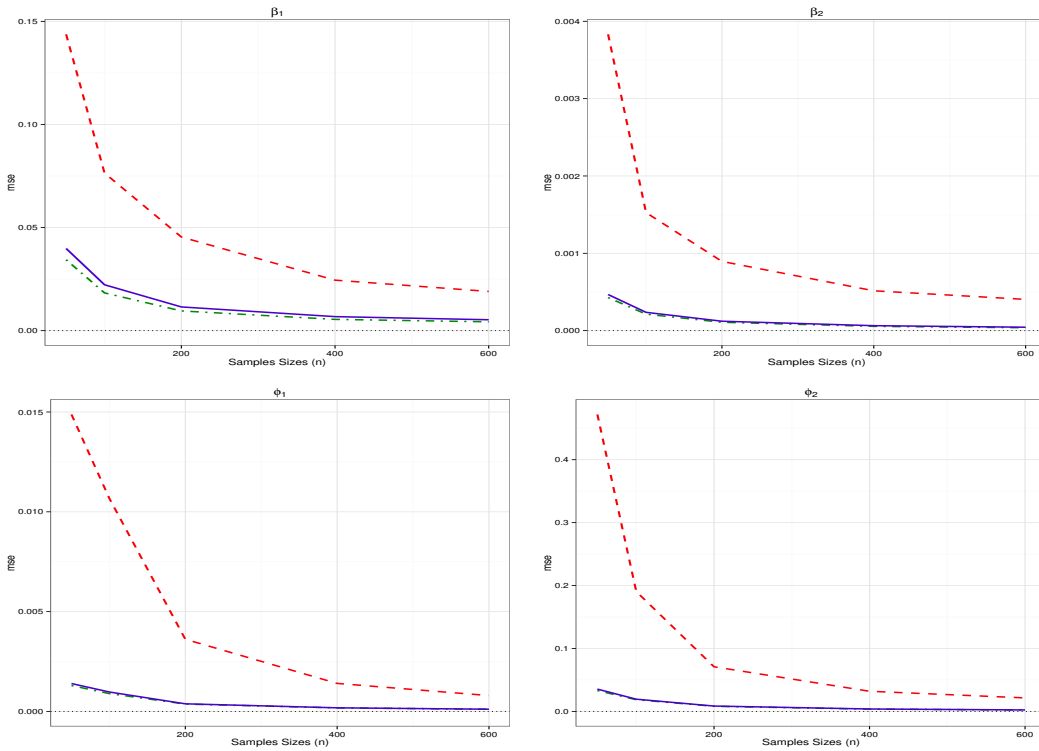


Figure 2: Simulation study - **Scenario 1**. Mean square error of the parameter estimates in the SMN-CR model under 10% of censoring level and different samples sizes. The solid line (blue) represents the T-CR model, the dotted line (red) represents the N-CR model and the dotdash line (green) represents the SL-CR model.

Table 1: Simulation study - **Scenario 1**. Results based on 100 simulated samples with 10% of censoring proportion. MC mean and MC Sd are the respective mean estimates and standard deviations from fitting SMN-CR models with different samples sizes. IM SE is the average value of the approximate standard error obtained through the empirical information-based method. MC AIC and MC BIC are the arithmetic averages of the respective model comparison measures.

		Censoring 10%								
Fit		Parameters					Criteria			
		β_1	β_2	σ^2	ϕ_1	ϕ_2	MC Loglik.	MC AIC	MC BIC	
$n = 50$	T	MC Mean	2.021	0.997	2.019	0.799	1.020	-539.049	1088.098	1106.617
		IM SE	0.235	0.024	0.359	0.046	0.203			
		MC Sd	0.200	0.022	0.351	0.038	0.188			
	SL	MC Mean	2.015	0.997	1.446	0.800	1.022			
		IM SE	0.246	0.025	0.206	0.045	0.197			
		MC Sd	0.186	0.021	0.210	0.036	0.182			
	N	MC Mean	2.121	1.003	16.052	0.801	1.184			
		IM SE	1.012	0.089	1.554	0.050	0.278			
		MC Sd	0.361	0.062	8.093	0.123	0.665			
$n = 100$	T	MC Mean	2.023	0.997	1.973	0.799	1.015	-1075.346	2160.692	2182.676
		IM SE	0.163	0.017	0.242	0.031	0.139			
		MC Sd	0.148	0.015	0.253	0.031	0.139			
	SL	MC Mean	2.016	0.998	1.420	0.800	1.018			
		IM SE	0.171	0.017	0.139	0.030	0.134			
		MC Sd	0.135	0.014	0.159	0.030	0.136			
	N	MC Mean	2.088	1.009	16.134	0.790	1.051			
		IM SE	0.633	0.053	0.688	0.020	0.081			
		MC Sd	0.263	0.038	6.519	0.103	0.436			
$n = 200$	T	MC Mean	2.023	0.997	1.972	0.801	1.010	-2152.751	4315.502	4340.953
		IM SE	0.114	0.012	0.169	0.022	0.096			
		MC Sd	0.105	0.011	0.172	0.020	0.092			
	SL	MC Mean	2.016	0.997	1.421	0.801	1.013			
		IM SE	0.120	0.012	0.097	0.021	0.093			
		MC Sd	0.097	0.010	0.106	0.019	0.089			
	N	MC Mean	2.068	1.011	16.115	0.798	1.016			
		IM SE	0.419	0.034	0.375	0.009	0.040			
		MC Sd	0.203	0.028	4.570	0.060	0.267			
$n = 400$	T	MC Mean	2.023	0.997	1.972	0.802	1.010	-4309.241	8628.483	8657.399
		IM SE	0.081	0.008	0.119	0.015	0.068			
		MC Sd	0.080	0.008	0.109	0.013	0.061			
	SL	MC Mean	2.017	0.998	1.423	0.802	1.013			
		IM SE	0.084	0.009	0.068	0.015	0.065			
		MC Sd	0.072	0.007	0.069	0.013	0.059			
	N	MC Mean	2.061	1.011	16.166	0.803	1.002			
		IM SE	0.289	0.023	0.237	0.005	0.025			
		MC Sd	0.145	0.020	3.125	0.038	0.180			
$n = 600$	T	MC Mean	2.020	0.998	1.972	0.802	1.010	-6468.005	12946.01	12976.95
		IM SE	0.066	0.007	0.097	0.012	0.055			
		MC Sd	0.070	0.006	0.080	0.010	0.047			
	SL	MC Mean	2.014	0.998	1.423	0.803	1.013			
		IM SE	0.069	0.007	0.056	0.012	0.053			
		MC Sd	0.064	0.006	0.051	0.010	0.045			
	N	MC Mean	2.057	1.012	16.228	0.803	0.988			
		IM SE	0.235	0.019	0.187	0.004	0.019			
		MC Sd	0.126	0.016	2.459	0.028	0.147			

We present three scenarios considering the same probability distribution and correlation structure for the datasets. The responses follow a contaminated normal distribution with parameter $\boldsymbol{\nu} = (\nu, \gamma)^\top = (0.1, 0.1)^\top$ and DEC structure with $\phi_1 = 0.8$ and $\phi_2 = 1$. The simulated data are generated following the model defined in Subsection 3.1, with parameters setting at $\beta_1 = 2$ $\beta_2 = 1$, $\sigma^2 = 2$ and time points set as $t_i = (1, 3, 5, 7, 10, 14)$, for $i = 1, \dots, n$.

Scenario 1: A censoring proportion of 10% and different sample sizes, say, $n = 50, 100, 200, 400$ and 600. Under each setting, we fitted the N-CR model, the T-CR model with 4 degrees of freedom

and the SL-CR model with $\nu = 2$. The goal in this study is to show the asymptotic behavior of the ML estimates obtained via the proposed SAEM algorithm.

Scenario 2: A sample of size $n = 200$ and different censoring proportions, say, 0, 5, 10, 20 and 30%. As in the previous case, the N-, T- and SL-CR models are fitted. We aim at studying the behavior of the SMN-CR models under different proportions of censoring.

Scenario 3: We consider a data set of sample size $n = 100$ and a censoring level of 5% to show the convergence of the SAEM algorithm and the imputation performance of censored values.

Note that, for scenarios 1 and 2, there are 30 different simulation settings with 100 simulated Monte Carlo datasets for each one. The ML estimates and their associate standard errors together with the AIC and BIC values were recorded. For all the fitted models, the initial estimates are chosen by fitting a linear regression for all the parameters and we fixed the number maximum of iterations $W = 300$ and a cut point $c = 0.25$.

Scenario 1

To study the finite sample properties, we compute the absolute bias (Bias) and mean square error (MSE) of the regression coefficient estimates obtained from the SMN-CR models under different sample sizes. These measures are defined as:

$$\text{Bias} = \frac{1}{100} \sum_{j=1}^{100} |\hat{\theta}_i^{(j)} - \theta_i| \quad \text{and} \quad \text{MSE} = \frac{1}{100} \sum_{j=1}^{100} \left(\hat{\theta}_i^{(j)} - \theta_i \right)^2, \quad (27)$$

where $\hat{\theta}_i^{(j)}$ is the ML estimate of the parameter θ_i for the j -th sample, $j = 1, \dots, 100$. The main objective of this simulation is to provide empirical evidence about consistence of the ML estimates. It is apparently seen in Figure 2 that the MSE tends to zero as the sample size increases. Similar results are obtained after the analysis of the absolute bias (see Figure 5 in Appendix C). In general, for all models, the SAEM algorithm provides estimates with good asymptotic properties. In addition, Table 1 presents the summary statistics for parameter estimation under this scenario. As expected, censored models with heavy-tailed distributions have better performance than the normal one in recovering the true parameter values independently of sample sizes.

Scenario 2

In this scenario, we intend to study the behavior of the SMN-CR models under different proportions of censoring. It can be found from Table 2 that the heavy-tailed models outperforms the normal one for all levels of censoring. In fact, those models have smaller standard deviations. In addition, Monte Carlo means of the model comparison criteria (MC AIC and MC BIC) strongly favor the heavy-tailed ones.

Table 2 provides the Monte Carlo standard errors of the SAEM estimates obtained through the empirical information matrix described in Section 4 (IM SE). Note that, comparing to the Monte Carlo standard deviation (MC Sd) for the parameters of interest, it is evident that the proposed asymptotic approximation for the variances of the parameters obtained through Equation (25) is reliable. Furthermore, it is readily seen that the estimates of the scale parameter σ^2 obtained from the heavy-tailed models are less sensitive to the variation in the censoring level, concluding that these models are not only robust to model misspecification but also for different levels of censoring.

Scenario 3

The aim of this last simulation study is monitor the convergence of the SAEM algorithm as well the performance of the imputation procedure. To this end, an arbitrary simulated dataset is considered, where

Table 2: Simulation study - **Scenario 2**. Results based on 100 simulated samples with sample size 200. MC mean and MC Sd are the respective mean estimates and standard deviations from fitting SMN-CRM with different settings of censoring proportions. IM SE is the average value of the approximate standard error obtained through the information-based method. MC AIC and MC BIC are the arithmetic averages of the respective model comparison measures.

		$n = 200$								
Censoring	Fit	Parameters					Criteria			
		β_1	β_2	σ^2	ϕ_1	ϕ_2	MC Loglik.	MC AIC	MC BIC	
0%	T	MC Mean	2.010	0.998	1.994	0.801	1.009	-2357.915	4725.830	4751.280
		IM SE	0.111	0.012	0.168	0.021	0.093			
		MC Sd	0.104	0.011	0.172	0.018	0.087			
	SL	MC Mean	2.007	0.999	1.425	0.801	1.012			
		IM SE	0.116	0.012	0.095	0.020	0.090			
		MC Sd	0.096	0.010	0.106	0.018	0.086			
	N	MC Mean	2.031	0.999	22.820	0.800	1.033			
		IM SE	0.392	0.040	0.397	0.008	0.035			
		MC Sd	0.313	0.032	6.086	0.050	0.259			
5%	T	MC Mean	2.011	0.998	1.988	0.802	1.013	-2221.685	4453.371	4478.821
		IM SE	0.112	0.012	0.170	0.021	0.094			
		MC Sd	0.106	0.011	0.171	0.018	0.086			
	SL	MC Mean	2.008	0.998	1.425	0.803	1.016			
		IM SE	0.117	0.012	0.096	0.020	0.091			
		MC Sd	0.098	0.010	0.106	0.018	0.085			
	N	MC Mean	2.300	0.988	16.115	0.800	1.040			
		IM SE	0.413	0.034	0.350	0.008	0.038			
		MC Sd	0.226	0.028	4.549	0.060	0.274			
10%	T	MC Mean	2.023	0.997	1.972	0.801	1.010	-2152.751	4315.502	4340.953
		IM SE	0.114	0.012	0.169	0.022	0.096			
		MC Sd	0.105	0.011	0.172	0.020	0.092			
	SL	MC Mean	2.016	0.997	1.421	0.801	1.013			
		IM SE	0.120	0.012	0.097	0.021	0.093			
		MC Sd	0.097	0.010	0.106	0.019	0.089			
	N	MC Mean	2.068	1.011	16.115	0.798	1.016			
		IM SE	0.419	0.034	0.375	0.009	0.040			
		MC Sd	0.203	0.028	4.570	0.060	0.267			
20%	T	MC Mean	2.091	0.991	1.968	0.797	1.012	-1987.323	3984.646	4010.097
		IM SE	0.129	0.013	0.171	0.024	0.106			
		MC Sd	0.105	0.010	0.169	0.021	0.101			
	SL	MC Mean	2.071	0.993	1.420	0.798	1.017			
		IM SE	0.134	0.013	0.099	0.023	0.102			
		MC Sd	0.097	0.010	0.108	0.020	0.096			
	N	MC Mean	1.627	1.051	16.496	0.788	0.976			
		IM SE	0.464	0.036	0.425	0.011	0.045			
		MC Sd	0.202	0.030	4.783	0.061	0.256			
30%	T	MC Mean	2.290	0.974	1.998	0.796	1.055	-1804.398	3618.796	3644.246
		IM SE	0.158	0.015	0.176	0.027	0.124			
		MC Sd	0.134	0.012	0.169	0.022	0.115			
	SL	MC Mean	2.262	0.976	1.421	0.796	1.056			
		IM SE	0.165	0.016	0.100	0.026	0.119			
		MC Sd	0.128	0.012	0.106	0.022	0.113			
	N	MC Mean	1.394	1.071	16.922	0.774	0.928			
		IM SE	0.538	0.039	0.486	0.013	0.051			
		MC Sd	0.274	0.035	4.975	0.063	0.250			

the conditional expectation $E[\mathbf{y}_{cens} | \mathbf{y}_{obs}]$ of the censored values is computed via Equation (22). Figure 3 shows the plot of the imputed values $E[\mathbf{y}_{cens} | \mathbf{y}_{obs}]$ as a function of the true censored (simulated) values \mathbf{y} . As expected, the SAEM algorithm provides a satisfactory imputation for these censored values when heavy-tailed distributions are considered.

Finally, Figures 6, 7 and 8, given in Appendix C, show the convergence of the SAEM algorithm for all the parameters and SMN distributions for this simulated dataset. Observing these figures, the estimates converge swiftly to a neighborhood of the ML estimates during the first 75 iterations for all models. The next few iterations ensure the almost sure convergence of the sequence to these estimates.

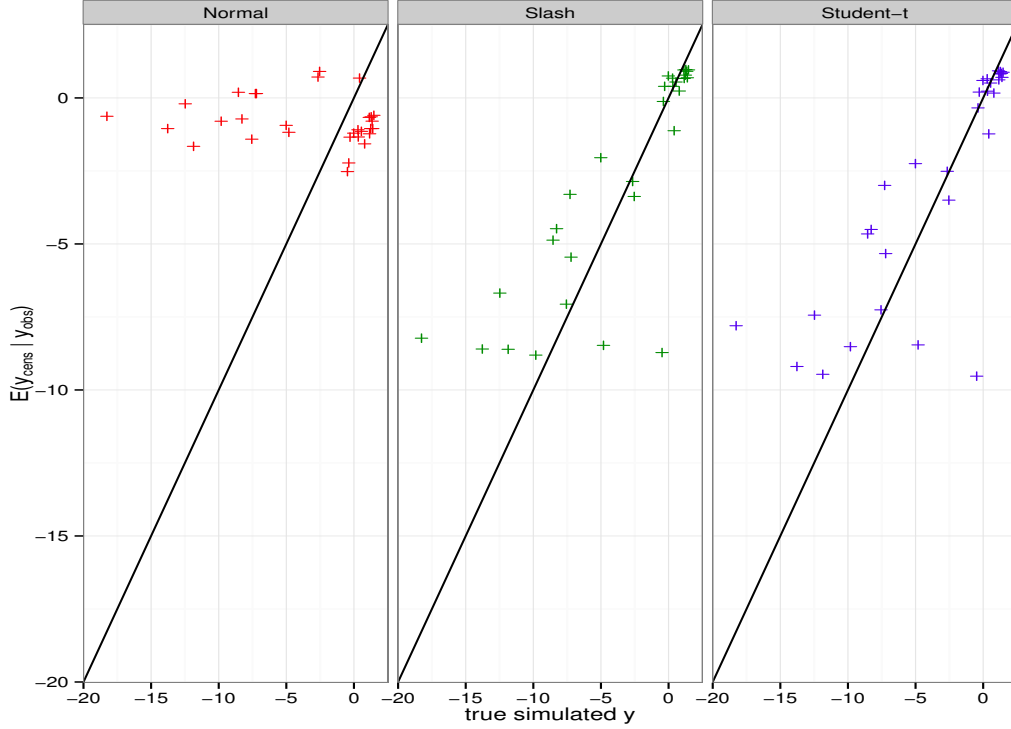


Figure 3: Simulation study - **Scenario 3**. Conditional expectation of the censored values ($E[y_{cens} | y_{obs}]$) evaluated by the SAEM algorithm as a function of the true censored simulated values y .

5.2 Real Data - UTI Data

The application considered in this section is referred to a study of 72 perinatally HIV-infected children (Saitoh et al. 2008). This dataset is available in the R package (R Development Core Team 2015) through the library *lme4*. All subjects in the study had received ARV therapy for at least 6 months before the interruption, and the medication was discontinued for more than 3 months. Out of 362 observations, 26 were below the detection limits (50 or 400 copies/mL) and considered left-censored at those values.

We consider the SMN-CR models with DEC structure $\Sigma_i = \sigma^2 E_i$ defined in Subsection 3.1 to fit this dataset. Here, y_{ij} is the \log_{10} HIV 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 $t_8 = 24$.

For the Student's- t , slash and contaminated normal models, the degrees of freedom ν are assumed to be unknown but fixed. According to the AIC (or BIC) values, the appropriate values of ν vary under different types correlation structures. Observing Table 3, the CN-CR model with $\nu = (0.2, 0.1)$ outperforms all other SMN competitors. Moreover, for these models, the estimated values of ν are fairly small, indicating a lack of adequacy of the normal assumption for the UTI data.

Table 4 reports the ML estimates and standard errors for the model parameters from the four fitted SMN models. Note that the estimates of β_1 , β_2 , and β_3 (the slope parameters corresponding to time points 0, 1, and 3 months) for the SMN models are quite close to each other and those for the time points further away, *i.e.*, $\beta_4 \dots, \beta_8$, are also reasonably close to each other. The standard error estimates of β are smaller than those in the normal model, indicating that the three heavy-tailed models are capable of producing more precise estimates. The variance components are not comparable since they are on different scales. The regression coefficients β_j , for $j = 1, \dots, 8$, increase gradually under these models. This gives the evidence of the negative effect of the antiretroviral therapy interruption 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 our best model (CN-CR), the convergence of the estimates obtained through the SAEM algorithm are shown in Figures 9 and 10 (Appendix C). As can be seen, the convergence can be achieved very quickly.

Table 3: **UTI data.** Information criteria for the SMN-CR models under DEC structure.

Distribution	Criteria	Structure				
		DEC	AR(1)	MA(1)	SYM	UNC
T	ℓ_{max}	-363.08	-406.98	-468.31	-364.21	-473.92
	AIC	748.15	833.96	956.62	748.43	965.84
	BIC	790.96	872.87	995.53	787.34	1000.86
	ν	2.3	2.1	2.1	2.3	2.1
SL	ℓ_{max}	-359.72	-403.08	-470.46	-360.90	-476.12
	AIC	741.44	826.15	960.92	741.79	970.24
	BIC	784.25	865.07	999.84	780.71	1005.26
	ν	0.8	0.7	1.0	0.8	1.0
CN	ℓ_{max}	-351.32	-396.56	-481.87	-353.37	-487.92
	AIC	724.64	813.12	983.74	726.75	993.83
	BIC	767.44	852.04	1022.66	765.66	1028.86
	ν	(0.2,0.1)	(0.3,0.1)	(0.1,0.1)	(0.2,0.1)	(0.1,0.1)
N	ℓ_{max}	-411.93	-463.05	-516.52	-412.06	-524.17
	AIC	845.87	946.11	1053.03	844.11	1066.34
	BIC	888.68	985.02	1091.95	883.03	1101.37
	ν	-	-	-	-	-

Now, we are interested in investigating the performance of the prediction for future values described in Section 4. Toward this, we compare the predicted values under the four fitted models, say, T-CR, SL-CR, CN-CR and N-CR with DEC structure. In order to do that, we exclude the last two measurements of each individual in the datasets with more than 6 observations (total of 29 individuals). To evaluate the predictive accuracy, we compute the mean absolute error (MAE) and the mean square error (MSE), defined as

$$MAE = \frac{1}{58} \sum_{i,j} |y_{ij} - y_{ij}^*| \quad \text{and} \quad MSE = \frac{1}{58} \sum_{i,j} (y_{ij} - y_{ij}^*)^2, \quad (28)$$

where y_{ij} is the original value and y_{ij}^* is the predicted value, for $i = 1, \dots, 29$ and $j = 1, 2$. Table 5 shows the comparison between the predicted values and real ones under the SMN-CR models. We can see from these results that the CN-CR model outperforms its competitors.

Finally, for the CN-CR model (our best model), we present in Figure 4 a comparison between the predicted values and the real ones considering five different correlation structures, say, AR(1), MA(1), SYM, UNC, and DEC (without fixing parameters ϕ_1 and ϕ_2). From this figure we can see that the CN-CR model with DEC structure has a better performance in terms of prediction than the other ones.

Table 4: **UTI data.** ML estimates with standard errors for the SMN-CR models under DEC structure.

Parameter	T		SL		CN		N	
	Estimative	SE	Estimative	SE	Estimative	SE	Estimative	SE
β_1	4.040	(0.096)	4.020	(0.096)	3.993	(0.097)	3.625	(0.136)
β_2	4.321	(0.107)	4.312	(0.107)	4.303	(0.111)	4.185	(0.178)
β_3	4.354	(0.111)	4.344	(0.115)	4.332	(0.119)	4.259	(0.212)
β_4	4.533	(0.115)	4.498	(0.117)	4.487	(0.119)	4.375	(0.201)
β_5	4.675	(0.130)	4.649	(0.129)	4.638	(0.122)	4.579	(0.223)
β_6	4.670	(0.147)	4.646	(0.141)	4.623	(0.139)	4.582	(0.243)
β_7	4.688	(0.136)	4.670	(0.140)	4.657	(0.152)	4.688	(0.218)
β_8	4.871	(0.183)	4.842	(0.189)	4.791	(0.206)	4.806	(0.378)
σ^2	0.544	(0.139)	0.282	(0.065)	0.543	(0.100)	1.090	(0.134)
ϕ_1	0.812	(0.040)	0.820	(0.038)	0.823	(0.038)	0.700	(0.043)
ϕ_2	0.094	(0.083)	0.096	(0.082)	0.121	(0.085)	0.028	(0.071)

Table 5: **UTI data.** Evaluation of the prediction accuracy for the SMN-CR models under DEC correlation structure.

	T	SL	CN	N
MSE	0.219	0.227	0.197	0.240
MAE	0.357	0.361	0.340	0.383

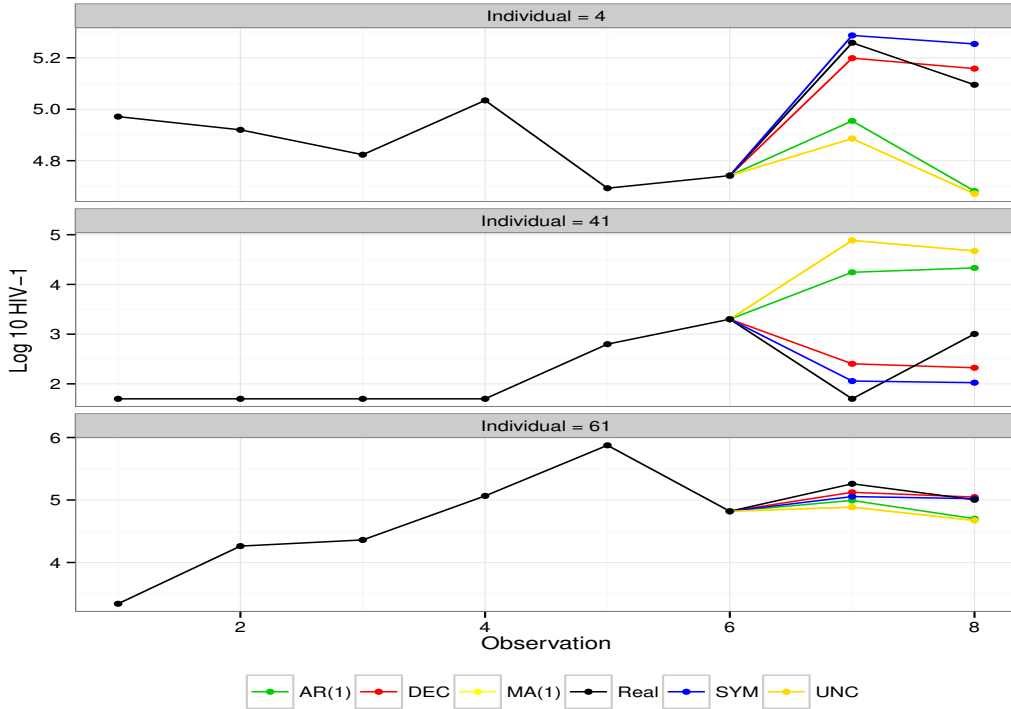


Figure 4: **UTI data.** Evaluation of the prediction performance for three random subjects, considering the CN-CRM model under different correlation structures.

6 Conclusions

In this paper, we have introduced a robust multivariate censored regression model for longitudinal data under the SMN class of distributions, extending the recent work by Garay et al. (2014) and Garay et al. (2015) to a multivariate context. For modeling the autocorrelation existing among irregularly observed measures, a damped exponential correlation structure was adopted as proposed by Muñoz et al. (1992). The main advantage of the proposed SMN-CR model is that it can reduce the negative impact of distributional misspecification and outliers in the parameters estimation. Moreover, the SMN class admits a convenient framework for the implementation of the SAEM algorithm, leading to an efficient ML estimation of the parameters. The computing code is available from the authors upon request.

We applied our methods to an AIDS study (freely downloadable from \mathbb{R}). Numerical results signify that when the antiretroviral therapy is interrupted, the HIV-1 RNA levels in blood increase consistently along the period of evaluation. We also undertake a simulation study to demonstrate the superiority of SMN-CR model on the provision of more adequate results when the available data have censored components. Furthermore, the simulation results reveal that our method gives very competitive performance in terms of imputation when the DEC structure is imposed. Therefore, it is noteworthy to mention that the use of the SMN-CR model with DEC structure can offer a better fit, protection against outliers, and more precise inferences.

Future extensions of the work include the use of scale mixtures of skew-normal distributions (Lachos et al. 2010) to accommodate both skewness and heavy-tailed feature, or the development of some

diagnostics and tests for the model adequacy. Incorporating measurement error models within our robust framework for related HIV viral load covariates (namely, CD4 cell counts) is also part of our future research.

Acknowledgements

L. Matos acknowledges support from FAPESP-Brazil (Grant 2011/22063-9 and Grant 2015/05385-3). The research of V. H. Lachos was partially supported by CNPq-Brazil (Grant 305054/2011-2) and FAPESP-Brazil (Grant 2014/02938-9). L.M. Castro acknowledges support from Grant FONDECYT 1130233 from the Chilean government, INNÓVATE-PERÚ (Grant ECIP-1-P-043-14) from Peruvian government and CONICYT-Chile through BASAL project CMM, Universidad de Chile. T.I. Lin acknowledges support from the Ministry of Science and Technology of Taiwan (Grant MOST 103-2118-M-005-001-MY2)

Appendix

Appendix A. Proof of Corollary 1

Proof. Let $\kappa(u_i) = \kappa(u)$, $p_1 = n_i^o$, $p_2 = n_i^c$, $\mathbf{y}_1 = \mathbf{y}_i^o$, $\mathbf{y}_2 = \mathbf{V}_i^c$, $\boldsymbol{\mu}_1 = \mathbf{X}_i^o \boldsymbol{\beta}$, $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_i$, $\boldsymbol{\Sigma}_1 = \boldsymbol{\Omega}_i^{oo}$ and $\boldsymbol{\Sigma}_2 = \mathbf{S}_i$. The likelihood contributed by subject i is given by

$$\begin{aligned} L_i(\boldsymbol{\theta}) &= \int_0^\infty \phi_{n_i^o}(\mathbf{y}_i^o; \mathbf{X}_i^o \boldsymbol{\beta}, \kappa(u_i) \boldsymbol{\Omega}_i^{oo}) \Phi_{n_i^c}(\mathbf{V}_i^c; \boldsymbol{\mu}_i, \kappa(u_i) \mathbf{S}_i) dH(u) \\ &= \int_0^\infty \phi_{p_1}(\mathbf{y}_1; \boldsymbol{\mu}_1, \kappa(u) \boldsymbol{\Sigma}_1) \Phi_{p_2}(\mathbf{y}_2; \boldsymbol{\mu}_2, \kappa(u) \boldsymbol{\Sigma}_2) dH(u). \end{aligned}$$

(a) For the *multivariate normal* distribution:

The proof is straightforward since U is degenerated in 1.

(b) For the *multivariate Student- t* distribution:

$$\begin{aligned} L(\boldsymbol{\theta}) &= \int_0^\infty \frac{1}{\sqrt{(2\pi)^{p_1} |\frac{1}{u} \boldsymbol{\Sigma}_1|}} \exp \left\{ -\frac{u}{2} (\mathbf{y}_1 - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_1^{-1} (\mathbf{y}_1 - \boldsymbol{\mu}_1) \right\} \\ &\quad \Phi_{p_2} \left(\mathbf{y}_2; \boldsymbol{\mu}_2, \frac{\boldsymbol{\Sigma}_2}{u} \right) \frac{(\frac{\nu}{2})^{\frac{\nu}{2}} u^{\frac{\nu}{2}-1}}{\Gamma(\frac{\nu}{2})} \exp \left\{ -\frac{\nu}{2} u \right\} du. \end{aligned}$$

Let $d(\mathbf{y}_1) = (\mathbf{y}_1 - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_1^{-1} (\mathbf{y}_1 - \boldsymbol{\mu}_1)$. After some algebraic manipulations, we can deduce that

$$\begin{aligned} L(\boldsymbol{\theta}) &= t_{p_1}(\mathbf{y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu) \int_0^\infty \left(\nu + \frac{d(\mathbf{y}_1)}{2} \right)^{\frac{(p_1+\nu)}{2}} \frac{1}{\Gamma(\frac{p_1+\nu}{2})} \\ &\quad \exp \left\{ -\frac{u}{2} (d(\mathbf{y}_1) + \nu) \right\} u^{\frac{(p_1+\nu)}{2}-1} \Phi_{p_2} \left(\mathbf{y}_2; \boldsymbol{\mu}_2, \frac{\boldsymbol{\Sigma}_2}{u} \right) du \\ &= t_{p_1}(\mathbf{y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu) \int_0^\infty f(u) \Phi_{p_2} \left(\mathbf{y}_2; \boldsymbol{\mu}_2, \frac{\boldsymbol{\Sigma}_2}{u} \right) du \\ &\quad \left(U \sim \text{Gamma} \left(\frac{p_1 + \nu}{2}, \frac{d(\mathbf{y}_1) + \nu}{2} \right) \right) \\ &= t_{p_1}(\mathbf{y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu) \int_0^\infty f(u) \Phi_{p_2} \left(\sqrt{U} \boldsymbol{\Sigma}_2^{-1/2} (\mathbf{y}_2 - \boldsymbol{\mu}_2); \mathbf{0}, \mathbf{I}_{p_2} \right) du \\ &= t_{p_1}(\mathbf{y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu) E_U \left\{ \Phi_{p_2} \left(\sqrt{U} \boldsymbol{\Sigma}_2^{-1/2} (\mathbf{y}_2 - \boldsymbol{\mu}_2) \right) \right\}. \end{aligned}$$

It follows from Lemma 1 of Prates et al. (2014) that

$$\begin{aligned} L(\boldsymbol{\theta}) &= t_{p_1}(\mathbf{y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu) T_{p_2} \left(\sqrt{\frac{d(\mathbf{y}_1) + \nu}{p_1 + \nu}} \boldsymbol{\Sigma}_2^{-1/2} (\mathbf{y}_2 - \boldsymbol{\mu}_2) \middle| \mathbf{0}, \mathbf{I}_{p_2}, p_1 + \nu \right) \\ &= t_{p_1}(\mathbf{y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu) T_{p_2} \left(\mathbf{y}_2 \middle| \boldsymbol{\mu}_2, \frac{d(\mathbf{y}_1) + \nu}{p_1 + \nu} \boldsymbol{\Sigma}_2, p_1 + \nu \right). \end{aligned}$$

(c) For the *multivariate contaminated normal* distribution:

$$\begin{aligned}
L_i(\boldsymbol{\theta}) &= \int_0^\infty \phi_{p_1}(\mathbf{y}_1; \boldsymbol{\mu}_1, \kappa(u)\boldsymbol{\Sigma}_1)\Phi_{p_2}(\mathbf{y}_2; \boldsymbol{\mu}_2, \kappa(u)\boldsymbol{\Sigma}_2)dH(u) \\
&= \nu \left[\frac{1}{\sqrt{(2\pi)^{p_1}|\frac{1}{\gamma}\boldsymbol{\Sigma}_1|}} \exp\left\{-\frac{1}{2}(\mathbf{y}_1 - \boldsymbol{\mu}_1)^\top \left(\frac{\boldsymbol{\Sigma}_1^{-1}}{\gamma}\right)(\mathbf{y}_1 - \boldsymbol{\mu}_1)\right\} \Phi_{p_2}\left(\mathbf{y}_2; \boldsymbol{\mu}_2, \frac{\boldsymbol{\Sigma}_2}{\gamma}\right) \right] \\
&+ (1 - \nu) \left[\frac{1}{\sqrt{(2\pi)^{p_1}|\boldsymbol{\Sigma}_1|}} \exp\left\{-\frac{1}{2}(\mathbf{y}_1 - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_1^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)\right\} \Phi_{p_2}(\mathbf{y}_2; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) \right] \\
&= \nu [\phi_{p_1}(\mathbf{y}_1; \boldsymbol{\mu}_1, \gamma^{-1}\boldsymbol{\Sigma}_1)\Phi_{p_2}(\mathbf{y}_2; \boldsymbol{\mu}_2, \gamma^{-1}\boldsymbol{\Sigma}_2)] \\
&+ (1 - \nu) [\phi_{p_1}(\mathbf{y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)\Phi_{p_2}(\mathbf{y}_2; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)].
\end{aligned}$$

Appendix C. Complementary results of simulation study

C1. Scenario 1: Absolute bias of parameter estimates in the SMN-CR model

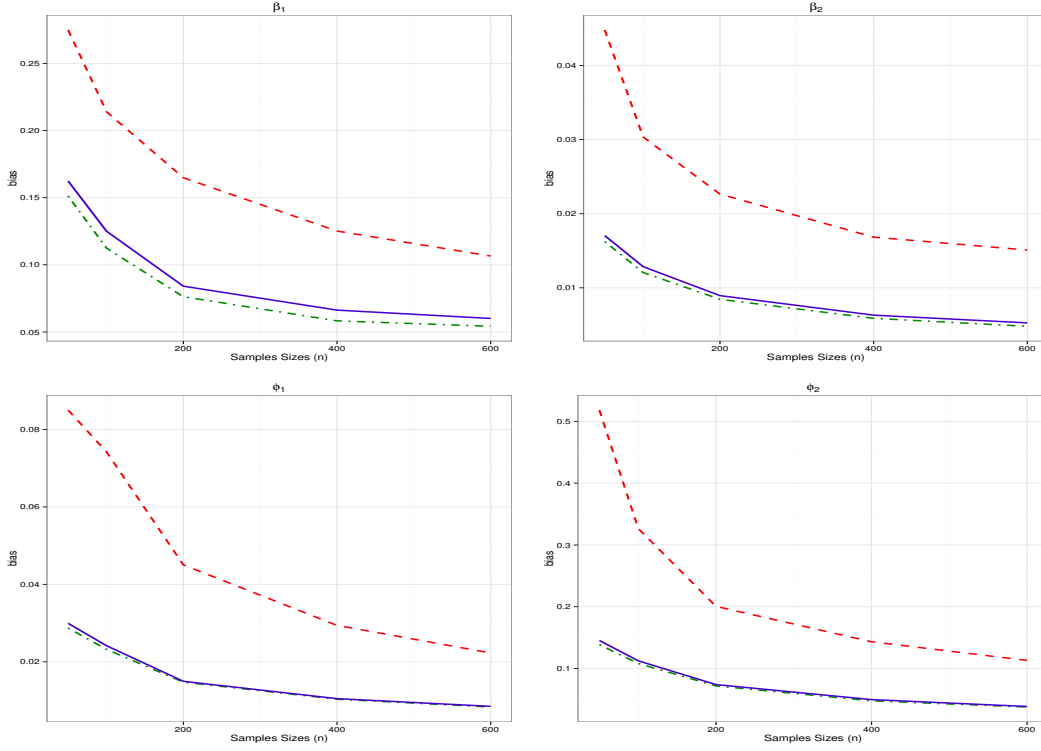


Figure 5: Simulation study - **Scenario 1**. Absolute bias of the parameter estimates in the SMN-CR model under 10% of censoring and different samples sizes. The solid line (blue) represents the T-CR model, the dotted line (red) represents the N-CR model and the dotted line (green) represents the SL-CR model.

C2. Scenario 3: Convergence of the parameters estimates

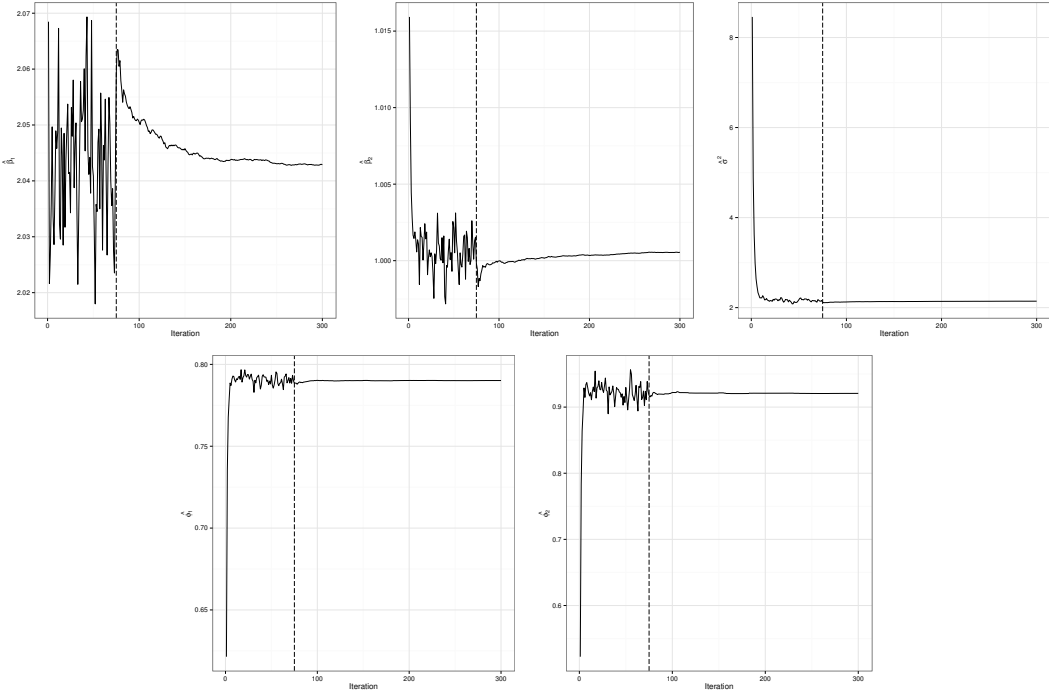


Figure 6: Simulation study - **Scenario 3**. Convergence of the SAEM parameters estimates for the T-CR model.

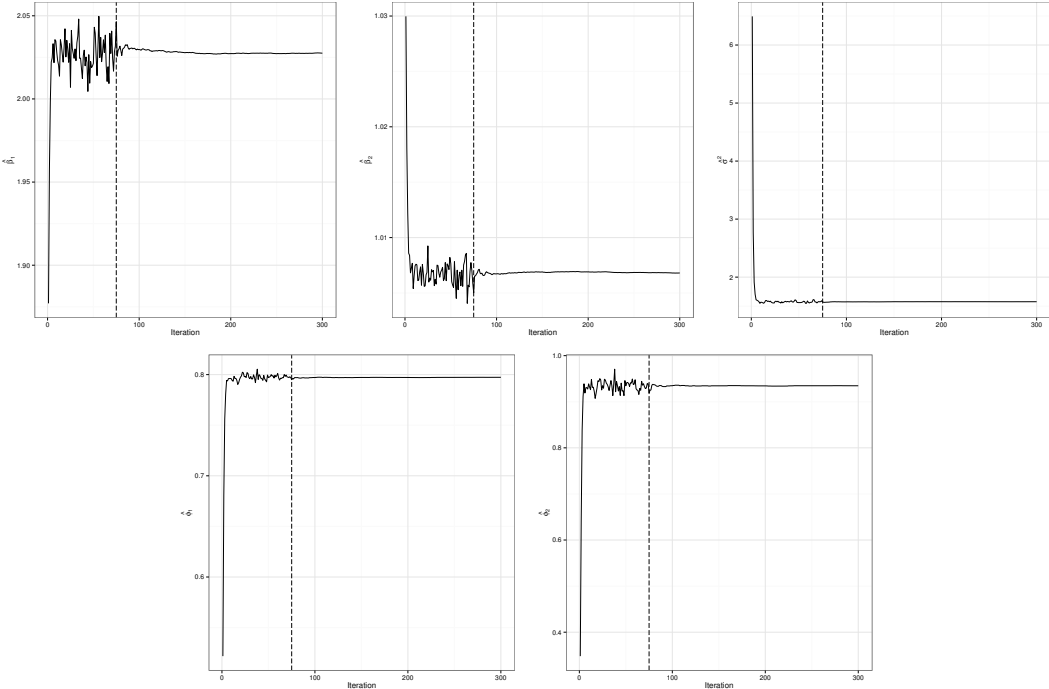


Figure 7: Simulation study - **Scenario 3**. Convergence of the parameters estimates for the SL-CR model.

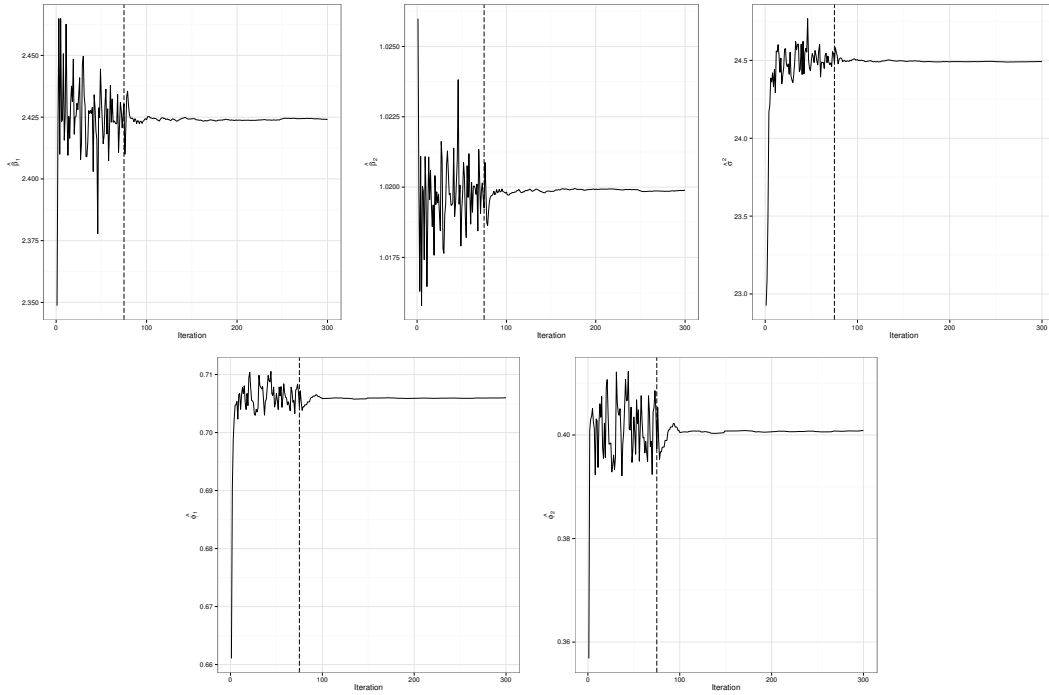


Figure 8: Simulation study - **Scenario 3**. Convergence of the SAEM parameters estimates for the N-CR model.

Appendix D. Complementary results of the application: convergence of the parameters estimates

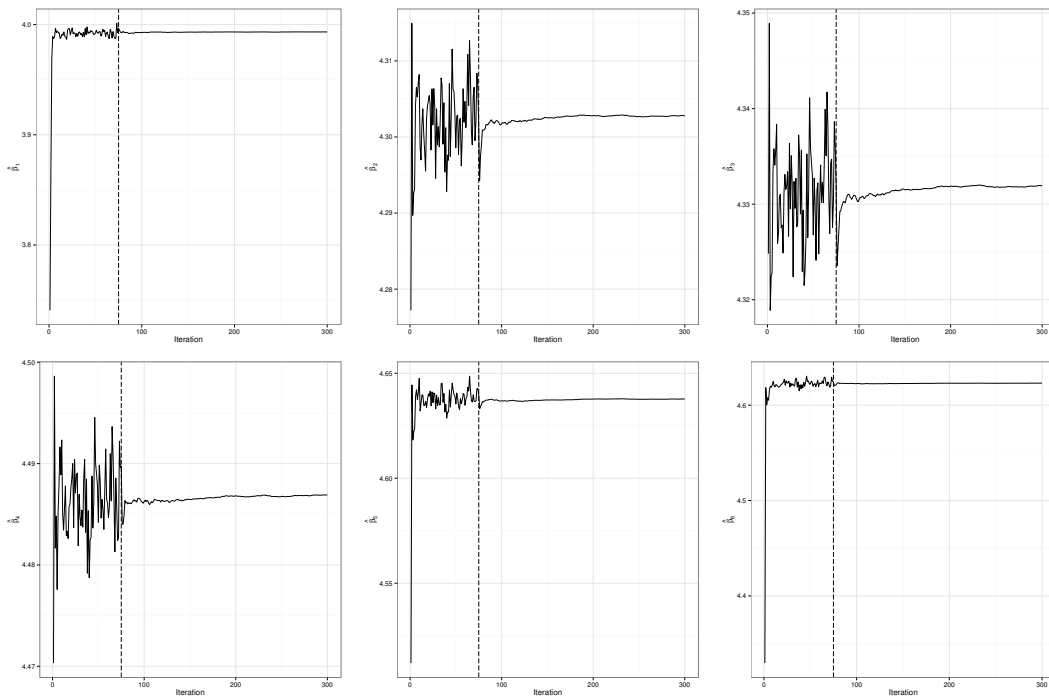


Figure 9: **UTI data**. Convergence of the SAEM parameters estimates.

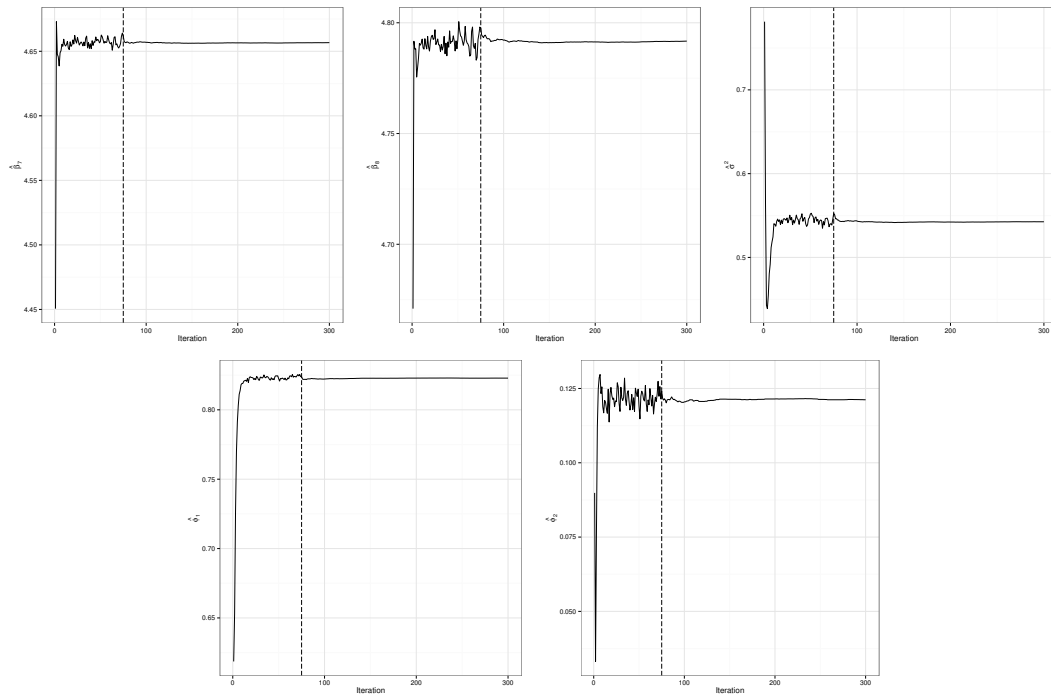


Figure 10: UTI data. Convergence of the SAEM parameters estimates (cont.).

References

- Andrews, D. F. and C. L. Mallows (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B* 36(1), 99–102.
- Antunes, R., S. Figueiredo, I. Bártolo, M. Pinheiro, L. Rosado, I. Soares, H. Lourenço, and N. Taveira (2003). Evaluation of the clinical sensitivities of three viral load assays with plasma samples from a pediatric population predominantly infected with human immunodeficiency virus type 1 subtype G and BG recombinants forms. *Journal of Clinical Microbiology* 41(7), 3361–3367.
- Arellano-Valle, R. B., L. M. Castro, G. González-Farías, and K. A. Muñoz-Gajardo (2012). Student-t censored regression model: properties and inference. *Statistical Methods & Applications* 21(4), 453–473.
- Delyon, B., M. Lavielle, and E. Moulines (1999). Convergence of a stochastic approximation version of the EM algorithm. *Annals of Statistics*, 94–128.
- Dempster, A., N. Laird, and D. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1–38.
- Galarza, C. E., D. Bandyopadhyay, and V. H. Lachos (2015). Quantile regression in linear mixed models: A stochastic approximation EM approach. *Submitted*.
- Garay, A. M., L. M. Castro, J. Leskow, and V. H. Lachos (2014). Censored linear regression models for irregularly observed longitudinal data using the multivariate-t distribution. *Statistical Methods in Medical Research*, DOI: 10.1177/0962280214551191.
- Garay, A. M., V. H. Lachos, H. Bolfarine, and C. R. Cabral (2015). Linear censored regression models with scale mixtures of normal distributions. *Statistical Papers* DOI: 10.1007/s00362-015-0696-9.
- Jacqmin-Gadda, H., R. Thiebaut, G. Chene, and D. Commenges (2000). Analysis of left-censored longitudinal data with application to viral load in HIV infection. *Biostatistics* 1(4), 355.

- Jank, W. (2006). Implementing and diagnosing the stochastic approximation EM algorithm. *Journal of Computational and Graphical Statistics* 15(4), 803–829.
- Kuhn, E. and M. Lavielle (2004). Coupling a stochastic approximation version of EM with an MCMC procedure. *ESAIM: Probability and Statistics* 8, 115–131.
- Kuhn, E. and M. Lavielle (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics & Data Analysis* 49(4), 1020–1038.
- Lachos, V., F. Labra, H. Bolfarine, and P. Ghosh (2010). Multivariate measurement error models based on scale mixtures of the skew-normal distribution. *Statistics* 44(6), 541–556.
- Lachos, V. H., D. Bandyopadhyay, and D. K. Dey (2011). Linear and nonlinear mixed-effects models for censored hiv viral loads using normal/independent distributions. *Biometrics* 67, 1594–1604.
- Lange, K., R. Little, and J. Taylor (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, 881–896.
- Lange, K. L. and J. S. Sinsheimer (1993). Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics* 2, 175–198.
- Lavielle, M. and C. Mbogning (2014). An improved SAEM algorithm for maximum likelihood estimation in mixtures of non linear mixed effects models. *Statistics and Computing* 24(5), 693–707.
- Lin, T. and J. Lee (2007). Bayesian analysis of hierarchical linear mixed modeling using the multivariate t distribution. *Journal of Statistical Planning and Inference* 137(2), 484–495.
- Lin, T.-I. (2010). Robust mixture modeling using multivariate skew t distributions. *Statistics and Computing* 20(3), 343–356.
- Lin, T. I. and J. C. Lee (2006). A robust approach to t linear mixed models applied to multiple sclerosis data. *Statistics in Medicine* 25(8), 1397–1412.
- Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 226–233.
- Lucas, A. (1997). Robustness of the student t based M-estimator. *Communications in Statistics-Theory and Methods* 26(5), 1165–1182.
- Massuia, M. B., C. R. B. Cabral, L. A. Matos, and V. H. Lachos (2015). Influence diagnostics for Student-t censored linear regression models. *Statistics* 49, 1074–1094.
- Matos, L., M. Prates, M.-H. Chen, and V. Lachos (2013). Likelihood based inference for linear and nonlinear mixed-effects models with censored response using the multivariate-t distribution. *Statistica Sinica* 23, 1323–1345.
- Meza, C., F. Osorio, and R. De la Cruz (2012). Estimation in nonlinear mixed-effects models using heavy-tailed distributions. *Statistics and Computing* 22(1), 121–139.
- Muñoz, A., V. Carey, J. P. Schouten, M. Segal, and B. Rosner (1992). A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics* 48, 733–742.
- Pinheiro, J. C., C. H. Liu, and Y. N. Wu (2001). Efficient algorithms for robust estimation in linear mixed-effects models using a multivariate t-distribution. *Journal of Computational and Graphical Statistics* 10, 249–276.
- Prates, M. O., D. R. Costa, and V. H. Lachos (2014). Generalized linear mixed models for correlated binary data with t-link. *Statistics and Computing* 24(6), 1111–1123.

- R Development Core Team (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Rosa, G., C. Padovani, and D. Gianola (2003). Robust linear mixed models with normal/independent distributions and Bayesian MCMC implementation. *Biometrical Journal* 45(5), 573–590.
- Saitoh, A., M. Foca, R. Viani, S. Heffernan-Vacca, F. Vaida, J. Lujan-Zilbermann, P. Emmanuel, J. Deville, and S. Spector (2008). Clinical outcomes after an unstructured treatment interruption in children and adolescents with perinatally acquired HIV infection. *Pediatrics* 121(3), e513.
- Samson, A., M. Lavielle, and F. Mentré (2006). Extension of the SAEM algorithm to left-censored data in nonlinear mixed-effects model: application to HIV dynamics model. *Computational Statistics & Data Analysis* 51(3), 1562–1574.
- Swenson, L., B. Cobb, A. Geretti, P. Harrigan, M. Poljak, C. Seguin-Devaux, C. Verhofstede, M. Wirden, A. Amendola, J. Boni, T. Bourlet, J. Huder, J. Karasi, S. Lepej, M. Lunar, O. Mukabayire, R. Schuurman, J. Tomažič, K. Van Laethem, L. Vandekerckhove, and A. Wensing (2014). Comparative performances of HIV-1 RNA load assays at low viral load levels: results of an international collaboration. *Journal of Clinical Microbiology* 52(2), 517–523.
- Vaida, F., A. Fitzgerald, and V. DeGruttola (2007). Efficient hybrid EM for linear and nonlinear mixed effects models with censored response. *Computational Statistics & Data Analysis* 51(12), 5718–5730.
- Vaida, F. and L. Liu (2009). Fast implementation for normal mixed effects models with censored response. *Journal of Computational and Graphical Statistics* 18(4), 797–817.
- Wang, W.-L. (2013). Multivariate t linear mixed models for irregularly observed multiple repeated measures with missing outcomes. *Biometrical Journal* 55(4), 554–571.
- Wei, G. C. and M. A. Tanner (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association* 85(411), 699–704.
- Wu, L. (2010). *Mixed Effects Models for Complex Data*. Boca Raton, FL: Chapman & Hall/CRC.