UNIVERSIDADE ESTADUAL DE
CAMPINAS

Instituto de Matemática, Estatística e
Computação Científica

LARISSA AVILA MATOS

Estimation and diagnostics in multivariate models for censored data

Estimação e diagnóstico em modelos multivariados para dados censurados

Campinas

2016

Larissa Avila Matos

Estimation and diagnostics in multivariate models for censored data

Estimação e diagnóstico em modelos multivariados para dados censurados

Tese apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor(a) em Estatística.

e

Thesis presented to the Institute of Mathematics, Statistics and Scientific Computing of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Statistics.

Orientador: Víctor Hugo Lachos Dávila

Coorientador: Luis Mauricio Castro Cepero

Este exemplar corresponde à versão final da Tese defendida pela aluna Larissa Avila Matos e orientada pelo Prof. Dr. Víctor Hugo Lachos Dávila.

Campinas

2016

Agência(s) de fomento e nº(s) de processo(s): FAPESP, 2011/22063-9; FAPESP, 2015/05385-3

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Maria Fabiana Bezerra Muller - CRB 8/6162

Matos, Larissa Avila, 1987-
M428e Estimation and diagnostics in multivariate models for censored data / Larissa Avila Matos. – Campinas, SP : [s.n.], 2016.

Orientador: Víctor Hugo Lachos Dávila.
Coorientador: Luis Mauricio Castro Cepero.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Análise multivariada. 2. Modelos mistos. 3. Algoritmos de esperança-maximização. 4. Misturas de escala. I. Lachos Dávila, Víctor Hugo,1973-. II. Castro Cepero, Luis Mauricio. III. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Estimação e diagnóstico em modelos multivariados para dados censurados
Palavras-chave em inglês:
Multivariate analysis
Mixed models
Expectation-maximization algorithms
Scale mixtures
Área de concentração: Estatística
Titulação: Doutora em Estatística
Banca examinadora:
Mariana Rodrigues Motta
Filidor Edilfonso Vilca Labra
Gilberto Alvarenga Paula
Clécio da Silva Ferreira
Data de defesa: 30-05-2016
Programa de Pós-Graduação: Estatística

Tese de Doutorado defendida em 30 de maio de 2016 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.

Prof(a). Dr(a). VÍCTOR HUGO LACHOS DÁVILA

Prof(a). Dr(a). MARIANA RODRIGUES MOTTA

Prof(a). Dr(a). FILIDOR EDILFONSO VILCA LABRA

Prof(a). Dr(a). GILBERTO ALVARENGA PAULA

Prof(a). Dr(a). CLÉCIO DA SILVA FERREIRA

A Ata da defesa com as respectivas assinaturas dos membros
encontra-se no processo de vida acadêmica do aluno.

This thesis is dedicated to my parents

Acknowledgements

This thesis would not have been possible without the help and support of many people.

First of all, I would like to thank God, for providing me this opportunity and by giving me strength and courage to do this work.

I would like to express my deepest gratitude to my advisor, Professor Victor Hugo Lachos Dávila, for his friendship, patience, crucial advices and suggestions. Without him it would be impossible to finish this thesis. I also would like to express my sincere thanks to Prof. Luis Mauricio Castro Cepero, my co-advisor, for his amity, valuable advices, help and corrections. My appreciation must also be dedicated to Prof. Dr. Ming-Hui Chen for his great support and generosity.

I would like to thank the financial support from FAPESP-Brazil (Grants 2011/22063-9 and 2015/05385-3).

Also, I would like to thank the members of the reading committee, Profa. Dra. Mariana Rodrigues Motta, Prof. Dr. Filidor Edilfonso Vilca Labra, Prof. Dr. Gilberto Alvarenga Paula and Prof. Dr. Clécio da Silva Ferreira for the critical and comprehensive examination of this thesis.

Furthermore, I would like to thank all the members and professors of IMEC-C/UNICAMP, for their help and support of my PhD work. I also apologize to everyone who may not be mentioned personally, here.

Finally, I would like to thank all of my family for their continued support, but I would like to send a special and kindly acknowledgement to my parents (Rita and Francisco) and my sister (Luana) for supporting me spiritually throughout writing this thesis and my life in general. To all my friends, thank you for your understanding and encouragement in my many moments of crisis. Your friendship makes my life a wonderful experience. I cannot list all the names here, but you are always on my mind.

Last but not least, I especially thank Marcos for supporting me, standing by my side all the time. He has faith in me and in my intellect even when I didn't have faith in myself. Thank you!

Resumo

Em alguns ensaios clínicos da síndrome da imunodeficiência adquirida (AIDS), as medições dos ácidos ribonucleicos do vírus da imunodeficiência humana (HIV-1) são coletadas periodicamente ao longo do tempo e muitas vezes estão sujeitas a limites de detecção inferiores ou superiores, dependendo dos ensaios de quantificação que foram utilizados. Assim, estas respostas podem ser censuradas à esquerda ou à direita. Na prática, dados longitudinais provenientes de estudos de acompanhamento do HIV, podem ser modelados utilizando modelos lineares e não-lineares de efeitos mistos censurados e também modelos de regressão censurados com estruturas de correlação específicas sobre os erros. Uma complicação adicional surge quando duas ou mais variáveis respostas são coletadas de forma irregular e repetidamente em cada sujeito durante um certo período de tempo. Os modelos lineares multivariados de efeitos mistos com respostas censuradas são ferramentas bastante utilizadas para análise conjunta de mais de uma série de respostas de dados longitudinais. Nesta tese desenvolvemos métodos inferenciais para lidar com dados censurados com estrutura longitudinal sob uma perspectiva clássica. Como resultado, conclusões importantes foram obtidas a partir da análise dos modelos propostos.

Palavras-chave: Modelos de regressão. Modelos de efeitos mistos. Dados censurados. Algoritmo EM. Algoritmo SAEM. Distribuições de misturas de escala normal.

Abstract

In some acquired immunodeficiency syndrome (AIDS) clinical trials, the human immunodeficiency virus-1 ribonucleic acid measurements are collected irregularly over time and are often subject to some upper and lower detection limits, depending on the quantification assays. Hence, these responses are either left- or right-censored. In practice, longitudinal data coming from those follow-up studies can be modelled using censored linear and nonlinear mixed-effects models and also censored regression models with a specific correlation structures on the error terms. A complication arises when more than one series of responses are repeatedly collected on each subject at irregularly occasions over a period of time. The multivariate censored linear mixed model is a frequently used tool for a joint analysis of more than one series of longitudinal data. In this thesis we develop a series of essays in which different models and techniques to deal with censored data are applied. As result, we had several works to carry out censored data.

Keywords: Regression models. Mixed-effects models. Censored data. EM algorithm. SAEM algorithm. Scale mixtures of normal distributions.

List of Figures

List of Tables

List of abbreviations and acronyms

SMN	Scale mixtures of normal
ML	Maximum-likelihood
pdf	Probability density function
cdf	Cumulative distribution function
MCMC	Markov Chain Monte Carlo
DEC	Damping exponential correlation.
UNC	Uncorrelated
AR(1)	Continuous-time autoregressive of order 1
CS	Compound symmetric
LME	Linear mixed effects model
NLME	Nonlinear mixed effects model
LMEC	Linear mixed effects model with censored responses
NLMEC	Nonlinear mixed effects model with censored responses
MAE	Mean absolute error
MSE	Mean square error

Contents

Preface

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 have been the scope of many areas of the statistic. They have received significant attention in the biomedical literature in recent years.

Particularly, we are interest 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. 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 censored regression models with a specific correlation structure on the error term (Garay et al., 2014).

The statistical modelling of viral load can be challenging. First, as mentioned, depending on the diagnostic assays used, the viral load measures may be subjected to upper or lower detection limits (hence, left or right censored), *i.e.*, below and above in which they are not quantifiable (Wu, 2002). Under non-trivial censoring proportions, considering ad-hoc alternatives (Huang and Dagne, 2011) might lead to bias in fixed effects and variance components estimates. As alternatives to these crude imputation techniques, Vaida and Liu (2009) proposed expectation-maximization (EM) schemes for linear and nonlinear mixed-effects (LME/NLME) models with censored responses (henceforth LMEC/NLMEC). However, all these methods assume normality of the between-subject random effects and within-subject errors. Even though normality is mostly a reasonable model assumption, it may lack robustness in parameter estimation under departures from normality, namely, presence of heavy tails and outliers (Pinheiro et al., 2001). Secondy, censored HIV viral loads can exhibit heavy-tailed behaviour (Lachos et al., 2011). A variety of proposals (both classical and Bayesian) exist in this direction that uses the univariate or multivariate Student's-t distribution (Pinheiro et al., 2001; Lin and Lee, 2006, 2007) in the context of LME/NLME models. Some Bayesian propositions in the context of heavy-tailed LMEC/NLMEC models include Lachos et al. (2011) who advocated the use of the normal/independent density (Lange and Sinsheimer, 1993), while Bandyopadhyay et al. (2012, 2015) studied the LMEC model considering both skewness and heavy-tails. Very recently, Matos et al. (2013b) proposed a full maximum likelihood (ML) based inference using a computationally convenient exact ECM algorithm for the LMEC/NLMEC models using the multivariate Student's-t distribution (henceforth, the t-LMEC/NLMEC model).

Finally, it happens quite commonly that more than one series of responses are repeatedly measured on each subject across time. For analyzing the so-called multivariate longitudinal data, the multivariate linear mixed effect (MLME) model proposed by Shah et al. (1997) has become a widely used tool. Wang (2013) proposed the multivariate t linear mixed model (tMLME), which has been shown to be a robust approach to modeling multioutcome continuous repeated measures in the presence of outliers or heavy-tailed noises. In the context of censored responses, Wang et al. (2015) extended the tMLME to allow the analysis of multiple longitudinal censored outcomes and heavy-tails (tMLMEC), where an exact EM algorithm for maximum likelihood (ML) estimation is developed based on the mean and variance of a truncated multivariate Student-t distribution developed by Ho et al. (2012).

This thesis is devoted to a series of chapters that use different models and techniques to deal with censored data, in particular AIDS - studies. As a result, we had different works to carried out the censored data. The organization of the thesis is as follows:

Chapter 1: We provide some background material. We review some definitions and methodologies used throughout the thesis.

Chapter 2: We provide and presents a framework for fitting linear and nonlinear mixed-effects censored (LMEC/NLMEC) models with response variables recorded at irregular intervals. To address the serial correlation among the within-subject errors, a damped exponential correlation structure is considered in the random error and an EM-type algorithm is developed for computing the maximum likelihood estimates, obtaining as a byproduct the standard errors of the fixed effects and the likelihood value. The proposed methods are illustrated with simulations and the analysis of two real AIDS case studies.

Chapter 3: Recently Matos et al. (2013b) proposed an exact EM-type algorithm for linear and nonlinear mixed-effects censored (LMEC/NLMEC) models using a multivariate Student's-t distribution, with closed-form expressions at the E-step. In this work, we develop influence diagnostics for LMEC/NLMEC models using the multivariate Student's-t density, based on the conditional expectation of the complete data log-likelihood. This partially eliminates the complexity associated with the approach of Cook (1977, 1986) for censored mixed-effects models. The new methodology is illustrated via an application to a longitudinal HIV dataset. In addition, a simulation study explores the accuracy of the proposed measures in detecting possible influential observations for heavy-tailed censored data under different perturbation and censoring schemes.

Chapter 4: We propose a robust nonlinear censored regression model based on the scale mixtures of normal (SMN) distributions. To take into account the autocorrelation existing among irregularly observed measures, a damped exponential correlation structure is considered. A stochastic approximation of the EM (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 method is exemplified using both simulated and real data.

Chapter 5: The multivariate censored linear mixed model (MLEMC) is a frequently used tool for a joint analysis of more than one series of longitudinal data. Motivated by a concern of sensitivity to potential outliers or data with longer-than-normal tails and possible serial correlation, we develop a robust generalization of the MLMEC that is constructed by using the scale mixtures of normal (SMN) distributions. To take into account the autocorrelation existing among irregularly observed measures, a damped exponential correlation (DEC) structure is considered. For this complex longitudinal structure, we propose an exact estimation procedure to obtain the maximum likelihood estimates of the fixed effects and variance components, using a stochastic approximation of the EM (SAEM) algorithm. This approach allow us to estimate the parameters of interest in an easy and fast way, obtaining as a by-product the standard errors of the fixed effects, predictions of unobservable values of the response and the log-likelihood function. The methodology is illustrated through an application to a set of AIDS data and a small simulation study.

Chapter 6: We present final remarks and further researches related to this thesis.

The majority of the material in this thesis is based on original publications. Below, we give a list of the parts of the thesis based principally on those publications.

Chapter 2: Matos, L. A., L. M. Castro, and V. H. Lachos (2016). Censored mixed-effects models for irregularly observed repeated measures with applications to HIV viral loads. *Test*, DOI: 10.1007/s11749-016-0486-2.

Chapter 3: Matos, L. A., D. Bandyopadhyay, L. M. Castro, and V. H. Lachos (2015). Influence assessment in censored mixed-effects models using the multivariate student's-t distribution. *Journal of multivariate analysis* 141, 104–117.

Chapter 4: Matos, L. A., V. H. Lachos, T.-I Lin, and L. M. Castro. Heavy-tailed longitudinal regression models for censored data: A likelihood based perspective. (Submitted)

Chapter 5: Lachos, V. H., **L. A. Matos**, L. M. Castro, and M.-H. Chen. Heavy-tailed longitudinal linear mixed models for multiple censored responses data.

ANNEX: Matos, L. A., L. M. Castro, C. R. B. Cabral, and V. H. Lachos. Multivariate measurement error models based on student-t distribution under censored responses. (Submitted)

Part I

Background Material

1 An overview

In this chapter we present some background material. We review the scale mixtures of multivariate normal distributions and the damping exponential correlation (DEC) structure. We also describe the algorithms for ML estimation, the EM and SAEM algorithms, and how to compute the empirical information matrix. Finally, we describe the motivating datasets, which will be analyzed in the course of this work.

1.1 Preliminaries

1.1.1 Scale mixtures of normal distributions (SMN)

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

$$\mathbf{Y} = \boldsymbol{\mu} + \kappa(U)^{1/2}\mathbf{Z}, \tag{1.1}$$

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*) $\mathrm{H}(u \mid \boldsymbol{\nu})$ 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. 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

$$\mathrm{SMN}_p(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) = \int_0^\infty \phi_p(\mathbf{y}; \boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma}) d\mathrm{H}(u \mid \boldsymbol{\nu}),$$

where $\phi_p(.; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ 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 $\mathbf{Y} \sim \mathrm{SMN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu})$. Three members of the scale mixtures of normal class of distributions are commonly used for robust estimation:

- *The multivariate Student-t distribution*, $\mathrm{T}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, where ν is called the degrees of freedom, can be derived from the mixture model (1.1), arises when 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

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

  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. When $\nu \uparrow \infty$, the Student-t distribution reduces to the normal distribution as the limiting case.

- *The multivariate slash distribution*, $\mathrm{SL}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, 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

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

  The slash distribution reduces to the normal distribution when $\nu \uparrow \infty$.

- *The multivariate contaminated normal distribution*, $\mathrm{CN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu_1, \nu_2)$, where $\nu_1, \nu_2 \in (0, 1)$. Here, $\kappa(u) = 1/u$ and U is a discrete random variable taking one of two states and has pdf given by

$$h(u \mid \boldsymbol{\nu}) = \nu_1 \mathbb{I}_{\{\nu_2\}}(u) + (1 - \nu_1)\mathbb{I}_{\{1\}}(u),$$

  where $\boldsymbol{\nu} = (\nu_1, \nu_2)$ 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

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) = \nu_1 \phi_p(\mathbf{y}; \boldsymbol{\mu}, \nu_2^{-1}\boldsymbol{\Sigma}) + (1 - \nu_1)\phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

  The parameter ν_1 can be interpreted as the proportion of outliers while ν_2 may be interpreted as a scale factor (Osorio et al., 2007). In this case, the contaminated normal distribution reduces to the normal distribution when $\nu_2 = 1$.

Table 1 – Summary of some SMN distributions.

Distribution	$\kappa(u)$	U	Density function $f(\mathbf{y})$
$\mathrm{T}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$	$\frac{1}{u}$	$U \overset{\text{ind.}}{\sim} \mathrm{Gamma}(\nu/2, \nu/2)$, $u > 0, \nu > 0$	$\frac{\Gamma(\frac{p+\nu}{2})}{\Gamma(\frac{\nu}{2})\pi^{p/2}} \nu^{-p/2} \lvert\boldsymbol{\Sigma}\rvert^{-1/2} \left(1 + \frac{d}{\nu}\right)^{-(p+\nu)/2}$
$\mathrm{SL}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$	$\frac{1}{u}$	$U \overset{\text{ind.}}{\sim} \mathrm{Beta}(\nu, 1)$, $u \in (0, 1)$, $\nu > 0$	$\nu \int_0^1 u^{\nu-1} \phi_p(\mathbf{y}; \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}) du$
$\mathrm{CN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu_1, \nu_2)$	$\frac{1}{u}$	$h(u\lvert\boldsymbol{\nu})$, $\boldsymbol{\nu} = (\nu_1, \nu_2)$, $u = \nu_2, 1$ and $\nu_1, \nu_2 \in (0, 1)$	$\nu_1 \phi_p(\mathbf{y}; \boldsymbol{\mu}, \nu_2^{-1}\boldsymbol{\Sigma}) + (1 - \nu_1)\phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$

with $d = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ and $h(u|\boldsymbol{\nu}) = \nu_1 \mathbb{I}_{\{\nu_2\}}(u) + (1 - \nu_1)\mathbb{I}_{\{1\}}(u)$

In this work, we will focus on the distributions described on Table 1, where $\kappa(u) = 1/u$ for all distributions, then we have that

$$\mathbf{Y} = \boldsymbol{\mu} + U^{-1/2}\mathbf{Z}. \tag{1.2}$$

Remark: *The multivariate normal distribution*, $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, also belongs to the scale mixtures of normal distributions. In this case, the random variable U is degenerate in 1, *i.e.*, $P(U = 1) = 1$.

1.1.2 Damped exponential correlation structure (DEC)

Following Muñoz et al. (1992), the DEC (damped exponential correlation) structure is defined as:

$$\mathbf{E}_i = \mathbf{E}_i(\boldsymbol{\phi}, \mathbf{t}_i) = \left[\phi_1^{|t_{ij}-t_{ik}|^{\phi_2}}\right], \; i = 1, \ldots, n, \;\; j, k = 1, \ldots, n_i, \tag{1.3}$$

where $\mathbf{t}_i = (\mathbf{t}_{i1}, \ldots, \mathbf{t}_{in_i})$ is a vector of time points for subject i and $\boldsymbol{\phi} = (\phi_1, \phi_2)^\top$. 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:

1. if $\phi_2 = 0$, then $\mathbf{E}_i$ generates the compound symmetry correlation structure (CS);
2. 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;
3. if $\phi_2 = 1$, then $\mathbf{E}_i$ generates an AR(1) structure;
4. when $\phi_2 > 1$, $\mathbf{E}_i$ presents a decay rate faster than the AR(1) structure; and
5. if $\phi_2 \to \infty$, 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).

1.1.3 The EM/SAEM algorithm

The EM algorithm, introduced by (Dempster et al., 1977), is a 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, $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)$, until it reaches the convergence. Each iteration is performed as follows:

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

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

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 the SAEM algorithm. In this procedure, 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 algorithm at iteration k proceeds as follows:

E-Step:

1. Simulation-step:

   (a) draw $\mathbf{q}^{(k,l)}$ $(l = 1, \ldots, m)$ from the conditional distribution $f(\mathbf{q} \mid \mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k-1)})$, or

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

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

$$Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) = Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k-1)}) + \delta_k \left[\frac{1}{m} \sum_{l=1}^{m} \ell_c(\boldsymbol{\theta} \mid \mathbf{q}^{(k,l)}, \mathbf{y}) - Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k-1)}) \right], \quad (1.4)$$

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

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = \underset{\theta}{\text{argmax}} Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}).$$

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 work we will refer the MCMC-SAEM as SAEM algorithm. As proposed by Galarza et al. (2015) we will consider the following smoothing parameter

$$\delta_k = \begin{cases} 1, & \text{if} \quad 1 \leqslant k \leqslant cW; \\ \dfrac{1}{k - cW}, & \text{if} \quad cW + 1 \leqslant k \leqslant W, \end{cases} \tag{1.5}$$

where W is the maximum number of iterations and c is a cut point ($0 \leqslant c \leqslant 1$) which determines the percentage of the initial iterations. By Equation (1.4), 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 $m \leqslant 20$) 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 $(\mathbf{q}^{(k,l)})$, $l = 1, \ldots, m$ but some or all previous simulations, where this "memory" property is set by the smoothing parameter δ_k.

1.1.4 The empirical information matrix

To compute the asymptotic covariance of the ML estimates we follow Lin (2010). According to large sample theory, the asymptotic covariance matrix of the ML estimates can be approximated by (Meilijson, 1989)

$$\mathbf{I}_e(\boldsymbol{\theta} \mid \mathbf{y}) = \sum_{i=1}^{n} \mathbf{s}(\mathbf{y}_i \mid \boldsymbol{\theta}) \mathbf{s}^{\top}(\mathbf{y}_i \mid \boldsymbol{\theta}) - \frac{1}{n} \mathbf{S}(\mathbf{y} \mid \boldsymbol{\theta}) \mathbf{S}^{\top}(\mathbf{y} \mid \boldsymbol{\theta}), \tag{1.6}$$

where $\mathbf{S}(\mathbf{y} \mid \boldsymbol{\theta}) = \sum_{i=1}^{n} \mathbf{s}(\mathbf{y}_i \mid \boldsymbol{\theta})$ and $\mathbf{s}(\mathbf{y}_i \mid \boldsymbol{\theta})$ is the empirical score function for subject i. According to Louis (1982), it is possible to relate the score function of the incomplete data log-likelihood with the conditional expectation of the complete data log-likelihood function. Therefore, the individual score can be determined as

$$\mathbf{s}(\mathbf{y}_i \mid \boldsymbol{\theta}) = \frac{\partial \log f(\mathbf{y}_i \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = E\left(\frac{\partial \ell_i(\boldsymbol{\theta} \mid \mathbf{y}_c)}{\partial \boldsymbol{\theta}} \mid \mathbf{y}_i, \boldsymbol{\theta} \right), \tag{1.7}$$

where $\mathbf{y}_c$ is the vector of complete data and $\ell_i(\boldsymbol{\theta} \mid \mathbf{y}_c)$ is the complete data log-likelihood function formed from the i-th observation. Using the ML estimates $\widehat{\boldsymbol{\theta}}$, that is, $\mathbf{S}(\mathbf{y} \mid \widehat{\boldsymbol{\theta}}) = 0$,

it follows that (1.6) can be approximated by

$$\mathbf{I}_e(\widehat{\boldsymbol{\theta}} \mid \mathbf{y}) = \sum_{i=1}^{n} \widehat{\mathbf{s}}_i \widehat{\mathbf{s}}_i^{\top}, \tag{1.8}$$

where $\widehat{\mathbf{s}}_i = \mathbf{s}(\mathbf{y}_i \mid \widehat{\boldsymbol{\theta}})$.

1.2 Case studies

In this section we present the motivating datasets, which will be analysed in this thesis.

1.2.1 UTI data

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 (R Development Core Team, 2009) through the library *lmec*. 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 from the closest time points at 0, 1, 3, 6, 9, 12, 18, 24 months after UTI were studied. 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 1.

1.2.2 ACTG 315 data

The ACTG 315 protocol considers 46 HIV-1 infected patients treated with a potent antiretroviral drug cocktail based on the protease inhibitor ritonavir and two reverse transcriptase inhibitor drugs (zidovudine and lamivudine). Before initiating the antiretroviral therapy, all patients discontinued their own antiretroviral regimen for five weeks as a “washout" period. The aim of this antiretroviral regimen is to show that immunity can be partially restored in people with moderately advanced HIV disease.

The viral load was quantified on days 0, 2, 7, 10, 14, 21, 28, 56, 84, 168 and 196 after starting treatment. The dataset includes 361 observations. An immunologic marker known as CD4+ cell count was also measured along with viral load and 72 out of 361 (20%) CD4 values were missing due to a mismatch of the CD4 and the viral load measurement schedules. The number of measurements per subject varied from 4 to 10.

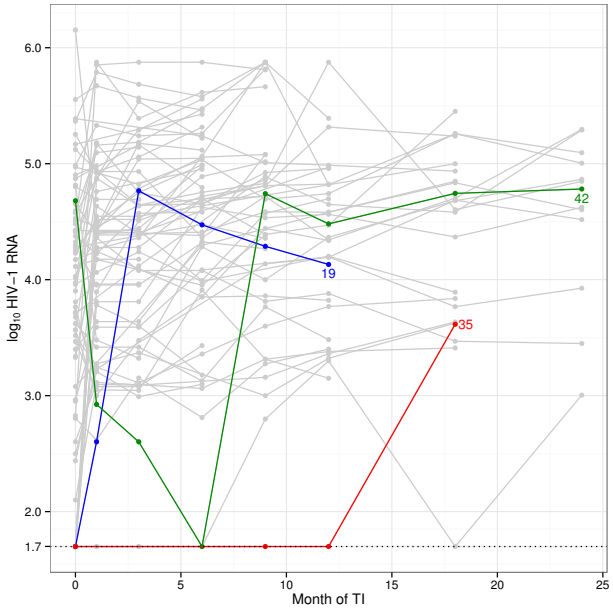

Figure 1 – **UTI data.** Individual profiles (in $\log_{10}$ scale) for HIV viral load at different follow-up times. Trajectories for some censored individuals are indicated in different colors.

Viral load measurements below the detectable threshold of 100 copies/mL (40 out of 361, 11%) were considered left-censored, and the censoring process assumed independence of the complete data. The individual profiles are shown in Figure 2.

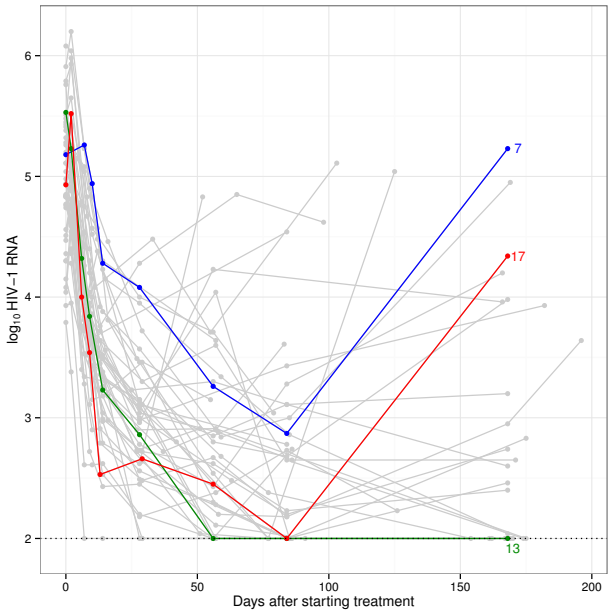

Figure 2 – **ACTG 315 data.** Individual profiles (in $\log_{10}$ scale) for HIV viral load at different follow-up times. Trajectories for some censored individuals are indicated in different colors.

1.2.3 AIEDRP data

This AIDS case study is from the AIEDRP program. This program, which is a large multicenter observational study of subjects with acute and early HIV infection, covers areas such as the evaluation of immune responses to HIV during acute infection, the assessment of thymic function and T-cell turnover during acute HIV infection and the assessment of transmission and prevalence of HIV resistance among treatment-naive subjects. The aim of this study was to help design future vaccines and to learn the implications of new anti-HIV treatments.

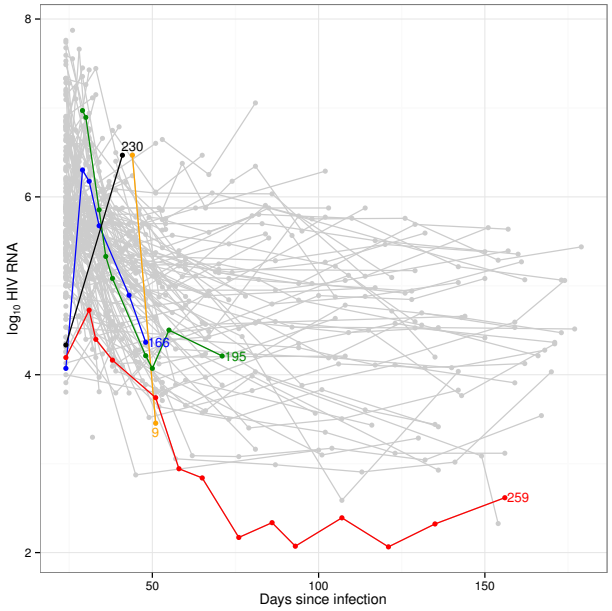

Figure 3 – **AIEDRP data.** Individual profiles (in $\log_{10}$ scale) for HIV viral load at different follow-up times. Trajectories for some censored individuals are indicated in different colors.

We consider 320 untreated individuals with acute HIV infection (see Vaida and Liu (2009) for more details). Of the 830 recorded observations, 185 (22%) were above the limit of assay quantification. This limit was between 75,000 and 500,000 copies/milliliter, depending on the assay. The individual profiles are shown in Figure 3.

1.2.4 A5055 data

This research is from the AIDS clinical trial study – A5055 (Wang, 2013), which involves a total of 44 infected patients with the human immunodeficiency virus type 1 (HIV-1). These patients were treated with one of the two potent antiretroviral (ARV) therapies, namely IDV 800 mg twice daily (q12h) plus RTV 200 mg q12h (treatment 1), and IDV 400 mg q12h plus RTV 400 mg q12h (treatment 2). In AIDS research, the number of RNA copies (viral load) in blood plasma and its evolutionary trajectories play a prominent role in the diagnosis of HIV-1 disease progression after a treatment of ARV

regimen (Paxton et al., 1997). Additionally, other two immunologic markers frequently used for monitoring disease progression in AIDS studies are the cluster of differentiation 4 (CD4) and cluster of differentiation 8 (CD8) T cells.

The dataset consists of plasma viral load measurements (in copies per milliliter), CD4 and CD8 cell counts measured roughly at days 0, 7, 14, 28, 56, 84, 112, 140, and 168 of follow-up for each patient. In this study, we focus on investigating the longitudinal trajectories for RNA viral load (in log-base-10 scale), denoted by $\log_{10}$(RNA), and CD4/CD8 ratio. In this study the lower detection limit for RNA viral load is 50 copies/milliliter, and therefore 33.5% (106 out of 316) of measurements lying below the limits of assay quantification (left-censored). Figure 4 shows the trajectories of the two immunologic responses along the time visit.

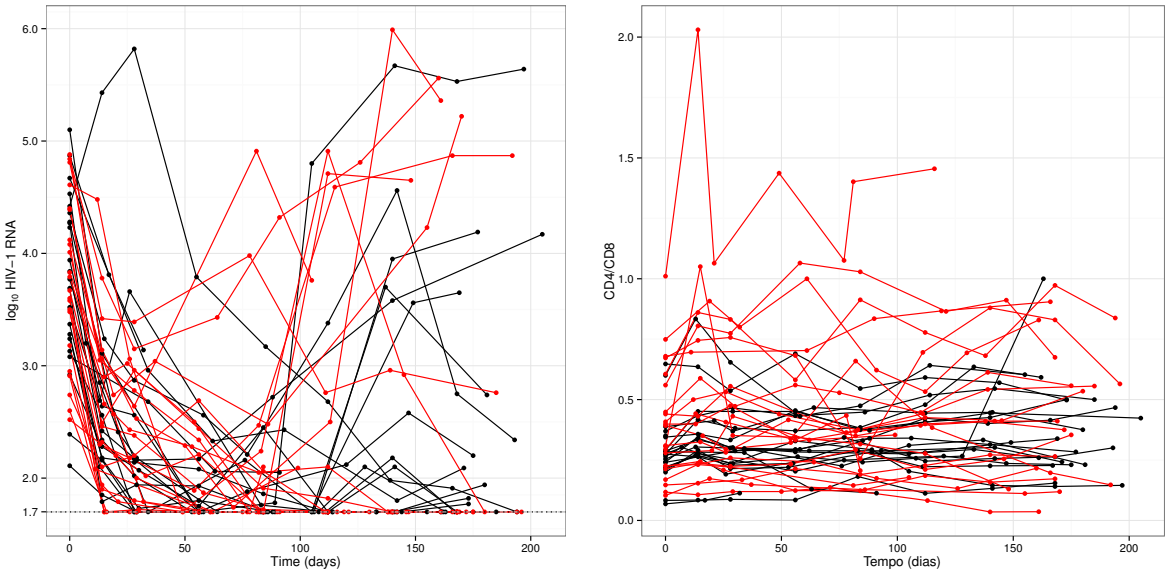

Figure 4 – **A5055 data.** Trajectories of $\log_{10}$ RNA (left panel) and CD4/CD8 ratio (right panel) for 44 HIV-1 infected patients who were randomized in two IDV-RTV regimens. Black lines indicate patients in under treatment 1 and red lines indicate patients under treatment 2. Dotted line in left panel indicates the censoring level.

Part II

Estimation and diagnostics in multivariate models for censored data

2 Censored mixed-effects models for irregularly observed repeated measures with applications to HIV viral load

2.1 Introduction

The study of acquired immunodeficiency syndrome (AIDS) and understanding of the dynamics of the human immunodeficiency virus (HIV) have become a focus of biomedical and biostatistical research. As mentioned by many researchers, HIV is an extremely dynamic and variable virus having new subtypes and recombinant forms, about which the scientific community knows little or nothing. HIV/AIDS clinical trials aim to find new ways to prevent, detect and/or treat AIDS by determining whether a new anti-retroviral (ARV) agent/therapy is safe and effective in people. Most of these clinical trials assess the quantitative rates/changes of HIV-1 ribonucleic acid (RNA) levels in plasma (or simply HIV-1 viral load), since this is an important surrogate marker to assess the risk of disease progression and to monitor response to ARV therapy in routine medical care of infected patients.

However, modeling HIV-1 viral load presents many challenges from the statistical point of view. Three are of particular importance. First, the viral load measurements are often left or right censored (undetected) due to a lower and/or upper detection limit of quantification. This happens because some quantification assays cannot accurately quantify HIV-1 RNA above/below a specific level. Particularly, lower detection limits ranging from 400 to 500 RNA copies/mL are considered for standard assays while the range is 50 to 100 RNA copies/mL for ultra-sensitive assays. For example, the Amplicor HIV-1 monitor test 1.5 and Nuclisens HIV-1 QT assay consider a lower detection limit of 400 copies/mL (Antunes et al., 2003), while the Roche Cobas Amplicor HIV-1 Monitor test (versions 1 and 1.5) considers a detection limit of 50 HIV-1 RNA copies/ml and the TaqMan assay (versions 1 and 2) considers a lower limit of quantification of 40 and 20 copies/ml respectively (see Swenson et al., 2014).

Second, as a result of unscheduled follow-up visits of patients and/or missed visits, the viral loads are usually recorded at irregular intervals. As an example of this situation, Ciesielski and Metler (1997) studied the duration of time between exposure and seroconversion in healthcare workers with occupationally acquired HIV infection. In this study, the authors mentioned that “because many of the healthcare workers had follow-up testing at irregular intervals, with long periods between tests, it was not possible to define

precisely when seroconversion occurred." Another example of this situation was reported by Azevedo et al. (2010), were a patient diagnosed with HIV-infection in 2003 during her first pregnancy made follow-up visits at irregular intervals. In this particular case, the antiretrovirals were given only for prophylaxis of vertical HIV-infection transmission during pregnancy. Finally, since the viral load is measured longitudinally over time, the between-subject and within-subject variations have to be taken into account.

Recently, some alternatives for modeling the irregular observation responses and correlations induced by longitudinal data have been proposed in the statistical literature. These proposals consider not only the correlation structure induced by the random effects term but also by other types of correlation in the error term. Particularly, Wang (2013) introduces the multivariate t (Student-t) linear mixed model (MtLMM) for outcome variables recorded on irregular occasions considering a damping exponential correlation (DEC) structure, as proposed by Muñoz et al. (1992). This correlation structure takes into account the autocorrelation generated by the within-subject dependence among irregular occasions. On the other hand, Lin and Wang (2013) consider a multivariate Student-t distribution for nonlinear mixed models with multiple outcomes in the presence of missing data. To capture the serial correlation among the observations, the authors consider a DEC structure of the error vector. Moreover, Wang and Fan (2011) consider the multivariate Student-t linear mixed with autoregressive of order p (AR(p)) dependence structure for the within-subject errors in the case of multiple outcomes.

In the case of censored responses, there are several alternatives in the literature to deal with them. For example, Arellano-Valle et al. (2012) extend the classic Tobit model (Tobin, 1958) by considering a Student-t distribution for the error term and proposing an EM-type algorithm for the parameter estimation. More recently, Rocha et al. (2015) put forward an errors-in-variable Student-t censored model, obtaining the maximum likelihood estimates (MLE) of the model through an EM algorithm, and Müller and van de Geer (2014) study a censored linear model for high dimensional data. In the context of linear/nonlinear mixed-effects (LME/NLME) models, Hughes (1999) proposes a likelihood-based Monte Carlo EM algorithm (MCEM) for LME with censored responses (LMEC). Wu (2002) suggests a Monte Carlo EM and a linearization procedure to estimate the parameters of a censored NLME model. In turn, Vaida et al. (2007) and Vaida and Liu (2009) extend the work of Hughes, proposing a more efficient EM algorithm than Hughes's. An extended review of these proposals can be found in the book by Wu (2010). Recently, Matos et al. (2013a), Matos et al. (2013b) and Matos et al. (2015) have proposed a likelihood-based estimation and influence analysis for LMEC/NLMEC models, respectively.

Moreover, stochastic versions of EM such as Monte Carlo EM (Levine and Casella, 2001), SAEM (Delyon et al., 1999) and many other approximations have been proposed to deal with NLME models under censoring. In fact, Samson et al. (2006) put

forward an extension of the SAEM algorithm to left-censored data in NLME model. However, to the best of our knowledge, there is no work considering irregular observations, damping exponential correlation and censored longitudinal responses simultaneously in the context of LMEC/NLMEC models using an exact EM algorithm. Consequently, the aim of this chapter is to study the impact of censoring and irregularly timed observed responses under Gaussian LMEC and NLMEC models.

For this purpose, we consider the analysis of two AIDS case studies. The first one investigated the effect of a highly active antiretroviral therapy (HAART) in persons with moderately advanced HIV-1 infection. This case study presented 11% of observations below the detection limits (left-censored). The second case study evaluated the immune responses to HIV during acute infection, presenting about 22% of measurements lying above the limits of assay quantification (right-censored). In both studies, the viral loads were irregularly measured over time.

The rest of the chapter is organized as follows. Section 2.2 introduces the model (DEC-LMEC) and the likelihood function. In Section 2.3, the related likelihood-based inference is presented, including estimation of the random effects and the expected information matrix. The method for predicting future observations is discussed in Section 2.4. Section 2.5 describes the extension to the nonlinear case (DEC-NLMEC). The application of the proposed method is presented in Sections 2.6 and 2.7 through a simulation study and the analysis of two case studies of HIV viral load. Finally, Section 2.8 concludes with a short discussion of issues raised by this study and some possible directions for future research.

2.2 Model formulation

In the non-censored case, a Gaussian LME model is specified as follows

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \tag{2.1}$$

where $\mathbf{b}_i \overset{iid}{\sim} \mathrm{N}_q(\mathbf{0}, \mathbf{D})$ is independent of $\boldsymbol{\epsilon}_i \overset{ind.}{\sim} \mathrm{N}_{n_i}(\mathbf{0}, \boldsymbol{\Omega}_i)$, $i = 1, \ldots, n$; the subscript i is the subject index; $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$ is an $n_i \times 1$ vector of observed continuous responses for subject i measured at particular time points $\mathbf{t}_i = (t_{i1}, \ldots, t_{n_i})^\top$; $\mathbf{X}_i$ is the $n_i \times p$ design matrix corresponding to the fixed effects, $\boldsymbol{\beta}$, of dimension $p \times 1$; $\mathbf{Z}_i$ is the $n_i \times q$ design matrix corresponding to the $q \times 1$ vector of random effects $\mathbf{b}_i$; $\boldsymbol{\epsilon}_i$, with dimension $(n_i \times 1)$, is the vector of random errors; and the dispersion matrix $\mathbf{D} = \mathbf{D}(\boldsymbol{\alpha})$ depends on the unknown and reduced parameters $\boldsymbol{\alpha}$. 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, such as the ACTG 315 and AIEDRP datasets, it is necessary to consider

a parsimonious parameterization of the matrix $\mathbf{E}_i$. Following Muñoz et al. (1992), we adopt a DEC structure for $\mathbf{E}_i$ described in Section 1.1.2.

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 $(\mathbf{V}_i, \mathbf{C}_i)$ be the observed data for the i-th subject, where $\mathbf{V}_i$ represents the vector of uncensored readings or censoring level and $\mathbf{C}_i$ is the vector of censoring indicators, such that

$$\begin{aligned} y_{ij} &\leqslant V_{ij} \quad \text{if} \quad C_{ij} = 1, \\ y_{ij} &= V_{ij} \quad \text{if} \quad C_{ij} = 0. \end{aligned} \tag{2.2}$$

Note that since the observed response y_{ij} is defined over the real line, extensions to right-censored data are straightforward. In fact, the right-censored problem can be represented by a left-censored problem by simultaneously transforming the response y_{ij} and censoring level V_{ij} to $-y_{ij}$ and $-V_{ij}$. The model defined in (2.1)-(2.2), is henceforth called DEC-LMEC.

2.2.1 The log-likelihood function

Following Vaida and Liu (2009), classic inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top, \boldsymbol{\phi}^\top)^\top$ is based on the marginal distribution of $\mathbf{y}_i$. For complete data, the marginal distribution of the vector $\mathbf{y}_i$, for $i = 1, \ldots, n$, is $\mathrm{N}_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i)$, where $\boldsymbol{\Sigma}_i = \boldsymbol{\Omega}_i + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top$. The strategy followed to compute the likelihood function associated with models (2.1) and (2.2) 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 $C_{ij} = 1$ for all elements in $\mathbf{y}_i^c$. After reordering, $\mathbf{y}_i$, $\mathbf{V}_i$, $\mathbf{X}_i$, and $\boldsymbol{\Sigma}_i$ can be partitioned as follows

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

In this setup, the operator $vec(\cdot)$ denotes the function that stacks vectors or matrices with the same number of columns. Consequently, from the marginal-conditional decomposition of the multivariate normal distribution, $\mathbf{y}_i^o \sim \mathrm{N}_{n_i^o}(\mathbf{X}_i^o\boldsymbol{\beta}, \boldsymbol{\Sigma}_i^{oo})$ and $\mathbf{y}_i^c \mid \mathbf{y}_i^o \sim \mathrm{N}_{n_i^c}(\boldsymbol{\mu}_i, \mathbf{S}_i)$, where $\boldsymbol{\mu}_i = \mathbf{X}_i^c\boldsymbol{\beta} + \boldsymbol{\Sigma}_i^{co}(\boldsymbol{\Sigma}_i^{oo})^{-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\boldsymbol{\beta})$ and $\mathbf{S}_i = \boldsymbol{\Sigma}_i^{cc} - \boldsymbol{\Sigma}_i^{co}(\boldsymbol{\Sigma}_i^{oo})^{-1}\boldsymbol{\Sigma}_i^{oc}$. Now, 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 $\mathrm{N}_{n_i}(\mathbf{a}, \mathbf{A})$ computed at vector $\mathbf{u}$. From Vaida and Liu (2009) and Matos et al. (2013a), the likelihood function for subject i (using conditional probability arguments) is given by

$$\begin{aligned} L_i(\boldsymbol{\theta}) &= f(\mathbf{y}_i^c \leqslant \mathbf{V}_i^c \mid \mathbf{y}_i^o = \mathbf{V}_i^o, \boldsymbol{\theta}) f(\mathbf{y}_i^o = \mathbf{V}_i^o \mid \boldsymbol{\theta}) \\ &= f(\mathbf{y}_i^c \leqslant \mathbf{V}_i^c \mid \mathbf{y}_i^o, \boldsymbol{\theta}) f(\mathbf{y}_i^o \mid \boldsymbol{\theta}) \\ &= \Phi_{n_i^c}(\mathbf{V}_i^c; \boldsymbol{\mu}_i, \mathbf{S}_i) \phi_{n_i^o}(\mathbf{y}_i^o; \mathbf{X}_i^o\boldsymbol{\beta}, \boldsymbol{\Sigma}_i^{oo}), \end{aligned} \tag{2.3}$$

which can be easily evaluated computationally.

The log-likelihood function for the observed data, given by

$$\ell(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^{n}\{\log L_i(\boldsymbol{\theta})\},$$

is used to compute different model selection criteria, such as

$$\text{AIC} = 2\,m - 2\,\ell_{max} \;\text{ and }\; \text{BIC} = m\,log\,N - 2\,\ell_{max},$$

where m is the number of model parameters, $N = \sum_{i=1}^{n} n_i$ and ℓ_{max} is the maximized log-likelihood value.

2.3 The EM algorithm

This section describes in detail how the proposed model specified in (2.1)-(2.2) can be fitted by using the ECM algorithm (Meng and Rubin, 1993). The EM algorithm (proposed originally by Dempster et al., 1977) has several appealing features, such as stability of monotone convergence with each iteration, increasing the likelihood and simplicity of implementation. Due to the computational difficulty at the M-step, we use the ECM algorithm (an extension of the EM algorithm), which shares the appealing features of the EM and converges faster than the original algorithm.

Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$, $\mathbf{V} = vec(\mathbf{V}_1, \ldots, \mathbf{V}_n)$ and $\mathbf{C} = vec(\mathbf{C}_1, \ldots, \mathbf{C}_n)$. Considering $\mathbf{b}$ as the hypothetical missing data, the complete data are denoted by

$$\mathbf{y}_c = (\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{y}^\top, \mathbf{b}^\top)^\top.$$

Hence, the ECM algorithm is applied to the complete data log-likelihood function

$$\begin{aligned}\ell_i(\boldsymbol{\theta} \mid \mathbf{y}_c) &= -\frac{1}{2}\Big[n_i \log\sigma^2 + \log(|\mathbf{E}_i|) + \frac{1}{\sigma^2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top\mathbf{E}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)\\ &+ \log|\mathbf{D}| + \mathbf{b}_i^\top\mathbf{D}^{-1}\mathbf{b}_i\Big] + K, \qquad (2.4)\end{aligned}$$

with K being a constant that does not depend on the parameter vector $\boldsymbol{\theta}$. Given the current estimate $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, the E-step calculates the conditional expectation of the complete data log-likelihood function, given by

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

where

$$
\begin{aligned}
Q_{1i}\left(\boldsymbol{\beta},\sigma^2,\boldsymbol{\phi}\mid\widehat{\boldsymbol{\theta}}^{(k)}\right) &= -\frac{n_i}{2}\log\widehat{\sigma^2}^{(k)} - \frac{1}{2\widehat{\sigma^2}^{(k)}}\left[\widehat{a}_i^{(k)}(\widehat{\boldsymbol{\phi}}^{(k)}) + \widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top\widehat{\mathbf{E}}_i^{-1(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\right. \\
&- \left. 2\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top\widehat{\mathbf{E}}_i^{-1(k)}\left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{\mathbf{b}}_i^{(k)}\right)\right] - \frac{1}{2}\log(|\widehat{\mathbf{E}}_i^{(k)}|), \qquad (2.5)\\
Q_{2i}\left(\boldsymbol{\alpha}\mid\widehat{\boldsymbol{\theta}}^{(k)}\right) &= -\frac{1}{2}\log|\widehat{\mathbf{D}}^{(k)}| - \frac{1}{2}tr\left(\widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)}\widehat{\mathbf{D}}^{-1(k)}\right), \qquad (2.6)
\end{aligned}
$$

and $\widehat{a}_i^{(k)}(\boldsymbol{\phi}) = tr\left(\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)}\mathbf{E}_i^{-1} - 2\widehat{\mathbf{y}_i\mathbf{b}_i^\top}^{(k)}\mathbf{Z}_i^\top\mathbf{E}_i^{-1} + \widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)}\mathbf{Z}_i^\top\mathbf{E}_i^{-1}\mathbf{Z}_i\right)$,

$$
\begin{aligned}
\widehat{\mathbf{b}_i}^{(k)} &= E\left\{\mathbf{b}_i\mid\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\right\} = \widehat{\boldsymbol{\varphi}}_i^{(k)}(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}),\\
\widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)} &= E\left\{\mathbf{b}_i\mathbf{b}_i^\top\mid\mathbf{V}_i,\mathbf{C}_i,\ \widehat{\boldsymbol{\theta}}^{(k)}\right\}\\
&= \widehat{\boldsymbol{\Lambda}}_i^{(k)} + \widehat{\boldsymbol{\varphi}}_i^{(k)}(\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} - \widehat{\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\widehat{\mathbf{y}}_i^{(k)\top} + \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top)\widehat{\boldsymbol{\varphi}}_i^{(k)\top},\\
\widehat{\mathbf{y}_i\mathbf{b}_i^\top}^{(k)} &= E\{\mathbf{y}_i\mathbf{b}_i^\top\mid\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\} = (\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} - \widehat{\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top)\widehat{\boldsymbol{\varphi}}_i^{(k)\top},
\end{aligned}
$$

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

It is easy to see from (2.5) and (2.6) that the E-step reduces only to the computation of

$$\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} = E\{\mathbf{y}_i\mathbf{y}_i^\top\mid\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\} \text{ and } \widehat{\mathbf{y}}_i^{(k)} = E\{\mathbf{y}_i\mid\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\}.$$

These conditional expectations rely on the first and second moments of a multivariate truncated normal distribution and can be determined in closed-form (for more details on the computation of these moments, see Vaida and Liu, 2009).

The conditional maximization step (CM) conditionally maximizes $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$ obtaining a new estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$, as follows

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}^{(k+1)} &= \left(\sum_{i=1}^n\mathbf{X}_i^\top\widehat{\mathbf{E}}_i^{-1(k)}\mathbf{X}_i\right)^{-1}\sum_{i=1}^n\mathbf{X}_i^\top\widehat{\mathbf{E}}_i^{-1(k)}\left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{\mathbf{b}_i}^{(k)}\right), \qquad (2.7)\\
\widehat{\sigma^2}^{(k+1)} &= \frac{1}{N}\sum_{i=1}^n\left[\widehat{a}_i^{(k)} - 2\widehat{\boldsymbol{\beta}}^{(k+1)\top}\mathbf{X}_i^\top\widehat{\mathbf{E}}_i^{-1(k)}\left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{\mathbf{b}_i}^{(k)}\right)\right.\\
&+ \left.\widehat{\boldsymbol{\beta}}^{(k+1)\top}\mathbf{X}_i^\top\widehat{\mathbf{E}}_i^{-1(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k+1)}\right], \qquad (2.8)\\
\widehat{\mathbf{D}}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^n\widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)}, \qquad (2.9)\\
\widehat{\boldsymbol{\phi}}^{(k+1)} &= \underset{\boldsymbol{\phi}\in(0,1)\times\mathbb{R}^+}{\operatorname{argmax}}\left(-\frac{1}{2}\log(|\mathbf{E}_i|) - \frac{1}{2\widehat{\sigma^2}^{(k+1)}}\left[\widehat{a}_i^{(k)}(\boldsymbol{\phi})\right.\right.\\
&- \left.\left.2\widehat{\boldsymbol{\beta}}^{(k+1)\top}\mathbf{X}_i^\top\mathbf{E}_i^{-1}\left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{\mathbf{b}}_i^{(k)}\right) + \widehat{\boldsymbol{\beta}}^{(k+1)\top}\mathbf{X}_i^\top\mathbf{E}_i^{-1}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k+1)}\right]\right). \qquad (2.10)
\end{aligned}
$$

2.3.1 Estimation of random effects and standard errors

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, $E\{\mathbf{b}_i \mid \mathbf{V}_i, \mathbf{C}_i\}$. Thus, for a given value of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top, \boldsymbol{\phi}^\top)^\top$, the conditional mean of $\mathbf{b}_i$ given $\mathbf{V}_i$ and $\mathbf{C}_i$ is

$$\widehat{\mathbf{b}}_i(\boldsymbol{\theta}) = E\{\mathbf{b}_i \mid \mathbf{V}_i, \mathbf{C}_i\} = \boldsymbol{\varphi}_i(\widehat{\mathbf{y}}_i - \mathbf{X}_i\boldsymbol{\beta}), \tag{2.11}$$

where $\boldsymbol{\varphi}_i = \boldsymbol{\Lambda}_i\mathbf{Z}_i^\top\mathbf{E}_i^{-1}/\sigma^2$ and $\boldsymbol{\Lambda}_i = (\mathbf{D}^{-1} + \mathbf{Z}_i^\top\mathbf{E}_i^{-1}\mathbf{Z}_i/\sigma^2)^{-1}$. Note that $\widehat{\mathbf{y}}_i = E\{\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i\}$ is given by the first moment of a multivariate truncated normal distribution. In practice, the estimator of $\mathbf{b}_i$, $\widehat{\mathbf{b}}_i$, can be obtained by substituting the ML estimate $\widehat{\boldsymbol{\theta}}$ into (2.11), leading to $\widehat{\mathbf{b}}_i = \widehat{\mathbf{b}}_i(\widehat{\boldsymbol{\theta}})$. On the other hand, the conditional covariance matrix of $\mathbf{b}_i$ given $\mathbf{V}_i$ and $\mathbf{C}_i$ is

$$Var\{\mathbf{b}_i \mid \mathbf{V}_i, \mathbf{C}_i\} = E\{\mathbf{b}_i\mathbf{b}_i^\top \mid \mathbf{V}_i, \mathbf{C}_i\} - \widehat{\mathbf{b}}_i(\boldsymbol{\theta})\widehat{\mathbf{b}}_i(\boldsymbol{\theta})^\top = \boldsymbol{\Lambda}_i + \boldsymbol{\varphi}_i Var(\mathbf{y}_i \mid \mathbf{V}_i, \mathbf{C}_i)\boldsymbol{\varphi}_i^\top.$$

Note that $Var(\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i)$ can be easily obtained as a byproduct of the proposed ECM algorithm developed in Section 2.3.

The empirical information matrix

Following the methodology described in (1.1.4), we compute the asymptotic covariance of the ML estimates through the empirical information matrix. So, we have that

$$\mathbf{I}_e(\widehat{\boldsymbol{\theta}} \mid \mathbf{y}) = \sum_{i=1}^{n} \widehat{\mathbf{s}}_i\widehat{\mathbf{s}}_i^\top, \tag{2.12}$$

where $\widehat{\mathbf{s}}_i = \mathbf{s}(\mathbf{y}_i \mid \widehat{\boldsymbol{\theta}}) = E\left(\dfrac{\partial \ell_i(\boldsymbol{\theta} \mid \mathbf{y}_c)}{\partial \boldsymbol{\theta}} \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\right) = \left(\widehat{\mathbf{s}}_{i,\boldsymbol{\beta}}^\top, \widehat{\mathbf{s}}_{i,\sigma^2}, \widehat{\mathbf{s}}_{i,\boldsymbol{\alpha}}^\top, \widehat{\mathbf{s}}_{i,\boldsymbol{\phi}}^\top\right)^\top$ has elements given by

$$\begin{aligned}
\widehat{\mathbf{s}}_{i,\boldsymbol{\beta}} &= (\widehat{\mathbf{s}}_{i,\beta_1}, \ldots, \widehat{\mathbf{s}}_{i,\beta_p})^\top = \frac{1}{\widehat{\sigma^2}}\left[\mathbf{X}_i^\top\widehat{\mathbf{E}}_i^{-1}\left(\widehat{\mathbf{y}}_i - \mathbf{Z}_i\widehat{\mathbf{b}_i}\right) - \mathbf{X}_i^\top\widehat{\mathbf{E}}_i^{-1}\mathbf{X}_i\widehat{\boldsymbol{\beta}}\right], \\
\widehat{\mathbf{s}}_{i,\sigma^2} &= -\frac{n_i}{2\widehat{\sigma^2}} + \frac{1}{2\widehat{\sigma^4}}\left[\widehat{a}_i - 2\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_i^\top\widehat{\mathbf{E}}_i^{-1}\left(\widehat{\mathbf{y}}_i - \mathbf{Z}_i\widehat{\mathbf{b}}_i\right) + \widehat{\boldsymbol{\beta}}^\top\mathbf{X}_i^\top\widehat{\mathbf{E}}_i^{-1}\mathbf{X}_i\widehat{\boldsymbol{\beta}}\right], \\
\widehat{\mathbf{s}}_{i,\boldsymbol{\alpha}} &= (\widehat{\mathbf{s}}_{i,\alpha_1}, \ldots, \widehat{\mathbf{s}}_{i,\alpha_r})^\top, \\
\widehat{\mathbf{s}}_{i,\boldsymbol{\phi}} &= (\widehat{\mathbf{s}}_{i,\phi_1}, \widehat{\mathbf{s}}_{i,\phi_2})^\top,
\end{aligned}$$

with $\widehat{a}_i = tr\left(\widehat{\mathbf{y}_i\mathbf{y}_i^\top}\widehat{\mathbf{E}}_i^{-1} - 2\widehat{\mathbf{y}_i\mathbf{b}_i^\top}\mathbf{Z}_i^\top\widehat{\mathbf{E}}_i^{-1} + \widehat{\mathbf{b}_i\mathbf{b}_i^\top}\mathbf{Z}_i^\top\widehat{\mathbf{E}}_i^{-1}\mathbf{Z}_i\right)$, $\widehat{\mathbf{s}}_{i,\alpha_r} = -\dfrac{1}{2}tr\left(\widehat{\mathbf{D}}^{-1}\dot{\mathbf{D}}(r)\widehat{\mathbf{D}}^{-1}(\widehat{\mathbf{D}} - \widehat{\mathbf{b}_i\mathbf{b}_i^\top})\right)$,

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

where $\dot{\mathbf{D}}(r) = \dfrac{\partial \mathbf{D}}{\partial \alpha_r}|_{\boldsymbol{\alpha}=\widehat{\boldsymbol{\alpha}}}$, $r = 1, \ldots, dim(\boldsymbol{\alpha})$; and $\dot{\mathbf{E}}_i(s) = \dfrac{\partial \mathbf{E}_i}{\partial \phi_s}|_{\boldsymbol{\phi}=\widehat{\boldsymbol{\phi}}}$, $s = 1, 2$. For the DEC structure we have that

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

2.4 Prediction of future observations

The problem related to the prediction of future values has a great impact on many practical applications. Rao (1987) pointed out that the predictive accuracy of future observations can be taken as an alternative measure of "goodness-of-fit". In order to propose a strategy to generate predicted values from the DEC-LMEC model, we use the approach proposed by Wang (2013). Thus, let $\mathbf{y}_{i,obs}$ be an observed response vector of dimension $n_{i,obs} \times 1$ for a new subject i over the first portion of time and $\mathbf{y}_{i,pred}$ be the corresponding $n_{i,pred} \times 1$ response vector over the future portion of time. Let $\bar{\mathbf{X}}_i = (\mathbf{X}_{i,obs}, \mathbf{X}_{i,pred})$ be the $(n_{i,obs} + n_{i,pred}) \times p$ design matrix corresponding to $\bar{\mathbf{y}}_i = (\mathbf{y}_{i,obs}^\top, \mathbf{y}_{i,pred}^\top)$.

To deal with the censored values existing in $\mathbf{y}_{i,obs}$, we use the imputation procedure, by replacing the censored values by $\widehat{\mathbf{y}}_i = E\{\mathbf{y}_i | \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\}$ obtained from the EM algorithm. Therefore, when the censored values are imputed, a complete dataset, denoted by $\mathbf{y}_{i,obs*}$, is obtained. The reason to use the imputation procedure is that it avoids computing truncated conditional expectations of the multivariate normal distribution originated by the censoring scheme. Hence, we have that

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

where the matrix $\boldsymbol{\Sigma}_i$, can be represented by $\boldsymbol{\Sigma}_i = \begin{pmatrix} \boldsymbol{\Sigma}_i^{obs*,obs*} & \boldsymbol{\Sigma}_i^{obs*,pred} \\ \boldsymbol{\Sigma}_i^{pred,obs*} & \boldsymbol{\Sigma}_i^{pred,pred} \end{pmatrix}$. As mentioned in Wang (2013), the best linear predictor of $\mathbf{y}_{i,pred}$ with respect to the minimum mean squared error (MSE) criterion is the conditional expectation of $\mathbf{y}_{i,pred}$ given $\mathbf{y}_{i,obs*}$, which is given by

$$\widehat{\mathbf{y}}_{i,pred}(\boldsymbol{\theta}) = \mathbf{X}_{i,pred}\boldsymbol{\beta} + \boldsymbol{\Sigma}_i^{pred,obs*}\boldsymbol{\Sigma}_i^{obs*,obs*-1}\left(\mathbf{y}_{i,obs*} - \mathbf{X}_{i,obs*}\boldsymbol{\beta}\right). \tag{2.13}$$

Therefore, $\mathbf{y}_{i,pred}$ can be estimated directly by substituting $\widehat{\boldsymbol{\theta}}$ into (2.13), leading to $\widehat{\mathbf{y}_{i,pred}} = \widehat{\mathbf{y}}_{i,pred}(\widehat{\boldsymbol{\theta}})$.

2.5 The nonlinear case

As mentioned in the Introduction, some approximations based on the EM algorithm have been proposed in the statistical literature to deal with NLME models. In

this context, we use an approximation of the nonlinear functions mentioned by Vaida and Liu (2009). This approximation (2.15) was considered by Matos et al. (2013a) in the context of censored nonlinear mixed-effects models. In that paper, simulation studies revealed that the approximation can efficiently estimate the model parameters. Recently, Wang (2013) used this approximation to implement an ECM algorithm to carry out ML estimation in Student-t nonlinear mixed-effects models for multi-outcome longitudinal data with missing values. Consequently, we conclude that this approximation is robust, stable, and does not produce any severe consequences in inference when applied to other types of (censored) nonlinear models.

The NLME (without censoring) of Pinheiro and Bates (2000) is defined as

$$\mathbf{y}_i = \eta(\boldsymbol{\psi}_i, \mathbf{X}_i) + \boldsymbol{\epsilon}_i, \;\; \boldsymbol{\psi}_i = \mathbf{A}_i\boldsymbol{\beta} + B_i\mathbf{b}_i, \;\; i = 1, \ldots, n, \tag{2.14}$$

where $\mathbf{b}_i \overset{iid}{\sim} \mathrm{N}_q(0, \mathbf{D})$ and $\boldsymbol{\epsilon}_i \overset{ind}{\sim} \mathrm{N}_{n_i}(0, \sigma^2\mathbf{E}_i)$ are independent; $\mathbf{y}_i$ is an $(n_i \times 1)$ vector of observed responses for subject i; η is a nonlinear function of the individual random parameter $\boldsymbol{\psi}_i$; $\mathbf{A}_i$ and $\mathbf{B}_i$ are known design matrices of dimensions $r \times p$ and $r \times q$, respectively, possibly depending on some covariate values, $\mathbf{X}_i$; and $\boldsymbol{\beta}$ is the $(p \times 1)$ vector of fixed effects and $\mathbf{b}_i$ is the $(q \times 1)$ vector of random effects.

As mentioned by Vaida and Liu (2009), the linearization (L) procedure to obtain the approximate MLE of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top, \boldsymbol{\phi}^\top)^\top$ involves taking the first-order Taylor expansion of η_i around the current parameter estimate $\widetilde{\boldsymbol{\beta}}$ and the random effect estimates $\widetilde{\mathbf{b}}_i$ (empirical predictors). This procedure is equivalent to iteratively solving the following LME model (L-step)

$$\widetilde{\mathbf{Y}}_i = \widetilde{\mathbf{W}}_i\boldsymbol{\beta} + \widetilde{\mathbf{H}}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \;\; i = 1, \ldots, n, \tag{2.15}$$

where $\mathbf{b}_i \overset{iid}{\sim} \mathrm{N}_q(0, \mathbf{D})$ and $\boldsymbol{\epsilon}_i \overset{ind}{\sim} \mathrm{N}_{n_i}(\mathbf{0}, \sigma^2\mathbf{E}_i)$; and $\widetilde{\mathbf{Y}}_i = \mathbf{Y}_i - \widetilde{\eta}(\boldsymbol{\psi}(\widetilde{\boldsymbol{\beta}}, \widetilde{\mathbf{b}}_i), \mathbf{X}_i)$, with

$$\widetilde{\mathbf{H}}_i = \frac{\partial\eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i)}{\partial\mathbf{b}_i^\top}\Big|_{\mathbf{b}_i=\widetilde{\mathbf{b}}_i}, \quad \widetilde{\mathbf{W}}_i = \frac{\partial\eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i)}{\partial\boldsymbol{\beta}^\top}\Big|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}},$$

and $\widetilde{\eta}(\boldsymbol{\psi}(\widetilde{\boldsymbol{\beta}}, \widetilde{\mathbf{b}}_i), \mathbf{X}_i) = \eta(\mathbf{A}_i\widetilde{\boldsymbol{\beta}} + \mathbf{B}_i\widetilde{\mathbf{b}}_i, \mathbf{X}_i) - \widetilde{\mathbf{H}}_i\widetilde{\mathbf{b}}_i - \widetilde{\mathbf{W}}_i\widetilde{\boldsymbol{\beta}}$. Thus, in the censored case, the model in (2.15) is an LME with censored data that can be fitted using the strategy explained in Section 2.3. The model matrices in (2.15) depend on the current parameter value, and need to be recalculated at each iteration. The algorithm iterates between the L-, E- and CM-steps until convergence.

2.6 Analysis of case studies

This section illustrates the performance of the proposed methods with the analysis of two HIV datasets, previously analyzed by Wu (2002) and Vaida and Liu (2009), respectively.

2.6.1 ACTG 315 data

Here we reanalyze the HIV viral load data from clinical trial ACTG 315 (Wu, 2002), considering four different correlation structures, namely the uncorrelated structure (UNC), damped exponential correlation (DEC), continuous-time autoregressive of order 1 (AR(1)) and compound symmetric structure (CS). As mentioned in Section 1.2, the dataset consists of 46 HIV-1 infected patients treated with a potent ARV therapy. The viral load was repeatedly quantified on days 0, 2, 7, 10, 14, 21, 28, 56, 84, 168, and 196 after start of treatment, with a total of 361 observations. The viral load detectable limit is 100 copies/mL, and 40 out of 361 (11%) of all viral load measurements are below the detection limit. Wu and Ding (1999) proposed the use of a biphasic model

$$V(t) = e^{\varphi_1 - \varphi_2 t} + e^{\varphi_3 - \varphi_4 t}, \tag{2.16}$$

where $V(t)$ is the viral load at time t. The parameters φ_2 and φ_4 are called the first and second phase viral decay rates, which can represent the minimum turnover rate of productively infected cells and that of latently or long-lived infected cells, respectively. The parameters φ_1 and φ_3 are macro-parameters and $e^{\varphi_1} + e^{\varphi_3}$ is the baseline viral load at time $t = 0$.

Table 2 – **ACTG 315 data.** Model selection criterion for the NLMEC model under different correlation structures.

	NLMEC			
Criterion	UNC	DEC	AR(1)	CS
ℓ_{max}	-281.31	**-255.83**	-264.99	-279.33
AIC	594.61	**547.66**	563.97	592.66
BIC	656.83	**617.66**	630.08	658.77

As noted by Wu and Ding (1999), the inter-subject variation of observed viral loads motivates the use of a NLME model. The viral load trajectories initially exhibit rapid decay (known as first-phase decay), followed by a phase of slow decay for some (the second phase) with the others rebounding back to the original levels (Liu and Wu, 2012). Therefore, following Wu (2002) we consider the following NLME model to reflect the dynamics of the HIV viral load

$$y_{ij} = \log_{10}(e^{\varphi_{1i} - \varphi_{2i} t_{ij}} + e^{\varphi_{3i} - \varphi_{4ij} t_{ij}}) + \epsilon_{ij}, \tag{2.17}$$

$$\beta_{1ij} = \varphi_{1i} = \beta_1 + b_{1i}, \quad \beta_{3ij} = \varphi_{3i} = \beta_3 + b_{3i}, \tag{2.18}$$

$$\beta_{2ij} = \varphi_{2i} = \beta_2 + b_{2i}, \quad \beta_{4_{ij}} = \varphi_{4ij} = \beta_4 + \beta_5 \text{CD4}_{ij} + b_{4i}, \tag{2.19}$$

where y_{ij} is the $\log_{10}$-transformation of the viral load for the i-th subject at time t_{ij} $(i = 1, 2, \ldots, n, j = 1, 2, \ldots, n_i)$ and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})^\top$ represents the vector of within-individual random errors; CD4_{ij} indicates the observed CD4 values up to time t_{ij}; $\boldsymbol{\beta}_{ij} =$

$(\beta_{1ij}, \beta_{2ij}, \beta_{3ij}, \beta_{4ij})^\top$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_5)^\top$ are individual parameters for the i-th subject at time t_{ij} and population parameters, respectively and $\mathbf{b}_i = (b_{1i}, \ldots, b_{4i})^\top$ is the random effects vector for subject i.

Table 3 – **ACTG 315 data.** ML estimates with standard errors for the NLMEC model under DEC structure.

Fixed effects			Between-subject variances			Within-subject variances		
Parameter	Estimate	SE	Parameter	Estimate	SE	Parameter	Estimate	SE
β_1	11.552	0.266	α_{11}	0.155	0.045	σ^2	0.407	0.094
β_2	31.549	0.040	α_{12}	-0.808	0.127	ϕ_1	0.188	0.152
β_3	6.861	0.325	α_{22}	5.753	0.045	ϕ_2	0.647	0.084
β_4	-0.994	0.810	α_{13}	0.020	0.099			
β_5	0.612	0.195	α_{23}	0.110	0.114			
			α_{33}	0.258	0.215			
			α_{14}	-0.714	0.11			
			α_{24}	4.625	0.069			
			α_{34}	0.598	0.121			
			α_{44}	5.654	0.03			

The values of ℓ_{max}, AIC and BIC for the four considered models are presented in Table 2. Note that based on these criteria, the model presenting the best fit is the model with a damped exponential correlation structure (DEC). Furthermore, the likelihood ratio test (LRT) for the hypotheses $H_0 : \phi_2 = 1$ and $H_1 : \phi_2 \neq 1$ is performed. The resulting LRT statistic is 18.32 with p-value 0.00002, which is significant compared to $\chi^2_{1,0.05}$, suggesting that the DEC structure is more appropriate than the AR(1) for modeling the dependence among the within-subject errors. Figure 5 shows some individual profiles (in log10 scale) for HIV viral load at different follow-up times and the smoothed means of residuals from model fits.

ML estimates corresponding to the best model are presented in Table 3. Using these estimates, one can quantify the population decay rates and viral load parameters. The first- and second-phase decay rates can be approximated as $\widehat{\varphi}_2 = 31.549$ and $\widehat{\varphi}_4(t) = -0.994 + 0.612\,\mathrm{CD4}$. The population viral load process can be represented as $\widehat{V}(t) = \exp\{11.552 - \widehat{\varphi}_2(t)t\} + \exp\{6.861 - \widehat{\varphi}_4(t)t\}$. The SE values of the parameter estimates are obtained using the empirical information matrix (Section 2.3.1). Finally, using a bootstrap procedure, one can conclude that all the fixed effects considered in the model are statistically significant at $\alpha = 0.05$.

2.6.2 AIEDRP data

The second case study is taken from the AIEDRP program, a large multicenter observational study of subjects with acute and early HIV infection, consisting of 320 untreated individuals with acute HIV infection. Of the 830 recorded observations, 185 (22%) were above the limit of assay quantification. Therefore, in the spirit of Vaida and Liu (2009), we consider a right-censored five-parameter NLME model (inverted S-shaped

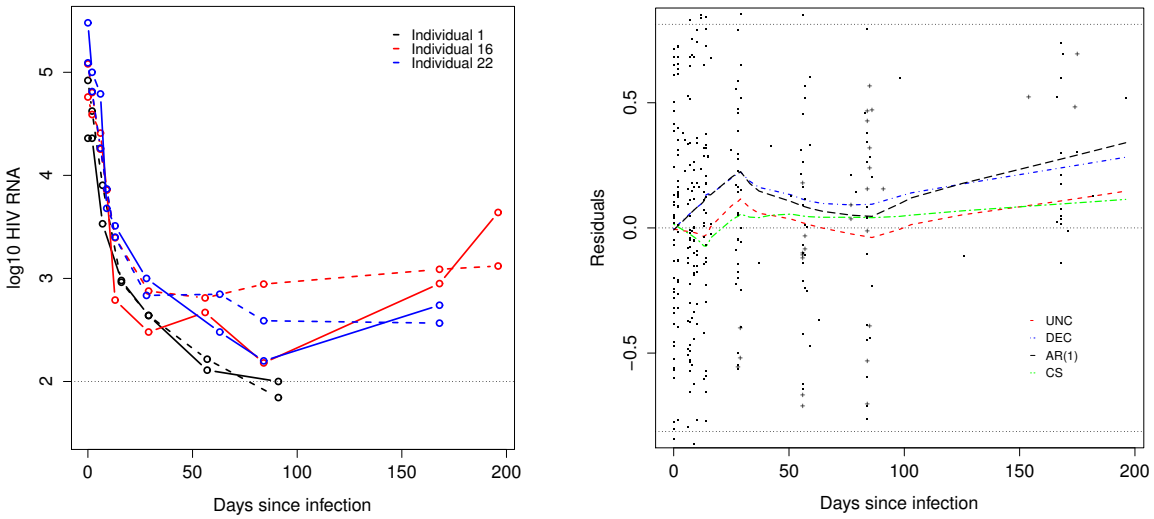

Figure 5 – **ACT315 data.** (Left panel) Individual profiles (in log10 scale) for HIV viral load at different follow-up times for some subjects. The dashed lines are the respective fitted profile. (Right panel) Smoothed means of residuals from model fits. The residuals from the model with autoregressive of order 1 correlation appear as points.

curve) as follows

$$y_{ij} = \lambda_{1i} + \frac{\lambda_2}{1 + \exp((t_{ij} - \lambda_3)/\lambda_4)} + \lambda_{5i}(t_{ij} - 50) + \epsilon_{ij}, \tag{2.20}$$

where y_{ij} is the $\log_{10}$ of the viral load for subject i at time t_{ij}. The parameters λ_{1i} and λ_2 represent the subject-specific set-point values and decrease from the maximum HIV-1 RNA. The location parameter λ_3 indicates the time point at which half of the change in HIV-1 RNA is attained, λ_4 is a scale parameter modeling the rate of decline and λ_{5i} allows increasing the HIV-1 RNA trajectory after day 50. The reparameterization given by $\beta_{1i} = \log(\lambda_{1i}) = \beta_1 + b_{1i}$; $\beta_k = \log(\lambda_k)$, $k = 2, 3, 4$, and $\lambda_{5i} = \beta_5 + b_{2i}$ is adopted to assure positive values for the model parameters.

Table 4 – **AIEDRP data.** Model selection criterion for the NLMEC model under different correlation structures.

	LMEC			
Criterion	UNC	DEC	AR(1)	CS
ℓ_{max}	-783.79	-769.81	**-770.10**	-775.62
AIC	1585.59	1561.63	**1560.19**	1571.25
BIC	1628.08	1613.56	**1607.41**	1618.46

As in Section 2.6.1, the correlation structures UNC, DEC, AR(1) and CS are considered. Table 4 summarizes the values of ℓ_{max}, AIC and BIC for all considered models. Note that the values of ℓ_{max} for the DEC and AR(1) models are close. This is

explained because the estimated values of ϕ_1 and ϕ_2 under the DEC model are 0.83 and 1.15 respectively. Based on this observation and the criteria, the best (most parsimonious) fit is obtained using the continuous-time autoregressive of order 1 correlation (AR(1)). The likelihood ratio tests (LRT) of hypothesis $H_0 : \phi_2 = 1$ and $H_1 : \phi_2 \neq 1$ are also performed. The resulting LRT statistic is 0.58 with p-value 0.446, which is not significant compared to $\chi^2_{1,0.05}$, suggesting that the AR(1) structure is more appropriate than the DEC for modeling the dependence among the within-subject errors. Moreover, the model fit of the AR(1) (and DEC) model is slightly better than the CS model, with the smoothed mean residual curve in Figure 6 (right panel) always being closer to zero.

Table 5 – **AIEDRP data.** ML estimates with standard errors for the NLMEC model under AR(1) structure.

Fixed effects			Between-subject variances			Within-subject variances		
Parameter	Estimate	SE	Parameter	Estimate	SE	Parameter	Estimate	SE
β_1	1.614	0.011	α_{11}	0.01658	0.00307	σ^2	0.308	0.024
β_2	0.128	0.003	α_{12}	0.00020	0.00016	ϕ_1	0.808	0.033
β_3	3.516	0.025	α_{22}	0.00003	0.00001			
β_4	1.118	0.001						
β_5	-0.004	0.001						

The ML estimates in this model are presented in Table 5. As in the previous case, the SE values for the parameter estimates are obtained using the empirical information matrix. One can use the AR(1) model with reasonable confidence for predictions of viral load. For example, at 6 months since infection, the average viral load is 4.537 $\log_{10}$ units. The individual 6-month viral load estimates vary between 1.794 and 6.469, with 5th and 95th quantiles at 3.466 and 5.549. The average slope after day 50 is negative, $\beta_{5i} = -0.004$ $\log_{10}$HIV/day, with 95% CI(-0.006,-0.002). And, for the individual slopes α_{5i} the 5th and 95th quantiles are -0.0061 and -0.0015. We performed a bootstrap procedure for hypothesis testing of the significance of the fixed effects ($\alpha = 0.05$), concluding that all of them are statistically significant (different from zero).

2.7 Simulation Studies

In order to examine the performance of the proposed method, here we report three simulation studies to investigate (a) the consequences for parameter estimation, (b) the behavior of the prediction when the correlation structure of the error term is misspecified, and (c) the asymptotic behavior of the parameter estimates. For this purpose and simplicity's sake, we consider a logistic model similar to that studied in Section 2.6.2, with random set-points λ_{1i} and random decline rates λ_{4i}, as follows

$$y_{ij} = \lambda_{1i} + \frac{\lambda_2}{1 + \exp((t_{ij} - \lambda_3)/\lambda_{4i})} + \epsilon_{ij}, \tag{2.21}$$

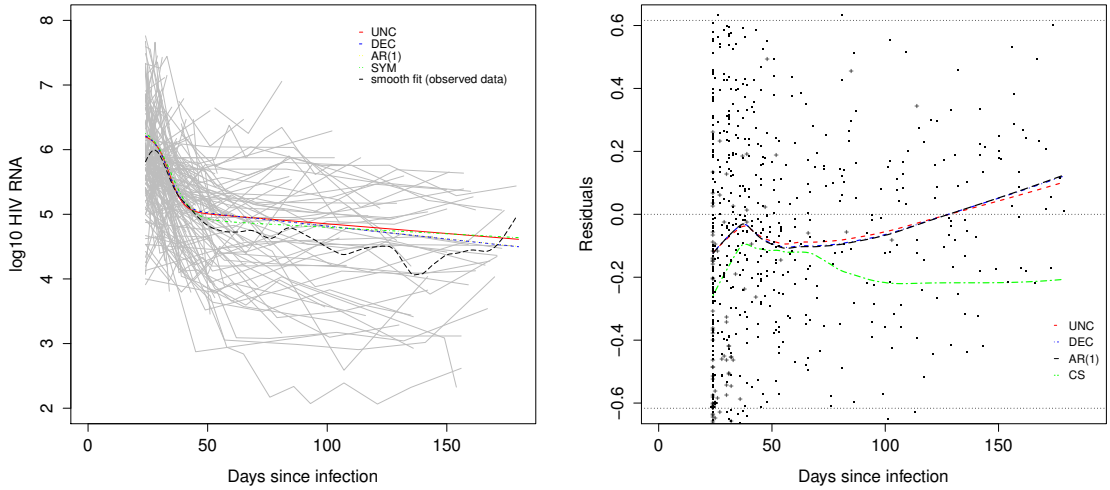

Figure 6 – **AIEDRP data.** (Left panel) Individual profiles (in log10 scale) for HIV viral load at different follow-up times with the model fits. (Right panel) Smoothed means of residuals from model fits. The residuals from the model with autoregressive of order 1 correlation appear as points.

with $i = 1, \ldots, 100$, $j = 1, \ldots, 10$, $\lambda_{1i} = \exp(\beta_1 + b_{1i})$, $\beta_k = \log(\lambda_k)$, $k = 2, 3$, $\lambda_{4i} = \exp(\beta_4 + b_{2i})$, $(b_{1i}, b_{2i}) \overset{ind.}{\sim} \mathrm{N}_2(\mathbf{0}, \mathbf{D})$, and $\epsilon_{ij} \overset{ind.}{\sim} \mathrm{N}_{n_i}(\mathbf{0}, \boldsymbol{\Omega}_i)$.

The parameters are set at $\boldsymbol{\beta} = (1.6094, 0.6931, 3.8067, 2.3026)^\top$, $\sigma^2 = 0.55$, and $\mathbf{D}$ with elements $\alpha_{11} = 0.05$ and $\alpha_{22} = 0.1$.

For the first study, we simulated several datasets considering different values of the parameter ϕ_1 under the correlation structure AR(1), with the aim to discover the effect of the correlation level on the estimation. For each value of ϕ_1, we simulated 100 datasets. In addition, we considered 5% and 20% of censored observations for each value of ϕ_1. Once the simulated datasets were generated, we fitted the proposed model assuming the uncorrelated (UNC) and AR(1) structures. The model selection criteria (AIC and BIC) as well as the estimates of the model parameters were stored for each simulation. Summary statistics such as the mean estimate (MC mean), the mean of the approximate standard error obtained through the information-based method described in Section 2.3.1 (IM SE), the empirical standard error (MC Sd) and the coverage probability at 95% (MC CP) are presented in Tables 6 and 7.

From the results shown in these tables, it can be observed when the AR(1) is chosen as the true model. Moreover, in general, the MC CP values are higher than those obtained under the uncorrelated model, even when the correlation parameter ϕ_1 is small (0.3). It is important to remark that for some values of the correlation parameter, the MC CP values for β_1, β_2 and β_4 obtained under the uncorrelated structure are a bit higher than those obtained in the AR(1) case. This situation occurs in general when the correlation parameter takes high values (0.7, 0.8 and 0.9). Note that although the MC CP remains

Table 6 – **Simulation study (5% censored).** Summary statistics based on 100 simulated AR(1) samples.

			Parameter estimates					Criterion	
ϕ_1	Corr. Structure		β_1	β_2	β_3	β_4	σ^2	MC AIC	MC BIC
	UNC	MC Mean	1.67	0.53	3.73	2.11	0.55	3020	3056
		IM SE	0.04	0.11	0.05	0.20			
		MC Sd	0.02	0.07	0.03	0.15			
		MC CP	79%	78%	66%	89%			
0.3									
	AR(1)	MC Mean	1.61	0.71	3.83	2.27	0.55	3024	3065
		IM SE	0.12	0.22	0.10	0.26			
		MC Sd	0.10	0.22	0.10	0.27			
		MC CP	84%	88%	94%	91%			
	UNC	MC Mean	1.66	0.54	3.74	2.12	0.55	3015	3050
		IM SE	0.04	0.11	0.05	0.20			
		MC Sd	0.02	0.07	0.03	0.16			
		MC CP	82%	80%	67%	91%			
0.5									
	AR(1)	MC Mean	1.60	0.71	3.83	2.27	0.55	3018	3058
		IM SE	0.12	0.23	0.10	0.26			
		MC Sd	0.11	0.23	0.10	0.27			
		MC CP	84%	88%	93%	91%			
	UNC	MC Mean	1.66	0.56	3.74	2.16	0.54	3004	3039
		IM SE	0.04	0.11	0.05	0.20			
		MC Sd	0.02	0.08	0.03	0.17			
		MC CP	82%	86%	69%	92%			
0.6									
	AR(1)	MC Mean	1.60	0.71	3.83	2.27	0.55	3003	3044
		IM SE	0.13	0.23	0.10	0.26			
		MC Sd	0.12	0.23	0.10	0.28			
		MC CP	84%	88%	93%	91%			
	UNC	MC Mean	1.65	0.62	3.73	2.27	0.52	2978	30134
		IM SE	0.04	0.11	0.05	0.19			
		MC Sd	0.02	0.09	0.03	0.18			
		MC CP	90%	91%	68%	94%			
0.7									
	AR(1)	MC Mean	1.59	0.72	3.84	2.27	0.55	2962	3002
		IM SE	0.20	0.25	0.11	0.28			
		MC Sd	0.17	0.25	0.11	0.29			
		MC CP	84%	88%	91%	92%			
	UNC	MC Mean	1.62	0.75	3.74	2.51	0.47	2912	2948
		IM SE	0.04	0.10	0.04	0.17			
		MC Sd	0.03	0.08	0.05	0.17			
		MC CP	99%	96%	50%	76%			
0.8									
	AR(1)	MC Mean	1.60	0.72	3.84	2.27	0.55	2840	2881
		IM SE	0.17	0.25	0.11	0.29			
		MC Sd	0.14	0.25	0.11	0.30			
		MC CP	84%	88%	93%	91%			
	UNC	MC Mean	1.60	0.90	3.73	2.77	0.36	2673	2708
		IM SE	0.03	0.07	0.03	0.14			
		MC Sd	0.04	0.09	0.06	0.16			
		MC CP	95%	17%	13%	12%			
0.9									
	AR(1)	MC Mean	1.61	0.70	3.83	2.26	0.53	2453	2493
		IM SE	0.12	0.21	0.09	0.26			
		MC Sd	0.11	0.22	0.10	0.28			
		MC CP	83%	88%	94%	91%			

Table 7 – **Simulation study (20% censored).** Summary statistics based on 100 simulated AR(1) samples.

			Parameter estimates					Criterion	
ϕ_1	Corr. Structure		β_1	β_2	β_3	β_4	σ^2	MC AIC	MC BIC
	UNC	MC Mean	1.67	0.50	3.72	2.08	0.55	2796	2832
		IM SE	0.04	0.12	0.05	0.21			
		MC Sd	0.02	0.07	0.03	0.16			
		MC CP	68%	69%	63%	87%			
0.3									
	AR(1)	MC Mean	1.59	0.72	3.84	2.28	0.55	2800	2841
		IM SE	0.23	0.26	0.11	0.27			
		MC Sd	0.19	0.26	0.12	0.28			
		MC CP	87%	90%	92%	91%			
	UNC	MC Mean	1.67	0.51	3.72	2.09	0.54	2791	2827
		IM SE	0.04	0.12	0.05	0.21			
		MC Sd	0.02	0.07	0.03	0.16			
		MC CP	76%	73%	63%	87%			
0.5									
	AR(1)	MC Mean	1.60	0.71	3.83	2.27	0.55	2794	2835
		IM SE	0.17	0.25	0.11	0.27			
		MC Sd	0.15	0.25	0.11	0.27			
		MC CP	87%	90%	92%	91%			
	UNC	MC Mean	1.67	0.52	3.72	2.12	0.54	2781	2816
		IM SE	0.04	0.12	0.05	0.21			
		MC Sd	0.02	0.08	0.03	0.18			
		MC CP	78%	77%	64%	88%			
0.6									
	AR(1)	MC Mean	1.60	0.71	3.83	2.27	0.55	2780	2821
		IM SE	0.19	0.26	0.11	0.28			
		MC Sd	0.16	0.25	0.11	0.28			
		MC CP	87%	90%	91%	92%			
	UNC	MC Mean	1.66	0.58	3.72	2.22	0.52	2757	2792
		IM SE	0.04	0.12	0.05	0.20			
		MC Sd	0.02	0.09	0.04	0.20			
		MC CP	85%	87%	67%	92%			
0.7									
	AR(1)	MC Mean	1.59	0.72	3.83	2.27	0.55	2742	2783
		IM SE	0.20	0.27	0.12	0.29			
		MC Sd	0.17	0.26	0.11	0.28			
		MC CP	86%	90%	91%	94%			
	UNC	MC Mean	1.63	0.71	3.72	2.48	0.48	2703	2739
		IM SE	0.04	0.11	0.04	0.18			
		MC Sd	0.03	0.09	0.06	0.21			
		MC CP	96%	99%	45%	80%			
0.8									
	AR(1)	MC Mean	1.62	0.68	3.82	2.24	0.55	2637	2677
		IM SE	0.11	0.23	0.10	0.29			
		MC Sd	0.10	0.21	0.09	0.27			
		MC CP	86%	89%	93%	94%			
	UN	MC Mean	1.61	0.85	3.72	2.73	0.36	2484	2520
		IM SE	0.03	0.07	0.03	0.16			
		MC Sd	0.03	0.10	0.07	0.16			
		MC CP	98%	43%	24%	26%			
0.9									
	AR(1)	MC Mean	1.62	0.67	3.81	2.21	0.53	2290	2331
		IM SE	0.09	0.20	0.09	0.26			
		MC Sd	0.08	0.20	0.09	0.25			
		MC CP	86%	89%	95%	94%			

constant for the AR(1) case, in the UNC case the MC CP presents a lot of variation, going from 12% ($\rho = 0.9$, 5% censoring) to 99% ($\rho = 0.8$, 5% and 20% censoring). One possible explanation for such phenomenon is the effect of the bias in the estimation, particularly in the parameters β_1 and β_2 for high values of the correlation parameter. The biases of fixed effects estimates under the AR(1) structure are lower than those obtained under the uncorrelated structure (see Figures 7 and 8) for different values of the ϕ_1 parameter. The model selection criterion chose the true model (AR(1)) for moderate values of the ϕ_1 parameter (greater than 0.5) for the two levels of censoring considered.

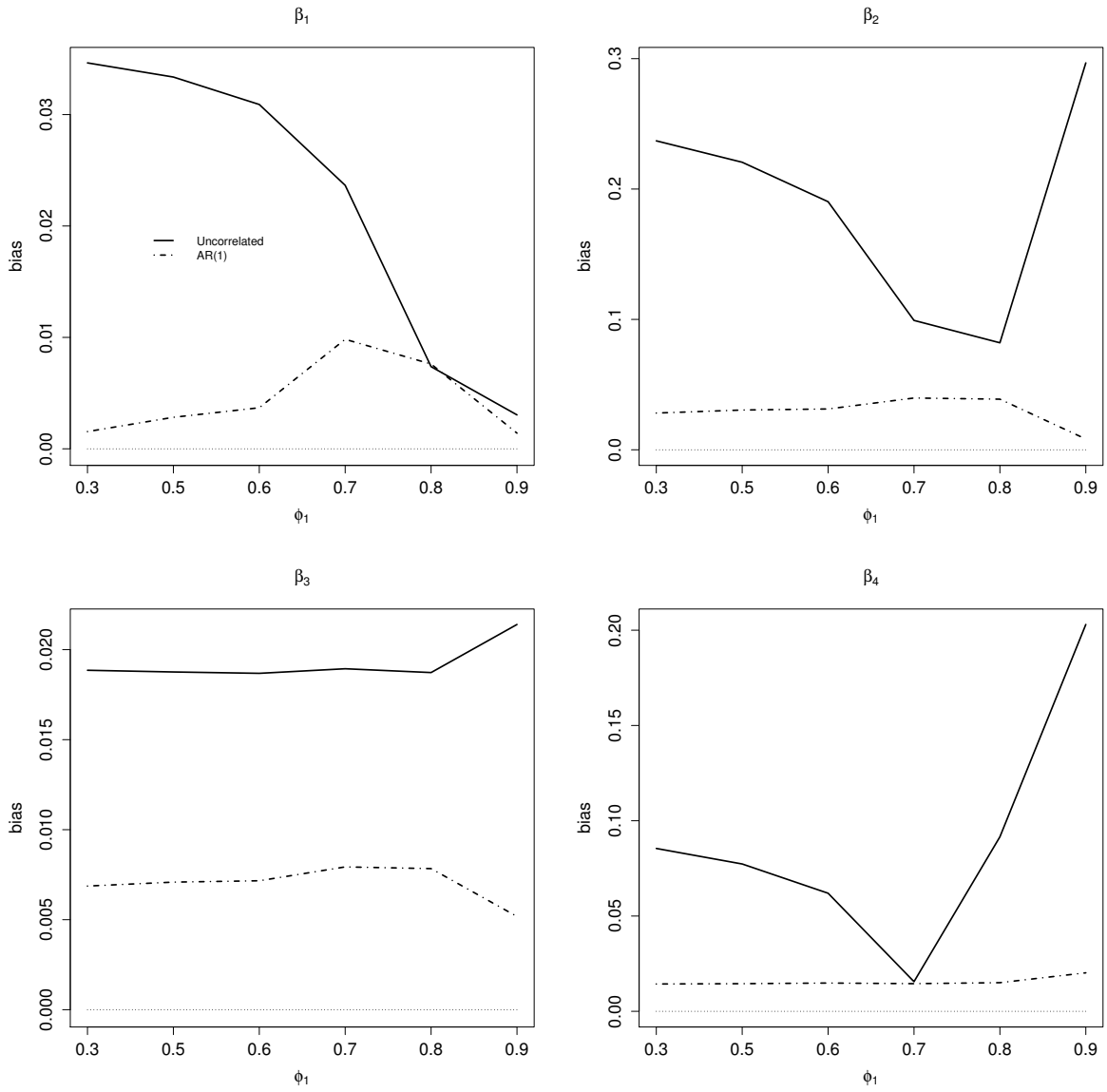

Figure 7 – **Simulation study (5% censored).** Bias of $\boldsymbol{\beta}$ estimates under the uncorrelated and AR(1) models for 6 different values of ϕ_1.

The second simulation study analyzes the performance of the prediction of future values described in Section 2.4. For this purpose, we compared the prediction of the NLMEC model in (2.21) under the UNC and AR(1) structures. As in the first study,

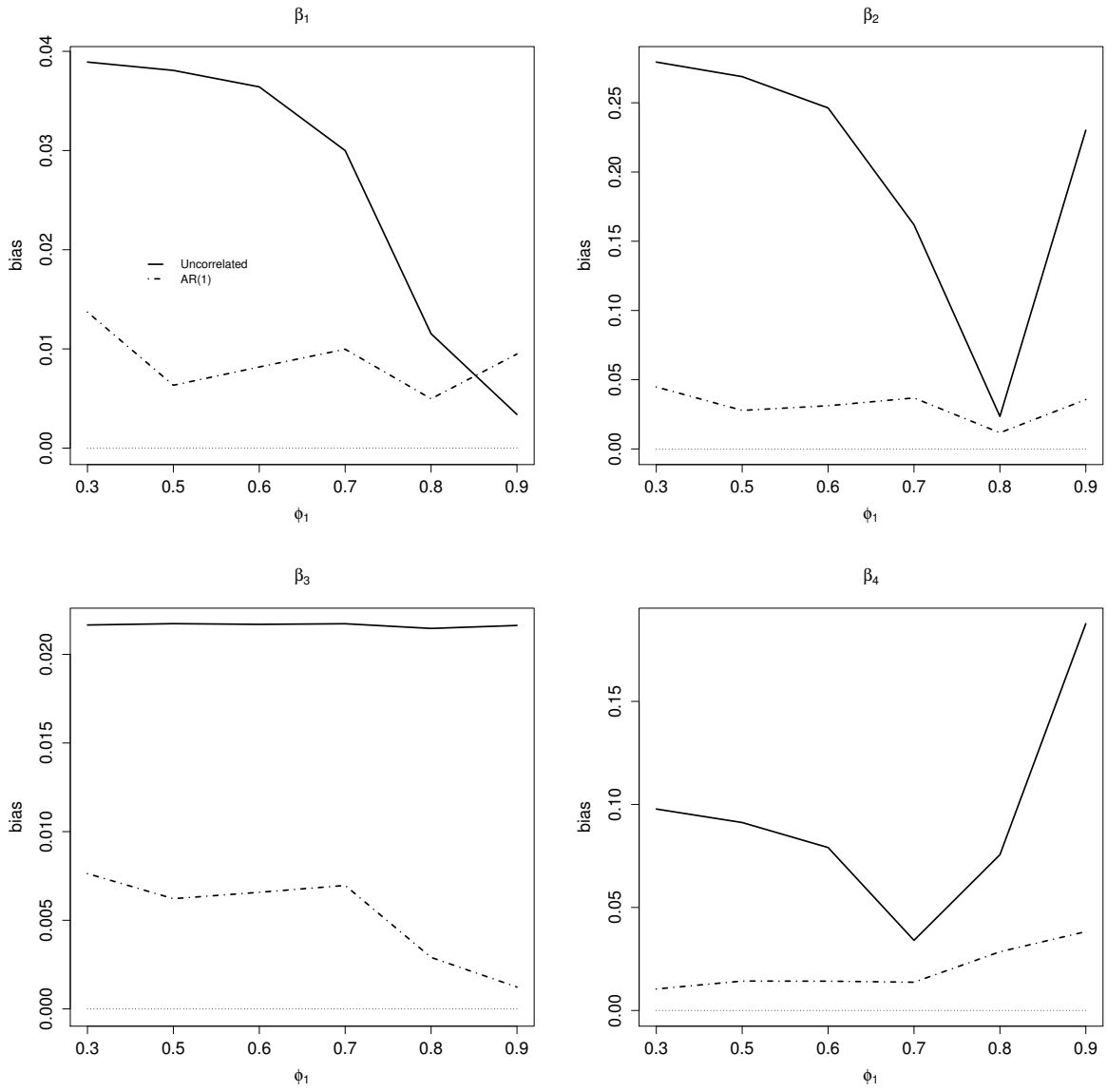

Figure 8 – **Simulation study (20% censored).** Bias of $\boldsymbol{\beta}$ estimates under the uncorrelated and AR(1) models for 6 different values of ϕ_1.

we generated 100 datasets of size $n = 100$ under the AR(1) structure with parameter $\phi_1 = 0.9$, considering two different settings of censoring proportions, 5% and 20%. For the prediction, we excluded the last two measurements of each simulated individual in the datasets. To compare the performance of the prediction, we considered two empirical discrepancy measures, namely the MAE (mean absolute error) and MSE (mean square error). These measures are given by

$$\text{MAE} = \frac{1}{200}\sum_{i,j}|y_{ij} - y_{ij}^{\star}| \quad \text{and} \quad \text{MSE} = \frac{1}{200}\sum_{i,j}(y_{ij} - y_{ij}^{\star})^2, \tag{2.22}$$

where y_{ij} is the original value and $y_{ij}^{\star}$ is the predicted value, for $i = 1, \ldots, 100$ and $j = 1, \ldots, 2$. Table 8 shows the comparison between the predicted values and real ones under the NLMEC model considering the UNC and AR(1) structures. One can see from

these results that the model with AR(1) structure generates predictive values close to the real ones.

Table 8 – **Simulation study.** Evaluation of the prediction accuracy for the NLMEC model with different correlation structures.

	5% censored		20% censored	
Corr. Structure	MAE	MSE	MAE	MSE
UNC	0.5507	0.4739	0.6418	0.6746
AR(1)	0.5169	0.4299	0.6073	0.6165

Finally, we analyzed the absolute bias (Bias) and mean square error (MSE) of the fixed effects and variance components estimates obtained from the DEC-LMEC model with different sample sizes. The idea of this simulation is to provide empirical evidence about the consistency of the ML estimates. The bias and MSE measures are defined as

$$\text{Bias} = \frac{1}{J}\sum_{j=1}^{J}|\widehat{\boldsymbol{\theta}}_i^{(j)} - \boldsymbol{\theta}_i| \quad \text{and} \quad \text{MSE} = \frac{1}{J}\sum_{j=1}^{J}\left(\widehat{\boldsymbol{\theta}}_i^{(j)} - \boldsymbol{\theta}_i\right)^2, \tag{2.23}$$

where $\widehat{\boldsymbol{\theta}}_i^{(j)}$ is the ML estimate of the parameter θ_i for the j-th sample, $j = 1, \ldots, J$.

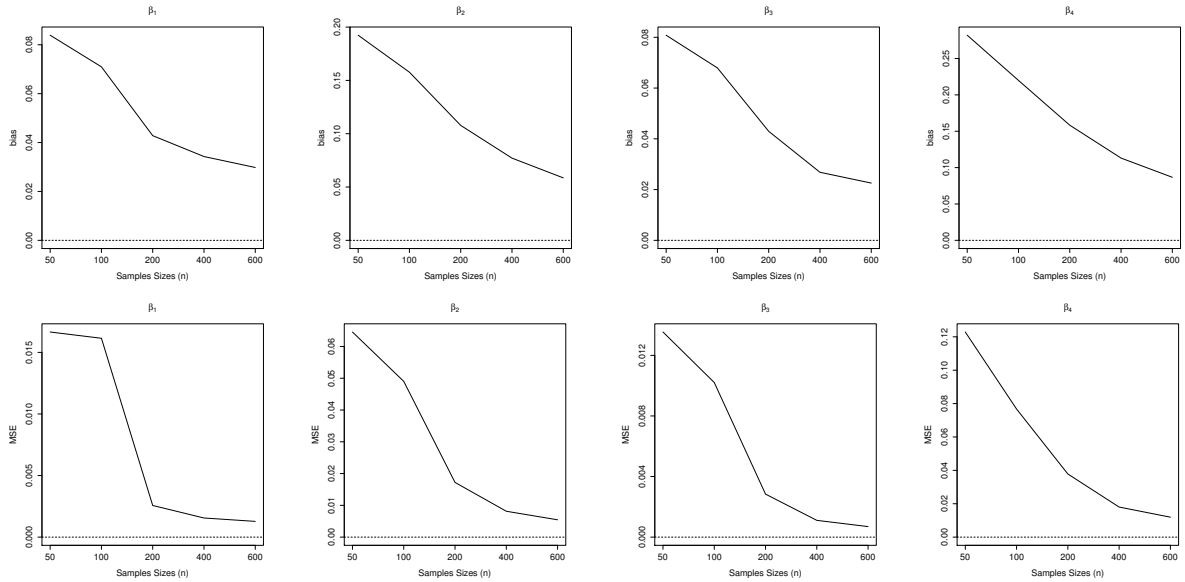

Figure 9 – **Simulation study (10% censored).** Bias and MSE of $\boldsymbol{\beta}$ estimates under the AR(1) model for different sample sizes

The censoring proportion was fixed at 10% and different sample sizes were considered, namely $n = 50, 100, 200, 400$ and 600. Also, we considered $J = 100$, *i.e.*, we simulated 100 samples of size n. For this simulation, an AR (1) structure with parameter $\phi_1 = 0.8$ was considered.

Figures 9 and 10 show that the MSE of the parameter estimates of $\boldsymbol{\beta}$, σ^2 and $\boldsymbol{\alpha}$ tends to zero as the sample size increases. Note that similar results are obtained after

the analysis of the absolute bias. In conclusion, the results provide empirical evidence about the consistency of the ML estimates of the DEC-LMEC model, even considering the linearization procedure described in (2.15).

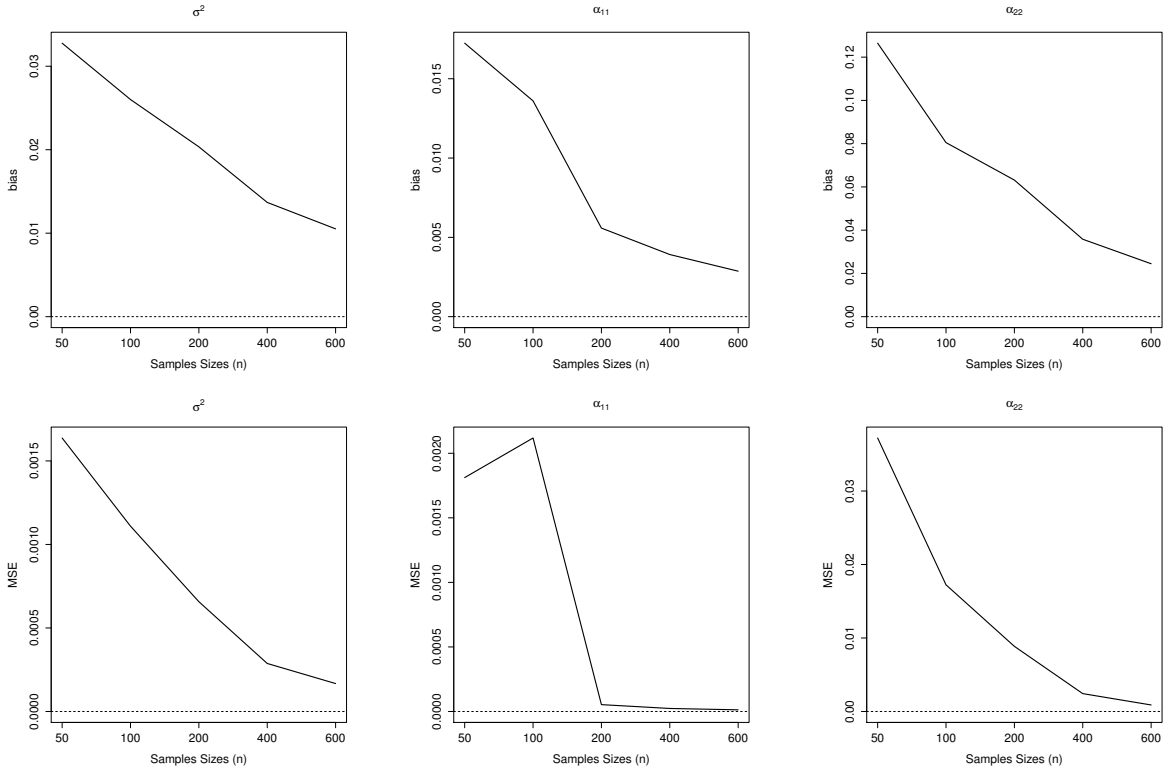

Figure 10 – **Simulation study (10% censored).** Bias and MSE of σ^2 and $\boldsymbol{\alpha}$ estimates under the AR(1) model for different sample sizes

2.8 Conclusions

This chapter proposes a mixed-effects model with censored observations based on the multivariate normal distribution. We adopted a DEC structure as proposed by Muñoz et al. (1992) to model the autocorrelation existing among irregularly observed measures. This structure is flexible, since 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. An ECM algorithm to obtain the ML estimates was developed by using the statistical properties of the multivariate truncated normal distribution. The proposed algorithm has a closed-form expression for the E-step, based on the first two moments of the truncated normal distribution. In this context, the DEC structure can be easily implemented using the exact EM algorithm, making the proposed approach easy to implement by practitioners. The `R` codes are available upon request. The proposed methods were applied on two AIDS case studies and a simulation study was performed, showing the effects of misspecification on the correlation structure over the

fixed effects estimates. The DEC structure provided better results than the uncorrelated structure in terms of parameter estimation and prediction.

Although the LMEC/NLMEC models showed great flexibility to model symmetric data, they can be seriously affected by the presence of outliers. Recently, Garay et al. (2014) proposed a remedy to accommodate outliers using a Student-t regression model with DEC structure. Our methods can be extended by considering the Student-t in the context of LMEC/NLMEC models as in Matos et al. (2015), providing satisfactory results at the expense of additional complexity in implementation. Further, it is also of interest to develop an effective Markov chain Monte Carlo algorithm for the DEC-LMEC/NLMEC in a fully Bayesian treatment.

3 Influence assessment in censored mixed-effects models using the multivariate Student's-t distribution

3.1 Introduction

In AIDS research, the study of the human immunodeficiency virus (HIV) dynamics has received significant attention in the biomedical literature allowing us to understand the pathogenesis of HIV, and assess the effectiveness of the anti-retroviral (ARV) therapy. Most of the clinical trials on ARV therapy assess the rates/changes of viral loads/HIV-1 RNA copies (the amount of actively replicating virus) collected longitudinally over time. The viral load is considered a key primary endpoint because its monitoring is mostly available, a failure in the treatment can be defined virologically, and a new regimen of therapy is recommended as soon as virological rebound occurs (Ndembi et al., 2010). Since the individual viral load trajectories yield large between-subject variations, statistical modeling often focus in formulating the correct linear and nonlinear mixed-effects models (LME/NLME) to estimate these trajectories, and quantify within- and between-subject variations (Wu, 2005; Wu et al., 2010; Qiu and Wu, 2010).

The statistical modelling of viral load can be challenging. First, depending on the diagnostic assays used, the viral load measures may be subjected to upper or lower detection limits (hence, left or right censored), below and above which they are not quantifiable (Wu, 2002). Under non-trivial censoring proportions, considering ad-hoc alternatives (Huang and Dagne, 2011) might lead to bias in fixed effects and variance components estimates. As alternatives to these crude imputation techniques, Vaida and Liu (2009) proposed expectation-maximization (EM) schemes for LME/NLME with censored responses (henceforth LMEC/NLMEC). However, all these methods assume normality of the between-subject random effects and within-subject errors. Even though normality is mostly a reasonable model assumption, it may lack robustness in parameter estimation under departures from normality, namely, presence of heavy tails and outliers (Pinheiro et al., 2001). Censored HIV viral loads do exhibit heavy-tailed behaviour (Lachos et al., 2011). This is also revealed from the raw histogram and QQ plots of viral loads from our motivating example (see Figure 11, panels a and b in Section 3.5.1). Although popular data transformations (say, Box-Cox) might render normality, or close to normality with reasonable empirical results, various issues still persist with these transformations (Lachos et al., 2011). Hence, an appropriate theoretical but 'robust' framework that avoids data

transformation is desirable. A variety of proposals (both classical and Bayesian) exist in this direction that uses the univariate or multivariate Student's-t distribution (Pinheiro et al., 2001; Lin and Lee, 2006, 2007) in the context of LME/NLME models. Some Bayesian propositions in the context of heavy-tailed LMEC/NLMEC models include Lachos et al. (2011) who advocated the use of the normal/independent density (Lange and Sinsheimer, 1993), while Bandyopadhyay et al. (2012, 2015) studied the LMEC model considering both skewness and heavy-tails. Very recently, Matos et al. (2013b) proposed a full maximum-likelihood (ML) based inference using a computationally convenient exact ECM algorithm for the LMEC/NLMEC models using the multivariate Student's-t distribution (henceforth, the t-LMEC/NLMEC model). Here, the E-step yields closed-form expressions, and all parameters are updated in the M-step by considering the random components and the censored observations as missing data.

A vast majority of model development in the literature for LMEC/NLMEC models focus on estimating the mean function. Hence, developing influence diagnostics is a key in assessing the effect of a single observation on the predicted scores for other observations, and consequently the overall parameter estimates, all based on the mean function. Although diagnostics for the traditional normality based LME and LMEC (Matos et al., 2013a) models exist, those for heavy-tailed LMEC/NLMEC models are not well developed. Influence analysis is generally conducted using two primary approaches. The first one is the case-deletion approach (Cook, 1977) based on the well-known Cook's distance. Under normality assumptions for LME, Banerjee and Frees (1997), Hodges (1998), Tan et al. (2001) focused on case-deletion diagnostics for fixed effects, while Christensen et al. (1992) considered a one-step approximation to Cook's distance for the variance components. The other approach is the computationally attractive local influence approach (Cook, 1986), which is a general technique used to assess the stability of the estimation outputs with respect to the model inputs. For elliptical mixed-effects models, this method had been discussed in the literature by Beckman et al. (1987); Lesaffre and Verbeke (1998); Zhu and Lee (2001); Lee and Xu (2004); Osorio et al. (2007); Russo et al. (2009), among others.

Developing influence diagnostics for LMEC/NLMEC models in the spirit of Cook (1977, 1986) leads to the underlying observed log-likelihood functions involving intractable integrals. This renders the direct application of Cook's approach to be very difficult if not impossible, since the measures involve first and second derivatives of these functions. In this context, Zhu and Lee (2001) and Zhu et al. (2001) developed an unified approach for performing local influence and case-deletion diagnostics, respectively, for general missing data models based on the Q-function, *i.e.*, the conditional expectation of the complete-data log likelihood at the E-step in the EM algorithm. This was extended to generalized linear and NLME models by Lee and Xu (2004) and Xu et al. (2006), respectively. This Q-function approach produces result similar to those obtained using

the Cook's approach. Recently, Matos et al. (2013a) used this Q-function approach for developing influence diagnostics for LMEC/NLMEC models. Stemming from the same difficulty with intractable integrals (for example, the *pdfs* of truncated multivariate Student's-t distributions) in implementing the Cook's diagnostics for the t-LMEC/NLMEC model of Matos et al. (2013b), we develop case-deletion and influence diagnostics measures using the approach of Zhu et al. (2001) (see also Lee and Xu, 2004).

The rest of this chapter is organised as follows. Section 3.2 develops the t-LMEC model specification and an EM-type algorithm for ML estimation. Section 3.3 presents the global and local influence approaches for the t-LMEC model. For local influence, various perturbation schemes for both subject- and observation-level diagnostics are considered. In Section 3.4, the t-NLMEC model is defined. The methodology is illustrated in Section 3.5 using a motivating HIV dataset. Section 3.6 presents a simulation study evaluating the efficiency of our method in detecting outliers under various degrees of data perturbation and censoring. Finally, Section 3.7 presents some concluding remarks, with some possible directions for future research.

3.2 Censored linear mixed-effect model

Ignoring censoring for the moment, the t-LME model of Matos et al. (2013b) is specified as:

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

where

$$\begin{pmatrix} \mathbf{b}_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \overset{ind.}{\sim} t_{n_i+q}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \sigma^2\mathbf{I}_{n_i} \end{pmatrix}, \nu\right), i = 1, \ldots, n,$$

which implies that, marginally,

$$\mathbf{b}_i \overset{iid}{\sim} t_q(\mathbf{0}, \mathbf{D}, \nu) \quad \text{and} \quad \boldsymbol{\epsilon}_i \overset{ind.}{\sim} t_{n_i}(\mathbf{0}, \sigma^2\mathbf{I}_{n_i}, \nu), \quad i = 1, \ldots, n, \tag{3.2}$$

where $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ denotes the *pdf* of a multivariate Student's-t distribution with location vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$ and degrees of freedom ν. The subscript i refers to the subject index; $\mathbf{I}_p$ denotes the $p \times p$ identity matrix; $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$ is a vector of observed continuous responses for subject i of dimension $n_i \times 1$; $\mathbf{X}_i$ is the $n_i \times p$ design matrix associated with the $p\times 1$ vector of fixed-effects $\boldsymbol{\beta}$; $\mathbf{Z}_i$ is the $n_i \times q$ design matrix corresponding to the $q \times 1$ vector of random effects $\mathbf{b}_i$; $\boldsymbol{\epsilon}_i$ is the $(n_i \times 1)$ vector of random errors and the random effects dispersion matrix $\mathbf{D} = \mathbf{D}(\boldsymbol{\alpha})$ depends on unknown parameters $\boldsymbol{\alpha}$. Following Matos et al. (2013b), we consider the case where the response Y_{ij} is not fully observed for all i, j. Consequently, the observed data for the i-th subject is $(\mathbf{V}_i, \mathbf{C}_i)$, where $\mathbf{V}_i$ is the vector of censoring level and $\mathbf{C}_i$ is the vector of censoring indicators such that

$$\begin{aligned} y_{ij} &\leqslant V_{ij} \quad \text{if} \quad C_{ij} = 1, \\ y_{ij} &= V_{ij} \quad \text{if} \quad C_{ij} = 0. \end{aligned} \tag{3.3}$$

For simplicity, we assume that the data are left censored. Extensions to other arbitrary censoring patterns are immediate.

3.2.1 The likelihood function

Classical inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top, \nu)^\top$ is based on the marginal distribution of $\mathbf{y}_i$, $\mathbf{y}_i \sim t_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i, \nu)$, for $i = 1, \ldots, n$, where $\boldsymbol{\Sigma}_i = \sigma^2\mathbf{I}_{n_i} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top$. For computing the likelihood function associated with model (3.1)–(3.3), 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, $\mathbf{y}_i$, $\mathbf{V}_i$, $\mathbf{X}_i$, and $\boldsymbol{\Sigma}_i$ can be partitioned as: $\mathbf{y}_i = vec(\mathbf{y}_i^o, \mathbf{y}_i^c)$, $\mathbf{V}_i = vec(\mathbf{V}_i^o, \mathbf{V}_i^c)$, $\mathbf{X}_i^\top = (\mathbf{X}_i^o, \mathbf{X}_i^c)$ 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 $vec(\cdot)$ 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 and Bolfarine, 1995), we have $\mathbf{y}_i^o \sim t_{n_i^o}(\mathbf{X}_i^o\boldsymbol{\beta}, \boldsymbol{\Sigma}_i^{oo}, \nu)$, and $\mathbf{y}_i^c|\mathbf{y}_i^o, \sim t_{n_i^c}(\boldsymbol{\mu}_i^{co}, \mathbf{S}_i^{co}, \nu + n_i^o)$, where

$$\boldsymbol{\mu}_i^{co} = \mathbf{X}_i^c\boldsymbol{\beta} + \boldsymbol{\Sigma}_i^{co}\boldsymbol{\Sigma}_i^{oo-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\boldsymbol{\beta}), \;\; \mathbf{S}_i^{co} = \left(\frac{\nu + Q(\mathbf{y}_i^o)}{\nu + n_i^o}\right)\boldsymbol{\Sigma}_i^{cc.o}, \tag{3.4}$$

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

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

where $T_p(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ denotes the cumulative distribution function (*cdf*) of the multivariate Student's-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). As in Matos et al. (2013b), we assume the degrees of freedom and the shape parameters for Student-t to be fixed, and we use a model selection procedure based on AIC or BIC to choose the most appropriate value of ν. Then, we consider that the parameter vector is $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top)^\top$.

3.2.2 The EM algorithm

The observed log-likelihood function involves complex expressions, making it very difficult to work directly with $\ell(\boldsymbol{\theta}|\mathbf{y})$, either for the ML estimation, or the corresponding influence analysis. As mentioned above, Matos et al. (2013b) developed an EM-type algorithm for the t-LMEC/NLMEC models by treating $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$,

and $\mathbf{u} = (u_1, \ldots, u_n)^\top$ as hypothetical missing data, and augmenting those to the observed data vector $(\mathbf{V}, \mathbf{C})$, where $\mathbf{V} = vec(\mathbf{V}_1, \ldots, \mathbf{V}_n)$, and $\mathbf{C} = vec(\mathbf{C}_1, \ldots, \mathbf{C}_n)$. Thus, the resulting complete data is $\mathbf{y}_c = (\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{y}^\top, \mathbf{b}^\top, \mathbf{u}^\top)^\top$, and the EM-type algorithm is applied to the complete data log-likelihood function $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}|\mathbf{y}_c)$, where

$$\begin{aligned} \ell_i(\boldsymbol{\theta}|\mathbf{y}_c) &= -\frac{1}{2}\Big[n_i \log\sigma^2 + \frac{u_i}{\sigma^2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i) \\ &+ \log|\mathbf{D}| + u_i\mathbf{b}_i^\top\mathbf{D}^{-1}\mathbf{b}_i\Big] + h(u_i|\nu) + C, \end{aligned}$$

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 a current value $\widehat{\boldsymbol{\theta}}^{(k)}$ of $\boldsymbol{\theta}$, the Q function (the conditional expectation of the complete data log-likelihood function) is given by

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^{n} Q_{1i}(\boldsymbol{\beta}, \sigma^2|\widehat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^{n} Q_{2i}(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}^{(k)}), \tag{3.5}$$

where

$$\begin{aligned} Q_{1i}(\boldsymbol{\beta}, \sigma^2|\widehat{\boldsymbol{\theta}}^{(k)}) &= -\frac{n_i}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\Big[\widehat{a}_i^{(k)} - 2\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top(\widehat{u\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{u\mathbf{b}}_i^{(k)}) \\ &+ \widehat{u}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\Big] \quad \text{and} \\ Q_{2i}(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}^{(k)}) &= -\frac{1}{2}\log|\mathbf{D}| - \frac{1}{2}\mathrm{tr}\left(\widehat{u\mathbf{b}_i^2}^{(k)}\mathbf{D}^{-1}\right). \end{aligned}$$

Here, $\widehat{a}_i^{(k)} = \mathrm{tr}\Big(\widehat{u\mathbf{y}_i^2}^{(k)} - 2\widehat{u\mathbf{y}\mathbf{b}_i}^{(k)}\mathbf{Z}_i^\top + \widehat{u\mathbf{b}_i^2}^{(k)}\mathbf{Z}_i^\top\mathbf{Z}_i\Big)$; $\widehat{u\mathbf{b}_i^2}^{(k)} = E\{u_i\mathbf{b}_i\mathbf{b}_i^\top|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\} = \widehat{\sigma^2}^{(k)}\widehat{\boldsymbol{\Lambda}}_i^{(k)} + \widehat{\boldsymbol{\varphi}}_i^{(k)}(\widehat{u\mathbf{y}_i^2}^{(k)} - \widehat{u\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\widehat{u\mathbf{y}}_i^{(k)\top} + \widehat{u}_i^{(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top)\widehat{\boldsymbol{\varphi}}_i^\top$; $\widehat{u\mathbf{b}}_i^{(k)} = E\{u_i\mathbf{b}_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\} = \widehat{\boldsymbol{\varphi}}_i^{(k)}(\widehat{u\mathbf{y}}_i^{(k)} - \widehat{u}_i^{(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)})$; $\widehat{u\mathbf{y}\mathbf{b}_i}^{(k)} = E\{u_i\mathbf{y}_i\mathbf{b}_i^\top|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\} = (\widehat{u\mathbf{y}_i^2}^{(k)} - \widehat{u\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top)\widehat{\boldsymbol{\varphi}}_i^\top$, with $\widehat{\boldsymbol{\Lambda}}_i^{(k)} = (\widehat{\sigma^2}^{(k)}\widehat{\mathbf{D}}^{-1(k)} + \mathbf{Z}_i^\top\mathbf{Z}_i)^{-1}$ and $\widehat{\boldsymbol{\varphi}}_i^{(k)} = \widehat{\boldsymbol{\Lambda}}_i^{(k)}\mathbf{Z}_i^\top$.

It is easy to observe that the E-step reduces to the computation of $\widehat{u\mathbf{y}_i^2} = E\{u_i\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\}$, $\widehat{u\mathbf{y}}_i = E\{u_i\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\}$, and $\widehat{u}_i = E\{u_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\}$. These expected values are available in closed form using Propositions available in Matos et al. (2013b).

Next, the conditional maximization step (CM-step) maximizes $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ conditionally with respect to $\boldsymbol{\theta}$ to obtain new estimates $\widehat{\boldsymbol{\theta}}^{(k+1)}$ as follows:

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \left(\sum_{i=1}^{n}\widehat{u}_i^{(k)}\mathbf{X}_i^\top\mathbf{X}_i\right)^{-1}\sum_{i=1}^{n}\mathbf{X}_i^\top\left(\widehat{u\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{u\mathbf{b}}_i^{(k)}\right), \tag{3.6}$$

$$\widehat{\sigma^2}^{(k+1)} = \frac{1}{N}\sum_{i=1}^{n}\left[\widehat{a}_i^{(k)} - 2\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top(\widehat{u\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{u\mathbf{b}}_i^{(k)}) + \widehat{u}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\right], \tag{3.7}$$

$$\widehat{\mathbf{D}}^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\widehat{u\mathbf{b}_i^2}^{(k)}, \tag{3.8}$$

where $N = \sum_{i=1}^{n} n_i$, and the scale matrix $\mathbf{D}$ is unstructured with $\boldsymbol{\alpha}$ the upper triangular elements of $\mathbf{D}$. The algorithm is iterated until the distance involving two successive

evaluations of the log-likelihood $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)})/\ell(\widehat{\boldsymbol{\theta}}^{(k)}) - 1|$ is sufficiently small. Here, we do not focus on the ML estimation, and the interested might refer to Matos et al. (2013b) for further details. In the following section, we derive influence diagnostic measures, given the ML estimate $\widehat{\boldsymbol{\theta}}$.

3.3 Influence analysis

Influence diagnostics are routinely used in statistical modelling to identify aberrant observations and assess their impact on model fitting and parameter estimation. Recognising the difficulties following the Cook (1977, 1986)'s approach (described in Section 1), we use the Q-function of Zhu et al. (2001) to develop case-deletion measures, leading to the influence measures for the t-LMEC model.

3.3.1 Global influence

The case-deletion approach is a commonly used scheme to study the effects of deleting the ith case/observation from the data set. Henceforth, the subscript '$[i]$' will denote the original data set with the ith case deleted. Consequently, the log-likelihood function corresponding to the remaining data is denoted by $\ell(\boldsymbol{\theta}|\mathbf{Y}_{c[i]})$. In order to assess the influence of the ith case on the ML estimate $\widehat{\boldsymbol{\theta}}$, we need to compare the difference between $\widehat{\boldsymbol{\theta}}_{[i]}$ and $\widehat{\boldsymbol{\theta}}$, where $\widehat{\boldsymbol{\theta}}_{[i]} = (\widehat{\boldsymbol{\beta}}_{[i]}^\top, \widehat{\sigma}^2_{[i]}, \widehat{\boldsymbol{\alpha}}_{[i]}^\top)^\top$ is the maximizer of the function $Q_{[i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = E\{\ell(\boldsymbol{\theta}|\mathbf{Y}_{c[i]})|\mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}\}$, with $\widehat{\boldsymbol{\theta}}$ being the ML estimate of $\boldsymbol{\theta}$. An observation is regarded as influential if its deletion generates considerable influence on model estimates. In other words, if $\widehat{\boldsymbol{\theta}}_{[i]}$ is fairly far from $\widehat{\boldsymbol{\theta}}$, then the ith observation could considered as influential. Note that, since the estimator $\widehat{\boldsymbol{\theta}}_{[i]}$ is needed for every case, this scheme requires a considerable computational effort, particularly for large sample sizes. For that reason, a one-step approximation (see Cook and Weisberg, 1982; Zhu et al., 2001) is used to reduce the burden. This approximation follows:

$$\widehat{\boldsymbol{\theta}}^1_{[i]} = \widehat{\boldsymbol{\theta}} + \{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1}\dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \tag{3.9}$$

where $\dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \dfrac{\partial Q_{[i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$, and $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \dfrac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ represents the Hessian matrix, $i = 1, \ldots, n$, with its elements given by

$$\dot{Q}_{[i]\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \partial Q_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\beta} = \frac{1}{\widehat{\sigma}^2}E_{1[i]}, \tag{3.10}$$

$$\dot{Q}_{[i]\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \partial Q_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})/\partial\sigma^2 = -\frac{1}{2\widehat{\sigma}^2}E_{2[i]}, \tag{3.11}$$

$$\dot{Q}_{[i]\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \partial Q_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\alpha}, \tag{3.12}$$

where $E_{1[i]} = \sum_{j\neq i}\mathbf{X}_j^\top(\widehat{u\mathbf{y}}_j - \mathbf{Z}_j\widehat{u\mathbf{b}}_j - \widehat{u}_j\mathbf{X}_j\widehat{\boldsymbol{\beta}})$ and $E_{2[i]} = \sum_{j\neq i}(n_j - \dfrac{A_j}{\widehat{\sigma}^2})$, with $A_j = \mathrm{tr}(\widehat{u\mathbf{y}_j^2} - 2\widehat{u\mathbf{y}\mathbf{b}_j}\mathbf{Z}_j^\top + \widehat{u\mathbf{b}_j^2}\mathbf{Z}_j^\top\mathbf{Z}_j) - 2\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_j^\top(\widehat{u\mathbf{y}}_j - \mathbf{Z}_j\widehat{u\mathbf{b}}_j) + \widehat{u}_j\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_j^\top\mathbf{X}_j\widehat{\boldsymbol{\beta}}$. Finally, the elements of

$\dot{Q}_{[i]\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ are of the form

$$\dot{Q}_{[i]\alpha_r}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = -\frac{1}{2}\sum_{j\neq i}\mathrm{tr}[\mathbf{D}^{-1}\dot{\mathbf{D}}(r) - \mathbf{D}^{-1}\dot{\mathbf{D}}(r)\mathbf{D}^{-1}\widehat{u\mathbf{b}_j^2}],\ r = 1,\ldots,p^*,\ \text{with } p^* = \dim(\boldsymbol{\alpha}).$$

It is necessary to compute the Hessian matrix $\ddot{Q}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n}\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ to develop case-deletion, local influence and any particular perturbation schemes, following Zhu and Lee (2001). The Hessian matrix $\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ has the following elements:

$$\begin{aligned}
\frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} &= -\frac{1}{\sigma^2}\mathbf{X}_i^\top\widehat{u}_i\mathbf{X}_i, \quad \frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\sigma^2} = -\frac{1}{\sigma^4}\mathbf{X}_i^\top(\widehat{u\mathbf{y}}_i - \mathbf{Z}_i\widehat{u\mathbf{b}}_i - \widehat{u}_i\mathbf{X}_i\boldsymbol{\beta}),\\
\frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\alpha_r} &= \mathbf{0}, \quad \frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\sigma^2} = \frac{1}{2\sigma^4}[n_i - \frac{2}{\sigma^2}A_i],\\
\frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\alpha_r} &= 0, \quad \frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\alpha_s\partial\alpha_r} = \frac{1}{2}\mathrm{tr}(\mathbf{A}(sr)) - \frac{1}{2}\mathrm{tr}(\mathbf{B}(sr)\widehat{u\mathbf{b}_i^2}),
\end{aligned}$$

where $\mathbf{A}(sr) = \mathbf{D}^{-1}[\dot{\mathbf{D}}(s)\mathbf{D}^{-1}\dot{\mathbf{D}}(r) - \ddot{\mathbf{D}}(s,r)]$ and $\mathbf{B}(sr) = \mathbf{D}^{-1}[\dot{\mathbf{D}}(s)\mathbf{D}^{-1}\dot{\mathbf{D}}(r) + \dot{\mathbf{D}}(r)\mathbf{D}^{-1}\dot{\mathbf{D}}(s) - \ddot{\mathbf{D}}(s,r)]\mathbf{D}^{-1}$, with $\dot{\mathbf{D}}(r) = \partial\mathbf{D}/\partial\alpha_r$, $\ddot{\mathbf{D}}(s,r) = \partial^2\mathbf{D}/\partial\alpha_s\partial\alpha_r, r, s = 1,\ldots,p^*$, $p^* = \dim(\boldsymbol{\alpha})$ and $i = 1,\ldots,n$. After some rearrangement and evaluating these derivatives at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, we obtain the Hessian matrix $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ (see Appendix A.1) as block-diagonal of the form $\ddot{Q}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \mathrm{diag}(\ddot{Q}_\beta(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \ddot{Q}_{\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \ddot{Q}_\alpha(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}))$ (the normal case given in Matos et al., 2013a), where $\ddot{Q}_\beta(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = -\frac{1}{\widehat{\sigma^2}}\sum_{i=1}^{n}\mathbf{X}_i^\top\widehat{u}_i\mathbf{X}_i$, $\ddot{Q}_{\sigma^2}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = b/2(\widehat{\sigma^2})^2$ and $\ddot{Q}_{\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n}\partial^2 Q_i(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})/\partial\alpha_s\partial\alpha_r$, with $\mathbf{X} = (\mathbf{X}_1^\top,\ldots,\mathbf{X}_n^\top)^\top$ and $b = \sum_{i=1}^{n}(n_i - 2A_i/\widehat{\sigma^2})$. Using (3.9), the next result proposes the one-step pseudo approximation of $\widehat{\boldsymbol{\theta}}_{[i]} = (\widehat{\boldsymbol{\theta}}_{[i]}^\top, \widehat{\sigma^2}_{[i]}, \widehat{\boldsymbol{\alpha}}_{[i]}^\top)^\top$, $i = 1,\ldots,n$. Its proof is straightforward and is therefore omitted.

Proposição 3.1. *The one-step pseudo approximation for the parameter estimates of the t-LMEC model with the ith case deleted is given by*

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{[i]}^1 &= \widehat{\boldsymbol{\beta}} + (\sum_{i=1}^{n}\mathbf{X}_i^\top\widehat{u}_i\mathbf{X}_i)^{-1}E_{1[i]}\\
\widehat{\sigma^2}_{[i]}^1 &= \widehat{\sigma^2}\left(1 + \frac{E_{2[i]}}{b}\right)\\
\widehat{\boldsymbol{\alpha}}_{[i]}^1 &= \widehat{\boldsymbol{\alpha}} + \{-\ddot{Q}_{\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1}\dot{Q}_{[i]\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})
\end{aligned}$$

where $E_{1[i]}$, $E_{2[i]}$ and $\dot{Q}_{[i]\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ are as in (3.10), (3.11) and (3.12) respectively, $b = \sum_{i=1}^{n}(n_i - 2A_i/\widehat{\sigma^2})$ and $\ddot{Q}_{\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n}\partial^2 Q_i(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})/\partial\alpha_s\partial\alpha_r$.

Note that Proposition 3.1 allows a straightforward influence assessment via the case-deletion approach for the t-LMEC model. One needs to compute the ML estimate $\widehat{\boldsymbol{\theta}}$

for the complete data, the ML estimate $\widehat{\boldsymbol{\theta}}_{[i]}$ with the ith case deleted, and compare both estimates using some metric such as the Cook's or likelihood distance. If the difference between them is fairly large, then the ith case is regarded as influential. The generalized Cook distance (Zhu and Lee, 2001) is defined as

$$GD_i(\boldsymbol{\theta}) = (\widehat{\boldsymbol{\theta}}_{[i]} - \widehat{\boldsymbol{\theta}})^\top \{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}(\widehat{\boldsymbol{\theta}}_{[i]} - \widehat{\boldsymbol{\theta}}), i = 1, \ldots, n, \tag{3.13}$$

substituting (3.9) into (3.13), we have the approximation

$$GD_i^1(\boldsymbol{\theta}) = \dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}})^\top \{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1} \dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}),\ i = 1, \ldots, n.$$

Since $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ is a diagonal matrix, this approximation can be written as $GD_i^1(\boldsymbol{\theta}) = \sum_{k=1}^{p} GD_i^1(\theta_k)$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top$ (for details see Xie et al., 2007). Consequently, for our t-LMEC model we have

$$GD_i^1(\boldsymbol{\theta}) = GD_i^1(\boldsymbol{\beta}) + GD_i^1(\sigma^2) + GD_i^1(\boldsymbol{\alpha}). \tag{3.14}$$

3.3.2 Local Influence

In this section, we consider local influence analysis (Cook, 1986) focusing on the following perturbation schemes: the case-weight, scale matrix and response perturbation. Here, we consider both subject-level and observation-level diagnostics. The subject-level diagnostics identify if a subject is considered influential or not, and is carried out considering a perturbation function for the ith subject. However, in modelling longitudinal data, we have two level of responses, namely, the subject-level and observation level, and intuitively, an influential subject may/may not contain influential observations (Pan et al., 2014). Hence, exploring atypical observations at both levels are warranted. The observation-level diagnostics consider a perturbation in the jth observation of the ith subject.

The theoretical developments in this section proceed in the framework of Cook (1986) and Zhu and Lee (2001). Let $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_g)^\top$ be a perturbation vector varying in an open region $\boldsymbol{\Omega} \subset \mathbb{R}^g$ and $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{y}_c)$, the complete-data log-likelihood with respect to the perturbed model induced by $\boldsymbol{\omega}$. We assume there exists $\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}$, such that $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}_0|\mathbf{y}_c) = \ell_c(\boldsymbol{\theta}|\mathbf{y}_c)$ for all $\boldsymbol{\theta}$. The Q-displacement function $f_Q(\boldsymbol{\omega})$ is defined as $f_Q(\boldsymbol{\omega}) = 2\left[Q\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right) - Q\left(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})|\widehat{\boldsymbol{\theta}}\right)\right]$, where $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ is the maximum of the function $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \mathrm{E}[\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{y}_c)|\mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}]$. The local behavior of the Q-displacement function can be analyzed by using the normal curvature $C_{f_Q,\mathbf{d}}$ of $\boldsymbol{\alpha}(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, f_Q(\boldsymbol{\omega}))^\top$ at $\boldsymbol{\omega}_0$ in the direction of some unit vector $\mathbf{d}$. It follows that

$$C_{f_Q,\mathbf{d}} = -2\mathbf{d}^\top \ddot{Q}_{\boldsymbol{\omega}_o}\mathbf{d} \text{ and } -\ddot{Q}_{\boldsymbol{\omega}_0} = \boldsymbol{\Delta}_{\boldsymbol{\omega}_0}^\top \left\{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\right\}^{-1} \boldsymbol{\Delta}_{\boldsymbol{\omega}_0},$$

where $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \dfrac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ and $\boldsymbol{\Delta}_{\boldsymbol{\omega}} = \dfrac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}$. For our t-LMEC model, we consider $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0} = (\boldsymbol{\Delta}_\beta^\top, \boldsymbol{\Delta}_{\sigma^2}^\top, \boldsymbol{\Delta}_\alpha^\top)^\top$, where $\boldsymbol{\Delta}_\beta = \dfrac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^\top}|\boldsymbol{\omega}_o$, $\boldsymbol{\Delta}_{\sigma^2} = \dfrac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\boldsymbol{\omega}^\top}|\boldsymbol{\omega}_o$ and $\boldsymbol{\Delta}_\alpha = (\boldsymbol{\Delta}_{\alpha 1}^\top, \ldots, \boldsymbol{\Delta}_{\alpha p^*}^\top)^\top$, with $\boldsymbol{\Delta}_{\alpha r} = \dfrac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\alpha_r\partial\boldsymbol{\omega}^\top}|\boldsymbol{\omega}_o,\ r = 1, \ldots, p^*$.

3.3.2.1 Subject-level diagnostics

Case weight perturbation

We consider an arbitrary attribution of weights for the expected value of the complete-data log-likelihood function (perturbed Q-function), which may capture departures in general directions, by writing

$$Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \mathrm{E}[\ell_c(\boldsymbol{\theta},\boldsymbol{\omega}|\mathbf{y}_c)|\mathbf{V},\mathbf{C},\widehat{\boldsymbol{\theta}}] = \sum_{i=1}^{n}\omega_i \mathrm{E}[\ell_i(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{V},\mathbf{C},\widehat{\boldsymbol{\theta}}] = \sum_{i=1}^{n}\omega_i Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}).$$

Here, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ is an $n \times 1$ vector and $\boldsymbol{\omega}_o = (1, \ldots, 1)^\top$. Note that the local influence analysis for this perturbation scheme is equivalent to the case-deletion approach discussed in Section 3.3.1 (see Appendix A.2). Under this perturbation scheme, we have $\boldsymbol{\Delta}_\beta = \dfrac{1}{\sigma^2}\mathbf{X}^\top D(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n)$, $\boldsymbol{\Delta}_{\sigma^2} = -\dfrac{1}{2\sigma^2}\mathbf{n}^\top + \dfrac{1}{2\sigma^4}\mathbf{m}^\top$, $\boldsymbol{\Delta}_{\alpha_r} = [\dfrac{\partial Q_1(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\alpha_r}, \ldots, \dfrac{\partial Q_n(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\alpha_r}]$ for $r = 1, \ldots, p^*$, where $\mathbf{n} = (n_1, \ldots, n_n)^\top$, $\mathbf{m} = (A_1, \ldots, A_n)^\top$, $D(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n)$ is a block-diagonal matrix, with $\boldsymbol{\epsilon}_i = \widehat{u\mathbf{y}}_i - \mathbf{Z}_i\widehat{u\mathbf{b}}_i - \hat{u}_i\mathbf{X}_i\widehat{\boldsymbol{\beta}}$ and $\dfrac{\partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\alpha_r} = -\dfrac{1}{2}\mathrm{tr}[\mathbf{D}^{-1}\dot{\mathbf{D}}(r) - \mathbf{D}^{-1}\dot{\mathbf{D}}(r)\mathbf{D}^{-1}\widehat{u\mathbf{b}_i^2}]$.

Scale matrix perturbation

In order to study the effects of perturbation on the scale matrix $\boldsymbol{\Sigma}_i = \sigma^2\mathbf{I}_{n_i} + \mathbf{Z}\mathbf{D}\mathbf{Z}_i^\top$, we consider $\mathbf{D}(\omega_i) = \omega_i^{-1}\mathbf{D}$, or $\sigma^2(\omega_i) = \omega_i^{-1}\sigma^2$, for $i = 1, \ldots, n$. The non-perturbed model arises when $\boldsymbol{\omega}_o = (1, \ldots, 1)^\top$. The perturbed Q-function follows (3.5), with $\mathbf{D}(\omega_i)$ and $\sigma^2(\omega_i)$ in place of $\mathbf{D}$ and σ^2, respectively. Considering a perturbation on $\mathbf{D}$ (matrix of random effects), we have $\boldsymbol{\Delta}_\beta = \mathbf{0}$, $\boldsymbol{\Delta}_{\sigma^2} = \mathbf{0}$ and $\boldsymbol{\Delta}_{\alpha_r} = \dfrac{1}{2}[g_1, \ldots, g_n]$, where $g_i = \mathrm{tr}(\mathbf{D}^{-1}\dot{\mathbf{D}}(r)\mathbf{D}^{-1}\widehat{u\mathbf{b}_i^2})$, $r = 1, \ldots, p^*$. Perturbation on σ^2 (the random error variance) yields $\boldsymbol{\Delta}_\beta = \dfrac{1}{\sigma^2}\mathbf{X}^\top D(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n)$, $\boldsymbol{\Delta}_{\sigma^2} = \dfrac{1}{2\sigma^4}\mathbf{m}^\top$ and $\boldsymbol{\Delta}_{\boldsymbol{\alpha}} = \mathbf{0}$.

Response perturbation

A general way for perturbing the response variables Q_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, n_i$, is introduced by considering $Q_{ij}(\omega) = Q_{ij} + \omega_i s_{ij}$, where s_{ij} is a known constant. Hence, for the t-LMEC model, the perturbed response is obtained as $y_{ij}(\omega) \leqslant Q_{ij}$ if $C_{ij} = 1$, and $y_{ij}(\omega) = Q_{ij}$ if $C_{ij} = 0$, where $\mathbf{y}_{ij}(\omega) = \mathbf{y}_{ij} - \omega_i s_{ij}$. Again, the perturbed

Q-function follows (3.5), with $\widehat{u\mathbf{y}}_i$, $\widehat{u\mathbf{y}_i^2}$ and $\widehat{u\mathbf{y}\mathbf{b}_i}$ replaced by $\widehat{u\mathbf{y}_{i\omega}} = \widehat{u\mathbf{y}}_i - \omega_i\mathbf{s}_i\widehat{u}_i$, $\widehat{u\mathbf{y}_{i\omega}^2} = \widehat{u\mathbf{y}_i^2} - \omega_i(\widehat{u\mathbf{y}}_i\mathbf{s}_i^\top + \mathbf{s}_i\widehat{u\mathbf{y}_i^\top}) + \omega_i^2\mathbf{s}_i\mathbf{s}_i^\top$ and $\widehat{u\mathbf{y}\mathbf{b}_{i\omega}} = \widehat{u\mathbf{y}\mathbf{b}_i} - \omega_i\mathbf{s}_i\widehat{u\mathbf{b}_i^\top}$, respectively, where $\mathbf{s}_i = (s_{i1}, \ldots, s_{in_i})^\top$. The vector $\boldsymbol{\omega}_0 = \mathbf{0}$ represents no perturbation. Finally, we have $\boldsymbol{\Delta}_\beta = -\frac{1}{\sigma^2}[\mathbf{X}_1^\top\widehat{u}_1\mathbf{s}_1, \ldots, \mathbf{X}_n^\top\widehat{u_n}\mathbf{s}_n]$, $\boldsymbol{\Delta}_{\sigma^2} = -\frac{1}{\sigma^4}[(\widehat{u\mathbf{y}_1} - \mathbf{Z}_1\widehat{u\mathbf{b}_1} - \widehat{u}_1\mathbf{X}_1\boldsymbol{\beta})^\top\mathbf{s}_1, \ldots, (\widehat{u\mathbf{y}_n} - \mathbf{Z}_n\widehat{u\mathbf{b}_n} - \widehat{u_n}\mathbf{X}_n\boldsymbol{\beta})^\top\mathbf{s}_n]$, and $\boldsymbol{\Delta}_\alpha = \mathbf{0}$.

3.3.2.2 Observation-level diagnostics

We proceed as above considering a perturbation vector $\boldsymbol{\omega} = (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_g)^\top$, where $\boldsymbol{\omega}_i = (\omega_{i1}, \ldots, \omega_{in_i})^\top$, and noting that all the previous results for the subject-level diagnostics hold for the observation-level cases as well. Also, we denote $\mathbf{u}_i = (u_{i1}, \ldots, u_{in_i})^\top$, $\mathbf{v}_i = (v_{i1}, \ldots, v_{in_i})^\top$ and $\mathbf{g}_i = (g_{i1}, \ldots, g_{in_i})^\top$.

Case weight perturbation

In this case, we have $\boldsymbol{\Delta}_\beta = \frac{1}{\sigma^2}[\mathbf{u}_1, \ldots, \mathbf{u}_n]$, with $u_{ij} = X_{ij}^\top(\widehat{u\mathbf{y}}_{ij} - \mathbf{Z}_{ij}\widehat{u\mathbf{b}}_i - \widehat{u}_i\mathbf{X}_{ij}\widehat{\boldsymbol{\beta}})$; $\boldsymbol{\Delta}_{\sigma^2} = -\frac{1}{2\sigma^2}[\mathbf{v}_1, \ldots, \mathbf{v}_n]$ with $v_{ij} = 1 - \frac{1}{\sigma^2}A_{ij}$ and $A_{ij} = \mathrm{tr}(\widehat{u\mathbf{y}_{ij}^2} - 2\widehat{u\mathbf{y}\mathbf{b}_{ij}}\mathbf{Z}_{ij}^\top + \widehat{u\mathbf{b}_i^2}\mathbf{Z}_{ij}^\top\mathbf{Z}_{ij}) - 2\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_{ij}^\top(\widehat{u\mathbf{y}}_{ij} - \mathbf{Z}_{ij}\widehat{u\mathbf{b}}_i) + \widehat{u}_i\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_{ij}^\top\mathbf{X}_{ij}\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\Delta}_{\alpha_r} = -\frac{1}{2}[\mathbf{g}_1, \ldots, \mathbf{g}_n]$, with $g_{ij} = \mathrm{tr}(\mathbf{D}^{-1}\dot{\mathbf{D}}(r)\mathbf{D}^{-1}(\mathbf{D} - \widehat{u\mathbf{b}_i^2}))$, $r = 1, \ldots, p^*$.

Scale matrix perturbation

Similar to the subject-level, we consider perturbations on $\mathbf{D}$ and σ^2. Consequently, for $\mathbf{D}$ we have that $\boldsymbol{\Delta}_\beta = \mathbf{0}$, $\boldsymbol{\Delta}_{\sigma^2} = \mathbf{0}$ and $\boldsymbol{\Delta}_{\alpha_r} = \frac{1}{2}[\mathbf{g}_1, \ldots, \mathbf{g}_n]$, with $g_{ij} = \mathrm{tr}(\mathbf{D}^{-1}\dot{\mathbf{D}}(r)\mathbf{D}^{-1}\widehat{u\mathbf{b}_i^2})$, $r = 1, \ldots, p^*$. In addition, a perturbation on σ^2 generates $\boldsymbol{\Delta}_\beta = \frac{1}{\sigma^2}[\mathbf{u}_1, \ldots, \mathbf{u}_n]$, with $u_{ij} = X_{ij}^\top(\widehat{u\mathbf{y}}_{ij} - \mathbf{Z}_{ij}\widehat{u\mathbf{b}}_i - \widehat{u}_i\mathbf{X}_{ij}\widehat{\boldsymbol{\beta}})$; $\boldsymbol{\Delta}_{\sigma^2} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$, with $v_{ij} = \frac{1}{2\sigma^4}A_{ij}$ and $A_{ij} = \mathrm{tr}(\widehat{u\mathbf{y}_{ij}^2} - 2\widehat{u}_i\mathbf{X}_{ij}\widehat{\boldsymbol{\beta}} + \widehat{u\mathbf{b}_i^2}\mathbf{Z}_{ij}^\top\mathbf{Z}_{ij}) - 2\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_{ij}^\top(\widehat{u\mathbf{y}}_{ij} - \mathbf{Z}_{ij}\widehat{u\mathbf{b}}_i) + \widehat{u}_i\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_{ij}^\top\mathbf{X}_{ij}\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\Delta}_\alpha = \mathbf{0}$.

Response perturbation

Finally, for the response perturbation case, we have $\boldsymbol{\Delta}_\beta = -\frac{1}{\sigma^2}[\mathbf{u}_1, \ldots, \mathbf{u}_n]$, with $u_{ij} = X_{ij}^\top$; $\boldsymbol{\Delta}_{\sigma^2} = -\frac{1}{\sigma^4}[\mathbf{v}_1, \ldots, \mathbf{v}_n]$, with $v_{ij} = (\widehat{u\mathbf{y}}_{ij} - \mathbf{Z}_{ij}\widehat{u\mathbf{b}}_i - \widehat{u}_i\mathbf{X}_{ij}\widehat{\boldsymbol{\beta}})$ and $\boldsymbol{\Delta}_{\alpha_r} = \mathbf{0}$.

As the reader can note, it is impossible to give details for all perturbation schemes that would be of interest. However, if we can find an appropriate $\boldsymbol{\omega}$ such that the perturbed complete data log-likelihood function $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{y}_c)$ is smooth enough and the pertinent derivatives in the diagnostic measures are well-defined, we can conduct the local

influence analysis without much difficulty.

In order to quantify the influence of a case in the data, we follow the method based on the function $M(0)_l = \sum_{k=1}^{r} \tilde{\zeta}_k \boldsymbol{\varepsilon}_{kl}^2$, where $\tilde{\zeta}_k = \zeta_k/(\zeta_1 + \ldots + \zeta_r)$ and $\boldsymbol{\varepsilon}_k^2 = (\boldsymbol{\varepsilon}_{k1}^2, \ldots, \boldsymbol{\varepsilon}_{kg}^2)^\top$ with $\{(\zeta_k, \boldsymbol{\varepsilon}_k), k = 1, \ldots, g\}$ the eigenvalue-eigenvector pairs of $-2\ddot{Q}_{\boldsymbol{\omega}_0}$, where $\zeta_1 \geqslant \ldots \geqslant \zeta_r > \zeta_{r+1} = \ldots = 0$ and the eigenvectors $\{\boldsymbol{\varepsilon}_k, k = 1, \ldots, g\}$ are orthonormal (for details see Matos et al., 2013a). The l*th* case may be regarded as influential if $M(0)_l$ is larger than the benchmark (cut-off).

Based on the work of Zhu and Lee (2001), we use the following conformal normal curvature $B_{f_Q,\mathbf{d}}(\boldsymbol{\theta}) = C_{f_Q,\mathbf{d}}(\boldsymbol{\theta})/tr[-2\ddot{Q}_{\boldsymbol{\omega}_0}]$, whose computation is quite simple and also has the property that $0 \leqslant B_{f_Q,\mathbf{d}}(\boldsymbol{\theta}) \leqslant 1$. Let $\mathbf{d}_l$ be a basic perturbation vector with lth entry as 1 and all other entries as zero. Zhu and Lee (2001) showed that for all l, $M(0)_l = B_{f_Q,\mathbf{d}_l}$. Thus, we can obtain $M(0)_l$ via $B_{f_Q,\mathbf{d}_l}$. Following Lee and Xu (2004), we consider our benchmark as $\overline{M}(0) + c^* SM(0)$, where $\overline{M}(0)$ and $SM(0)$ are the mean and standard error of $\{M(0)_l : l = 1, \ldots, g\}$ respectively; and c^* is a selected constant. The choice of c^* is subjective. In this work, we will consider $c^* = 4$; following Russo et al. (2009) and Zeller et al. (2010).

3.4 Censored nonlinear mixed-effects model

In this section, we develop the censored nonlinear mixed-effects model under the Student's-t distribution (henceforth, t-NLMEC). Similar to the t-LMEC model, we denote the number of subjects by n, and the number of measurements on the ith subject by n_i. Ignoring censoring for the moment, let us consider x_{ij} the vector incorporating explanatory variables (covariates), the longitudinal time component t_{ij}, $\boldsymbol{\beta}_{ij} = (\beta_{1ij}, \ldots, \beta_{sij})^\top$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top (p > s)$. The Student's-$t$ nonlinear mixed-effect model (t-NLME model) can be written as:

$$\mathbf{y}_i = \eta_i(t_{ij}, \boldsymbol{\beta}_{ij}) + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\beta}_{ij} = d(x_{ij}, \boldsymbol{\beta}, \mathbf{b}_i), \tag{3.15}$$

where $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$, with y_{ij} the response for subject i at time t_{ij}, $\eta_i(t_{ij}, \boldsymbol{\beta}_{ij}) = (\eta(t_{i1}, \boldsymbol{\beta}_{i1})^\top, \ldots, \eta(t_{in_i}, \boldsymbol{\beta}_{in_i}))^\top$, with $\eta(\cdot)$ being a nonlinear (known) but differentiable function of vector-valued mixed-effects parameters $\boldsymbol{\beta}_{ij}$, $\boldsymbol{\epsilon} = (\epsilon_{i1}, \ldots, \epsilon_{in_i})^\top$ is the random error vector, $d(\cdot)$ is an s-dimensional linear function, and $\mathbf{b}_i = (b_{1i}, \ldots, b_{qi})^\top$ is the vector of random effects ($q \leqslant s$). The joint distribution of $(\mathbf{b}_i, \boldsymbol{\epsilon}_i)$ follows (3.2). From Matos et al. (2013b), the marginal distribution is given by

$$f(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=1}^{n} \int_0^\infty \int_{\mathbb{R}^q} \phi_{n_i}(\mathbf{y}_i, \eta_i(t_{ij}, d(x_{ij}, \boldsymbol{\beta}, \mathbf{b}_i)), u_i^{-1}\sigma^2 \mathbf{I}_{n_i}) \phi_q(\mathbf{b}_i; 0, u_i^{-1}\mathbf{D}) \times G(u_i|\nu/2, \nu/2) d\mathbf{b}_i du_i,$$

where $G(\cdot|a,b)$ denotes the density of a Gamma(a,b) distribution with mean a/b. The marginal distribution $f(\mathbf{y}|\boldsymbol{\theta})$ does not have a closed form because the model function is not linear in the random effects. However, in order to use all the theory on influence diagnostics developed above for the LMEC model, we use the following approximation proposed by Matos et al. (2013b) which linearizes the t-NLMEC likelihood in terms of $\mathbf{b}_i$ and $\boldsymbol{\beta}$.

Proposição 3.2. *Let $\widetilde{\mathbf{b}}_i$ and $\widetilde{\boldsymbol{\beta}}$ be expansion points in the neighborhood of $\mathbf{b}_i$ and $\boldsymbol{\beta}$, respectively. Then, the t-NLME model as defined in (3.2) and (3.15) has the following t-LME form*

$$\widetilde{\mathbf{y}}_i = \widetilde{\mathbf{W}}_i\boldsymbol{\beta} + \widetilde{\mathbf{H}}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1,\ldots,n, \tag{3.16}$$

where $\widetilde{\mathbf{y}}_i = \mathbf{y}_i - \widetilde{\eta}_i(\widetilde{\boldsymbol{\beta}},\widetilde{\mathbf{b}}_i)$, $\mathbf{b}_i \overset{ind}{\sim} t_q(0,\mathbf{D},\nu)$, $\boldsymbol{\epsilon}_i \overset{ind.}{\sim} t_{n_i}(\mathbf{0},\sigma^2\mathbf{I}_{n_i},\nu)$, $\widetilde{\mathbf{H}}_i = \frac{\partial\eta_i(t_{ij},d(x_{ij},\widetilde{\boldsymbol{\beta}},\mathbf{b}_i))}{\partial\mathbf{b}_i^\top}|_{\mathbf{b}_i=\widetilde{\mathbf{b}}_i}$, $\widetilde{\mathbf{W}}_i = \frac{\partial\eta_i(t_{ij},d(x_{ij},\boldsymbol{\beta},\widetilde{\mathbf{b}}_i))}{\partial\boldsymbol{\beta}_i^\top}|_{\boldsymbol{\beta}_i=\widetilde{\boldsymbol{\beta}}_i}$ *and* $\widetilde{\eta}(\widetilde{\boldsymbol{\beta}},\widetilde{\mathbf{b}}_i) = \eta_i(t_{ij},d(x_{ij},\widetilde{\boldsymbol{\beta}},\widetilde{\mathbf{b}}_i)) - \widetilde{\mathbf{H}}_i\widetilde{\mathbf{b}}_i - \widetilde{\mathbf{W}}_i\widetilde{\boldsymbol{\beta}}$.

Proof: Based on the first-order Taylor expansion of the function η_i around $\widetilde{\mathbf{b}}_i$ and $\widetilde{\boldsymbol{\beta}}$, we have that

$$\eta_i(t_{ij},d(x_{ij},\boldsymbol{\beta},\mathbf{b}_i)) \approx [\eta_i(t_{ij},d(x_{ij},\widetilde{\boldsymbol{\beta}},\widetilde{\mathbf{b}}_i)) + \widetilde{\mathbf{H}}_i\mathbf{b}_i - \widetilde{\mathbf{H}}_i\widetilde{\mathbf{b}}_i + \widetilde{\mathbf{W}}_i\boldsymbol{\beta} - \widetilde{\mathbf{W}}_i\widetilde{\boldsymbol{\beta}}]$$

with $\widetilde{\mathbf{H}}_i = \frac{\partial\eta_i(t_{ij},d(x_{ij},\boldsymbol{\beta},\mathbf{b}_i))}{\partial\mathbf{b}_i^\top}|_{\mathbf{b}_i=\widetilde{\mathbf{b}}_i}$ and $\widetilde{\mathbf{W}}_i = \frac{\partial\eta_i(t_{ij},d(x_{ij},\boldsymbol{\beta},\mathbf{b}_i))}{\partial\boldsymbol{\beta}^\top}|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}$. It follows that

$$\begin{aligned}
\boldsymbol{\epsilon}_i &= \mathbf{y}_i - \eta_i(t_{ij},d(x_{ij},\boldsymbol{\beta},\mathbf{b}_i)) \\
&\approx \mathbf{y}_i - [\eta_i(t_{ij},d(x_{ij},\widetilde{\boldsymbol{\beta}},\widetilde{\mathbf{b}}_i)) + \widetilde{\mathbf{H}}_i\mathbf{b}_i - \widetilde{\mathbf{H}}_i\widetilde{\mathbf{b}}_i + \widetilde{\mathbf{W}}_i\boldsymbol{\beta} - \widetilde{\mathbf{W}}_i\widetilde{\boldsymbol{\beta}}] \\
&= \mathbf{y}_i - [\widetilde{\eta}(\widetilde{\boldsymbol{\beta}},\widetilde{\mathbf{b}}_i) + \widetilde{\mathbf{W}}_i\boldsymbol{\beta} + \widetilde{\mathbf{H}}_i\mathbf{b}_i] = \widetilde{\mathbf{y}}_i - [\widetilde{\mathbf{W}}_i\boldsymbol{\beta} + \widetilde{\mathbf{H}}_i\mathbf{b}_i],
\end{aligned}$$

which concludes the proof.

For the censored case, model (3.16) is a t-LMEC model with the same structure as (3.1)-(3.3). The model matrices in (3.16) depend on the current parameter value, and need to be recalculated at each iteration. The algorithm iterates between the L-, E- and CM-steps until convergence. Moreover, the influence diagnostics for t-LMEC discussed earlier in Section 3.3 can be incorporated along with the approximation in (3.16) to obtain approximate influence diagnostics for t-NLMEC.

The approximation (3.16) was initially proposed in Matos et al. (2013a) in the context of censored nonlinear mixed-effects models. In particular, simulation studies in that paper revealed that this approximation can efficiently detect outliers contaminating the generated data. More recently, Wang and Lin (2014) used this approximation to implement an efficient ECM algorithm for carrying out ML estimation in Student's-t nonlinear mixed-effects models for multi-outcome longitudinal data with missing values. Consequently, we conclude that this approximation is robust, stable, and we do not anticipate any severe consequences in inference when applied to other types of (censored) nonlinear models.

3.5 Application

3.5.1 AIEDRP Dataset

In this section, we consider the study from the AIEDRP program in order to illustrate the proposed influence analysis. We consider 320 untreated individuals with HIV infection (see Vaida and Liu, 2009, for more detalis). The dataset consists of 830 observations, with 185 (22%) lying above the limit of assay quantification. The individual profiles are shown in Figure 3. As was proposed in Vaida and Liu (2009), we consider a right-censored five-parameter NLMEC model as follows:

$$y_{ij} = \lambda_{1i} + \frac{\lambda_2}{1 + \exp((t_{ij} - \lambda_3)/\lambda_4)} + \lambda_{5i}(t_{ij} - 50) + \epsilon_{ij}, \tag{3.17}$$

where y_{ij} is the $\log_{10}$ of the viral load for subject i at time t_{ij}. The parameters λ_{1i} and λ_2 represent the subject-specific random setpoints value and decrease from the maximum HIV RNA, respectively. In the absence of treatment (following acute infection), the HIV RNA varies around a setpoint, which may differ among individuals; hence the setpoint is chosen to be subject specific. The location parameter λ_3 indicates the time point at which half of the change in HIV RNA is attained, λ_4 is a scale parameter modelling the rate of decline and λ_{5i} allows for increasing HIV RNA trajectory after day 50. The reparameterization given by $\beta_{1i} = \log(\lambda_{1i}) = \beta_1 + b_{1i}$; $\beta_k = \log(\lambda_k)$, $k = 2, 3, 4$, and $\lambda_{5i} = \beta_5 + b_{2i}$ is adopted to assure positive values for the model parameters. Figure 11 (Panels left and right) presents

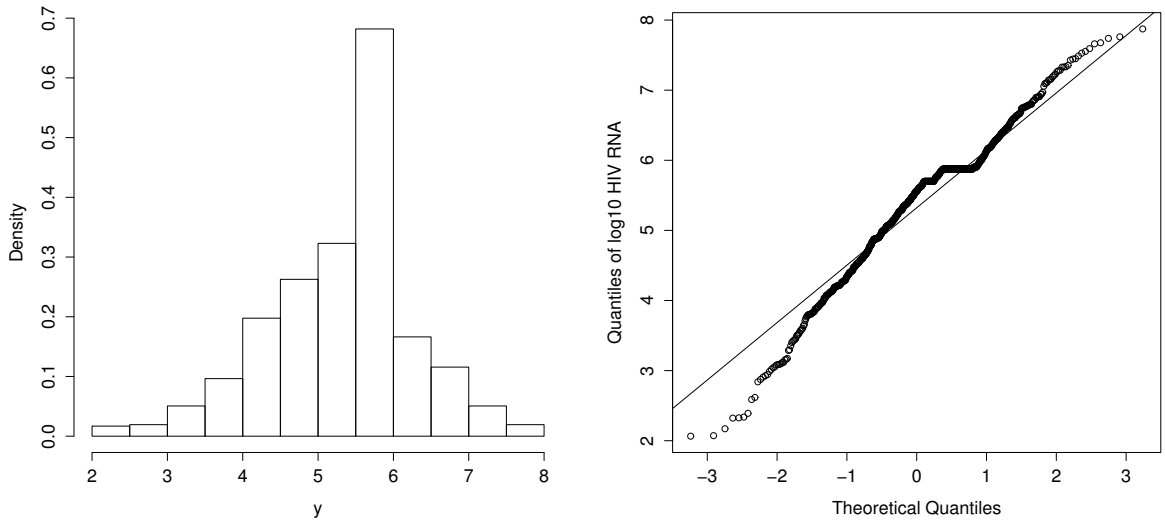

Figure 11 – **AIEDRP data.** Plots of raw density histogram (Left panel) and Q-Q plot (Right panel) of viral load.

raw histogram and Q-Q plot of the log viral load measures, respectively. These plots reveal that viral loads exhibit heavy-tail behaviour, and presence of possible outliers. Hence, to

accommodate these features, we fit the t-NLMEC model defined in (3.15) considering the structure given in (3.17).

3.5.2 ML estimates using EM algorithm

The model fitting uses the approximated ML method given in Proposition 3.2 and the ECM algorithm presented in Section 3.2.2. The degrees of freedom ν is assumed to be known. Using the AIC criterion, we choose $\nu = 10$ which maximises the profile log-likelihood (see, Figure 12, Left panel). This reveals that a fit using a normality-based LMEC model might be inadequate. Further model comparison between the normal and t-NLMEC models using the AIC/BIC criteria presented in Table 9 show that the t-NLMEC model provided a much improved model fit than the normal one.

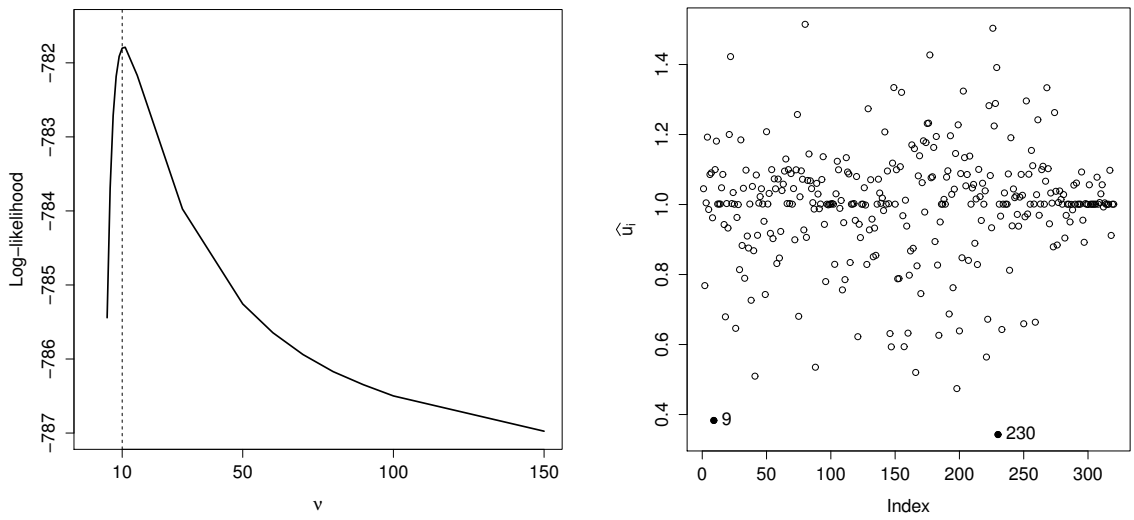

Figure 12 – **AIEDRP data.** Plot of the profile log-likelihood versus the degrees of freedom ν (Left panel), and estimated weight $\hat{u}_i$ for the t-NLMEC fit (Right panel), with the influential observations numbered.

Because we currently focus on exploring influence diagnostics, details on the estimation and interpretation of the parameter estimates $\boldsymbol{\beta}$ are omitted for brevity. From Figure 12 (Right panel), we observe that the t-NLMEC model insulates the overall parameter estimation by assigning smaller weights $\hat{u}_i$ to the possible influential observations, which are described later in more details.

3.5.3 Global influence

In order to evaluate the effect on the ML estimates when some observation is deleted, we analyse the $GD_i^1(\boldsymbol{\theta})$ plot in Figure 13 (Panel a). The plot reveals that two cases (#195, #259) are potentially influential on the parameter estimates. Figures 13 (Panels

Table 9 – **AIEDRP data.** ML estimates and model comparison criteria for normal and t-NLMEC models. SE are the estimated asymptotic standard errors.

	N-NLMEC		t-NLMEC	
Parameter	MLE	SE	MLE	SE
β_1	1.6093	0.0137	1.6109	0.0133
β_2	0.1449	0.0953	0.1636	0.0854
β_3	3.5256	0.0237	3.5233	0.0207
β_4	1.0599	0.2666	0.9910	0.2450
β_5	-0.0035	0.0015	-0.0031	0.0015
σ^2	0.2621		0.2053	
α_{11}	0.01766		0.01611	
α_{12}	0.00017		0.00014	
α_{22}	0.00005		0.00005	
ν			10	
log-like	-783.8905		-781.8017	
AIC	1585.7812		1581.6034	
BIC	1628.2740		1624.0963	

b-d) present plots of $GD_i^1(\boldsymbol{\beta})$, $GD_i^1(\sigma^2)$ and $GD_i^1(\boldsymbol{\alpha})$ respectively, using Proposition 3.1. From these figures, we infer that subject #195 is influential for $\boldsymbol{\beta}$, #9 and #230 are influential for σ^2, and #259 is influential for α.

3.5.4 Local influence

Next, we focus on the local influence analysis for the dataset based on $M(0)$, with interest focussing on $\boldsymbol{\theta}$. We study both the subject-level and observation-level diagnostics. It is important to stress that in local influence analysis, there are no general rules so far for selecting the benchmark (Lee and Xu, 2004). Hence, we follow the criterion suggested by Lee and Xu (2004), *i.e* $M(0)_i > \overline{M}(0) + 3.5SM(0)$, $i = 1, \ldots, 320$, to discriminate whether an observation is influential or not.

3.5.4.1 Subject-level diagnostics

Figure 14 presents the index plots of $M(0)$ under the perturbation schemes discussed in Section 3.3.2.1. We find that subjects #195 and #259 appears influential under case weight perturbation scheme. Moreover, subjects #133 and #159 are potentially influential under perturbation on $\mathbf{D}$. For perturbation on σ^2, we find that observations #166, #195 and #259 appear as influential. Finally, for response variable perturbation, observations #174, #175 and #176 are considered as potentially influential.

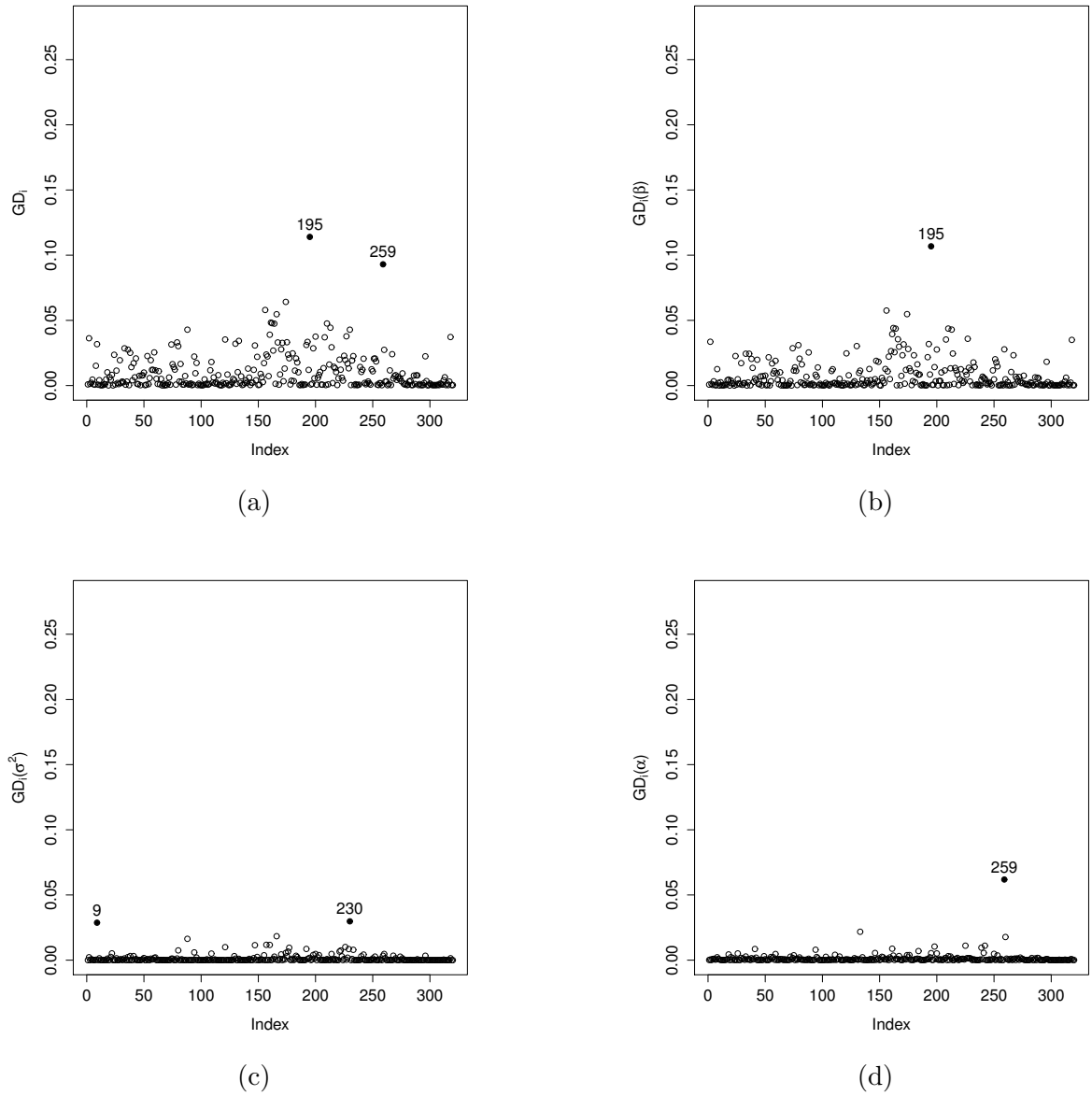

Figure 13 – **AIEDRP data.** Global influence. Approximate generalized Cook's distance $GD_i^1(\boldsymbol{\theta})$ (Panel a), GD_i^1 for subset $\boldsymbol{\beta}$ (Panel b), GD_i^1 for subset σ^2 (Panel c), and GD_i^1 for subset $\boldsymbol{\alpha}$ (Panel d). The influential observations are numbered.

To assess the individual impact of these possible influential observations on the ML estimates, we refitted the t-NLMEC model multiple times by removing one of the following observations: 9, 133, 166, 174, 175, 176 ,195, 230 and 259, identified as possibly influential, each time. Table 10 presents the % relative changes (RC) in the parameter estimates presented in Table 9 compared to the parameter estimates obtained after removing the influential observations. Specifically, the RC measure is defined as $\text{RC}_{\widehat{\delta}} = \left|\dfrac{\widehat{\delta} - \widehat{\delta}_{[i]}}{\widehat{\delta}}\right|$, where $\delta = \beta_1, \ldots, \beta_5, \sigma^2$, $\boldsymbol{\alpha}$ and $\widehat{\delta}_{[i]}$ denotes the ML estimate of δ with the ith observation removed. From Table 10, we observe that these observations generates greater changes in the RC, in particularly for parameters β_2, α_{12} and α_{22}. These findings

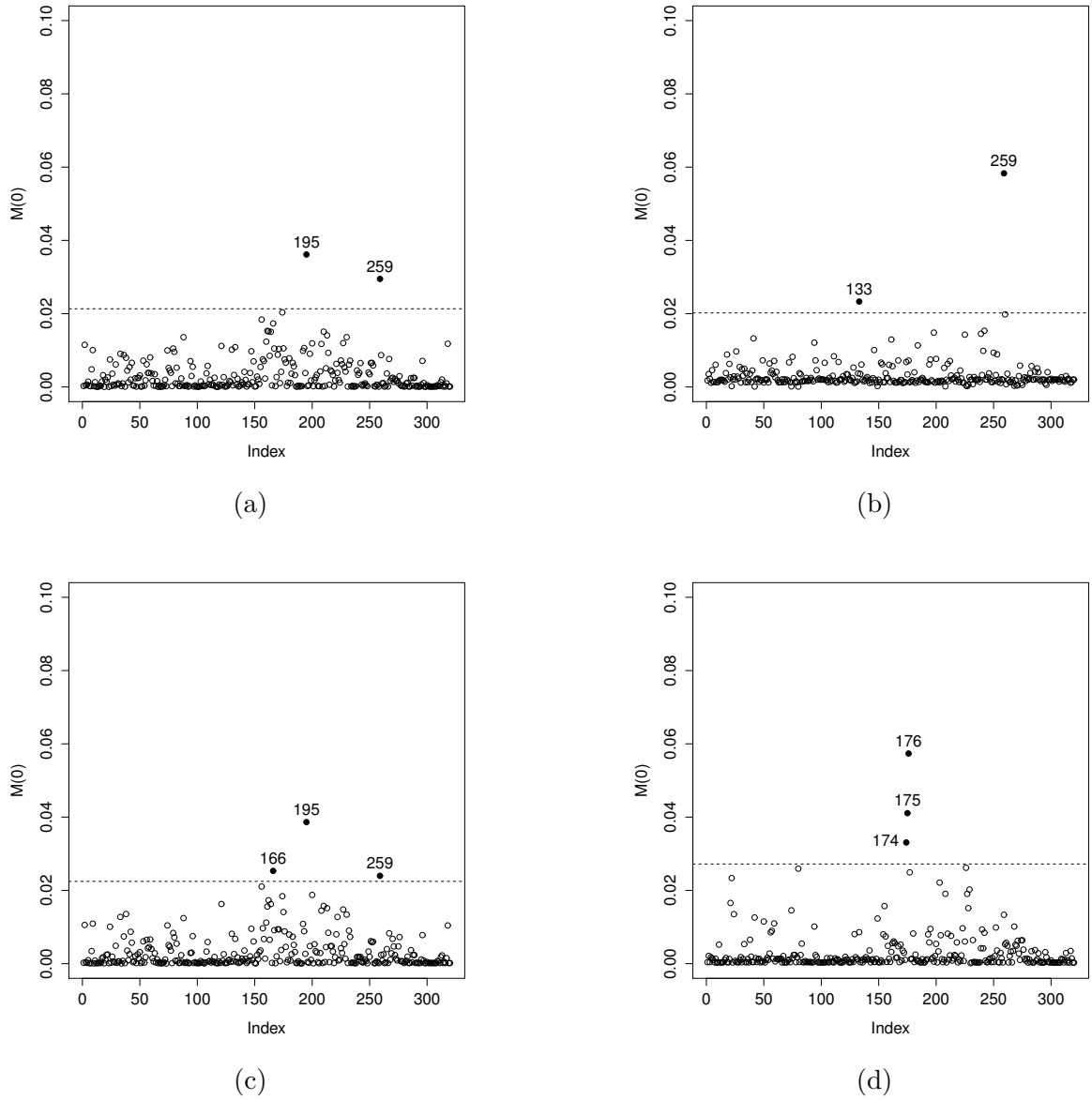

Figure 14 – **AIEDRP data.** Index plot of $M(0)$ for assessing local influence on $\boldsymbol{\theta}$ under case weight perturbation (Panel a), perturbation on $\mathbf{D}$ (Panel b), perturbation on σ^2 (Panel c), and perturbation on the response variable (Panel d). The influential observations are numbered.

are in agreement with the results showed in Figure 13.

3.5.4.2 Observation-level diagnostics

Using the perturbation schemes described in Section 3.3.2.2, Figure 15 presents the observation-level diagnostics for the dataset. Note that, in the case weight and σ^2 perturbation schemes, the observations #402, #403, #404, #410 (subject #174), #412 (subject #175), #422 (subject #176) and #512, #513, #514 and #515 (subject #203) can be considered influential. For perturbation on $\mathbf{D}$, we find that all observations between

Table 10 – **AIEDRP data.** RC (in %).

Dropped	$RC_{\hat{\beta}_1}$	$RC_{\hat{\beta}_2}$	$RC_{\hat{\beta}_3}$	$RC_{\hat{\beta}_4}$	$RC_{\hat{\beta}_5}$	$RC_{\widehat{\sigma^2}}$	$RC_{\hat{\alpha}_{11}}$	$RC_{\hat{\alpha}_{12}}$	$RC_{\hat{\alpha}_{22}}$
9	0.0124	0.9169	0.0170	0.9082	0.0000	2.8738	0.8690	0.0000	20.0000
133	0.1862	8.8020	0.0426	2.6438	3.2258	0.5845	3.5382	0.0000	20.0000
166	0.0062	6.6626	0.1334	3.1887	0.0000	2.4355	1.1173	0.0000	20.0000
174	0.0931	6.9071	0.1845	2.8355	0.0000	1.5100	1.1794	21.4286	0.0000
175	0.0621	1.7726	0.0993	3.0071	3.2258	0.6332	0.8070	7.1429	20.0000
176	0.1800	7.7017	0.0511	2.9162	0.0000	1.0229	1.1794	0.0000	20.0000
195	0.2421	9.4743	0.1760	2.7447	0.0000	0.2923	0.3724	0.0000	0.0000
230	0.0621	4.7066	0.0284	0.7164	0.0000	2.8251	1.0552	0.0000	0.0000
259	0.2111	7.7017	0.1306	4.0464	3.2258	1.1203	6.2073	35.7143	0.0000

#680 and #693 can be considered influential. Note, these observations correspond to subject #259, which was considered as possibly influential using the diagnostic tools proposed previously (see Sections 3.5.3 and 3.5.4.1). Finally, in the case of the perturbation on the response variable, we find that observations #44 (subject #22), #182 and #186 (subject #80), #420 (subject #175), #529 (subject #208), #596 (subject #226), #604 (subject #227) and #616 (subject #229) appear as influential. All these observations with the exception of observation #181 corresponds to the last time observed for the subjects.

3.6 Simulation studies

In order to assess the finite sample performance of the proposed diagnostic measures for identifying outliers, we conduct a simulation study focussing on subject-level diagnostics. We consider the nonlinear mixed-effects model given by

$$y_{ij} = \frac{\beta_1 + b_{i1}}{1 + \exp(-[t_{ij} - (\beta_2 + b_{i2})]/\beta_3)} + \epsilon_{ij}, \quad i = 1, \ldots, 50, \ j = 1, \ldots, 10, \tag{3.18}$$

where $t_{ij} = 100, 267, 433, 600, 767, 933, 1100, 1267, 1433, 1600$ for all i. The random effects $\mathbf{b}_i = (b_{i1}, b_{i2})^\top$, and the error term $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{i10})^\top$ are non-correlated with

$$\begin{pmatrix} \mathbf{b}_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \overset{ind.}{\sim} t_{12}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \sigma^2\mathbf{I}_{10} \end{pmatrix}, 8\right), \ i = 1, \ldots, 15.$$

We set the fixed-effects $\boldsymbol{\beta}^\top = (\beta_1, \beta_2, \beta_3) = (200, 700, 350)$, the between-subject covariance matrix $\mathbf{D} = \begin{pmatrix} 4 & -2 \\ -2 & 25 \end{pmatrix}$, and the within-subject variance $\sigma^2 = 25$. Under this model we consider the following perturbation schemes:

(a) Replace the fixed effects $\boldsymbol{\beta}$ by $2\boldsymbol{\beta}$ to generate the responses of the 1st subject $\mathbf{y}_1$,

(b) Replace $\boldsymbol{\beta}$ by $3\boldsymbol{\beta}$ and,

(c) Replace $\boldsymbol{\beta}$ by $4\boldsymbol{\beta}$.

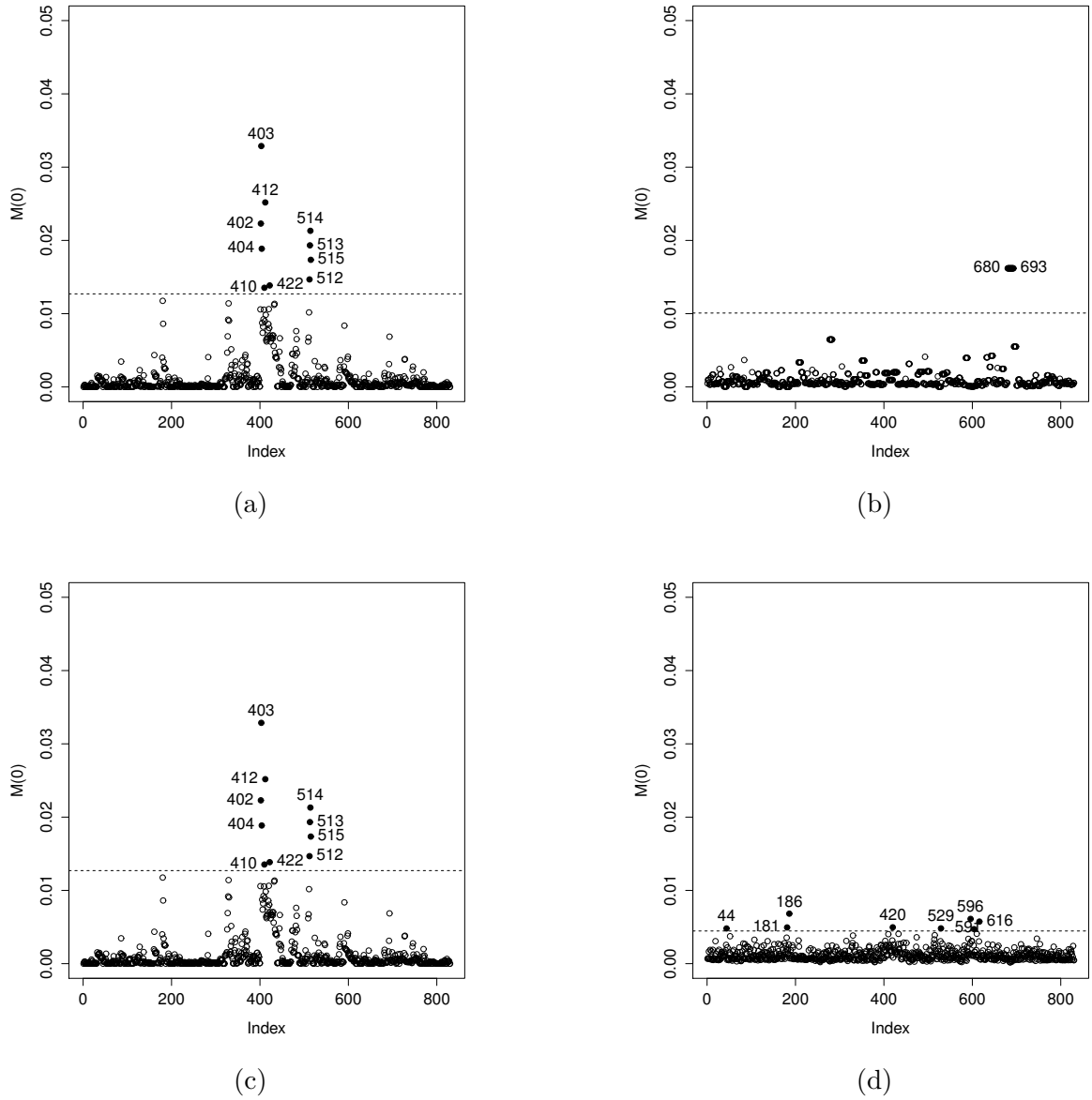

Figure 15 – **AIEDRP data.** Index plot of $M(0)$ for assessing local influence on $\boldsymbol{\theta}$ under case weight perturbation (Panel a), perturbation on $\mathbf{D}$ (Panel b), perturbation on σ^2 (Panel c), and perturbation on the response variable (Panel d). The influential observations are numbered.

The diagnostic measures were computed for 500 simulated data sets under various censoring proportions, say 0%, 5%, 10%, 20% and 30%. Table 11 reports (in percentage) the number of times the measures correctly identifies $\mathbf{y}_1$ as the most influential.

As expected, the percentage of correctly detecting atypical observations increases for increasing perturbation rates (*i.e.*, for $3\boldsymbol{\beta}$ or $4\boldsymbol{\beta}$) as compared to $2\boldsymbol{\beta}$, and with increased rate of censoring. Interestingly, the % of correct detection when the influence analysis is focused on σ^2 is not appealing, with a lower percentage of correct detection when the perturbation rate increases. However, higher % of correct detection when the influence

Table 11 – **Simulation study.** The values in the table denotes the % of correctly identifying the influential observations using case-deletion, case weight, σ^2 perturbation and matrix $\mathbf{D}$ perturbation from 500 simulated datasets under the t-NLMEC model specified in (3.18).

	% of censoring				
Case-deletion measure (GD_i)	0%	5%	10%	20%	30%
Pert. $2\boldsymbol{\beta}$	66.8	66.8	74.8	75.8	81.8
Pert. $3\boldsymbol{\beta}$	83.0	83.4	85.8	91.6	94.8
Pert. $4\boldsymbol{\beta}$	93.0	93.2	94.2	97.4	98.4
Case-weight perturbation	0%	5%	10%	20%	30%
Pert. $2\boldsymbol{\beta}$	66.8	66.8	74.8	75.8	81.8
Pert. $3\boldsymbol{\beta}$	83.0	83.4	85.8	91.6	94.8
Pert. $4\boldsymbol{\beta}$	93.0	93.2	94.2	97.4	98.4
Perturbation on σ^2	0%	5%	10%	20%	30%
Pert. $2\boldsymbol{\beta}$	13.0	14.4	18.8	19.2	15.2
Pert. $3\boldsymbol{\beta}$	3.60	3.60	4.60	6.00	6.00
Pert. $4\boldsymbol{\beta}$	0.40	0.60	0.80	1.00	0.60
Perturbation on $\mathbf{D}$	0%	5%	10%	20%	30%
Pert. $2\boldsymbol{\beta}$	83.8	83.6	83.2	83.0	84.8
Pert. $3\boldsymbol{\beta}$	95.0	94.6	94.0	94.8	97.4
Pert. $4\boldsymbol{\beta}$	97.2	97.8	97.6	98.8	99.0

analysis is focused on $\mathbf{D}$ is detected. A possible explanation for this fact is that a perturbation on the fixed-effects of one subject contributes to increasing the between-subject variance, but the within-subject variance remains the same.

3.7 Conclusions

This chapter proposes diagnostic tools for detecting outliers and/or influential observations in the context of linear and nonlinear mixed-effects censored model where the joint distribution of the random effects and random errors follow the Student's-t distribution. The results presented here supplement the robust likelihood-based inference developed by Matos et al. (2013b) for LMEC/NLMEC models, appropriate for longitudinal HIV data. Our proposed estimation method relies on the Q-function and the corresponding ECM algorithm. The NLME formulation is mathematically (and computationally) feasible through a linearisation. The methodology is implemented using the `R` software (codes are available upon request), providing practitioners with a convenient tool for further applications in their domain.

For ease of implementation, our current proposal considers an independent

within-subject covariance structure, viz. $\sigma^2 I_{n_i}$. Nevertheless, it can be extended to different unstructured covariance matrices (such as AR(1), or ante-dependence) following the work of Pan et al. (2014). This issue is currently under investigation, and we plan to tackle it in a future work.

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

4.1 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 viewpoint, such an assumption may not be realistic. In this context, some recent works in censored models (Matos et al., 2013b; 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. 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 multivariate Student's-t linear mixed model (t-LME). Lin and Lee (2006, 2007) developed some additional tools for the t-LME from likelihood-based and Bayesian perspectives, respectively. It is important to stress that, from a 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

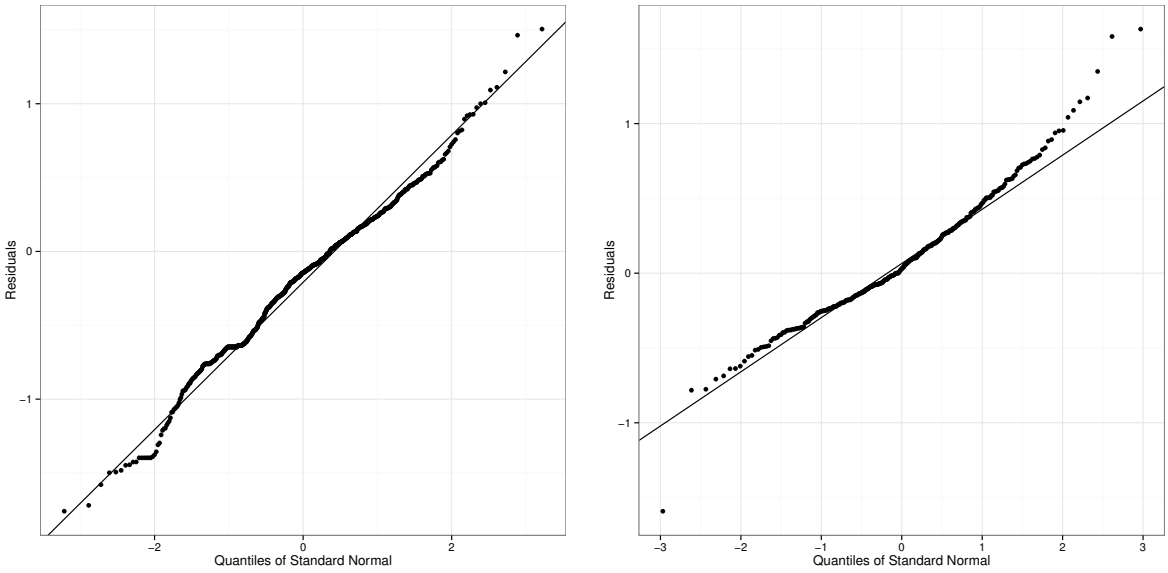

Figure 16 – **AIDS studies data.** Normal Q–Q plot for model residuals obtained by fitting a censored (Gaussian) mixed-effect model (AIEDRP data left panel/UTI data right panel).

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. For instance, Garay et al. (2014) and Matos et al. (2013b) utilized the Student's-t distribution in the context of censored regression (t-CR) and linear and nonlinear mixed-effects (t-LMEC) models for censored responses respectively. They considered exact EM algorithms for maximum likelihood (ML) estimation, relying on the mean and variance of a truncated multivariate Student's-t distribution. However, the use of others heavy-tailed distributions has not been explored in the context of censored longitudinal models. In this regard, the aim of this work is to consider the multivariate scale mixtures of normal (SMN) distributions as the distribution of the random error in the framework of the nonlinear censored regression (SMN-NCR) model for longitudinal data. Our contribution extends the recent works of Garay et al. (2014) and Garay et al. (2015) since they used only the Student's-t distribution which is a member of the SMN/NI class.

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) is generally 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.

In order to evaluate the performance of our proposal, we consider the analysis of two AIDS case studies. The first study evaluated the immune responses to HIV during acute infection, presenting about 22% of measurements lying above the limits of assay quantification (right-censored). The viral loads were irregularly measured over time. The individual profiles (in $\log_{10}$ scale) of HIV viral load at different follow-up times are displayed in Figure 3. The corresponding normal quantile-quantile (QQ) plot (Figure 16 left panel) for the HIV viral load after fitting the Gaussian nonlinear censored mixed-effect model suggests that the normality assumption for the within-subject errors might be inappropriate.

The second study 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 dataset presents about 7% of observations below the detection limits of assay quantifications (left censored). Figure 1 presents the individual profiles of viral load at different follow-up times after UTI. In addition, a normal QQ plot for the residuals (Figure 16 right panel) obtained by fitting a normal censored mixed-effect model is presented.

Since the outcome variables were recorded at irregular occasions in both studies, 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.

The chapter is organized as follows. Section 4.2 proposes the SMN-NCR model and shows how the ML estimates through the SAEM algorithm are computed. In Section 4.3, we formulate analytically the empirical information matrix of model parameters. The issue concerning the prediction of future observations is also discussed. In Section 4.4, our proposed techniques are compared with the normality-based approach using simulated data and illustrated with the analysis of the AIDS case studies. Section 4.5 concludes with a short discussion of issues raised by our methods and some possible directions for a future

research.

4.2 Regression models for irregularly observed longitudinal data

4.2.1 The statistical model

Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$ 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 $\mathbf{t}_i = (t_{i1}, \ldots, t_{in_i})$. Formally, the nonlinear regression model is given by

$$\begin{aligned} \mathbf{y}_i &= \mathbf{g}(\boldsymbol{\varphi}_i, \mathbf{t}_i) + \boldsymbol{\epsilon}_i, \\ \boldsymbol{\varphi}_i &= \mathbf{A}_i\boldsymbol{\beta} \end{aligned} \tag{4.1}$$

where $\mathbf{g}(\boldsymbol{\varphi}_i, \mathbf{t}_i)$ is a nonlinear vector-valued differentiable function of the parameter $\boldsymbol{\varphi}_i$; $\mathbf{A}_i$ is a known design matrix of dimension $r \times p$, possibly depending on some covariate vector $\mathbf{X}_i$; $\boldsymbol{\beta}$ is the $p \times 1$ vector of fixed effects; and $\boldsymbol{\epsilon}_i$ is the vector of random errors of dimension $(n_i \times 1)$ with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Omega}_i$. Instead of the usual assumption of normality, we replace the multivariate normal distribution by the scale mixtures of multivariate normal distributions. Therefore, it follows that

$$\boldsymbol{\epsilon}_i \overset{\text{ind.}}{\sim} \text{SMN}_{n_i}(\mathbf{0}, \boldsymbol{\Omega}_i, \boldsymbol{\nu}), \quad i = 1, \ldots, n. \tag{4.2}$$

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 structure for $\mathbf{E}_i$, described in Section 1.1.2.

Using the stochastic representation (1.1), the hierarchical representation (two-stages) of the nonlinear regression model defined in (4.1) - (4.2) is given by

$$\begin{aligned} \mathbf{y}_i \mid U_i = u_i &\overset{\text{ind.}}{\sim} \text{N}_{n_i}(\mathbf{g}(\boldsymbol{\varphi}_i, \mathbf{t}_i), \kappa(u_i)\boldsymbol{\Omega}_i), \\ U_i &\overset{\text{iid.}}{\sim} h(u_i \mid \boldsymbol{\nu}). \end{aligned} \tag{4.3}$$

For simplicity we will denote $\boldsymbol{\mu}_i(\boldsymbol{\beta}) = \mathbf{g}(\boldsymbol{\varphi}_i, \mathbf{t}_i)$.

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

$$\begin{aligned} y_{ij} &\leqslant V_{ij} \quad if \quad C_{ij} = 1, \\ y_{ij} &= V_{ij} \quad if \quad C_{ij} = 0, \end{aligned} \tag{4.4}$$

where V_{ij} represents the uncensored observation or limit of quantification and C_{ij} is the censoring indicator whose value is equal to one if y_{ij} is a censored observation and zero if y_{ij} is an uncensored observation. Consequently, 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 extended to other censoring types. The formulations defined in (4.1) – (4.4) will be called the SMN-NCR model.

4.2.2 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 $\mathbf{y}_i$, $i = 1, \ldots, n$. For the SMN-NCR model with complete data, we have that, marginally,

$$\mathbf{y}_i \overset{\text{ind.}}{\sim} \text{SMN}_{n_i}(\boldsymbol{\mu}_i(\boldsymbol{\beta}), \boldsymbol{\Omega}_i, \boldsymbol{\nu}), \quad i = 1, \ldots, n. \tag{4.5}$$

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 Definition 4.1 below.

Definição 4.1. *Let* $\mathbf{y}$ *be partitioned as* $\mathbf{y}_i = vec(\mathbf{y}_i^o, \mathbf{y}_i^c)$ *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* $vec(\cdot)$ *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* $\mathbf{V}_i$, $\boldsymbol{\mu}_i(\boldsymbol{\beta})$, *and* $\boldsymbol{\Omega}_i$ *also be partitioned as follows:* $\mathbf{V}_i = vec(\mathbf{V}_i^o, \mathbf{V}_i^c)$, $\boldsymbol{\mu}_i(\boldsymbol{\beta})^\top = (\boldsymbol{\mu}_i^o(\boldsymbol{\beta}), \boldsymbol{\mu}_i^c(\boldsymbol{\beta}))^\top$, *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* $\mathbf{y}_i \mid u_i \sim N_{n_i}(\boldsymbol{\mu}_i(\boldsymbol{\beta}), \kappa(u_i)\boldsymbol{\Omega}_i)$, *where*

$$\mathbf{y}_i^o \mid u_i \sim N_{n_i^o}(\boldsymbol{\mu}_i^o(\boldsymbol{\beta}), \kappa(u_i)\boldsymbol{\Omega}_i^{oo}) \quad \textit{and} \quad \mathbf{y}_i^c \mid \mathbf{y}_i^o, u_i \sim N_{n_i^c}(\boldsymbol{\mu}_i^{c\cdot o}, \kappa(u_i)\mathbf{S}_i), \tag{4.6}$$

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

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

Definição 4.2. *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*

$$\begin{aligned} L_i(\boldsymbol{\theta}) &= f(\mathbf{y}_i^o \mid \boldsymbol{\theta}) f(\mathbf{y}_i^c \leqslant \mathbf{V}_i^c \mid \mathbf{y}_i^o, \boldsymbol{\theta}) \\ &= \int_0^\infty f(\mathbf{y}_i^o \mid u_i, \boldsymbol{\theta}) f(\mathbf{y}_i^c \leqslant \mathbf{V}_i^c \mid u_i, \mathbf{y}_i^o, \boldsymbol{\theta}) dH(u_i) \\ &= \int_0^\infty \phi_{n_i^o}(\mathbf{y}_i^o; \boldsymbol{\mu}_i^o(\boldsymbol{\beta}), \kappa(u_i)\boldsymbol{\Omega}_i^{oo}) \Phi_{n_i^c}(\mathbf{V}_i^c; \boldsymbol{\mu}_i^{c\cdot o}, \kappa(u_i)\mathbf{S}_i) dH(u_i). \end{aligned} \tag{4.7}$$

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

Proposição 4.1. *The likelihood function for special elements of the SMN class are given by.*

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

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

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

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

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

3. *(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 $h(u|\boldsymbol{\nu}) = \nu_1\mathbb{I}_{\{\nu_2\}}(u) + (1-\nu_1)\mathbb{I}_{\{1\}}(u)$, then*

$$\begin{aligned} L_i(\boldsymbol{\theta}) &= \nu_1\left[\phi_{n_i^o}(\mathbf{y}_i^o; \boldsymbol{\mu}_i^o(\boldsymbol{\beta}), \nu_2^{-1}\boldsymbol{\Omega}_i^{oo})\Phi_{n_i^c}(\mathbf{V}_i^c; \boldsymbol{\mu}_i^{c\cdot o}, \nu_2^{-1}\mathbf{S}_i)\right] \\ &+ (1-\nu_1)\left[\phi_{n_i^o}(\mathbf{y}_i^o; \boldsymbol{\mu}_i^o(\boldsymbol{\beta}), \boldsymbol{\Omega}_i^{oo})\Phi_{n_i^c}(\mathbf{V}_i^c; \boldsymbol{\mu}_i^{c\cdot o}, \mathbf{S}_i)\right]. \end{aligned}$$

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 work, we assume that the parameter $\boldsymbol{\nu}$ is fixed. The most appropriate value of $\boldsymbol{\nu}$ (see Lange et al., 1989; Meza et al., 2012) are chosen on AIC or BIC. The entire parameter vector is $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\phi}^\top)^\top$ hereafter.

4.2.3 Maximum likelihood estimation

In this subsection, we develop the MCMC-SAEM (hereafter SAEM) algorithm for ML estimation of the parameters in the SMN-NCR model defined previously. Consider the model defined in (4.1) – (4.4), $\mathbf{u} = (u_1, \ldots, u_n)^\top$, $\mathbf{V} = vec(\mathbf{V}_1, \ldots, \mathbf{V}_n)$, and $\mathbf{C} = vec(\mathbf{C}_1, \ldots, \mathbf{C}_n)$ such that we observe $(\mathbf{V}_i, \mathbf{C}_i)$ for the i-th subject. 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 (4.3), as $\ell_c(\boldsymbol{\theta} \mid \mathbf{y}_c) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta} \mid \mathbf{y}_c)$,

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

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) $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, the conditional expectation of the complete data log-likelihood function is given by:

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

where

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

with

$$\widehat{\kappa\mathbf{y}_i^2}^{(k)} = E\left[\kappa^{-1}(u_i)\mathbf{y}_i\mathbf{y}_i^\top \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right], \tag{4.8}$$

$$\widehat{\kappa\mathbf{y}_i}^{(k)} = E\left[\kappa^{-1}(u_i)\mathbf{y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right], \tag{4.9}$$

$$\widehat{\kappa}_i^{(k)} = E\left[\kappa^{-1}(u_i) \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right]. \tag{4.10}$$

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

In the traditional EM algorithm, we evaluate the conditional expectations given in Equations (4.8) – (4.10). As there are no closed-form expressions for them, two intermediate steps are introduced, including the simulation and approximation steps. In the simulation step, for the i-th subject, we generate samples from the full conditional distributions of the latent variables $(u_i, \mathbf{y}_i)$ via the Gibbs sampler algorithm according to the following scheme (at the k-th iteration):

Step 1: Sample $\mathbf{y}_i^{c(k,l)}$ from $f(\mathbf{y}_i^c \mid \mathbf{V}_i^c, \mathbf{y}_i^o, u_i, \widehat{\boldsymbol{\theta}}^{(k-1)})$, which is a truncated normal distribution. Using Definition 4.1 and conditioning on the censored components, we obtain

$$\mathbf{y}_i^c \mid \mathbf{V}_i^c, \mathbf{y}_i^o, u_i, \boldsymbol{\theta} \sim \mathrm{TN}_{n_i^c}(\boldsymbol{\mu}_i^{c\cdot o}, \kappa(u_i)\mathbf{S}_i; \mathbb{A}_i),$$

with $\mathbb{A}_i = \{\mathbf{y}_i^c = (y_{i1}^c, \ldots, y_{in_i^c}^c)^\top \mid y_{i1}^c \leqslant V_{i1}^c, \ldots, y_{in_i^c}^c \leqslant V_{in_i^c}^c\}$, $\boldsymbol{\mu}_i^{c\cdot o} = \boldsymbol{\mu}_i^c(\boldsymbol{\beta}) + \boldsymbol{\Omega}_i^{co}(\boldsymbol{\Omega}_i^{oo})^{-1}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o(\boldsymbol{\beta}))$ 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 $\mathbf{y}_i^{(k,l)} = (y_{i1}^{c(k,l)}, \ldots, y_{in_i^c}^{c(k,l)}, y_{n_i^c+1}, \ldots, y_{n_i})$ 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 $f(u_i \mid \mathbf{y}_i^{(k,l)}, \widehat{\boldsymbol{\theta}}^{(k-1)})$. Let $D^2_{\boldsymbol{\epsilon}_i} = (\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}))^\top \boldsymbol{\Omega}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}))$, this gives rise to

(a) Student's-t: $u_i \mid \mathbf{y}_i, \boldsymbol{\theta} \sim Gamma\left(\dfrac{\nu + n_i}{2}, \dfrac{\nu + D^2_{\boldsymbol{\epsilon}_i}}{2}\right)$;

(b) Slash: $u_i \mid \mathbf{y}_i, \boldsymbol{\theta} \sim TGamma\left(\nu + \dfrac{n_i}{2}, \dfrac{D^2_{\boldsymbol{\epsilon}_i}}{2}; (0,1)\right)$, which follows a truncated gamma distribution lying on the interval (0,1);

(c) Contaminated normal: $f(u_i \mid \mathbf{y}_i, \boldsymbol{\theta})$ is a discrete distribution taking values ν_2 with probability $\dfrac{p_1}{p_1 + p_2}$ and 1 with probability $\dfrac{p_2}{p_1 + p_2}$, where

$$p_1 = \nu_1 \nu_2^{\frac{n_i}{2}} \exp\left(-\frac{\nu_2}{2} D^2_{\boldsymbol{\epsilon}_i}\right), \quad \text{and} \quad p_2 = (1 - \nu_1) \exp\left(-\frac{1}{2} D^2_{\boldsymbol{\epsilon}_i}\right).$$

The next step is the **Stochastic Approximation**. Since the sequence $(\mathbf{y}_i^{(k,l)}, u_i^{(k,l)})$ for $l = 1, \ldots, m$ is collected at the k-th iteration, we replace the conditional expectations given in (4.8) –(4.10) with the following stochastic approximations:

$$\widehat{\kappa \mathbf{y}_i^2}^{(k)} = \widehat{\kappa \mathbf{y}_i^2}^{(k-1)} + \delta_k \left[\frac{1}{m}\sum_{l=1}^{m} \kappa^{-1}(u_i^{(k,l)}) \mathbf{y}_i^{(k,l)} \mathbf{y}_i^{(k,l)\top} - \widehat{\kappa \mathbf{y}_i^2}^{(k-1)}\right], \tag{4.11}$$

$$\widehat{\kappa \mathbf{y}_i}^{(k)} = \widehat{\kappa \mathbf{y}_i}^{(k-1)} + \delta_k \left[\frac{1}{m}\sum_{l=1}^{m} \kappa^{-1}(u_i^{(k,l)}) \mathbf{y}_i^{(k,l)} - \widehat{\kappa \mathbf{y}_i}^{(k-1)}\right], \tag{4.12}$$

$$\widehat{\kappa}_i^{(k)} = \widehat{\kappa}_i^{(k-1)} + \delta_k \left[\frac{1}{m}\sum_{l=1}^{m} \kappa^{-1}(u_i^{(k,l)}) - \widehat{\kappa}_i^{(k-1)}\right]. \tag{4.13}$$

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 $m \leqslant 10$ 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:

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \widehat{\boldsymbol{\beta}}^{(k)} + \left(\sum_{i=1}^{n} \widehat{\kappa}_i{}^{(k)} \widehat{\mathbf{J}}_i^{(k)\top} \widehat{\mathbf{E}}_i^{-1(k)} \widehat{\mathbf{J}}_i^{(k)}\right)^{-1} \sum_{i=1}^{n} \widehat{\mathbf{J}}_i^{(k)\top} \widehat{\mathbf{E}}_i^{-1(k)} \left(\widehat{\kappa \mathbf{y}_i}^{(k)} - \boldsymbol{\mu}_i(\widehat{\boldsymbol{\beta}}^{(k)})\right) \quad (4.14)$$

$$\begin{aligned} \widehat{\sigma^2}^{(k+1)} &= \frac{1}{N} \sum_{i=1}^{n} \left[tr\left(\widehat{\kappa \mathbf{y}_i^2}^{(k)} \widehat{\mathbf{E}}_i^{-1(k)}\right) - 2\boldsymbol{\mu}_i^{\top}(\widehat{\boldsymbol{\beta}}^{(k)}) \widehat{\mathbf{E}}_i^{-1(k)} \widehat{\kappa \mathbf{y}_i}^{(k)} \right. \\ &+ \left. \widehat{\kappa}_i{}^{(k)} \boldsymbol{\mu}_i^{\top}(\widehat{\boldsymbol{\beta}}^{(k)}) \widehat{\mathbf{E}}_i^{-1(k)} \boldsymbol{\mu}_i(\widehat{\boldsymbol{\beta}}^{(k)}) \right], \end{aligned} \quad (4.15)$$

$$\begin{aligned} \widehat{\boldsymbol{\phi}}^{(k+1)} &= \underset{\boldsymbol{\phi} \in (0,1) \times \mathbb{R}^+}{\operatorname{argmax}} \left(-\frac{1}{2\widehat{\sigma^2}^{(k)}} \sum_{i=1}^{n} \left[tr\left(\widehat{\kappa \mathbf{y}_i^2}^{(k)} \mathbf{E}_i^{-1}\right) - 2\boldsymbol{\mu}_i^{\top}(\widehat{\boldsymbol{\beta}}^{(k)}) \mathbf{E}_i^{-1} \widehat{\kappa \mathbf{y}_i}^{(k)} \right. \right. \\ &+ \left. \left. \widehat{\kappa}_i{}^{(k)} \boldsymbol{\mu}_i^{\top}(\widehat{\boldsymbol{\beta}}^{(k)}) \mathbf{E}_i^{-1} \boldsymbol{\mu}_i^{\top}(\widehat{\boldsymbol{\beta}}^{(k)}) \right] - \frac{1}{2} \sum_{i=1}^{n} \log(|\mathbf{E}_i^{-1}|) \right), \end{aligned} \quad (4.16)$$

where $\mathbf{J}_i = \dfrac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^{\top}}$ and $\widehat{\kappa \mathbf{y}_i^2}^{(k)}$, $\widehat{\kappa \mathbf{y}_i}^{(k)}$ and $\widehat{\kappa}_i{}^{(k)}$ rely on minimal sufficient statistics.

It is important to stress that, 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.

4.2.4 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 $\tilde{\mathbf{y}}_i^{c(k)}$, are calculated as

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

where

$$\tilde{\mathbf{y}}_i^{c(k)} = \tilde{\mathbf{y}}_i^{c(k-1)} + \delta_k \left[\frac{1}{m} \sum_{l=1}^{m} \mathbf{y}_i^{c(k,l)} - \tilde{\mathbf{y}}_i^{c(k)} \right] \quad (4.17)$$

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

4.3 Standard errors and prediction of future observations

4.3.1 Empirical information matrix

The asymptotic covariance matrix of the ML estimates can be approximated by 1.6. As a result, the empirical information matrix $\mathbf{I}_e(\boldsymbol{\theta} \mid \mathbf{y})$ is reduced to

$$\mathbf{I}_e(\widehat{\boldsymbol{\theta}} \mid \mathbf{y}) = \sum_{i=1}^{n} \widehat{\mathbf{s}}_i \widehat{\mathbf{s}}_i^\top = \begin{pmatrix} \widehat{\mathbf{s}}_{i,\boldsymbol{\beta}} \\ \widehat{\mathbf{s}}_{i,\sigma 2} \\ \widehat{\mathbf{s}}_{i,\boldsymbol{\phi}} \end{pmatrix} \begin{pmatrix} \widehat{\mathbf{s}}_{i,\boldsymbol{\beta}} & \widehat{\mathbf{s}}_{i,\sigma 2} & \widehat{\mathbf{s}}_{i,\boldsymbol{\phi}} \end{pmatrix}, \tag{4.18}$$

where $\widehat{\mathbf{s}}_i = \mathbf{s}(\mathbf{y}_i \mid \widehat{\boldsymbol{\theta}}) = E\left(\dfrac{\partial \ell_i(\boldsymbol{\theta} \mid \mathbf{y}_c)}{\partial \boldsymbol{\theta}} \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\right)$, then

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

with

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

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

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

4.3.2 Prediction

For generating predicted values from the SMN-NCR 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 $\boldsymbol{\mu}_i(\boldsymbol{\beta}) = (\boldsymbol{\mu}_{i,obs}(\boldsymbol{\beta}), \boldsymbol{\mu}_{i,pred}(\boldsymbol{\beta}))^\top$ be the $(n_{i,obs} + n_{i,pred}) \times 1$ nonlinear vector 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 (4.17). Therefore, after this imputation step, a complete data set, $\mathbf{y}_{i,obs*}$, is obtained. We obtain

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

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

$$\widehat{\mathbf{y}}_{i,pred}(\boldsymbol{\theta}) = \boldsymbol{\mu}_{i,pred}(\boldsymbol{\beta}) + \boldsymbol{\Omega}_i^{pred,obs*}\boldsymbol{\Omega}_i^{obs*,obs*-1}\left(\mathbf{y}_{i,obs*} - \boldsymbol{\mu}_{i,obs*}(\boldsymbol{\beta})\right). \tag{4.19}$$

Consequently, $\mathbf{y}_{i,pred}$ can be estimated directly by substituting $\widehat{\boldsymbol{\theta}}$ into (4.19).

4.4 Application

In this section, we illustrate the performance of the proposed techniques through simulated datasets. Afterward, we apply the methods to the analysis of two HIV datasets previously analyzed by Vaida and Liu (2009) and Matos et al. (2013b).

4.4.1 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.

We present three scenarios considering the same probability distribution and correlation structure for the datasets. The responses follow a contaminated normal distribution with parameter $\boldsymbol{\nu} = (\nu_1, \nu_2)^\top = (0.1, 0.1)^\top$ and DEC structure with $\phi_1 = 0.8$ and $\phi_2 = 1$. The simulated data are generated following the model defined in Subsection 4.2.1, where $\mathbf{A}_i = \begin{bmatrix} \mathbf{1}_{n_i} & \mathbf{t}_i^\top \end{bmatrix}$ and $\mathbf{g}_i$ is the identity function, with parameters setting at $\beta_1 = 2$ $\beta_2 = 1$, $\sigma^2 = 2$ and time points set as $\mathbf{t}_i = (1, 3, 5, 7, 10, 14)$, for $i = 1, \ldots, n$.

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-NCR model, the T-NCR model with 4 degrees of freedom and the SL-NCR 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-NCR models are fitted. We aim to study the behavior of the SMN-NCR 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 $c = 0.25$.

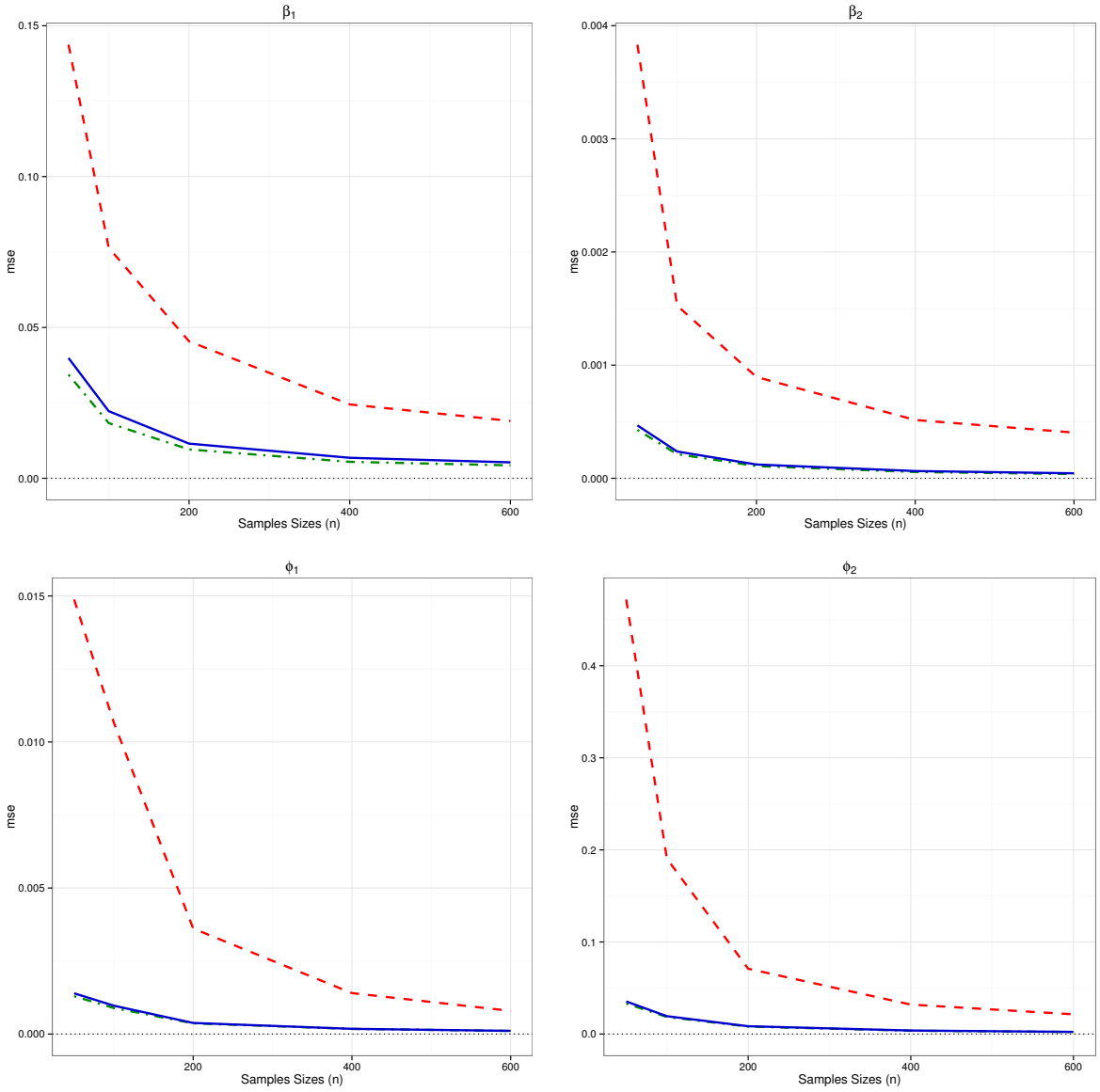

Figure 17 – **Simulation study - Scenario 1.** Mean square error of the parameter estimates in the SMN-NCR model under 10% of censoring level and different samples sizes. The solid line (blue) represents the T-NCR model, the dotted line (red) represents the N-NCR model and the dotdash line (green) represents the SL-NCR model.

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-NCR models under different sample sizes. These measures are defined as:

$$\text{Bias} = \frac{1}{100}\sum_{j=1}^{100}|\widehat{\boldsymbol{\theta}}_i^{(j)} - \boldsymbol{\theta}_i| \quad \text{and} \quad \text{MSE} = \frac{1}{100}\sum_{j=1}^{100}\left(\widehat{\boldsymbol{\theta}}_i^{(j)} - \boldsymbol{\theta}_i\right)^2, \tag{4.20}$$

where $\widehat{\boldsymbol{\theta}}_i^{(j)}$ is the ML estimate of the parameter θ_i for the j-th sample, $j = 1, \ldots, 100$. The main objective of this simulation is to provide empirical evidence about consistence of the ML estimates. It is apparently seen in Figure 17 that the MSE tends to zero as the sample size increases. Similar results are obtained after the analysis of the absolute bias (see Figure 25 in Appendix B.2). In general, for all models, the SAEM algorithm provides estimates with good asymptotic properties. In addition, Table 12 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-NCR models under different proportions of censoring. It can be found from Table 13 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 13 provides the Monte Carlo standard errors of the SAEM estimates obtained through the empirical information matrix described in Section 4.3 (IM SE). 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 (4.18) 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 conduct the experimental study ,an arbitrary simulated dataset is considered, where the conditional expectation $E\left[\mathbf{y}_{cens} \mid \mathbf{y}_{obs}\right]$ of the censored values is computed using Equation (4.17). Figure 18 shows the plot of the imputed values $E\left[\mathbf{y}_{cens} \mid \mathbf{y}_{obs}\right]$ as a function of the true censored (simulated)

Table 12 – **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-NCR 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%										
			Parameters					Criteria		
Distribution			β_1	β_2	σ^2	ϕ_1	ϕ_2	MC Loglik.	MC AIC	MC BIC
		MC Mean	2.021	0.997	2.019	0.799	1.020	-539.049	1088.098	1106.617
	T	IM SE	0.235	0.024	0.359	0.046	0.203			
		MC Sd	0.200	0.022	0.351	0.038	0.188			
		MC Mean	2.015	0.997	1.446	0.800	1.022	-540.094	1090.188	1108.707
$n = 50$	SL	IM SE	0.246	0.025	0.206	0.045	0.197			
		MC Sd	0.186	0.021	0.210	0.036	0.182			
		MC Mean	2.121	1.003	16.052	0.801	1.184	-714.865	1439.731	1458.250
	N	IM SE	1.012	0.089	1.554	0.050	0.278			
		MC Sd	0.361	0.062	8.093	0.123	0.665			
		MC Mean	2.023	0.997	1.973	0.799	1.015	-1075.346	2160.692	2182.676
	T	IM SE	0.163	0.017	0.242	0.031	0.139			
		MC Sd	0.148	0.015	0.253	0.031	0.139			
		MC Mean	2.016	0.998	1.420	0.800	1.018	-1077.510	2165.019	2187.004
$n = 100$	SL	IM SE	0.171	0.017	0.139	0.030	0.134			
		MC Sd	0.135	0.014	0.159	0.030	0.136			
		MC Mean	2.088	1.009	16.134	0.790	1.051	-1449.267	2908.533	2930.518
	N	IM SE	0.633	0.053	0.688	0.020	0.081			
		MC Sd	0.263	0.038	6.519	0.103	0.436			
		MC Mean	2.023	0.997	1.972	0.801	1.010	-2152.751	4315.502	4340.953
	T	IM SE	0.114	0.012	0.169	0.022	0.096			
		MC Sd	0.105	0.011	0.172	0.020	0.092			
		MC Mean	2.016	0.997	1.421	0.801	1.013	-2156.934	4323.867	4349.318
$n = 200$	SL	IM SE	0.120	0.012	0.097	0.021	0.093			
		MC Sd	0.097	0.010	0.106	0.019	0.089			
		MC Mean	2.068	1.011	16.115	0.798	1.016	-2923.747	5857.494	5882.944
	N	IM SE	0.419	0.034	0.375	0.009	0.040			
		MC Sd	0.203	0.028	4.570	0.060	0.267			
		MC Mean	2.023	0.997	1.972	0.802	1.010	-4309.241	8628.483	8657.399
	T	IM SE	0.081	0.008	0.119	0.015	0.068			
		MC Sd	0.080	0.008	0.109	0.013	0.061			
		MC Mean	2.017	0.998	1.423	0.802	1.013	-4317.693	8645.387	8674.303
$n = 400$	SL	IM SE	0.084	0.009	0.068	0.015	0.065			
		MC Sd	0.072	0.007	0.069	0.013	0.059			
		MC Mean	2.061	1.011	16.166	0.803	1.002	-5873.152	11756.305	11785.221
	N	IM SE	0.289	0.023	0.237	0.005	0.025			
		MC Sd	0.145	0.020	3.125	0.038	0.180			
		MC Mean	2.020	0.998	1.972	0.802	1.010	-6468.005	12946.01	12976.95
	T	IM SE	0.066	0.007	0.097	0.012	0.055			
		MC Sd	0.070	0.006	0.080	0.010	0.047			
		MC Mean	2.014	0.998	1.423	0.803	1.013	-6480.928	12971.86	13002.80
$n = 600$	SL	IM SE	0.069	0.007	0.056	0.012	0.053			
		MC Sd	0.064	0.006	0.051	0.010	0.045			
		MC Mean	2.057	1.012	16.228	0.803	0.988	-8826.036	17662.07	17693.02
	N	IM SE	0.235	0.019	0.187	0.004	0.019			
		MC Sd	0.126	0.016	2.459	0.028	0.147			

Table 13 – **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-NCRM 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			Parameters					Criteria		
	Fit		β_1	β_2	σ^2	ϕ_1	ϕ_2	MC Loglik.	MC AIC	MC BIC
		MC Mean	2.010	0.998	1.994	0.801	1.009	-2357.915	4725.830	4751.280
	T	IM SE	0.111	0.012	0.168	0.021	0.093			
		MC Sd	0.104	0.011	0.172	0.018	0.087			
		MC Mean	2.007	0.999	1.425	0.801	1.012	-2360.530	4731.060	4756.510
0%	SL	IM SE	0.116	0.012	0.095	0.020	0.090			
		MC Sd	0.096	0.010	0.106	0.018	0.086			
		MC Mean	2.031	0.999	22.820	0.800	1.033	-3342.554	6695.108	6720.559
	N	IM SE	0.392	0.040	0.397	0.008	0.035			
		MC Sd	0.313	0.032	6.086	0.050	0.259			
		MC Mean	2.011	0.998	1.988	0.802	1.013	-2221.685	4453.371	4478.821
	T	IM SE	0.112	0.012	0.170	0.021	0.094			
		MC Sd	0.106	0.011	0.171	0.018	0.086			
		MC Mean	2.008	0.998	1.425	0.803	1.016	-2224.916	4459.832	4485.282
5%	SL	IM SE	0.117	0.012	0.096	0.020	0.091			
		MC Sd	0.098	0.010	0.106	0.018	0.085			
		MC Mean	2.300	0.988	16.115	0.800	1.040	-3038.035	6086.069	6111.519
	N	IM SE	0.413	0.034	0.350	0.008	0.038			
		MC Sd	0.226	0.028	4.549	0.060	0.274			
		MC Mean	2.023	0.997	1.972	0.801	1.010	-2152.751	4315.502	4340.953
	T	IM SE	0.114	0.012	0.169	0.022	0.096			
		MC Sd	0.105	0.011	0.172	0.020	0.092			
		MC Mean	2.016	0.997	1.421	0.801	1.013	-2156.934	4323.867	4349.318
10%	SL	IM SE	0.120	0.012	0.097	0.021	0.093			
		MC Sd	0.097	0.010	0.106	0.019	0.089			
		MC Mean	2.068	1.011	16.115	0.798	1.016	-2923.747	5857.494	5882.944
	N	IM SE	0.419	0.034	0.375	0.009	0.040			
		MC Sd	0.203	0.028	4.570	0.060	0.267			
		MC Mean	2.091	0.991	1.968	0.797	1.012	-1987.323	3984.646	4010.097
	T	IM SE	0.129	0.013	0.171	0.024	0.106			
		MC Sd	0.105	0.010	0.169	0.021	0.101			
		MC Mean	2.071	0.993	1.420	0.798	1.017	-1992.863	3995.727	4021.177
20%	SL	IM SE	0.134	0.013	0.099	0.023	0.102			
		MC Sd	0.097	0.010	0.108	0.020	0.096			
		MC Mean	1.627	1.051	16.496	0.788	0.976	-2687.424	5384.848	5410.299
	N	IM SE	0.464	0.036	0.425	0.011	0.045			
		MC Sd	0.202	0.030	4.783	0.061	0.256			
		MC Mean	2.290	0.974	1.998	0.796	1.055	-1804.398	3618.796	3644.246
	T	IM SE	0.158	0.015	0.176	0.027	0.124			
		MC Sd	0.134	0.012	0.169	0.022	0.115			
		MC Mean	2.262	0.976	1.421	0.796	1.056	-1811.364	3632.728	3658.178
30%	SL	IM SE	0.165	0.016	0.100	0.026	0.119			
		MC Sd	0.128	0.012	0.106	0.022	0.113			
		MC Mean	1.394	1.071	16.922	0.774	0.928	-2430.493	4870.987	4896.437
	N	IM SE	0.538	0.039	0.486	0.013	0.051			
		MC Sd	0.274	0.035	4.975	0.063	0.250			

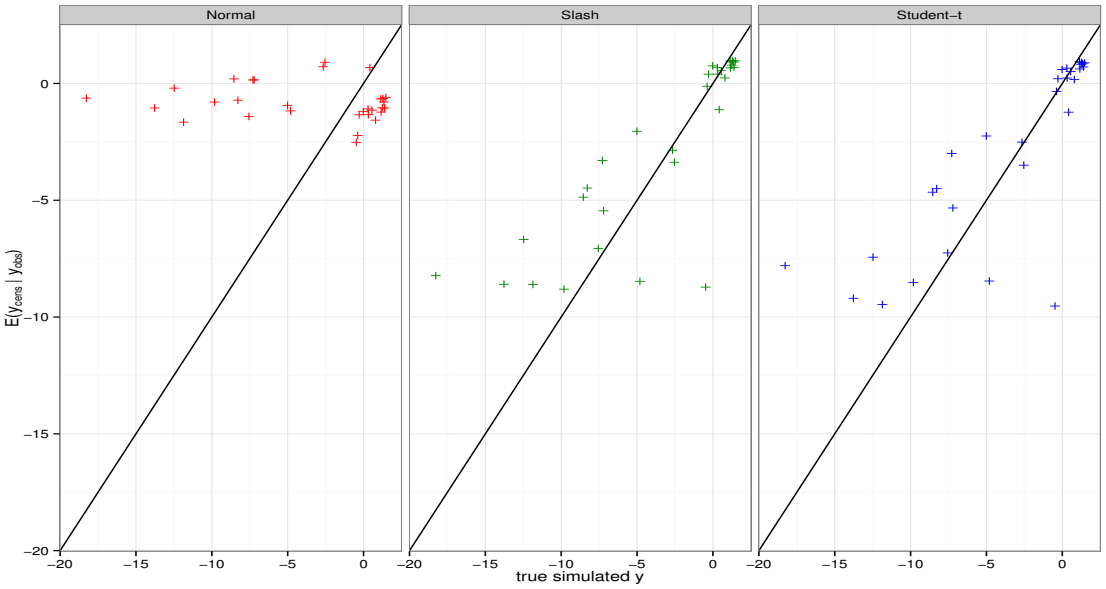

Figure 18 – **Simulation study - Scenario 3.** Conditional expectation of the censored values ($E[\mathbf{y}_{cens} \mid \mathbf{y}_{obs}]$) evaluated by the SAEM algorithm as a function of the true censored simulated values $\mathbf{y}$.

values $\mathbf{y}$. As expected, the SAEM algorithm provides a satisfactory imputation for these censored values when heavy-tailed distributions are used.

Figures 26, 27 and 28, given in Appendix B.2, 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.

4.4.2 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 describe in Section 1.2.

We consider the SMN-CR models with DEC structure defined in Subsection 4.2.1 to fit this dataset. We considered five different correlation structures, namely the uncorrelated structure (UNC), continuous-time autoregressive of order 1 (AR(1)), first-order moving average (MA(1)), compound symmetric structure (CS) and damped exponential correlation (DEC) (without fixing parameters ϕ_1 and ϕ_2). Here, $y_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i$ where y_i is the $\log_{10}$ HIV RNA for subject i from follow-up times, with $t_1 = 0$, $t_2 = 1$, $t_3 = 3$, $t_4 = 6$, $t_5 = 9$, $t_6 = 12$, $t_7 = 18$, and $t_8 = 24$; and $\mathbf{X}_i$ the design matrix.

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 of correlation structures. Observing

Table 14 – **UTI data.** Information criteria for the SMN-CR models under different structures.

Distribution	Criteria	Structure				
		DEC	AR(1)	MA(1)	CS	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
	$\boldsymbol{\nu}$	**(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	-	-	-	-	-

Table 14, the CN-CR model with $\boldsymbol{\nu} = (0.2, 0.1)$ and DEC structure outperforms all other competitors. Moreover, for these models, the estimated values of $\boldsymbol{\nu}$ are fairly small, indicating a lack of adequacy of the normal assumption for the UTI data.

Table 15 reports the ML estimates and standard errors for the model parameters from the four fitted SMN models under DEC structure . 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.*, $\beta_4 \ldots, \beta_8$, are also reasonably close to each other. The standard error estimates of $\boldsymbol{\beta}$ 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 $j = 1, \ldots, 8$, increase gradually under these models. This signifies 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 Figure 29 (Appendix B.3). As can be seen, the convergence can be achieved very quickly.

We are also interested in investigating the performance of the prediction for future values described in Section 4.3. 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. We exclude the last two measurements of each individual in the datasets with more than 6

Table 15 – **UTI data.** ML estimates with standard errors for the SMN-CR models under DEC structure.

	T		SL		CN		N	
Parameter	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)

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

$$\text{MAE} = \frac{1}{m}\sum_{i,j}|y_{ij} - y^*_{ij}| \quad \text{and} \quad \text{MSE} = \frac{1}{m}\sum_{i,j}(y_{ij} - y^*_{ij})^2, \tag{4.21}$$

where y_{ij} is the original value and y^*_{ij} is the predicted value, for $i = 1, \ldots, 29$, $j = 1, 2$ and $m = 58$. Table 16 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.

Table 16 – **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

In addition, for the CN-CR model (our best model), we present in Figure 19 a comparison between the predicted values and the real ones considering the five different correlation structures. 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.

4.4.3 Real Data - AIEDRP study

This study is taken from the AIEDRP program describe in Section 1.2. For this data we consider the same model of Vaida and Liu (2009) and Matos et al. (2013b) without the random effects. We fit a right-censored five-parameter SMN-NCR model with DEC structure, as follows

$$\mathbf{y}_i = \mathbf{g}(\boldsymbol{\varphi}_i, \mathbf{t}_i) + \boldsymbol{\epsilon}_i, \tag{4.22}$$

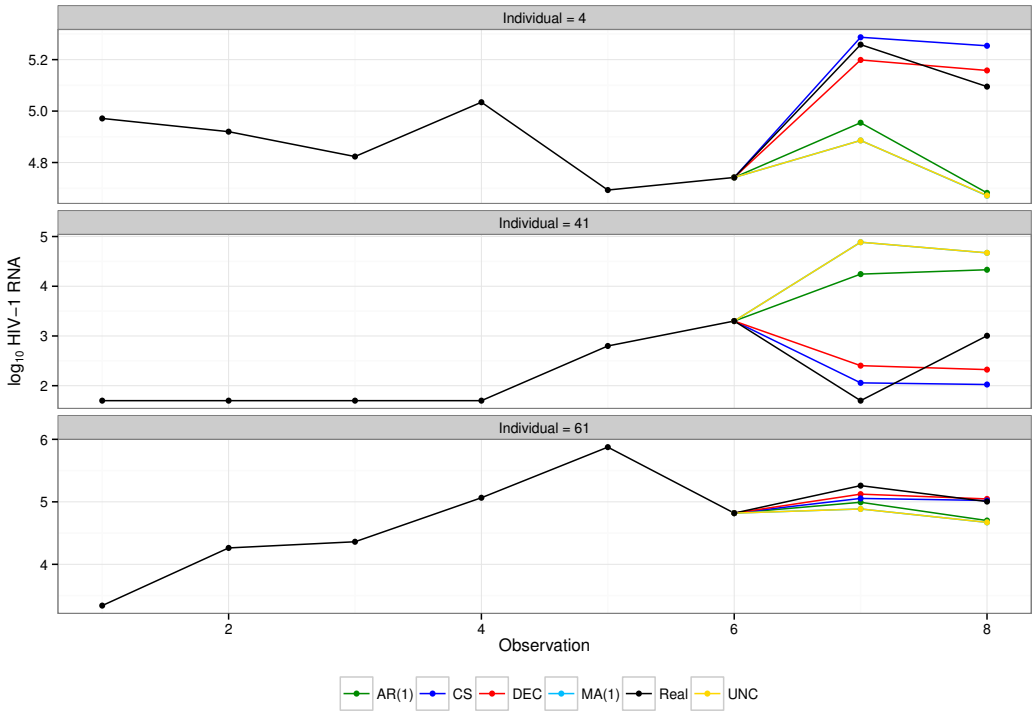

Figure 19 – **UTI data.** Evaluation of the prediction performance for three random subjects, considering the CN-CR model under different correlation structures.

where $\varphi_i = \mathbf{A}_i\boldsymbol{\beta}$, with $\mathbf{A}_i = \mathbf{1}_5$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_5)^\top$, and

$$\mathbf{g}(\varphi_i, \mathbf{t}_i) = e^{\varphi_1} + \frac{e^{\varphi_2}}{1 + \exp((\mathbf{t}_i - e^{\varphi_3})/e^{\varphi_4})} + e^{\varphi_5}(\mathbf{t}_i - 50). \tag{4.23}$$

In this study, y_{ij} is the $\log_{10}$ of the viral load for subject i at time t_{ij}. The parameters φ_1 and φ_2 represent the subject-specific set-point values and decrease from the maximum HIV-1 RNA. The location parameter φ_3 indicates the time point at which half of the change in HIV-1 RNA is attained, φ_4 is a scale parameter modeling the rate of decline and φ_5 allows increasing the HIV-1 RNA trajectory after day 50. We adopted the exponential function for each model parameter to avoid negative values.

Table 17 – **AIEDRP data.** Model selection criterion for the NCR model under DEC structure.

	Distribution			
Criterion	N	T	SL	CN
ℓ_{max}	-769.54	**-762.13**	-762.46	-762.60
AIC	1555.07	**1540.27**	1540.91	1541.19
BIC	1592.85	**1578.04**	1578.68	1578.961
$\boldsymbol{\nu}$	-	**10**	2.4	(0.1,0.3)

As in the first real data, the degrees of freedom ($\boldsymbol{\nu}$) for the Student's-t, slash and contaminated normal models are assumed to be unknown but fixed. According to the AIC (or BIC) values, the appropriate values of $\boldsymbol{\nu}$ vary under different types of correlation

structures. For all SMN distribution (N, T, SL and CN), the DEC structure fitted better than the others correlation structures. Observing Table 17, the T-NCR model with DEC structure and $\nu = 10$ outperforms all the other SMN competitors.

Table 18 summarizes the ML estimates and standard errors for the model parameters from the four fitted SMN models. As in the simulation study, the SE values for the parameter estimates are obtained using the empirical information matrix. From this table, the standard errors under the heavy-tailed models are smaller than the normal one, reflecting that the heavy-tailed models produces more precise estimates.

Table 18 – **AIEDRP data.** ML estimates with standard errors for the SMN-NCR models under DEC structure.

	N		T		SL		CN	
Parameter	Estimative	SE	Estimative	SE	Estimative	SE	Estimative	SE
β_1	1.580	0.021	**1.590**	**0.017**	1.588	0.018	1.587	0.018
β_2	0.387	0.155	**0.327**	**0.119**	0.338	0.123	0.349	0.128
β_3	3.543	0.034	**3.541**	**0.025**	3.536	0.026	3.528	0.027
β_4	1.603	0.258	**1.413**	**0.225**	1.390	0.227	1.426	0.232
β_5	-0.002	0.002	**-0.003**	**0.002**	-0.003	0.002	-0.003	0.002
σ^2	0.733	0.061	**0.642**	**0.064**	0.477	0.045	0.645	0.058
ϕ_1	0.841	0.028	**0.872**	**0.026**	0.875	0.025	0.876	0.025
ϕ_2	0.342	0.064	**0.383**	**0.070**	0.389	0.068	0.394	0.067

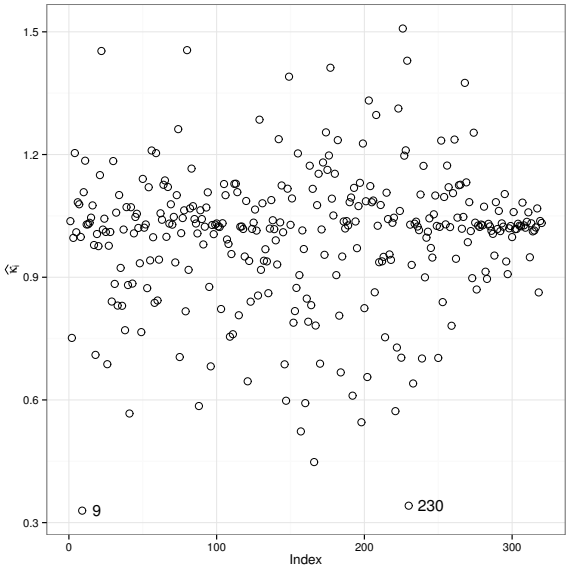

Figure 20 – **AIEDRP data.** Estimated weight $\widehat{\kappa}_i$ for the T-NCR fit. The influential observations are numbered.

It is well known that outlying observations may affect the estimation of the parameters under the normality assumption. If we use the heavy-tailed distributions, the SAEM algorithm allows one to accommodate discrepant observations attributing small weights to them in the estimation procedure. The estimated weights ($\widehat{\kappa}_i$, $i = 1, \ldots, 320$) for the T-NCR model with DEC structure (our best model) are presented in Figure 20.

We found that the observations #9 and #230 seems to be possible outliers receiving small weight.

Table 19 – **AIEDRP data.** Evaluation of the prediction accuracy for the T-NCR model under different correlation structures.

	DEC	AR(1)	CS	MA(1)	UNC
MSE	**0.212**	0.516	0.280	0.640	0.639
MAE	**0.323**	0.539	0.395	0.618	0.618

To compare the performance of the prediction for future values, we compute the predicted values under T-NCR model with five types of correlation structures (AR(1), MA(1), CS, UNC, and DEC). As in the first application, we exclude the last two measurements of each individual in the datasets with more than 6 observations (total of 36 individuals), namely $i = 1, \ldots, 36$, $j = 1, 2$ and $m = 72$. Table 19 shows the comparison between the predicted values and real ones under the T-NCR model. The MAE and MSE values indicate that the T-NCR model with DEC structure outperforms its competitors.

Besides, for the T-NCR model, we present in Figure 21 a comparison between the predicted values and the real ones considering the five different correlation structures. It is clearly seen that the T-NCR model with DEC structure has a better performance in terms of prediction than the other ones.

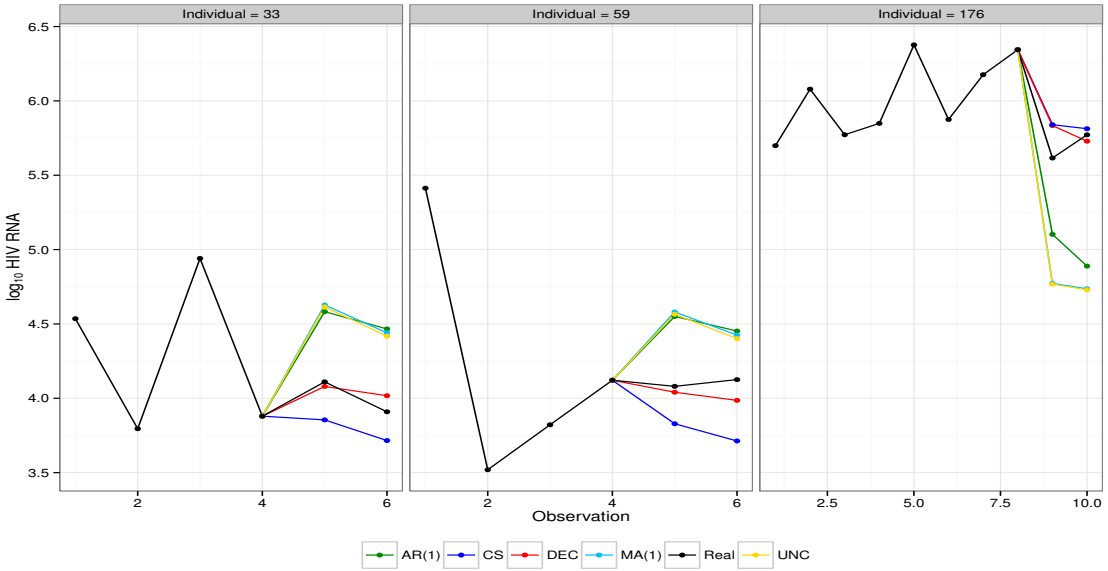

Figure 21 – **AIEDRP data.** Evaluation of the prediction performance for three random subjects, considering the T-NCR model under different correlation structures.

4.5 Conclusions

In this chapter, 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 and a nonlinear case. 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-NCR 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.

We applied our methods to two AIDS studies and undertake a simulation study to demonstrate the superiority of SMN-NCR 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-NCR 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.

5 Heavy-tailed longitudinal linear mixed models for multiple censored responses data

5.1 Introduction

According to Alder (2001), acquired immune deficiency syndrome (AIDS) is defined as an illness characterised by the laboratory evidence of the human immunodeficiency virus (HIV) infection in the human body. This type of evidence is commonly determined by using anti-HIV tests based on antibodies to HIV (HIV antigens) and also by detecting HIV genome sequences through the polymerase chain reaction (PCR) technique. The PCR amplification of HIV RNA is the most accurate way to detect and quantify the virus present in blood. Moreover, PCR amplification also provides rapid access to the HIV genome and can lead to characterisation of an HIV isolate to strain level (Mortimer and Loveday, 2001).

However, this sophisticated molecular technique quantifies the viral load in blood subjected to a limit of detection. In other words, assays based on the PCR amplification presents a threshold values below or above which the measurements of the HIV RNA are not quantifiable. For example, and as was mentioned by Barletta et al. (2004), the Roche Amplicor HIV-1 Monitor Test (versions 1.0 and 1.5; Roche Molecular Systems, Basel, Switzerland) presents a range of detection from 400 to 750,000 HIV-1 RNA copies per milliliter when using the Standard Specimen Processing Procedure (200 μL of sample), varying from 50 to 100,000 copies per milliliter when using the Ultrasensitive Specimen Processing Procedure (500 μL of sample centrifuged at 23,600g for 1 hour to concentrate virions) .

From the statistical viewpoint, the situation described above deals with the problem of censored statistical models in which the observed response lies in a restricted interval (to the left, right or both of them). Moreover, since the HIV data is generally obtained from follow-up studies (AIDS clinical trials), censored linear and nonlinear mixed effects models (see for example Wu, 2010, and references therein) and regression models with a specific correlation structures on the error term (see Matos et al., 2016) are considered to study in detail the effects of specific antiretroviral (ARV) therapies in infected persons.

Generally, in the statistical literature those models considered a Gaussian assumption for the random components of the model motivated by the computational easiness in the parameter estimation (see Vaida and Liu, 2009). However in some cases, specifically AIDS studies, such an assumption could not be realistic because the heavy-tailed

behaviour of the data. In particular, atypical observations and/or the misspecification of the parametric distributions for both random effects errors affect the likelihood based inference. Recently, Matos et al. (2013b, 2015) have considered the use of Student-t distribution in the context of censored linear and nonlinear mixed-effects (tLMEC/tNLMEC) models, including influence diagnostics with different perturbation schemes. Even though the tLMEC and tNLMEC models have been used for the analysis of longitudinal data, they are restricted to univariate censored repeated measures data.

In longitudinal studies, it is quite commonly to observe more than one series of responses repeatedly measured on each subject across time. This type of data is the so-called multivariate longitudinal data, being analyzed (in general) using the multivariate linear mixed-effect (MLME) model proposed by Shah et al. (1997). As an extension of this model, Wang (2013) proposed the multivariate t linear mixed model (tMLME), which has been considered to be a robust approach for modeling multioutcome continuous repeated measures in the presence of outliers or heavy-tailed noises. From the censoring viewpoint, Wang et al. (2015) extended the tMLME allowing the analysis of multiple longitudinal censored outcomes and heavy-tails, arising the censored multivariate t linear mixed (tMLMEC) model. For estimating the parameters of the tMLMEC, an exact EM algorithm for maximum likelihood (ML) estimation is developed based on the mean and variance of a truncated multivariate Student-t distribution provided by Ho et al. (2012).

It is important to stress that a drawback of those proposals is that they are not appropriate when the probability distribution of the errors and random effects terms is not the same, causing that such models do not consider different levels of heaviness in the tails of the distributions of the random effects and errors respectively. Consequently, our proposed method deals with this situation by considering the use of the class of scale mixtures of normal (SMN) distributions Andrews and Mallows (1974), assuming that the probability distributions of the error and random effects terms are different but belonging to the SMN class. This approach allows us to introduce more flexibility to the current model proposed by Wang (2013), accommodating influential and/or outlying observations generated by the misspecification of the distributions of the errors and/or random effects terms.

The motivating dataset considered here is obtained from the AIDS clinical trial – A5055 (Wang, 2013). This dataset was described in Section 1.2 and involves a total of 44 infected patients with the human immunodeficiency virus type 1 (HIV-1), where each patients were treated with one of the two potent antiretroviral (ARV) therapies. In this study, we focus on investigating the longitudinal trajectories for RNA viral load (in log-base-10 scale), denoted by $\log_{10}(\text{RNA})$, and CD4/CD8 ratio. As was mentioned early, in this case the lower detection limit for RNA viral load is 50 copies/milliliter, and therefore 33.5% (106 out of 316) of measurements lying below the limits of assay

quantification (left-censored). Figure 22 shows the trajectories of the two immunologic responses along the time visit.

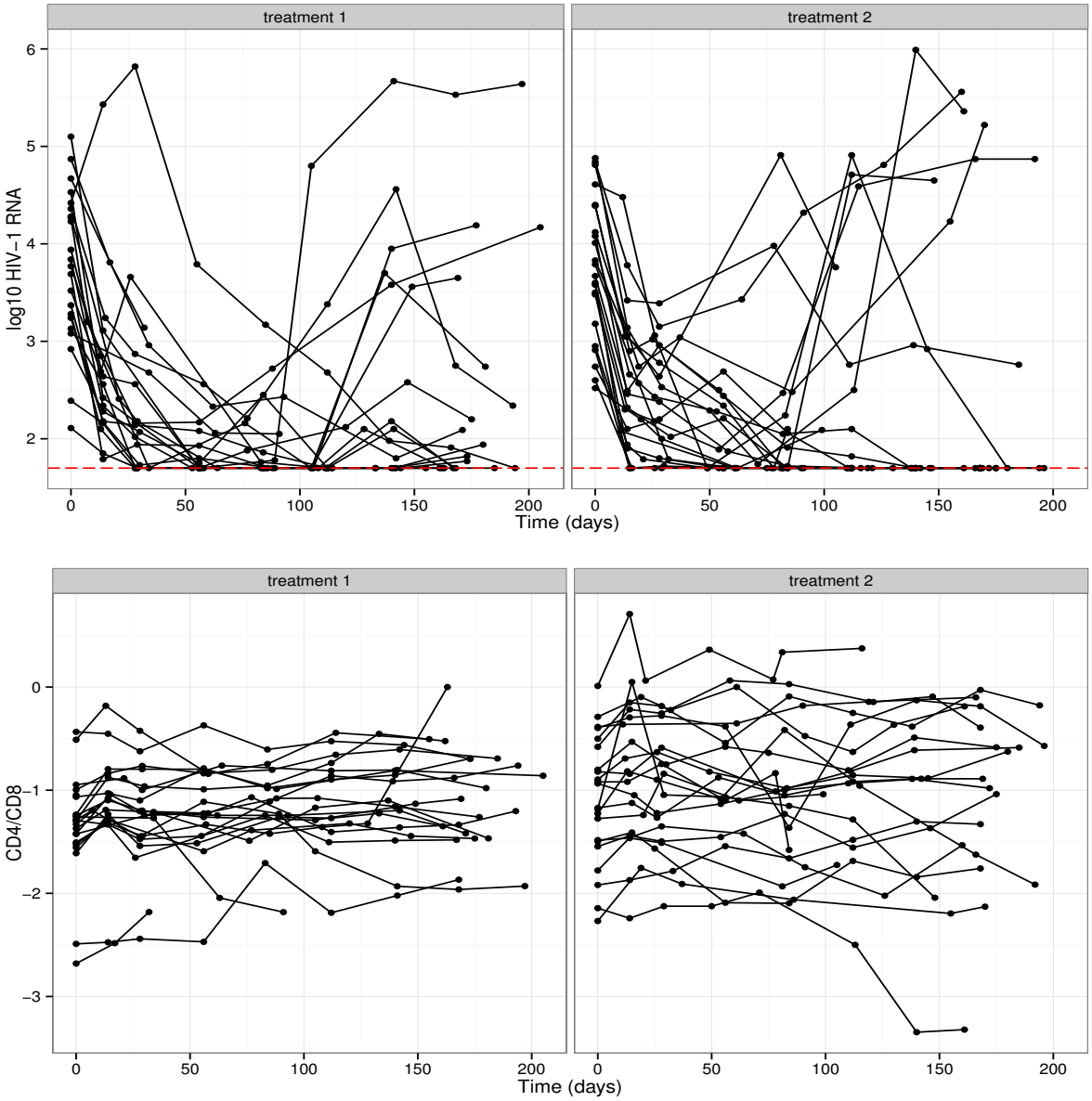

Figure 22 – **A5055 data.** Individual profiles for $\log_{10}$ RNA and CD4/CD8 ratio under two treatments.

A typical algorithm for maximum likelihood (ML) estimation in censoring models involving the class of SMN distributions is the EM-type algorithm and its variants (see, for example, Matos et al., 2013b). However, in some cases, EM-type algorithms are not appropriate due to the computational difficulty in the E-step, which involves the computation of expected quantities that cannot be obtained analytically. To deal with this problem, Delyon et al. (1999) proposed a stochastic approximation version of the EM algorithm called SAEM algorithm. This algorithm consists of replacing the E-step by a stochastic approximation obtained using simulated data, while the M-step remains unchanged. Jank (2006) showed that the computational effort of SAEM is much smaller

and reaches convergence in just a fraction of the simulation size when compared to Monte Carlo EM (MCEM). In this work, we develop a full likelihood approach for a censored linear mixed model based on the SMN distribution, including the implementation of the SAEM algorithm for ML estimation. Our approach also consider the prediction of unobservable values of the response and the asymptotic standard errors of the parameter estimates.

The rest of the chapter is organized as follows. The censored linear mixed-effects model using the SMN class of distributions and the ML estimation procedure based in the SAEM algorithm are described in Section 5.2. In Section 5.3, the estimation of the likelihood function as well the computation of the approximated standard errors is discussed. To examine the performance of our proposed methods, we present a small simulation study in Section 5.4. Section 5.5 is devoted to the analysis of a the AIDS clinical trial A5055. Finally, Section 5.6 closes the chapter with a short discussion of issues raised by our study and some possible directions for a future research.

5.2 Linear mixed models for multiple censored responses data

5.2.1 The statistical model

In this section we introduce the SMN multivariate linear mixed model for censored responses (*the SMN-MLMEC* model). Following Wang et al. (2015), suppose that r outcome variables of interest together with several covariates are repeatedly measured for each of n subjects at irregularly occasions over a period of time. For the ith subject $(i = 1, \ldots, n)$, let $\mathbf{Y}_i = [\mathbf{y}_{i1} : \ldots : \mathbf{y}_{ir}]$ be a $n_i \times r$ outcome matrix, where each column vector $\mathbf{y}_{ij} = (y_{ij1}, \ldots, y_{ijn_i})^\top$ is a $n_i \times 1$ vector of the jth outcome $(j = 1, \ldots, r)$ over occasion k $(k = 1, \ldots, n_i)$. Let $\mathbf{X}_{ij}$ be a $n_i \times p_j$ design matrix for fixed effects corresponding to the jth outcome of the ith subject. Further, let $\mathbf{Z}_{ij}$, which is generally a subset of $\mathbf{X}_{ij}$, be a $n_i \times q_j$ design matrix for random effects. In a block-diagonal matrix form, we write $\mathbf{X}_i = \text{diag}\{\mathbf{X}_{i1}, \ldots, \mathbf{X}_{ir}\}$, $\mathbf{Z}_i = \text{diag}\{\mathbf{Z}_{i1}, \ldots, \mathbf{Z}_{ir}\}$ and $\mathbf{E}_i = [\boldsymbol{\epsilon}_{i1} : \ldots : \boldsymbol{\epsilon}_{ir}]$, which is a $n_i \times r$ matrix of within-subject errors, where the jth column $\boldsymbol{\epsilon}_{ij}$ corresponds to the error for the jth outcome $\mathbf{y}_{ij}$.

For writing the model we ignore the censoring scheme for the moment, using the vec$(\cdot)$ operator to stack all columns of a matrix vertically. Then, $\mathbf{y}_i = \text{vec}(\mathbf{Y}_i) = (\mathbf{y}_{i1}^\top, \ldots, \mathbf{y}_{ir}^\top)^\top$, and $\boldsymbol{\epsilon}_i = \text{vec}(\mathbf{E}_i) = (\boldsymbol{\epsilon}_{i1}^\top, \ldots, \boldsymbol{\epsilon}_{ir}^\top)^\top$, which are of dimension $s_i = n_i \times r$. The linear mixed-effect model for the ith subject can be written as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n, \tag{5.1}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_r^\top)^\top$ is the $p \times 1$ vector of fixed effects associated with the design matrix $\mathbf{X}_i$ and $\mathbf{b}_i = (\mathbf{b}_{i1}^\top, \ldots, \mathbf{b}_{ir}^\top)^\top$ is the $q \times 1$ vector of random effects associated with

the design matrix $\mathbf{Z}_i$, with $p = \sum_{j=1}^{r} p_j$ and $q = \sum_{j=1}^{r} q_j$.

Instead of the usual assumption of normality for the errors and random effects, we replace the multivariate normal distribution by the multivariate SMN distributions. Therefore, the model can be expressed as

$$\begin{aligned} \mathbf{y}_i \mid \mathbf{b}_i &\overset{\text{ind.}}{\sim} \text{SMN}_{s_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \mathbf{R}_i; \text{H}_1), \\ \mathbf{b}_i &\overset{\text{ind.}}{\sim} \text{SMN}_q(\mathbf{0}, \mathbf{D}; \text{H}_2), \quad i = 1, \ldots, n. \end{aligned} \tag{5.2}$$

Using the stochastic representation (1.1), the hierarchical representation (four-stages) to the model defined in (5.1) – (5.2) is given by

$$\begin{aligned} \mathbf{y}_i \mid \mathbf{b}_i, \kappa_i &\overset{\text{ind.}}{\sim} \text{N}_{s_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \kappa_i^{-1}\mathbf{R}_i), \\ \mathbf{b}_i \mid \tau_i &\overset{\text{ind.}}{\sim} \text{N}_q(\mathbf{0}, \tau_i^{-1}\mathbf{D}), \\ \kappa_i &\overset{\text{ind.}}{\sim} \text{H}_1(\boldsymbol{\nu}) \\ \tau_i &\overset{\text{ind.}}{\sim} \text{H}_2(\boldsymbol{\eta}), \quad i = 1, \ldots, n, \end{aligned} \tag{5.3}$$

where $\mathbf{D} = \mathbf{D}(\boldsymbol{\alpha}) = [\mathbf{D}_{jj'}]$ is a $q \times q$ dispersion matrix that depends on the unknown and reduced parameters $\boldsymbol{\alpha}$, with $D_{jj'}$ being a partition matrix, in particular for $j = j'$, D_{jj} is a covariance structure of random effects for the jth outcome, and for $j \neq j'$, $D_{jj'}$ is the covariance for a pair of outcome variables. The variables κ_i and τ_i are assumed to be mutually independent, in which $\text{H}_1 = \text{H}_1(\cdot, \boldsymbol{\nu})$ and $\text{H}_2 = \text{H}_2(\cdot, \boldsymbol{\eta})$ are the *cdf* generator that determines the specific SMN model that are chosen. We also assume that the within-subject errors for the response at different time points have serial correlation described by the $n_i \times n_i$ autocorrelation matrix $\boldsymbol{\Omega}_i = \boldsymbol{\Omega}_i(\boldsymbol{\phi}; \mathbf{t}_i)$. This matrix presents a parsimonious dependence structure involving only parameter $\boldsymbol{\phi}$ and measurement time $\mathbf{t}_i$ of subject i, and that for the multiple responses at a particular occasion are correlated with an $r \times r$ variance-covariance matrix $\boldsymbol{\Sigma} = [\sigma^2_{jj'}]$. Accordingly, $\mathbf{R}_i = \boldsymbol{\Sigma} \otimes \boldsymbol{\Omega}_i$, where $\otimes$ denotes the Kronecker product.

Now, we include the censoring scheme, focusing on the left-censored case. Specifically, we assume that the observations are of the form

$$\begin{aligned} y_{ijk} &\leqslant V_{ijk} \quad if \quad C_{ijk} = 1, \\ y_{ijk} &= V_{ijk} \quad if \quad C_{ijk} = 0, \end{aligned} \tag{5.4}$$

where V_{ijk} represents the uncensored observation or limit of quantification and C_{ijk} is the censoring indicator whose value equals to one if censored observation and zero if uncensored observation. The observed data for the i-th subject is represented by $(\mathbf{V}_i, \mathbf{C}_i)$, where $\mathbf{V}_i = vec(\mathbf{V}_{i1}, \ldots, \mathbf{V}_{ir})$ is a s_i vector and $\mathbf{C}_i = vec(\mathbf{C}_{i1}, \ldots, \mathbf{C}_{ir})$ is a $s_i \times 1$ vector, with $\mathbf{V}_{ij} = (V_{ij1}, \ldots, V_{ijn_i})^\top$ and $\mathbf{C}_{ij} = (C_{ij1}, \ldots, C_{ijn_i})^\top$.

For modelling the correlation structure between the observations of a specific subject, we follow Muñoz et al. (1992) adopting a DEC (damped exponential correlation) structure for $\boldsymbol{\Omega}_i$ defined in (1.3).

5.2.2 Maximum likelihood estimation

This section is devoted to the SAEM algorithm for ML estimation of the parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\sigma}^\top, \boldsymbol{\alpha}^\top, \boldsymbol{\phi}^\top, \boldsymbol{\nu}^\top, \boldsymbol{\eta}^\top)^\top$ in the *SMN-MLMEC* model defined previously. In this case, $\boldsymbol{\alpha}$ is the upper triangular parameters of $\mathbf{D}$ and $\boldsymbol{\sigma}$ is the upper triangular parameters of $\boldsymbol{\Sigma}$.

Consider the model defined in (5.1) – (5.4), $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_n)^\top$, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n)^\top$, $\mathbf{V} = vec(\mathbf{V}_1, \ldots, \mathbf{V}_n)$, and $\mathbf{C} = vec(\mathbf{C}_1, \ldots, \mathbf{C}_n)$ such that we observe $(\mathbf{V}_i, \mathbf{C}_i)$ for the i-th subject. Treating $\mathbf{y}$, $\mathbf{b}$, $\boldsymbol{\kappa}$ and $\boldsymbol{\tau}$ as missing data, and augmenting with the observed data $\mathbf{V}, \mathbf{C}$, we set $\mathbf{y}_c = (\mathbf{V}^\top, \mathbf{C}^\top, \mathbf{y}^\top, \mathbf{b}^\top, \boldsymbol{\kappa}^\top, \boldsymbol{\tau}^\top)^\top$ as the complete data. Therefore, the complete data log-likelihood function for all subjects can be written, using the representation defined in (5.3), as $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}|\mathbf{y}_c)$,

$$\begin{aligned}
\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) &= \sum_{i=1}^{n} [\log f(\mathbf{y}_i|\mathbf{b}_i, \kappa_i) + \log f(\mathbf{b}_i|\tau_i) + \log h_1(\kappa_i|\boldsymbol{\nu}) + \log h_2(\tau_i|\boldsymbol{\eta})] \\
&= -\frac{1}{2}\sum_{i=1}^{n} \log|\mathbf{R}_i| - \frac{1}{2}\sum_{i=1}^{n} \kappa_i (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i) \\
&- \frac{1}{2}\sum_{i=1}^{n} \log|\mathbf{D}| - \frac{1}{2}\sum_{i=1}^{n} \tau_i \mathbf{b}_i^\top \mathbf{D}^{-1}\mathbf{b}_i + \sum_{i=1}^{n} \log h_1(\kappa_i|\boldsymbol{\nu}) + \sum_{i=1}^{n} \log h_2(\tau_i|\boldsymbol{\eta}) + K,
\end{aligned}$$

with K 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) $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, the conditional expectation of the complete log-likelihood function is given by:

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

where

$$\begin{aligned}
Q_i\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}\right) &= E\left[\log h_1(\kappa_i|\boldsymbol{\nu})|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] + E\left[\log h_2(\tau_i|\boldsymbol{\eta})|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] \\
&- \frac{1}{2}\log|\widehat{\mathbf{D}}^{(k)}| - \frac{1}{2}E\left[\tau_i \mathbf{b}_i^\top \mathbf{R}_i^{-1}\mathbf{b}_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] - \frac{1}{2}\sum_{i=1}^{n} \log|\widehat{\mathbf{R}}_i^{(k)}| \\
&- \frac{1}{2}E\left[\kappa_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \widehat{\ell h_{1i}}^{(k)} + \widehat{\ell h_{2i}}^{(k)} - \frac{1}{2}\log|\widehat{\mathbf{D}}^{(k)}| - \frac{1}{2}\mathrm{tr}\Big(\widehat{\tau\mathbf{b}_i^2}^{(k)}\widehat{\mathbf{D}}_i^{-1(k)}\Big) - \frac{1}{2}\sum_{i=1}^{n}\log|\widehat{\mathbf{R}}_i^{(k)}| \\
&- \frac{1}{2}\Bigg[\mathrm{tr}\Big(\widehat{\kappa\mathbf{y}_i^2}^{(k)}\widehat{\mathbf{R}}_i^{-1(k)}\Big) - 2\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top\widehat{\mathbf{R}}_i^{-1(k)}\widehat{\kappa\mathbf{y}}_i^{(k)} + 2\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top\widehat{\mathbf{R}}_i^{-1(k)}\mathbf{Z}_i\widehat{\kappa\mathbf{b}_i}^{(k)} \\
&- 2\mathrm{tr}\Big(\mathbf{Z}_i^\top\widehat{\mathbf{R}}_i^{-1(k)}\widehat{\kappa\mathbf{y}\mathbf{b}_i}^{(k)}\Big) + \mathrm{tr}\Big(\mathbf{Z}_i^\top\widehat{\mathbf{R}}_i^{-1(k)}\mathbf{Z}_i\widehat{\kappa\mathbf{b}_i^2}^{(k)}\Big) + \widehat{\kappa_i}^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top\widehat{\mathbf{R}}_i^{-1(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\Bigg],
\end{aligned}
$$

with

$$
\begin{aligned}
&\widehat{\ell h_{1i}}^{(k)} = E\Big[\log h_1(\kappa_i|\boldsymbol{\nu})|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\Big], \quad \widehat{\ell h_{2i}}^{(k)} = E\Big[\log h_2(\tau_i|\boldsymbol{\eta})|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\Big] \\
&\widehat{\kappa\mathbf{y}_i^2}^{(k)} = E\Big[\kappa_i\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\Big], \qquad \widehat{\kappa\mathbf{y}}_i^{(k)} = E\Big[\kappa_i\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\Big], \\
&\widehat{\kappa\mathbf{b}_i^2}^{(k)} = E\Big[\kappa_i\mathbf{b}_i\mathbf{b}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\Big], \qquad \widehat{\kappa\mathbf{b}_i}^{(k)} = E\Big[\kappa_i\mathbf{b}_i|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\Big], \\
&\widehat{\tau\mathbf{b}_i^2}^{(k)} = E\Big[\tau_i\mathbf{b}_i\mathbf{b}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\Big], \qquad \widehat{\kappa\mathbf{y}\mathbf{b}_i}^{(k)} = E\Big[\kappa_i\mathbf{y}_i\mathbf{b}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\Big], \\
&\widehat{\kappa_i}^{(k)} = E\Big[\kappa_i|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\Big].
\end{aligned} \tag{5.5}
$$

Note that the conditional expectations proposed above do not have a closed form. Consequently, we need to consider two intermediate steps, namely the simulation (S) and approximation (A) steps. In this context, for the i-th subject, a Gibbs sampler step is considered for generating samples from the full conditional distributions of the latent variables $(\mathbf{y}_i, \mathbf{b}_i, \kappa_i, \tau_i)$:

Step S1: Generate $\mathbf{y}_i^{c(k,l)}$ from $f(\mathbf{y}_i^c|\mathbf{V}_i^c,\mathbf{y}_i^o,\mathbf{b}_i^{(k,l-1)},\kappa_i^{(k,l-1)},\tau_i^{(k,l-1)},\widehat{\boldsymbol{\theta}}^{(k)})$ considering a partition of $\mathbf{y}_i$ given by $\mathbf{y}_i = vec(\mathbf{y}_i^o,\mathbf{y}_i^c)$ with $dim(\mathbf{y}_i^o) = s_i^o$, $dim(\mathbf{y}_i^c) = s_i^c$ and $s_i^o + s_i^c = s_i$, where $C_{ijk} = 0$ for all elements in $\mathbf{y}_i^o$, and 1 for all elements in $\mathbf{y}_i^c$. In addition, let $\mathbf{V}_i$, $\mathbf{X}_i$, $\mathbf{Z}_i$ and $\mathbf{R}_i$ also be partitioned as follows: $\mathbf{V}_i = vec(\mathbf{V}_i^o,\mathbf{V}_i^c)$, $\mathbf{X}_i^\top = (\mathbf{X}_i^o,\mathbf{X}_i^c)$, $\mathbf{Z}_i^\top = (\mathbf{Z}_i^o,\mathbf{Z}_i^c)$, and $\mathbf{R}_i = \begin{pmatrix}\mathbf{R}_i^{oo} & \mathbf{R}_i^{oc}\\ \mathbf{R}_i^{co} & \mathbf{R}_i^{cc}\end{pmatrix}$. Then, we have $\mathbf{y}_i \mid \mathbf{b}_i,\kappa_i \equiv \mathbf{y}_i \mid \mathbf{b}_i,\kappa_i,\tau_i$, in which $\mathbf{y}_i \mid \mathbf{b}_i,\kappa_i \sim \mathrm{N}_{s_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \kappa_i^{-1}\mathbf{R}_i)$, where

$$
\mathbf{y}_i^o \mid \mathbf{b}_i,\kappa_i \sim \mathrm{N}_{s_i^o}(\mathbf{X}_i^o\boldsymbol{\beta} + \mathbf{Z}_i^o\mathbf{b}_i, \kappa_i^{-1}\mathbf{R}_i^{oo}) \quad \text{and} \quad \mathbf{y}_i^c \mid \mathbf{y}_i^o,\mathbf{b}_i,\kappa_i \sim \mathrm{N}_{s_i^c}(\boldsymbol{\mu}_i, \kappa_i^{-1}\mathbf{S}_i), \tag{5.6}
$$

with $\boldsymbol{\mu}_i = (\mathbf{X}_i^c\boldsymbol{\beta} + \mathbf{Z}_i^c\mathbf{b}_i) + \mathbf{R}_i^{co}(\mathbf{R}_i^{oo})^{-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\boldsymbol{\beta} - \mathbf{Z}_i^o\mathbf{b}_i)$ and $\mathbf{S}_i = \mathbf{R}_i^{cc} - \mathbf{R}_i^{co}(\mathbf{R}_i^{oo})^{-1}\mathbf{R}_i^{oc}$. Conditioning on the censoring information, we have that

$$
f(\mathbf{y}_i^c|\mathbf{V}_i^c,\mathbf{y}_i^o,\mathbf{b}_i,\kappa_i,\tau_i) = \mathrm{TN}_{s_i^c}(\boldsymbol{\mu}_i,\kappa_i^{-1}\mathbf{S}_i;\mathbb{A}_i),
$$

which is a truncated normal distribution with $\mathbb{A}_i = \{\mathbf{y}_i^c = (y_{i1}^c,\ldots,y_{is_i^c}^c)^\top \mid y_{i1}^c \leqslant V_{i1}^c,\ldots,y_{is_i^c}^c \leqslant V_{is_i^c}^c\}$. The observation $\mathbf{y}_i^{(k,l)} = (y_{i1},\ldots,y_{is_i^o},y_{is_i^o+1}^{c(k,l)},\ldots,y_{is_i}^{c(k,l)})$ is a sample generated from a truncated normal distribution for the s_i^c censored cases and the observed values (uncensored cases).

Step S2: Generate $\mathbf{b}_i^{(k,l)}$ from $f(\mathbf{b}_i|\mathbf{y}_i^{(k,l)},\kappa_i^{(k,l-1)},\tau_i^{(k,l-1)},\widehat{\boldsymbol{\theta}}^{(k)})$ using the fact that

$$
f(\mathbf{b}_i|\mathbf{y}_i,\kappa_i,\tau_i) = \mathrm{N}_q(\boldsymbol{\Psi}_i\mathbf{Z}_i^\top\mathbf{R}_i^{-1}\kappa_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}),\boldsymbol{\Psi}_i),
$$

where $\boldsymbol{\Psi}_i = (\kappa_i \mathbf{Z}_i^\top \mathbf{R}_i^{-1} \mathbf{Z}_i + \tau_i \mathbf{D}^{-1})^{-1}$.

Step S3: Generate $\kappa_i^{(k,l)}$ from $f(\kappa_i | \mathbf{y}_i^{(k,l)}, \mathbf{b}_i^{(k,l)}, \tau_i^{(k,l-1)}, \widehat{\boldsymbol{\theta}}^{(k)})$.

Step S4: Generate $\tau_i^{(k,l)}$ from $f(\tau_i | \mathbf{y}_i^{(k,l)}, \mathbf{b}_i^{(k,l)}, \kappa_i^{(k,l)}, \widehat{\boldsymbol{\theta}}^{(k)})$.

Remark: Note that, given $\mathbf{y}_i \mid \mathbf{b}_i$ independent of τ_i; $\mathbf{b}_i$ independent of κ_i; and κ_i and τ_i mutually independent, then we have that $f(\kappa_i | \mathbf{y}_i, \mathbf{b}_i, \tau_i) \propto f(\mathbf{y}_i | \mathbf{b}_i, \kappa_i) f(\kappa_i)$ and $f(\tau_i | \mathbf{y}_i, \mathbf{b}_i, \kappa_i) \propto f(\mathbf{b}_i | \tau_i) f(\tau_i)$. Thus, for some particular cases it is possible to compute the full conditional distributions of $(\kappa_i | \mathbf{y}_i, \mathbf{b}_i, \tau_i)$ and $(\tau_i | \mathbf{y}_i, \mathbf{b}_i, \kappa_i)$ (see Tables 25 and 26 given in Appendix C.1).

The **Stochastic Approximation** step is performed considering the sequence $(\mathbf{y}_i^{(k,l)}, \mathbf{b}_i^{(k,l)}, \kappa_i^{(k,l)}, \tau_i^{(k,l)})$, $l = 1, \ldots, m$, at the k-th iteration, and replacing the conditional expectations given in (5.5) by their corresponding stochastic approximations.

Step A:

$$
\begin{aligned}
\widehat{\kappa \mathbf{y}_i^2}^{(k)} &= \widehat{\kappa \mathbf{y}_i^2}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^{m} \kappa_i^{(k,l)} \mathbf{y}_i^{(k,l)} \mathbf{y}_i^{(k,l)\top} - \widehat{\kappa \mathbf{y}_i^2}^{(k-1)} \right), \\
\widehat{\kappa \mathbf{y}_i}^{(k)} &= \widehat{\kappa \mathbf{y}_i}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^{m} \kappa_i^{(k,l)} \mathbf{y}_i^{(k,l)} - \widehat{\kappa \mathbf{y}_i}^{(k-1)} \right), \\
\widehat{\kappa \mathbf{b}_i^2}^{(k)} &= \widehat{\kappa \mathbf{b}_i^2}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^{m} \kappa_i^{(k,l)} \mathbf{b}_i^{(k,l)} \mathbf{b}_i^{(k,l)\top} - \widehat{\kappa \mathbf{b}_i^2}^{(k-1)} \right), \\
\widehat{\kappa \mathbf{b}_i}^{(k)} &= \widehat{\kappa \mathbf{b}_i}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^{m} \kappa_i^{(k,l)} \mathbf{b}_i^{(k,l)} - \widehat{\kappa \mathbf{b}_i}^{(k-1)} \right), \\
\widehat{\kappa \mathbf{y} \mathbf{b}_i}^{(k)} &= \widehat{\kappa \mathbf{y} \mathbf{b}_i}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^{m} \kappa_i^{(k,l)} \mathbf{y}_i^{(k,l)} \mathbf{b}_i^{(k,l)\top} - \widehat{\kappa \mathbf{y} \mathbf{b}_i}^{(k-1)} \right), \\
\widehat{\tau \mathbf{b}_i^2}^{(k)} &= \widehat{\tau \mathbf{b}_i^2}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^{m} \tau_i^{(k,l)} \mathbf{b}_i^{(k,l)} \mathbf{b}_i^{(k,l)\top} - \widehat{\tau \mathbf{y}_i^2}^{(k-1)} \right), \\
\widehat{\kappa_i}^{(k)} &= \widehat{\kappa_i}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^{m} \kappa_i^{(k,l)} - \widehat{\kappa_i}^{(k-1)} \right), \\
\widehat{\ell h_{1i}}^{(k)} &= \widehat{\ell h_{1i}}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^{m} \log h_1(\kappa_i^{(k,l)} | \boldsymbol{\nu}) - \widehat{\ell h_{1i}}^{(k-1)} \right), \\
\widehat{\ell h_{2i}}^{(k)} &= \widehat{\ell h_{2i}}^{(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^{m} \log h_2(\tau_i^{(k,l)} | \boldsymbol{\eta}) - \widehat{\ell h_{2i}}^{(k-1)} \right).
\end{aligned}
$$

As was mentioned earlier, 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 algorithm. Although some authors claim that $m \leqslant 10$ is large enough we consider $m = 20$ for being more conservative.

The conditional maximization step leads to the update of $\widehat{\boldsymbol{\theta}}^{(k)}$ as follows:

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}^{(k+1)} &= \left(\sum_{i=1}^{n} \widehat{\kappa}_i^{(k)} \mathbf{X}_i^\top \widehat{\mathbf{R}}_i^{-1(k)} \mathbf{X}_i\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i^\top \widehat{\mathbf{R}}_i^{-1(k)} \left(\widehat{\kappa \mathbf{y}_i}^{(k)} - Z_i \widehat{\kappa \mathbf{b}_i}^{(k)}\right), \\
\widehat{\sigma_{jl}^2}^{(k+1)} &= \begin{cases} (\sum_{i=1}^{n} n_i)^{-1} \sum_{i=1}^{n} \operatorname{tr}\left(\widehat{\boldsymbol{\Omega}}_i^{-1(k)} \widehat{\kappa \boldsymbol{\epsilon}}_{ijl}^{(k)}\right) & \text{for } j = l, \\ (2\sum_{i=1}^{n} n_i)^{-1} \sum_{i=1}^{n} \operatorname{tr}\left[\widehat{\boldsymbol{\Omega}}_i^{-1(k)} \left(\widehat{\kappa \boldsymbol{\epsilon}}_{ijl}^{(k)} + \widehat{\kappa \boldsymbol{\epsilon}}_{ilj}^{(k)}\right)\right] & \text{for } j \neq l, \end{cases} \\
\widehat{\boldsymbol{\phi}}^{(k+1)} &= \underset{\boldsymbol{\phi} \in (0,1) \times \mathbb{R}^+}{\operatorname{argmax}} \left\{ -\frac{r}{2} \sum_{i=1}^{n} \log |\boldsymbol{\Omega}_i(\boldsymbol{\phi}, \mathbf{t}_i)| \frac{1}{2} \sum_{i=1}^{n} \operatorname{tr}\left[\left(\widehat{\boldsymbol{\Sigma}}^{(k)} \otimes \boldsymbol{\Omega}_i(\boldsymbol{\phi}, \mathbf{t}_i)\right)^{-1} \widehat{\kappa \mathbf{E}_i}\right] \right\}, \\
\widehat{\mathbf{D}}^{(k+1)} &= \frac{1}{n} \sum_{i=1}^{n} \widehat{\tau \mathbf{b}_i^2}^{(k)}, \\
\widehat{\boldsymbol{\nu}}^{(k+1)} &= \underset{\boldsymbol{\nu}}{\operatorname{argmax}} \sum_{i=1}^{n} \widehat{\ell h_{1i}}^{(k)}(\boldsymbol{\nu}), \\
\widehat{\boldsymbol{\eta}}^{(k+1)} &= \underset{\boldsymbol{\eta}}{\operatorname{argmax}} \sum_{i=1}^{n} \widehat{\ell h_{2i}}^{(k)}(\boldsymbol{\eta}),
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{\kappa \mathbf{E}_i} &= \widehat{\kappa \mathbf{y}_i^2}^{(k)} - 2\mathbf{X}_i \boldsymbol{\beta} \widehat{\kappa \mathbf{y}}_i^{(k)^\top} - 2\mathbf{Z}_i \widehat{\kappa \mathbf{y} \mathbf{b}_i}^{(k)\top} + \widehat{\kappa}_i^{(k)} \mathbf{X}_i \boldsymbol{\beta} \boldsymbol{\beta}^\top \mathbf{X}_i^\top + 2\mathbf{X}_i \boldsymbol{\beta} \widehat{\kappa \mathbf{b}_i}^{(k)\top} \mathbf{Z}_i^\top \\
&+ \widehat{\kappa}_i^{(k)} \mathbf{X}_i \boldsymbol{\beta} \boldsymbol{\beta}^\top \mathbf{X}_i^\top + \mathbf{Z}_i \widehat{\kappa \mathbf{b}_i^2}^{(k)} \mathbf{Z}_{il}^\top,
\end{aligned}
$$

and for $j, l = 1, \ldots, r$,

$$
\begin{aligned}
\widehat{\kappa \boldsymbol{\epsilon}}_{ijl}^{(k)} &= \widehat{\kappa \mathbf{y}_{i\,[j:l]}^2}^{(k)} - \widehat{\kappa \mathbf{y}}_{ij}^{(k)} \boldsymbol{\beta}^\top \mathbf{X}_{il}^\top - \widehat{\kappa \mathbf{y} \mathbf{b}}_{i[j:l]}^{(k)} \mathbf{Z}_{il}^\top - \mathbf{X}_{ij} \boldsymbol{\beta} \widehat{\kappa \mathbf{y}}_{il}^{(k)^\top} + \widehat{\kappa}_{ij}^{(k)} \mathbf{X}_{ij} \boldsymbol{\beta} \boldsymbol{\beta}^\top \mathbf{X}_{il}^\top \\
&+ \mathbf{X}_{ij} \boldsymbol{\beta} \widehat{\kappa \mathbf{b}}_{il}^{(k)\top} \mathbf{Z}_{il}^\top - \mathbf{Z}_{ij} \widehat{\kappa \mathbf{y} \mathbf{b}}_{i[l:j]}^{(k)\top} + \mathbf{Z}_{ij} \widehat{\kappa \mathbf{b}}_{ij}^{(k)} \boldsymbol{\beta}^\top \mathbf{X}_{il}^\top + \mathbf{Z}_{ij} \widehat{\kappa \mathbf{b}_{i\,[j:l]}^2}^{(k)} \mathbf{Z}_{il}^\top,
\end{aligned}
$$

with $\widehat{\kappa \mathbf{y}}_{ij}^{(k)}$ being a $n_i \times 1$ subvector consisting of $((j-1)n_i + 1)$th to (jn_i)th entries of $\widehat{\kappa \mathbf{y}}_i^{(k)}$; $\widehat{\kappa \mathbf{b}}_{ij}^{(k)}$ being a $q_j \times 1$ subvector consisting of $(\sum_{v \leqslant (j-1)} q_v + 1)$th to $\sum_{v \leqslant j} q_v$th entries of $\widehat{\kappa \mathbf{b}}_i^{(k)}$; $\widehat{\kappa \mathbf{y}_{i\,[j:l]}^2}^{(k)}$ being a $n_i \times n_i$ submatrix consisting of the $((j-1)n_i + 1)$th to (jn_i)th rows and the $((l-1)s_i + 1)$th to (ls_i)th columns of $\widehat{\kappa \mathbf{y}_i^2}^{(k)}$; $\widehat{\kappa \mathbf{y} \mathbf{b}}_{i[j:l]}^{(k)}$ is a $s_i \times q_l$ submatrix consisting of the $((j-1)s_i + 1)$th to (js_i)th rows and the $(\sum_{v \leqslant (l-1)} q_v + 1)$th to $\sum_{v \leqslant l} q_v$th columns of $\widehat{\kappa \mathbf{y} \mathbf{b}_i}^{(k)}$; and $\widehat{\kappa \mathbf{b}_{i\,[j:l]}^2}^{(k)}$ is a $q_j \times q_l$ submatrix consisting of $(\sum_{v \leqslant (j-1)} q_v + 1)$th to $\sum_{v \leqslant j} q_v$th rows and $(\sum_{v \leqslant (l-1)} q_v + 1)$th to $\sum_{v \leqslant l} q_v$th columns of $\widehat{\tau \mathbf{b}_i^2}^{(k)}$.

Depending on the distribution for the conditional response vector and for the random effects, we have close expressions for the parameters estimates $\widehat{\nu}$ and $\widehat{\eta}$. In Appendix C.2 we present these expressions.

Remark: If we consider that the parameters $\boldsymbol{\nu}$ and $\boldsymbol{\eta}$ associated with the mixture variables κ_i and τ_i respectively are known, we have that $\widehat{\kappa \mathbf{y}_i^2}^{(k)}$, $\widehat{\kappa \mathbf{y}_i}^{(k)}$ and $\widehat{\kappa_i}^{(k)}$ depend on the minimal sufficient statistics, since the complete likelihood function belongs to the exponential family. Therefore, the parameters estimates of the SAEM algorithm converges as was mentioned by Kuhn and Lavielle (2005) and Samson et al. (2006). However, in this work we consider these parameters to be unknown. Then $\widehat{\kappa \mathbf{y}_i^2}^{(k)}$, $\widehat{\kappa \mathbf{y}_i}^{(k)}$ and $\widehat{\kappa_i}^{(k)}$ do not depend on minimal sufficient statistics and we can not guarantee the convergence of the parameters estimates. To deal with this problem, we conduct a simulation study to show the well behaviour of the SAEM algorithm.

5.2.3 Imputation of censored components

Let $\mathbf{y}_i^c$ be the true unobserved response vector for the censored components. Now, as a by-product of the SAEM algorithm, the prediction of the censored components is

$$\tilde{\mathbf{y}}_i^{c(k)} = \tilde{\mathbf{y}}_i^{c(k-1)} + \delta_k \left(\frac{1}{m} \sum_{l=1}^{m} \mathbf{y}_i^{c(k,l)} - \tilde{\mathbf{y}}_i^{c(k)} \right).$$

Note that $\mathbf{y}_i^{c(k,l)}$ is obtained from the **Step S1** of the proposed SAEM algorithm.

5.3 Estimation of the likelihood and standard errors

5.3.1 Likelihood estimation

The likelihood function for the observed data can be computed as

$$L_o(\boldsymbol{\theta}; \mathbf{y}^{obs}) = \prod_{i=1}^{n} \int \left[\int_0^{\infty} f(\mathbf{y}_i|\mathbf{b}_i, \kappa_i; \boldsymbol{\theta}) h_1(\kappa_i|\boldsymbol{\nu}) d\kappa_i \right] f(\mathbf{b}_i|\boldsymbol{\theta}) d\mathbf{b}_i.$$

Partitioning $\mathbf{y}_i$ as in (5.6), and we have that

$$\begin{aligned} L_o(\boldsymbol{\theta}; \mathbf{y}^{obs}) &= \prod_{i=1}^{n} \int \left[\int_0^{\infty} \phi_{s_i^o}(\mathbf{y}_i^o; \mathbf{X}_i^c\boldsymbol{\beta} - \mathbf{Z}_i^c\mathbf{b}_i, \kappa_i^{-1}\mathbf{R}_i^{oo}) \Phi_{s_i^c}(\mathbf{V}_i^c; \boldsymbol{\mu}_i, \kappa_i^{-1}\mathbf{S}_i) h_1(\kappa_i|\boldsymbol{\nu}) d\kappa_i \right] \\ &\times f(\mathbf{b}_i|\boldsymbol{\theta}) d\mathbf{b}_i = \prod_{i=1}^{n} \int g(\mathbf{y}_i|\mathbf{b}_i, \kappa_i; \boldsymbol{\theta}) f(\mathbf{b}_i|\boldsymbol{\theta}) d\mathbf{b}_i \end{aligned} \tag{5.7}$$

where $g(\mathbf{y}_i|\mathbf{b}_i, \kappa_i; \boldsymbol{\theta}) = \int_0^{\infty} \phi_{s_i^o}(\mathbf{y}_i^o; \mathbf{X}_i^c\boldsymbol{\beta} - \mathbf{Z}_i^c\mathbf{b}_i, \kappa_i^{-1}\mathbf{R}_i^{oo}) \Phi_{s_i^c}(\mathbf{V}_i^c; \boldsymbol{\mu}_i, \kappa_i^{-1}\mathbf{S}_i) h_1(\kappa_i|\boldsymbol{\nu}) d\kappa_i$.

The integral involved in (5.7) can be compute using a importance sampling strategy for any continuous distribution $f^{\star}$. In fact, we have that

$$L_o(\boldsymbol{\theta}; \mathbf{y}^{obs}) = \prod_{i=1}^{n} \int g(\mathbf{y}_i|\mathbf{b}_i, \kappa_i; \boldsymbol{\theta}) \frac{f(\mathbf{b}_i|\boldsymbol{\theta})}{f^{\star}(\mathbf{b}_i|\boldsymbol{\theta})} d\mathbf{b}_i,$$

where the distribution $f^\star$ is the importance distribution. Consequently, $L_o(\boldsymbol{\theta}; \mathbf{y}_i^{obs})$ is estimated through the following approximation

$$L_o(\boldsymbol{\theta}; \mathbf{y}^{obs}) = \prod_{i=1}^{n} \left[\frac{1}{M} \sum_{m=1}^{M} g(\mathbf{y}_i | \mathbf{b}_{im}, \kappa_i; \boldsymbol{\theta}) \frac{f(\mathbf{b}_{im}|\boldsymbol{\theta})}{f^\star(\mathbf{b}_{im}|\boldsymbol{\theta})} \right],$$

with $\mathbf{b}_{i1}, \ldots, \mathbf{b}_{im}$ being draw from $f^\star(\mathbf{b}_i|\boldsymbol{\theta})$.

5.3.2 Model selection criteria

Follow the idea of Zhang et al. (2014), our goal is to assess the contribution of using the joint model, *i.e*, the model with multiple outcome. In this context, the model selection criteria used in this work are

$$\text{AIC} = 2\,m - 2\,\ell_{max} \;\text{ and }\; \text{BIC} = m\,log\,N - 2\,\ell_{max},$$

where m is the number of model parameters, $N = \sum_{i=1}^{n} n_i$ and ℓ_{max} is the maximized log-likelihood value.

To assess the contribution of using the joint model, we need to decompose both model selection criteria into two parts. Let $\mathbf{y}_{i1}^\star = (\mathbf{y}_{i1}^\top, \ldots, \mathbf{y}_{ir^\star}^\top)^\top$ and $\mathbf{y}_{i2}^\star = (\mathbf{y}_{ir^\star+1}^\top, \ldots, \mathbf{y}_{ir}^\top)^\top$, where $\mathbf{y}_i = (\mathbf{y}_{i1}^{\star\top}, \mathbf{y}_{i2}^{\star\top})^\top$ and $r^\star \in \{1, \ldots, r\}$. Also let $f(\mathbf{y}_{i1}^\star; \boldsymbol{\theta}_1^\star)$ be the marginal density of $\mathbf{y}_{i1}^\star$ and $f(\mathbf{y}_{i2}^\star|\mathbf{y}_{i1}^\star; \boldsymbol{\theta}_2^\star; \boldsymbol{\theta}_1^\star)$ be the conditional density of the partition $\mathbf{y}_{i2}^\star$ given $\mathbf{y}_{i1}^\star$, where $\ell(\boldsymbol{\theta}_1^\star; \mathbf{y}_i)$ and $\ell(\boldsymbol{\theta}_2^\star; \mathbf{y}_i)$ are the log-likelihood function of $\mathbf{y}_{i2}^\star|\mathbf{y}_{i1}^\star$ and $\mathbf{y}_{i1}^\star$ respectively (see Appendix C.3 for more details about these conditional distributions). Then the AIC and BIC has the following decomposition

$$\text{AIC} = \text{AIC}_{\mathbf{y}_1^\star} + \text{AIC}_{\mathbf{y}_2^\star|\mathbf{y}_1^\star} \;\text{ and }\; \text{BIC} = \text{BIC}_{\mathbf{y}_1^\star} + \text{BIC}_{\mathbf{y}_2^\star|\mathbf{y}_1^\star},$$

where $\text{AIC}_{\mathbf{y}_1^\star} = 2\dim(\boldsymbol{\theta}_1^\star) - 2\sum_{i=1}^{n} \ell(\widehat{\boldsymbol{\theta}}_1^\star; \mathbf{y}_i)$, $\text{AIC}_{\mathbf{y}_2^\star|\mathbf{y}_1^\star} = 2\dim(\boldsymbol{\theta}_2^\star) - 2\sum_{i=1}^{n} \ell(\widehat{\boldsymbol{\theta}}_2^\star; \mathbf{y}_i)$, $\text{BIC}_{\mathbf{y}_1^\star} = \dim(\boldsymbol{\theta}_1^\star)\,log\,N - 2\sum_{i=1}^{n} \ell(\widehat{\boldsymbol{\theta}}_1^\star; \mathbf{y}_i)$, $\text{BIC}_{\mathbf{y}_2^\star|\mathbf{y}_1^\star} = \dim(\boldsymbol{\theta}_2^\star)\,log\,N - 2\sum_{i=1}^{n} \ell(\widehat{\boldsymbol{\theta}}_2^\star; \mathbf{y}_i)$, and $\widehat{\boldsymbol{\theta}}_1^\star$ and $\widehat{\boldsymbol{\theta}}_2^\star$ are the estimates of $\boldsymbol{\theta}$. Following Zhang et al. (2014), $\dim(\boldsymbol{\theta}_2^\star) = \dim(\boldsymbol{\theta}) - \dim(\boldsymbol{\theta}_1^\star)$.

Finally, we define the model assessment criteria

$$\Delta\text{AIC} = \text{AIC}_{\mathbf{y}_{2,0}^\star} - \text{AIC}_{\mathbf{y}_2^\star|\mathbf{y}_1^\star} \;\text{ and }\; \Delta\text{BIC} = \text{BIC}_{\mathbf{y}_{2,0}^\star} - \text{BIC}_{\mathbf{y}_2^\star|\mathbf{y}_1^\star}, \tag{5.8}$$

where $\text{AIC}_{\mathbf{y}_{2,0}^\star}$ and $\text{BIC}_{\mathbf{y}_{2,0}^\star}$ are calculated considering only $\mathbf{y}_2^\star$. In this case, greater the ΔAIC and ΔBIC better the model.

5.3.3 Empirical information matrix

The individual score can be determined as

$$\mathbf{s}(\mathbf{y}_i \mid \boldsymbol{\theta}) = \frac{\partial \log f(\mathbf{y}_i \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = E\left(\frac{\partial \ell_i(\boldsymbol{\theta} \mid \mathbf{y}_c)}{\partial \boldsymbol{\theta}} \mid \mathbf{V}_i, \mathbf{C}_i, \boldsymbol{\theta} \right),$$

where $\ell_i(\boldsymbol{\theta} \mid \mathbf{y}_c)$ is the complete data log-likelihood function formed from the i-th observation of $\mathbf{y}_c$ (see Louis, 1982). Then, the empirical information matrix is given by

$$\mathbf{I}_e(\widehat{\boldsymbol{\theta}} \mid \mathbf{y}) = \sum_{i=1}^{n} \widehat{\mathbf{s}}_i \widehat{\mathbf{s}}_i^{\top}, \tag{5.9}$$

where

$$\widehat{\mathbf{s}}_i = (\widehat{\mathbf{s}}_{i,\beta_1}, \ldots, \widehat{\mathbf{s}}_{i,\beta_p}, \widehat{\mathbf{s}}_{i,\sigma_1^2}, \ldots, \widehat{\mathbf{s}}_{i,\sigma_{r*}^2}, \widehat{\mathbf{s}}_{i,\alpha_1}, \ldots, \widehat{\mathbf{s}}_{i,\alpha_{q*}}, \widehat{\mathbf{s}}_{i,\phi_1}, \widehat{\mathbf{s}}_{i,\phi_2}, \widehat{\mathbf{s}}_{i,\nu}, \widehat{\mathbf{s}}_{i,\eta})^{\top}. \tag{5.10}$$

The elements of $\widehat{\mathbf{s}}_i$ are given in Appendix C.4.

5.4 Simulation study

In this section, we conduct an extensive simulation scheme to study the empirical performance of the parameter estimates. We consider three types of models, namely, M1, M2 and M3, where M1 is the model with the parameters $\boldsymbol{\nu}$ and $\boldsymbol{\eta}$ considered to be unknown; M2 is the model considering $\boldsymbol{\nu}$ and $\boldsymbol{\eta}$ fixed at the true values and M3 is the model considering $\boldsymbol{\nu}$ and $\boldsymbol{\eta}$ fixed at different values than the true ones. Our goal in this study is to compare how affected are the estimates when the parameters associated with the mixture variables are either estimated or considered as fixed.

We generated 570 simulated data sets from a bivariate longitudinal model with censored responses, considering the Slash distribution as the distribution of the error term and random effect. Each data set includes the information of $n = 50$ subjects. The time points t_{ij}'s at which the longitudinal measures were taken, were fixed at $(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10)$. On one hand, the design matrix of fixed effects for each outcome X_{ij} contains the intercept and time point t_{ij}. On the other hand, the design matrix of random effects Z_{ij} contains only the intercept component. Therefore, $\mathbf{X}_i = \mathbf{I}_2 \otimes X_{ij}$ and $\mathbf{Z}_i = \mathbf{I}_2 \otimes Z_{ij}$, with $\mathbf{I}_2$ being the identity matrix of dimension 2×2. For the simulation, we considered the following values of the model parameters:

$$\boldsymbol{\beta} = \begin{pmatrix} 6 \\ -0.2 \\ 4 \\ -0.1 \end{pmatrix}, \ \boldsymbol{\Sigma} = \begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix}, \ \mathbf{D} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \ \boldsymbol{\Omega}_i = \boldsymbol{\Omega}_i((0.8, 1), \mathbf{t}_i), \ \nu = 3 \text{ and } \eta = 3.$$

Note that, the specification of $\boldsymbol{\Omega}_i$ generates an AR(1) structure. Additionally, we considered two different schemes of left censoring proportions, say, 10% and 20% in each data set. In the case of the SAEM algorithm, the maximum number of iterations was $W = 300$ and a cut point $c = 0.25$ was considered.

The model selection criteria (AIC and BIC) as well as the estimates of the model parameters were registered for each simulation. Summary statistics such as the

Monte Carlo mean estimate (MC mean), the mean of the approximate standard error obtained through the information-based method described in Section 5.3.3 (IM SE) and the Monte Carlo empirical standard error (MC Sd) are presented in Table 20. It can be observed from Table 20 that, in all models, the fixed effects the estimates are closer to the real values. Moreover, in general, the estimates of the variance components are over estimated under the model M3. In fact, these estimates are always greater than those obtained under models M1 and M2. A possible explanation for this fact is that model M3 considers the parameters $\nu = 8$ and $\eta = 8$ fixed at different values than the true ones. Finally, after observe the Monte Carlo standard deviation (MC Sd) for the parameters of interest, it can be concluded that the proposed asymptotic approximation for the variances of the parameters obtained through Equation (5.9) generates similar results.

Table 20 – **Simulation study**. Parameter estimates based on 570 simulated samples. MC mean, MC SD are the respective mean estimates and standard deviations. 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.

		M1			M2			M3		
Parameters		MC Mean	IM SE	MC SD	MC Mean	IM SE	MC SD	MC Mean	IM SE	MC SD
		Censoring 10%								
β_{10}	(6.0)	5.922	0.497	0.419	5.927	0.483	0.420	5.937	0.493	0.426
β_{11}	(-0.2)	-0.193	0.055	0.054	-0.193	0.054	0.054	-0.193	0.056	0.054
β_{20}	(4.0)	4.012	0.495	0.454	4.015	0.481	0.456	4.021	0.490	0.460
β_{21}	(-0.1)	-0.101	0.055	0.056	-0.101	0.054	0.056	-0.102	0.055	0.056
σ_{11}	(4.0)	3.999	0.987	0.815	3.909	0.875	0.752	4.814	1.052	0.950
σ_{12}	(2.0)	1.987	0.536	0.430	1.942	0.480	0.393	2.394	0.582	0.495
σ_{22}	(4.0)	3.986	0.985	0.789	3.899	0.876	0.730	4.803	1.057	0.928
α_{11}	(2.0)	2.099	3.727	1.039	2.002	1.185	0.928	2.590	1.517	1.249
α_{12}	(1.0)	1.097	2.100	0.789	1.038	0.833	0.705	1.334	1.060	0.952
α_{22}	(2.0)	2.109	3.723	0.974	1.993	1.178	0.829	2.563	1.498	1.132
ϕ_1	(0.8)	0.789	0.044	0.039	0.790	0.043	0.039	0.790	0.043	0.039
ν	(3.0)	3.684								
η	(3.0)	4.101								
MC AIC			3505.619			3502.972			3511.686	
MC BIC			3569.420			3556.957			3565.671	
		Censoring 20%								
β_{10}	(6.0)	5.925	0.484	0.442	5.929	0.470	0.442	5.944	0.479	0.447
β_{11}	(-0.2)	-0.191	0.055	0.047	-0.191	0.054	0.047	-0.191	0.055	0.048
β_{20}	(4.0)	3.988	0.486	0.432	3.992	0.472	0.431	4.005	0.482	0.438
β_{21}	(-0.1)	-0.097	0.054	0.050	-0.097	0.053	0.050	-0.098	0.055	0.050
σ_{11}	(4.0)	3.952	0.984	0.808	3.823	0.860	0.695	4.697	1.022	0.867
σ_{12}	(2.0)	1.957	0.530	0.428	1.893	0.469	0.373	2.327	0.561	0.466
σ_{22}	(4.0)	3.952	0.981	0.779	3.823	0.856	0.663	4.699	1.019	0.837
α_{11}	(2.0)	2.036	3.575	0.962	1.860	1.140	0.828	2.378	1.439	1.095
α_{12}	(1.0)	1.024	1.931	0.698	0.942	0.799	0.631	1.193	1.011	0.839
α_{22}	(2.0)	2.129	3.703	0.975	1.950	1.150	0.847	2.490	1.451	1.128
ϕ_1	(0.8)	0.787	0.045	0.036	0.787	0.043	0.036	0.787	0.043	0.036
ν	(3.0)	4.003								
η	(3.0)	4.433								
MC AIC			3381.656			3379.021			3386.550	
MC BIC			3445.457			3433.007			3440.535	

We also analyzed the absolute bias (Bias) and mean square error (MSE) of the fixed effects and variance components estimates obtained from the models under study. In

this case, our aim is to compare the Bias and MSE of the ML estimates from all models, showing that they are the same when the parameters ν and η (those parameters associated with the mixture variables) are either estimated or considered as fixed. The Bias and MSE measures are defined as

$$\text{Bias} = \frac{1}{J}\sum_{j=1}^{J}|\widehat{\boldsymbol{\theta}}_i^{(j)} - \boldsymbol{\theta}_i| \quad \text{and} \quad \text{MSE} = \frac{1}{J}\sum_{j=1}^{J}\left(\widehat{\boldsymbol{\theta}}_i^{(j)} - \boldsymbol{\theta}_i\right)^2, \tag{5.11}$$

where $\widehat{\boldsymbol{\theta}}_i^{(j)}$ is the ML estimate of the parameter θ_i for the j-th sample, $j = 1, \ldots, J$.

Table 21 show the bias and MSE of the parameter estimates. It can be seen that these measures are similar in the case of the fixed effects for all models under study. However, the Bias and MSE seem to be higher in the case of the variance components, specifically in model M3. This fact could reveal the effect of the misspecification related to parameters ν and η. It important to note that similar results are obtained for the two levels of censoring proportions considered in the study.

Table 21 – **Simulation study**. Bias and MSE of the parameter estimates.

	Censoring 10%						Censoring 20%					
	Bias			MSE			Bias			MSE		
Parameters	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
β_1	0.338	0.337	0.339	0.181	0.180	0.184	0.364	0.363	0.366	0.201	0.200	0.203
β_2	0.044	0.044	0.044	0.003	0.003	0.003	0.039	0.039	0.039	0.002	0.002	0.002
β_3	0.368	0.372	0.372	0.205	0.207	0.210	0.350	0.350	0.353	0.186	0.186	0.191
β_4	0.044	0.044	0.044	0.003	0.003	0.003	0.040	0.040	0.040	0.002	0.002	0.003
σ_{11}	0.643	0.612	0.997	0.659	0.570	1.559	0.648	0.588	0.888	0.655	0.514	1.237
σ_{12}	0.332	0.311	0.506	0.184	0.156	0.399	0.343	0.317	0.450	0.185	0.150	0.324
σ_{22}	0.630	0.596	0.989	0.618	0.539	1.500	0.619	0.559	0.869	0.608	0.470	1.188
α_{11}	0.850	0.767	1.096	1.082	0.854	1.897	0.775	0.686	0.887	0.925	0.704	1.340
α_{12}	0.627	0.563	0.789	0.627	0.494	1.011	0.545	0.500	0.656	0.487	0.400	0.740
α_{22}	0.775	0.682	0.991	0.953	0.681	1.589	0.774	0.676	0.948	0.965	0.719	1.511
ϕ_1	0.032	0.032	0.032	0.002	0.002	0.002	0.030	0.030	0.030	0.001	0.001	0.001

Finally, Figure 23 shows the box plots corresponding to the parameter estimates over the 570 simulated datasets. Note that, the estimates for the three models under study are, in general, equivalents. However, the estimates of variance components obtained from models M1 and M2 seem to be more close to the true values than those obtained under model M3. Similar results are obtained after the analysis of the data with 10% of censoring (see Figure 30 in Appendix C.5).

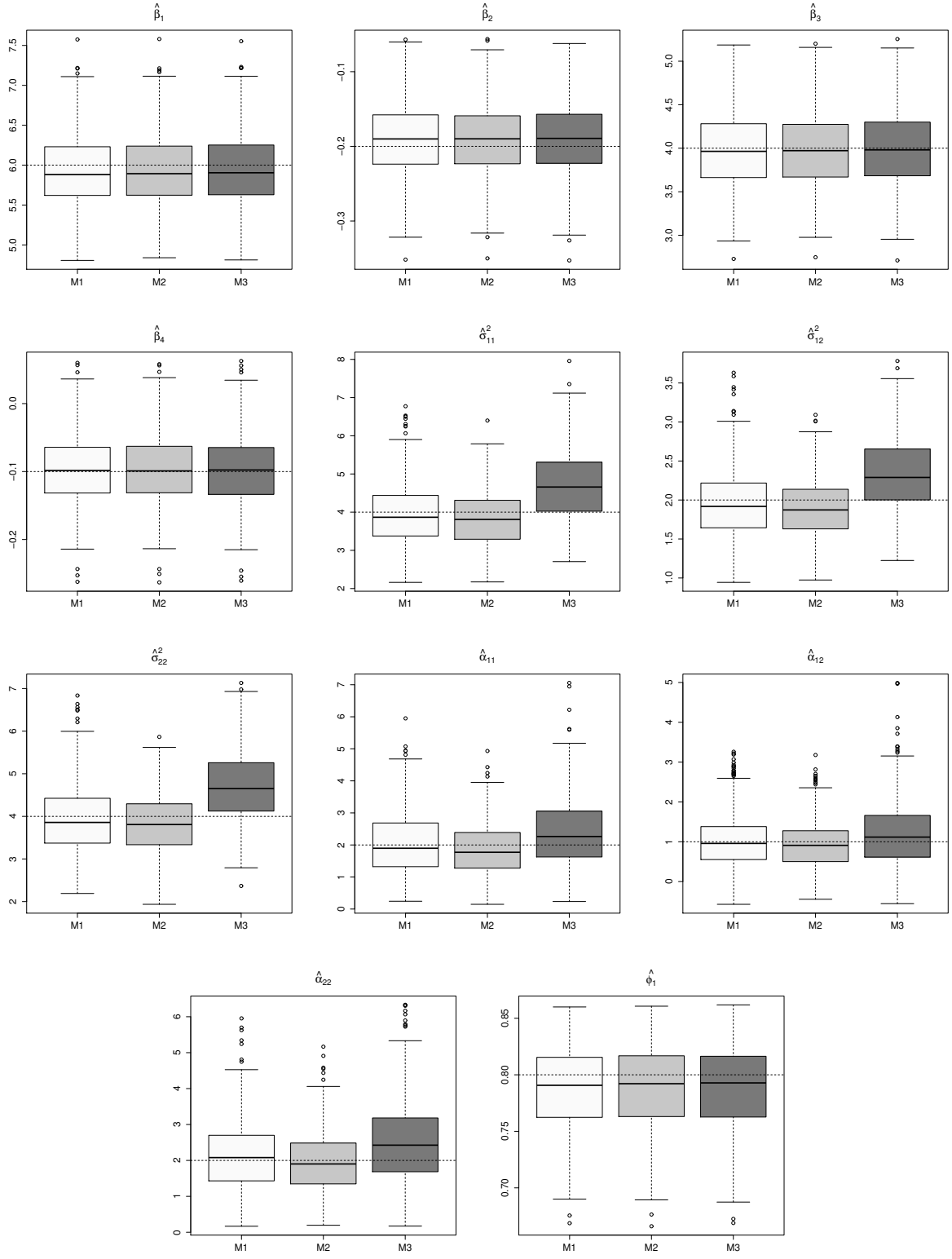

Figure 23 – **Simulation study**: (20% censored). Boxplots of the parameter estimates. Dotted lines indicate the true parameter value.

5.5 Analysis of A5055 clinical trial

This section illustrates the performance of the proposed methods with the analysis of a HIV dataset, previously analyzed by (Wang, 2013). Specifically, we study the HIV viral load data from clinical trial study A5055, considering two different correlation structures, namely the uncorrelated structure (UNC) and damped exponential correlation (DEC); and different distributions belonging to the SMN class. As was mentioned in the Introduction, the dataset consists of 44 HIV-1 infected patients treated with two potent antiretroviral therapies (treatment 1 and treatment 2). The viral load was quantified on days 0, 7, 14, 28, 56, 84, 112, 140, and 168 of follow-up for each patient, the immunologic marker CD4 and CD8 were also measured. The dataset includes 316 observations, where the viral load detectable limit is 50 copies/mL, and 106 out of 316 (33 %) of all viral load measurements are below the detection limit.

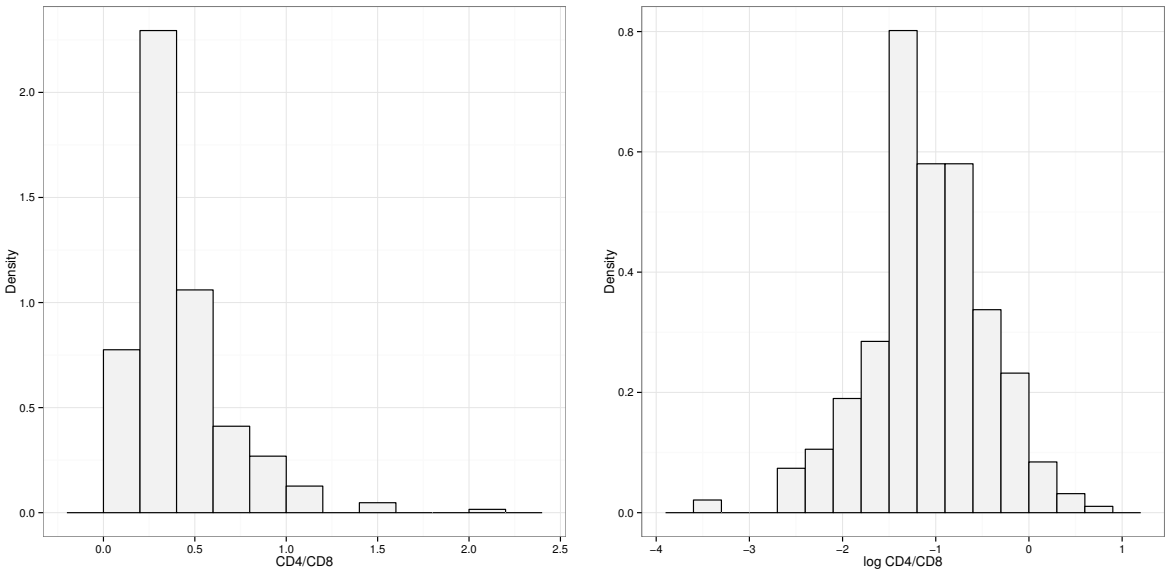

Figure 24 – **A5055 data.** Histogram of the CD4/CD8 ratio (left panel) and histogram of the log CD4/CD8 (right panel).

Before fitting the model, we applied a log transformation to the ratio CD4/CD8 in order to generate a positively skewed distribution. Figure 24 shows the histogram for the variable CD4/CD8 and log transformation of CD4/CD8.

Let y_{i1k} and y_{i2k} be $\log_{10}$(RNA) and log(CD4/CD8), respectively, for subject i ($i = 1, \ldots, 44$) measured roughly at day$_{ik}$. The model considered for modeling the A5055 dataset is given bt

$$y_{i1k} = \beta_{10} + \beta_{11} t_{ik} + \beta_{12}\text{treat}_i + \beta_{13} t_{ik}^{0.5} + \beta_{14}\text{treat}_i \times t_{ik} + b_{i10} + b_{i11} t_{ik} + e_{i1k}, \quad (5.12)$$

$$y_{i2k} = \beta_{20} + \beta_{21} t_{ik} + \beta_{22}\text{treat}_i + \beta_{23}\text{treat}_i \times t_{ik} + b_{i20} + b_{i21} t_{ik} + e_{i2k}, \quad (5.13)$$

where $t_{ik} = \text{day}_{ik}/7$ (week), for $k = 1, \ldots, s_i$, treat$_i$ is a treatment indicator (= 0 for treatment 1; = 1 for treatment 2), $t_{ik}^{0.5}$ is the square root at time t_{ik}, and b_{ij0} and b_{ij1} are

the random intercept and random slope, respectively, for y_{ijk}, $j = 1, 2$.

Table 22 – **A5055 data.** Information criteria for the *SMN-MLMEC* models under DEC structure.

	Distribution $\boldsymbol{\epsilon}$ / Distribution **b**								
	N/N	SL/N	T/N	N/SL	N/T	SL/SL	SL/T	T/SL	T/T
AIC	789.85	742.18	**739.59**	791.98	792.29	744.47	744.54	741.85	741.51
BIC	896.62	853.41	**850.81**	903.20	903.51	860.14	860.21	857.52	857.19

To specify the distribution of the error term and the random effects, we have considered many combinations of possibles distributions. Table 22 shows the AIC and BIC values obtained for these combinations. From this table, we conclude that the best model is the model considering the Student-t (T) distribution for the error and Normal (N) distribution for the random effects.

Table 23 – **A5055 data.** ML estimates with standard errors for the *SMN-LMMC* model under the Student-t/Normal distribution.

	Fixed effects			Between-subject variances			Within-subject variances		
Structure	Parameters	Estimate	SE	Parameters	Estimate	SE	Parameters	Estimate	SE
DEC	β_{10}	3.743	0.134	d_{11}	0.1446	0.0829	σ_{11}	0.409	0.076
	β_{11}	0.130	0.026	d_{21}	0.0011	0.0133	σ_{21}	-0.039	0.020
	β_{12}	-0.005	0.067	d_{22}	-0.0884	0.1182	σ_{22}	0.050	0.011
	β_{13}	-0.957	0.098	d_{31}	-0.0011	0.0033	ϕ_1	0.704	0.065
	β_{14}	-0.007	0.025	d_{32}	0.0034	0.0027	ϕ_2	0.632	0.131
	β_{20}	-1.284	0.077	d_{33}	-0.0122	0.0116	ν	4.737	0.003
	β_{21}	0.005	0.005	d_{41}	-0.0004	0.0004			
	β_{22}	0.252	0.084	d_{42}	0.2727	0.0861			
	β_{23}	-0.003	0.007	d_{43}	0.0008	0.0015			
				d_{44}	0.0001	0.0001			
	loglik	-344.79		AIC	739.59		BIC	850.81	
UNC	β_{10}	3.718	0.135	d_{11}	0.4089	0.1463	σ_{11}	0.263	0.053
	β_{11}	0.129	0.026	d_{21}	-0.0112	0.0153	σ_{21}	-0.024	0.012
	β_{12}	0.003	0.091	d_{22}	-0.0964	0.1251	σ_{22}	0.028	0.005
	β_{13}	-0.955	0.075	d_{31}	0.0002	0.0030	ν	4.340	0.004
	β_{14}	-0.008	0.027	d_{32}	0.0054	0.0029			
	β_{20}	-1.278	0.076	d_{33}	-0.0132	0.0116			
	β_{21}	0.005	0.004	d_{41}	-0.0006	0.0004			
	β_{22}	0.286	0.081	d_{42}	0.2953	0.0785			
	β_{23}	-0.006	0.006	d_{43}	0.0002	0.0015			
				d_{44}	0.0001	0.0001			
	loglik	-357.97		AIC	761.94		BIC	864.26	

Table 23 presents the ML estimates of the parameters of interest under the best model, considering two different correlation structure namely, the uncorrelated (UNC) and DEC structure. This table also presents the corresponding standard errors of the parameters estimates. It can be observed that, according to the AIC or BIC values, the model with DEC structure provides a much improved model fit overcoming the one under the UNC structure. In fact, the maximum log-likelihood was -357.97 for the UNC and -344.79 for the DEC structure, which gives a likelihood ratio statistic of 26.36 (p-value < 0.0001). Consequently, the model with DEC structure fits the data substantially better

than the model with UNC structure. Moreover, for those models, the estimated values of ν are fairly small, indicating a lack of adequacy of the normal assumption for the error term.

Note that, for the best model (T/N-MLMEC with DEC structure), the estimates of β_{11} and β_{21} (0.130, 0.005) reveal that RNA viral loads and the log(CD4/CD8) change over the time. From the estimate of β_{22} (0.252) we conclude that patients receiving different treatment have difference on baseline log(CD4/CD8), while from the estimate -0.005 of β_{12} we do not have significant difference on baseline RNA viral loads. From the estimates of treatment and time iteration (β_{14}, β_{23}), the patients receiving treatment 2 exhibit little decaying quantities on $\log_{10}$ RNA than those receiving treatment 1 by an expected 0.007 unit per week, and less growing quantities on log(CD4/CD8) than those receiving treatment 2 by an expected 0.003 unit per week. For our best model, the convergence of the estimates obtained through the SAEM algorithm are shown in Appendix C.6 (Figures 31, 32 and 33). As can be seen, the convergence can be achieved very quickly.

Table 24 – **A5055 data.** Decomposition of AIC and BIC for the best *SMN-MLMEC* model.

AIC	739.59	BIC	850.81
$\text{AIC}_{\mathbf{y}_2^\star\|\mathbf{y}_1^\star}$	92.65	$\text{BIC}_{\mathbf{y}_2^\star\|\mathbf{y}_1^\star}$	158.80
$\text{AIC}_{\mathbf{y}_{2,0}^\star}$	125.26	$\text{BIC}_{\mathbf{y}_{2,0}^\star}$	166.58
ΔAIC	32.61	ΔBIC	7.77

Table 24 shows the AIC and BIC decomposition described in Subsection 5.3.2. We can see from these measures (ΔAIC and ΔBIC) that the contribution of considering a joint model (log(CD4/CD8) with RNA viral load) is significative.

5.6 Conclusions

In this work, we have introduced a robust multivariate linear mixed model for multiple censored responses based on the class of SMN distributions, extending the recent work proposed by Wang et al. (2015). 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-MLMEC* 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. An additional characteristic of our proposed model is that it allows to consider different distributions for the error terms, thereby overcoming the foregoing limitation of the MLMEC model and broadening the scope of censored mixed models. Based in Zhang et al. (2014), we also propose a decomposition of the AIC and BIC

criteria, which was shown to be very effective to assess the effect of the joint modelling of censored outcomes. The computing code is available from the authors upon request.

Although the SMN-MLMEC considered here has shown great flexibility for modeling symmetric data, its robustness against outliers can be seriously affected by the presence of skewness. Recently, Bandyopadhyay et al. (2012) proposed a remedy to accommodate skewness and heavy- tails simultaneously in the context of LMEC models by using scale mixtures of skew-normal (SMSN) distributions. We conjecture that our methods can be used under MLMEC models and should yield satisfactory results at the expense of additional complexity in the implementation. An in-depth investigation of such extensions is beyond the scope of the present work, but it is an interesting topic for further research.

6 Final Considerations

In this thesis, we discuss a frequentist approach to model censored responses under the class SMN distributions, which is a generalization of the works by Matos et al. (2013a), Matos et al. (2013b), Garay et al. (2014), Wang et al. (2015), among others. The class of SMN distributions, proposed by Andrews and Mallows (1974), is attractive since it has a stochastic representation for easy implementation of ML algorithms and it also facilitates the study of many useful properties. This extension result in a flexible class of models for robust estimation since it contains distributions such as the normal, the Student- t, slash and contaminated normal distribution.

The EM algorithm (Dempster et al., 1977) and a stochastic approximation of the EM (SAEM) algorithm (Delyon et al., 1999) was developed to obtain the maximum likelihood estimates for the parameters of the models. Furthermore, the standard errors of the fixed effects and predictions of unobservable values of the response are obtained as a by-product. We also applied our methods to four AIDS real data and undertake several simulation studies to demonstrate the proposed methods.

The proposed methods and techniques can be extended to another general framework, such as measurement error models, semiparametric regression models, spatial models, spatial and temporal models, among many others. An in-depth investigation of such extensions is certainly an interesting topic for future research. Although all models considered in this thesis has shown great flexibility for modelling symmetric data, it can be seriously affected by the presence of skewness. Recently, Bandyopadhyay et al. (2012) proposed a remedy to accommodate skewness and heavy-tailedness simultaneously in the context of LMEC models by using scale mixtures of skew-normal distributions. The models presented in this thesis can be extended to incorporate this class of distributions and become interesting topics of future researches.

Bibliography

Abarin, T., H. Li, L. Wang, and L. Briollais (2014). On method of moments estimation in linear mixed effects models with measurement error on covariates and response with application to a longitudinal study gene-environment interaction. *Statistics in Biosciences 6*, 1–18.

Alder, M. (2001). Development of the epidemic. In M. Alder (Ed.), *ABC of AIDS*. BMJ Publishing Group.

Andrews, D. F. and C. L. Mallows (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 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 recombinant forms. *Journal of Clinical Microbiology 41*(7), 3361–3367.

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

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.

Ash, R. B. (2000). *Probability and Measure Theory*. San Diego: Academic Press.

Azevedo, K. M. L., S. Setúbal, V. G. S. Lopes, L. A. B. Camacho, and S. A. Oliveira (2010). Congenital toxoplasmosis transmitted by human immunodeficiency-virus infected women. *Brazilian Journal of Infectious Diseases 14*(2), 186–189.

Bandyopadhyay, D., L. M. Castro, V. H. Lachos, and H. Pinheiro (2015). Robust joint nonlinear mixed-effects models and diagnostics for censored HIV viral loads with CD4 measurement error. *Journal of Agricultural, Biological and Environmental Statistics 20*, 121–139.

Bandyopadhyay, D., V. H. Lachos, L. M. Castro, and D. K. Dey (2012). Skew-normal/independent linear mixed models for censored responses with applications to HIV viral loads. *Biometrical Journal 54*(3), 405–425.

Banerjee, M. and E. W. Frees (1997). Influence diagnostics for linear longitudinal models. *Journal of the American Statistical Association 92*(439), 999–1005.

Barletta, J., D. Edelman, and N. Constantine (2004). Lowering the detection limits of HIV-1 viral load using real-time inmuno PCR for HIV-1 p24 antigen. *American Journal of Clinical Pathology 112*(20-27).

Barnett, V. (1969). Simultaneous pairwise linear structural relationships. *Biometrics*, 129–142.

Beckman, R. J., C. J. Nachtsheim, and R. D. Cook (1987). Diagnostics for mixed–model analysis of variance. *Technometrics 29*(4), 413–426.

Buonaccorsi, J. (2010). *Measurement Error: Models, Methods and Applications*. Chapman & Hall/CRC, Boca Raton.

Buonaccorsi, J., E. Demidenko, and T. Tosteson (2000). Estimation in longitudinal random effects models with measurement error. *Statistica Sinica 10*, 885–903.

Cabral, C. R. B., V. H. Lachos, and C. B. Zeller (2014). Multivariate measurement error models using finite mixtures of skew-student t distributions. *Journal of Multivariate Analysis 124*, 179–198.

Carroll, R. J., X. Lin, and N. Wang (1997). Generalized linear mixed measurement error models. In *Modellling Longitudinal and Spatially Correlated Data*, Volume 122 of *Lecture Notes in Statistics*, pp. 321–330. Springer–Verlag.

Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu (2006). *Mesurement Error in Nonlinear Models* (2 ed.). Chapman & Hall/CRC, Boca Raton.

Castro, L. M., D. R. Costa, M. O. Prates, and V. H. Lachos (2015). Likelihood-based inference for Tobit confirmatory factor analysis using the multivariate student-t distribution. *Statistics and Computing 25*(6), 1163–1183.

Cheng, C. L. and J. W. Van Ness (1999). *Statistical Regression with Measurement Error*. Arnold, London.

Chipkevitch, E., R. T. Nishimura, D. G. Tu, and M. Galea-Rojas (1996). Clinical measurement of testicular volume in adolescents: comparison of the reliability of 5 methods. *The Journal of Urology 156*(6), 2050–2053.

Christensen, R., L. Pearson, and W. Johson (1992). Case-deletion diagnostics for mixed-models. *Technometrics 34*, 38–45.

Ciesielski, C. A. and R. P. Metler (1997). Duration of time between exposure and seroconversion in healthcare workers with occupationally acquired infection with human immunodeficiency virus. *The American Journal of Medicine 102*(5), 115–116.

Cook, R. (1977). Detection of influential observation in linear regression. *Technometrics*, 15–18.

Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society, Series B, 48*, 133–169.

Cook, R. D. and S. Weisberg (1982). *Residuals and Influence in Regression*. Boca Raton, FL: Chapman & Hall/CRC.

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.

Dumitrescu, L. (2010). Estimation for a longitudinal linear model with measurement errors. *Electronic Journal of Statistics 4*, 486–524.

Fuller, W. A. (1987). *Measurement Error Models*. New York: John Wiley and Sons.

Galarza, C. E., D. Bandyopadhyay, and V. H. Lachos (2015). Quantile regression in linear mixed models: A stochastic approximation EM approach. *Statistics and its Interface*, In press.

Galea-Rojas, M., H. Bolfarine, and M. de Castro (2002). Local influence in comparative calibration models. *Biometrical Journal 44*(1), 59–81.

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.

Ho, H. J., T.-I. Lin, H. Y. Chen, and W.-L. Wang (2012). Some results on the truncated multivariate t distribution. *Journal of Statistical Planning and Inference 142*, 25–40.

Hodges, J. S. (1998). Some algebra and geometry for hierarchical models, applied to diagnostics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60*(3), 497–536.

Huang, Y. and G. Dagne (2011). A bayesian approach to joint mixed-effects models with a skew-normal distribution and measurement errors in covariates. *Biometrics 67*(1), 260–269.

Hughes, J. (1999). Mixed effects models with censored data with application to HIV RNA levels. *Biometrics 55*(2), 625–629.

Jank, W. (2006). Implementing and diagnosing the stochastic approximation EM algorithm. *Journal of Computational and Graphical Statistics 15*(4), 803–829.

Kotz, S. and S. Nadarajah (2004). *Multivariate t Distributions and Their Applications*. Cambridge: Cambridge University Press.

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. 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.

Lachos, V. H., F. V. 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.

Lange, K. L., R. J. Little, and J. M. Taylor (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association 84*(408), 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.

Lee, S. Y. and L. Xu (2004). Influence analysis of nonlinear mixed-effects models. *Computational Statistics and Data Analysis 45*, 321–341.

Lesaffre, E. and G. Verbeke (1998). Local influence in linear mixed models. *Biometrics 54*, 570–582.

Levine, R. and G. Casella (2001). Implementations of the Monte Carlo EM algorithm. *Journal of Computational and Graphical Statistics 10*(3), 422–439.

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.

Lin, T.-I. and J. C. 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. and W.-L. Wang (2013). Multivariate skew-normal at linear mixed models for multi-outcome longitudinal data. *Statistical Modelling 13*(3), 199–221.

Liu, W. and L. Wu (2012). Two-step and likelihood methods for HIV viral dynamic models with covariate measurement errors and missing data. *Journal of Applied Statistics 39*(5), 963–978.

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. A., D. Bandyopadhyay, L. M. Castro, and V. H. Lachos (2015). Influence assessment in censored mixed-effects models using the multivariate Student's-t distribution. *Journal of Multivariate Analysis 141*, 104–117.

Matos, L. A., L. M. Castro, and V. H. Lachos (2016). Censored mixed-effects models for irregularly observed repeated measures with applications to HIV viral loads. *Test*, DOI: 10.1007/s11749–016–0486–2.

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

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

Meilijson, I. (1989). A fast improvement to the EM algorithm on its own terms. *Journal of the Royal Statistical Society. Series B (Methodological)*, 127–138.

Meng, X. and D. B. Rubin (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika 81*, 633–648.

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.

Mortimer, P. and C. Loveday (2001). The virus and the tests. In *ABC of AIDS*. BMJ Publishing Group.

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.

Müller, P. and S. van de Geer (2014). Censored linear model in high dimensions. *Test*, 1–18.

Ndembi, N., R. Goodall, D. Dunn, A. McCormick, A. Burke, F. Lyagoba, P. Munderi, P. Katundu, C. Kityo, V. Robertson, D. Yirrell, A. Walker, D. Gibb, C. Gilks, P. Kaleebu, and D. Pillay (2010). Viral rebound and emergence of drug resistance in the absence of viral load testing: A randomized comparison between Zidovudine-Lamivudine plus Nevirapine and Zidovudine-Lamivudine plus Abacavir. *Journal of Infectious Diseases 201*(1), 106–113.

Osorio, F., G. A. Paula, and M. Galea (2007). Assessment of local influence in elliptical linear models with longitudinal structure. *Computational Statistics and Data Analysis 51*, 4354–4368.

Pan, J., Y. Fei, and P. Foster (2014). Case-deletion diagnostics for linear mixed models. *Technometrics 56*(3), 269–281.

Paxton, W. B., R. W. Coombs, M. J. McElrath, M. C. Keefer, J. Hughes, F. Sinangil, D. Chernoff, L. Demeter, B. Williams, and L. Corey (1997). Longitudinal analysis of quantitative virologic measures in human immunodeficiency virus-infected subjects with $\geqslant$ 400 CD4 lymphocytes: implications for applying measurements to individual patients. *Journal of Infectious Diseases 175*(2), 247–254.

Pinheiro, J. C. and M. Bates, Douglas (2000). *Mixed-Effects Models in S and S-PLUS*. New York, NY: Springer.

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.

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

R Development Core Team (2009). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rao, C. R. (1987). Prediction of future observations in growth curve models. *Statistical Science 2*, 434–447.

Rocha, G. H., R. B. Arellano-Valle, and R. H. Loschi (2015). Maximum likelihood methods in a robust censored errors-in-variables model. *Test 24*(4), 857–877.

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.

Russo, C., G. Paula, and R. Aoki (2009). Influence diagnostics in nonlinear mixed-effects elliptical models. *Computational Statistics & Data Analysis 53*(12), 4143–4156.

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.

Shah, A., N. Laird, and D. Schoenfeld (1997). A random-effects model for multiple characteristics with possibly missing data. *Journal of the American Statistical Association 92*(438), 775–779.

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.

Tan, F. E., M. J. Ouwens, and M. P. Berger (2001). Detection of influential observations in longitudinal mixed effects regression models. *Journal of the Royal Statistical Society: Series D (The Statistician) 50*(3), 271–284.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica 26*, 24–36.

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, J. and M. G. Genton (2006). The multivariate skew-slash distribution. *Journal of Statistical Planning and Inference 136*, 209–220.

Wang, W.-L. (2013). Multivariate t linear mixed models for irregularly observed multiple repeated measures with missing outcomes. *Biometrical Journal 55*(4), 554–571.

Wang, W.-L. and T.-H. Fan (2011). Estimation in multivariate t linear mixed models for multivariate longitudinal data. *Statistica Sinica 21*, 1857–1880.

Wang, W.-L. and T.-I. Lin (2014). Multivariate t nonlinear mixed-effects models for multi-outcome longitudinal data with missing values. *Statistics in Medicine 33*(17), 3029–3046.

Wang, W.-L., T.-I. Lin, and V. H. Lachos (2015). Extending multivariate-t linear mixed models for multiple longitudinal data with censored responses and heavy tails. *Statistical Methods in Medical Research*, DOI: 0962280215620229.

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, H. (2005). Statistical methods for HIV dynamic studies in AIDS clinical trials. *Statistical Methods in Medical Research 14*(2), 171.

Wu, H. and A. Ding (1999). Population HIV-1 dynamics in vivo: applicable models and inferential tools for virological data from AIDS clinical trials. *Biometrics 55*(2), 410–418.

Wu, L. (2002). A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to AIDS studies. *Journal of the American Statistical Association 97*(460), 955–964.

Wu, L. (2010). *Mixed Effects Models for Complex Data*. Boca Raton, FL: Chapman & Hall/CRC.

Wu, L., W. Liu, and X. Hu (2010). Joint inference on HIV viral dynamics and immune suppression in presence of measurement errors. *Biometrics*, 327–335.

Xie, F., B. Wei, and J. Lin (2007). Case-deletion influence measures for the data from multivariate t distributions. *Journal of Applied Statistics 34*, 907–921.

Xu, L., S. Lee, and W. Poon (2006). Deletion measures for generalized linear mixed effects models. *Computational Statistics & Data Analysis 51*(1131-1146).

Zeller, C. B., F. V. Labra, V. H. Lachos, and N. Balakrishnan (2010). Influence analyses of skew-normal/independent linear mixed models. *Computational Statistics & Data Analysis 54*(5), 1266–1280.

Zhang, D., M.-H. Chen, J. G. Ibrahim, M. E. Boye, P. Wang, and W. Shen (2014). Assessing model fit in joint models of longitudinal and survival data with applications to cancer clinical trials. *Statistics in Medicine 33*(27), 4715–4733.

Zhang, S., D. Midthune, P. M. Guenther, S. M. Krebs-Smith, V. Kipnis, K. W. Dodd, D. W. Buckman, J. A. Tooze, L. Freedman, and R. J. Carroll (2011). A new multivariate measurement error model with zero-inflated dietary data, and its application to dietary assessment. *The Annals of Applied Statistics 5*, 1456–1487.

Zhu, H. and S. Lee (2001). Local influence for incomplete-data models. *Journal of the Royal Statistical Society, Series B (Statistical Methodology) 63*, 111–126.

Zhu, H., S. Lee, B. Wei, and J. Zhou (2001). Case-deletion measures for models with incomplete data. *Biometrika 88*, 727–737.

Part III

Supplementary Material

APPENDIX A – Supplementary Material for Chapter 3

A.1 $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ is a block-diagonal matrix

From the EM-algorithm, we have $\partial Q(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = 0$. Consequently, for a t-LMEC model:

$$\begin{aligned}
\sum_{i=1}^{n}\mathbf{X}_i^\top(\widehat{u\mathbf{y}}_i - \mathbf{Z}_i\widehat{u\mathbf{b}}_i) &= \sum_{i=1}^{n}\widehat{u}_i\mathbf{X}_i^\top\mathbf{X}_i\widehat{\boldsymbol{\beta}},\\
\sum_{i=1}^{n}(n_i\widehat{\sigma^2} - \mathrm{tr}(\widehat{u\mathbf{y}_i^2} - 2\widehat{u\mathbf{y}\mathbf{b}_i}\mathbf{Z}_i^\top + \widehat{u\mathbf{b}_i^2}\mathbf{Z}_i^\top\mathbf{Z}_i)) &= \sum_{i=1}^{n}(2\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_i^\top(\widehat{u\mathbf{y}}_i - \mathbf{Z}_i\widehat{u\mathbf{b}}_i) - \widehat{u}_i\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_i^\top\mathbf{X}_i\widehat{\boldsymbol{\beta}}),\\
\partial Q(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\alpha} &= 0,
\end{aligned}$$

Finally, from above,

$$\frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\sigma^2} = -\frac{1}{\sigma^4}\mathbf{X}_i^\top(\widehat{u\mathbf{y}}_i - \mathbf{Z}_j\widehat{u\mathbf{b}}_i - \widehat{u}_i\mathbf{X}_i\boldsymbol{\beta}) = 0$$

and hence, the matrix $\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ is block-diagonal.

A.2 Equivalence of GD_i^1 and the local influence based on the case weights scheme

For the i-th subject, the normal curvature is given by $C_i = 2\boldsymbol{\Delta}_i^\top\left\{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\right\}^{-1}\boldsymbol{\Delta}_i$, $i = 1,\ldots,n$, where for the case weights pertubation $\boldsymbol{\Delta}_i = \dfrac{\partial^2 Q_i(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}}|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \dfrac{\partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}$. Since $\dot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = 0$, we can show that

$$\dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = -\dot{Q}_i(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = -\frac{\partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}.$$

Then, $\boldsymbol{\Delta}_i = -\dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ and, as a result, $C_i = 2\dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\left\{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\right\}^{-1}\dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$.
Hence, from $GD_i^1(\boldsymbol{\theta}) = \dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}})^\top\{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}^{-1}\dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}})$, $i = 1,\ldots,n$, we have that $C_i = 2GD_i^1$, and consequently GD_i^1 is equivalent to the local influence based on the case weights scheme.

APPENDIX B – Supplementary Material for Chapter 4

B.1 Proof of Proposition 4.1

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

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

(a) For the *multivariate normal* distribution:
The proof is straightforward since U is degenerated in 1.

(b) For the *multivariate Student's-t* distribution:

$$\begin{aligned} L_i(\boldsymbol{\theta}) &= \int_0^\infty \frac{1}{\sqrt{(2\pi)^{p_1}|\frac{1}{u}\boldsymbol{\Sigma}_1|}} \exp\Big\{ -\frac{u}{2}(\mathbf{y}_1 - \boldsymbol{\mu}_1)^\top \Sigma_1^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)\Big\} \\ &\quad \Phi_{p_2}\left(\mathbf{y}_2; \boldsymbol{\mu}_2, \frac{\boldsymbol{\Sigma}_2}{u}\right) \frac{(\frac{\nu}{2})^{\frac{\nu}{2}} u^{\frac{\nu}{2}-1}}{\Gamma(\frac{\nu}{2})} \exp\Big\{ -\frac{\nu}{2}u\Big\} du. \end{aligned}$$

Let $d(\mathbf{y}_1) = (\mathbf{y}_1 - \boldsymbol{\mu}_1)^\top \Sigma_1^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$. After some algebraic manipulations, we can deduce that

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

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

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

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

$$\begin{aligned} L_i(\boldsymbol{\theta}) &= \int_0^\infty \phi_{p_1}(\mathbf{y}_1; \boldsymbol{\mu}_1, \kappa(u)\boldsymbol{\Sigma}_1)\Phi_{p_2}(\mathbf{y}_2; \boldsymbol{\mu}_2, \kappa(u)\boldsymbol{\Sigma}_2) dH(u) \\ &= \nu_1\left[\frac{1}{\sqrt{(2\pi)^{p_1}|\frac{1}{\nu_2}\boldsymbol{\Sigma}_1|}}\, \mathrm{e}^{\{-\frac{1}{2}(\mathbf{y}_1-\boldsymbol{\mu}_1)^\top \nu_2 \boldsymbol{\Sigma}_1^{-1}(\mathbf{y}_1-\boldsymbol{\mu}_1)\}}\Phi_{p_2}\left(\mathbf{y}_2; \boldsymbol{\mu}_2, \frac{\boldsymbol{\Sigma}_2}{\nu_2}\right)\right] \\ &+ (1-\nu_1)\left[\frac{1}{\sqrt{(2\pi)^{p_1}|\boldsymbol{\Sigma}_1|}}\, \mathrm{e}^{\{-\frac{1}{2}(\mathbf{y}_1-\boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_1^{-1}(\mathbf{y}_1-\boldsymbol{\mu}_1)\}}\Phi_{p_2}(\mathbf{y}_2; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)\right] \\ &= \nu_1\left[\phi_{p_1}(\mathbf{y}_1; \boldsymbol{\mu}_1, \nu_2^{-1}\boldsymbol{\Sigma}_1)\Phi_{p_2}(\mathbf{y}_2; \boldsymbol{\mu}_2, \nu_2^{-1}\boldsymbol{\Sigma}_2)\right] \\ &+ (1-\nu_1)\left[\phi_{p_1}(\mathbf{y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)\Phi_{p_2}(\mathbf{y}_2; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)\right]. \end{aligned}$$

□

B.2 Complementary results of simulation study

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

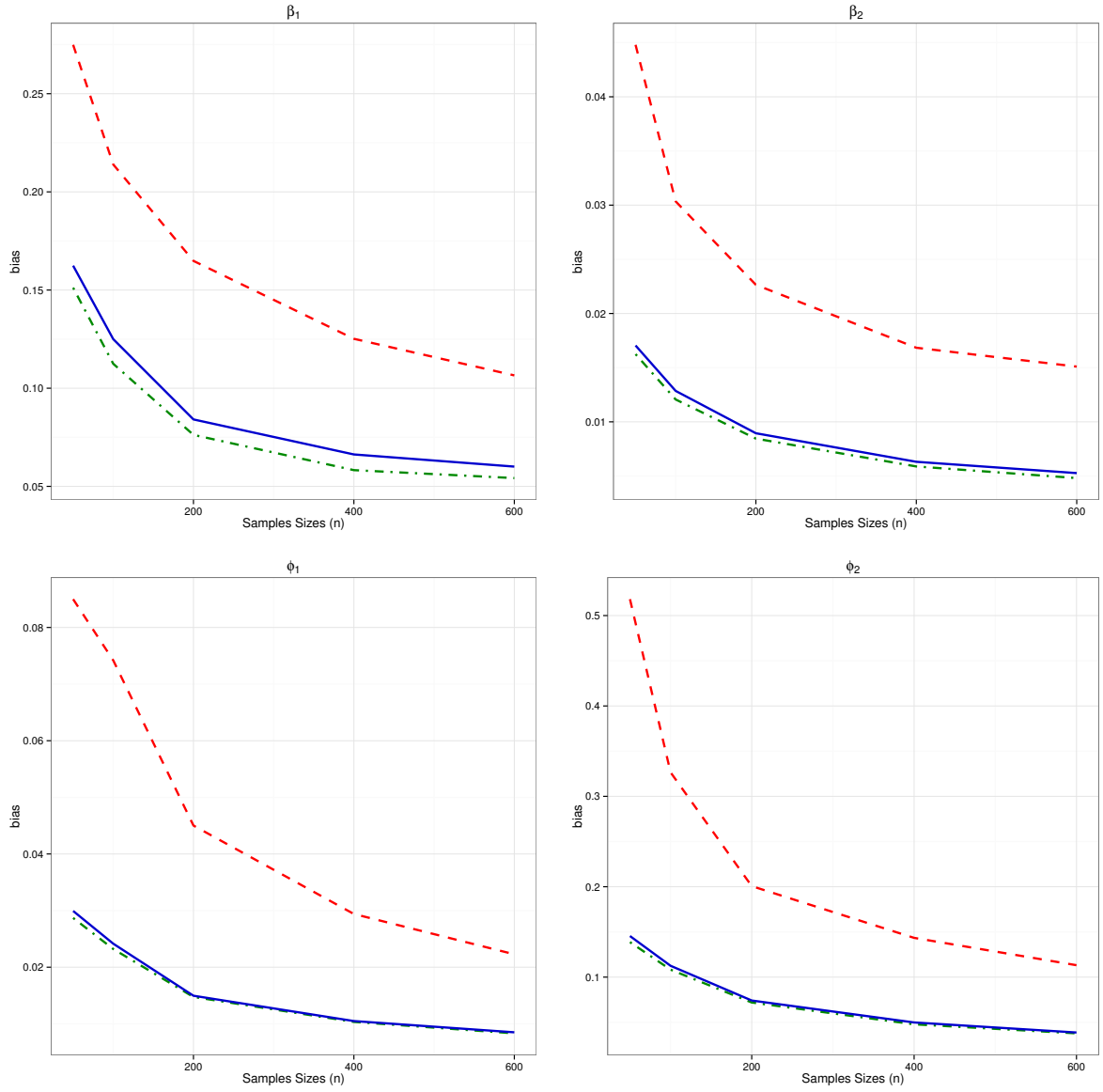

Figure 25 – **Simulation study - Scenario 1.** Absolute bias of the parameter estimates in the SMN-CR model under 10% of censoring and different samples sizes. The solid line (blue) represents the T-CR model, the dotted line (red) represents the N-CR model and the dotdashed line (green) represents the SL-CR model.

Scenario 3: Convergence of the parameters estimates

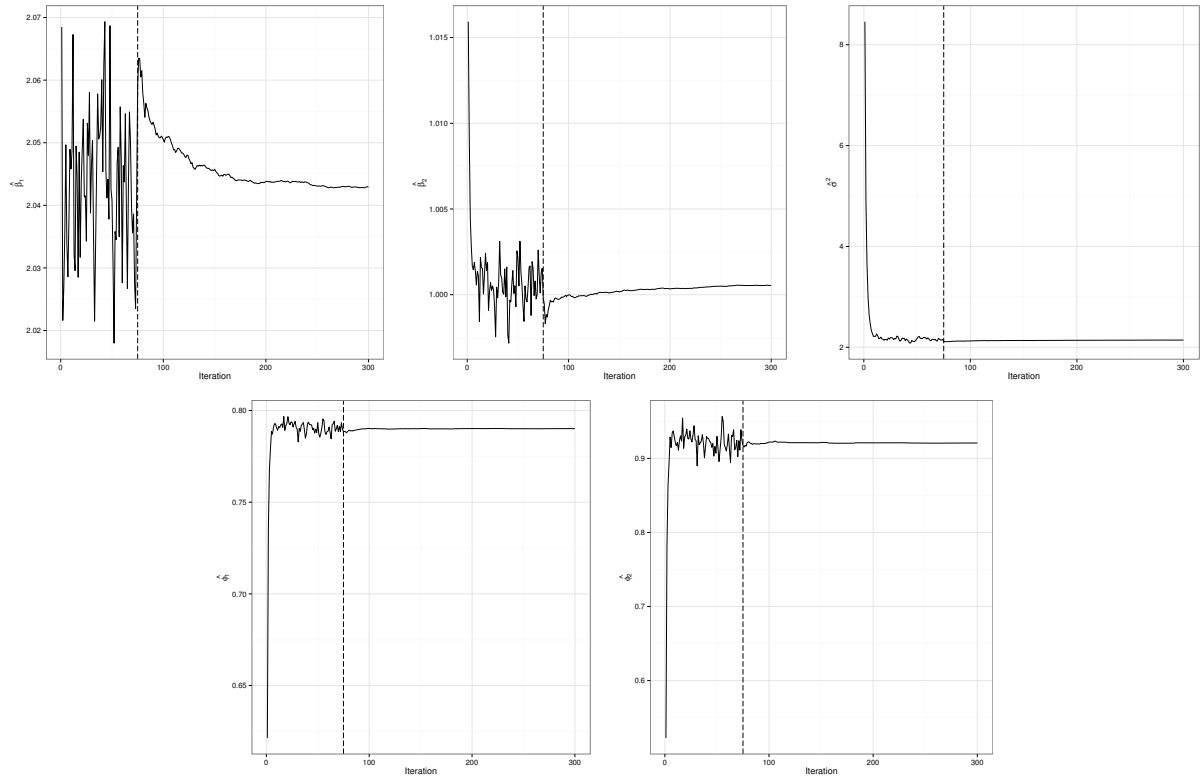

Figure 26 – **Simulation study - Scenario 3.** Convergence of the SAEM parameters estimates for the T-CR model.

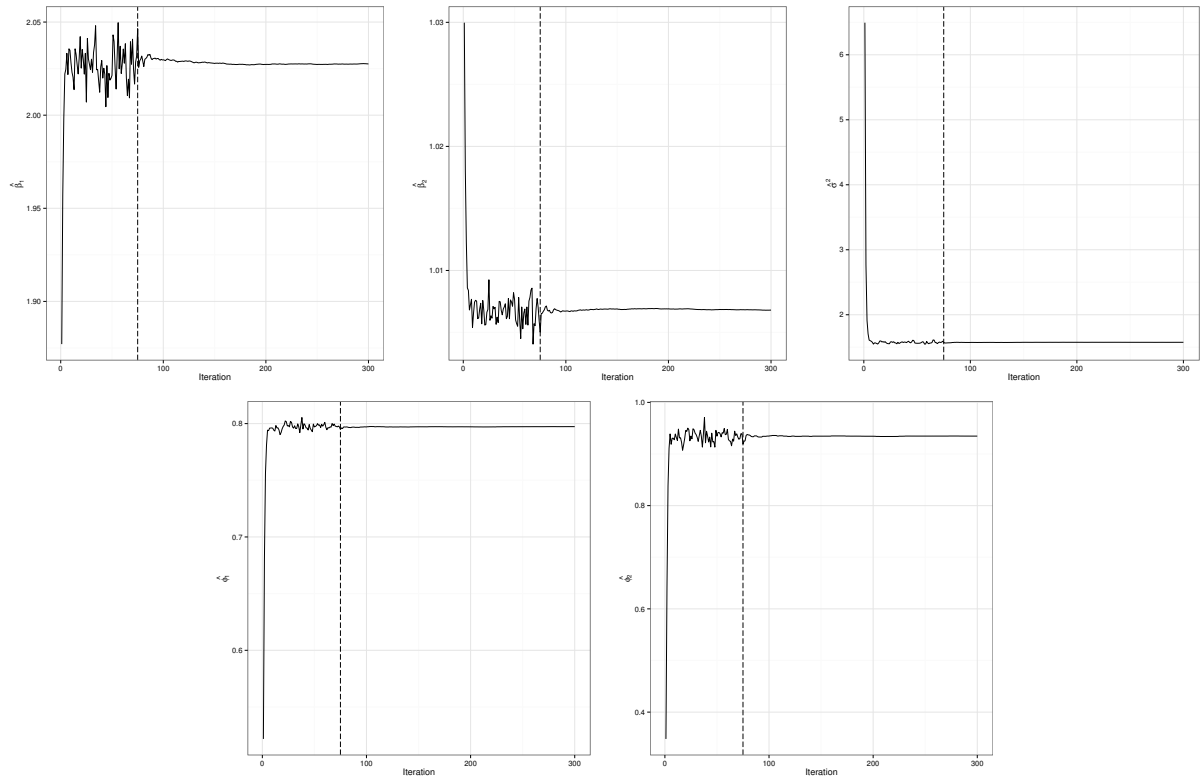

Figure 27 – **Simulation study - Scenario 3.** Convergence of the parameters estimates for the SL-CR model.

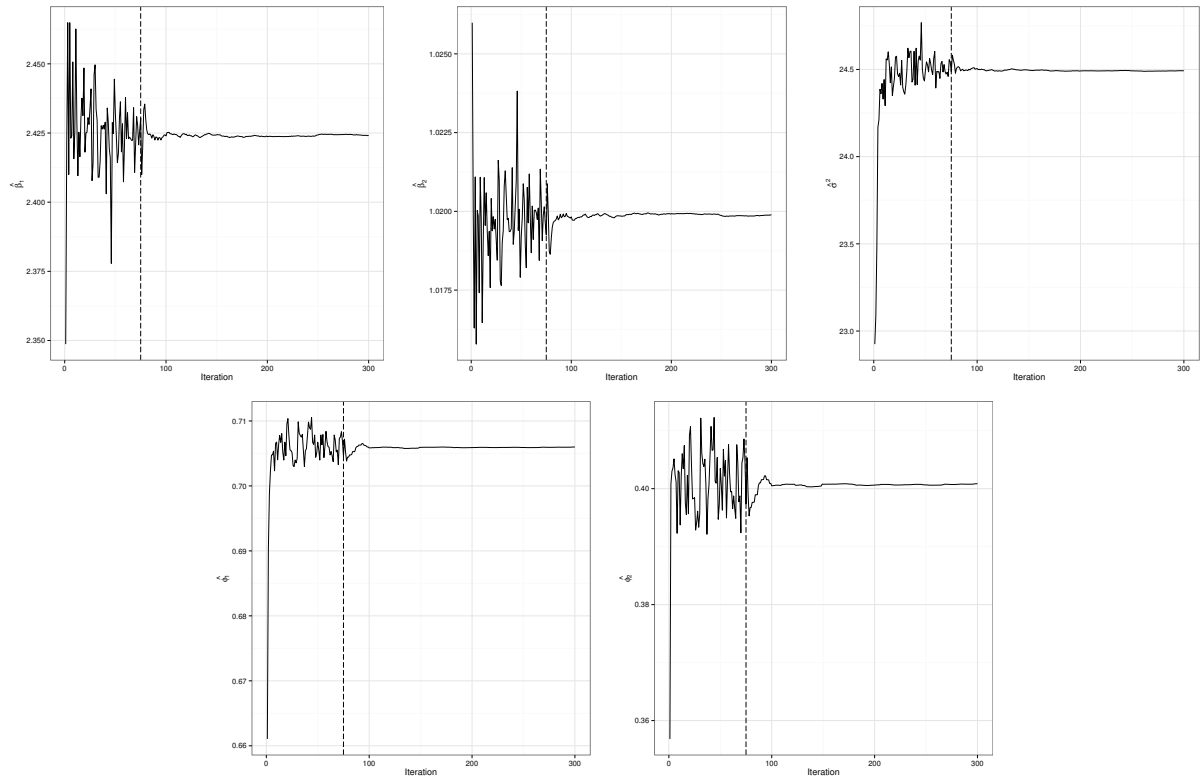

Figure 28 – **Simulation study - Scenario 3.** Convergence of the SAEM parameters estimates for the N-CR model.

B.3 Complementary results of the UTI data: convergence of the parameters estimates

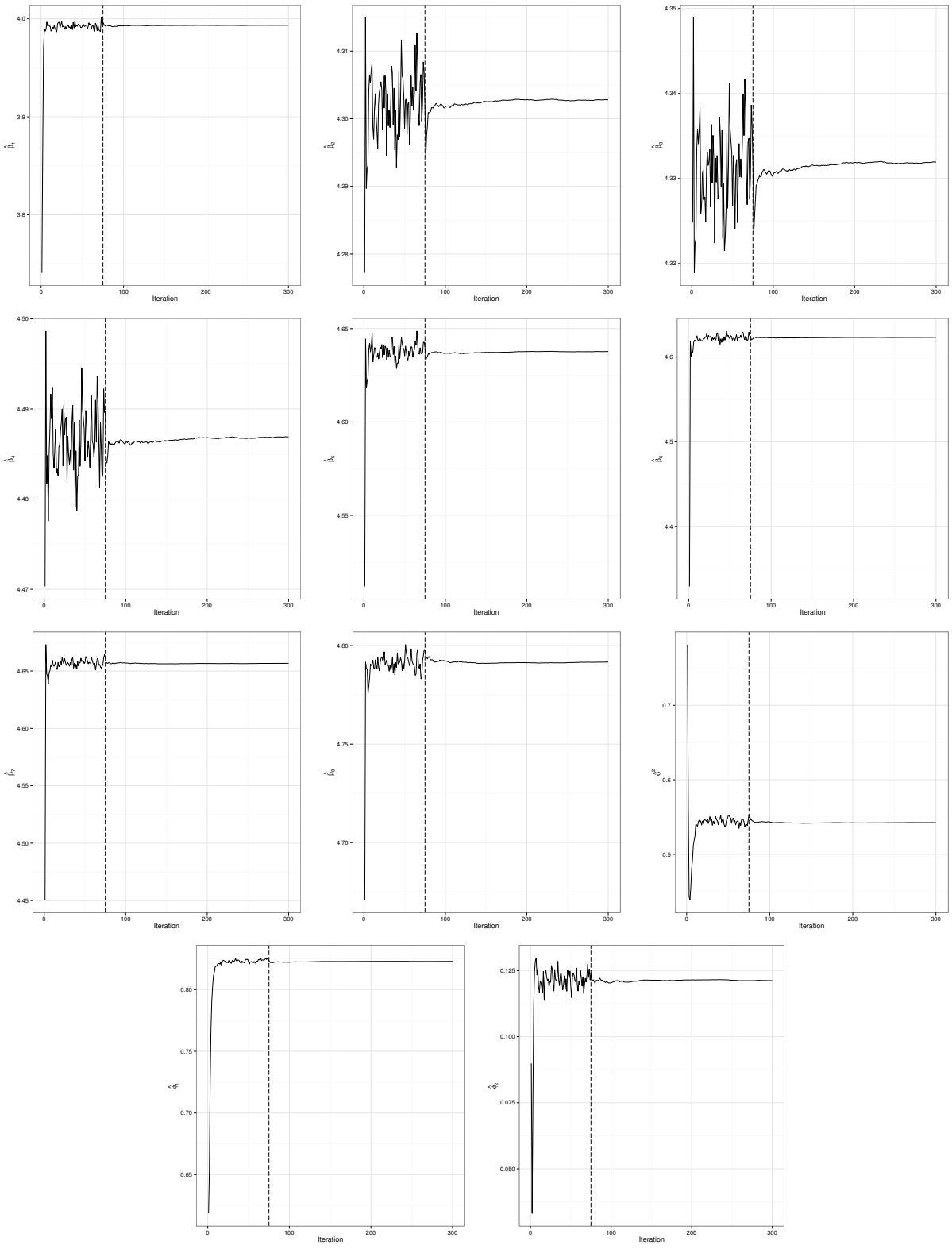

Figure 29 – **UTI data.** Convergence of the SAEM parameters estimates.

APPENDIX C – Supplementary Material for Chapter 5

C.1 Full conditional distributions

Table 25 – Full conditional distributions of $\kappa_i|\mathbf{y}_i, \mathbf{b}_i, \tau_i$.

Distribution of $\boldsymbol{\epsilon}_i$	Distribution of κ_i	Distribution of $\kappa_i|\mathbf{y}_i, \mathbf{b}_i, \tau_i$
$\mathrm{T}_{si}(\mathbf{0}, \mathbf{R}_i, \nu)$	$\mathrm{Gamma}(\nu/2, \nu/2)$	$\mathrm{Gamma}\left((\nu + s_i)/2, (D^2_{e_i} + \nu)/2\right)$
$\mathrm{SL}_{si}(\mathbf{0}, \mathbf{R}_i, \nu)$	$\mathrm{Beta}(\nu, 1)$	$\mathrm{TGamma}\left(\nu + s_i/2, D^2_{e_i}/2, 1\right)$
$\mathrm{CN}_{si}(\mathbf{0}, \mathbf{R}_i, \nu_1, \nu_2)$	$\nu_1 \mathbb{I}_{\{\nu_2\}}(\kappa_i) + (1-\nu_1)\mathbb{I}_{\{1\}}(\kappa_i)$	$P(\kappa_i = \nu_2) = 1 - P(\kappa_i = 1) = p_1/p_1 + p_2$ $p_1 = \nu_1 \nu_2^{s_i/2} exp\{-\frac{\nu_2}{2} D^2_{e_i}\}$ $p_2 = (1-\nu_1) exp\{-\frac{1}{2} D^2_{e_i}\}$

where $D^2_{e_i} = (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)$

Table 26 – Full conditional distributions of $\tau_i|\mathbf{y}_i, \mathbf{b}_i, \kappa_i$.

Distribution of $\mathbf{b}_i$	Distribution of τ_i	Distribution of $\tau_i|\mathbf{y}_i, \mathbf{b}_i, \kappa_i$
$\mathrm{T}_q(\mathbf{0}, \mathbf{D}, \eta)$	$\mathrm{Gamma}(\eta/2, \eta/2)$	$\mathrm{Gamma}\left((\eta + q)/2, (D^2_{\mathbf{b}_i} + \eta)/2\right)$
$\mathrm{SL}_q(\mathbf{0}, \mathbf{D}, \eta)$	$\mathrm{Beta}(\eta, 1)$	$\mathrm{TGamma}\left(\eta + q/2, D^2_{\mathbf{b}_i}/2, 1\right)$
$\mathrm{CN}_q(\mathbf{0}, \mathbf{D}, \eta_1, \eta_2)$	$\eta_1 \mathbb{I}_{\{\eta_2\}}(\tau_i) + (1-\eta_1)\mathbb{I}_{\{1\}}(\tau_i)$	$P(\tau_i = \eta_2) = 1 - P(\tau_i = 1) = q_1/q_1 + q_2$ $q_1 = \eta_1 \eta_2^{q/2} exp\{-\frac{\eta_2}{2} D^2_{\mathbf{b}_i}\}$ $q_2 = (1-\eta_1) exp\{-\frac{1}{2} D^2_{\mathbf{b}_i}\}$

where $D^2_{\mathbf{b}_i} = \mathbf{b}_i^\top \mathbf{D}^{-1}\mathbf{b}_i$

C.2 Maximum likelihood estimation

In this section, we propose the derivation of the parameters estimates, $\widehat{\boldsymbol{\nu}}$ and $\widehat{\boldsymbol{\eta}}$, for the some particular distributions. To find the estimates of $\widehat{\boldsymbol{\nu}}$ and $\widehat{\boldsymbol{\eta}}$, we just need to

solve

$$\frac{\partial Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\nu}} = \frac{\partial}{\partial \boldsymbol{\nu}} \sum_{i=1}^{n} \widehat{\ell h_{1i}} = 0 \quad \text{and} \quad \frac{\partial Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\eta}} = \frac{\partial}{\partial \boldsymbol{\eta}} \sum_{i=1}^{n} \widehat{\ell h_{2i}} = 0,$$

where $\widehat{\ell h_{1i}} = E\left[\log h_1(\kappa_i|\boldsymbol{\nu})|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\right]$ and $\widehat{\ell h_{2i}} = E\left[\log h_2(\tau_i|\boldsymbol{\eta})|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\right]$.

- If the response vector has a Student-t distribution, we have that $\kappa_i \sim \text{Gamma}(\nu/2, \nu/2)$, and

$$\widehat{\ell h_{1i}} = -\log\left(\Gamma\left(\frac{\nu}{2}\right)\right) + \frac{\nu}{2}\log\left(\frac{\nu}{2}\right) + \left(\frac{\nu}{2} - 1\right)\widehat{\log \kappa_i} - \frac{\nu\widehat{\kappa}_i}{2},$$

  with $\widehat{\kappa}_i = E\left[\kappa_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\right]$ and $\widehat{\log \kappa_i} = E\left[\log(\kappa_i)|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\right]$. Then, $\widehat{\nu}$ can be obtained by using a one-dimensional Newton-Raphson algorithm to solve

$$\log\left(\frac{\nu}{2}\right) - \Psi\left(\frac{\nu}{2}\right) + 1 + \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\log \kappa_i} - \widehat{\kappa_i}\right) = 0,$$

  where $\Psi(z) = \dfrac{d}{dz}\log(\Gamma(z))$ denotes the digamma function.

- If the response vector have Slash distribution, we have that $\kappa_i \sim \text{Beta}(\nu, 1)$, and

$$\widehat{\ell h_{1i}} = \log(\nu) + (\nu - 1)\widehat{\log \kappa_i}.$$

  So,

$$\widehat{\nu} = -\frac{n}{\sum_{i=1}^{n}\widehat{\log \kappa_i}}.$$

For the random effects we follow the same strategy.

- If the random effect have Student-t distribution, we have that $\tau_i \sim \text{Gamma}(\eta/2, \eta/2)$, and

$$\widehat{\ell h_{2i}} = -\log\left(\Gamma\left(\frac{\eta}{2}\right)\right) + \frac{\eta}{2}\log\left(\frac{\eta}{2}\right) + \left(\frac{\eta}{2} - 1\right)\widehat{\log \tau_i} - \frac{\eta\widehat{\tau}_i}{2},$$

  with $\widehat{\tau}_i = E\left[\tau_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\right]$ and $\widehat{\log \tau_i} = E\left[\log(\tau_i)|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\right]$. Then, $\widehat{\eta}$ can be obtained by using a one-dimensional Newton-Raphson algorithm to solve

$$\log\left(\frac{\eta}{2}\right) - \Psi\left(\frac{\eta}{2}\right) + 1 + \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\log \tau_i} - \widehat{\tau}_i\right) = 0.$$

- If the random effects have Slash distribution, we have that $\tau_i \sim \text{Beta}(\eta, 1)$, and

$$\widehat{\ell h_{2i}} = \log(\eta) + (\eta - 1)\widehat{\log \tau_i}.$$

  So,

$$\widehat{\eta} = -\frac{n}{\sum_{i=1}^{n}\widehat{\log \tau_i}}.$$

We replace the conditional expectations, $\widehat{\kappa}_i$, $\widehat{\log \kappa_i}$, $\widehat{\tau}_i$ and $\widehat{\log \tau_i}$, by the stochastic approximations described in Section 5.3.

C.3 AIC and BIC decomposition

Let $\mathbf{y}_{i1}^{\star}$ be the s_{i1}-vector with $r^{\star}$ outcomes and $\mathbf{y}_{i2}^{\star}$ be the s_{i2}-vector with $r^{\star}-r$ outcomes, where $s_{i1}=n_i\cdot r^{\star}$ and $s_{i1}=n_i\cdot(r-r^{\star})$, with $s_i=s_{i1}+s_{i2}$. After reordering, $\mathbf{y}$, $\mathbf{V}_i$, $\mathbf{X}_i$, $\mathbf{Z}_i$ and $\mathbf{R}_i$ can be partitioned as $\mathbf{y}_i=vec(\mathbf{y}_{i1}^{\star},\mathbf{y}_{i2}^{\star})$, $\mathbf{V}_i=vec(\mathbf{V}_i^1,\mathbf{V}_i^2)$, $\mathbf{X}_i^{\top}=(\mathbf{X}_i^1,\mathbf{X}_i^2)$, $\mathbf{Z}_i^{\top}=(\mathbf{Z}_i^1,\mathbf{Z}_i^2)$ and $\mathbf{R}_i=\begin{pmatrix}\mathbf{R}_i^{11} & \mathbf{R}_i^{12}\\ \mathbf{R}_i^{21} & \mathbf{R}_i^{22}\end{pmatrix}$. Without considering the censoring for the moment, we know that $\mathbf{y}_i \mid \mathbf{b}_i,\kappa_i \sim \mathrm{N}_{s_i}(\mathbf{X}_i\boldsymbol{\beta}+\mathbf{Z}_i\mathbf{b}_i,\kappa_i^{-1}\mathbf{R}_i)$, and we have that,

$$\begin{aligned}\mathbf{y}_{i1}^{\star} \mid \mathbf{b}_i,\kappa_i &\sim \mathrm{N}_{s_{i1}}(\mathbf{X}_i^1\boldsymbol{\beta}+\mathbf{Z}_i^1\mathbf{b}_i,\kappa_i^{-1}\mathbf{R}_i^{11}), &\text{(C.1)}\\ \mathbf{y}_{i2}^{\star} \mid \mathbf{y}_{i1}^{\star},\mathbf{b}_i,\kappa_i &\sim \mathrm{N}_{s_{i2}}(\boldsymbol{\mu}_i^{\star},\kappa_i^{-1}\mathbf{S}_i^{\star}), &\text{(C.2)}\end{aligned}$$

where $\boldsymbol{\mu}_i^{\star}=(\mathbf{X}_i^2\boldsymbol{\beta}+\mathbf{Z}_i^2\mathbf{b}_i)+\mathbf{R}_i^{21}(\mathbf{R}_i^{11})^{-1}(\mathbf{y}_{i1}^{\star}-\mathbf{X}_i^1\boldsymbol{\beta}-\mathbf{Z}_i^1\mathbf{b}_i)$ and $\mathbf{S}_i^{\star}=\mathbf{R}_i^{22}-\mathbf{R}_i^{21}(\mathbf{R}_i^{11})^{-1}\mathbf{R}_i^{12}$.

Then, using (C.2) and (C.1) the distributions $f(\mathbf{y}_{i1}^{\star};\boldsymbol{\theta}_1^{\star})$ and $f(\mathbf{y}_{i2}^{\star}|\mathbf{y}_{i1}^{\star})$ are given by

$$\begin{aligned}f(\mathbf{y}_{i1}^{\star};\boldsymbol{\theta}_1^{\star}) &= \int\int f(\mathbf{y}_{i1}^{\star}|\mathbf{b}_i,\kappa_i,\boldsymbol{\theta}_1^{\star})f(\mathbf{b}_i|\boldsymbol{\theta}_1^{\star})f(\kappa_i|\boldsymbol{\theta}_1^{\star})d\kappa_i d\mathbf{b}_i,\\ f(\mathbf{y}_{21}^{\star};\boldsymbol{\theta}_2^{\star}) &= \int\int f(\mathbf{y}_{i2}^{\star}|\mathbf{y}_{i1}^{\star},\mathbf{b}_i,\kappa_i,\boldsymbol{\theta}_2^{\star})f(\mathbf{b}_i|\boldsymbol{\theta}_2^{\star})f(\kappa_i|\boldsymbol{\theta}_2^{\star})d\kappa_i d\mathbf{b}_i.\end{aligned}$$

C.4 Empirical information matrix

The elements of $\widehat{\mathbf{s}}_i$ in Equation (5.10) are given by

$$\begin{aligned}\widehat{\mathbf{s}}_{i,\boldsymbol{\beta}} &= (\widehat{\mathbf{s}}_{i,\beta_1},\ldots,\widehat{\mathbf{s}}_{i,\beta_p})^{\top}=\mathbf{X}_i^{\top}\widehat{\mathbf{R}}_i^{-1}\left(\widehat{\kappa\mathbf{y}}_i-\mathbf{Z}_i\widehat{\kappa\mathbf{b}}_i\right)-\widehat{\kappa}_i\mathbf{X}_i^{\top}\widehat{\mathbf{R}}_i^{-1}\mathbf{X}_i\widehat{\boldsymbol{\beta}},\\ \widehat{\mathbf{s}}_{i,\alpha_r} &= -\frac{1}{2}\mathrm{tr}\left(\widehat{\mathbf{D}}^{-1}\dot{\mathbf{D}}_r\widehat{\mathbf{D}}^{-1}(\widehat{\mathbf{D}}-\widehat{\tau\mathbf{b}_i^2})\right),\\ \widehat{\mathbf{s}}_{i,\sigma_j^2} &= -\frac{n_i}{2}\log|\widehat{\boldsymbol{\Sigma}}_i^{-1}\dot{\boldsymbol{\Sigma}}_j|+\frac{1}{2}\mathrm{tr}\left[\left(\widehat{\boldsymbol{\Sigma}}^{-1}\dot{\boldsymbol{\Sigma}}_j\widehat{\boldsymbol{\Sigma}}^{-1}\otimes\boldsymbol{\Omega}_i^{-1}(\widehat{\boldsymbol{\phi}})\right)\widehat{\kappa\mathbf{E}_i}\right],\\ \widehat{\mathbf{s}}_{i,\phi_s} &= -\frac{r}{2}\log|\boldsymbol{\Omega}_i^{-1}(\widehat{\boldsymbol{\phi}})\dot{\boldsymbol{\Omega}}_{is}|+\frac{1}{2}\mathrm{tr}\left[\left(\widehat{\boldsymbol{\Sigma}}^{-1}\otimes\boldsymbol{\Omega}_i^{-1}(\widehat{\boldsymbol{\phi}})\dot{\boldsymbol{\Omega}}_{is}\boldsymbol{\Omega}_i^{-1}(\widehat{\boldsymbol{\phi}})\right)\widehat{\kappa\mathbf{E}_i}\right],\end{aligned}$$

where $\dot{\mathbf{D}}_r=\frac{\partial\mathbf{D}}{\partial\alpha_r}|_{\boldsymbol{\alpha}=\widehat{\boldsymbol{\alpha}}}$, $r=1,\ldots,\dim(\boldsymbol{\alpha})$; $\dot{\boldsymbol{\Sigma}}_r=\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_j^2}|_{\boldsymbol{\sigma}=\widehat{\boldsymbol{\sigma}}}$, $j=1,\ldots,\dim(\boldsymbol{\sigma})$; and $\dot{\boldsymbol{\Omega}}_{is}=\frac{\partial\boldsymbol{\Omega}_i}{\partial\phi_s}|_{\boldsymbol{\phi}=\widehat{\boldsymbol{\phi}}}$, $s=1,2$. For the DEC structure we have that

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

As was showed in Appendix C.2, for computing $\widehat{\mathbf{s}}_{i,\nu}$ and $\widehat{\mathbf{s}}_{i,\eta}$, we need to define the distribution of the error term and random effect. For the Student-t and Slash distribution, we have that:

- If the response vector have Student-t distribution,

$$\widehat{\mathbf{s}}_{i,\nu} = -\frac{1}{2}\left[\Psi(\nu/2) - \log(\nu/2) - 1 - (\widehat{\log \kappa_i} - \widehat{\kappa}_i)\right];$$

- If the response vector have Slash distribution,

$$\widehat{\mathbf{s}}_{i,\nu} = \frac{1}{\nu} + \widehat{\log \kappa_i};$$

- If the random effects have Student-t distribution,

$$\widehat{\mathbf{s}}_{i,\eta} = -\frac{1}{2}\left[\Psi(\eta/2) - \log(\eta/2) - 1 - (\widehat{\log \tau_i} - \widehat{\tau}_i)\right];$$

- If the random effects have Slash distribution,

$$\widehat{\mathbf{s}}_{i,\eta} = \frac{1}{\eta} + \widehat{\log \tau_i}.$$

C.5 Complementary results of the simulation study

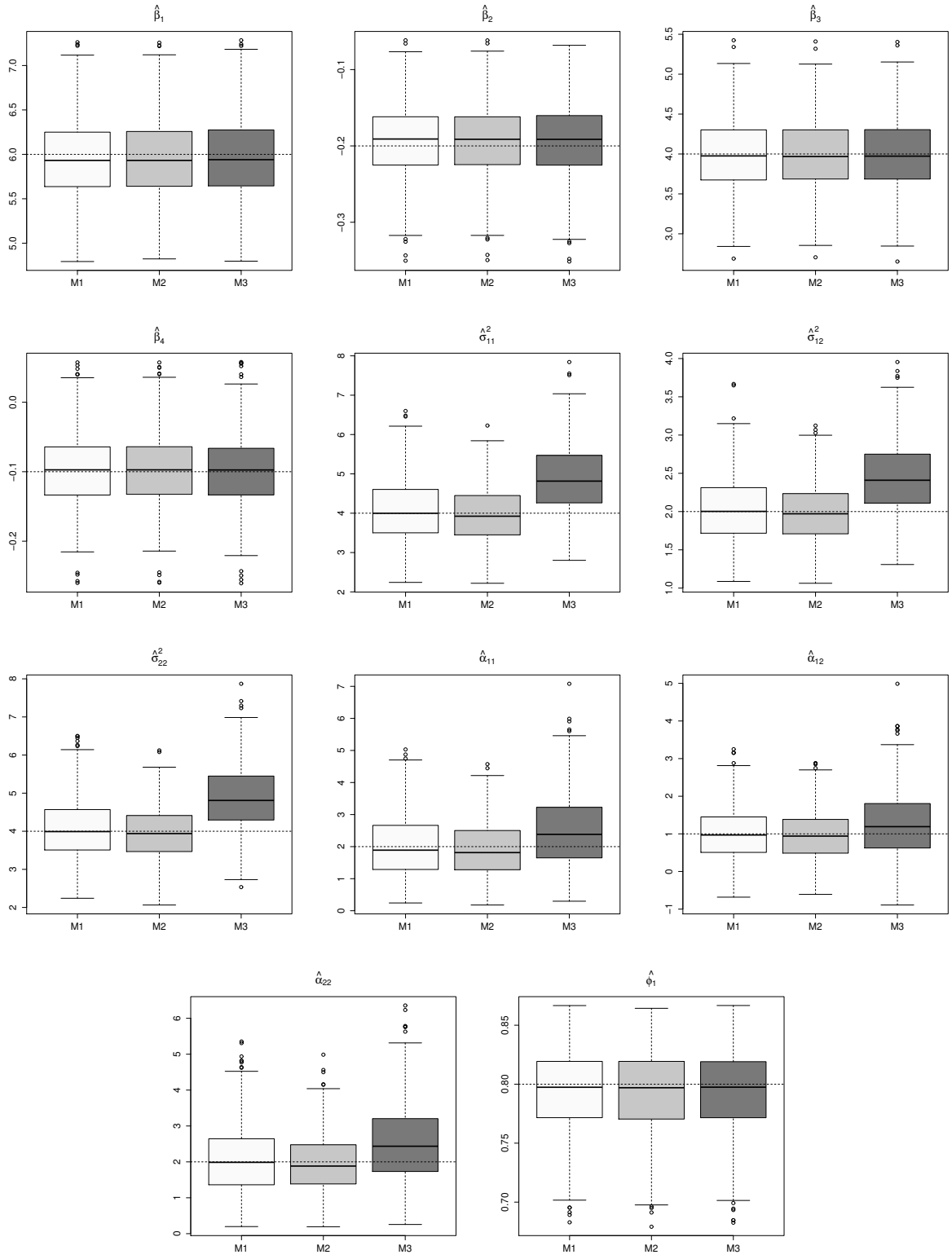

Figure 30 – **Simulation study**: Boxplots of the parameter estimates. Dotted lines indicate the true parameter value. The censoring proportion considered is 10%.

C.6 Complementary results of the aplpication: convergence of the parameters estimates

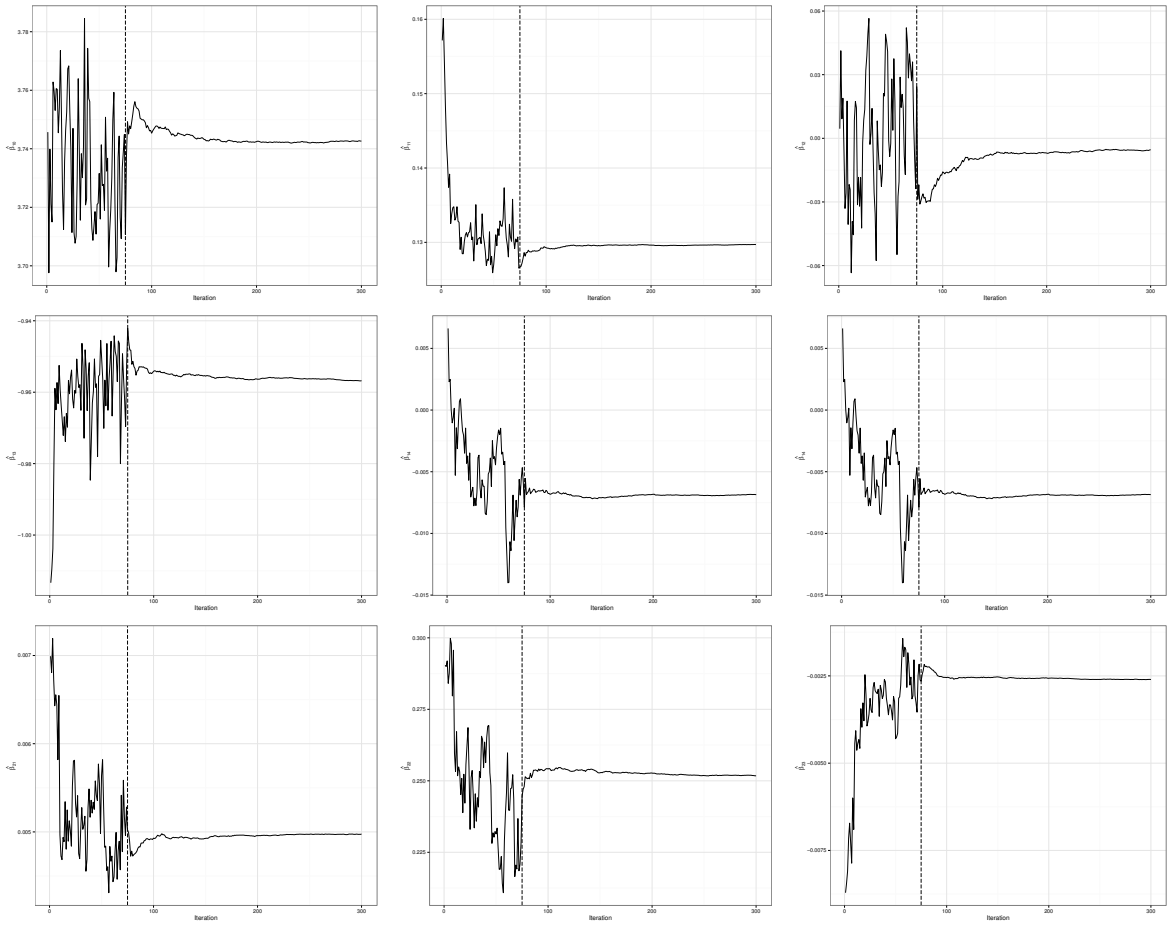

Figure 31 – **A5055 data.** Convergence of the SAEM parameters estimates $\widehat{\beta}_{10}$, $\widehat{\beta}_{11}$, $\widehat{\beta}_{12}$, $\widehat{\beta}_{13}$, $\widehat{\beta}_{14}$, $\widehat{\beta}_{20}$, $\widehat{\beta}_{21}$, $\widehat{\beta}_{22}$ and $\widehat{\beta}_{23}$.

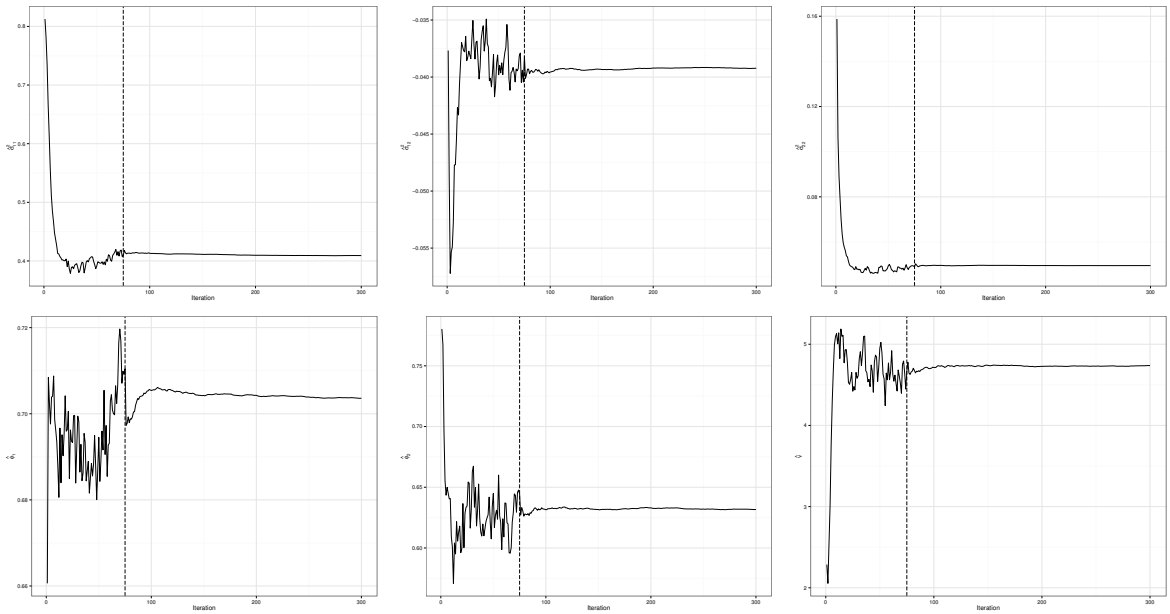

Figure 32 – **A5055 data.** Convergence of the SAEM parameters estimates $\widehat{\sigma^2_{11}}$, $\widehat{\sigma^2_{12}}$, $\widehat{\phi}_1$, $\widehat{\phi}_2$ and $\widehat{\nu}$.

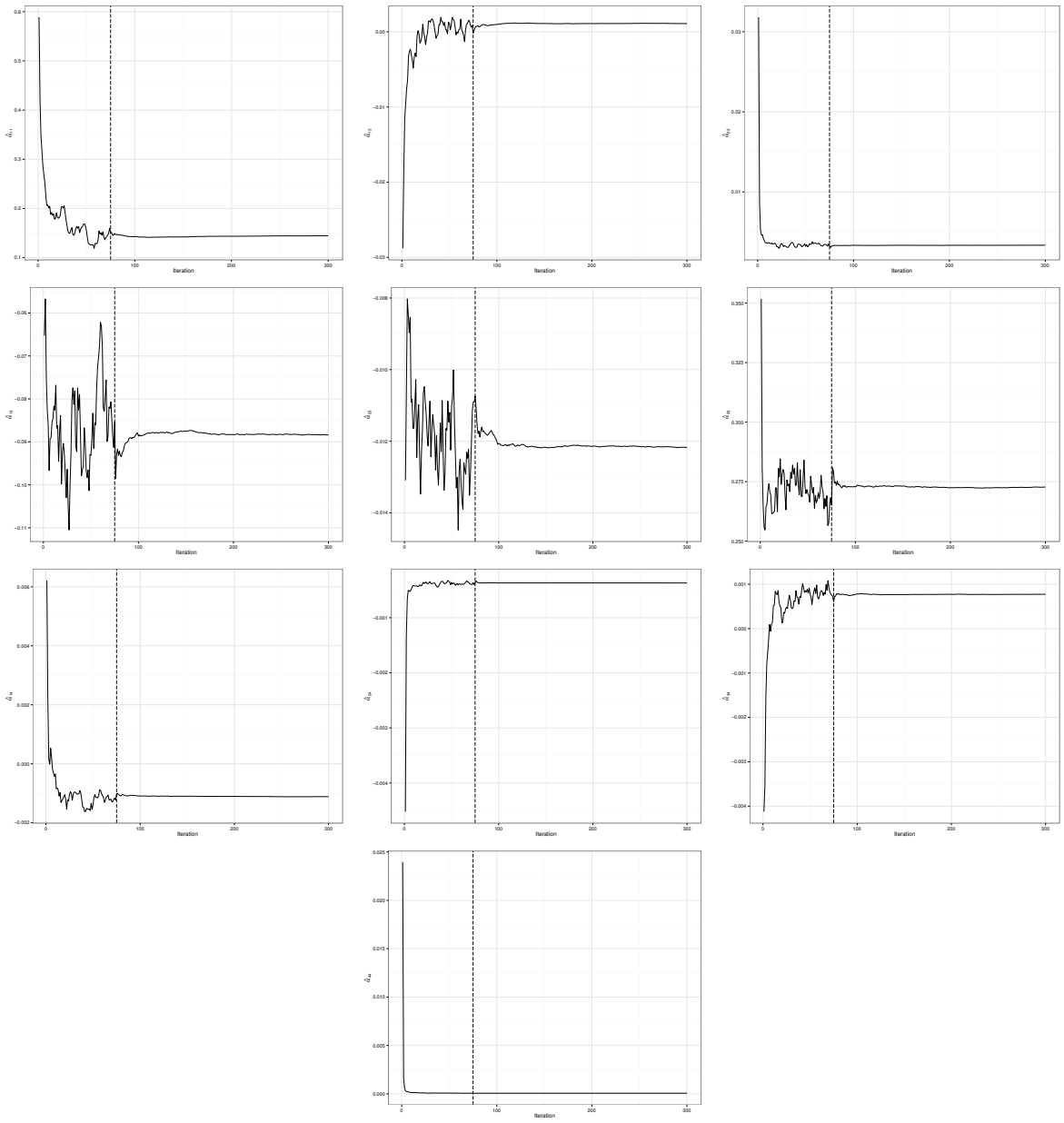

Figure 33 – **A5055 data.** Convergence of the SAEM parameters estimates $\widehat{\alpha}_{11}$, $\widehat{\alpha}_{12}$, $\widehat{\alpha}_{22}$, $\widehat{\alpha}_{13}$, $\widehat{\alpha}_{23}$, $\widehat{\alpha}_{33}$, $\widehat{\alpha}_{14}$, $\widehat{\alpha}_{24}$, $\widehat{\alpha}_{34}$ and $\widehat{\alpha}_{44}$.

Annex

ANNEX A – Multivariate Measurement Error Models Based on Student-t Distribution under Censored Responses

Abstract

Measurement error models constitute a wide class of models, that include linear and nonlinear regression models. They are very useful to model many real life phenomena, particularly in the medical and biological areas. The great advantage of these models is that, in some sense, they can be represented as mixed effects models, allowing to us the implementation of well-known techniques, like the EM-algorithm for the parameter estimation. In this work, we consider a class of multivariate measurement error models where the observed response and/or covariate are not fully observed, *i.e.*, the observations are subject to certain threshold values below or above which the measurements are not quantifiable. Consequently, these observations are considered censored. We assume a Student-t distribution for the unobserved true values of the mismeasured covariate and the error term of the model, providing a robust alternative for parameter estimation. Our approach relies on a likelihood-based inference using the EM-algorithm. The proposed method is illustrated through simulation studies and the analysis of a dataset consisting on the measurements of the testicular volume of 42 adolescents.

A.1 Introduction

Measurement error – hereafter ME – models (also known as error-in-variables models) are defined as regression models where the covariates cannot be measured/observed directly, or are measured with a substantial error. From a practical point of view, such models are very useful because they take into account some notions of randomness inherent to the covariates. For example, in AIDS studies, linear and nonlinear mixed-effects models are typically considered to study the relationship between the viral load (HIV-1 RNA) measures and CD4 (T-cells) cell count. However, as pointed out by many authors (see for instance Wu (2010); Bandyopadhyay et al. (2015), among others), this covariate is measured (in general) with substantial error. This is because, in most HIV clinical trials, cell counts are measured periodically with a substantial amount of variability.

A wide variety of proposals exist in the statistical literature trying to deal with the presence of ME in multivariate data. For example, Carroll et al. (1997) proposed a

generalized linear mixed ME model and Buonaccorsi et al. (2000) (see also Dumitrescu (2010)) studied estimation of the variance components in a linear mixed-effect model with ME in a time varying covariate. Zhang et al. (2011) introduced a multivariate ME model including the presence of zero inflation. Recently, Abarin et al. (2014) proposed a method of moments for the parameter estimation in the linear mixed-effect with ME model. Moreover, Cabral et al. (2014) studied a multivariate ME model using finite mixtures of skew Student-t distributions. A comprehensive review of ME models can be found in the books of Fuller (1987), Cheng and Van Ness (1999), Carroll et al. (2006) and Buonaccorsi (2010).

Although many models for multivariate data consider the existence of mismeasured covariates, many of them do not consider censored observations or detection limits for the response variable. This aspect is relevant, since in many studies the observed response is subject to maximum/minimum detection limits. For that reason, clearly there is a need for a new methodology that takes into account censored responses in multivariate data and mismeasured covariates at the same time. We propose an approach where the random observational errors and the unobserved latent variable are jointly modeled by a Student-t distribution, which has heavier tails than the normal one. Besides this, our estimation approach relies on an exact EM-type algorithm, providing explicit expressions for the E and M steps, obtaining as byproduct the asymptotic covariance of the maximum likelihood estimates. To illustrate the applicability of the method, we analyze a real dataset consisting of measurements of the testicular volume of 42 adolescents using five different techniques.

The work is organized as follows. Section A.2 presents some results about the multivariate Student-t distribution, focusing on its truncated version. Section A.3 proposes the ME model for censored multivariate responses under the Student-t distribution. Sections A.4 and A.5 present the likelihood-based estimation and standard errors of the parameter estimates in the proposed model via an EM-type algorithm, respectively. The analysis of a real dataset is presented in Section A.6. Section A.7 presents the results of simulation studies conducted to examine the performance of the proposed method with respect to the asymptotic properties of the ML estimates, and the consequences of the inappropriateness of the normality assumption. Finally, the work closes with some conclusions in Section A.8

A.2 The multivariate Student-t distribution and truncated related ones

We say that the random vector $\mathbf{Y} : p \times 1$ has a Student-t distribution with location vector $\boldsymbol{\mu}$, dispersion matrix $\boldsymbol{\Sigma}$ and ν degrees of freedom, when its probability

density function (*pdf*) is given by

$$\mathrm{t}_p(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) = \frac{\Gamma(\frac{p+\nu}{2})}{\Gamma(\frac{\nu}{2})\pi^{p/2}}\nu^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\left(1+\frac{\mathrm{d}_{\boldsymbol{\Sigma}}(\mathbf{y},\boldsymbol{\mu})}{\nu}\right)^{-(p+\nu)/2}, \tag{A.1}$$

where $\Gamma(\cdot)$ is the standard gamma function and

$$\mathrm{d}_{\boldsymbol{\Sigma}}(\mathbf{y},\boldsymbol{\mu}) = (\mathbf{y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu}),$$

is the Mahalanobis distance. The cumulative distribution function (*cdf*) of $\mathbf{Y}$ is denoted by $\mathrm{T}_p(\cdot \mid \boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$. If $\nu>1$, $\boldsymbol{\mu}$ is the mean of $\mathbf{Y}$, and if $\nu>2$, $\nu(\nu-2)^{-1}\boldsymbol{\Sigma}$ is its covariance matrix. We use the notation $\mathbf{Y}\sim\mathrm{t}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$.

It is possible to show that $\mathbf{Y}$ admits the stochastic representation

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

where $\mathbf{Z}$ and U are independent, and $\mathrm{Gamma}(a,b)$ denotes the gamma distribution with mean a/b. As ν tends to infinity, U converges to one with probability one and $\mathbf{Y}$ is approximately distributed as a $\mathrm{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ distribution. From this representation we can easily deduce that an affine transformation $\mathbf{AY}+\mathbf{b}$ has a $\mathrm{t}_q(\mathbf{A}\boldsymbol{\mu}+\mathbf{b},\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top,\nu)$ distribution, where $\mathbf{A}$ is a $q\times p$ matrix and $\mathbf{b}$ is a $q-$dimensional vector. For a reference with extensive material regarding the multivariate Student-t distribution, see Kotz and Nadarajah (2004).

The following result shows that the Student-t family of distributions is closed under marginalization and conditioning. The proof can be found in (Matos et al., 2013b, Prop. 1).

Proposição A.1. *Let* $\mathbf{Y}\sim\mathrm{t}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$. *Consider the partition* $\mathbf{Y}=(\mathbf{Y}_1^\top,\mathbf{Y}_2^\top)^\top$, *with* $\mathbf{Y}_1 : p_1\times 1$ *and* $\mathbf{Y}_2 : p_2\times 1$. *Accordingly, consider the partitions* $\boldsymbol{\mu}=(\boldsymbol{\mu}_1^\top,\boldsymbol{\mu}_2^\top)^\top$ *and* $\boldsymbol{\Sigma}=(\boldsymbol{\Sigma}_{ij})$, $i,j=1,2$. *Then*

(i) $\mathbf{Y}_1\sim\mathrm{t}_{p_1}(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11},\nu)$,

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

where

$$\boldsymbol{\mu}_{2.1}=\boldsymbol{\mu}_2+\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1-\boldsymbol{\mu}_1), \quad \widetilde{\boldsymbol{\Sigma}}_{22.1}=\frac{\nu+\mathrm{d}_{\boldsymbol{\Sigma}_{11}}(\mathbf{y}_1,\boldsymbol{\mu}_1)}{\nu+p_1}\boldsymbol{\Sigma}_{22.1}, \quad \textit{and}$$
$$\boldsymbol{\Sigma}_{22.1}=\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}.$$

Let $\mathbf{Y}\sim\mathrm{t}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$ and $\mathbb{D}$ be a Borel set in $\mathbb{R}^p$. We say that the random vector $\mathbf{Z}$ has a truncated Student-t distribution on $\mathbb{D}$ when $\mathbf{Z}$ has the same distribution as $\mathbf{Y}|(\mathbf{Y}\in\mathbb{D})$. In this case, the *pdf* of $\mathbf{Z}$ is given by

$$\mathrm{Tt}_p(\mathbf{Z}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu;\mathbb{D}) = \frac{\mathrm{t}_p(\mathbf{Z}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}{P(\mathbf{Y}\in\mathbb{D})}\mathbb{I}_{\mathbb{D}}(\mathbf{Z}),$$

where $\mathbb{I}_{\mathbb{D}}(\cdot)$ is the indicator function of $\mathbb{D}$, that is, $\mathbb{I}_{\mathbb{D}}(\mathbf{Z}) = 1$ if $\mathbf{Z} \in \mathbb{D}$ and $\mathbb{I}_{\mathbb{D}}(\mathbf{Z}) = 0$ otherwise. We use the notation $\mathbf{Z} \sim \mathrm{Tt}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{D})$. If $\mathbb{D}$ has the form

$$\mathbb{D} = \{(x_1, \ldots, x_p) \in \mathbb{R}^p;\ \ x_1 \leqslant d_1, \ldots, x_p \leqslant d_p\}, \tag{A.3}$$

then we use the notation $(\mathbf{Y} \in \mathbb{D}) = (\mathbf{Y} \leqslant \mathbf{d})$, where $\mathbf{d} = (d_1, \ldots, d_p)^\top$. In this case, $P(\mathbf{Y} \leqslant \mathbf{d}) = \mathrm{T}_p(\mathbf{d}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. Notice that we can have $d_i = +\infty$, $i = 1, \ldots, p$.

The following propositions are crucial to obtain the expectations in the E step of the EM type algorithm, which will be used to compute maximum likelihood estimates of the parameters in the model proposed in this work. Proofs can be found in (Matos et al., 2013b, Prop. 2 and 3). We will use the notations $\mathbf{Z}^{(0)} = 1$, $\mathbf{Z}^{(1)} = \mathbf{Z}$ and $\mathbf{Z}^{(2)} = \mathbf{Z}\mathbf{Z}^\top$.

Proposição A.2. *Let $\mathbf{Z} \sim \mathrm{Tt}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{D})$, where $\mathbb{D}$ is as in (A.3). Then, for $k = 0, 1, 2$,*

$$\mathrm{E}\left[\left(\frac{\nu + p}{\nu + d_{\boldsymbol{\Sigma}}(\mathbf{Z}, \boldsymbol{\mu})}\right)^r \mathbf{Z}^{(k)}\right] = c_p(\nu, r)\frac{\mathrm{T}_p(\mathbf{d}|\boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \nu + 2r)}{\mathrm{T}_p(\mathbf{d}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}\mathrm{E}[\mathbf{Y}^{(k)}], \tag{A.4}$$

where $\nu + 2r > 0$ and

$$\begin{aligned} \mathbf{Y} &\sim \mathrm{Tt}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \nu + 2r; \mathbb{D}), \\ \boldsymbol{\Sigma}^* &= \frac{\nu}{\nu + 2r}\boldsymbol{\Sigma}, \\ c_p(\nu, r) &= \left(\frac{\nu + p}{\nu}\right)^r \left(\frac{\Gamma((p + \nu)/2)\Gamma((\nu + 2r)/2)}{\Gamma(\nu/2)\Gamma((p + \nu + 2r)/2)}\right). \end{aligned} \tag{A.5}$$

Observe that the computation of the expectation on the left side of (A.4) is reduced to the computation of the moments of the truncated Student-t distribution in (A.5). These moments are available in closed form in Ho et al. (2012) and the implementations were done using the `R` package *TTmoment()*, available on CRAN.

Proposição A.3. *Let $\mathbf{Z} \sim \mathrm{Tt}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{D})$, where $\mathbb{D}$ is as in (A.3). Consider the partition $\mathbf{Z} = (\mathbf{Z}_1^\top, \mathbf{Z}_2^\top)^\top$, with $\mathbf{Z}_1 : p_1 \times 1$ and $\mathbf{Z}_2 : p_2 \times 1$. Accordingly, consider the partitions $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top$ and $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_{ij})$, $i, j = 1, 2$. Then ,*

$$\begin{aligned} \mathrm{E}\left[\left(\frac{\nu + p}{\nu + d_{\boldsymbol{\Sigma}}(\mathbf{Z}, \boldsymbol{\mu})}\right)^r \mathbf{Z}_2^{(k)} \middle| \mathbf{Z}_1 = \mathbf{Z}_1\right] &= \frac{h_p(p_1, \nu, r)}{(\nu + d_{\boldsymbol{\Sigma}_{11}}(\mathbf{Z}_1, \boldsymbol{\mu}_1))^r} \\ &\quad \times \frac{\mathrm{T}_{p_2}(\mathbf{d}_2|\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}^*_{22.1}, \nu + p_1 + 2r)}{\mathrm{T}_{p_2}(\mathbf{d}_2|\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}_{22.1}, \nu + p_1)}\mathrm{E}[\mathbf{Y}^{(k)}], \end{aligned}$$

where $\nu + p_1 + 2r > 0$, $\mathbf{d}_2 = (d_{p_1+1}, \ldots, d_p)^\top$,

$$\begin{aligned} \mathbf{Y} &\sim \mathrm{Tt}_{p_2}(\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}^*_{22.1}, \nu + p_1 + 2r; \mathbb{D}_2), \\ \mathbb{D}_2 &= \{(x_{p_1+1}, \ldots, x_p) \in \mathbb{R}^{p_2};\ \ x_{p_1+1} \leqslant d_{p_1+1}, \ldots, x_p \leqslant d_p\}, \\ \widetilde{\boldsymbol{\Sigma}}^*_{22.1} &= \frac{\nu + \mathrm{d}_{\boldsymbol{\Sigma}_{11}}(\mathbf{Z}_1, \boldsymbol{\mu}_1)}{\nu + p_1 + 2r}\boldsymbol{\Sigma}_{22.1}, \\ h_p(p_1, \nu, r) &= (\nu + p)^r \left(\frac{\Gamma((p + \nu)/2)\Gamma((p_1 + \nu + 2r)/2)}{\Gamma((p_1 + \nu)/2)\Gamma((p + \nu + 2r)/2)}\right), \end{aligned}$$

$\boldsymbol{\mu}_{2.1}$, $\boldsymbol{\Sigma}_{22.1}$ and $\widetilde{\boldsymbol{\Sigma}}_{22.1}$ are given in Proposition A.1.

A.3 Model specification

Let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ir})^\top$ be the vector of responses for the ith experimental unit, where Y_{ij} is the jth observed response of unit i (for $i = 1, \ldots, n$ and $j = 1, \ldots, r$). Let X_i be the ith observed value of the covariate and x_i be the unobserved (true) covariate value for unit i. Following Barnett (1969), the multivariate ME model is formulated as

$$X_i = x_i + \xi_i \tag{A.6}$$

and

$$\mathbf{Y}_i = \boldsymbol{\alpha} + \boldsymbol{\beta} x_i + \mathbf{e}_i, \tag{A.7}$$

where $\mathbf{e}_i = (e_{i1}, \ldots, e_{ir})^\top$ is a vector of measurement errors, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r)^\top$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r)^\top$ are vectors with regression parameters. Let $\boldsymbol{\epsilon}_i = (\xi_i, \mathrm{e}_i^\top)^\top$ and $\mathbf{Z}_i = (X_i, \mathbf{Y}_i^\top)^\top = (Z_{i1}, \ldots, Z_{ip})^\top$. Then, equations (A.6) and (A.7) imply

$$\mathbf{Z}_i = \mathbf{a} + \mathbf{b} x_i + \boldsymbol{\epsilon}_i = \mathbf{a} + \mathbf{B}\mathbf{r}_i, \;\; i = 1, \ldots, n, \tag{A.8}$$

where $\mathbf{a} = (0, \boldsymbol{\alpha}^\top)^\top$ and $\mathbf{b} = (1, \boldsymbol{\beta}^\top)^\top$ are $p \times 1$ vectors, with $p = r + 1$, $\mathbf{B} = [\mathbf{b}; \mathbf{I}_p]$ is a $p \times (p+1)$ matrix and $\mathbf{r}_i = (x_i, \boldsymbol{\epsilon}_i^\top)^\top$. Thus, from equation (A.8), the distribution of $\mathbf{Z}_i$ becomes specified once the distribution of $\mathbf{r}_i$ is specified. Usually, a normality assumption is made, such that

$$\mathbf{r}_i \overset{\text{iid.}}{\sim} \mathrm{N}_{1+p}\left(\begin{pmatrix} \mu_x \\ \mathbf{0}_p \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \mathbf{0}_p^\top \\ \mathbf{0}_p & \boldsymbol{\Omega} \end{pmatrix} \right), \quad i = 1, \ldots, n, \tag{A.9}$$

where $\mathbf{0}_p = (0, \ldots, 0)^\top : p \times 1$, $\boldsymbol{\Omega} = \mathrm{diag}(\omega_1^2, \ldots, \omega_p^2)$ and $\overset{\text{iid.}}{\sim}$ denotes independent and identically distributed random vectors. Marginally, we have that $x_i \overset{\text{iid.}}{\sim} \mathrm{N}(\mu_x, \sigma^2)$ and $\boldsymbol{\epsilon}_i \overset{\text{iid.}}{\sim} \mathrm{N}_r(\mathbf{0}, \boldsymbol{\Omega})$ are independent for all $i = 1, \ldots, n$. For more details see, for example, (Fuller, 1987, Sec. 4.1).

In order to obtain robust estimation of the parameters in the model, we propose to replace assumption (A.9) by

$$\mathbf{r}_i = \begin{bmatrix} x_i \\ \boldsymbol{\epsilon}_i \end{bmatrix} \overset{\text{iid.}}{\sim} \mathrm{t}_{1+p}\left(\begin{pmatrix} \mu_x \\ \mathbf{0}_p \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \mathbf{0}_p^\top \\ \mathbf{0}_p & \boldsymbol{\Omega} \end{pmatrix}, \nu \right), \quad i = 1, \ldots, n. \tag{A.10}$$

By (A.2), this formulation implies

$$\begin{aligned} \begin{bmatrix} x_i \\ \boldsymbol{\epsilon}_i \end{bmatrix} \mid U_i = u_i \;\; &\sim \;\; \mathrm{N}_{1+p}\left(\begin{pmatrix} \mu_x \\ \mathbf{0}_p \end{pmatrix}, u_i^{-1} \begin{pmatrix} \sigma_x^2 & \mathbf{0}_p^\top \\ \mathbf{0}_p & \boldsymbol{\Omega} \end{pmatrix} \right), \\ U_i \;\; &\sim \;\; \mathrm{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2} \right), \end{aligned}$$

for $i = 1, \ldots, n$. Consequently,

$$x_i \mid U_i = u_i \overset{\text{ind.}}{\sim} \mathrm{N}(\mu_x, u_i^{-1}\sigma_x^2) \;\; \text{and,} \tag{A.11}$$

$$\boldsymbol{\epsilon}_i \mid U_i = u_i \overset{\text{ind.}}{\sim} \mathrm{N}_p(\mathbf{0}_p, u_i^{-1}\boldsymbol{\Omega}). \tag{A.12}$$

Besides this, $\boldsymbol{\epsilon}_i$ and x_i have Student-t marginal distributions, with $\boldsymbol{\epsilon}_i \sim \mathrm{t}_p(\mathbf{0}, \boldsymbol{\Omega}, \nu)$ and $x_i \sim \mathrm{t}(\mu_x, \sigma_x^2, \nu)$.

Since for each i, $\boldsymbol{\epsilon}_i$ and x_i are indexed by the same scale mixing factor U_i, they are not independent in general. The independence corresponds to the case where $U_i = 1$ (normal case). However, conditional on U_i, $\boldsymbol{\epsilon}_i$ and x_i are independent for each $i = 1, \ldots, n$, which implies that $\boldsymbol{\epsilon}_i$ and x_i are not correlated, since $\mathrm{Cov}(\boldsymbol{\epsilon}_i, x_i) = \mathrm{E}[\boldsymbol{\epsilon}_i x_i | U_i] = 0$. By (A.8), $\mathbf{Z}_i$ is an affine transformation of $\mathbf{r}_i$. Thus, its distribution is given by

$$\mathbf{Z}_i \sim \mathrm{t}_p(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z, \nu), \quad i = 1, \ldots, n, \tag{A.13}$$

where

$$\boldsymbol{\mu}_z = \mathbf{a} + \mathbf{b}\mu_x \quad \text{and} \quad \boldsymbol{\Sigma}_z = \sigma_x^2 \mathbf{b}\mathbf{b}^\top + \boldsymbol{\Omega}. \tag{A.14}$$

As mentioned earlier, our model considers censored observations. Following Matos et al. (2013b), we consider the case in which the response Z_{ij} is not fully observed for all i, j. What we truly observe, for each $i = 1, \ldots, n$, is the random vector $\mathbf{V}_i = (V_{i1}, \ldots, V_{ip})^\top$, such that $V_{ij} = \max\{Z_{ij}, \kappa_{ij}\}$, where κ_{ij} is a censoring level, that is,

$$V_{ij} = \begin{cases} Z_{ij} & \text{if} \quad Z_{ij} > \kappa_{ij} \\ \kappa_{ij} & \text{if} \quad Z_{ij} \leqslant \kappa_{ij}. \end{cases} \tag{A.15}$$

The model defined by Equations (A.6), (A.7) along with (A.10) and (A.15) is named *the Student-t Censored Measurement Error Model* – hereafter t-MEC model.

A.3.1 The likelihood function

In this section we present the likelihood function, which will be used in the model selection computations to compare fitted models.

First, let us partition $\mathbf{Z}_i$ into the observed and censored components, namely, $\mathbf{Z}_i = \mathrm{vec}(\mathbf{Z}_i^o, \mathbf{Z}_i^c)$, where $\mathbf{Z}_i^o : p_o \times 1$ corresponds to the former case, $\mathbf{Z}_i^c : p_c \times 1$ corresponds to the latter and $\mathrm{vec}(\cdot)$ denotes the function which stacks vectors or matrices of the same number of columns. Accordingly, let us consider $\mathbf{V}_i = \mathrm{vec}(\mathbf{V}_i^o, \mathbf{V}_i^c)$ and, recalling that $\mathbf{Z}_i \sim \mathrm{t}_p(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z, \nu)$, see (A.13), $\boldsymbol{\mu}_z = \mathrm{vec}(\boldsymbol{\mu}_z^o, \boldsymbol{\mu}_z^c)$ and $\boldsymbol{\Sigma}_z = \begin{pmatrix} \boldsymbol{\Sigma}_z^{oo} \boldsymbol{\Sigma}_z^{oc} \\ \boldsymbol{\Sigma}_z^{co} \boldsymbol{\Sigma}_z^{cc} \end{pmatrix}$. $\boldsymbol{\kappa}_i^c$ is the vector with the corresponding censoring levels for $\mathbf{Z}_i^c$. By Proposition A.1, we have

$$\mathbf{Z}_i^o \sim \mathrm{t}_{p_o}(\boldsymbol{\mu}_z^o, \boldsymbol{\Sigma}_z^{oo}, \nu) \quad \text{and} \quad \mathbf{Z}_i^c | \mathbf{Z}_i^o = \mathbf{Z}_i^o, \sim \mathrm{t}_{p_c}(\boldsymbol{\mu}_z^{co}, \mathbf{S}_z^{co}, \nu + p_o), \tag{A.16}$$

where

$$\boldsymbol{\mu}_z^{co} = \boldsymbol{\mu}_z^c + \boldsymbol{\Sigma}_z^{co}(\boldsymbol{\Sigma}_z^{oo})^{-1}(\mathbf{Z}_i^o - \boldsymbol{\mu}_z^o), \tag{A.17}$$

$$\mathbf{S}_z^{co} = \left(\frac{\nu + \mathrm{d}_{\boldsymbol{\Sigma}_z^{oo}}(\mathbf{Z}_i^o, \boldsymbol{\mu}_z^o)}{\nu + p^o} \right) \boldsymbol{\Sigma}_z^{cc.o}, \tag{A.18}$$

$$\boldsymbol{\Sigma}_z^{cc.o} = \boldsymbol{\Sigma}_z^{cc} - \boldsymbol{\Sigma}_z^{co} \boldsymbol{\Sigma}_z^{oo-1} \boldsymbol{\Sigma}_z^{oc}. \tag{A.19}$$

The observed sample for the experimental unit i is $\{\mathbf{Z}_i^o, \boldsymbol{\kappa}_i^c\}$. The associated likelihood is

$$L_i(\boldsymbol{\theta}) = P(\mathbf{V}_i^c = \boldsymbol{\kappa}_i^c | \mathbf{Z}_i^o = \mathbf{Z}_i^o) f(\mathbf{Z}_i^o),$$

where $f(\cdot)$ is the marginal density of $\mathbf{Z}_i^o$. But $\mathbf{V}_i^c = \boldsymbol{\kappa}_i^c$ if and only if $\mathbf{Z}_i^c \leqslant \boldsymbol{\kappa}_i^c$. By (A.16), we obtain

$$L_i(\boldsymbol{\theta}) = \mathrm{T}_{p_c}(\boldsymbol{\kappa}_i^c | \boldsymbol{\mu}_z^{co}, \mathbf{S}_z^{co}, \nu + p_o) \mathrm{t}_{p_o}(\mathbf{Z}_i^o | \boldsymbol{\mu}_z^o, \boldsymbol{\Sigma}_z^{oo}, \nu).$$

The log-likelihood associated with the whole sample is

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log L_i(\boldsymbol{\theta}). \tag{A.20}$$

A.4 The ECM algorithm

In this section, we describe how the t-MEC model can be fitted by using the ECM algorithm (Meng and Rubin (1993)). This algorithm considers a simple modification of the traditional EM algorithm initially proposed by Dempster et al. (1977) and is an efficient tool to obtain the maximum likelihood estimates under a missing data framework.

The t-MEC model can be formulated in a flexible hierarchical representation that is useful for theoretical derivations. It is easily obtained through Equations (A.8), (A.11) and (A.12) and is given by

$$\mathbf{Z}_i \mid x_i, U_i = u_i \overset{\text{ind.}}{\sim} \mathrm{N}_p(\mathbf{a} + \mathbf{b}x_i, u_i^{-1}\boldsymbol{\Omega}), \tag{A.21}$$

$$x_i \mid U_i = u_i \overset{\text{ind.}}{\sim} \mathrm{N}(\mu_x, u_i^{-1}\sigma_x^2), \tag{A.22}$$

$$U_i \overset{\text{iid.}}{\sim} \mathrm{Gamma}\left(\nu/2, \nu/2\right), \ i = 1, \ldots, n. \tag{A.23}$$

Following the suggestions of Lange et al. (1989) and Lucas (1997), who pointed out difficulties in estimating ν due to problems of unbounded and local maxima in the likelihood function, we consider the value of ν to be known.

Now, we enunciate two important results that will be useful in the E step of the EM algorithm.

Proposição A.4. *Consider the hierarchical representation of the t-MEC model given in (A.21)–(A.23). Then,*

$$x_i | U_i = u_i, \mathbf{Z}_i = \mathbf{Z}_i \sim \mathrm{N}\left(\frac{\mu_x + \sigma_x^2 \mathbf{b}'\boldsymbol{\Omega}^{-1}(\mathbf{Z}_i - \mathbf{a})}{1 + \sigma_x^2 \mathbf{b}'\boldsymbol{\Omega}^{-1}\mathbf{b}}, \frac{\sigma_x^2}{u_i(1 + \sigma_x^2 \mathbf{b}'\boldsymbol{\Omega}^{-1}\mathbf{b})}\right).$$

The proof follows directly from the relation $f(x_i|u_i, \mathbf{Z}_i) \propto f(\mathbf{Z}_i|x_i, u_i) f(x_i|u_i)$, where $f(\cdot)$ denotes a generic *pdf*.

Proposição A.5. *For the t-MEC model,*

$$\mathrm{E}(U_i | \mathbf{Z}_i = \mathbf{Z}_i) = \frac{p + \nu}{\mathrm{d}_{\boldsymbol{\Sigma}_z}(\mathbf{Z}_i, \boldsymbol{\mu}_z) + \nu}.$$

To prove this result, recall that $\mathbf{Z}_i \sim \mathrm{t}_p(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z, \nu)$, which implies $\mathbf{Z}_i|U_i = u_i \sim \mathrm{N}_p(\boldsymbol{\mu}_z, u_i^{-1}\boldsymbol{\Sigma}_z)$ and $U_i \sim \mathrm{Gamma}(\nu/2, \nu/2)$ – see (A.2). Using the relation

$$f(u_i|\mathbf{Z}_i) \propto f(\mathbf{Z}_i|u_i) f(u_i),$$

we can prove that $U_i|\mathbf{Z}_i = \mathbf{Z}_i \sim \mathrm{Gamma}\left(\frac{p+\nu}{2}, \frac{1}{2}\left(\mathrm{d}_{\boldsymbol{\Sigma}_z}(\mathbf{Z}_i, \boldsymbol{\mu}_z) + \nu\right)\right)$, and the result follows.

A.4.1 The E Step

Let $\mathbf{Z} = (\mathbf{Z}_1^\top, \ldots, \mathbf{Z}_n^\top)^\top$, $\mathbf{X} = (x_1, \ldots, x_n)^\top$ and $\mathbf{u} = (u_1, \ldots, u_n)^\top$. Let $\boldsymbol{\theta}$ be the vector with all the parameters in the model. Apart from constants which do not depend on $\boldsymbol{\theta}$, the complete log-likelihood associated with the complete data $\mathbf{Z}_c = \{\mathbf{Z}, \mathbf{x}, \mathbf{u}\}$ is given by

$$\begin{aligned}\ell_c(\boldsymbol{\theta}|\mathbf{Z}_c) = &-\frac{n}{2}\sum_{i=1}^{p}\log\omega_j^2 - \frac{1}{2}\sum_{i=1}^{n} u_i(\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i)^\top \boldsymbol{\Omega}^{-1}(\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i) \\ &- \frac{n}{2}\log\sigma_x^2 - \frac{1}{2\sigma_x^2}\sum_{i=1}^{n} u_i(x_i - \mu_x)^2.\end{aligned}$$

Suppose that at the kth stage of the algorithm we obtain an estimate $\widehat{\boldsymbol{\theta}}^{(k)}$ of $\boldsymbol{\theta}$. The E step consists of the computation of the conditional expectation

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \mathrm{E}_{\widehat{\boldsymbol{\theta}}^{(k)}}\left[\ell_c(\boldsymbol{\theta}|\mathbf{Z}_c)|\mathbf{V}\right],$$

where $\mathrm{E}_{\widehat{\boldsymbol{\theta}}^{(k)}}$ means that the expectation is being affected using $\widehat{\boldsymbol{\theta}}^{(k)}$ as the true parameter value and $\mathbf{V} = (\mathbf{V}_1^\top, \ldots, \mathbf{V}_n^\top)^\top$. The M step consists of maximizing $Q(\cdot|\widehat{\boldsymbol{\theta}}^{(k)})$ in $\boldsymbol{\theta}$. To do so, first observe that the function $Q(\cdot|\widehat{\boldsymbol{\theta}}^{(k)})$ can be decomposed into

$$Q\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}\right) = Q_1\left(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}^{(k)}\right) + Q_2\left(\mu_x, \sigma_x^2|\widehat{\boldsymbol{\theta}}^{(k)}\right), \tag{A.24}$$

where $\boldsymbol{\omega} = (\omega_1^2, \ldots, \omega_p^2)^\top$,

$$\begin{aligned}&Q_1\left(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}^{(k)}\right) = \\ &\quad \mathrm{E}_{\widehat{\boldsymbol{\theta}}^{(k)}}\left[-\frac{n}{2}\sum_{i=1}^{p}\log\omega_j^2 - \frac{1}{2}\sum_{i=1}^{n} u_i(\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i)^\top \boldsymbol{\Omega}^{-1}(\mathbf{Z}_i - \mathbf{a} - \mathbf{b}x_i)|\mathbf{V}\right] \quad \text{and}\end{aligned} \tag{A.25}$$

$$Q_2\left(\mu_x, \sigma_x^2|\widehat{\boldsymbol{\theta}}^{(k)}\right) = \mathrm{E}_{\widehat{\boldsymbol{\theta}}^{(k)}}\left[-\frac{n}{2}\log\sigma_x^2 - \frac{1}{2\sigma_x^2}\sum_{i=1}^{n} u_i(x_i - \mu_x)^2|\mathbf{V}, \right].$$

Given this decomposition, we can reduce the problem to the maximization of two independent functions, searching for critical points of $Q_1(\cdot|\widehat{\boldsymbol{\theta}}^{(k)})$ and $Q_2(\cdot|\widehat{\boldsymbol{\theta}}^{(k)})$ separately.

Expanding the expressions of $Q_1(\cdot|\widehat{\boldsymbol{\theta}}^{(k)})$ and $Q_2(\cdot|\widehat{\boldsymbol{\theta}}^{(k)})$ and taking expectations, it follows that

$$Q_1\left(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}^{(k)}\right) = -\frac{n}{2}\sum_{i=1}^{p}\log\omega_j^2 - \frac{1}{2}\left\{\sum_{i=1}^{n}\left(\mathrm{tr}\{\boldsymbol{\Omega}^{-1}\widehat{u\mathbf{Z}_i^2}\} - 2\mathbf{a}^\top\boldsymbol{\Omega}^{-1}\widehat{u\mathbf{Z}_i}\right.\right.$$
$$\left.\left. -2\widehat{ux\mathbf{Z}_i}\boldsymbol{\Omega}^{-1}\mathbf{b} + \mathbf{a}^\top\boldsymbol{\Omega}^{-1}\mathbf{a}\widehat{u}_i + 2\mathbf{a}^\top\boldsymbol{\Omega}^{-1}\mathbf{b}\widehat{ux}_i + \mathbf{b}^\top\boldsymbol{\Omega}^{-1}\mathbf{b}\widehat{ux_i^2}\right)\right\},$$
$$Q_2(\mu_x,\sigma_x^2|\widehat{\boldsymbol{\theta}}^{(k)}) = -\frac{n}{2}\log\sigma_x^2 - \frac{1}{2\sigma_x^2}\left\{\sum_{i=1}^{n}\left(\widehat{ux_i^2} - 2\mu_x\widehat{ux}_i + \mu_x^2\widehat{u}_i\right)\right\},$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix,

$$\widehat{u\mathbf{Z}_i^2} = \mathrm{E}[U_i\mathbf{Z}_i\mathbf{Z}_i^\top|\mathbf{V}_i],\quad \widehat{u\mathbf{Z}_i} = \mathrm{E}[U_i\mathbf{Z}_i|\mathbf{V}_i],$$
$$\widehat{u}_i = \mathrm{E}[U_i|\mathbf{V}_i],\qquad \widehat{ux\mathbf{Z}_i} = \mathrm{E}[U_ix_i\mathbf{Z}_i^\top|\mathbf{V}_i],$$
$$\widehat{ux}_i = \mathrm{E}[U_ix_i|\mathbf{V}_i],\quad \widehat{ux_i^2} = \mathrm{E}[U_ix_i^2|\mathbf{V}_i],$$

and we have omitted the subscript $\widehat{\boldsymbol{\theta}}^{(k)}$ to simplify the notation. To obtain expressions for these expectations, we will use a result from probability theory called *the tower property of conditional expectation*: if $\mathbf{X}$ and$\mathbf{Y}$ are arbitrary random vectors and $f(\cdot)$ is a measurable function, then $\mathrm{E}[\mathrm{E}(\mathbf{X}|\mathbf{Y})|f(\mathbf{Y})] = \mathrm{E}[\mathbf{X}|f(\mathbf{Y})]$. For a proof, see (Ash, 2000, Theo. 5.5.10). Now, observe that, by (A.15), $\mathbf{V}_i$ is a function of $\mathbf{Z}_i$. Then, by this property, we can write

$$\widehat{u\mathbf{Z}_i^2} = \mathrm{E}\{\mathrm{E}[U_i\mathbf{Z}_i\mathbf{Z}_i^\top|\mathbf{Z}_i]|\mathbf{V}_i]\},\quad \widehat{u\mathbf{Z}_i} = \mathrm{E}\{\mathrm{E}[U_i\mathbf{Z}_i|\mathbf{Z}_i]|\mathbf{V}_i]\},\quad \text{and}\quad \widehat{u}_i = \mathrm{E}\{\mathrm{E}[U_i|\mathbf{Z}_i]|\mathbf{V}_i\}. \tag{A.26}$$

Proposition A.5 gives the conditional expectation $\mathrm{E}[U_i|\mathbf{Z}_i]$ and, from this result and formulas (A.26) we obtain the following expressions for $\widehat{u}_i$, $\widehat{u\mathbf{Z}_i}$ and $\widehat{u\mathbf{Z}_i^2}$ (as we will see soon, all expectations involved in the E step are written as functions of these), considering three different cases:

(i) Individual i does not have censored components. In this case, $\mathbf{V}_i = \mathbf{Z}_i$ – see Equation (A.15) –. Thus,

$$\widehat{u}_i = \frac{p+\nu}{\mathrm{d}_{\boldsymbol{\Sigma}_z}(\mathbf{Z}_i,\boldsymbol{\mu}_z)+\nu},\quad \widehat{u\mathbf{Z}_i} = \frac{p+\nu}{\mathrm{d}_{\boldsymbol{\Sigma}_z}(\mathbf{Z}_i,\boldsymbol{\mu}_z)+\nu}\mathbf{Z}_i,\quad \text{and}\quad \widehat{u\mathbf{Z}_i^2} = \frac{p+\nu}{\mathrm{d}_{\boldsymbol{\Sigma}_z}(\mathbf{Z}_i,\boldsymbol{\mu}_z)+\nu}\mathbf{Z}_i\mathbf{Z}_i^\top.$$

(ii) Individual i has only censored components. By Equation (A.15), this fact occurs if and only if $\mathbf{Z}_i \leqslant \boldsymbol{\kappa}_i$, where $\boldsymbol{\kappa}_i$ is the vector with the censoring levels for individual i. Thus,

$$\widehat{u}_i = \mathrm{E}\{\mathrm{E}[U_i|\mathbf{Z}_i]|\mathbf{Z}_i \leqslant \boldsymbol{\kappa}_i\} = \mathrm{E}\left[\frac{p+\nu}{\mathrm{d}_{\boldsymbol{\Sigma}_z}(\mathbf{Z}_i,\boldsymbol{\mu}_z)+\nu}\Big|\mathbf{Z}_i \leqslant \boldsymbol{\kappa}_i\right].$$

By (A.13) and the definition of a truncated Student-t distribution, we have that $\mathbf{Z}_i|(\mathbf{Z}_i \leqslant \boldsymbol{\kappa}_i) \sim \mathrm{Tt}_p(\boldsymbol{\mu}_z,\boldsymbol{\Sigma}_z,\nu;\mathbb{D}_i)$, where $\mathbb{D}_i$ is like in (A.3) with $\mathbf{d} = \boldsymbol{\kappa}_i$. Using $r = 1$ and $k = 0$ in Proposition A.2, we get

$$\widehat{u}_i = \frac{\mathrm{T}_p(\boldsymbol{\kappa}_i|\boldsymbol{\mu}_z,\boldsymbol{\Sigma}_z^*,\nu+2r)}{\mathrm{T}_p(\boldsymbol{\kappa}_i|\boldsymbol{\mu}_z,\boldsymbol{\Sigma}_z,\nu)},$$

where $\boldsymbol{\Sigma}_z^* = (\nu/(\nu+2))\boldsymbol{\Sigma}_z$. Using $r=1$ and $k=1$ in Proposition A.2, we obtain

$$\widehat{u\mathbf{Z}_i} = \frac{\mathrm{T}_p(\boldsymbol{\kappa}_i|\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z^*, \nu+2r)}{\mathrm{T}_p(\boldsymbol{\kappa}_i|\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z, \nu)}\mathrm{E}[\mathbf{Y}_i],$$

where $\mathbf{Y}_i \sim \mathrm{Tt}_p(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z^*, \nu+2; \mathbb{D}_i)$. Finally, $r=1$ and $k=2$ in Proposition A.2 imply

$$\widehat{u\mathbf{Z}_i^2} = \frac{\mathrm{T}_p(\boldsymbol{\kappa}_i|\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z^*, \nu+2r)}{\mathrm{T}_p(\boldsymbol{\kappa}_i|\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z, \nu)}\mathrm{E}[\mathbf{Y}_i\mathbf{Y}_i^\top].$$

(iii) Individual i has censored and uncensored components. As we commented before in Section A.3.1, in this case, we decompose the vector $\mathbf{V}_i$ into two subvectors, $\mathbf{Z}_i^o$ and $\boldsymbol{\kappa}_i^c$, corresponding to the uncensored observations and the censoring levels, respectively. Accordingly, we partition the vector $\mathbf{Z}_i$ as $\mathbf{Z}_i = \mathrm{vec}(\mathbf{Z}_i^o, \mathbf{Z}_i^c)$. The components are censored if and only if $\mathbf{Z}_i^c \leqslant \boldsymbol{\kappa}_i^c$. Thus,

$$\widehat{u}_i = \mathrm{E}\{\mathrm{E}[U_i|\mathbf{Z}_i]|\mathbf{Z}_i^o = \mathbf{Z}_i^o, \mathbf{Z}_i^c \leqslant \boldsymbol{\kappa}_i^c\} = \mathrm{E}\left[\frac{p+\nu}{\mathrm{d}_{\boldsymbol{\Sigma}_z}(\mathbf{Z}_i, \boldsymbol{\mu}_z)+\nu}\Big|\mathbf{Z}_i^o = \mathbf{Z}_i^o, \mathbf{Z}_i^c \leqslant \boldsymbol{\kappa}_i^c\right].$$

In this case, we have that $\mathbf{Z}_i|(\mathbf{Z}_i^c \leqslant \boldsymbol{\kappa}_i^c) \sim \mathrm{Tt}_p(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z, \nu; \mathbb{D}_i^c)$, with

$$\mathbb{D}_i^c = \{(x_1, \ldots, x_p) \in \mathbb{R}^p;\ \ x_i \leqslant \boldsymbol{\kappa}_i^c,\ \ i \in \mathcal{C}\}, \tag{A.27}$$

where $\mathcal{C}$ is the set of indices for the censored components. Consequently, we make $d_i = +\infty$ for $i \notin \mathcal{C}$ in (A.3). Thus, $\widehat{u}_i$ can be calculated using Proposition A.3, with $\mathbf{Z}_i^o$ and $\mathbf{Z}_i^c$ playing the role of $\mathbf{Z}_1$ and $\mathbf{Z}_2$, respectively, taking $r=1$ and $k=0$. Then, we get

$$\widehat{u}_i = \frac{p^o+\nu}{\nu+\mathrm{d}_{\boldsymbol{\Sigma}_z^{oo}}(\mathbf{Z}_i^o, \boldsymbol{\mu}_z^o)}\frac{\mathrm{T}_{p^c}(\boldsymbol{\kappa}_i^c|\boldsymbol{\mu}_z^{co}, \widetilde{\mathbf{S}}_z^{co}, \nu+p_o+2)}{\mathrm{T}_{p^c}(\boldsymbol{\kappa}_i^c|\boldsymbol{\mu}_z^{co}, \mathbf{S}_z^{co}, \nu+p_o)},$$

where p^o and p^c are the dimensions of the vectors $\mathbf{Z}_i^o$ and $\mathbf{Z}_i^c$, respectively, $\nu+p^o+2 > 0$,

$$\widetilde{\mathbf{S}}_z^{co} = \frac{\nu+\mathrm{d}_{\boldsymbol{\Sigma}_z^{oo}}(\mathbf{Z}_i^o, \boldsymbol{\mu}_z^o)}{\nu+p^o+2}\boldsymbol{\Sigma}_z^{cc.o},$$

$\boldsymbol{\mu}_z^o$, $\mathbf{S}_z^{co}$ and $\boldsymbol{\Sigma}_z^{cc.o}$ are given in (A.17), (A.18) and (A.19), respectively. Regarding $\widehat{u\mathbf{Z}_i}$, we have that

$$\begin{aligned}
\widehat{u\mathbf{Z}_i} &= \mathrm{E}\left[\frac{p+\nu}{\mathrm{d}_{\boldsymbol{\Sigma}_z}(\mathbf{Z}_i, \boldsymbol{\mu}_z)+\nu}\mathrm{vec}(\mathbf{Z}_i^o, \mathbf{Z}_i^c)\Big|\mathbf{Z}_i^o = \mathbf{Z}_i^o, \mathbf{Z}_i^c \leqslant \boldsymbol{\kappa}_i^c\right] \\
&= \mathrm{vec}\left(\mathrm{E}\left[\frac{p+\nu}{\mathrm{d}_{\boldsymbol{\Sigma}_z}(\mathbf{Z}_i, \boldsymbol{\mu}_z)+\nu}\mathbf{Z}_i^o\Big|\mathbf{Z}_i^o = \mathbf{Z}_i^o, \mathbf{Z}_i^c \leqslant \boldsymbol{\kappa}_i^c\right],\right. \\
&\qquad \left.\mathrm{E}\left[\frac{p+\nu}{\mathrm{d}_{\boldsymbol{\Sigma}_z}(\mathbf{Z}_i, \boldsymbol{\mu}_z)+\nu}\mathbf{Z}_i^c\Big|\mathbf{Z}_i^o = \mathbf{Z}_i^o, \mathbf{Z}_i^c \leqslant \boldsymbol{\kappa}_i^c\right]\right) \\
&= \mathrm{vec}(\widehat{u}_i\mathbf{Z}_i^o, \mathrm{E}[\mathbf{Y}_i]),
\end{aligned}$$

where

$$\mathbf{Y}_i \sim\ \mathrm{Tt}_{p^c}(\boldsymbol{\mu}_z^{co}, \widetilde{\mathbf{S}}_z^{co}, \nu+p^o+2; \mathbb{D}_i^c). \tag{A.28}$$

Finally, to compute $\widehat{u\mathbf{Z}_i^2}$, observe that

$$\begin{aligned}\widehat{u\mathbf{Z}_i^2} &= \mathrm{E}\left[\frac{p+\nu}{\mathrm{d}_{\boldsymbol{\Sigma}_z}(\mathbf{Z}_i,\boldsymbol{\mu}_z)+\nu}\begin{pmatrix}\mathbf{Z}_i^o\mathbf{Z}_i^{o\top} & \mathbf{Z}_i^o\mathbf{Z}_i^{c\top}\\ \mathbf{Z}_i^c\mathbf{Z}_i^{o\top} & \mathbf{Z}_i^c\mathbf{Z}_i^{c\top}\end{pmatrix}\Big|\mathbf{Z}_i^o=\mathbf{Z}_i^o,\mathbf{Z}_i^c\leqslant\boldsymbol{\kappa}_i^c\right]\\ &= \begin{pmatrix}\widehat{u}_i\mathbf{Z}_i^o\mathbf{Z}_i^{o\top} & \widehat{u}_i\mathbf{Z}_i^o\mathrm{E}[\mathbf{Y}_i]^\top\\ \widehat{u}_i\mathrm{E}[\mathbf{Y}_i]\mathbf{Z}_i^{o\top} & \widehat{u}_i\mathrm{E}[\mathbf{Y}_i\mathbf{Y}_i^\top]\end{pmatrix},\end{aligned}$$

where $\mathbf{Y}_i$ is as in (A.28).

Regarding the remaining expectations, we have

$$\begin{aligned}\mathrm{E}[x_iU_i|\mathbf{V}_i=\mathbf{v}_i] &= \iint x_iu_i\pi(x_i,u_i|\mathbf{v}_i)\,dx_i\,du_i\\ &= \int x_i\pi(x_i|u_i,\mathbf{v}_i)\,dx_i\int u_i\pi(u_i|\mathbf{v}_i)du_i\\ &= \mathrm{E}[x_i|U_i=u_i,\mathbf{V}_i=\mathbf{v}_i]\mathrm{E}[U_i|\mathbf{V}_i=\mathbf{v}_i].\end{aligned}\tag{A.29}$$

By the tower property, we have

$$\mathrm{E}[x_i|U_i,\mathbf{V}_i]=\mathrm{E}[\mathrm{E}(x_i|U_i,\mathbf{Z}_i)|U_i,\mathbf{V}_i].$$

Consequently,

$$\begin{aligned}\widehat{ux}_i=\mathrm{E}[x_iU_i|\mathbf{V}_i] &= \mathrm{E}\left[\frac{\mu_x+\sigma_x^2\mathbf{b}'\boldsymbol{\Omega}^{-1}(\mathbf{Z}_i-\mathbf{a})}{1+\sigma_x^2\mathbf{b}'\boldsymbol{\Omega}^{-1}\mathbf{b}}|U_i,\mathbf{V}_i\right]\mathrm{E}[U_i|\mathbf{V}_i]\\ &= \frac{\mu_x\mathrm{E}[U_i|\mathbf{V}_i]+\sigma_x^2\mathbf{b}'\boldsymbol{\Omega}^{-1}\mathrm{E}[\mathbf{Z}_i|U_i,\mathbf{V}_i]\mathrm{E}[U_i|\mathbf{V}_i]-\mathbf{a}\mathrm{E}[U_i|\mathbf{V}_i]}{1+\sigma_x^2\mathbf{b}'\boldsymbol{\Omega}^{-1}\mathbf{b}}\\ &= \frac{\mu_x\widehat{u}_i+\sigma_x^2\mathbf{b}'\boldsymbol{\Omega}^{-1}\widehat{u\mathbf{Z}_i}-\sigma_x^2\mathbf{b}'\boldsymbol{\Omega}^{-1}\mathbf{a}\widehat{u}_i}{1+\sigma_x^2\mathbf{b}'\boldsymbol{\Omega}^{-1}\mathbf{b}}\\ &= \mu_x\widehat{u}_i+\boldsymbol{\varphi}(\widehat{u\mathbf{Z}_i}-\boldsymbol{\mu}_z\widehat{u}_i),\end{aligned}\tag{A.30}$$

where

$$\boldsymbol{\varphi}=\frac{\sigma_x^2\mathbf{b}^\top\boldsymbol{\Omega}^{-1}}{1+\sigma_x^2\mathbf{b}^\top\boldsymbol{\Omega}^{-1}\mathbf{b}}\tag{A.31}$$

and the equality in (A.30) is obtained by proving that $\mathrm{E}[\mathbf{Z}_i|U_i,\mathbf{V}_i]\mathrm{E}[U_i|\mathbf{V}_i]=\mathrm{E}[U_i\mathbf{Z}_i|\mathbf{V}_i]\equiv\widehat{u\mathbf{Z}_i}$, which can be done following the same paths that led to (A.29), replacing x_i with $\mathbf{Z}_i$.

In a similar fashion, we get

$$\widehat{ux_i^2}=\Lambda+\mu_x^2\widehat{u}_i+2\boldsymbol{\varphi}\left[\widehat{u\mathbf{z}_i}-\boldsymbol{\mu}_z\widehat{u}_i\right]+\boldsymbol{\varphi}\left[\widehat{u\mathbf{z}_i^2}-\widehat{u\mathbf{z}_i}\boldsymbol{\mu}_z^\top-\boldsymbol{\mu}_z\widehat{u\mathbf{z}_i}^\top+\boldsymbol{\mu}_z\boldsymbol{\mu}_z^\top\widehat{u}_i\right]\boldsymbol{\varphi}^\top,\ \text{and}$$

$$\widehat{ux\mathbf{z}_i}=\mu_x\widehat{u\mathbf{z}_i}+\boldsymbol{\varphi}\left[\widehat{u\mathbf{z}_i^2}-\boldsymbol{\mu}_z\widehat{u\mathbf{z}_i}\right],$$

with

$$\Lambda=\frac{\sigma_x^2}{1+\sigma_x^2\mathbf{b}^\top\boldsymbol{\Omega}^{-1}\mathbf{b}}.\tag{A.32}$$

A.4.2 The CM Step

Given the current estimate $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$ at the *kth* stage, the CM-step of the ECM algorithm (Meng and Rubin (1993)) consists of the *conditional maximization* of the Q function given in (A.24). More precisely, ECM replaces each M-step of the EM algorithm of Dempster et al. (1977) by a sequence of S conditional maximization steps, called CM-steps, each of which maximizes the Q function over $\boldsymbol{\theta}$ but with some vector function of $\boldsymbol{\theta}$, $(g_1(\boldsymbol{\theta}), \ldots, g_S(\boldsymbol{\theta}))$ say, fixed at its previous value. In our case, for example, we first maximize conditionally the function $Q_1\left(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\omega} | \widehat{\boldsymbol{\theta}}^{(k)}\right)$ in (A.25) over $\boldsymbol{\alpha}$ fixing the values $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}^{(k)}$ and $\boldsymbol{\omega} = \widehat{\boldsymbol{\omega}}^{(k)}$. Then we maximize $Q_1\left(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\omega} | \widehat{\boldsymbol{\theta}}^{(k)}\right)$ over $\boldsymbol{\beta}$ fixing the values $\boldsymbol{\alpha} = \widehat{\boldsymbol{\alpha}}^{(k+1)}$ and $\boldsymbol{\omega} = \widehat{\boldsymbol{\omega}}^{(k)}$ and so on. We get the following closed expressions:

$$
\begin{aligned}
\widehat{\boldsymbol{\alpha}}^{(k+1)} &= \overline{\mathbf{z}}_u^{(k)} - \overline{x}_u^{(k)} \widehat{\boldsymbol{\beta}}^{(k)}, \\
\widehat{\boldsymbol{\beta}}^{(k+1)} &= \frac{n\overline{\widehat{u}^{(k)}} \sum_{i=1}^n \widehat{ux\mathbf{z}_i}^{\star(k)} - \sum_{i=1}^n \widehat{u\mathbf{z}_i}^{\star(k)} \sum_{i=1}^n \widehat{ux_i}^{(k)}}{n\overline{\widehat{u}^{(k)}} \sum_{i=1}^n \widehat{ux_i^2}^{(k)} - \left(\sum_{i=1}^n \widehat{ux_i}^{(k)}\right)^2}, \\
\widehat{\omega_1^2}^{(k+1)} &= \frac{1}{n} \sum_{i=1}^n (\widehat{u\mathbf{z}_i^2}_{11}^{(k)} - 2\widehat{ux\mathbf{z}_i}_1^{(k)} + \widehat{ux_i^2}^{(k)}), \\
\widehat{\omega_{j+1}^2}^{(k+1)} &= \frac{1}{n} \sum_{i=1}^n \left(\widehat{u\mathbf{z}_i^2}_{(j+1)(j+1)}^{(k)} + \widehat{u}_i^{(k)} \widehat{\alpha}_j^{2(k+1)} + \widehat{ux_i^2}^{(k)} \widehat{\beta}_j^{2(k+1)} + 2\widehat{ux_i}^{(k)} \widehat{\alpha}_j^{(k+1)} \widehat{\beta}_j^{(k+1)} \right. \\
&\quad \left. -2\widehat{ux\mathbf{z}_i}_{(j+1)}^{(k)} \widehat{\beta}_j^{(k+1)} - 2\widehat{u\mathbf{z}_i}_{(j+1)}^{(k)} \widehat{\alpha}_j^{(k+1)} \right), \; j = 1, \ldots, r, \\
\widehat{\mu}_x^{(k+1)} &= \overline{x}_u^{(k)}, \\
\widehat{\sigma}_x^{2(k+1)} &= \frac{1}{n} \sum_{i=1}^n (\widehat{ux_i^2}^{(k)} - 2\widehat{ux_i}^{(k)} \widehat{\mu}_x^{(k+1)} + \widehat{ux_i}^{(k)} \widehat{\mu}_x^{2(k+1)}),
\end{aligned}
$$

where $\overline{\mathbf{z}}_u^{(k)} = \frac{\sum_{i=1}^n \widehat{u\mathbf{z}_i}^{\star(k)}}{\sum_{i=1}^n \widehat{u}_i^{(k)}}, \overline{x}_u^{(k)} = \frac{\sum_{i=1}^n \widehat{ux_i}^{(k)}}{\sum_{i=1}^n \widehat{u}_i^{(k)}}$ and $\overline{\widehat{u}^{(k)}} = \frac{1}{n} \sum_{i=1}^n \widehat{u}_i^{(k)}$, with $\widehat{u\mathbf{z}_i}^{\star(k)} = (\widehat{u\mathbf{z}}_{i2}, \ldots, \widehat{u\mathbf{z}}_{ip})^\top$ and $\widehat{ux\mathbf{z}_i}^{\star(k)} = (\widehat{ux\mathbf{z}}_{i2}, \ldots, \widehat{ux\mathbf{z}}_{ip})^\top$.

A.4.3 Imputation of censored components

Let $\mathbf{Z}_i^{(c)}$ be the true (partially or completely unobserved) response vector for the censored components of the ith unit. We define a predictor for $\mathbf{Z}_i^{(c)}$ as

$$\widetilde{\mathbf{Z}}_i^{(c)} = \mathrm{E}[\mathbf{Z}_i | \mathbf{V}_i = \mathbf{v}_i].$$

We have the following particular cases:

1. If unit i has only censored components then, if we make $r = 0$ and $k = 1$ in Proposition A.2, we get

$$\widetilde{\mathbf{Z}}_i^{(c)} = \mathrm{E}[\mathbf{Y}_i], \quad \text{with} \quad \mathbf{Y}_i \sim \mathrm{Tt}_p(\widehat{\boldsymbol{\mu}}_z, \widehat{\boldsymbol{\Sigma}}_z, \nu; \mathbb{D}_i),$$

$\widehat{\boldsymbol{\mu}}_z$ and $\widehat{\boldsymbol{\Sigma}}_z$ are the EM estimates of $\boldsymbol{\mu}_z$ and $\boldsymbol{\Sigma}_z$, respectively, and $\mathbb{D}_i$ is as in (A.3) with $\mathbf{d} = \boldsymbol{\kappa}_i$, where $\boldsymbol{\kappa}_i$ is the vector with censoring levels for unit i.

2. Unit i has uncensored and censored components. In this case, we partition the vector $\mathbf{Z}_i$ as $\mathbf{Z}_i = \text{vec}(\mathbf{Z}_i^o, \mathbf{Z}_i^c)$. Components are censored if and only if $\mathbf{Z}_i^c \leqslant \boldsymbol{\kappa}_i^c$, such that

$$\widetilde{\mathbf{Z}}_i^{(c)} = \text{E}[\text{vec}(\mathbf{Z}_i^o, \mathbf{Z}_i^c) | \mathbf{Z}_i^o = \mathbf{Z}_i^o, \mathbf{Z}_i^c \leqslant \boldsymbol{\kappa}_i^c] = \text{vec}(\mathbf{Z}_i^o, \widehat{\mathbf{y}}_i^c),$$

where, by Proposition A.3 with $r = 0$ and $k = 1$,

$$\widehat{\mathbf{y}}_i^c = \text{E}[\mathbf{Y}_i], \quad \text{with} \quad \mathbf{Y}_i \sim \text{Tt}_{p^c}(\boldsymbol{\mu}_z^{co}, \mathbf{S}_z^{co}, \nu + p^o; \mathbb{D}_i^c), \tag{A.33}$$

where $\boldsymbol{\mu}_z^{co}$ and $\mathbf{S}_z^{co}$ are given in (A.17) and (A.18), respectively, and $\mathbb{D}_i^c$ is given in (A.27).

A.4.4 Estimation of x_i

Following Lin and Lee (2006), Ho et al. (2012) and recently Castro et al. (2015), we consider the conditional mean to estimate the unobserved latent covariate. Using the tower property and Proposition A.4, we have that an estimator for x_i can be obtained through

$$\begin{aligned}
\widehat{x}_i &= \text{E}[x_i | \mathbf{V}_i] = \text{E}[\text{E}(x_i | U_i, \mathbf{Z}_i) | \mathbf{V}_i] \\
&= \text{E}\left[\frac{\mu_x + \sigma_x^2 \mathbf{b}' \boldsymbol{\Omega}^{-1} (\mathbf{Z}_i - \mathbf{a})}{1 + \sigma_x^2 \mathbf{b}' \boldsymbol{\Omega}^{-1} \mathbf{b}} \Big| \mathbf{V}_i \right] \\
&= \mu_x + \boldsymbol{\varphi}(\widehat{\mathbf{Z}}_i - \mathbf{a} - \mathbf{b}\mu_x),
\end{aligned} \tag{A.34}$$

where $\boldsymbol{\varphi}$ is given in (A.31) and $\widehat{\mathbf{Z}}_i = \text{E}[\mathbf{Z}_i | \mathbf{V}_i]$. Observe that, if individual i does not have censored components, then $\widehat{\mathbf{Z}}_i$ is the first moment of a $\text{t}_p(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z, \nu)$ distribution. If all its components are censored, then $\text{E}[\mathbf{Z}_i | \mathbf{V}_i] = \text{E}[\mathbf{Z}_i | \mathbf{Z}_i \leqslant \boldsymbol{\kappa}_i]$, which can be computed using Proposition A.2 with $r = 0$ and $k = 1$. Finally, if it has censored and uncensored components, then $\text{E}[\mathbf{Z}_i | \mathbf{V}_i] = \text{vec}(\mathbf{Z}_i^o, \widehat{\mathbf{y}}_i^c)$, see (A.33). The parameter values in (A.34) must be replaced with the respective EM estimates.

Moreover, the conditional covariance matrix of x_i given $\mathbf{V}_i$ is

$$\text{Var}[x_i | \mathbf{V}_i] = \text{E}[x_i^2 | \mathbf{V}_i] - (\text{E}[x_i | \mathbf{V}_i])^2.$$

By the tower property and Proposition A.4, we have

$$\begin{aligned}
\text{E}[x_i^2 | \mathbf{V}_i] &= \text{E}\{\text{E}[\text{E}(x_i^2 | U_i, \mathbf{Z}_i) | \mathbf{Z}_i] | \mathbf{V}_i\} \\
&= \text{E}\left\{ \text{E}\left[\frac{\sigma_x^2}{U_i(1 + \sigma_x^2 \mathbf{b}' \boldsymbol{\Omega}^{-1} \mathbf{b})} | \mathbf{Z}_i \right] \Big| \mathbf{V}_i \right\} + \text{E}\left[\left(\frac{\mu_x + \sigma_x^2 \mathbf{b}' \boldsymbol{\Omega}^{-1} (\mathbf{Z}_i - \mathbf{a})}{1 + \sigma_x^2 \mathbf{b}' \boldsymbol{\Omega}^{-1} \mathbf{b}} \right)^2 \Big| \mathbf{V}_i \right].
\end{aligned}$$

It is easy to show that $\mathrm{E}[U_i^{-1}|\mathbf{Z}_i] = (\mathrm{d}_{\boldsymbol{\Sigma}_z}(\mathbf{Z}_i, \boldsymbol{\mu}_z) + \nu)/(p + \nu - 2)$ – recall that $U_i|\mathbf{Z}_i = \mathbf{Z}_i \sim \text{Gamma}\left(\frac{p+\nu}{2}, \frac{1}{2}\left(\mathrm{d}_{\boldsymbol{\Sigma}_z}(\mathbf{Z}_i, \boldsymbol{\mu}_z) + \nu\right)\right)$–, see the result after Proposition A.5. After some lengthy algebra, we can prove that

$$\mathrm{Var}[x_i|\mathbf{V}_i] = \Lambda\mathrm{E}\left[\frac{\mathrm{d}_{\boldsymbol{\Sigma}_z}(\mathbf{Z}_i, \boldsymbol{\mu}_z) + \nu}{p + \nu - 2}\Big|\mathbf{V}_i\right] + \Lambda^2\mathbf{b}'\boldsymbol{\Omega}^{-1}\mathrm{Var}[\mathbf{Z}_i\big|\mathbf{V}_i]\boldsymbol{\Omega}^{-1}\mathbf{b}, \tag{A.35}$$

where Λ is given in (A.32) and

$$\mathrm{Var}[\mathbf{Z}_i\big|\mathbf{V}_i] = \mathrm{E}[\mathbf{Z}_i\mathbf{Z}_i^\top\big|\mathbf{V}_i] - \mathrm{E}[\mathbf{Z}_i\big|\mathbf{V}_i]\mathrm{E}[\mathbf{Z}_i\big|\mathbf{V}_i]^\top.$$

If individual i has only uncensored components, then expression (A.35) can be computed using the moments of the $\mathrm{t}_p(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z, \nu)$ distribution using Proposition A.2: it is enough to make $d_1 = \cdots = d_p = +\infty$, $r = 1$ and $k = 0$ to obtain the first expectation in (A.35) and $r = 0$, $k = 1$ ($k = 2$) to obtain the other one. If the components are all censored, we again use Proposition A.2, but now considering the moments of a $\mathrm{Tt}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{D}_i)$ distribution. Finally, if there are censored and uncensored components, then the expectations can be computed through Proposition A.3, using the partition $\mathbf{Z}_i = \mathrm{vec}(\mathbf{Z}_i^o, \mathbf{Z}_i^c)$. Besides this, the parameter values in (A.35) must be replaced with the respective EM estimates.

A.5 The observed information matrix

Under some general regularity conditions, we follow Lin (2010) to provide an information-based method to obtain the asymptotic covariance of ML estimates of the t-MEC model's parameters. As defined by Meilijson (1989), the empirical information matrix can be computed as

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

where $\mathbf{S}(\mathbf{Z}_i \mid \boldsymbol{\theta}) = \sum_{i=1}^{n}\mathbf{s}(\mathbf{Z}_i \mid \boldsymbol{\theta})$ and $\mathbf{s}(\mathbf{Z}_i \mid \boldsymbol{\theta})$ is the empirical score function for the ith unit. According to Louis (1982) it is possible to relate the score function of the incomplete data log-likelihood with the conditional expectation of the complete data log-likelihood function. Therefore, the individual score can be determined as

$$\mathbf{s}(\mathbf{Z}_i \mid \boldsymbol{\theta}) = \frac{\partial \log f(\mathbf{Z}_i \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathrm{E}\left[\frac{\partial \ell_{ic}(\boldsymbol{\theta} \mid \mathbf{Z}_i^c)}{\partial \boldsymbol{\theta}} \mid \mathbf{V}_i, \mathbf{C}_i, \boldsymbol{\theta}\right], \tag{A.36}$$

where $\ell_{ic}(\boldsymbol{\theta} \mid \mathbf{Z}_i^c)$ is the complete data log-likelihood formed from the single observation $\mathbf{Z}_i$, $i = 1, \ldots, n$. Using the EM estimates $\widehat{\boldsymbol{\theta}}$, $\mathbf{S}(\mathbf{Z}_i \mid \widehat{\boldsymbol{\theta}}) = 0$, and then (A.36) is given by

$$\mathbf{I}_e(\widehat{\boldsymbol{\theta}} \mid \mathbf{Z}) = \sum_{i=1}^{n}\widehat{\mathbf{s}}_i\widehat{\mathbf{s}}_i^\top, \tag{A.37}$$

where $\widehat{\mathbf{s}}_i = (\widehat{\mathbf{s}}_{i,\boldsymbol{\alpha}}, \widehat{\mathbf{s}}_{i,\boldsymbol{\beta}}, \widehat{\mathbf{s}}_{i,\boldsymbol{\omega}}, \widehat{\mathbf{s}}_{i,\mu_x}, \widehat{\mathbf{s}}_{i,\sigma_x^2})^\top$ is a $3p$-dimensional vector, with components given by

$$\begin{aligned}
\widehat{\mathbf{s}}_{i,\boldsymbol{\alpha}} &= (\widehat{\mathbf{s}}_{i,\alpha_1}, \ldots, \widehat{\mathbf{s}}_{i,\alpha_r})^\top = \mathbb{I}_{(p)}\widehat{\boldsymbol{\Omega}}^{-1}(\widehat{u\mathbf{z}}_i - \widehat{u}_i\widehat{\mathbf{a}} - \widehat{ux}_i\widehat{\mathbf{b}}),\\
\widehat{\mathbf{s}}_{i,\boldsymbol{\beta}} &= (\widehat{\mathbf{s}}_{i,\beta_1}, \ldots, \widehat{\mathbf{s}}_{i,\beta_r})^\top = \mathbb{I}_{(p)}\widehat{\boldsymbol{\Omega}}^{-1}(\widehat{ux\mathbf{z}}_i - \widehat{ux}_i\widehat{\mathbf{a}} - \widehat{ux_i^2}\widehat{\mathbf{b}}),\\
\widehat{\mathbf{s}}_{i,\boldsymbol{\omega}} &= (\widehat{\mathbf{s}}_{i,\omega_1^2}, \ldots, \widehat{\mathbf{s}}_{i,\omega_p^2})^\top = -\frac{1}{2}\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{1}_p + \frac{1}{2}\widehat{\boldsymbol{\Omega}}^{-2}diag(\widehat{a}_i),\\
\widehat{\mathbf{s}}_{i,\mu_x} &= = \frac{1}{\widehat{\sigma_x^2}}(\widehat{ux}_i - \widehat{u}_i\widehat{\mu}_x),\\
\widehat{\mathbf{s}}_{i,\sigma_x^2} &= = -\frac{1}{2\widehat{\sigma_x^2}} + \frac{1}{2\widehat{\sigma_x^4}}(\widehat{ux_i^2} - 2\widehat{ux}_i\widehat{\mu}_x + \widehat{u}_i\widehat{\mu}_x^2),
\end{aligned}$$

with $\mathbb{I}_{(p)} = [\mathbf{0}, \mathbf{I}_{p-1}]_{(p-1)\times p}$, $\mathbf{1}_p = (1, \ldots, 1)^\top_{p\times 1}$ and $\widehat{a}_i = \widehat{u\mathbf{z}_i^2} - 2\widehat{u\mathbf{z}}_i\widehat{\mathbf{a}}^\top - 2\widehat{ux\mathbf{z}}_i\widehat{\mathbf{b}}^\top + 2\widehat{ux}_i\widehat{\mathbf{a}}\widehat{\mathbf{b}}^\top + \widehat{u}_i\widehat{\mathbf{a}}\widehat{\mathbf{a}}^\top + \widehat{ux_i^2}\widehat{\mathbf{b}}\widehat{\mathbf{b}}^\top$.

A.6 Testicular volume data

We illustrate the proposed method with a dataset from Chipkevitch et al. (1996). The data consist of measurements of the testicular volume of 42 adolescents by using five different techniques: ultrasound (US), graphical method proposed by the authors (I), dimensional measurement (II), Prader orchidometer (III), and ring orchidometer (IV). The ultrasound approach was assumed to be the reference measurement device. Galea-Rojas et al. (2002) analyzed the same dataset by fitting the usual normal ME model and recommended considering a data transformation in order to obtain normality. Lachos et al. (2010) also analyzed this dataset with the aim of providing a better fit, attempting to avoid possibly unnecessary data transformation. In fact, they considered a joint model of the latent variable and observational errors by using the scale mixtures of skew-normal (SMSN) class of distributions. They also showed evidence of the heavy-tailed behavior of the data (see also, Cabral et al. (2014)).

To apply our method to this dataset, we censored (randomly) 10% (21 observations) of the data. As a consequence, the detection limit κ_{ij} was fixed at 4.4 for all i and j. Table 27 shows the testicular volume data with the true value in parentheses for the censored observations. We fitted the t-MEC (with $\nu = 6$) and N-MEC models. The EM estimates for the parameters of the two model, as well as their corresponding standard errors (SE) obtained via the empirical information matrix are reported in Table 28. This table shows that the estimates of $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, $\boldsymbol{\omega}$ for the t-MEC and N-MEC models are close. However, the standard errors (SE) of the t-MEC are smaller than those of the N-MEC model, indicating that the our robust model seem to produce more precise estimates.

Table 29 compares the fit of the two models using the model selection criteria

Table 27 – **Chipkevitch data.** Testicular volume data (in ml).

	Methods						Methods				
i	US	I	II	III	IV	i	US	I	II	III	IV
1	5.0	7.5	5.9	8.0	9.0	22	16.5	10.0	15.3	15.0	15.0
2	5.7	5.0	4.8	6.0	10.0	23	4.5	**4.4** (3.5)	**4.4** (3.9)	6.0	7.0
3	7.4	5.0	6.8	9.0	12.0	24	5.6	5.0	4.5	4.5	6.0
4	**4.4** (2.6)	**4.4** (3.5)	**4.4** (3.1)	**4.4** (4.0)	**4.4** (4.0)	25	11.0	7.5	9.7	9.0	11.0
5	5.7	5.0	5.0	6.0	7.0	26	9.2	10.0	11.3	12.0	13.5
6	6.1	5.0	**4.4** (4.4)	7.0	8.0	27	8.5	7.5	8.8	12.0	12.0
7	6.2	5.0	6.0	8.0	9.0	28	5.4	5.0	6.1	8.0	8.0
8	10.4	10.0	8.8	10.0	10.0	29	6.7	7.5	7.2	10.0	8.0
9	9.1	7.5	7.9	10.0	11.0	30	5.3	5.0	5.9	8.0	10.0
10	14.8	10.0	13.0	12.0	15.0	31	20.0	20.0	16.3	25.0	22.5
11	16.4	12.5	10.3	17.5	17.5	32	18.8	15.0	16.3	20.0	25.0
12	9.6	7.5	8.2	10.0	11.0	33	13.9	12.5	12.2	15.0	17.5
13	15.7	15.0	19.8	20.0	20.0	34	9.4	10.0	10.3	12.0	13.5
14	**4.4** (3.0)	**4.4** (2.0)	**4.4** (2.0)	**4.4** (3.0)	**4.4** (4.0)	35	9.1	7.5	10.8	12.0	12.0
15	16.4	15.0	17.3	20.0	20.0	36	14.1	15.0	13.0	13.5	15.0
16	17.6	15.0	17.3	20.0	22.5	37	9.3	10.0	8.4	10.0	10.0
17	10.0	7.5	7.9	12.0	12.0	38	20.9	20.0	22.1	25.0	25.0
18	**4.4** (4.1)	**4.4** (3.5)	**4.4** (4.4)	**4.4** (4.0)	6.0	39	11.5	10.0	10.6	15.0	13.5
19	12.7	10.0	11.4	12.0	12.0	40	9.7	10.0	9.7	11.0	12.0
20	**4.4** (2.7)	**4.4** (3.5)	**4.4** (4.1)	**4.4** (2.5)	6.0	41	13.7	12.5	11.6	17.5	15.0
21	10.2	10.0	11.1	12.0	13.5	42	8.9	10.0	8.1	12.0	12.0

Table 28 – **Chipkevitch data.** ML and SE for parameter estimates.

	t-MEC		N-MEC	
	Estimate	SE	Estimate	SE
α_1	-0.0510	1.1501	-0.0584	1.0995
α_2	-0.6674	0.9077	-0.4205	1.2257
α_3	0.2815	0.9361	0.1172	0.9931
α_4	1.9037	0.9853	1.8075	1.0288
β_1	0.9067	0.1166	0.8959	0.0997
β_2	1.0214	0.0809	0.9792	0.0848
β_3	1.1400	0.1017	1.1371	0.0951
β_4	1.0645	0.1038	1.0619	0.0954
μ_x	9.1089	0.8979	9.9222	1.0681
σ_x^2	18.4174	0.0168	25.0263	0.0124
ω_1	1.1068	0.6503	1.4442	0.8227
ω_2	1.1179	0.4447	1.4313	0.5369
ω_3	1.1339	0.3945	1.9156	0.6718
ω_4	0.9437	0.4708	1.1390	0.5460
ω_5	1.1536	0.4261	1.5493	0.5580

Table 29 – **Chipkevitch data.** Model comparison criteria.

	t-MEC	N-MEC
Log-likelihood	-398.4389	-401.4635
AIC	826.8777	832.9269
BIC	877.0843	883.1336

(AIC and BIC) discussed in Subsection A.7.2. Note that, as expected, the t-MEC model outperform the normal one.

A.7 Simulation studies

In order to study the performance of our proposed method, we present three simulation studies. The first one shows the asymptotic behavior of the EM estimates for the proposed model. The second one investigates the consequences on parameter inference when the normality assumption is inappropriate. Finally, the third one is designed to investigate the effect of including the censoring component in the model.

A.7.1 Asymptotic properties

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

$$\text{BIAS}_k = \frac{1}{M}\sum_{j=1}^{M} |\widehat{\boldsymbol{\theta}}_k^{(j)} - \boldsymbol{\theta}_k| \quad \text{and} \quad \text{MSE}_k = \frac{1}{M}\sum_{j=1}^{M} \left(\widehat{\boldsymbol{\theta}}_k^{(j)} - \boldsymbol{\theta}_k\right)^2, \tag{A.38}$$

where $\widehat{\boldsymbol{\theta}}_k^{(j)}$ is the EM estimate of the parameter $\boldsymbol{\theta}_k$, $k = 1, \ldots, 3p$, for the j-th sample. The key idea of this simulation is to provide empirical evidence about consistency of the EM estimators under the proposed t-MEC model. For each sample size, we generate $M = 100$ datasets with 10% censoring proportion. Using the ECM algorithm, the absolute bias and mean squared error for each parameter over the 100 datasets were computed. The parameter setup (see Section A.3), is

$$\boldsymbol{\alpha} = (3, 2, 1, 2)^\top, \ \boldsymbol{\beta} = (1.5, 1, 1.5, 1)^\top, \ \mu_x = 4, \ \sigma_x^2 = 2 \ \text{ and } \ \boldsymbol{\Omega} = \text{diag}(0.5, 0.5, 0.5, 0.5, 0.5). \tag{A.39}$$

The degrees of freedom were fixed at the value $\nu = 5$.

The results are presented in Figure 34. From this figure we can see that the MSE tends to zero as the sample size increases. Similar results were obtained after the analysis of the absolute bias (BIAS) as can be seen from Figure 37 in the Appendix A.9. As expected, the proposed ECM algorithm provides ML estimates with good asymptotic properties for the t-MEC model.

A.7.2 Parameter inference

In this study we investigate the consequences on parameter inference when the normality assumption is inappropriate, as well the ability of some model choice criteria (AIC and BIC) to select the correct model. In addition, we study the effect of different censoring proportions on the EM estimates. For this purpose, we consider a heavy-tail distribution for the random errors. In this context, we generate $M = 100$ datasets coming

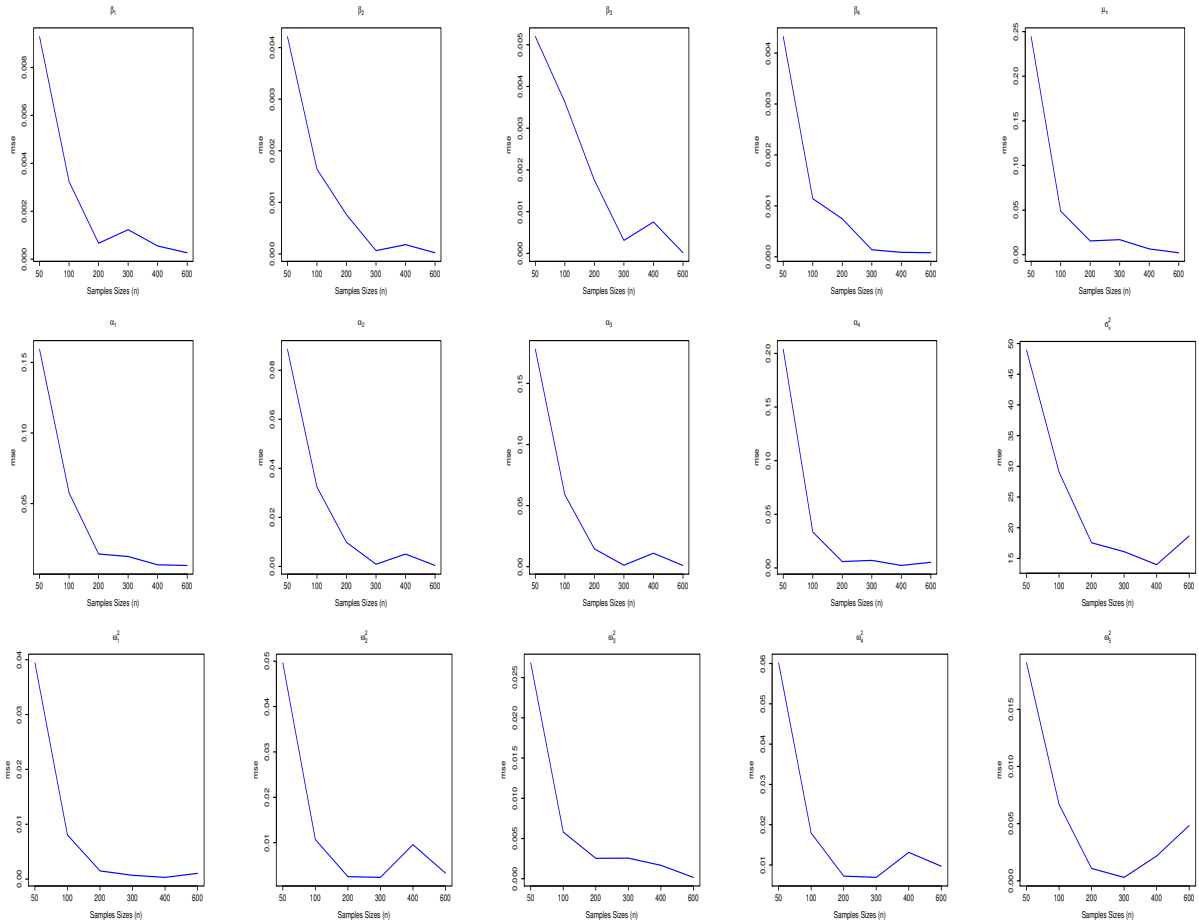

Figure 34 – **Simulation study A.7.1.** MSE of parameter estimates under the t-MEC model considering 10% censoring.

from a slash distribution with parameter $\nu = 1.5$ and censoring proportions 0%, 10%, 20% and 30%. The slash distribution arises when we change the distribution of U in (A.2) to $U \sim \text{Beta}(\nu, 1)$, with *pdf* $f(u|\nu) = \nu u^{\nu-1}$, $u \in (0, 1)$, and $\nu > 0$. See Wang and Genton (2006) for details. The parameter values are set as in the previous experimental study.

For each simulated dataset we fitted the t-MEC (with $\nu = 5$ degrees of freedom) and the N-MEC models. The model selection criteria AIC and BIC as well as the estimates of the model parameters were recorded at each simulation. Summary statistics such as the Monte Carlo mean estimate (MC mean), coverage probability (MC CP) and the approximate standard error obtained through the information-based method (IM SE), discussed in Section 5, for the parameter estimates are presented in Table 30.

From these results we can observe that for all considered levels of censoring, the t-MEC model is chosen as the correct model. Under the t-MEC model, the MC CP for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are stable, but the MC CP of μ_x is lower than the nominal level (95%). In general, the MC CP values are higher than those obtained under the normal model. Figure 35 shows the MSE for some parameter estimates (the biases are presented in Figure 39 in the Appendix A.9). Note that, the MSE under the t-MEC model is lower than the

Table 30 – **Simulation study A.7.2.** Summary statistics based on 100 simulated samples from the slash distribution for different levels of censoring (0%, 10%, 20%, 30%).

			Simulated data									
Censoring	Fit		α_1	α_2	α_3	α_4	β_1	β_2	β_3	β_4	μ_x	σ_x^2
0%	Normal	MC Mean	3.096	2.457	1.274	2.238	1.482	0.933	1.466	0.988	3.463	212.602
		IM SE	0.332	0.243	0.343	0.302	0.038	0.029	0.045	0.038	4.783	0.002
		MC CP	100%	24%	97%	100%	100%	24%	100%	100%	100%	
	Student-t	MC Mean	2.712	1.853	0.964	1.894	1.542	1.022	1.521	1.022	4.586	9.299
		IM SE	0.375	0.245	0.331	0.258	0.066	0.044	0.060	0.046	0.404	0.022
		MC CP	100%	100%	100%	100%	100%	100%	100%	100%	79%	
10%	Normal	MC Mean	2.440	1.570	0.887	1.701	1.608	1.097	1.535	1.090	4.797	31.343
		IM SE	0.345	0.328	0.456	0.405	0.044	0.051	0.072	0.062	0.847	0.010
		MC CP	43%	61%	100%	61%	71%	68%	100%	100%	100%	
	Student-t	MC Mean	2.612	1.626	0.880	1.849	1.565	1.069	1.540	1.031	4.539	7.927
		IM SE	0.415	0.312	0.397	0.294	0.073	0.058	0.072	0.053	0.351	0.026
		MC CP	100%	100%	100%	100%	100%	100%	100%	100%	75%	
20%	Normal	MC Mean	2.435	1.539	0.905	1.600	1.610	1.103	1.533	1.103	4.754	32.254
		IM SE	0.415	0.417	0.532	0.492	0.051	0.063	0.081	0.076	0.832	0.010
		MC CP	55%	61%	100%	61%	83%	85%	100%	98%	100%	
	Student-t	MC Mean	2.506	1.605	0.816	1.783	1.580	1.073	1.549	1.041	4.576	7.599
		IM SE	0.503	0.399	0.494	0.357	0.084	0.070	0.084	0.061	0.350	0.027
		MC CP	99%	100%	100%	100%	100%	100%	100%	100%	64%	
30%	Normal	MC Mean	2.511	1.490	0.999	1.532	1.608	1.112	1.528	1.114	4.703	32.890
		IM SE	0.521	0.543	0.662	0.606	0.061	0.077	0.096	0.093	0.841	0.011
		MC CP	61%	61%	100%	61%	87%	89%	100%	88%	100%	
	Student-t	MC Mean	2.522	1.578	0.864	1.840	1.580	1.080	1.545	1.034	4.554	7.211
		IM SE	0.622	0.506	0.641	0.433	0.096	0.082	0.099	0.069	0.352	0.029
		MC CP	100%	100%	100%	100%	100%	99%	100%	100%	84%	

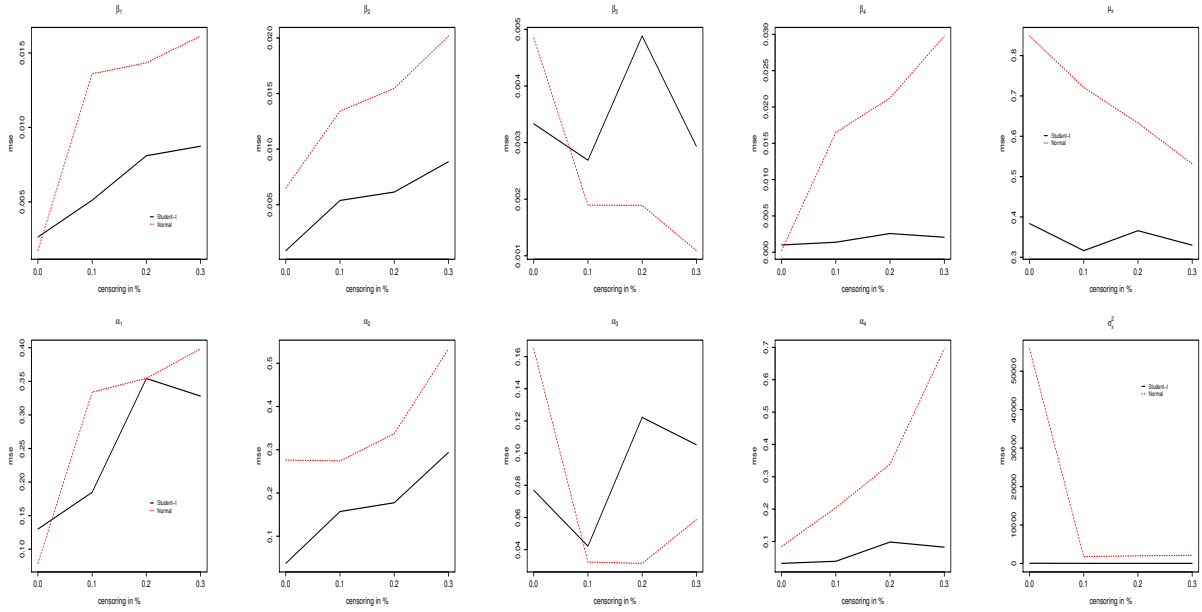

Figure 35 – **Simulation study A.7.2.** MSE of $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, μ_x, σ_x^2 estimates under normal and Student-t models for different levels of censoring (0%, 10%, 20%, 30%).

obtained under the normal, for different levels of censoring.

Regarding the model choice, the t-MEC model was chosen as the best by the two criteria for all samples.

A.7.3 Censored model

In this section, the main goal is to study the effect of taking into account censored data on the parameter estimates. We generated $M = 100$ samples from the t-MEC model with $\nu = 5$, setting the censoring level at 20%. The other parameter values are set as in (A.39). For each dataset, we fitted two models: in *case 1* we use a naive model, where censored responses are not taken into account. In *case 2* we fit a t-MEC model.

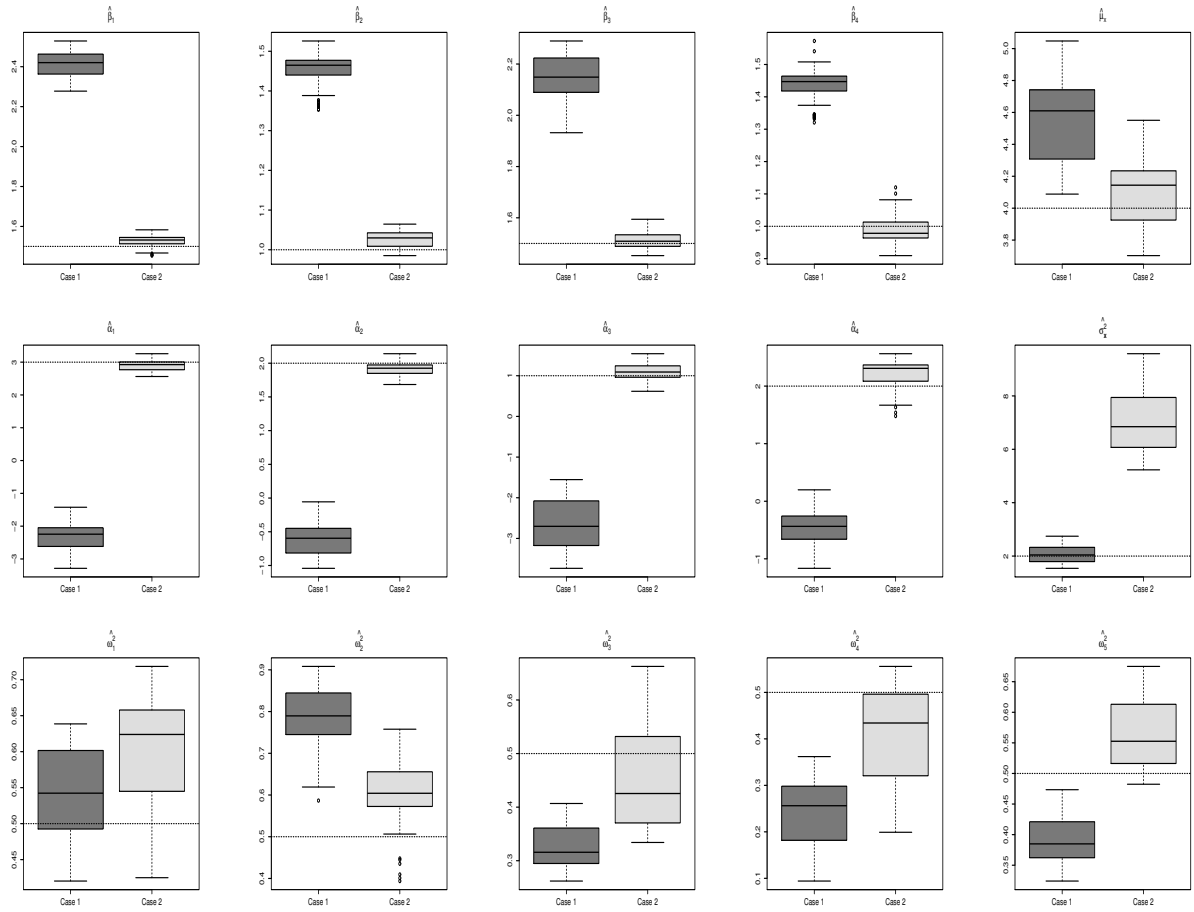

Figure 36 – **Simulation study A.7.3.** Boxplots of the parameter estimates. Dotted lines indicate the true parameter value.

Figure 36 shows the box plots corresponding to each parameter estimate considering the $M = 100$ datasets. Note that, the estimates in *case 2* are, in general, more precise than those obtained in *case 1*. It is also possible to note that, in *case 2*, the variability observed in the estimations is smaller than in *case 1*, except for some dispersion parameters. We point out that it is important to consider the effect of censoring in data modeling, avoiding ad-hoc methods.

A.8 Conclusions

In this work, we introduce the multivariate ME model with censored responses based on the Student-t distribution, the so-called t-MEC model. This model considers the possibility of censoring in the surrogate covariate and the response. Moreover, we assume that the latent unobserved covariate and random observational errors follow a multivariate Student-t distribution, which provides a robust alternative to the usual Gaussian model. For the parameter estimation, an ECM algorithm based on some statistical properties of the multivariate truncated Student-t distribution is developed to obtain ML estimates. Some simulation studies revealed that our proposed method generates less biased estimates of model parameters than the case when the censoring scheme is not taken into account. Moreover, we showed that the use of the Student-t distribution generates better results than the normal one, in the context of the censored ME models.

Of course, further extensions of the current work are possible. For example, the proposed method can be naturally extended by considering the family of scale mixtures of normal (SMN) and skew-normal (SMSN) distributions. An efficient estimation procedure to obtain ML estimates of model parameters can be implemented by using a stochastic approximation of the traditional EM (SAEM) algorithm. Other extensions include, a Bayesian treatment via Markov chain Monte Carlo (MCMC) sampling methods in the context of SMN-MEC and SMSN-MEC models (Lachos et al. (2010)).

A.9 Appendix

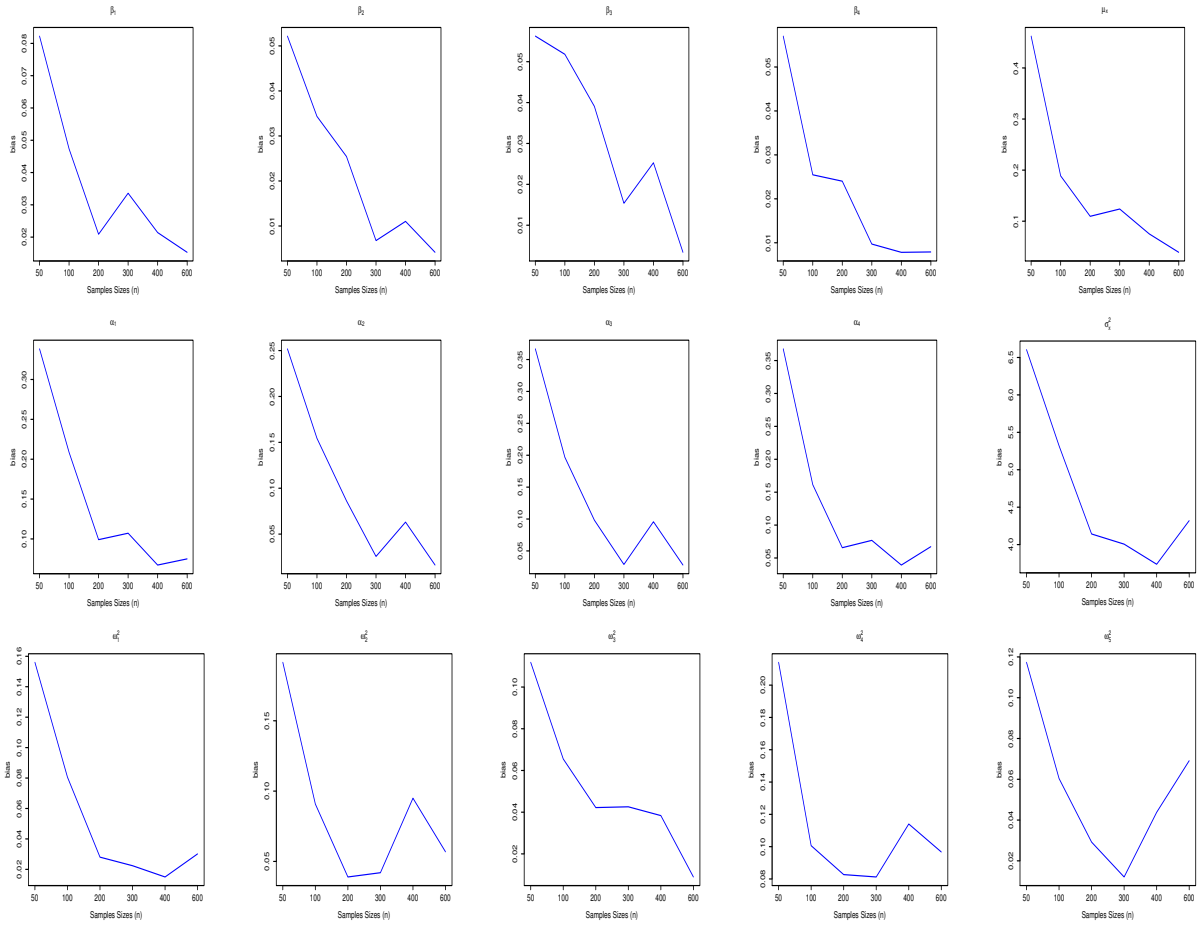

Figure 37 – **Simulation A.7.1.** Bias of parameter estimates under the t-MEC model considering 10% of censoring.

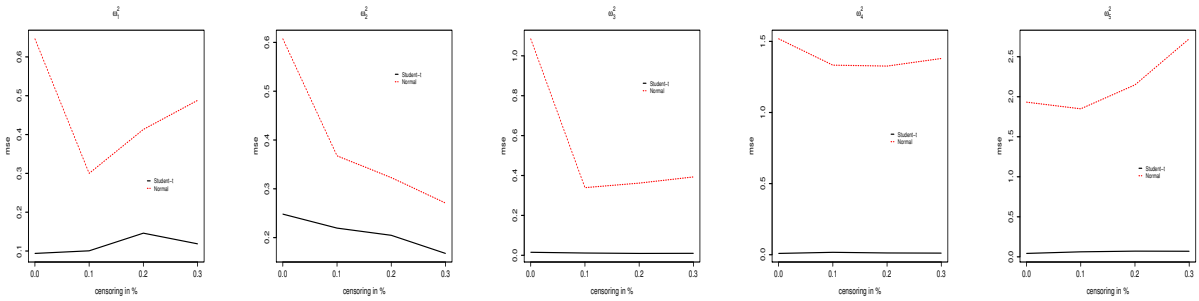

Figure 38 – **Simulation A.7.2.** MSE of parameter estimates under the t-MEC model considering 10% of censoring.

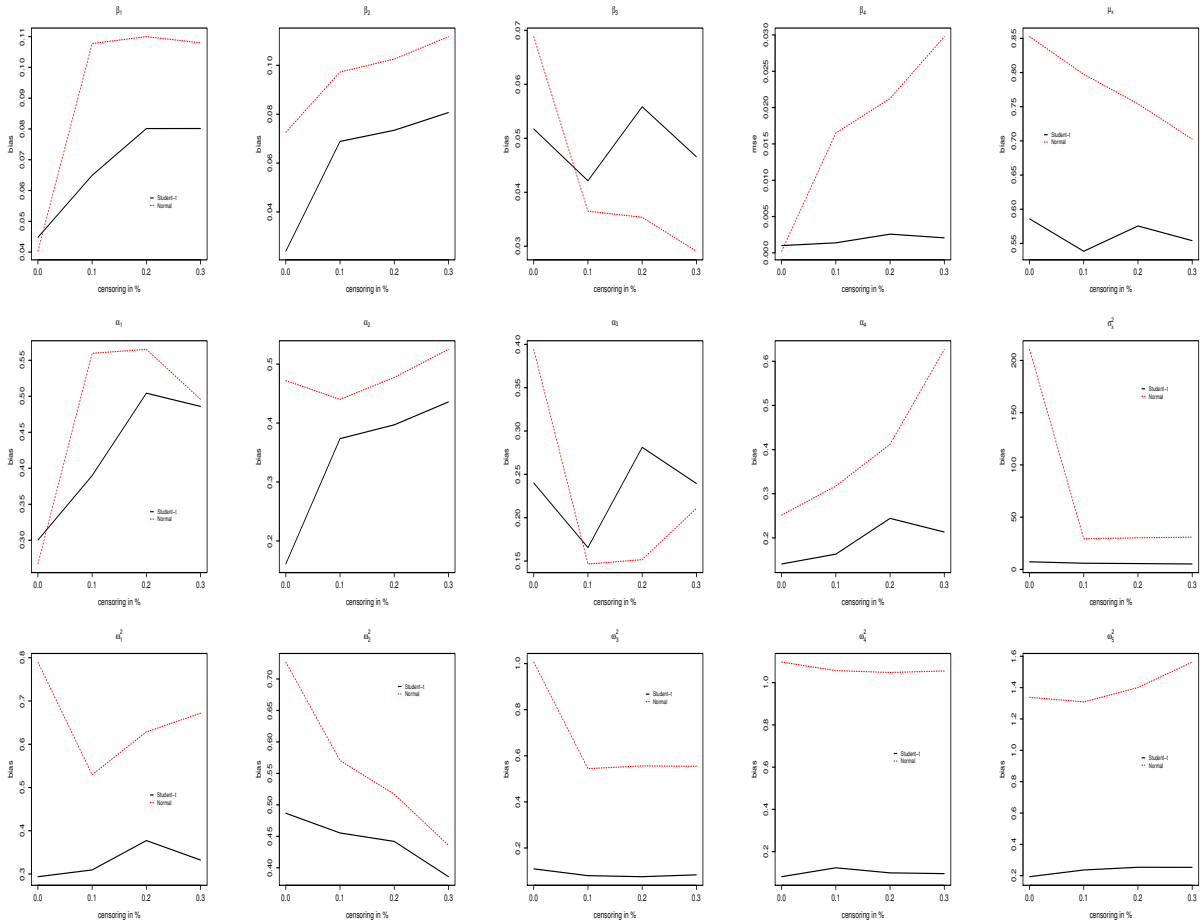

Figure 39 – **Simulation A.7.2.** Bias of parameter estimates under the t-MEC model considering different levels of censoring .