UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Matemática, Estatística e Computação Científica

THALITA DO BEM MATTOS

Parametric and semiparametric mixed-effects models for longitudinal censored data

Modelos paramétricos e semiparamétricos de efeitos mistos para dados longitudinais censurados

Campinas

2020

Thalita do Bem Mattos

Parametric and semiparametric mixed-effects models for longitudinal censored data

Modelos paramétricos e semiparamétricos de efeitos mistos para dados longitudinais 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 Doutora em Estatística.

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.

Supervisor: Larissa Avila Matos

Co-supervisor: Victor Hugo Lachos Dávila

Este exemplar corresponde à versão final da Tese defendida pela aluna Thalita do Bem Mattos e orientada pela Profa. Dra. Larissa Avila Matos.

Campinas

2020

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

M436p

Mattos, Thalita do Bem, 1991-
Parametric and semiparametric mixed-effects models for longitudinal censored data / Thalita do Bem Mattos. – Campinas, SP : [s.n.], 2020.

Orientador: Larissa Avila Matos.
Coorientador: Víctor Hugo Lachos Dávila.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Observações censuradas (Estatística). 2. Estatística não paramétrica. 3. Distribuição normal assimétrica. I. Matos, Larissa Avila, 1987-. II. Lachos Dávila, Víctor Hugo, 1973-. III. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Modelos paramétricos e semiparamétricos de efeitos mistos para dados longitudinais censurados
Palavras-chave em inglês:
Censored observations (Statistics)
Nonparametric statistics
Skew-normal distributions
Área de concentração: Estatística
Titulação: Doutora em Estatística
Banca examinadora:
Larissa Avila Matos [Orientador]
Guilherme Vieira Nunes Ludwig
Camila Borelli Zeller
Cristian Marcelo Villegas Lobos
Luis Mauricio Castro Cepero
Data de defesa: 03-07-2020
Programa de Pós-Graduação: Estatística

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0003-4977-0498
- Currículo Lattes do autor: http://lattes.cnpq.br/9500569794852267

Tese de Doutorado defendida em 03 de julho de 2020 e aprovada

pela banca examinadora composta pelos Profs. Drs.

Prof(a). Dr(a). LARISSA AVILA MATOS

Prof(a). Dr(a). GUILHERME VIEIRA NUNES LUDWIG

Prof(a). Dr(a). CAMILA BORELLI ZELLER

Prof(a). Dr(a). CRISTIAN MARCELO VILLEGAS LOBOS

Prof(a). Dr(a). LUIS MAURICIO CASTRO CEPERO

A Ata da Defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós-Graduação do Instituto de Matemática, Estatística e Computação Científica.

This work is dedicated to the most important women in my life,
my mom and my sister.

Acknowledgements

Though it is impossible to express my gratitude to all individuals who supported me, I wish to extend my deepest appreciation to several special people who have helped and encouraged me in compiling this thesis. I owe them my sincere thanks.

First, I would like to thank God for never abandon me, giving the wisdom, strength, support and knowledge throughout this journey.

I am deeply grateful to my supervisors, Professor Larissa Avila Matos and Professor Víctor Hugo Lachos Dávila, for their expertise, support and mentorship throughout this long process. I consider myself extremely fortunate to have had the opportunity to learn from you both.

I would also like to thank the committee members, Professor Guilherme Vieira Nunes Ludwing, Professor Camila Borelli Zeller, Professor Cristian Marcelo Villegas Lobos and Professor Luis Mauricio Castro Cepero for their insightful comments, questions and remarks. They improved the quality of my work.

Furthermore, I would like to thank all my colleagues of the faculty, specially all the members of laboratory 105 - *"Modelos de Regressão"*. They encouraged and supported to me during my PhD life. I would also like to thank my friends outside the doctorate, Gisele, Isabela, Jéssica e Laila for providing the laughter, respite, and balance required to maintain focus and perspective throughout this degree.

Last but not least, I would love to cordially express my gratitude for all the support, understanding, help, and encouragement to my family, without whom I would not be where I am today. To my parents, Marilene and Milton, thank you for always believing in me and teaching me the value of hard work, determination, and perseverance. Thank you to my sister Areta and my nephew Luccas for their encouragement and unconditional support. To my fiancé, Francisco, thank you wholeheartedly for your endless support, patience, help, understanding, love, and helpful discussions. You are a true life partner. Thanks for everything!

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

"Wisdom, like an inheritance, is a good thing
and benefits those who see the sun.
Wisdom is a shelter as money is a shelter,
but the advantage of knowledge is this:
Wisdom preserves those who have it."
(Ecclesiastes 7:11-12)

Resumo

Estudos longitudinais envolvendo resultados de laboratório, medições repetidas podem ser censuradas devido aos limites de detecção do ensaio. Por exemplo, em ensaios clínicos de AIDS, as medições de RNA do HIV-1 são coletadas irregularmente ao longo do tempo e geralmente estão sujeitas a alguns limites superiores e inferiores de detecção, dependendo dos ensaios de quantificação. Para analisar esses dados, várias abordagens utilizando modelos de efeitos mistos lineares e não lineares censurados têm sido propostas na literatura. Uma complicação surge quando a forma paramétrica dos modelos de efeitos mistos parece muito restritiva para caracterizar a complexa relação entre uma variável de resposta e suas covariáveis ao longo do tempo. Nesta tese, propomos o uso de modelos mistos semiparamétricos para analisar dados longitudinais censurados, estendendo os modelos de efeitos mistos lineares censurados e fornecendo um esquema de modelagem mais flexível, permitindo modelar o valor esperado da variável resposta através de uma função arbitrária do tempo e de funções paramétricas das covariáveis. Além disso, uma suposição comum de distribuição para modelos de efeitos mistos é a distribuição normal para erros aleatórios e efeitos aleatórios. Essa suposição pode não ser robusta contra desvios da normalidade e pode levar a uma inferência enganosa ou tendenciosa. Portanto, também estendemos os modelos mistos semiparamétricos com erros normais a erros com distribuição t multivariada e também estendemos os modelos de efeitos mistos lineares/não lineares a erros com distribuição skew-normal, a fim de permitir distribuições com caudas mais pesadas que a normal.

Palavras-chave: Dados censurados. Algoritmo EM. Dados longitudinais. Modelos de efeitos mistos. Modelo misto semiparamétrico.

Abstract

Longitudinal studies involving laboratory results, repeated measurements can be censored due to assay detection limits. For example, in HIV/AIDS clinical trials, the HIV-1 RNA measurements are collected irregularly over time and are often subject to some upper and lower detection limits, depending on the quantification assays. For analyzing such data, several approaches using censored linear and nonlinear mixed-effects models have been proposed in the literature. A complication arises when the parametric form of mixed effects models appears too restrictive to characterize the complex relationship between a response variable and its covariates over time. In this thesis, we propose the use of semiparametric mixed models to analyze censored longitudinal data extend censored linear mixed-effects models and provide a more flexible modeling, allowing to model the expected value of the response variable through an arbitrary function of time and parametric functions of covariates. In addition, a common assumption of the distribution for mixed-effect models is the normal distribution for random errors and random effects. This assumption may not be robust against deviations from normality and may lead to a misleading or biased inference. Therefore, we also extend the semiparametric mixed models with normal errors to errors with multivariate-t distribution, and also, we extend the linear/nonlinear mixed-effects models to errors with skew-normal distribution, in order to allow distributions with tails heavier than normal.

Keywords: Censored data. EM algorithm. Longitudinal data. Mixed-effects models. Semiparametric mixed model.

List of Figures

List of Tables

List of abbreviations and acronyms

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
SMEC	Semiparametric mixed effects models for censored responses
ML	Maximum likelihood
MPL	Maximum penalized likelihood
REML	Restricted maximum likelihood
DEC	Damped exponential correlation
AR(1)	Continuous-time autoregressive of order 1
CS	Compound symmetric
MA(1)	Moving average of order 1
UNC	Uncorrelated
SN	Skew-normal distribution
ESN	Extended skew-normal distribution
AIC	Akaike information criterion
BIC	Bayesian information criterion

Contents

Chapter 1

Introduction

1.1 Background and motivation

Longitudinal studies are common in epidemiological and biomedical research. In these studies, measurements of one or more variables are made repeatedly over time for a group of subjects. A key feature of longitudinal data is that the repeated measurements of a variable within a subject tend to be correlated with each other, that is, there may be within-individual correlations. The measurements across subjects are usually assumed to be independent. Due to the within-individual correlations and the between-individual variations in longitudinal data, classical regression models for cross-sectional data, such as linear or generalized linear models, are not appropriate for longitudinal data analysis. Parametric regression models, such as linear mixed-effects (LME) models (Laird & Ware, 1982) have proved to be valuable tools for analyzing continuous longitudinal data. With the incorporation of subject-specific random effects, LMEs can properly model the correlation of longitudinal data.

As an example of longitudinal data, we may consider an AIDS (acquired immunodeficiency syndrome) study. The HIV (human immunodeficiency virus) progress status is usually measured via HIV-1 viral RNA (viral load) or CD4 cell count in the plasma. CD4 cell count is more often used as an endpoint for long follow-up trials or advanced patients population, but for trials with short follow-up periods, viral load is often used as a primary endpoint to quantify treatment effect, where CD4 cell count is viewed as a covariate to help predict virologic responses. Since the viral load is measured repeatedly from the beginning of the treatment, the measures obtained from the same subject may be correlated but can be assumed to be independent if obtained across different subjects. A powerful tool to handle such longitudinal data is mixed-effects modeling, where linear and nonlinear mixed-effects (LME/NLME) modeling approaches have been proposed in HIV dynamics (Wu *et al.*, 1998; Wu, 2005).

Another complexity of longitudinal data is when the response in a longitudinal study is a laboratory-based outcome, in this study censoring may occur due to the upper and/or lower detection limits of an assay. Examples of such censored longitudinal data arise from a variety of research areas (Moulton & Halsey, 1995; Singh & Nocerino, 2002). Censoring can be left, right or interval-censored, typical examples of left censored longitudinal data are from HIV studies, where the detection of viral load in blood compartment is often limited by the sensitivity of a laboratory-performed assay.

Many statistical approaches have been developed to deal with longitudinal data containing censored measurements within the mixed-effects model framework. A simple and *ad hoc* approach is to substitute the censored measurements with the value of the full or half detection limit, which was shown to produce biased estimates (Hughes, 1999; Jacqmin-Gadda *et al.*, 2000). As alternatives to these crude imputation techniques, Vaida & Liu (2009) proposed expectation-maximization (EM) schemes for LME/NLME models with censored responses (LMEC/NLMEC).

Although LMEs are useful tools for analyzing longitudinal data, an important assumption of LMEs is that the response variable is linearly related to its covariates by a known function. Commonly, this linear regression function is not straightforward to derive due to the lack of sufficient understanding of scientific problems. In other situations, the linear parametric form of LMEs appears too restrictive to be used to address the complex relationship between a response variable and covariates. To overcome this difficulty, a more general and robust modeling tool is needed, which motivates the development of nonparametric regression models.

In the last years, nonparametric and semiparametric regression models, that provide great flexibility in modeling covariate effects of longitudinal data, have been extensively investigated. Instead of using a linear predictor, these models formulate the relationship between the response variable and certain covariates through arbitrary functions, and the unknown functions are estimated using nonparametric smoothing techniques. Hence, semiparametric regression models have gained increasing attention in longitudinal data analysis due to their flexible structure. As implied by the name, semiparametric regression models incorporate both parametric and nonparametric forms of covariate effects, and therefore enjoy the flexibility of nonparametric regression models while retaining nice properties such as easy implementation and good interpretability of parametric models. There are a rich literature on the development of semiparametric regression models for longitudinal data analysis, for example, Zeger & Diggle (1994); Ruppert *et al.* (2003); Arribas-Gil *et al.* (2015); Szczesniak *et al.* (2015).

However, these developments are in general made on the assumption of normal errors. Some works have investigated alternative distributions for the errors in multivariate and repeated-measures problems with indications of light- or heavy-tailed distributions. Pinheiro *et al.* (2001) proposed a robust hierarchical linear mixed model in which the random effects and the within-subject errors have a multivariate t-distribution. Lachos *et al.* (2010) proposed a robust generalization of LME, called the skew normal/independent linear mixed (SNI-LME) model, by assuming a skew normal/independent (SNI) distribution (Branco & Dey, 2001) for the random effects and a normal/independent distribution for the random errors. Ibacache-Pulgar *et al.* (2012) extend semiparametric mixed linear models with normal errors to elliptical errors in order to permit distributions with heavier and lighter tails than the normal ones. In the context of censored responses, Matos *et al.* (2013b) proposed an EM algorithm for linear and nonlinear mixed-effects models with censored response using the multivariate Student's t-distribution (t-LMEC/t-NLMEC). Bayesian proposals in the context of heavy-tailed include Lachos *et al.* (2011) who adopted a Bayesian approach to carry out posterior inference for censored linear and nonlinear mixed-effects models considering a class of thick-tail distributions (the so called normal/independent proposed by Lange & Sinsheimer (1993)) as the joint distribution of the error term and random effects, while Bandyopadhyay *et al.* (2012, 2015) studied the LMEC model considering both skewness and heavy-tails. Most recently, Castro *et al.* (2019) proposed a Bayesian flexible semiparametric approach to model censored longitudinal data, where the random error of the model is normally distributed and the random effects follow a skew-normal distribution.

This thesis is devoted to a series of chapters that use different models and techniques to deal with censored data, in particular, HIV/AIDS clinical trials. As a result, we have implemented different approaches to modeling longitudinal the censored data. The organization of the thesis is as follows:

Chapter 2: We propose semiparametric mixed models to analyze censored longitudinal data with irregularly observed repeated measures. The proposed model extends the censored LME model and provides more flexible modeling schemes by allowing the time effect to vary nonparametrically over time. We develop an EM algorithm for maximum penalized likelihood (MPL) estimation of model parameters and the nonparametric component. Further, as a byproduct of the EM algorithm, the smoothing parameter is estimated using a modified LME model, which is faster than alternative methods such as the restricted maximum likelihood (REML) approach. Finally, the performance of the proposed approach is evaluated through extensive simulation studies as well as application to dataset from AIDS study.

Chapter 3: We extended the semiparametric mixed model for longitudinal censored data with normal errors to Student's-t erros. This models allows flexible functional dependence of an outcome variable on covariates by using nonparametric regression, while accounting for correlation between observations by using random effects. Penalized likelihood equations are applied to derive the maximum likelihood estimates which appear to be robust against outlying observations in the sense of the Mahalanobis distance. We estimate nonparametric functions by using smoothing splines jointly estimate smoothing parameter by the EM algorithm. Finally, the performance of the proposed approach is evaluated through extensive simulation studies as well as application to dataset from AIDS study.

Chapter 4: Although normal distributions are commonly assumed for random effects, such assumption may be unrealistic obscuring important features of among-individual variation. We relax this assumption by consider a likelihood-based inference for linear and nonlinear mixed effects models with censored response (NLMEC/LMEC) based on the multivariate skew-normal distribution. An ECM algorithm is developed for computing the maximum likelihood estimates for NLMEC/LMEC with the standard errors of the fixed effects and the exact likelihood value as a by-product. The algorithm uses closed-form expressions at the E-step, that rely on formulas for the mean and variance of a truncated multivariate skew-normal distribution. It is applied to analyze longitudinal HIV viral load data in two recent AIDS studies. In addition, a simulation study is conducted to examine the performance of the proposed methods.

Chapter 5: We present final remarks and perspectives for future research related to this thesis.

In the following sections of Chapter 1, we provide a brief description of the EM algorithm, which will be used in our developments to find, for instance, the maximum likelihood estimates of the model parameters and the damping exponential correlation (DEC) structure. Also, we describe the data sets used in the applications of the thesis.

1.2 The EM algorithm

The EM algorithm (Dempster *et al.*, 1977) is a popular iterative algorithm for maximum likelihood (ML) estimation in models when the data has missing/censored observations and/or latent variables. More specifically, let $\mathbf{y}_{obs}$ denote the observed data and $\mathbf{y}_{mis}$ denote the missing data. The complete data $\mathbf{y}_{com} = (\mathbf{y}_{obs}, \mathbf{y}_{mis})$ is $\mathbf{y}_{obs}$ augmented with $\mathbf{y}_{mis}$, and $\ell_{com}(\boldsymbol{\theta}|\mathbf{y}_{com}) = \log(f(\mathbf{y}_{com}|\boldsymbol{\theta}))$ the complete-data log-likelihood function of a parameter vector $\boldsymbol{\theta} \in \Theta$. The EM algorithm consists basically of two steps: the *Expectation* step (E-step) and the *Maximization* step (M-step). Each iteration is performed as follows:

E-step: Calculate the conditional expectation $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \mathbb{E}\left[\ell_{com}(\boldsymbol{\theta}|\mathbf{y}_{com})|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}\right]$, where $\widehat{\boldsymbol{\theta}}^{(k)}$ is the estimate of $\boldsymbol{\theta}$ at the k-th iteration.

M-step: Find $\boldsymbol{\theta}^{(k+1)}$ such that $Q(\boldsymbol{\theta}^{(k+1)}|\boldsymbol{\theta}^{(k)}) = \max\limits_{\boldsymbol{\theta}\in\Theta} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$.

These steps are performed iteratively in $\ell_{com}(\boldsymbol{\theta}|\mathbf{y}_{com})$ until it reaches the convergence.

When the M-step in the EM algorithm is difficult to implement, it is often useful to replace it with a sequence of constrained maximization (CM) steps, each of which maximizes $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$ over $\boldsymbol{\theta}$ with some function of $\boldsymbol{\theta}$ held fixed. The sequence of CM-steps is such that the over all maximization is over the full parameter space. This leads to a simple extension of the EM algorithm, called the ECM algorithm (Meng & Rubin, 1993). A further extension of the EM algorithm is the ECME algorithm (Liu & Rubin, 1994). This algorithm replaces each CM-step of ECM with a CM-step that maximizes either the constrained Q-function, as in ECM, or the correspondingly constrained likelihood function. Liu & Rubin (1994) showed that ECME typically shares with EM the simplicity and stability, but has a faster rate of convergence, especially for the Student's t-distribution with unknown degrees-of-freedom.

1.3 Damped exponential correlation structure (DEC)

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

$$\mathbf{E}_i = \mathbf{E}_i(\boldsymbol{\phi}, \mathbf{t}_i) = \left[\phi_1^{|t_{ij}-t_{ik}|^{\phi_2}}\right], \quad i = 1, \ldots, n, j, k = 1, \ldots, n_i, \tag{1.1}$$

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 is the correlation between observations separated by one t-unit in time, and the "scale parameter" ϕ_2 permits attenuation or acceleration of the exponential decay of the autocorrelation function defining a continuous-time autoregressive (AR) model. From a practical point of view and in order to avoid computational problems, the parameter space of ϕ_1 and ϕ_2 is confined within $\boldsymbol{\phi} = \{(\phi_1, \phi_2) : 0 < \phi_1 < 1, \phi_2 \geqslant 0\}$.

For nonnegative ϕ_1, the correlation structure given in (1.1) produces a variety of standard correlation structures upon fixing the damping parameter ϕ_2, as follows:

1. if $\phi_2 = 0$, then $\mathbf{E}_i$ is the compound symmetry correlation structure (CS);
2. when $0 < \phi_2, 1$, then $\mathbf{E}_i$ presents a decay rate between the compound symmetry and AR(1) model;
3. if $\phi_2 = 1$, then $\mathbf{E}_i$ generates an AR(1) structure;
4. when $\phi_2 > 1$, then $\mathbf{E}_i$ presents a decay rate faster than that of AR(1).
5. if $\phi_2 \to \infty$, $\mathbf{E}_i$ yields MA(1), the moving average model of order 1.

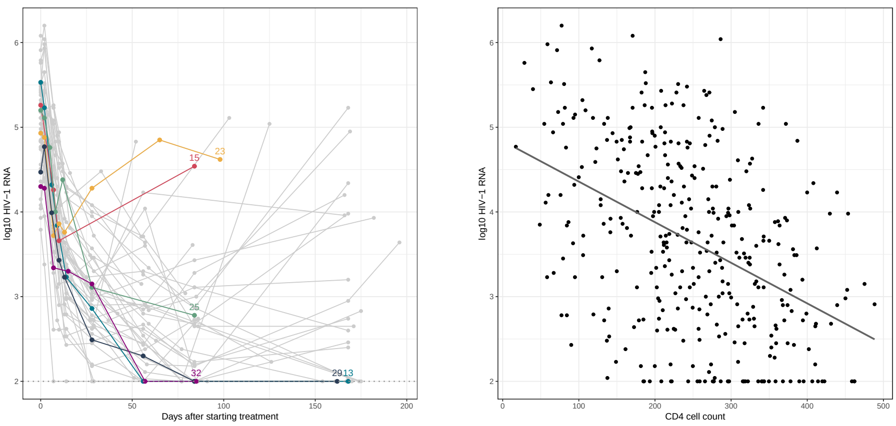
1.4 Case studies

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

1.4.1 ACTG 315 data

The AIDS Clinical Trials Group (ACTG) protocol 315 considers 46 HIV-1 infected patients treated with a potent antiretroviral regimen consisting of protease inhibitor and reverse transcriptase inhibitor drugs. Before initiating the antiretroviral regimen, 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 irregularly on days 0, 2, 7, 10, 14, 21, 28, 56, 84, 168 and 196 after start of treatment, generating 361 observations. $CD4^+$ cell counts were also measured along with viral loads. Measurements below the detectable threshold of 100 copies/mL (40 out of 361, i.e, 11% censored observations) were considered left-censored, and the censoring mechanism was assumed to be independent of the complete data. The number of measurements per subject varied from 4 to 10. Figure 1a displays the individual profiles of the viral loads. As can be seen, the HIV-1 RNA levels changed over time in a nonlinear manner. Moreover, a variation in the intercept among individuals is also observed. In Figure 1b, we display a scatter plot of the viral load and $CD4^+$ cell counts, showing an inverse relationship between viral and the $CD4^+$ cell count, i.e., high $CD4^+$ cell count leads to lower levels of viral load. This is because the $CD4^+$ cells (also called T-cells) alert the immune system in case of invasion by viruses and/or bacteria. Consequently, a lower $CD4^+$ count means a weaker immune system. For a more detailed description of the HIV/AIDS study, we refer the interested reader to Landay *et al.* (1998) and Kotzin *et al.* (2000).



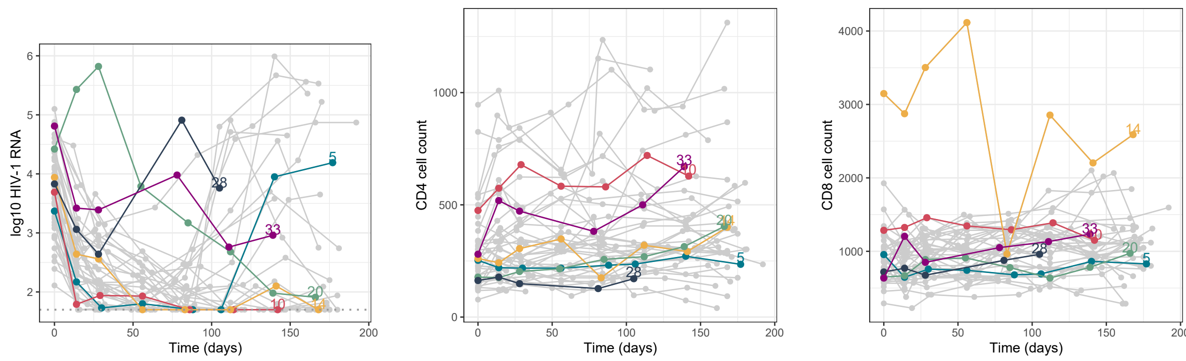
(a) (b)

Figure 1 – **ACTG 315 data**. (a) Individual profiles for HIV viral load (in $\log_{10}$ scale) at different follow-up times. Dotted line indicates the censoring level. (b) Scatter plot of the $CD4^+$ cell counts against viral loads (in $\log_{10}$ scale). Gray line is a regression line between $\mathbf{y} \sim CD4^+$.

1.4.2 A5055 data

The ACTG protocol A5055 was a phase I/II, randomized, open-label, 24-week comparative study of the pharmacokinetics, tolerability, safety and antiretroviral effects of two regimens of indinavir, ritonavir and two nucleoside analogue reverse transcriptase inhibitors on HIV-1 infected patients. A more detailed description of this study and data can be found in Acosta *et al.* (2004).

In this study, 44 patients were randomized in one of two regimens and plasma HIV-1 RNA (viral load) was measured (copies/mL) in blood samples collected irregularly on study days 0, 7, 14, 28, 56, 84, 112, 140, and 168 of follow-up. The nucleic acid sequence-based amplification assay (NASBA) was used to measure plasma HIV-1 RNA, with a lower limit of quantification of 50 copies/mL, and there were 102 out of 308 (around 33.12%) RNA viral load measurements below the detection limit, so there was left censoring. A series of potentially explanatory variables was collected at the same time. For the data analysis, we consider only the covariates $CD4^+$ and $CD8^+$ cell counts. The number of measurements per subject varied from 1 to 8. Figure 2a shows the longitudinal trajectories of RNA viral load (in log-base-10 scale) across days for patients. It can be noted that the viral load trajectory is complex and is substantially different across individuals. Figures 2b and 2c display the individual trajectories of $CD4^+$ and $CD8^+$ cell counts, respectively. Previous studies show that $CD4^+$ cells and $CD8^+$ cells are immunologic markers, providing a way of prognosticating the status of progression from HIV to AIDS. HIV-1 infection and AIDS are characterized by a significant and progressive destruction of $CD4^+$ cells, which results in a weakened immune system that can no longer fight infections. In addition to the steady decline of $CD4^+$ cells during infection, there is a concomitant increase in $CD8^+$ cells as part of the normal immune response to viral infection (Stevens *et al.*, 2006).



(a) (b) (c)

Figure 2 – **A5055 data**. Individual profiles for HIV viral load (in $\log_{10}$ scale), $CD4^+$ and $CD8^+$ cell count at different follow-up times. Dotted line indicates the censoring level.

1.4.3 AIEDRP data

The AIEDRP data set consists of longitudinal HIV RNA measurements taken on 320 subjects from the Acute Infection and Early Disease Research Program (AIEDRP), a large multicenter observational established to develop and evaluate data from studies of patients with acute or recent HIV infection. During the acute stage of infection, the large HIV RNA observations may lie above the limit of quantification of the assay, which we treat as right-censoring. This limit was between $75,000$ and $500,000$ copies/milliliter, depending on the assay. The time of infection was estimated at 24 days prior to first positive HIV RNA sample or detectable serum p24 antigen test. The subjects had between one and 14 observations: 129 had one, 82 had two, and 109 had three or more observations. Of the 830 recorded observations, 185 (22%) were above the limit of quantification of the

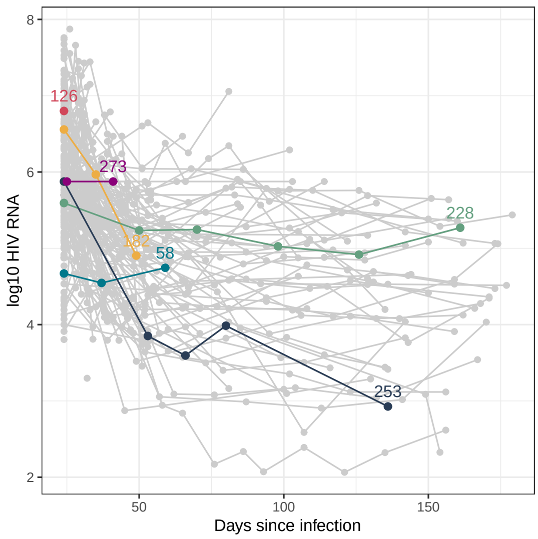


Figure 3 – **AIEDRP data**. Individual profiles (in $\log_{10}$ scale) for HIV viral load at different follow-up times.

assay. In the absence of treatment, following acute infection the HIV RNA decreases and then varies around a setpoint value. This setpoint value may differ between individuals, and is of central interest here. The viral setpoint characterizes the severity of infection, it may relate to the strength of the subject's immune system, and it may predict clinical progression of the disease. The individual profiles are shown in Figure 3.

1.4.4 UTI data

The UTI data is referred to a study of 72 children and adolescents who had HIV-1 infection and stopped their medications at 4 academic centers in the United States between January 2000 and September 2004. An unstructured treatment interruption (UTI) is an issue in the adolescent population, because the potential alternative of suboptimal adherence can lead to antiretroviral resistance and diminished treatment options in the future; however, there is little information on the clinical, virologic, and immunologic outcomes of UTI in pediatric and adolescent populations. The aim of this study was to monitor the HIV-1 viral laod (RNA) after unstructured treatment interruption. The subjects in the study had taken ARV therapy for at least 6 months before UTI, and the medication was discontinued for more than 3 months. The HIV viral load were studied from the closest time points at 0, 1, 3, 6, 9, 12, 18, 24 months after UTI. The number of observations from baseline(month 0) to month 24 are 71, 62, 58, 57, 43, 34, 24, and 13, respectively. Out of 362 observations, 26(7%) observations were below the detection limits (50 or 400 copies/mL) and were left-censored at these values. The individual profiles are shown in Figure 4.

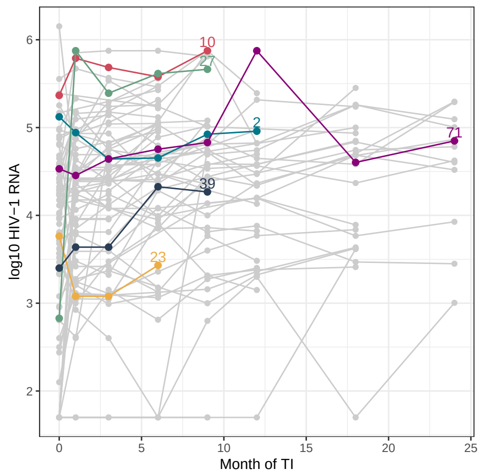


Figure 4 – **UTI data**. Individual profiles (in log10 scale) for HIV viral load at different follow-up times.

Chapter 2

A semiparametric mixed-effects model for censored longitudinal data

2.1 Introduction

Longitudinal studies are used in many fields of research, including epidemiology, clinical trials, and survey sampling. Parametric mixed-effects models are powerful tools to model the relationship between a response variable and covariates in longitudinal studies. LME models and nonlinear mixed-effects (NLME) models are the two most popular examples. These models have been extensively studied in the literature and applied to analyze longitudinal data (Davidian & Giltinan, 1995; Pinheiro & Bates, 2006; Diggle, 2002; Wu, 2010). One difficulty that arises in longitudinal data analysis is when the response is censored for some of the observations, that is, the measurements collected over time and the assay procedure may be subject to upper and lower detection limits. Typical examples of censored longitudinal data are from Human Immunodeficiency Virus (HIV) studies, where the detection of the viral load (the number or virus RNA copies) in the blood compartment is often limited by the sensitivity of a laboratory assay. With the advance of effective antiviral treatments, in some cases the HIV copy number can be extremely low and beyond the detection limit, which leads to left-censoring.

Several statistical approaches have been developed to deal with longitudinal data with censored measurements in the LME framework. Hughes (1999) proposed a likelihood method based on Monte Carlo EM (MCEM) for LME with censored responses (LMEC). Vaida & Liu (2009) proposed an EM algorithm to compute the maximum likelihood (ML) and restricted maximum likelihood (REML) for linear and nonlinear mixed effects models with censored responses (LMEC/NLMEC), which uses closed-form expressions at the E-step. Matos *et al.* (2013a) presented influence diagnostics and perturbation schemes for the LMEC and NLMEC models. For a robust estimation of longitudinal data in the presence of potential outliers or atypical observations, Pinheiro *et al.* (2001) proposed a robust extension of the LME model by considering a joint multivariate-t distribution for the random effects and within-subject errors, called the Student-t LME (t-LME) models. More recently, Matos *et al.* (2013b) developed an EM-type algorithm for computing the ML estimates for NLMEC/LMEC based on the multivariate Student's t-distribution, named t-NLMEC/t-LMEC. Furthermore, Lachos *et al.* (2019) proposed a flexible longitudinal LME model for multiple censored LME models based on the symmetric class of scale mixtures of normal (SMN) distributions, where an stochastic approximation of the EM (SAEM) algorithm is proposed to compute the ML estimates of the model parameters and to take into account the autocorrelation existing among irregular observations, and a damped exponential correlation (DEC) structure is considered.

Although LME models are useful tools for analyzing longitudinal data, an important assumption for these models is that the response variable is a known parametric function of both fixed effects and random effects. However, this assumption is not always satisfied in practical applications. To overcome this difficulty, a more general and robust modeling tool is needed, which motivates the development of nonparametric regression models (Green & Silverman, 1994; Wang, 1998a; Rice & Wu, 2001). Nonparametric models are more robust against the model assumptions but they are usually more complex and less efficient. Semiparametric models are a good compromise and retain nice features of both parametric and nonparametric models. In semiparametric models, the parametric components are often used to model important factors that affect the response and the nonparametric component is often used for nuisance factors. Semiparametric regression models for longitudinal data have gained increasing attention due to their flexible structure. For example, Zeger & Diggle (1994) proposed a semiparametric model where a nonparametric function is used to model the time effect, and a random intercept together with a Gaussian stochastic process is used to account for the within-subject correlation. Zhang *et al.* (1998) extended the Zeger & Diggle (1994) model to a more general class of models named semiparametric stochastic mixed models and proposed various stationary and nonstationary stochastic processes to model serial correlation. Vock *et al.* (2011) developed a mixed model framework for censored longitudinal data in which the random effects are represented by the flexible seminonparametric (SNP) density, and showed through simulations that this approach can lead to reduced bias and increased efficiency relative to assuming Gaussian random effects.

The literature contains many works about semiparametric models for longitudinal data, but to the best of our knowledge there are no studies of semiparametric mixed effects models for longitudinal irregularly observed censored data (SMEC). Motivated by this, the aim of this work is to perform a study of statistical inference in the SMEC model, in which the estimators of the regression coefficients and the nonparametric function of time are obtained using the EM algorithm for MPL estimation. A major challenge facing the penalized likelihood approach is estimation of the smoothing parameters (Wood, 2004).

There are several proposals to estimate this parameter, the most popular approaches being the generalized cross-validation (GCV), Akaike information criterion (AIC) and maximum restricted likelihood (REML). A drawback of these methods is that they are usually unstable and computationally expensive. In this article, as a byproduct of the EM algorithm, the smoothing parameter is estimated using a modified parametrization LME model, which is faster than those alternative methods. The autocorrelation existing among irregularly observed measures is modeled for the parametric damped exponential correlation (DEC) structure as proposed by Muñoz *et al.* (1992), this correlation structure allows us to deal with unequally spaced and unbalanced observations.

The rest of this chapter is structured as follows. Section 2.2 presents the multivariate normal and some of its keys properties. In Section 2.3, we introduce the SMEC model and the estimation and inference are outlined. Details of the EM algorithm as well as the derivation of the standard errors are also presented in this Section. A discussion about the estimation of the semiparametric degrees of freedom and the smoothing parameter are given in Section 2.4. Section 2.5 presents the results of simulation studies conducted to analyze the performance of the proposed methods. The analyses of two longitudinal datasets are presented in Section 2.6. Finally, some concluding remarks are given in Section 2.7 .

2.2 The multivariate normal

A random variable $\mathbf{Y}$ is said to follow a p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and variance matrix $\boldsymbol{\Sigma}$ (positive definite), denoted by $\mathrm{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$, if the probability density function (pdf) of $\mathbf{Y}$, is given by

$$\phi_p(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{p/2}}|\boldsymbol{\Sigma}|^{-1/2}\exp\left\{-\frac{1}{2}\left(\mathbf{y}-\boldsymbol{\mu}\right)^{\top}\boldsymbol{\Sigma}^{-1}\left(\mathbf{y}-\boldsymbol{\mu}\right)\right\},$$

where $\Phi_p(\cdot|\mathbf{a},\mathbf{A})$ and $\phi_p(\cdot|\mathbf{a},\mathbf{A})$ are the cdf and pdf, respectively, of $\mathrm{N}_p(\mathbf{a},\mathbf{A})$. In order to introduce some notation, for a normal random vector, we establish the following which is important for our subsequent research.

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

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

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

$$\mathbf{Y}_2|\mathbf{Y}_1=\mathbf{y}_1\sim N_{p_2}\left(\boldsymbol{\mu}_{2.1},\tilde{\boldsymbol{\Sigma}}_{22.1}\right),$$

where $\boldsymbol{\Sigma}_{22.1}=\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\mu}_{2.1}=\boldsymbol{\mu}_2+\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1-\boldsymbol{\mu}_1)$.

Now, let $\text{TN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbb{A})$ represent a p-variate truncated normal distribution for $\text{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ lying over the truncation region $\mathbb{A} = \{(y_1, \ldots, y_p) \in \mathbb{R}^p : a_1 \leqslant y_1 \leqslant b_1, \ldots, a_p \leqslant y_p \leqslant b_p\} = \{\mathbf{y} \in \mathbb{R}^p : \mathbf{a} \leqslant \mathbf{y} \leqslant \mathbf{b}\}$. Specifically, we say that the p-dimensional vector $\mathbf{Y} \sim \text{TN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbb{A})$, if its density is given by:

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbb{A}) = \frac{\phi_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\int_{\mathbb{A}} \phi_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) d\mathbf{y}} \mathbb{I}_{\mathbb{A}}(\mathbf{y}).$$

2.3 The semiparametric mixed effects model with censored responses

2.3.1 The model

Ignoring censoring for the moment, suppose that in a longitudinal study there are n subjects, with the ith subject having n_i observations over time. Denote by $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$ the vector of observed responses for the ith subject at time $\mathbf{t}_i = (t_{i1}, \ldots, t_{in_i})^\top$. The semiparametric mixed-effects model is specified as follows:

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

where $\mathbf{b}_i \overset{\text{iid.}}{\sim} \text{N}_q(\mathbf{0}, \mathbf{D})$ is independent of $\boldsymbol{\epsilon}_i \overset{\text{ind.}}{\sim} \text{N}_{n_i}(\mathbf{0}, \boldsymbol{\Omega}_i), i = 1, \ldots, n$, $\mathbf{X}_i$ is the $n_i \times p$ design matrix corresponding to the $p \times 1$ vector of fixed-effects $\boldsymbol{\beta}$, and $\mathbf{Z}_i$ is the $n_i \times q$ design matrix corresponding to the $q \times 1$ vector of random effects $\mathbf{b}_i$. Let $\mathbf{t}^0 = (t_1^0, \ldots, t_r^0)^\top$ be a vector of ordered distinct values of the time points t_{ij}. Then $\mathbf{N}_i$ is the incidence matrix $(n_i \times r)$ for the ith subject connecting $\mathbf{t}_i$ and $\mathbf{t}^0$ such that the (j, s)th element of $\mathbf{N}_i$ equals the indicator function $\mathbb{I}(t_{ij} = t_s^0)$ for $j = 1, \ldots, n_i$ and $s = 1, \ldots, r$, $\mathbf{f} = (f(t_1^0), \ldots, f(t_r^0))^\top$ is an $r \times 1$ vector, with $f(\cdot)$ an arbitrary twice-differentiable smooth function of time, and $\boldsymbol{\epsilon}_i$ is the $n_i \times 1$ vector of random errors. The dispersion matrix $\mathbf{D} = \mathbf{D}(\boldsymbol{\alpha})$ depends on the unknown and reduced parameter vector $\boldsymbol{\alpha}$ of dimension q. 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. Following Muñoz *et al.* (1992), as described in Section 1.3, we adopt a DEC structure for $\mathbf{E}_i$, which is defined as:

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

where $\boldsymbol{\phi} = \{(\phi_1, \phi_2) : 0 < \phi_1 < 1, \phi_2 \geqslant 0\}$, ϕ_1 is the correlation between observations separated by one t-unit in time and ϕ_2 is the "scale parameter", which permits attenuation or acceleration of the exponential decay of the autocorrelation function, defining a continuous-time autoregressive model. Examples of particular cases in this family of correlation structures include the compound symmetry (CS), AR(1), and MA(1) correlation structures when ϕ_2 takes the values 0,1, and ∞, respectively. A more detailed discussion of the DEC structure can be found in Muñoz *et al.* (1992).

Before proceeding further, we introduce some notation. Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{X} = (\mathbf{X}_1^\top, \ldots, \mathbf{X}_n^\top)^\top$, $\mathbf{N} = (\mathbf{N}_1^\top, \ldots, \mathbf{N}_n^\top)^\top$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^\top, \ldots, \boldsymbol{\epsilon}_n^\top)^\top$, and $\mathbf{Z} = \text{diag}(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$.

Then model (2.1) can be written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{f} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}, \tag{2.2}$$

where $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$ is $\mathrm{N}_{nq}(\mathbf{0}, \mathcal{D}(\boldsymbol{\alpha}))$, with $\mathcal{D}(\boldsymbol{\alpha}) = \mathrm{diag}(\mathbf{D}, \ldots, \mathbf{D})$ and $\boldsymbol{\epsilon}$ is $\mathrm{N}_N(\mathbf{0}, \boldsymbol{\Omega})$, with $\boldsymbol{\Omega} = \mathrm{diag}(\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_n)$. The matrix $[\mathbf{X}, \mathbf{N}\mathbf{T}]$ must be of full column rank, where $\mathbf{T} = [\mathbf{1}, \mathbf{t}^0]$ and $\mathbf{1}$ is an $r \times 1$ vector of 1's.

As mentioned earlier, the proposed model also considers censored observations, i.e., we assume that the response y_{ij} is not fully observed for all i, j. Let the observed data for the i-th subject be $(\mathbf{V}_i, \mathbf{C}_i)$, where $\mathbf{V}_i$ represents the vector of uncensored readings $(V_{ij} = V_{0i})$ or censoring interval (V_{1ij}, V_{2ij}), and $\mathbf{C}_i$ is the vector of censoring indicators, such that:

$$C_{ij} = \begin{cases} 1 & \text{if} \quad V_{1ij} \leqslant y_{ij} \leqslant V_{2ij}, \\ 0 & \text{if} \quad y_{ij} = V_{0i}, \end{cases} \tag{2.3}$$

for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, n_i\}$, i.e., $C_{ij} = 1$ if y_{ij} is located within a specific interval. Note that for a right-censored observation $V_{2ij} = \infty$, and for a left-censored observation $V_{1ij} = -\infty$. The model defined in (2.1)-(2.3) is henceforth called the DEC-SMEC model.

Notice that in the absence of the nonparametric function $\mathbf{N}_i\mathbf{f}$, $i = 1, \ldots, n$, in (2.1)-(2.3), the DEC-SMEC model reduces to the DEC-LMEC model proposed by Matos *et al.* (2016) (see also, Vaida & Liu, 2009), and in the absence of the random effects term $\mathbf{b}_i$ (other terms in (2.1) remaining intact), the DEC-SMEC model reduces to the well-known partial linear model (Ibacache-Pulgar *et al.*, 2013). Finally, when $\boldsymbol{\beta} = \mathbf{0}$, the DEC-SMEC model reduces to the nonparametric mixed model developed by Wang (1998a).

2.3.2 The log-likelihood function

Following Vaida & Liu (2009), frequentist inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \mathbf{f}^\top, \sigma^2, \boldsymbol{\alpha}^\top, \boldsymbol{\phi}^\top)^\top$ is based on the marginal distribution of $\mathbf{y}_i$. For complete data, we have marginally that $\mathbf{y}_i \overset{\text{ind.}}{\sim} \mathrm{N}_{n_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, where $\boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{N}_i\mathbf{f}$ and $\boldsymbol{\Sigma}_i = \boldsymbol{\Omega}_i + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top$, with the non-parametric component representing some fixed function. For responses with censoring pattern as in (2.3), we have $\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i \sim \mathrm{TN}_{n_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i; \mathbb{A}_i)$, where $\mathrm{TN}_{n_i}(.; \mathbb{A})$ denotes the truncated normal distribution on the interval $\mathbb{A}$, where $\mathbb{A}_i = A_{i1} \times \ldots \times A_{in_i}$, with A_{ij} being the interval $(-\infty, \infty)$ if $C_{ij} = 0$ and the interval $(V_{1ij}, V_{2ij}]$ if $C_{ij} = 1$. To compute the likelihood function associated with the model defined by (2.1) and (2.3), 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, $\mathbf{y}_i$, $\mathbf{V}_i$, $\boldsymbol{\mu}_i$, and $\boldsymbol{\Sigma}_i$ can be partitioned as follows:

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

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

$$\mathbf{y}_i^o \sim \text{N}_{n_i^o}(\boldsymbol{\mu}_i^o, \boldsymbol{\Sigma}_i^{oo}), \quad \mathbf{y}_i^c|\mathbf{y}_i^o \sim \text{N}_{n_i^c}(\boldsymbol{\mu}_{ico}, \mathbf{S}_i),$$

where $\boldsymbol{\mu}_{ico} = \boldsymbol{\mu}_i^c + \boldsymbol{\Sigma}_i^{co}(\boldsymbol{\Sigma}_i^{oo})^{-1}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o)$ and $\mathbf{S}_i = \boldsymbol{\Sigma}_i^{cc} - \boldsymbol{\Sigma}_i^{co}(\boldsymbol{\Sigma}_i^{oo})^{-1}\boldsymbol{\Sigma}_i^{oc}$. Thus, the likelihood function for subject i , using conditional probability arguments and following Vaida & Liu (2009) and Matos *et al.* (2013a), is given by:

$$\begin{aligned} L_i(\boldsymbol{\theta}) &= f(\mathbf{y}_i^o|\boldsymbol{\theta})P(\mathbf{V}_{1i}^c \leqslant \mathbf{y}_i^c \leqslant \mathbf{V}_{2i}^c|\mathbf{V}_i^o, \boldsymbol{\theta}) \\ &= \phi_{n_i^o}(\mathbf{y}_i^o; \boldsymbol{\mu}_i^o\boldsymbol{\beta}, \boldsymbol{\Sigma}_i^{oo}) \int_{\mathbf{V}_{1i}^c}^{\mathbf{V}_{2i}^c} \phi_p(\mathbf{y}_i^c; \boldsymbol{\mu}_{ico}, \mathbf{S}_i) d\mathbf{y}_i^c, \end{aligned} \tag{2.4}$$

where $\phi_p(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the probability distribution function (pdf) of the p-variate normal distribution $\text{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with mean vector $\boldsymbol{\mu}$ and covariate matrix $\boldsymbol{\Sigma}$.

The log-likelihood function for the observed data is thus given by $\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^{n}\{\log L_i\}$. Hence, the estimates obtained by maximizing the log-likelihood function $\ell(\boldsymbol{\theta})$ are the ML estimates. However, maximization of $\ell(\boldsymbol{\theta})$ without imposing restrictions on the function $\mathbf{f}(\cdot)$ may cause over-fitting and non-identification of $\boldsymbol{\beta}$ (see, for instance, Green, 1987). A well-known procedure that is based on the idea of log-likelihood penalization consists of incorporating a penalty function in the log-likelihood, such that:

$$\ell_p(\boldsymbol{\theta}, \lambda) = \ell(\boldsymbol{\theta}|\mathbf{y}) - \frac{\lambda}{2} J(\mathbf{f}), \tag{2.5}$$

where $J(\mathbf{f})$ denotes the penalty function over $\mathbf{f}(\cdot)$ and λ is a smoothing parameter that controls the tradeoff between goodness of fit and the smoothness of the estimated function. By maximizing (2.5), one obtains the MPL estimates.

2.3.3 The EM algorithm for MPL estimation

This Section describes in detail how the proposed DEC-SMEC model (2.1)-(2.3) can be fitted by using the ECM algorithm (Meng & Rubin, 1993).

Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$, $\mathbf{V} = \text{vec}(\mathbf{V}_1, \ldots, \mathbf{V}_n)$ and $\mathbf{C} = \text{vec}(\mathbf{C}_1, \ldots, \mathbf{C}_n)$, where $(\mathbf{V}_i, \mathbf{C}_i)$ is observed for the ith subject. In the estimation procedure, $\mathbf{b}$ and $\mathbf{Y}$ are treated as hypothetical missing data, and for augmentation with the observed data $\mathbf{V}$, $\mathbf{C}$, we set $\mathbf{y}_{\text{com}} = (\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{y}^\top, \mathbf{b}^\top)^\top$. Hence, the EM-type algorithm works over the complete-data log-likelihood function $\ell_c(\boldsymbol{\theta}|\mathbf{y}_{\text{com}}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}|\mathbf{y}_{\text{com}})$, where

$$\begin{aligned}\ell_i(\boldsymbol{\theta}|\mathbf{y}_{\text{com}}) &= -\frac{n_i}{2}\log\sigma^2 - \frac{1}{2}\log(|\mathbf{E}_i|) - \frac{1}{2\sigma^2}(\mathbf{y}_i - \boldsymbol{\mu}_i - \mathbf{Z}_i\mathbf{b}_i)^\top\mathbf{E}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i - \mathbf{Z}_i\mathbf{b}_i)\\ &- \frac{1}{2}\log|\mathbf{D}| - \frac{1}{2}\mathbf{b}_i^\top\mathbf{D}^{-1}\mathbf{b}_i + C,\end{aligned} \tag{2.6}$$

with C being a constant independent of 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:

$$\begin{aligned}Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) &= \mathbb{E}\left[\ell_c(\boldsymbol{\theta}|\mathbf{y}_{\text{com}})|\mathbf{V},\mathbf{C},\widehat{\boldsymbol{\theta}}^{(k)}\right]\\ &= \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})\\ &= \sum_{i=1}^{n} Q_{1i}(\boldsymbol{\beta},\mathbf{f},\sigma^2|\widehat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^{n} Q_{2i}(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}^{(k)}),\end{aligned} \tag{2.7}$$

where

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

and

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

with

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

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

It is clear that the E-step reduces only to the computation of

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

that is, the first and second moments of a truncated multivariate normal (MN) distribution. These expected values can be determined in closed form, using Proposition 1, as follows:

1. If the ith subject has only non-censored components, i.e, $\mathbf{y}_i = \mathbf{y}_i^o$ then

$$\widehat{\mathbf{y}_i\mathbf{y}_i^\top} = \mathbf{y}_i\mathbf{y}_i^\top, \quad \widehat{\mathbf{y}}_i = \mathbf{y}_i.$$

2. If the ith subject has only censored components, i.e, $\mathbf{y}_i = \mathbf{y}_i^c$, we have

$$\begin{aligned} \widehat{\mathbf{y}_i\mathbf{y}_i^\top} &= \mathbb{E}\left[\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right], \\ \widehat{\mathbf{y}}_i &= \mathbb{E}\left[\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right], \end{aligned}$$

   where $\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i \sim \mathrm{TN}_{n_i}(\widehat{\boldsymbol{\mu}}_i, \widehat{\boldsymbol{\Sigma}}_i; \mathbb{A}_i)$, $\widehat{\boldsymbol{\mu}}_i = \mathbf{X}_i\widehat{\boldsymbol{\beta}} + \mathbf{N}_i\widehat{\mathbf{f}}$, $\widehat{\boldsymbol{\Sigma}}_i = \widehat{\boldsymbol{\Omega}}_i + \mathbf{Z}_i\widehat{\mathbf{D}}\mathbf{Z}_i^\top$.

3. If the ith subject has censored and uncensored components, i.e, $\mathbf{y}_i = (\mathbf{y}_i^{c\top}, \mathbf{y}_i^{o\top})$. Then from Proposition 1 and by the fact that $\{\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i\}$, $\{\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i, \mathbf{y}_i^o\}$ and $\{\mathbf{y}_i^c|\mathbf{V}_i, \mathbf{C}_i, \mathbf{y}_i^o\}$ are equivalent processes, we have

$$\begin{aligned} \widehat{\mathbf{y}_i\mathbf{y}_i^\top} &= \mathbb{E}[\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}] = \begin{pmatrix} \mathbf{y}_i^o\mathbf{y}_i^{o\top} & \mathbf{y}_i^o\widehat{\mathbf{w}}_i^\top \\ \widehat{\mathbf{w}}_i\mathbf{y}_i^{o\top} & \widehat{\mathbf{w}_i\mathbf{w}_i^\top} \end{pmatrix}, \\ \widehat{\mathbf{y}}_i &= \mathbb{E}[\mathbf{y}_i|\mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}] = vec(\mathbf{y}_i^o, \widehat{\mathbf{w}}_i), \end{aligned}$$

   where $\widehat{\mathbf{w}}_i = \mathbb{E}[\mathbf{W}_i|\widehat{\boldsymbol{\theta}}]$, $\widehat{\mathbf{w}_i\mathbf{w}_i^\top} = \mathbb{E}[\mathbf{W}_i\mathbf{W}_i^\top|\widehat{\boldsymbol{\theta}}]$, with $\mathbf{W}_i \sim \mathrm{TN}_{n_i^c}(\boldsymbol{\mu}_{ico}, \mathbf{S}_i; \mathbb{A}_i)$, and $\mathbb{A}_i = A_{i1} \times \ldots \times A_{in_i}$, with A_{ij} being the interval $(-\infty, \infty)$ if $C_{ij} = 0$ and the interval $(V_{1ij}, V_{2ij}]$ if $C_{ij} = 1$.

For more details on the computation of these moments, see Vaida & Liu (2009) and Matos *et al.* (2013a). Alternatively, Kan & Robotti (2017) proposed an efficient algorithm to compute any arbitrary moment for the MN distribution. These can be obtained in the R package `MomTrunc` (Galarza *et al.*, 2020).

Following Green (1987), the MPL estimate of $\boldsymbol{\theta}$ is the value that maximizes the function

$$Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) - \frac{\lambda}{2}J(\mathbf{f}),$$

where $J(\mathbf{f})$ and λ are as defined in (2.5) and $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ is the conditional expectation function. Similarly to Ibacache-Pulgar *et al.* (2013), we will consider the following penalty function:

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

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

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

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}^{(k+1)} &= \left(\sum_{i=1}^n \mathbf{X}_i^\top \widehat{\mathbf{E}}_i^{-1(k)} \mathbf{X}_i\right)^{-1} \sum_{i=1}^n \mathbf{X}_i^\top \widehat{\mathbf{E}}_i^{-1(k)} \left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{N}_i \widehat{\mathbf{f}}^{(k)} - \mathbf{Z}_i \widehat{\mathbf{b}}_i^{(k)}\right), && (2.8)\\
\widehat{\mathbf{f}}^{(k+1)} &= \left(\sum_{i=1}^n \mathbf{N}_i^\top \widehat{\mathbf{E}}_i^{-1(k)} \mathbf{N}_i + \widehat{\sigma^2}^{(k)} \lambda \mathbf{K}\right)^{-1} \sum_{i=1}^n \mathbf{N}_i^\top \widehat{\mathbf{E}}_i^{-1(k)} \left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{X}_i \widehat{\boldsymbol{\beta}}^{(k+1)}\right. && (2.9)\\
&- \left.\mathbf{Z}_i \widehat{\mathbf{b}}_i^{(k)}\right),\\
\widehat{\sigma^2}^{(k+1)} &= \frac{1}{N} \sum_{i=1}^n \Big[\widehat{a}_i^{(k)} - 2(\mathbf{X}_i \widehat{\boldsymbol{\beta}}^{(k+1)} + \mathbf{N}_i \widehat{\mathbf{f}}^{(k+1)})^\top \widehat{\mathbf{E}}_i^{-1(k)} (\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i \widehat{\mathbf{b}}_i^{(k)})\\
&+ (\mathbf{X}_i \widehat{\boldsymbol{\beta}}^{(k+1)} + \mathbf{N}_i \widehat{\mathbf{f}}^{(k+1)})^\top \widehat{\mathbf{E}}_i^{-1(k)} (\mathbf{X}_i \widehat{\boldsymbol{\beta}}^{(k+1)} + \mathbf{N}_i \widehat{\mathbf{f}}^{(k+1)})\Big] && (2.10)\\
\widehat{\mathbf{D}}^{(k+1)} &= \frac{1}{n} \sum_{i=1}^n \widehat{\mathbf{b}_i \mathbf{b}_i^\top}^{(k)}, && (2.11)\\
\widehat{\boldsymbol{\phi}}^{(k+1)} &= \underset{\boldsymbol{\phi} \in (0,1) \times \mathcal{R}^+}{\arg\max} \left(-\frac{1}{2} \log(|\mathbf{E}_i|) - \frac{1}{2\widehat{\sigma^2}^{(k+1)}} \Big[\widehat{a}_i^{(k)} - 2\widehat{\boldsymbol{\mu}}_i^{(k+1)\top} \mathbf{E}_i^{-1} \left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i \widehat{\mathbf{b}}_i^{(k)}\right)\right.\\
&+ \left.\widehat{\boldsymbol{\mu}}_i^{(k+1)\top} \mathbf{E}_i^{-1} \widehat{\boldsymbol{\mu}}_i^{(k+1)}\Big]\right), && (2.12)
\end{aligned}
$$

where $N = \sum_{i=1}^n n_i$. This process is iterated until some distance between two successive evaluations of the actual penalized log-likelihood $\ell_p(\boldsymbol{\theta}, \lambda)$ in Section 2.3.2, such as $|\ell_p(\widehat{\boldsymbol{\theta}}^{(k+1)})/\ell_p(\widehat{\boldsymbol{\theta}}^{(k)}) - 1|$, becomes small enough, for example, $\epsilon = 10^-6$. A set of reasonable starting values may be achieved by computing $\widehat{\boldsymbol{\beta}}^{(0)}$, $\widehat{\sigma^2}^{(0)}$, $\widehat{\mathbf{D}}^{(0)}$ and $\widehat{\boldsymbol{\phi}}^{(0)}$ as the solution of the normal linear mixed-effects model, using the package `nlme` (Pinheiro *et al.*, 2020), and so, $\widehat{\mathbf{f}}^{(0)} = \left(\sum_{i=1}^n \mathbf{N}_i^\top \mathbf{N}_i + \widehat{\sigma^2}^{(0)} \lambda \mathbf{K}\right)^{-1} \sum_{i=1}^n \mathbf{N}_i^\top \left(\mathbf{y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}^{(0)}\right)$. In each iteration of the EM algorithm, the smoothing parameter, λ, can be estimated as described in Section 2.4.

2.3.4 Estimation of the random effects

To estimate the random effects, we consider the conditional mean of $\mathbf{b}_i$ given the observed data $\mathbf{V}_i$ and $\mathbf{C}_i$, that is, $\mathbb{E}[\mathbf{b}_i \mid \mathbf{V}_i, \mathbf{C}_i]$. Thus, for a given value of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \mathbf{f}^\top, \sigma^2, \boldsymbol{\alpha}^\top, \boldsymbol{\phi}^\top)^\top$, the conditional mean of $\mathbf{b}_i$ given $\mathbf{V}_i$ and $\mathbf{C}_i$ is

$$\widehat{\mathbf{b}}_i(\boldsymbol{\theta}) = \mathbb{E}[\mathbf{b}_i \mid \mathbf{V}_i, \mathbf{C}_i] = \boldsymbol{\varphi}_i(\widehat{\mathbf{y}}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{N}_i \mathbf{f}), \tag{2.13}$$

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 $\widehat{\mathbf{y}}_i = \mathbb{E}[\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i]$ is given by the first moment of a truncated MN distribution. In practice, the estimator of $\mathbf{b}_i$, $\widehat{\mathbf{b}}_i$, can be obtained by substituting the MPL estimate $\widehat{\boldsymbol{\theta}}$ into (2.13), leading to $\widehat{\mathbf{b}}_i = \widehat{\mathbf{b}}_i(\widehat{\boldsymbol{\theta}})$. Moreover, the conditional covariance matrix of $\mathbf{b}_i$ given $\mathbf{V}_i$ and $\mathbf{C}_i$ is

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

Note that $\text{Var}\left[\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i\right]$ can be easily obtained as a byproduct of the proposed ECM algorithm developed in Section 2.3.3.

2.3.5 Approximate standard errors

In the context of nonparametric regression, the covariance matrix of the MPL estimates can be evaluated by inverting the observed information matrix obtained by treating the penalized likelihood as a usual likelihood (Segal *et al.*, 1994). Within the framework of censoring, the variance of the parameter estimates can be obtained using the missing information principle (Louis, 1982), according to which:

observed information = complete information - missing information.

Following Segal *et al.* (1994) and Louis (1982), we derive the covariance matrix of $(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{f}})$ by using the inverse of the penalized observed information matrix. Thus, the approximate covariance matrix of $(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{f}})$ is given as:

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

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

$$\mathcal{I}_p(\boldsymbol{\beta}, \mathbf{f}) = \begin{pmatrix} \mathcal{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathcal{I}_{\boldsymbol{\beta}\mathbf{f}} \\ \mathcal{I}_{\boldsymbol{\beta}\mathbf{f}}^\top & \mathcal{I}_{\mathbf{ff}} \end{pmatrix}. \tag{2.14}$$

Thus, we obtain the variance of $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{f}}$ estimated at convergence, respectively, as:

$$\begin{aligned} \widehat{\text{Var}}_{\text{approx}}(\widehat{\boldsymbol{\beta}}) &= \left(\mathcal{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} - \mathcal{I}_{\boldsymbol{\beta}\mathbf{f}}\mathcal{I}_{\mathbf{ff}}^{-1}\mathcal{I}_{\boldsymbol{\beta}\mathbf{f}}^\top\right)\Big|_{\widehat{\boldsymbol{\theta}}}, \\ \widehat{\text{Var}}_{\text{approx}}(\widehat{\mathbf{f}}) &= \left(\mathcal{I}_{\mathbf{ff}} - \mathcal{I}_{\boldsymbol{\beta}\mathbf{f}}^\top\mathcal{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1}\mathcal{I}_{\boldsymbol{\beta}\mathbf{f}}\right)\Big|_{\widehat{\boldsymbol{\theta}}}, \end{aligned}$$

where

$$\begin{aligned} \mathcal{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} &= \sum_{i=1}^{n}\left\{\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i - \mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\text{Var}\left[\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i\right]\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\right\}, \\ \mathcal{I}_{\boldsymbol{\beta}\mathbf{f}} &= \sum_{i=1}^{n}\left\{\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i - \mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\text{Var}\left[\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i\right]\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i\right\}, \\ \mathcal{I}_{\mathbf{ff}} &= \sum_{i=1}^{n}\left\{\mathbf{N}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i - \mathbf{N}_i^\top\boldsymbol{\Sigma}_i^{-1}\text{Var}\left[\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i\right]\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i\right\} + \lambda^2\mathbf{K}\mathbf{f}\mathbf{f}^\top\mathbf{K}. \end{aligned}$$

Note that when $\mathbf{f} = \mathbf{0}$, we obtain the variance of the fixed effects in the approximate ML estimation given by Vaida & Liu (2009) and Hughes (1999).

2.4 Estimation of the smoothing parameter

The smoothing parameter λ has been assumed to be fixed, but in practice it should be estimated. Selecting a suitable value of λ is crucial to good curve fitting. A classic data-driven approach to selecting the smoothing parameter is cross-validation, which leaves out one subject's entire data at a time, but this approach is often computationally expensive. Likelihood-based smoothing parameter selection has been proposed as an alternative to prediction error-based approaches such as GCV or information criteria.

Several authors have shown the connection between a smoothing spline and a linear mixed-effects model for analysis of longitudinal data (see, for instance, Speed, 1991; Wang, 1998a). The authors treat the smoothing function as a linear combination of the fixed effects and random effects, so that the λ is variance component, which can be estimated by ML or restricted maximum likelihood (REML) (Wahba, 1985; Kohn *et al.*, 1991). Reiss & Ogden (2009) provided a theoretical comparison of GCV and REML with finite sample sizes, showing that GCV is prone to undersmoothing and is more likely to develop multiple minima and to give more variable λ estimates than REML. Zhang *et al.* (1998) formulated the semiparametric mixed model defined in (2.2) as a modified LME model and proposed to estimate the smoothing parameter λ and the variance component simultaneously using REML.

Following Green (1987) and Zhang *et al.* (1998), we can write $\mathbf{f}$ via a one-to-one linear transformation as:

$$\mathbf{f} = \mathbf{T}\boldsymbol{\delta} + \mathbf{B}\mathbf{d}, \tag{2.15}$$

where $\boldsymbol{\delta}$ and $\mathbf{d}$ are vectors with dimensions 2 and $r-2$, $\mathbf{B} = \mathbf{L}(\mathbf{L}^\top\mathbf{L})^{-1}$ and $\mathbf{L}$ is an $r \times (r-2)$ full-rank matrix satisfying $\mathbf{K} = \mathbf{L}\mathbf{L}^\top$ and $\mathbf{L}^\top\mathbf{T} = 0$. Given (2.15), Equation (2.2) can be reformulated as:

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

where $\mathbf{X}_* = [\mathbf{X}, \mathbf{NT}]$, $\mathbf{Z}_* = [\mathbf{NB}, \mathbf{Z}]$, $\boldsymbol{\beta}_* = (\boldsymbol{\beta}^\top, \boldsymbol{\delta}^\top)^\top$ are the regression coefficients and $\mathbf{b}_* = (\mathbf{d}^\top, \mathbf{b}^\top)^\top$ are mutually independent random effects with $\mathbf{d} \sim \mathrm{N}(\mathbf{0}, \frac{\sigma^2}{\lambda}\mathbf{I}_{r-2})$, and $\mathbf{b}$ and $\boldsymbol{\epsilon}$ have the same distributions as those given in Section 2.3.1.

Using the connection between the smoothing spline and the LME model, in this chapter we propose to estimate λ using the EM algorithm, due to its simplicity of implementation and stable monotone convergence.

This novel procedure is described as follows. Consider the following model:

$$\begin{aligned} \mathbf{y}|\mathbf{b}_* &\sim \mathrm{N}_N\left(\mathbf{X}_*\boldsymbol{\beta}_* + \mathbf{Z}_*\mathbf{b}_*, \boldsymbol{\Omega}\right), \\ \mathbf{b}_* &\sim \mathrm{N}_{(r-2+q)\times 1}\left(\mathbf{0}, \boldsymbol{\Psi}\right), \end{aligned}$$

where

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

In order to use the EM algorithm we consider the augmented data vector $\mathbf{y}_{comp*} = (\mathbf{y}^\top, \mathbf{b}_*^\top)^\top$, where $\mathbf{b}_*$ is assumed to be the missing variable. In this case, the log-likelihood function for the augmented data model, dropping all the terms that are not functions of λ, takes the form:

$$\ell(\lambda; \mathbf{y}_{comp*}) \propto -\frac{1}{2}\log|\mathbf{\Psi}| - \frac{1}{2}\mathbf{b}_*^\top \mathbf{\Psi}^{-1}\mathbf{b}_*.$$

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

Step 1: Obtain $\widehat{\boldsymbol{\theta}}^{(k+1)}$, as described in Subsection 2.3.3 ;

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

$$Q(\lambda|\widehat{\lambda}^{(k)}) = \mathbb{E}\left[\ell(\lambda; \mathbf{y}_{comp*})|\mathbf{y}, \widehat{\lambda}^{(k)}\right] = -\frac{1}{2}\log|\mathbf{\Psi}| - \frac{1}{2}\mathrm{tr}\left(\mathbf{\Psi}^{-1}\widehat{\mathbf{b}_*\mathbf{b}_*^\top}^{(k)}\right),$$

with $\widehat{\mathbf{b}_*\mathbf{b}_*^\top}^{(k)} = \mathbb{E}\left[\mathbf{b}_*\mathbf{b}_*^\top|\mathbf{y}, \widehat{\lambda}^{(k)}\right] = \mathbf{\Lambda}_i^* + \mathbf{\Lambda}_i^*\mathbf{Z}_i^{*\top}\mathbf{\Omega}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i^*\boldsymbol{\beta}^*)(\mathbf{y}_i - \mathbf{X}_i^*\boldsymbol{\beta}^*)^\top\mathbf{\Omega}_i^{-1}\mathbf{Z}_i^*\mathbf{\Lambda}_i^*$, $\mathbf{\Lambda}_i^* = (\mathbf{\Psi}^{-1} + \mathbf{Z}_i^{*\top}\mathbf{\Omega}_i^{-1}\mathbf{Z}_i^*)^{-1}$ (see, Matos *et al.*, 2013a);

Step 3: (M-step) Uptade λ by

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

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

2.4.1 Effective degrees of freedom for model selection

Degree of freedom is defined as approximately the number of effective parameters involved in modeling the nonparametric effects (Green & Silverman, 1994). It is useful in model selection criteria. Similarly to Tibshirani (1990), we can define the effective degrees of freedom as:

$$\begin{aligned} \mathrm{df}(\lambda) &= \mathrm{tr}\{\mathbf{N}\mathbf{S}_f\} \\ &= \mathrm{tr}\left\{\mathbf{N}\left(\mathbf{N}^\top\mathbf{E}^{-1}\mathbf{N} + \lambda\sigma^2\mathbf{K}\right)^{-1}\mathbf{N}^\top\mathbf{E}^{-1}\right\}, \end{aligned}$$

where $\mathbf{E} = \mathrm{diag}(\mathbf{E}_1, \ldots, \mathbf{E}_n)$.

The Akaike information criterion (AIC) is based on information theory and is useful for selecting an appropriate model given data with adequate sample size (Akaike, 1974). It is denoted by:

$$\mathrm{AIC} = -2\ell_p(\widehat{\boldsymbol{\theta}}, \lambda) + 2\{m + \mathrm{df}(\lambda)\},$$

where $\ell_p(\widehat{\boldsymbol{\theta}}, \lambda)$ denotes the penalized log-likelihood function available at $\widehat{\boldsymbol{\theta}}$ for a fixed λ and m is the number of parameters estimated $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\alpha}, \boldsymbol{\phi})$.

2.5 Simulation studies

In order to examine the performance of our proposed models and algorithm, we present two simulation studies. The first one investigates the performance of the MPL estimates of SMEC models for different correlation structures and the second one shows the asymptotic behavior of the MPL estimates as well as the consistency of the proposed standard errors of the MPL estimates.

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

For the simulations, we considered a DEC-SMEC model as defined in (2.1)-(2.3). We simulated data from the model

$$y_{ij} = \beta_1 x_{1_{ij}} + \beta_2 x_{2_{ij}} + f(t_{ij}) + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij}, \tag{2.16}$$

with $i = 1, \ldots, n$, $j = 1, \ldots, n_i$, $(b_{0i}, b_{1i}) \overset{\text{ind.}}{\sim} \mathrm{N}(\mathbf{0}, \mathbf{D})$, and $\epsilon_{ij} \overset{\text{ind.}}{\sim} \mathrm{N}_{n_i}(\mathbf{0}, \boldsymbol{\Omega}_i)$. The parameters were set at $\boldsymbol{\beta}^\top = (\beta_1, \beta_2) = (2, -1.5)$, $\sigma^2 = 0.55$, and $\mathbf{D}$ with elements $\alpha_{11} = 0.25$, $\alpha_{12} = 0.1$, and $\alpha_{22} = 0.2$. The values $\mathbf{x}_i^\top = (x_1, x_2)$ were generated independently from a uniform distribution in the intervals (0,1) and (-1,2), respectively, and those values were kept constant throughout the experiment.

In simulation studies, we generate samples with random censoring at a certain rate. For example, for a sample with 10% of censored observations: First, we generate $\mathbf{y}_i \overset{\text{ind.}}{\sim} \mathrm{N}_{n_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $i = 1, \ldots, n$. Then, we take the value of the 10th percentile of the sample, this value becomes the limit of censorship. Therefore, all values below the 10th percentile are now censored and their values are the censorship limit. Thus, we generated a random sample with 10% of censored observations.

2.5.1 Simulation study 1

For the first study, we simulated several datasets considering different values of the parameter ϕ_1 under the correlation structure AR(1). This study aims to discover the effect of the correlation level on the estimates. We chose a smooth function $f(t_{ij}) =$

2sin($0.25\pi t_{ij}$), where $t_{ij} = (4, 6, 9, 10, 11, 12)$, and for each value of ϕ_1, we simulated 500 datasets with sample sizes $n = 100$. In addition, we considered 10% and 20% of left-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) values as well as the estimates of the model parameters were stored for each simulation. To evaluate the estimates obtained from the DEC-SMEC model with different values of ϕ_1, we compared the bias (Bias) for each parameter over the 500 replicates. It is defined as:

$$\text{Bias}(\theta_k) = \frac{1}{500}\sum_{j=1}^{500}(\widehat{\theta}_k^{(j)} - \theta_k),$$

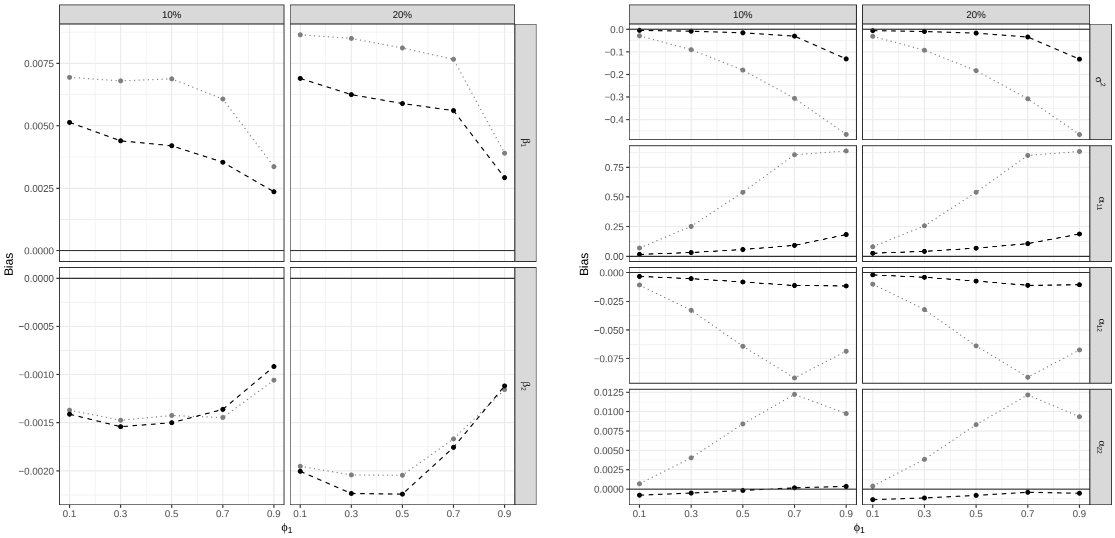
where $\widehat{\theta}_k^{(j)}$ is the estimate of θ_k from the jth sample for $j = 1, \ldots, 500$.

Table 1 has the summary statistics of the AIC values based on 500 replicates, where it can be observed that the model selection criterion chose the true model (AR(1)) in all 500 replicates when the parameter ϕ_1 is equal to or greater than 0.5. When $\phi_1 = 0.1$, the estimation of the AR(1) and UNC model is confounded.

Table 1 – **Simulation study 1**. Summary statistics of AIC values based on 500 simulated AR(1). The percentage of times that the AR (1) model was selected is in parentheses.

Censoring level		Structure	ϕ_1 = 0.1	0.3	0.5	0.7	0.9
10% censored	Mean	UNC	1772.426	1732.959	1663.156	1520.367	1124.027
		AR(1)	1772.360	1724.511	1639.541	1477.433	1065.243
	SD	UNC	32.8789	32.9188	33.9434	35.0294	36.0151
		AR(1)	32.8677	32.5091	32.7543	32.776	32.8190
			(**35%**)	(**96.8%**)	(**100%**)	(**100%**)	(**100%**)
20% censored	Mean	UNC	1637.034	1600.819	1537.841	1409.838	1058.398
		AR(1)	1637.119	1593.437	1516.984	1372.072	1007.159
	SD	UNC	30.7722	31.3998	32.2036	33.1929	34.1557
		AR(1)	30.8306	31.1223	31.2321	31.2826	31.6325
			(**31.8%**)	(**94%**)	(**100%**)	(**100%**)	(**100%**)

Figures 5a - 5b show the bias of each parameter estimated over 500 samples, considering 10 and 20% censored observations and in the uncorrelated and AR(1) models. Figure 5a shows that the correlation parameter did not affect the fixed-effects estimates, resulting in bias less than 0.009 for both levels of censoring. In Figure 6, for the both levels of censoring, the biases in the UNC structure are greater than those obtained in the AR(1) structure. In addition, as the value of the correlation parameter increases (ϕ_1) the bias in the UNC structure increases. It can also be noted that the estimates in both models were underestimated. With respect to the components of the variance (Figure 5b), it can be observed that the biases of parameter estimates (σ^2, α_{11}, α_{12} and α_{22}) in the AR(1) structure are lower than those obtained in the UNC structure for different values of ϕ_1. Also, the biases of the parameters in the UNC structure increase as the correlation parameter increases for the both levels of censoring. The biases in the AR(1) structure are smaller and are little affected by the correlation parameter. Therefore, we can conclude from this study that when the true model is fitted to the dataset, the correlation level does not affect the parameter estimates.



(a) (b)

Figure 5 – **Simulation study 1**. Bias of $\boldsymbol{\beta}$, σ^2 and $\boldsymbol{\alpha}$ estimates in the uncorrelated (gray, dotted line) and AR(1) (black, dashed line) structure for 5 different values of ϕ_1. First column: 10% of censored observations; second column: 20% of censored observations.

2.5.2 Simulation study 2

In this simulation study, the main focus is to evaluate the finite-sample performance of the MPL estimates of the regression coefficients and the nonparametric function with the smoothing parameter estimated using the proposed EM algorithm. Another goal is to examine the consistency of the standard errors for the MPL estimates of $\boldsymbol{\beta}$ and $\mathbf{f}$. In this study, parameter values were the same as mentioned in (2.16).

For this purpose, the left censoring proportion was fixed at 15% and sample sizes at $n = 60, 100, 200,$ and 400 were considered. For each sample size, we generated 500 samples of the DEC-SMEC model considering an AR(1) structure with parameter $\phi_1 = 0.6$. For this study, we chose a function $f(t_{ij}) = \cos(\pi\sqrt{t_{ij}})$, with $t_{ij} = (2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12)$. To evaluate the computational accuracy, we computed:

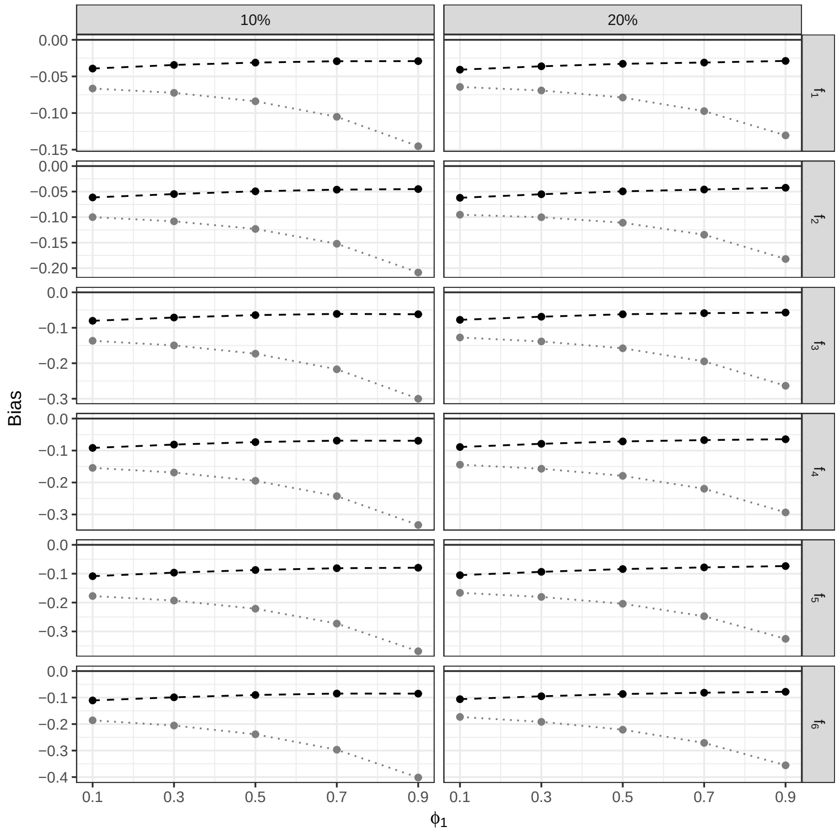


Figure 6 – **Simulation study 1**. Bias of $\mathbf{f}$ estimates in the uncorrelated (gray, dotted line) and AR(1) (black, dashed line) structure for 5 different values of ϕ_1. First column: 10% of censored observations; second column: 20% of censored observations.

- The arithmetic average of estimates:

$$\text{MC M}(\widehat{\theta}_k) = \frac{1}{500}\sum_{j=1}^{500} \widehat{\theta}_k^{\,(j)},$$

where $\widehat{\theta}_k^{(j)}$ is the estimate of θ_k from the j-th sample for $j = 1, \ldots, 500$.

- The absolute errors of estimates:

$$\text{MAE}(\widehat{\theta}_k) = \frac{1}{500}\sum_{j=1}^{500} |\widehat{\theta}_k^{(j)} - \theta_k|.$$

- The average values of the approximate standard errors obtained through the method described in Section 4.3.4 (MC SE).

- The standard deviations estimates of $\boldsymbol{\beta}$ and $\mathbf{f}$:

$$\text{MC SD}(\widehat{\theta}_i) = \text{sd}(\widehat{\theta}_k^{\,(1)}, \ldots, \widehat{\theta}_k^{\,(500)}).$$

- The coverage probability (CP), that is, the proportion of times that the 95% confidence interval of the estimated contains the true value:

$$CP(\widehat{\theta}_k) = \frac{1}{500}\sum_{j=1}^{500}\mathbb{I}(\theta_k \in [95\%\text{CI of } \widehat{\theta}_k^{(j)}]).$$

Evaluation of the parametric components

Table 2 has summarizes the simulation results for the parametric components of the model. It can be observed that the MAE tend to zero when n increases and the MC Mean approaches the true value of the parameter. It can also be seen that the estimation method of the standard errors (MC SE, Section 2.3.4) provides relatively close results for the standard deviation estimates (MC SD), suggesting that the derived standard errors work well. Figures 7 and 8 show the bias $(\widehat{\theta}_i^{(j)} - \theta_i, j = 1, \ldots, 500)$ of the estimates for each simulated dataset. It can be noted that as the sample size increases from 60 to 400, the range of the biases of the estimates becomes narrower, as expected. Also, $\widehat{\beta}_1$ and $\widehat{\beta}_2$ estimates are more accurate than estimates of the other parameters. Therefore, the results for the parametric components of the model indicate that the MPL estimates of the SMEC model do provide good asymptotic properties and that the proposed approximate standard errors are reliable.

Table 2 – **Simulation study 2**. Summary statistics based on 500 simulated AR(1) samples for the parametric components.

Parameter	$n = 60$					$n = 100$				
	MC M	MAE	MC SE	MC SD	CP	MC M	MAE	MC SE	MC SD	CP
$\beta_1 = 2$	2.0026	0.0623	0.0777	0.0804	93.2%	1.9997	0.0498	0.0608	0.0615	96.2%
$\beta_2 = -1.5$	-1.5001	0.0208	0.0245	0.0256	94.4%	-1.5017	0.0161	0.0206	0.0201	96.4%
$\sigma^2 = 0.55$	0.5394	0.0672				0.5403	0.0513			
$\alpha_{11} = 0.25$	0.2602	0.1342				0.2654	0.1188			
$\alpha_{12} = 0.1$	0.0890	0.0449				0.0932	0.0367			
$\alpha_{22} = 0.2$	0.1988	0.0330				0.1969	0.0242			
$\phi_1 = 0.6$	0.5907	0.0495				0.5918	0.0384			
	$n = 200$					$n = 400$				
	MC M	MAE	MC SE	MC SD	CP	MC M	MAE	MC SE	MC SD	CP
$\beta_1 = 2$	1.9998	0.0354	0.0430	0.0430	96.4%	2.0003	0.0256	0.0300	0.0320	92.6%
$\beta_2 = -1.5$	-1.5002	0.0117	0.0146	0.0147	94.4%	-1.5000	0.0086	0.0101	0.0107	94.6%
$\sigma^2 = 0.55$	0.5445	0.0384				0.5497	0.0276			
$\alpha_{11} = 0.25$	0.2555	0.0840				0.2515	0.0641			
$\alpha_{12} = 0.1$	0.0960	0.0258				0.0984	0.0191			
$\alpha_{22} = 0.2$	0.1984	0.0179				0.1984	0.0131			
$\phi_1 = 0.6$	0.5953	0.0280				0.5987	0.0199			

Evaluation of the nonparametric component

One of the principal objectives of this simulation is to see the effect of the nonparametric component, $f(t_{ij})$, and the performance of the smoothing parameter selected.

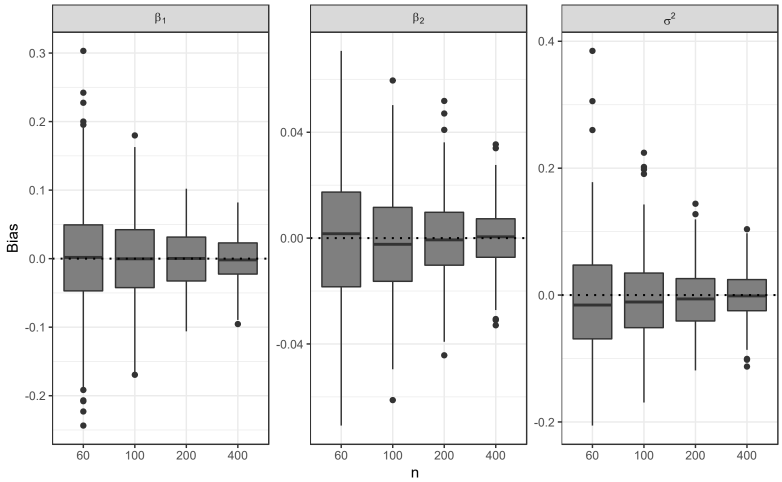


Figure 7 – **Simulation study 2**. Box-plots of the biases of $\boldsymbol{\beta}$ and σ^2 estimates.

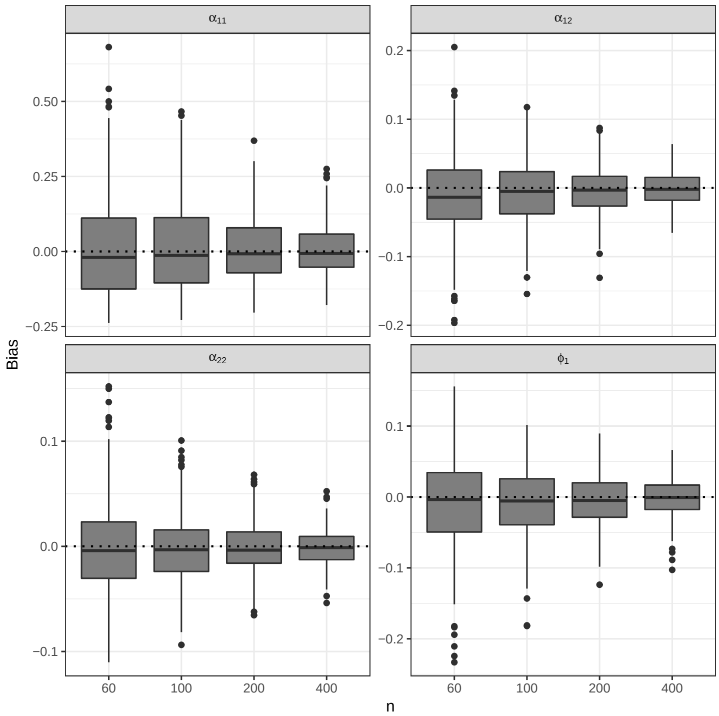


Figure 8 – **Simulation study 2**. Box-plots of the biases of $\boldsymbol{\alpha}$ and ϕ_1 estimates.

In practice, the estimates of the nonparametric component provide useful information.

Table 3 summarizes some simulation results from the estimation of $\mathbf{f}$ for $n = 60$, 100, 200 and 400. It can be observed that the MAE in the estimated nonparametric function is small for the first six components, and it becomes even smaller with increasing sample size. The MAE is relatively higher at values of t where the variance is larger, which may be due to the fact that the nonparametric function has large curvature at these points. The approximate standard errors (MC SE) obtained in Section 2.3.5 and the standard deviation estimates (MC SD) closely agree with each other, suggesting that the derived standard errors work well.

To investigate the accuracy of estimating the nonparametric function $f(t_{ij}) = \cos(\pi\sqrt{t_{ij}})$, the true shape of this function is plotted in Figure 9 with the 500 fitted curves and with the average estimates with four sample sizes. From Figure 9 it can be observed that the variability among the estimates of the nonparametric function declines as the sample size increases, and in addition, we can note that the shape of the average estimates of $f(t_{ij})$ is very close to the true function for all sample sizes. This is an indication of consistency of the nonparametric estimator as well as the capacity of the estimated smoothing parameter (discussed in Section 2.4) to capture the true unknown function.

Table 3 – **Simulation study 2**. Summary statistics based on 500 simulated AR(1) samples for the nonparametric components.

Parameter	$n = 60$					$n = 100$				
	MC M	MAE	MC SE	MC SD	CP	MC M	MAE	MC SE	MC SD	CP
$f_1(2) = -0.2663$	-0.2530	0.1442	0.1814	0.1802	95.6%	-0.2508	0.1180	0.1416	0.1451	94.4%
$f_2(3) = 0.6661$	0.6558	0.1754	0.2224	0.2198	95%	0.6697	0.1451	0.1731	0.1825	93.6%
$f_3(4) = 1$	0.9933	0.2134	0.2728	0.2689	94.6%	1.0078	0.1787	0.2120	0.2235	93.4%
$f_4(5) = 0.7374$	0.7444	0.2493	0.3288	0.3144	95.6%	0.7532	0.2124	0.2553	0.2651	94%
$f_5(6) = 0.1580$	0.1781	0.2954	0.3850	0.3718	94.4%	0.1820	0.2484	0.2987	0.3104	94.8%
$f_6(7) = -0.4421$	-0.4150	0.3343	0.4406	0.4223	95%	-0.4127	0.2811	0.3416	0.3517	94%
$f_7(8) = -0.8552$	-0.8322	0.3704	0.4964	0.4670	95.4%	-0.8242	0.3103	0.3847	0.3892	94.2%
$f_8(9) = -1$	-0.9724	0.4102	0.5531	0.5176	95.2%	-0.9665	0.3435	0.4285	0.4303	94.8%
$f_9(10) = -0.8728$	-0.8494	0.4551	0.6103	0.5740	95%	-0.8361	0.3823	0.4727	0.4791	94.8%
$f_{10}(11) = -0.5447$	-0.5242	0.4932	0.6677	0.6262	94.6%	-0.5030	0.4176	0.5171	0.5265	94.8%
$f_{11}(12) = -0.1125$	-0.0950	0.5329	0.7269	0.6807	94.6%	-0.0716	0.4502	0.5628	0.5670	94.8%
	$n = 200$					$n = 400$				
	MC M	MAE	MC SE	MC SD	CP	MC M	MAE	MC SE	MC SD	CP
$f_1(2) = -0.2663$	-0.2545	0.0802	0.1006	0.0989	95%	-0.2607	0.0622	0.0714	0.0769	93.8%
$f_2(3) = 0.6661$	0.6757	0.1017	0.1230	0.1274	94.2%	0.6720	0.0753	0.0872	0.0943	92.2%
$f_3(4) = 1$	1.0137	0.1266	0.1506	0.1584	93.2%	1.0095	0.0916	0.1067	0.1132	94%
$f_4(5) = 0.7374$	0.7567	0.1497	0.1814	0.1869	94%	0.7488	0.1080	0.1286	0.1357	93.6%
$f_5(6) = 0.1580$	0.1844	0.1767	0.2123	0.2184	95%	0.1737	0.1259	0.1504	0.1575	94.8%
$f_6(7) = -0.4421$	-0.4089	0.1967	0.2428	0.2441	95%	-0.4225	0.1427	0.1720	0.1770	95%
$f_7(8) = -0.8552$	-0.8221	0.2202	0.2735	0.2728	94.6%	-0.8348	0.1590	0.1937	0.1973	95.6%
$f_8(9) = -1$	-0.9633	0.2468	0.3047	0.3075	94.4%	-0.9761	0.1799	0.2158	0.2212	94.8%
$f_9(10) = -0.8728$	-0.8301	0.2762	0.3361	0.3420	93.8%	-0.8430	0.2006	0.2380	0.2475	95.4%
$f_{10}(11) = -0.5447$	-0.4954	0.3013	0.3678	0.3741	94.8%	-0.5104	0.2199	0.2605	0.2707	95.6%
$f_{11}(12) = -0.1125$	-0.0604	0.3272	0.4003	0.4053	94.6%	-0.0761	0.2381	0.2835	0.2919	95.2%

Two additional simulation studies can be found in Appendix A. In Section A.1, the simulation study verifies the behavior of the proposed model for different sizes of time dimensions. In Section A.2, the study assesses the behavior of the proposed model when compared to others in the literature.

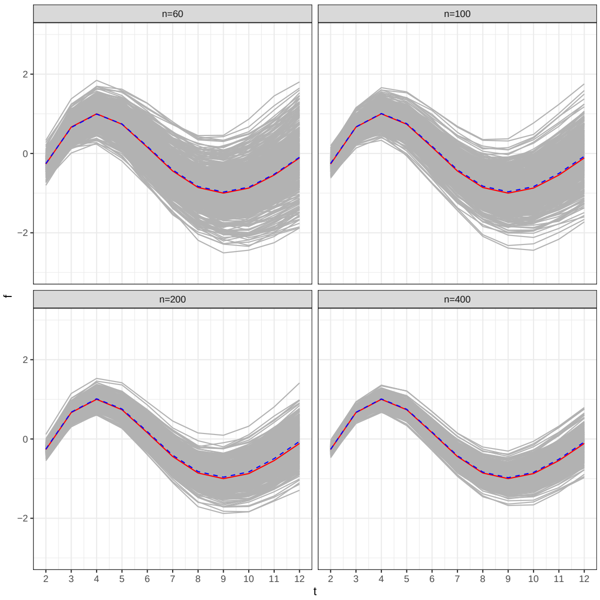


Figure 9 – **Simulation study 2**. Graphs of the nonparametric components with 500 replications. **Estimated curves** (gray lines), true curves (red lines) and the average estimates (blue lines, dashed).

2.6 Application

In this Section, we apply our proposed semiparametric linear mixed-effects model to the motivating ACTG 315 protocol HIV-1 RNA viral load dataset previously analyzed by Wu (2002). In HIV-AIDS research, it is hypothesized that the relationship between the viral load and the time of treatment within an antiviral regimen is nonlinear, whereas the relationship between the viral loads and certain immunological response such as $CD4^+$ cell count is linear. Since the viral load is recorded for patients at specific time points, mixed-effects models are typically used.

As mentioned in Section 1.4.1, the ACTG 315 dataset considers 46 HIV-1 infected patients treated with a potent ARV therapy. The viral load was measured on days 0, 2, 7, 10, 14, 28, 56, 84, 168 and 196 after start of treatment, with a total of 361 observations. Immunological markers known as $CD4^+$ cell counts were also measured along with viral load. Since one of our motivations is to investigate the relationship between virological and immunological responses in AIDS clinical trials, we consider the standardized version of $CD4^+$ cell count as a covariate for the parametric part of the model, whereas the time of treatment is modeled using splines. The predefined study day of viral load measurement (not the exact measured day) was used in our analysis.

We considered the following model:

$$y_{ij} = \mathrm{CD4}^+_{ij}\beta_1 + f(t_{ij}) + b_{0i} + b_{1i}t_{ij} + \epsilon_{ij} \tag{2.17}$$

where y_{ij} denotes the $\log_{10}$ transformation of the viral load for the ith subject at time t_{ij}

$(i = 1, 2, \ldots, 46\ ;\ j = 1, 2, \ldots, n_i)$, $f(t_{ij})$ is an arbitrary smoothing function, b_{0i}, b_{1i} are the random intercept and random slope, respectively for the i-th patient, and ϵ_{ij} are random errors. Following (2.1), one may express (2.17) in matrix form as:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{N}_i\mathbf{f} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \tag{2.18}$$

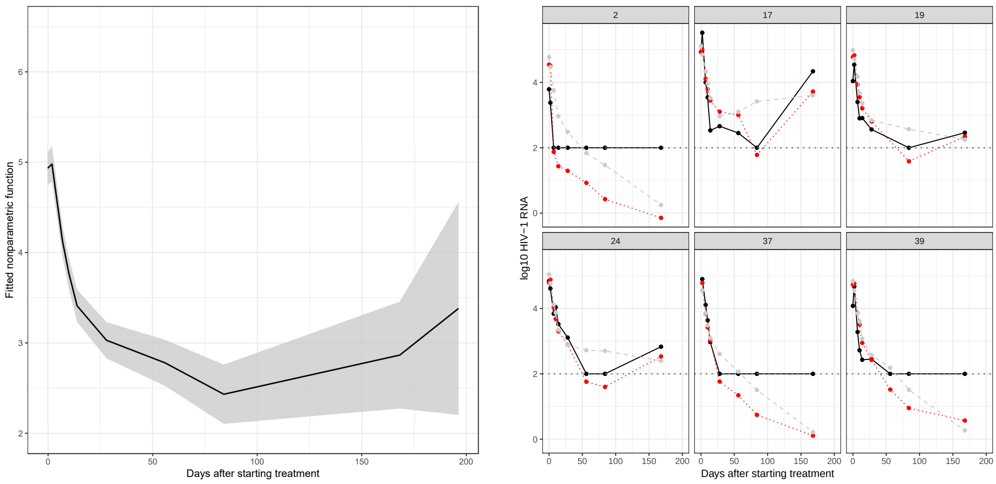
where $\mathbf{y}_i$ is an $(n_i \times 1)$ vector of responses for the i-th patient, $\mathbf{X}_i = [\mathrm{CD4}^+_{i1}, \ldots, \mathrm{CD4}^+_{in_i}]^\top$ where $\mathrm{CD4}^+_{ij}$ indicates a summary of the unobserved $\mathrm{CD4}^+$ values up to time t_{ij}, $\mathbf{N}_i$ is the incidence matrix, $\mathbf{f}$ is a (10×1) vector whose components are the function $f(\cdot)$ evaluated at the times in the set $\mathbf{t}^0 = (t^0_1 = 0, t^0_2 = 2, t^0_3 = 7, \ldots, t^0_{10} = 196)$, $\mathbf{Z}_i = [\mathbf{1}_{n_i}, \mathbf{t}_i]$, with $\mathbf{1}_{n_i}$ an $(n_i \times 1)$ vector of ones and $\mathbf{t}_i = [t_{i1}, \ldots, t_{in_i}]^\top$, $\mathbf{b}_i = (b_{0i}, b_{1i})^\top$ the random intercept and random slope, respectively, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})^\top$ represents the within-subject random error.

The MPL estimates of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \mathbf{f}^\top, \sigma^2, \boldsymbol{\alpha}^\top)^\top$, the smoothing parameter estimate (λ), the corresponding penalized log-likelihood function evaluated at $\widehat{\boldsymbol{\theta}}$ in the fitted models, and the values of AIC for the four models considered are presented in Table 4. These results reveal that the model with a DEC structure has lower AIC compared to the other structures. Moreover, and as expected, $\mathrm{CD4}^+$ cell counts are negatively correlated with HIV-1 RNA levels. The clinical interpretation is that as the count of CD4 cells increases, the immune systems of infected patients recover quickly and the viral load decreases with rapidly.

Table 4 – **ACTG 315 data**. Parameter estimates of the SMEC model (2.17) for the ACTG 315 data. SE indicates the standard errors.

	UNC		DEC		AR(1)		CS	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_1	-0.0703	0.0441	-0.0583	0.0392	-0.0617	0.0398	-0.0704	0.0441
f_1	4.9380	0.0918	4.9293	0.0912	4.9235	0.0952	4.9474	0.0926
f_2	4.9535	0.0778	4.9754	0.0996	4.9764	0.0945	4.9334	0.0834
f_3	4.1325	0.0842	4.1298	0.0870	4.1293	0.0898	4.1401	0.0829
f_4	3.7863	0.0833	3.7759	0.0867	3.7742	0.0900	3.7825	0.0821
f_5	3.4181	0.0893	3.4100	0.0904	3.4079	0.0928	3.4181	0.0875
f_6	3.0364	0.1009	3.0304	0.1017	3.0315	0.1022	3.0352	0.1005
f_7	2.7905	0.1269	2.7803	0.1294	2.7831	0.1286	2.7893	0.1268
f_8	2.4340	0.1647	2.4339	0.1666	2.4210	0.1666	2.4323	0.1647
f_9	2.9769	0.3025	2.8663	0.3008	2.8999	0.3034	2.9731	0.3024
f_{10}	3.4407	0.5585	3.3810	0.5995	3.3510	0.6102	3.4380	0.5531
σ^2	0.1449		0.2851		0.1991		0.2855	
α_{11}	0.2435		0.0507		0.1747		0.1034	
α_{12}	-0.0006		0.0008		-0.00003		-0.0006	
α_{22}	0.0001		0.0001		0.0001		0.0001	
ϕ_1			0.9		0.89		0.4914	
ϕ_2			0.6501		1		0	
λ	88.2971		63.7242		42.1648		174.4071	
loglikp	-275.481		-230.7881		-239.2762		-276.1796	
AIC	580.406		**495.4174**		510.4558		585.8651	

Figure 10a shows the nonparametric function for the ACTG data. From a medical point of view, it is important to identify when patients's RNA viral load levels decline and when this decline becomes slower and rebound occurs. From Figure 10a it can be observed that viral load decreases rapidly at the beginning of antiretroviral treatment and, after 84 days of therapy, viral load levels recover slightly. Clearly, the viral load changes with time in a nonlinear manner and the graph is U-shaped. Note that at later times the shaded area is larger due to the lower number of observations at these times. Figure 10b presents the estimated trajectories for 6 randomly chosen subjects after fitting the model given in (2.17) in the DEC structure. It is possible to see that the SMEC model provides better subject-specific estimated trajectories than the parametric DEC-NLMEC model. It can also be observed that in some individuals the DEC-NLMEC model overestimates the viral loads when observations are censored.



(a) (b)

Figure 10 – **ACTG 315 data**. (a) Fitted curve of nonparametric part. The shaded regions denote the 95% confidence intervals obtained by $\widehat{\mathbf{f}} \pm 1.96\sqrt{\widehat{\mathrm{Var}}(\widehat{\mathbf{f}})}$. (b) Viral loads in $\log_{10}$ scale (solid line) for 6 randomly chosen subjects and estimated trajectories (red, dotted line) for the SMEC model in the DEC structure. Gray line indicate the estimated trajectories in in the DEC-NLMEC model.

2.7 Conclusions

This chapter provides a theoretical framework for a semiparametric mixed model for longitudinal censored data, which can be considered a generalization of the normal linear/nonlinear mixed-effects models for censored data proposed by Matos *et al.* (2016) and Vaida & Liu (2009). We developed a method based on the EM algorithm to obtain MPL estimates of the regression coefficients of the parametric part and to estimate the nonparametric component as a natural cubic spline. We proposed the EM algorithm to estimate the smoothing parameter using a modification of the mixed model proposed by Green (1987). Simulation studies carried out suggest that the proposed method performs very well in estimating the parametric part and the nonparametric function. The variance of the MPL estimates can be estimated using the missing information principle, as described in Section 2.3.5. The approach was applied to analyze two HIV-AIDS studies, showing the advantage of the SMEC model to fit datasets with nonlinear subject-specific trajectories. The R code (R Core Team, 2020) is available from us upon request.

The multivariate normal assumption of SMEC models might not provide robust inference for the data which exhibit, even after being transformed, heavy-tailed and asymmetric features. It would thus also be interesting to consider a broader family of distributions such as the multivariate skew-normal distribution (Azzalini & Valle, 1996), the multivariate skew-t distribution (Azzalini & Genton, 2008) and the multivariate skew-normal/independent distributions (Lachos *et al.*, 2010), which could be more practical for the random effects and error terms.

Chapter 3

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

3.1 Introduction

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

Semiparametric mixed models have received great attention in the literature with approaches based on kernel smoothing (Zeger & Diggle, 1994), or, more often, on smoothing spline (Zhang *et al.*, 1998). However, these models (LME/NLME and semiparametric) are in general made on the assumption of Gaussian errors. Some studies have investigated alternative distributions for errors in LME/NLME, for example, Pinheiro *et al.* (2001) propose a robust hierarchical linear mixed model in which the random effects and the within-subject errors have a multivariate Student's t-distribution. Moreover, Meza *et al.* (2012) presented an extension of a Gaussian nonlinear mixed effects model considering a class of heavy tailed multivariate distributions for both random effects and residual errors. In the semiparametric context, Ibacache-Pulgar *et al.* (2012) extended semiparametric mixed linear models with normal errors to elliptical errors in order to permit distributions with heavier and lighter tails than the normal ones.

At the same time, longitudinal data can be complicated when the response is censored for some of the observations due to an assay detection limit used to quantify the marker. For example, this can occur when measuring the chemical content of a collection of samples (Palarea-Albaladejo & Martin-Fernandez, 2013), when measuring the concentration of some pollutants in environmental data (Helsel, 2011) or measuring Human Immunodeficiency Virus viral load in blood compartment (HIV RNA) (Hughes, 1999). Several methods have been proposed to deal with such limits of detection, censored mixed-effects models are frequently used in the analysis of longitudinal AIDS data. Lachos *et al.* (2011) considered a Bayesian treatment of the linear mixed model with censored responses (LMEC) and the nonlinear mixed model with censored responses (NLMEC) models based on the normal/independent distributions. Further, Matos *et al.* (2013b) developed a likelihood-based inference for LMEC and NLMEC based on the multivariate Student's t-distribution, named as t-LMEC and t-NLMEC.

The aim of this chapter is to consider the study of censored mixed-effects models using, simultaneously, semiparametric techniques such as smoothing splines and the multivariate Student's t-distribution, due to its capability of down-weighting out lying observations. This chapter is organized as follows. Section 3.2 describes the multivariate Student's t-distribution and some of its properties. In Section 3.3, the Student's-t semiparametric censored mixed-effects model is defined, where the estimation and inference procedures of the regression coefficients, nonparametric function, and scale parameter are presented. Results and discussions about the estimation of the smoothing parameter are given in Section 3.4. Moreover, in Section 3.5, the goodness of fit and model selection procedures are proposed to check the quality of fit. Simulation results are presented in Section 3.6 and an application to the data set of HIV viral loads is presented in Section 3.7. Finally, in Section 3.8 some concluding remarks are given with some future research directions.

3.2 The multivariate Student's t-distribution

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

A random variable $\mathbf{Y}$ having a p-variate Student's t-distribution with location vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$ (positive definite) and degrees of freedom ν ($\nu > 0$) denoted by $\mathbf{Y} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, has the probability density function (pdf):

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

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

$$\mathrm{T}_p(\mathbf{b}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \int_{-\infty}^{\mathbf{b}} t_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) d\mathbf{y} \tag{3.1}$$

and $\mathrm{T}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ is defined as

$$\mathrm{T}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \int_{\mathbf{a}}^{\mathbf{b}} t_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) d\mathbf{y}. \tag{3.2}$$

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

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

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

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

Proposition 2. *Let* $\mathbf{Y} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ *partitioned as* $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$, *with dim*$(\mathbf{Y}_1) = p_1$, *dim*$(\mathbf{Y}_2) = p_2$, *where* $p = p_1 + p_2$. *Let* $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top$ *and* $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$ *be the corresponding partions of* $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$. *Then, we have*

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

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

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

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

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

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

The following results provide the truncated moments of a Student's t random vector. The proofs of Proposition 3 and 4 are given in Matos *et al.* (2013b).

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

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

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

Proposition 4. *Let* $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; (\mathbf{a}, \mathbf{b}))$. *Consider the partition* $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ *with* $dim(\mathbf{Y}_1) = p_1$, $dim(\mathbf{Y}_2) = p_2$, $p_1 + p_2 = p$, *and the corresponding partitions of* $\mathbf{a}$, $\mathbf{b}$, $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$. *Then, under the notation of Proposition 2, the conditional* k*-th moment of* $\mathbf{Y}_2$ *is*

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

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

3.3 The Student-t semiparametric mixed effects model with censored responses

3.3.1 Model specification

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

$$y_{ij} = \mathbf{x}_{ij}^\top\boldsymbol{\beta} + f(t_{ij}) + \mathbf{z}_{ij}^\top\mathbf{b}_i + \epsilon_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n_i, \tag{3.3}$$

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

In order to write model (3.3) computationally more advantageous, we can express in a matrix form as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{N}_i\mathbf{f} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \tag{3.4}$$

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

In this work, we assume that the random effects and the errors follow a Student's t-distribution:

$$\begin{pmatrix} \mathbf{b}_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \overset{\text{ind.}}{\sim} t_{q+n_i}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_i \end{pmatrix}, \nu\right), i = 1, \ldots, n, \tag{3.5}$$

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

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

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

where ϕ_1 is the correlation between observations separated by one t-unit in time and ϕ_2 is the "scale parameter", which permits attenuation or acceleration of the exponential decay of the autocorrelation function, defining a continuous-time autoregressive model. Examples of particular cases in this family of correlation structures include the compound symmetry (CS), AR(1), and MA(1) - moving average of order 1, correlation structures when ϕ_2 takes the values 0,1, and ∞, respectively. A more detailed discussion of the DEC structure can be found in Muñoz *et al.* (1992) and in Section (1.3).

It follows that the semiparametric mixed model with Student's t-distribution assumes the following joint distribution:

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{b}_i \end{pmatrix} \overset{\text{ind.}}{\sim} t_{n_i+q}\left(\begin{pmatrix} \mathbf{X}_i\boldsymbol{\beta} + \mathbf{N}_i\mathbf{f} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top + \boldsymbol{\Omega}_i & \mathbf{Z}_i\mathbf{D} \\ \mathbf{D}\mathbf{Z}_i^\top & \mathbf{D} \end{pmatrix}, \nu\right). \tag{3.6}$$

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

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

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

As mentioned earlier, the proposed model also considers censored observations, i.e., we assume that the response y_{ij} is not fully observed for all i, j. Let the observed data for the i-th subject be $(\mathbf{V}_i, \mathbf{C}_i)$, where $\mathbf{V}_i$ represents the vector of uncensored readings $(V_{ij} = V_{0i})$ or censoring interval (V_{1ij}, V_{2ij}), and $\mathbf{C}_i$ is the vector of censoring indicators, such that:

$$C_{ij} = \begin{cases} 1 & \text{if} \quad V_{1ij} \leqslant y_{ij} \leqslant V_{2ij}, \\ 0 & \text{if} \quad y_{ij} = V_{0i}, \end{cases} \tag{3.7}$$

for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, n_i\}$, i.e., $C_{ij} = 1$ if y_{ij} is located within a specific interval. Note that for a right-censored observation $V_{2ij} = \infty$, and for a left-censored observation $V_{1ij} = -\infty$. The model defined in (3.3)-(3.7) is henceforth called the DEC-tSMEC model.

For responses with censoring pattern as in (3.7), we have that marginally

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

where $\mathrm{Tt}_{n_i}(.; \mathbb{A})$ denotes the truncated Student's t-distribution on the interval $\mathbb{A}$, $\mathbb{A}_i = A_{i1} \times \ldots \times A_{in_i}$, with A_{ij} being the interval $(-\infty, \infty)$ if $C_{ij} = 0$ and the interval $(V_{1ij}, V_{2ij}]$ if $C_{ij} = 1$.

3.3.2 The likelihood function

We are interested in maximum likelihood estimation of model (3.3) when $\mathbf{y}_i$ has a censored response. To compute the likelihood function associated with the model defined by (3.3)-(3.7), the first step is to treat separately the observed and censored components of $\mathbf{y}_i$. Let $\mathbf{y}_i^o$ be the n_i^o-vector of observed outcomes and $\mathbf{y}_i^c$ be the n_i^c-vector of censored observations for subject i with $(n_i = n_i^o + n_i^c)$, such that $C_{ij} = 0$ for all elements in $\mathbf{y}_i^o$ and 1 for all elements in $\mathbf{y}_i^c$. After reordering, $\mathbf{y}_i$, $\mathbf{V}_i$, $\boldsymbol{\mu}_i$, and $\boldsymbol{\Sigma}_i$ can be partitioned as follows:

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

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

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

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

where

$$\boldsymbol{\mu}_i^o = \mathbf{X}_i^o\boldsymbol{\beta} + \mathbf{N}_i^o\mathbf{f}, \quad \boldsymbol{\mu}_i^c = \mathbf{X}_i^c\boldsymbol{\beta} + \mathbf{N}_i^c\mathbf{f}, \quad \boldsymbol{\mu}_i^{co} = \boldsymbol{\mu}_i^c + \boldsymbol{\Sigma}_i^{co}\boldsymbol{\Sigma}_i^{oo-1}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o),$$

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

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

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

$$\begin{aligned} L_i(\boldsymbol{\theta}) = f(\mathbf{y}_i|\boldsymbol{\theta}) &= f(\mathbf{y}_i^o|\boldsymbol{\theta})P(\mathbf{V}_{1i}^c \leqslant \mathbf{y}_i^c \leqslant \mathbf{V}_{2i}^c|\mathbf{V}_i^o, \boldsymbol{\theta}) \\ &= t_{n_i^o}(\mathbf{V}_i^o; \boldsymbol{\mu}_i^o, \boldsymbol{\Sigma}_i^{oo}, \boldsymbol{\nu})\mathrm{T}_{n_i^c}(\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c; \boldsymbol{\mu}_i^{co}, \mathbf{S}_i^{co}, \nu + n_i^o) = L_i, \end{aligned} \tag{3.8}$$

where $\mathrm{T}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ is defined in (3.2).

Adopting the same idea of the work of Chapter 2, the log-likelihood function for the observed data is given by $\ell(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^{n} \log L_i$, and the estimates obtained by maximizing the log-likelihood function $\ell(\boldsymbol{\theta}|\mathbf{y})$ are the maximum likelihood estimates (MLEs). For the reason that $f(\cdot)$ is an infinite-dimensional parameter, the direct maximization of (3.8) without imposing restrictions over the function $f(\cdot)$ may cause overfitting and non-identifiability of $\boldsymbol{\beta}$ (see Green, 1987). A well-know procedure based on the idea of log-likelihood penalization and consists of incorporating a penalty function in the log-likelihood, such that

$$\ell_p(\boldsymbol{\theta}|\mathbf{y}) = \ell(\boldsymbol{\theta}|\mathbf{y}) - \frac{\lambda}{2}J(\mathbf{f}), \tag{3.9}$$

where $J(\mathbf{f})$ denotes the penalty function over $\mathbf{f}$ and $\lambda \geqslant 0$ is a smoothing parameter which controls the tradeoff between goodness of fit and the smoothness estimated function. By maximizing (3.9), one obtains the MPL estimate.

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

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

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

3.3.3 The EM algorithm for MPL estimation

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

The EM algorithm (Dempster *et al.*, 1977) is a popular iterative algorithm for ML estimation of models with incomplete data and has several appealing features such as stability of monotone convergence and simplicity of implementation. We adopt a variant of the the EM-type algorithm, called the ECME algorithm, for computing MPL estimates of model parameters. Liu & Rubin (1994) showed that ECME typically shares with EM the simplicity and stability, but has a faster rate of convergence, especially for multivariate Student's t-distribution with unknown degrees-of-freedom.

Based on the essential property of multivariate Student's t-distribution, the model (3.6) can be expressed in the following hierarchical model:

$$\begin{aligned} \mathbf{y}_i|\mathbf{b}_i, u_i &\overset{\text{ind.}}{\sim} \mathrm{N}_{n_i}(\boldsymbol{\mu}_i, u_i^{-1}\boldsymbol{\Omega}_i), \\ \mathbf{b}_i|u_i &\overset{\text{ind.}}{\sim} \mathrm{N}_q(\mathbf{0}, u_i^{-1}\mathbf{D}), \\ u_i &\overset{\text{ind.}}{\sim} \mathrm{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \end{aligned} \tag{3.10}$$

where Gamma(a, b) denotes the gamma distribution with mean a/b and variance a/b^2. Thus, it is possible to apply the penalized EM algorithm (Green, 1990) by assuming that $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)$, and $\mathbf{u} = (u_1, \ldots, u_n)^\top$ are hypothetical missing variables, and augmenting with the observed variables $(\mathbf{V}, \mathbf{C})$ where $\mathbf{V} = vec(\mathbf{V}_1, \ldots, \mathbf{V}_n)$, and $\mathbf{C} = vec(\mathbf{C}_1, \ldots, \mathbf{C}_n)$. Hence, the penalized log-likelihood function for the model based on complete data $\mathbf{y}_c = (\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{y}^\top, \mathbf{b}^\top, \mathbf{u}^\top)^\top$ is given by

$$\ell_{pc}(\boldsymbol{\theta}|\mathbf{y}_c) = \ell_c(\boldsymbol{\theta}|\mathbf{y}_c) - \frac{\lambda}{2}\mathbf{f}^\top\mathbf{K}\mathbf{f}, \tag{3.11}$$

with

$$\begin{aligned} \ell_c(\boldsymbol{\theta}|\mathbf{y}_c) &= \sum_{i=1}^{n}\Bigg[-\frac{n_i}{2}\log\sigma^2 - \frac{1}{2}\log(|\mathbf{E}_i|) - \frac{u_i}{2\sigma^2}(\mathbf{y}_i - \boldsymbol{\mu}_i - \mathbf{Z}_i\mathbf{b}_i)^\top\mathbf{E}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i - \mathbf{Z}_i\mathbf{b}_i) \\ &- \frac{1}{2}\log|\mathbf{D}| - \frac{u_i}{2}\mathbf{b}_i^\top\mathbf{D}^{-1}\mathbf{b}_i + \log h(u_i|\nu) + C\Bigg], \end{aligned} \tag{3.12}$$

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

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

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

where

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

and

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

with

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

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

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

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \left(\sum_{i=1}^n \widehat{u_i}^{(k)}\mathbf{X}_i^\top\widehat{\mathbf{E}}_i^{-1(k)}\mathbf{X}_i\right)^{-1}\sum_{i=1}^n \mathbf{X}_i^\top\widehat{\mathbf{E}}_i^{-1(k)}\left(\widehat{u_i\mathbf{y}_i}^{(k)} - \widehat{u_i}^{(k)}\mathbf{N}_i\widehat{\mathbf{f}}^{(k)} - \mathbf{Z}_i\widehat{u_i\mathbf{b}_i}^{(k)}\right) \tag{3.13}$$

$$\begin{aligned}
\widehat{\mathbf{f}}^{(k+1)} &= \left(\sum_{i=1}^n \widehat{u_i}^{(k)}\mathbf{N}_i^\top\widehat{\mathbf{E}}_i^{-1(k)}\mathbf{N}_i + \widehat{\sigma^2}^{(k)}\lambda\mathbf{K}\right)^{-1}\sum_{i=1}^n \mathbf{N}_i^\top\widehat{\mathbf{E}}_i^{-1(k)}\left(\widehat{u_i\mathbf{y}_i}^{(k)} - \widehat{u_i}^{(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k+1)}\right.\\
&\quad \left. - \mathbf{Z}_i\widehat{u_i\mathbf{b}}^{(k)}\right)
\end{aligned} \tag{3.14}$$

$$\widehat{\sigma^2}^{(k+1)} = \frac{1}{N}\sum_{i=1}^n\left[\widehat{a}_i^{(k)} - 2\widehat{\boldsymbol{\mu}}_i^{(k+1)\top}\mathbf{E}_i^{-1}\left(\widehat{u_i\mathbf{y}_i}^{(k)} - \mathbf{Z}_i\widehat{u_i\mathbf{b}_{ii}}^{(k)}\right) + \widehat{u_i}^{(k)}\widehat{\boldsymbol{\mu}}_i^{(k+1)\top}\mathbf{E}_i^{-1}\widehat{\boldsymbol{\mu}}_i^{(k+1)}\right] \tag{3.15}$$

$$\widehat{\mathbf{D}}^{(k+1)} = \frac{1}{m}\sum_{i=1}^n \widehat{u_i\mathbf{b}_i\mathbf{b}_i^\top}^{(k)} \tag{3.16}$$

$$\begin{aligned}
\widehat{\phi}^{(k+1)} &= \underset{\boldsymbol{\phi}\in(0,1)\times\mathcal{R}^+}{\arg\max}\left(-\frac{1}{2}\log(|\mathbf{E}_i|) - \frac{1}{2\widehat{\sigma^2}^{(k+1)}}\left[\widehat{a}_i^{(k)} - 2\widehat{\boldsymbol{\mu}}_i^{(k+1)\top}\mathbf{E}_i^{-1}\left(\widehat{u_i\mathbf{y}_i}^{(k)} - \mathbf{Z}_i\widehat{u_i\mathbf{b}_{ii}}^{(k)}\right)\right.\right.\\
&\quad \left.\left. + \widehat{u_i}^{(k)}\widehat{\boldsymbol{\mu}}_i^{(k+1)\top}\mathbf{E}_i^{-1}\widehat{\boldsymbol{\mu}}_i^{(k+1)}\right]\right)
\end{aligned} \tag{3.17}$$

$$\begin{aligned}
\widehat{\nu}^{(k+1)} &= \underset{\nu}{\arg\max}\left\{\sum_{i=1}^m \log \mathrm{T}_{n_i^c}\left(\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c; \boldsymbol{\mu}_i^{co(k+1)}, \mathbf{S}_i^{co(k+1)}, \nu + n_i^o\right)\right.\\
&\quad \left. + \sum_{i=1}^m \log t_{n_i^o}\left(\mathbf{V}_i^o; \boldsymbol{\mu}_i^{o(k+1)}, \boldsymbol{\Sigma}_i^{oo(k+1)}, \boldsymbol{\nu}\right)\right\},
\end{aligned} \tag{3.18}$$

where $N = \sum_{i=1}^n n_i$. The algorithm is iterated until a suitable convergence rule is satisfied, in this case, we adopt the distance involving two successive evaluations of the actual penalized log-likelihood. So, this process is iterated until some distance between two successive evaluations of the actual penalized log-likelihood $\ell_p(\boldsymbol{\theta},\lambda)$ in Section 3.3.2, such as $|\ell_p(\widehat{\boldsymbol{\theta}}^{(k+1)})/\ell_p(\widehat{\boldsymbol{\theta}}^{(k)}) - 1|$, becomes small enough, for example, $\epsilon = 10^-6$. A set of reasonable starting values may be achieved by computing $\widehat{\boldsymbol{\beta}}^{(0)}$, $\widehat{\sigma^2}^{(0)}$, $\widehat{\mathbf{D}}^{(0)}$ and $\widehat{\boldsymbol{\phi}}^{(0)}$ as the solution of the normal linear mixed-effects model, using the package `nlme` (Pinheiro *et al.*, 2020), and so, $\widehat{\mathbf{f}}^{(0)} = \left(\sum_{i=1}^{n} \mathbf{N}_i^\top \mathbf{N}_i + \widehat{\sigma^2}^{(0)} \lambda \mathbf{K}\right)^{-1} \sum_{i=1}^{n} \mathbf{N}_i^\top \left(\mathbf{y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}^{(0)}\right)$. In each iteration of the EM algorithm, the smoothing parameter, λ, can be estimated as described in Section 3.4.

It is important to stress that from equations (3.13) to (3.17), the E-step reduces to the computation of

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

that is, the first and second moments of a truncated multivariate Student's t-distribution. These expected values can be determined in closed form, using Propositions 3-4, as follows:

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

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

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

2. If the ith subject has only censored components then from Proposition 3, we have:

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

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

3. If the ith subject has censored and uncensored components and given that $(\mathbf{Y}_i | \mathbf{V}_i, \mathbf{C}_i)$, $(\mathbf{Y}_i | \mathbf{V}_i, \mathbf{C}_i, \mathbf{Y}_i^o)$, and $(\mathbf{Y}_i^c | \mathbf{V}_i, \mathbf{C}_i, \mathbf{Y}_i^o)$ are equivalent process, then from Proposition 4, we have

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

where $\mathbf{W}_i \sim \mathrm{Tt}_{n_i^c}(\boldsymbol{\mu}_i^{co}, \widetilde{\mathbf{S}}_i^{co}, \nu + n_i^o + 2, (\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c))$, $\widetilde{\mathbf{S}}_i^{co} = \left(\dfrac{\nu + \delta^2(\mathbf{y}_i^o)}{\nu + n_i^o + 2}\right) \mathbf{S}_i$ and $\mathbf{S}_i$, $\mathbf{S}_i^{co}$ and $\boldsymbol{\mu}_i^{co}$ are as in Section 3.3.2.

Formulas for $\mathbb{E}[\mathbf{W}]$ and $\mathbb{E}[\mathbf{W}\mathbf{W}^\top]$, where $\mathbf{W} \sim \mathrm{Tt}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$, have been recently developed using recurrence relations involving the density of multivariate Student's t-distribution. These can be obtained in the R package `MomTrunc` (Galarza *et al.*, 2020).

3.3.4 Estimation of the random effects

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

$$\widehat{\mathbf{b}}_i(\boldsymbol{\theta}) = \mathbb{E}\left[\mathbf{b}_i|\mathbf{V}_i, \mathbf{C}_i\right] = \boldsymbol{\varphi}_i(\widehat{\mathbf{y}}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{N}_i\mathbf{f}), \tag{3.19}$$

where $\boldsymbol{\varphi}_i$ is defined in Subsection 3.3.3 and $\widehat{\mathbf{y}}_i = \mathbb{E}[\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i]$ is the first moment of the truncated Student's t-distribution.

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

3.3.5 The expected information matrix

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

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

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

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

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

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

where

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

where $\mathbf{E}_1 = (\widehat{u_i\mathbf{y}_i} - \widehat{u}_i\boldsymbol{\mu}_i)(\widehat{u_i\mathbf{y}_i} - \widehat{u}_i\boldsymbol{\mu}_i)^\top$ and $\mathbf{E}_2 = (\widehat{u_i^2\mathbf{y}_i\mathbf{y}_i^\top} - \widehat{u_i^2\mathbf{y}_i}\boldsymbol{\mu}_i^\top - \boldsymbol{\mu}_i\widehat{u_i^2\mathbf{y}_i^\top} + \widehat{u_i^2}\boldsymbol{\mu}_i\boldsymbol{\mu}_i^\top)$. Note that $\mathbf{E}_1$ depend on the computation of $\widehat{u}_i$, $\widehat{u_i\mathbf{y}_i}$ that can be obtained in Subsection 3.3.3 and $\mathbf{E}_2$ depend on the following quantities

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

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

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

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

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

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

2. If the ith subject has only censored components then

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

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

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

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

where $d_p(n_i^o, \nu, 2) = \dfrac{(\nu + n_i)(n_i^o + \nu + 2)(n_i^o + \nu)}{n_i + \nu + 2}$, $\mathbf{W}_i \sim Tt_{n_i^c}(\boldsymbol{\mu}_i^{co}, \widetilde{\mathbf{S}}_i^{co}, \nu + n_i^o + 4; (\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c))$, $\widetilde{\mathbf{S}}_i^{co} = \left(\dfrac{\nu + \delta^2(\mathbf{y}_i^o)}{\nu + n_i^o + 4}\right)\mathbf{S}_i$ and $\mathbf{S}_i$, $\mathbf{S}_i^{co}$ and $\boldsymbol{\mu}_i^{co}$ are as in Subsection 3.3.2.

It can be noted that here we also need the first and second moments of truncated Student's t-distribution. And, as mentioned before, these moments can be obtained in the R package `MomTrunc` (Galarza *et al.*, 2020).

3.4 Estimation of the smoothing parameter

In the previous sections we considered the smoothing parameter λ fixed to make inference for the nonparametric function $\mathbf{f}$. However, in practice, this parameter need to be estimated from the data. Many authors have pointed out that the proper selection of smoothing parameters is essential for good a performance of the spline estimates (Green & Silverman, 1994).

Wahba & Wold (1975) examine how much the smoothing should be: If λ is too small, the spline is too wiggly and picks up too much noise (overfit); and if λ is too large, the spline is too smooth and the signal is lost (underfit). A classical data-driven approach to selecting the smoothing parameter is cross validation (CV), which leaves out one subject's entire data at a time, but this approach is often computationally expensive (Zeger & Diggle, 1994).

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

$$\mathbf{f} = \mathbf{T}\boldsymbol{\delta} + \mathbf{B}\mathbf{d}, \tag{3.20}$$

where $\boldsymbol{\delta}$ and $\mathbf{d}$ are vectors with dimensions 2 and $r-2$, $\mathbf{B} = \mathbf{L}(\mathbf{L}^\top\mathbf{L})^{-1}$ and $\mathbf{L}$ is an $r \times (r-2)$ full-rank matrix satisfying $\mathbf{K} = \mathbf{L}\mathbf{L}^\top$ and $\mathbf{L}^\top\mathbf{T} = 0$. Given (3.20), Equation (3.4) can be reformulated as:

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

where $\mathbf{X}_i^* = [\mathbf{X}_i, \mathbf{N}_i\mathbf{T}]$, $\mathbf{Z}_i = [\mathbf{N}_i\mathbf{B}, \mathbf{Z}_i]$, $\boldsymbol{\beta}^* = (\boldsymbol{\beta}^\top, \boldsymbol{\delta}^\top)^\top$ are the regression coefficients and $\mathbf{b}^* = (\mathbf{d}^\top, \mathbf{b}_i^\top)^\top$ are mutually independent random effects with $\mathbf{d} \sim t_{r-2}(\mathbf{0}, \dfrac{\sigma^2}{\lambda}\mathbf{I}_{r-2})$ and $\mathbf{b}_i$ and $\boldsymbol{\epsilon}_i$ have the same distributions as those given in Section 3.3.1.

Motivated by Zhang *et al.* (1998)'s results and using the connection between the smoothing spline and the linear mixed models, we propose to estimate λ using the

EM algorithm, due to its simplicity of implementation and stable monotone convergence. This novel procedure is described as follows. Consider the following model:

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

where

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

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

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

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

Step 1: Obtain $\widehat{\boldsymbol{\theta}}^{(k+1)}$, as described in Subsection 3.3.3;

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

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

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

Step 3: (M-step) Uptade λ by

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

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

3.5 Goodness of fit and model selection

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

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

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

and follows approximately a standard normal distribution. Thus, a Q-Q plot of the transformed distances, $F_i^{[z]}$, can be used to assess the fit of the multivariate Student's t-distribution.

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

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

where $\ell_p(\widehat{\boldsymbol{\theta}})$ corresponds to the logarithm of the penalized likelihood function, defined in Equation (3.9), p^* is the total number of parameters in the model, and N denotes the size of the sample.

3.6 Simulation studies

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

$$y_{ij} = \beta_1 x_{1ij} + \beta_2 x_{2_{ij}} + f(t_{ij}) + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij}, \quad i = 1, \ldots, n, \quad , j = 1, \ldots, n_i. \tag{3.21}$$

The parameters were set at $\boldsymbol{\beta}^\top = (\beta_1, \beta_2) = (2, -1.5)$, $\sigma^2 = 0.13$, and $\mathbf{D}$ with elements $\alpha_{11} = 0.25$, $\alpha_{12} = 0.01$, $\alpha_{22} = 0.1$. We chose a smooth function $f(t_{ij}) =$

$\exp(\text{sen}(0.3t_{ij})\cos(0.6t_{ij}))$, where $t_{ij} = (1,2,3,4,5,6,7)$. The values $\mathbf{x}_i^\top = (x_1, x_2)$ were generated independently from a uniform distribution in the intervals (0,1) and (-1,1), respectively, and those values were kept constant throughout the experiment.

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

3.6.1 Asymptotic properties

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

To evaluate the computational accuracy and to examine the consistency of the estimates of the standard erros suggested in Subsection 3.3.4, we computed the following measures:

- The arithmetic average of estimates:

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

- The average values of the estimates of the standard erros obtained through the method described in Subsection 3.3.4 using the expected information matrix (MC IM).

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

Table 5 summarize the simulation results based on 200 Monte Carlo data sets for the model parameters $(\boldsymbol{\beta}, \mathbf{f})$. It can be observed that the MC Mean approaches the true value for fixed components and when the sample size increases the value of MC SD decreases. It can also be seen that the approximate standard errors (MC IM) obtained in Subsection 3.3.4 and the standard deviation estimates (MC SD) closely agree with each other, suggesting that the derived standard errors works well. From Figure 11 it can be observed that the variability among the estimates of the nonparametric function declines as the sample size increases, and the censorship does not influence the estimation of the nonparametric part. Therefore, we can conclude that the t-SMEC model provides estimates with good asymptotic properties for the fixed components and the nonparametric part is able to capture the true unknown function.

Table 5 – **Simulation study - Asymptotic properties**. Results based on 200 simulated samples. MC IM, MC SD are the respective average of the approximate standard errors obtained using the expected information matrix, and the average of the approximate standard deviations from fitting t-SMEC model.

Cens. level	Parameter	n=50			n=100			n=300		
		MC Mean	MC IM	MC SD	MC Mean	MC IM	MC SD	MC Mean	MC IM	MC SD
10%	$\beta_1 = 2$	1.9993	0.0362	0.0389	2.0040	0.0263	0.0289	1.9996	0.0148	0.0145
	$\beta_2 = -1.5$	-1.4989	0.0186	0.0199	-1.5020	0.0131	0.0141	-1.4992	0.0073	0.0076
	$f(1) = 1.2762$	1.2759	0.1091	0.1128	1.2777	0.0787	0.0747	1.2713	0.0453	0.0462
	$f(2) = 1.2270$	1.2297	0.1384	0.1397	1.2301	0.1005	0.0933	1.2198	0.0576	0.0566
	$f(3) = 0.8370$	0.8470	0.1759	0.1820	0.8369	0.1281	0.1197	0.8270	0.0733	0.0713
	$f(4) = 0.5029$	0.5120	0.2176	0.2244	0.5020	0.1585	0.1507	0.4915	0.0906	0.0868
	$f(5) = 0.3725$	0.3866	0.2613	0.2675	0.3659	0.1904	0.1818	0.3568	0.1086	0.1059
	$f(6) = 0.4176$	0.4335	0.3061	0.3145	0.4077	0.2230	0.2111	0.3989	0.1272	0.1264
	$f(7) = 0.6549$	0.6785	0.3519	0.3616	0.6456	0.2561	0.2433	0.6349	0.1461	0.1433
20%	$\beta_1 = 2$	1.9981	0.0292	0.0466	2.0037	0.0209	0.0316	1.9987	0.0117	0.0165
	$\beta_2 = -1.5$	-1.4971	0.0150	0.0222	-1.5011	0.0104	0.0166	-1.4986	0.0058	0.0090
	$f(1) = 1.2762$	1.2666	0.0934	0.1129	1.2672	0.0654	0.0783	1.2582	0.0376	0.0477
	$f(2) = 1.2270$	1.2163	0.1184	0.1385	1.2156	0.0833	0.0957	1.1995	0.0478	0.0577
	$f(3) = 0.8370$	0.8301	0.1501	0.1815	0.8174	0.1061	0.1227	0.7999	0.0607	0.0722
	$f(4) = 0.5029$	0.4909	0.1853	0.2254	0.4777	0.1313	0.1543	0.4583	0.0749	0.0874
	$f(5) = 0.3725$	0.3629	0.2221	0.2698	0.3372	0.1577	0.1865	0.3163	0.0898	0.1067
	$f(6) = 0.4176$	0.4055	0.2599	0.3174	0.3731	0.1846	0.2166	0.3522	0.1051	0.1273
	$f(7) = 0.6549$	0.6476	0.2985	0.3649	0.6056	0.2120	0.2483	0.5820	0.1207	0.1443

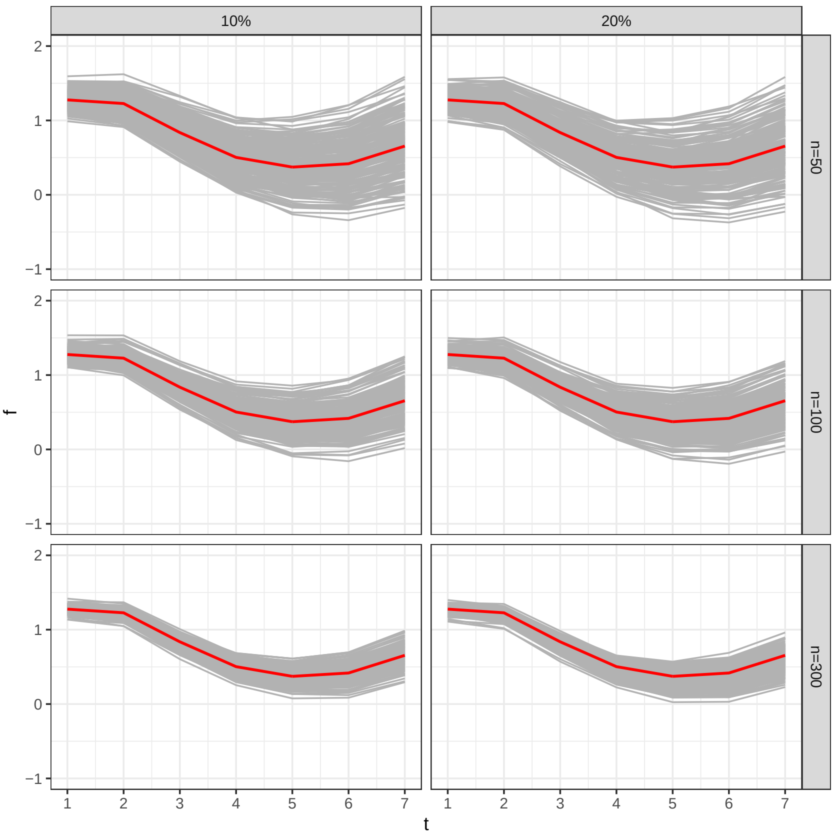


Figure 11 – **Simulation study - Asymptotic properties**. Graphs of the nonparametric components with 200 replications. Adjusted curves (gray lines) and true curves (red lines) for all scenarios.

3.6.2 Robustness of the estimates

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

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

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

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

The model selection criterion as well as the estimates of the model parameters were recorded for each simulation. The detailed numerical results under the scenarios considered, including the average BIC values and the MPL estimates are summarized in Table 6. From Table 6, can be noted that when the data generated follow the normal distribution the performances of the N-SMEC model and t-SMEC model are similar, indicating that the t-SMEC model gives reliable estimates. Also, to evaluate the use of the BIC criterion, the N-SMEC model was chosen by the criterion in 79.5% (159/200) of the samples generated as the best model. When the data generated follow the contamined normal, the t-SMEC model has better estimates and the standard errors are less than the N-SMEC model. Evaluating the BIC criterion, the N-SMEC model was chosen in 38.5% (77/200) of the samples.

Another important feature in our model is the ability to detect whether the distribution has heavy tails or not. It can be seen from Table 6 that when we fit the t-SMEC model to normal data, the estimate of ν on average is high, that is, the data does not have heavy-tails behavior. Now, when the data generated is contaminated normal, the estimated ν on average is small since we are dealing with a distribution with heavier tails that the normal distribution. Therefore, it can be observed that the t-SMEC model fits better than the N-SMEC model counterpart when the data have heavy tail behavior.

Table 6 – **Simulation study - Robustness of the estimates**. Summary statistics based on 200 simulated AR(1) samples for the estimates parameters.

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

3.7 Application

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

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

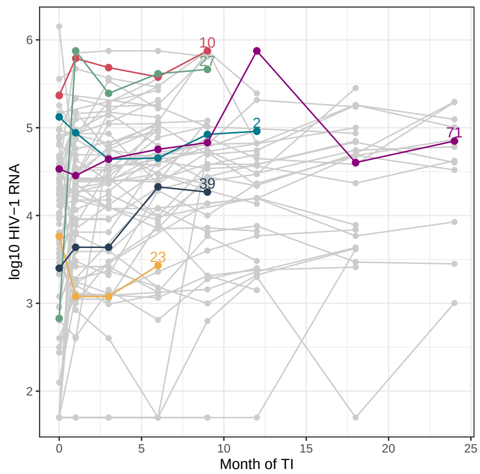


Figure 12 – **UTI data**. Individual profiles (in log10 scale) for HIV viral load at different follow-up times.

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

$$y_{ij} = f(t_{ij}) + b_i + \epsilon_{ij}, \tag{3.22}$$

where y_{ij} is the $\log_{10}$HIV RNA for subject i at time t_{ij} $(i = 1, 2, \ldots, 72;\ j = 1, 2, \ldots, n_i)$, $f(t_{ij})$ is an arbitrary smoothing function, b_i is the random intercept for the i-th subject, and ϵ_{ij} are random errors. The model (3.22) can be express in matrix form as:

$$\mathbf{y}_i = \mathbf{N}_i\mathbf{f} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \tag{3.23}$$

where $\mathbf{y}_i$ is a $(n_i \times 1)$ vector of responses for the i-th children, $\mathbf{N}_i$ is the incidence matrix, $\mathbf{f}$ is a (8×1) vector whose components are function $f(\cdot)$ evaluated at the times in the set $\mathbf{t}^0 = (t_1^0 = 0, t_2^0 = 1, t_3^0 = 3, \ldots, t_8^0 = 24)$, $\mathbf{Z}_i = \mathbf{1}_{n_i}$, with $\mathbf{1}_{n_i}$ a $(n_i \times 1)$ vector of ones and

$\mathbf{t}_i = [t_{i1}, \ldots, t_{in_i}]^\top$, $\mathbf{b}_i = b_i$ the random intercept and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})^\top$ represents the within-subject random error.

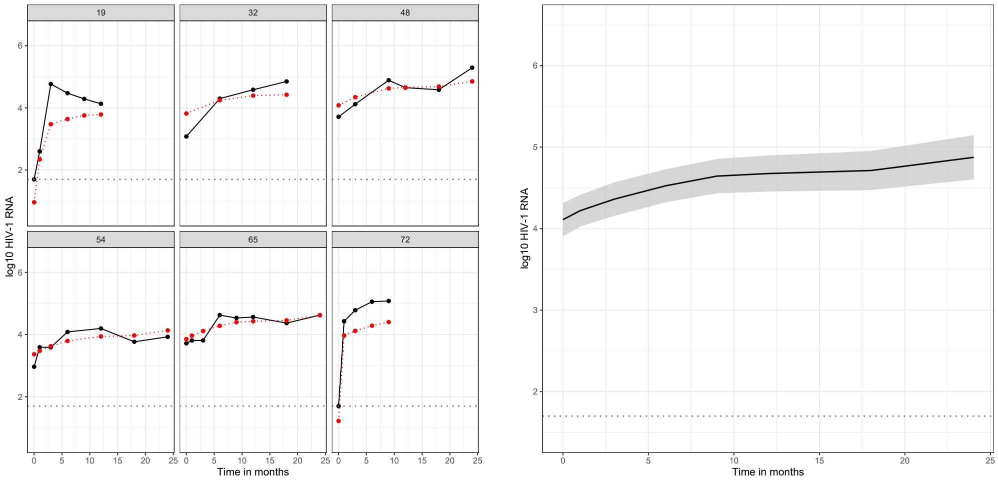
We acknowledge four cases of correlation structure to specify the matrix $\mathbf{E}_i$: the continuous-time AR(1) structure, the compound symmetry (CS), the damped exponential (DEC) and the uncorrelated (UNC). Table 7 represents the MPL estimates of $\boldsymbol{\theta} = (\mathbf{f}^\top, \sigma^2, \alpha, \boldsymbol{\phi}^\top, \nu)^\top$, the smoothing parameter estimate (λ), the corresponding penalized log-likelihood function evaluated at $\widehat{\boldsymbol{\theta}}$ in the fitted models, and the values of AIC and BIC. These results reveal that the model with an UNC structure has lower AIC and BIC compared to the other structures, that is, the measures over the time of the same subject are not correlated. From the fit of (3.22), considering the t-SMEC model under UNC correlation structure, estimates of individual profiles are shown for six subjects in Figure 13a, it can be seen that the model seems to provide a good fit.

Table 7 – **UTI dataset**. Parameter estimates of the t-SMEC model (3.22) for the UTI dataset. SE indicates the standard errors.

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

Again, for the t-SMEC model under UNC correlation structure (our best model), we present in Figure 13b the curve of the estimated nonparametric function and the corresponding confidence bands. It can be noted that the estimated nonparametric function increase gradually. This is the evidence of the negative effect of the antiretroviral therapy interruption on the viral load levels. It means, the viral load increments consistently along the time when the antiretroviral therapy begins to be interrupted. For the fit model, the mean viral load ($\mathbb{E}(y_{ij}) = f(t_{ij})$) increases from 4.09 at the time of UTI to 4.87 at 24 months.

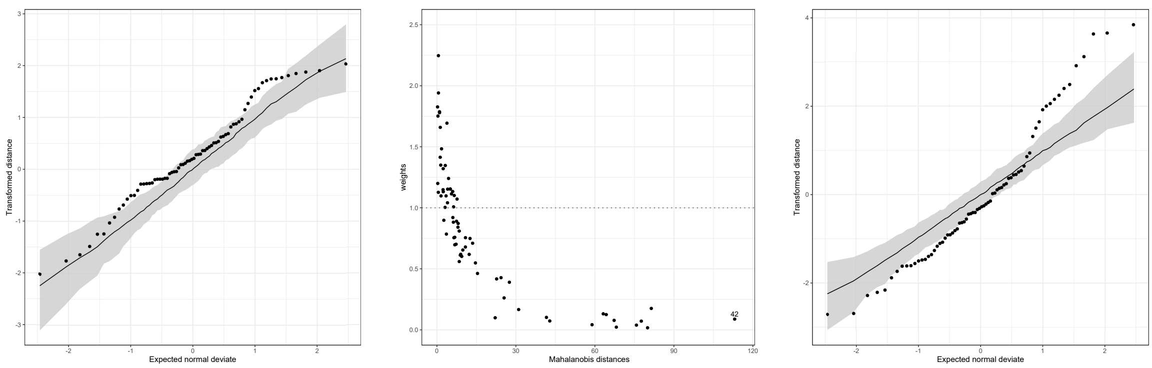
Figure 14 displays the transformed distance plots, for the Student-t (Figure 14a) and the normal (Figure 14c) models. The transformed distance under the Student-t



(a) (b)

Figure 13 – **UTI dataset**. (a) Viral loads in $\log_{10}$ scale (solid line) for 6 randomly chosen subjects and estimated trajectories (red, dotted line) for the t-SMEC model in the UNC structure. (b) Fitted curve of nonparametric part. The shaded regions denote the 95% confidence intervals obtained by $\widehat{\mathbf{f}} \pm 1.96\sqrt{\widehat{\mathrm{Var}}(\widehat{\mathbf{f}})}$.

model seems to be closer to normality than under the normal model. Therefore, it can be seen that the fitted model t-SMEC with the UNC correlation structure seems to present an adequate fit. Identification of outlying observations under the t-SMEC model may be performed, for instance, by the scatter plot between the estimated weight and the estimated Mahalanobis distance, Figure 14b. As can be seen, the subject 42 receive a smaller weight and the higher Mahalanobis distance. Besides, in this Figure, it can be observed that many observations present smaller weights, verifying the robust aspects of the MPL estimation under the Student's t-distribution.



(a) (b) (c)

Figure 14 – **UTI dataset**. (a) Normal probability plot for the transformed distance under the t-SMEC model with UNC structure. (b) Estimated weights ($\widehat{u}_i$) for the estimated t-SMEC model with UNC structure. (c) Normal probability plot for the transformed distance under the N-SMEC model with UNC structure.

3.8 Conclusion

In this chapter, we proposed a semiparametric mixed model for the analysis of longitudinal censored data, assuming that the within-individual measurement errors and the random effects were distributed with multivariate Student's t-distribution. This work can be considered as an extension of Matos *et al.* (2013b), where a linear/nonlinear mixed effects model was considered for censored data with Student's t-distribution.

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

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

Chapter 4

Likelihood-based inference for mixed-effects models with censored response using skew-normal distribution

4.1 Introduction

Longitudinal studies have attracted a considerable interest in clinical trials, biological psychology, environmental science, and medical research, as they enable the study of change over time of an outcome and the evaluation of determinants of change. Linear and nonlinear mixed-effect (N/LME) models are powerful tools for analyzing longitudinal data. In these models, random effects are incorporated to accommodate among-subject variation (Laird & Ware, 1982; Davidian & Giltinan, 1995). The random errors and/or random effects are routinely assumed to have a normal distribution due to their mathematical tractability and computational convenience.

Although the normality assumption may be reasonable for many situations, a serious departure of normality will cause a lack of robustness and subsequently lead to invalid inference and unreasonable estimates (Verbeke & Lesaffre, 1996). Specially non-normal characteristics such as skewness with heavy right or left tail appear often in virologic responses. For example, Figure 15a displays the histogram of the viral load measurements for 44 subjects enrolled in an AIDS clinical study - A5055, (refer to Subsections 1.4.2 and 4.6.1 for details of this data). From this figure, it can be seen that the viral load measurements are highly skewed, even after a $\log_{10}$ transformation.

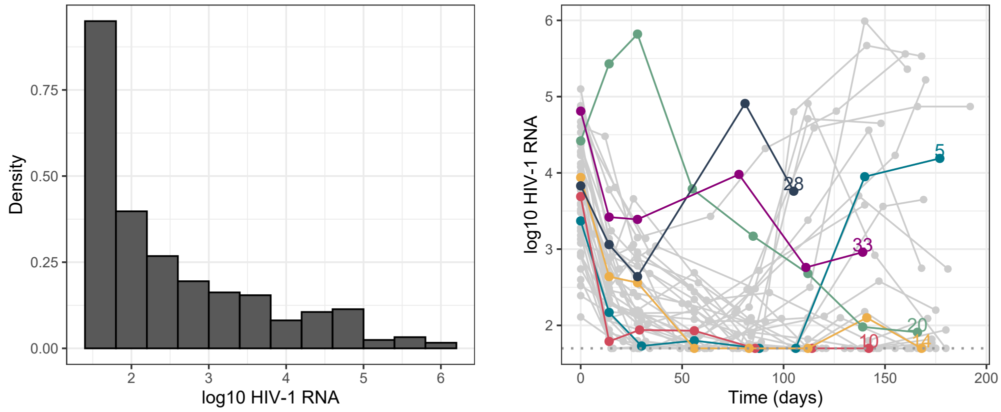
As an alternative to the weakness of unrealistic normality assumptions and eliminate the need for *ad hoc* data transformations, asymmetric distributions can be applied to consider this non-ignorable departure from normality. Lachos *et al.* (2010)

proposed a robust generalization of LME, called the skew normal/independent linear mixed (SNI-LME) model, by assuming a skew normal/independent (SNI) distribution (Branco & Dey, 2001) for the random effects and a normal/independent distribution for the random errors. HO & LIN (2010) proposed a model that provides flexibility in capturing the effects of skewness and heavy tails simultaneously among longitudinal data, they consider an extension of LME assuming a multivariate skew-t distribution for the random effects and a multivariate Student's t-distribution for the error terms.

Another complexity of longitudinal studies occurs when the response is censored for some of the observations, which often arises when assay measures are collected over time and the assay procedure is subject to limits of quantification. As a case in point, the HIV-1 viral load, which is currently the primary marker of HIV infection, has a lower and upper quantification limit, which depends on the type of assay used. The viral load of patients receiving anti-retroviral treatment will typically decline and stay for a longer period below the lower limit of quantification. Figure 15b shows the measurements of viral load for patients in the A5055 study. We can see that for some patients viral loads are below a limit of detection (50 copies/mL here). When response observations are below limits of quantification, a common practice is to impute the censored values by the detection limit or half the detection limit (Wu & Ding, 1999). Such *ad hoc* methods may produce biased results as pointed out by Hughes (1999), Jacqmin-Gadda *et al.* (2000), Matos *et al.* (2016), just to name a few.

In the literature, longitudinal data with censored observations have received considerable attention. Vaida & Liu (2009) proposed an exact EM algorithm for LME/NLME with censored response (LMEC/NLMEC), which uses closed-form expressions at the E-step, as opposed to Monte Carlo simulations. Robust extensions of LMEC and NLMEC based on the multivariate Student's t-distribution, named as tLMEC and tNLMEC, have been introduced by Matos *et al.* (2013b). On the other hand, and under a Bayesian framework, Bandyopadhyay *et al.* (2012) studied LMEC models considering both skewness and heavy tails, replacing the Gaussian assumptions with skew-normal/independent (SNI) distribution. However, to the best of our knowledge, no previous work have investigated LMEC/NLMEC models based on the skew-normal distributions from a likelihood based perspective.

In this chapter, we are devoted to presenting methodological developments of the skew-normal linear/nonlinear mixed model with censored responses (SN-LMEC/NLMEC) from a likelihood based perspective, which takes into account the skewness behaviour of the random effects. The SN-LMEC/NLMEC is defined by supposing that, for each subject, the random effects follow a SN distribution introduced by Azzalini & Valle (1996), while the within-subject errors follow a multivariate normal distribution to prevent identifiability problems. Like Matos *et al.* (2013b), we show that the E-step reduces to computing the



(a) (b)

Figure 15 – **A5055 data**. (a) Histogram for HIV viral load (in $\log_{10}$) scale. (b) Individual profiles for HIV viral load (in $\log_{10}$ scale).

first two moments of a truncated multivariate SN distribution. The likelihood function is easily computed as a byproduct of the E-step and is used for monitoring convergence and for model selection.

The organization of this chapter is outlined as follows. Section 4.2 presents the skew-normal distribution (SN) and some of its keys properties. Section 4.3 introduces the model (SN-LMEC) and describes an efficient EM algorithm for calculating maximum likelihood (ML) estimates of parameters. We also discuss the issues related to empirical Bayes estimates of the random effects and prediction of future responses. The extension to the nonlinear case (SN-NLMEC) is discussed in Section 4.4. A simulation study is conducted in Section 4.5 to evaluate the proposed method. In Section 4.6, two case studies of HIV viral load are analyzed in detail. We conclude the article with some discussions in Section 4.7.

4.2 The multivariate skew-normal distribution

In this section we present the multivariate skew-normal distribution (SN) and multivariate extended skew-normal (ESN) and some of its useful properties. Some versions, extensions, and unifications of the SN family are carefully surveyed in works such as Azzalini (2005) and Arellano-Valle *et al.* (2006).

Definition 1. *A random vector* $\mathbf{Y}$ *has multivariate skew-normal distribution with* $p \times 1$ *location vector* $\boldsymbol{\mu}$, $p \times p$ *positive definite dispersion matrix* $\boldsymbol{\Sigma}$ *and* $p \times 1$ *skewness parameter vector* $\boldsymbol{\lambda}$, *and we write* $\mathbf{Y} \sim SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, *if its density is given by*

$$SN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}) = 2\phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})), \tag{4.1}$$

where $\phi_p(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and* $\Phi_p(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ *denote, respectively, the probability distribution function (pdf) and the cumulative distribution function (cdf) of the p-variate normal distribution*

$N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *with mean vector* $\boldsymbol{\mu}$ *and covariate matrix* $\boldsymbol{\Sigma}$, *respectively, and* $\boldsymbol{\Sigma}^{-1/2}$ *is such that* $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Sigma}^{-1}$. *Note that if* $\boldsymbol{\lambda} = \mathbf{0}$, *then the density of* $\mathbf{Y}$ *reduces to the* $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *density.*

It is worth mentioning that the multivariate skew-normal distribution is not closed over conditioning. Next we present its extended version which holds this property, called, the multivariate ESN distribution.

Definition 2. *A random vector* $\mathbf{Y}$ *has multivariate ESN distribution with* $p \times 1$ *location vector* $\boldsymbol{\mu}$, $p \times p$ *positive definite dispersion matrix* $\boldsymbol{\Sigma}$, $p \times 1$ *skewness parameter vector* $\boldsymbol{\lambda}$, *and shift parameter* $\tau \in \mathcal{R}$, *denoted by* $\mathbf{Y} \sim ESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$, *if its density is given by*

$$ESN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau) = \xi^{-1}\phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1(\tau + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})), \tag{4.2}$$

with $\xi = \Phi_1(\tau/(1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2})$. *Note that when* $\tau = 0$, *we recover the skew-normal distribution defined in (4.1), that is,* $ESN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, 0) = SN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$.

Define,

$$\mathcal{L}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau) = \int_{\mathbf{a}}^{\mathbf{b}} \xi^{-1}\phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1(\tau + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu}))d\mathbf{y}.$$

When $\boldsymbol{\lambda} = 0$ and $\tau = 0$, we recover the multivariate normal case, and then

$$\mathcal{L}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{0}, 0) \equiv L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int_{\mathbf{a}}^{\mathbf{b}} \phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})d\mathbf{y}.$$

Proposition 5. *Let* $\mathbf{Y} \sim SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ *and* $\mathbf{Y}$ *be partitioned as* $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$, *with dimensions* p_1 *and* p_2, $p_1 + p_2 = p$, *respectively. Let*

$$\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad \boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^\top, \boldsymbol{\lambda}_2^\top)^\top \quad \text{and} \quad \boldsymbol{\varphi} = (\boldsymbol{\varphi}_1^\top, \boldsymbol{\varphi}_2^\top)^\top$$

be the corresponding partitions of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\lambda}$ *and* $\boldsymbol{\varphi} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda}$. *Then,*

(i) $\mathbf{Y}_1 \sim SN_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, c_{12}\boldsymbol{\Sigma}_{11}^{1/2}\tilde{\boldsymbol{\upsilon}})$; *and*

(ii) $\mathbf{Y}_2|\mathbf{Y}_1 = \mathbf{y}_1 \sim ESN_{p_2}(\boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{22.1}, \boldsymbol{\Sigma}_{22.1}^{1/2}\boldsymbol{\varphi}_2, \tau_{2.1})$,

where $c_{12} = (1 + \boldsymbol{\varphi}_2^\top \boldsymbol{\Sigma}_{22.1}\boldsymbol{\varphi}_2)^{-1/2}$, $\tilde{\boldsymbol{\upsilon}} = \boldsymbol{\varphi}_1 + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\varphi}_2$, $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$ *and* $\tau_{2.1} = \tilde{\boldsymbol{\upsilon}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1)$.

Proof. See Proposition 2 in Galarza *et al.* (2019). □

The mean and variance of a ESN random vector is given in the following lemma:

Lemma 1. *Let* $\mathbf{Y} \sim ESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$. *Then,*

i) $\mathbb{E}[\mathbf{Y}] = \boldsymbol{\mu} + \eta\boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}$,

ii) $\mathbb{E}[\mathbf{Y}\mathbf{Y}^\top] = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top + \eta\left(\boldsymbol{\mu}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{1/2} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}\boldsymbol{\mu}^\top\right) + \eta\tau\boldsymbol{\Sigma}^{1/2}\dfrac{\boldsymbol{\lambda}\boldsymbol{\lambda}^\top}{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}\boldsymbol{\Sigma}^{1/2}$,

iii) $Var(\mathbf{Y}) = \boldsymbol{\Sigma} - \eta\boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}\left(\eta\boldsymbol{\lambda} - \dfrac{\tau}{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}\boldsymbol{\lambda}\right)^\top\boldsymbol{\Sigma}^{1/2}$,

with $\eta = \phi_1(\tau; 0, 1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})/\xi$. *When* $\tau = 0$, *we recover* $\mathbb{E}[\mathbf{Y}], \mathbb{E}[\mathbf{Y}\mathbf{Y}^\top]$ *and* $Var(\mathbf{Y})$ *the skew-normal distribution.*

Definition 3. *Let* $\mathbf{X} \sim ESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$ *and* $\mathbb{P}(\mathbf{a} < \mathbf{X} < \mathbf{b}) > 0$. *A random vector* $\mathbf{Y}$ *has a truncated extended multivariate skew-normal (TESN) distribution in the interval* $[\mathbf{a}, \mathbf{b}]$, *denoted by* $\mathbf{Y} \sim TESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, [\mathbf{a}, \mathbf{b}])$, *if its density is given by*

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{ESN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)}{\int_{\mathbf{a}}^{\mathbf{b}} ESN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)d\mathbf{y}}, \quad \mathbf{a} \leqslant \mathbf{y} \leqslant \mathbf{b}.$$

For the special case $\tau = 0$, *we refer to this distribution as a truncated multivariate skew-normal (TSN) distribution, i.e.,* $TESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \mathbf{0}, [\mathbf{a}, \mathbf{b}]) \equiv TSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, [\mathbf{a}, \mathbf{b}])$.

The following properties of the multivariate truncated ESN distribution are useful for the implementation of the EM-algorithm in SN-LMEC/NLMEC models.

Lemma 2. *Let* $\mathbf{Y} \sim TESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, [\mathbf{a}, \mathbf{b}])$. *For any measurable function* $g(\cdot)$, *we have that*

$$\mathbb{E}\left[g(\mathbf{Y})\frac{\phi_1(\tau + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}-\boldsymbol{\mu}))}{\Phi_1(\tau + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}-\boldsymbol{\mu}))}\right] = \frac{\eta L}{\mathcal{L}}\mathbb{E}[g(\mathbf{W})],$$

with $\eta = \phi_1(\tau; 0, 1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})/\xi$, $L = L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu} - \boldsymbol{\mu}^*, \boldsymbol{\Psi})$, $\mathcal{L} = \mathcal{L}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$, $\boldsymbol{\Psi} = \boldsymbol{\Sigma}^{1/2}(\mathbf{I}_p + \boldsymbol{\lambda}\boldsymbol{\lambda}^\top)^{-1}\boldsymbol{\Sigma}^{1/2}$, $\boldsymbol{\mu}^* = \tau\boldsymbol{\Psi}\boldsymbol{\varphi}$, *and* $\mathbf{W} \sim TN_p(\boldsymbol{\mu} - \boldsymbol{\mu}^*, \boldsymbol{\Psi}, [\mathbf{a}, \mathbf{b}])$.

Proof. See Lemma 1 in Galarza *et al.* (2019). □

Corollary 1. *Setting* $\tau = 0$, *it follows that* $\mathbf{Y} \sim TSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, [\mathbf{a}, \mathbf{b}])$ *and*

$$\mathbb{E}\left[g(\mathbf{Y})\frac{\phi_1(\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}-\boldsymbol{\mu}))}{\Phi_1(\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}-\boldsymbol{\mu}))}\right] = \frac{L_0}{\sqrt{\frac{\pi}{2}(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}\mathcal{L}_0}\mathbb{E}[g(\mathbf{W}_0)],$$

with $L_0 = L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Psi})$, $\mathcal{L}_0 = \mathcal{L}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, 0)$ *and* $\mathbf{W}_0 \sim TN_p(\boldsymbol{\mu}, \boldsymbol{\Psi}, [\mathbf{a}, \mathbf{b}])$.

Proof. The proof is straightforward. Setting $\tau = 0$, it suffices to find that $\boldsymbol{\mu}^* = \mathbf{0}$ and $\eta = \sqrt{2/\pi(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}$. □

4.3 The skew-normal linear mixed effects model with censored responses

4.3.1 The statistical model

In order to allow symmetric-asymmetric properties in real data sets, the SN-LMEC is defined by extending the normal mixed-effects models presented by Vaida & Liu (2009). The model is specified as follows:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n, \tag{4.3}$$

where the subscript i is the subject index, $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^\top$ is a $n_i \times 1$ vector of observed continuous responses for sample unit i, $\mathbf{X}_i$ is the $n_i \times p$ design matrix corresponding to the $p \times 1$ vector of fixed-effects $\boldsymbol{\beta}$, and $\mathbf{Z}_i$ is the $n_i \times q$ design matrix corresponding to the $q \times 1$ vector of random effects $\mathbf{b}_i$, and $\boldsymbol{\epsilon}_i$ is the $n_i \times 1$ vector of random errors.

In this work, we assume that

$$\begin{pmatrix} \mathbf{b}_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \overset{\text{ind.}}{\sim} \mathrm{SN}_{q+n_i}\left(\begin{pmatrix} c\boldsymbol{\Delta} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_i \end{pmatrix}, \begin{pmatrix} \boldsymbol{\lambda} \\ \mathbf{0} \end{pmatrix} \right), \tag{4.4}$$

where $c = -\sqrt{2/\pi}$, $\boldsymbol{\Delta} = \mathbf{D}^{1/2}\boldsymbol{\delta}$, $\boldsymbol{\delta} = \dfrac{\boldsymbol{\lambda}}{(1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}}$. The value of the location parameter, $c\boldsymbol{\Delta}$, of $\mathbf{b}_i$ is chosen in order to obtain $\mathbb{E}[\mathbf{b}_i] = 0$, as in the normal model. The dispersion matrix $\mathbf{D} = \mathbf{D}(\boldsymbol{\alpha})$ models between-subjects variability, and depends on the unknown and reduced parameter vector $\boldsymbol{\alpha}$ of dimension q. 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. Thus, we adopt a DEC structure for $\boldsymbol{\Omega}_i$, as proposed by Muñoz *et al.* (1992). This correlation structure allows us to deal with unequally spaced and unbalanced observations and is defined as

$$\boldsymbol{\Omega}_i = \sigma^2\mathbf{E}_i = \sigma^2\mathbf{E}_i(\boldsymbol{\phi}; \mathbf{t}_i) = \sigma^2\left[\phi_1^{|t_{ij}-t_{ik}|^{\phi_2}}\right], \quad i = 1, \ldots, n, \quad j, k = 1, \ldots, n_i,$$

where ϕ_1 is the correlation parameter that describes the autocorrelation between observations separated by the absolute length of two time points, and ϕ_2 is the damping parameter which allows the 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 $\boldsymbol{\Phi} = \{(\phi_1, \phi_2) : 0 < \phi_1 < 1, \phi_2 > 0\}$. A more detailed discussion of the DEC structure can be found in Muñoz *et al.* (1992).

According to Lachos *et al.* (2010), model (4.3) can be written hierarchically as

$$\begin{aligned} \mathbf{Y}_i|\mathbf{b}_i &\overset{\text{ind.}}{\sim} \mathrm{N}_{n_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \boldsymbol{\Omega}_i), \\ \mathbf{b}_i|T_i = t_i &\overset{\text{ind.}}{\sim} \mathrm{N}_q(\boldsymbol{\Delta}t_i, \boldsymbol{\Gamma}), \\ T_i &\overset{\text{iid.}}{\sim} \mathrm{TN}(c, 1; (c, \infty)), \end{aligned} \tag{4.5}$$

where $\mathbf{\Gamma} = \mathbf{D} - \mathbf{\Delta}\mathbf{\Delta}^\top$.

It follows from (4.5) that the density of $\mathbf{Y}_i$ is

$$f(\mathbf{Y}_i) = 2\phi_{n_i}(\mathbf{Y}_i; \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i c\mathbf{\Delta}, \mathbf{\Sigma}_i)\Phi_1\left(\bar{\boldsymbol{\lambda}}_i^\top \mathbf{\Sigma}_i^{-1/2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i c\mathbf{\Delta})\right), \qquad (4.6)$$

i.e., $\mathbf{Y}_i \overset{\text{ind.}}{\sim} \text{SN}_{n_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i c\mathbf{\Delta}, \mathbf{\Sigma}_i, \bar{\boldsymbol{\lambda}}_i), i = 1, \ldots, n$, where $\mathbf{\Sigma}_i = \mathbf{\Omega}_i + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top$,

$$\mathbf{\Lambda}_i = (\mathbf{D}^{-1} + \mathbf{Z}_i^\top\mathbf{\Sigma}_i^{-1}\mathbf{Z}_i)^{-1} \quad \text{and} \quad \bar{\boldsymbol{\lambda}}_i = \frac{\mathbf{\Sigma}_i^{-1/2}\mathbf{Z}_i\mathbf{D}\boldsymbol{\zeta}}{\sqrt{1 + \boldsymbol{\zeta}^\top\mathbf{\Lambda}_i\boldsymbol{\zeta}}}, \quad \text{with} \quad \boldsymbol{\zeta} = \mathbf{D}^{-1/2}\boldsymbol{\lambda}.$$

As previously mentioned, the proposed model also considers censored observations, i.e., we assume that the response Y_{ij} is not fully observed for all i, j. Thus, we consider the approach proposed by Vaida & Liu (2009) to model the censored responses. Let the observed data for the i-th subject be $(\mathbf{V}_i, \mathbf{C}_i)$, where $\mathbf{V}_i$ represents the vector of uncensored readings $(V_{ij} = V_{0i})$ or censoring level, and $\mathbf{C}_i$ is the vector of censoring indicators, such that:

$$C_{ij} = \begin{cases} 1 & \text{if} \quad V_{1ij} \leqslant Y_{ij} \leqslant V_{2ij}, \\ 0 & \text{if} \quad Y_{ij} = V_{0i}, \end{cases} \qquad (4.7)$$

for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, n_i\}$, i.e., $C_{ij} = 1$ if Y_{ij} is located within a specific interval. The model defined in (4.3)-(4.7) is henceforth called the SN-LMEC model.

4.3.2 The likelihood function

To obtain the likelihood function of the SN-LMEC model, first we treat separately the observed and censored components of $\mathbf{Y}_i$, i.e. $\mathbf{Y}_i = (\mathbf{Y}_i^{o\top}, \mathbf{Y}_i^{c\top})^\top$, with $C_{ij} = 0$ for all elements in $\mathbf{Y}_i^o$, and $C_{ij} = 1$ for all elements in $\mathbf{Y}_i^c$. Analogous, we write $\mathbf{V}_i = \text{vec}(\mathbf{V}_i^o, \mathbf{V}_i^c)$, where $\mathbf{V}_i^c = (\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c)$ with

$$\boldsymbol{\mu}_i = (\boldsymbol{\mu}_i^{o\top}, \boldsymbol{\mu}_i^{c\top})^\top, \quad \mathbf{\Sigma}_i = \begin{pmatrix} \mathbf{\Sigma}_i^{oo} & \mathbf{\Sigma}_i^{oc} \\ \mathbf{\Sigma}_i^{co} & \mathbf{\Sigma}_i^{cc} \end{pmatrix}, \quad \bar{\boldsymbol{\lambda}}_i = (\bar{\boldsymbol{\lambda}}_i^{o\top}, \bar{\boldsymbol{\lambda}}_i^{c\top})^\top \quad \text{and} \quad \boldsymbol{\varphi}_i = (\boldsymbol{\varphi}_i^{o\top}, \boldsymbol{\varphi}_i^{c\top})^\top,$$

where $\boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i c\mathbf{\Delta}$ and $\boldsymbol{\varphi}_i = \mathbf{\Sigma}_i^{-1/2}\bar{\boldsymbol{\lambda}}_i$. Then, using Proposition 5, we have that

$$\mathbf{Y}_i^o \sim \text{SN}_{n_i^o}(\boldsymbol{\mu}_i^o, \mathbf{\Sigma}_i^{oo}, c_i^{co}\mathbf{\Sigma}_i^{oo^{1/2}}\tilde{\boldsymbol{\upsilon}}) \quad \text{and} \quad \mathbf{Y}_i^c|\mathbf{Y}_i^o = \mathbf{y}_i^o \sim \text{ESN}_{n_i^c}(\boldsymbol{\mu}_i^{co}, \mathbf{S}_i, \mathbf{S}_i^{1/2}\boldsymbol{\varphi}_i^c, \tau_i^{co}),$$

where

$$\boldsymbol{\mu}_i^{co} = \boldsymbol{\mu}_i^c + \mathbf{\Sigma}_i^{co}\mathbf{\Sigma}_i^{oo^{-1}}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o), \quad \mathbf{S}_i = \mathbf{\Sigma}_i^{cc} - \mathbf{\Sigma}_i^{co}(\mathbf{\Sigma}_i^{oo})^{-1}\mathbf{\Sigma}_i^{oc}, \quad c_i^{co} = (1 + \boldsymbol{\varphi}_i^{c\top}\mathbf{S}_i\boldsymbol{\varphi}_i^c)^{-1/2},$$

$$\tilde{\boldsymbol{\upsilon}} = \boldsymbol{\varphi}_i^o + \mathbf{\Sigma}_i^{oo^{-1}}\mathbf{\Sigma}_i^{oc}\boldsymbol{\varphi}_i^c \quad \text{and} \quad \tau_i^{co} = \tilde{\boldsymbol{\upsilon}}^\top(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o).$$

Thus, the likelihood for the i-th subject is given by

$$\begin{aligned} L_i(\boldsymbol{\theta}) = L_i &= f(\mathbf{y}_i^o|\boldsymbol{\theta})\mathbb{P}(\mathbf{V}_{1i}^c \leqslant \mathbf{y}_i^c \leqslant \mathbf{V}_{2i}^c|\mathbf{y}_i^o, \boldsymbol{\theta}) \\ &= \text{SN}_{n_i^o}(\boldsymbol{\mu}_i^o, \mathbf{\Sigma}_i^{oo}, c_i^{co}\mathbf{\Sigma}_i^{oo^{1/2}}\tilde{\boldsymbol{\upsilon}})\mathcal{L}_{n_i^c}(\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c; \boldsymbol{\mu}_i^{co}, \mathbf{S}_i, \mathbf{S}_i^{1/2}\boldsymbol{\varphi}_i^c, \tau_i^{co}), \end{aligned}$$

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

4.3.3 The EM algorithm

In this section, we describe how to use the EM-type algorithm to compute the Maximum Likelihood (ML) estimation of the DEC-SNLMEC model. The EM algorithm originally proposed 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.

Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$, $\mathbf{t} = (t_1, \ldots, t_n)^\top$, $\mathbf{V} = \mathrm{vec}(\mathbf{V}_1, \ldots, \mathbf{V}_n)$ and $\mathbf{C} = \mathrm{vec}(\mathbf{C}_1, \ldots, bC_n)$, where $(\mathbf{V}_i, \mathbf{C}_i)$ is observed for the ith subject. Treating $\mathbf{y}, \mathbf{b}$ and $\mathbf{t}$ as hypothetical missing data, and augmenting with the observed data $\mathbf{V}, \mathbf{C}$, we set $\mathbf{y}_c = (\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{y}^\top, \mathbf{b}^\top, \mathbf{t}^\top)^\top$ as the complete data. Hence, it follows from (4.5) that the complete-data log-likelihood function is of the form

$$
\begin{aligned}
\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) &= \sum_{i=1}^n [\log f(\mathbf{y}_i|\mathbf{b}_i) + \log f(\mathbf{b}_i|t_i) + \log f(t_i)] \\
&= \sum_{i=1}^n \Big\{ -\frac{1}{2}\log|\boldsymbol{\Omega}_i| - \frac{1}{2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top \boldsymbol{\Omega}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i) \\
&- \frac{1}{2}\log|\boldsymbol{\Gamma}| - \frac{1}{2}(\mathbf{b}_i - t_i\boldsymbol{\Delta})^\top \boldsymbol{\Gamma}^{-1}(\mathbf{b}_i - t_i\boldsymbol{\Delta}) \Big\} + C
\end{aligned}
$$

where C is a constant that is independent of the parameter vector $\boldsymbol{\theta}$.

The E-step evaluate the conditional expectation of complete-data log-likelihood function given the observed data $\mathbf{V}$, $\mathbf{C}$ and current values $\widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\boldsymbol{\beta}}^{(k)\top}, \widehat{\sigma^2}^{(k)}, \widehat{\boldsymbol{\alpha}}^{(k)\top}, \widehat{\boldsymbol{\phi}}^{(k)\top}, \widehat{\boldsymbol{\lambda}}^{(k)\top})^\top$, yielding the so-called Q-function

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

where

$$
\begin{aligned}
Q_{1i}\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}|\widehat{\boldsymbol{\theta}}^{(k)}\right) &= -\frac{n_i}{2}\log\widehat{\sigma^2}^{(k)} - \frac{1}{2}\log|\widehat{\mathbf{E}}_i^{(k)}| - \frac{1}{2\widehat{\sigma^2}^{(k)}}\Big[\widehat{a}_i^{(k)}(\widehat{\boldsymbol{\phi}}^{(k)}) \\
&+ \widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top \widehat{\mathbf{E}}_i^{-1^{(k)}}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)} - 2\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top\widehat{\mathbf{E}}_i^{-1^{(k)}}\left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{\mathbf{b}}_i^{(k)}\right)\Big], \quad (4.8) \\
Q_{2i}\left(\boldsymbol{\alpha}, \boldsymbol{\lambda}|\widehat{\boldsymbol{\theta}}^{(k)}\right) &= -\frac{1}{2}\log|\widehat{\boldsymbol{\Gamma}}^{(k)}| - \frac{1}{2}\mathrm{tr}\Big[\widehat{\boldsymbol{\Gamma}}^{-1^{(k)}}\Big(\widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)} - \widehat{\boldsymbol{\Delta}}^{(k)}\widehat{t_i\mathbf{b}_i^\top}^{(k)} - \widehat{t_i\mathbf{b}_i}^{(k)}\widehat{\boldsymbol{\Delta}}^{(k)\top} \\
&+ \widehat{t_i^2}^{(k)}\widehat{\boldsymbol{\Delta}}^{(k)}\widehat{\boldsymbol{\Delta}}^{(k)\top}\Big)\Big], \quad (4.9)
\end{aligned}
$$

with $\widehat{a}_i^{(k)}(\widehat{\boldsymbol{\phi}}^{(k)}) = \mathrm{tr}\left[\widehat{\mathbf{E}}_i^{-1^{(k)}}\left(\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} - 2\widehat{\mathbf{y}_i\mathbf{b}_i^\top}^{(k)}\mathbf{Z}_i^\top + \mathbf{Z}_i\widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)}\mathbf{Z}_i^\top\right)\right]$.

The following conditional distributions are useful for obtaining the conditional expectations of missing data. From Lachos *et al.* (2010), we have that

$$\begin{aligned}
\mathbf{b}_i | T_i = t_i, \mathbf{Y}_i = \mathbf{y}_i &\sim \mathrm{N}_q\left(\mathbf{s}_i t_i + \mathbf{B}_i \mathbf{Z}_i^\top \boldsymbol{\Omega}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}), \mathbf{B}_i\right), \\
T_i | \mathbf{Y}_i = \mathbf{y}_i &\sim \mathrm{TN}_1\left(c + m_i, M_i^2; (0,\infty)\right), \\
\mathbf{Y}_i &\sim \mathrm{SN}_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i, \bar{\boldsymbol{\lambda}}_i),
\end{aligned}$$

where $M_i = (1 + \boldsymbol{\Delta}^\top \mathbf{Z}_i^\top \boldsymbol{\Upsilon}_i^{-1}\mathbf{Z}_i\boldsymbol{\Delta})^{-1/2}$, $m_i = M_i^2 \boldsymbol{\Delta}^\top \mathbf{Z}_i^\top \boldsymbol{\Upsilon}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i c\boldsymbol{\Delta})$, $\mathbf{B}_i = (\boldsymbol{\Gamma}^{-1} + \mathbf{Z}_i^\top \boldsymbol{\Omega}_i^{-1}\mathbf{Z}_i)^{-1}$, $\mathbf{s}_i = (\mathbf{I}_q - \mathbf{B}_i\mathbf{Z}_i^\top\boldsymbol{\Omega}_i^{-1}\mathbf{Z}_i)\boldsymbol{\Delta}$, $\boldsymbol{\Upsilon}_i = \boldsymbol{\Omega}_i + \mathbf{Z}_i\boldsymbol{\Gamma}\mathbf{Z}_i^\top$.

Therefore, the Q-function is completely determined by the knowledge of the following expectations:

$$\begin{aligned}
\widehat{t_i}^{(k)} &= \mathbb{E}\left[T_i \middle| \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] \\
&= c + \widehat{M}_i^{2(k)} \widehat{\boldsymbol{\Delta}}^{(k)\top} \mathbf{Z}_i^\top \widehat{\boldsymbol{\Upsilon}}_i^{-1(k)} \left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)} - \mathbf{Z}_i c\widehat{\boldsymbol{\Delta}}^{(k)}\right) + \widehat{M}_i^{(k)}\widehat{\kappa}_i^{(k)}, \\
\widehat{t_i^2}^{(k)} &= \mathbb{E}\left[T_i^2 \middle| \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] \\
&= \widehat{M}_i^{4(k)} \widehat{\boldsymbol{\Delta}}^{(k)\top} \mathbf{Z}_i^\top \widehat{\boldsymbol{\Upsilon}}_i^{-1(k)} \widehat{\mathbf{R}}_i^{(k)} \widehat{\boldsymbol{\Upsilon}}_i^{-1(k)} \mathbf{Z}_i \widehat{\boldsymbol{\Delta}}^{(k)} + \widehat{M}_i^{2(k)} + c^2 \\
&+ 2c\widehat{M}_i^{2(k)} \widehat{\boldsymbol{\Delta}}^{(k)\top} \mathbf{Z}_i^\top \widehat{\boldsymbol{\Upsilon}}_i^{-1(k)} \left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)} - \mathbf{Z}_i c\widehat{\boldsymbol{\Delta}}^{(k)}\right) \\
&+ \widehat{M}_i^{3(k)} \widehat{\boldsymbol{\Delta}}^{(k)\top} \mathbf{Z}_i^\top \widehat{\boldsymbol{\Upsilon}}_i^{-1(k)} \left(\widehat{\kappa_i\mathbf{y}}_i^{(k)} - \widehat{\kappa}_i^{(k)}\left(\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)} + \mathbf{Z}_i c\widehat{\boldsymbol{\Delta}}^{(k)}\right)\right) + 2c\widehat{M}_i^{(k)}\widehat{\kappa}_i^{(k)}, \\
\widehat{t_i\mathbf{y}_i}^{(k)} &= \mathbb{E}\left[T_i\mathbf{Y}_i \middle| \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] \\
&= \widehat{M}_i^{2(k)} \left(\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} - \widehat{\mathbf{y}}_i^{(k)}\left(\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)} + \mathbf{Z}_i c\widehat{\boldsymbol{\Delta}}^{(k)}\right)^\top\right) \widehat{\boldsymbol{\Upsilon}}_i^{-1(k)} \mathbf{Z}_i \widehat{\boldsymbol{\Delta}}^{(k)} + \widehat{M}_i^{(k)} \widehat{\kappa_i\mathbf{y}_i}^{(k)} \\
&+ c\widehat{\mathbf{y}}_i^{(k)}, \\
\widehat{\mathbf{b}}_i^{(k)} &= \mathbb{E}\left[\mathbf{b}_i \middle| \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] \\
&= \widehat{\mathbf{s}}_i^{(k)} \widehat{t}_i^{(k)} + \widehat{\mathbf{B}}_i^{(k)} \mathbf{Z}_i^\top \widehat{\boldsymbol{\Omega}}_i^{-1(k)} \left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\right), \\
\widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)} &= \mathbb{E}\left[\mathbf{b}_i\mathbf{b}_i^\top \middle| \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] \\
&= \widehat{t_i^2}^{(k)} \widehat{\mathbf{s}}_i^{(k)} \widehat{\mathbf{s}}_i^{(k)\top} + 2\widehat{\mathbf{B}}_i^{(k)} \mathbf{Z}_i^\top \widehat{\boldsymbol{\Omega}}_i^{-1(k)} \left(\widehat{t_i\mathbf{y}_i}^{(k)} - \widehat{t}_i^{(k)} \mathbf{X}_i \widehat{\boldsymbol{\beta}}^{(k)}\right) \widehat{\mathbf{s}}_i^{(k)\top} \\
&+ \widehat{\mathbf{B}}_i^{(k)} \mathbf{Z}_i^\top \widehat{\boldsymbol{\Omega}}_i^{-1(k)} \widehat{\mathbf{r}}_i^{(k)} \widehat{\boldsymbol{\Omega}}_i^{-1(k)} \mathbf{Z}_i \widehat{\mathbf{B}}_i^{(k)} + \widehat{\mathbf{B}}_i^{(k)}, \\
\widehat{\mathbf{y}_i\mathbf{b}_i^\top}^{(k)} &= \mathbb{E}\left[\mathbf{Y}_i\mathbf{b}_i^\top \middle| \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] \\
&= \left(\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} - \widehat{\mathbf{y}}_i^{(k)} \widehat{\boldsymbol{\beta}}^{(k)\top} \mathbf{X}_i^\top\right) \widehat{\boldsymbol{\Omega}}_i^{-1(k)} \mathbf{Z}_i \widehat{\mathbf{B}}_i^{(k)} + \widehat{t_i\mathbf{y}_i}^{(k)} \widehat{\mathbf{s}}_i^{(k)\top}, \\
\widehat{t_i\mathbf{b}_i^\top}^{(k)} &= \mathbb{E}\left[T_i\mathbf{b}_i^\top \middle| \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] \\
&= \left(\widehat{t_i\mathbf{y}_i^\top}^{(k)} - \widehat{t}_i^{(k)} \widehat{\boldsymbol{\beta}}^{(k)\top} \mathbf{X}_i^\top\right) \widehat{\boldsymbol{\Omega}}_i^{-1(k)} \mathbf{Z}_i \widehat{\mathbf{B}}_i^{(k)} + \widehat{t_i^2}^{(k)} \widehat{\mathbf{s}}_i^{(k)\top},
\end{aligned}$$

with $\widehat{\mathbf{R}}_i^{(k)} = \widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} - 2\widehat{\mathbf{y}}_i^{(k)}\left(\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)} + \mathbf{Z}_i c\widehat{\boldsymbol{\Delta}}^{(k)}\right)^\top + \left(\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)} + \mathbf{Z}_i c\widehat{\boldsymbol{\Delta}}^{(k)}\right)\left(\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)} + \mathbf{Z}_i c\widehat{\boldsymbol{\Delta}}^{(k)}\right)^\top$,

$\widehat{\mathbf{r}}_i^{(k)} = \widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} - 2\widehat{\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top + \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top.$

It is easy to see that the E-step reduces only to the computation of $\widehat{\mathbf{y}}_i$, $\widehat{\mathbf{y}_i\mathbf{y}_i^\top}$, $\widehat{\kappa}_i$ and $\widehat{\kappa_i\mathbf{y}_i}$. These expected values can be determined in closed form using Lemma 2 and Corollary 1, as follows

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

$$\begin{aligned}
\widehat{\mathbf{y}}_i^{(k)} &= \mathbb{E}\left[\mathbf{Y}_i\middle|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] = \mathbf{y}_i,\\
\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} &= \mathbb{E}\left[\mathbf{Y}_i\mathbf{Y}_i^\top\middle|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] = \mathbf{y}_i\mathbf{y}_i^\top,\\
\widehat{\kappa}_i^{(k)} &= \mathbb{E}\left[\mathrm{W}_\Phi\left(\widehat{\bar{\boldsymbol{\lambda}}}_i\widehat{\boldsymbol{\Sigma}}_i^{-1/2}\left(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}} - \mathbf{Z}_i c\widehat{\boldsymbol{\Delta}}\right)\right)\middle|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right]\\
&= \mathrm{W}_\Phi\left(\widehat{\bar{\boldsymbol{\lambda}}}_i^{(k)}\widehat{\boldsymbol{\Sigma}}_i^{-1/2(k)}\left(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)} - \mathbf{Z}_i c\widehat{\boldsymbol{\Delta}}^{(k)}\right)\right),\\
\widehat{\kappa_i\mathbf{y}_i}^{(k)} &= \mathbb{E}\left[\mathbf{Y}_i\mathrm{W}_\Phi\left(\widehat{\bar{\boldsymbol{\lambda}}}_i\widehat{\boldsymbol{\Sigma}}_i^{-1/2}\left(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}} - \mathbf{Z}_i c\widehat{\boldsymbol{\Delta}}\right)\right)\middle|\mathbf{V}_i = \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right]\\
&= \mathbf{y}_i\widehat{\kappa}_i^{(k)},
\end{aligned}$$

with $\mathrm{W}_\Phi(x) = \phi_1(x)/\Phi(x), x \in \mathbb{R}$.

2. If the ith subject has only censored components then from Corollary 1,

$$\begin{aligned}
\widehat{\mathbf{y}}_i^{(k)} &= \mathbb{E}\left[\mathbf{Y}_i\middle|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] = \mathbb{E}\left[\mathbf{W}_i|\widehat{\boldsymbol{\theta}}^{(k)}\right],\\
\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} &= \mathbb{E}\left[\mathbf{Y}_i\mathbf{Y}_i^\top\middle|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] = \mathbb{E}\left[\mathbf{W}_i\mathbf{W}_i^\top|\widehat{\boldsymbol{\theta}}^{(k)}\right],\\
\widehat{\kappa}_i^{(k)} &= \mathbb{E}\left[\mathrm{W}_\Phi\left(\widehat{\bar{\boldsymbol{\lambda}}}_i\widehat{\boldsymbol{\Sigma}}_i^{-1/2}\left(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}} - \mathbf{Z}_i c\widehat{\boldsymbol{\Delta}}\right)\right)\middle|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right]\\
&= \frac{1}{\sqrt{\frac{\pi}{2}\left(1 + \widehat{\bar{\boldsymbol{\lambda}}}_i^{(k)\top}\widehat{\bar{\boldsymbol{\lambda}}}_i^{(k)}\right)}}\frac{L_{n_i}\left(\mathbf{V}_{1i}, \mathbf{V}_{2i}; \widehat{\boldsymbol{\mu}}_i^{(k)}, \widehat{\boldsymbol{\Psi}}_i^{(k)}\right)}{\mathcal{L}_{n_i}\left(\mathbf{V}_{1i}, \mathbf{V}_{2i}; \widehat{\boldsymbol{\mu}}_i^{(k)}, \widehat{\boldsymbol{\Sigma}}_i^{(k)}, \widehat{\bar{\boldsymbol{\lambda}}}_i^{(k)}, 0\right)},\\
\widehat{\kappa_i\mathbf{y}_i}^{(k)} &= \mathbb{E}\left[\mathbf{Y}_i\mathrm{W}_\Phi\left(\widehat{\bar{\boldsymbol{\lambda}}}_i\widehat{\boldsymbol{\Sigma}}_i^{-1/2}\left(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}} - \mathbf{Z}_i c\widehat{\boldsymbol{\Delta}}\right)\right)\middle|\mathbf{V}_i = \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right]\\
&= \mathbb{E}\left[\mathbf{W}_{0i}|\widehat{\boldsymbol{\theta}}^{(k)}\right]\widehat{\kappa}_i^{(k)},
\end{aligned}$$

where $\widehat{\boldsymbol{\Psi}}_i^{(k)} = \widehat{\boldsymbol{\Sigma}}_i^{(k)1/2}\left(\mathbf{I}_{n_i} + \widehat{\bar{\boldsymbol{\lambda}}}_i^{(k)}\widehat{\bar{\boldsymbol{\lambda}}}_i^{(k)\top}\right)^{-1}\widehat{\boldsymbol{\Sigma}}_i^{(k)1/2}$, $\mathbf{W}_{0i} \sim \mathrm{TN}_{n_i}\left(\widehat{\boldsymbol{\mu}}_i^{(k)}, \widehat{\boldsymbol{\Psi}}_i^{(k)}, [\mathbf{V}_{1i}, \mathbf{V}_{2i}]\right)$, and $\mathbf{W}_i \sim \mathrm{TSN}_{n_i}\left(\widehat{\boldsymbol{\mu}}_i^{(k)}, \widehat{\boldsymbol{\Sigma}}_i^{(k)}, \widehat{\bar{\boldsymbol{\lambda}}}_i^{(k)}, [\mathbf{V}_{1i}, \mathbf{V}_{2i}]\right)$.

3. If the ith subject has censored and uncensored components and given that $(\mathbf{Y}_i|\mathbf{V}_i, \mathbf{C}_i)$, $(\mathbf{Y}_i|\mathbf{V}_i, \mathbf{C}_i, \mathbf{Y}_i^o)$, and $(\mathbf{Y}_i^c|\mathbf{V}_i, \mathbf{C}_i, \mathbf{Y}_i^o)$ are equivalent process, then from Proposition 5 and Lemma 2, we have

$$\widehat{\mathbf{y}}_i^{(k)} = \mathbb{E}\left[\mathbf{Y}_i\middle|\mathbf{V}_i, \mathbf{C}_i, \mathbf{Y}_i^o, \widehat{\boldsymbol{\theta}}^{(k)}\right] = \mathrm{vec}(\mathbf{y}_i^o, \widehat{\mathbf{w}}_i^{(k)}),$$

$$
\begin{aligned}
\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} &= \mathbb{E}\left[\mathbf{Y}_i\mathbf{Y}_i^\top\middle|\mathbf{V}_i,\mathbf{C}_i,\mathbf{Y}_i^o,\widehat{\boldsymbol{\theta}}^{(k)}\right] = \begin{pmatrix} \mathbf{y}_i^o\mathbf{y}_i^{o\top} & \mathbf{y}_i^o\widehat{\mathbf{w}}_i^{(k)\top} \\ \widehat{\mathbf{w}}_i^{(k)}\mathbf{y}_i^{o\top} & \widehat{\mathbf{w}_i\mathbf{w}_i^\top} \end{pmatrix}, \\
\widehat{\kappa}_i^{(k)} &= \mathbb{E}\left[\mathrm{W}_\Phi\left(\widehat{\bar{\boldsymbol{\lambda}}}_i\widehat{\boldsymbol{\Sigma}}_i^{-1/2}\left(\mathbf{y}_i-\mathbf{X}_i\widehat{\boldsymbol{\beta}}-\mathbf{Z}_ic\widehat{\boldsymbol{\Delta}}\right)\right)\middle|\mathbf{V}_i,\mathbf{C}_i,\mathbf{Y}_i^o,\widehat{\boldsymbol{\theta}}^{(k)}\right] \\
&= \frac{\phi_1\left(\widehat{\tau}_i^{co(k)};0,1+\widehat{\boldsymbol{\lambda}}_i^{co(k)\top}\widehat{\boldsymbol{\lambda}}_i^{co(k)}\right)}{\Phi_1(\tilde{\tau})}\frac{L_{n_i^c}\left(\mathbf{V}_{1i}^c,\mathbf{V}_{2i}^c;\widehat{\tilde{\boldsymbol{\mu}}}_i^{(k)},\widehat{\tilde{\boldsymbol{\Psi}}}_i^{(k)}\right)}{\mathcal{L}_{n_i^c}\left(\mathbf{V}_{1i}^c,\mathbf{V}_{2i}^c;\widehat{\boldsymbol{\mu}}_i^{co(k)},\widehat{\mathbf{S}}_i^{(k)},\widehat{\boldsymbol{\lambda}}_i^{co(k)},\widehat{\tau}_i^{co(k)}\right)}, \\
\widehat{\kappa_i\mathbf{y}_i}^{(k)} &= \mathbb{E}\left[\mathbf{Y}_i\mathrm{W}_\Phi\left(\widehat{\bar{\boldsymbol{\lambda}}}_i\widehat{\boldsymbol{\Sigma}}_i^{-1/2}\left(\mathbf{y}_i-\mathbf{X}_i\widehat{\boldsymbol{\beta}}-\mathbf{Z}_ic\widehat{\boldsymbol{\Delta}}\right)\right)\middle|\mathbf{V}_i=\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\right] \\
&= \mathrm{vec}(\mathbf{y}_i^o,\widehat{\mathbf{w}}_{0i}^{(k)})\widehat{\kappa}_i^{(k)},
\end{aligned}
$$

where

$$
\widehat{\mathbf{w}}_i^{(k)} = \mathbb{E}\left[\mathbf{W}_i|\widehat{\boldsymbol{\theta}}^{(k)}\right], \quad \widehat{\mathbf{w}_i\mathbf{w}_i^\top} = \mathbb{E}\left[\mathbf{W}_i\mathbf{W}_i^\top|\widehat{\boldsymbol{\theta}}^{(k)}\right], \quad \widehat{\mathbf{w}}_{0i}^{(k)} = \mathbb{E}\left[\mathbf{W}_{0i}|\widehat{\boldsymbol{\theta}}^{(k)}\right],
$$

with $\mathbf{W}_i \sim \mathrm{TESN}_{n_i^c}\left(\widehat{\boldsymbol{\mu}}_i^{co(k)},\widehat{\mathbf{S}}_i^{(k)},\widehat{\boldsymbol{\lambda}}_i^{co(k)},\widehat{\tau}_i^{co(k)},[\mathbf{V}_{1i}^c,\mathbf{V}_{2i}^c]\right)$, $\mathbf{W}_{0i} \sim \mathrm{TN}_{n_i^c}\left(\widehat{\tilde{\boldsymbol{\mu}}}_i^{(k)},\widehat{\tilde{\boldsymbol{\Psi}}}_i^{(k)},[\mathbf{V}_{1i}^c,\mathbf{V}_{2i}^c]\right)$, and $\tilde{\tau} = \dfrac{\tau_i^{co}}{(1+\boldsymbol{\lambda}_i^{co\top}\boldsymbol{\lambda}_i^{co})^{1/2}}$, $\boldsymbol{\lambda}_i^{co} = \mathbf{S}_i^{1/2}\boldsymbol{\varphi}_i^c$, $\tilde{\boldsymbol{\mu}}_i = \boldsymbol{\mu}_i^{co} - \tau_i^{co}\tilde{\boldsymbol{\Psi}}_i\boldsymbol{\varphi}_i^c$, $\tilde{\boldsymbol{\Psi}}_i = \mathbf{S}_i^{1/2}(\mathbf{I}_{n_i^c}+\boldsymbol{\lambda}_i^{co}\boldsymbol{\lambda}_i^{co\top})^{-1}\mathbf{S}_i^{1/2}$.

It can be noted that we need the first and second moments of a TESN distribution. These can be determined in closed-form using recurrence relations. For more details on the computation of these moments, we refer to Galarza *et al.* (2019). These moments can be obtained in the R package `MomTrunc` (Galarza *et al.*, 2020).

The M-step then conditionally maximizes $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$ and obtains a new estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$, as follows:

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}^{(k+1)} &= \left(\sum_{i=1}^n \mathbf{X}_i^\top\widehat{\mathbf{E}}_i^{-1(k)}\mathbf{X}_i\right)^{-1}\sum_{i=1}^n \mathbf{X}_i^\top\widehat{\mathbf{E}}_i^{-1(k)}\left(\widehat{\mathbf{y}}_i^{(k)}-\mathbf{Z}_i\widehat{\mathbf{b}}_i^{(k)}\right), \\
\widehat{\boldsymbol{\Delta}}^{(k+1)} &= \frac{\sum_{i=1}^n \widehat{t_i\mathbf{b}_i}^{(k)}}{\sum_{i=1}^n \widehat{t_i^2}^{(k)}}, \\
\widehat{\boldsymbol{\Gamma}}^{(k+1)} &= \frac{1}{N}\sum_{i=1}^n\left(\widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)} - \widehat{t_i\mathbf{b}_i}^{(k)}\boldsymbol{\Delta}^{(k+1)\top} - \boldsymbol{\Delta}^{(k+1)}\widehat{t_i\mathbf{b}_i}^{(k)\top} + \widehat{t_i^2}^{(k)}\boldsymbol{\Delta}^{(k+1)}\boldsymbol{\Delta}^{(k+1)\top}\right), \\
\widehat{\sigma^2}^{(k+1)} &= \frac{1}{N}\sum_{i=1}^n\left[\widehat{a}_i^{(k)} + \widehat{\boldsymbol{\beta}}^{(k+1)^\top}\mathbf{X}_i^\top\widehat{\mathbf{E}}_i^{-1^{(k)}}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k+1)} - 2\widehat{\boldsymbol{\beta}}^{(k+1)^\top}\mathbf{X}_i^\top\widehat{\mathbf{E}}_i^{-1^{(k)}}\left(\widehat{\mathbf{y}}_i^{(k)}-\mathbf{Z}_i\widehat{\mathbf{b}}_i^{(k)}\right)\right], \\
\widehat{\phi}^{(k+1)} &= \underset{\phi\in(0,1)\times\mathcal{R}^+}{\arg\max}\left(-\frac{1}{2}\log(|\mathbf{E}_i|) - \frac{1}{2\widehat{\sigma^2}^{(k+1)}}\left[\widehat{a}_i^{(k)} - 2\widehat{\boldsymbol{\beta}}^{(k+1)^\top}\mathbf{X}_i^\top\widehat{\mathbf{E}}_i^{-1^{(k)}}\left(\widehat{\mathbf{y}}_i^{(k)}-\mathbf{Z}_i\widehat{\mathbf{b}}_i^{(k)}\right)\right.\right. \\
&+ \left.\left.\widehat{\boldsymbol{\beta}}^{(k+1)^\top}\mathbf{X}_i^\top\widehat{\mathbf{E}}_i^{-1^{(k)}}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k+1)}\right]\right),
\end{aligned}
$$

where $N = \sum_{i=1}^{n} n_i$. The skewness parameter vector, and the parameters of the scale matrix of the random effects $\mathbf{b}$, can be estimated by noting that

$$\widehat{\mathbf{D}}^{(k)} = \widehat{\boldsymbol{\Gamma}}^{(k)} + \widehat{\boldsymbol{\Delta}}^{(k)}\widehat{\boldsymbol{\Delta}}^{(k)\top} \quad \text{and} \quad \widehat{\boldsymbol{\lambda}}^{(k)} = \frac{\widehat{\mathbf{D}}^{(k)^{-1/2}}\widehat{\boldsymbol{\Delta}}^{(k)}}{\left(1 - \widehat{\boldsymbol{\Delta}}^{(k)\top}\widehat{\mathbf{D}}^{(k)^{-1}}\widehat{\boldsymbol{\Delta}}^{(k)}\right)^{1/2}}.$$

This process is iterated until some distance between two successive evaluations of the log-likelihood $\ell(\boldsymbol{\theta}|\mathbf{y})$ in Section 4.3.2, such as $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)}) - \ell(\widehat{\boldsymbol{\theta}}^{(k)})|$ or $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)})/\ell(\widehat{\boldsymbol{\theta}}^{(k)}) - 1|$, becomes small enough, for example, $\epsilon = 10^{-6}$.

4.3.4 Approximate standard errors

In what follows, we reparameterize $\mathbf{D} = \mathbf{F}^2$ for ease of computation and theoretical derivation, where $\mathbf{F}$ is the square root of $\mathbf{D}$, i.e. $\mathbf{F}^{1/2}$, containing $q(q+1)/2$ distinct elements $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{q(q+1)/2})^\top$.

Following Louis (1982), the individual score is determined as

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

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

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

where $\widehat{\mathbf{s}}_i = \left(\widehat{\mathbf{s}}_i(\boldsymbol{\beta})^\top, \widehat{\mathbf{s}}_i(\sigma^2), \widehat{\mathbf{s}}_i(\boldsymbol{\alpha})^\top, \widehat{\mathbf{s}}_i(\boldsymbol{\phi})^\top, \widehat{\mathbf{s}}_i(\boldsymbol{\lambda})^\top\right)^\top$, has elements given by

$$\begin{aligned}
\widehat{\mathbf{s}}_i(\boldsymbol{\beta}) &= \frac{1}{\widehat{\sigma^2}}\left[\mathbf{X}_i^\top \widehat{\mathbf{E}}_i^{-1}\left(\widehat{\mathbf{y}}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}} - \mathbf{Z}_i\widehat{\mathbf{b}}_i\right)\right], \\
\widehat{\mathbf{s}}_i(\sigma^2) &= -\frac{n_i}{2\widehat{\sigma^2}} + \frac{1}{2\widehat{\sigma^2}^2}\left[\widehat{a}_i - 2\widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i^\top \widehat{\mathbf{E}}_i^{-1}\left(\widehat{\mathbf{y}}_i - \mathbf{Z}_i\widehat{\mathbf{b}}_i\right) + \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i^\top \widehat{\mathbf{E}}_i^{-1}\mathbf{X}_i\widehat{\boldsymbol{\beta}}\right], \\
\widehat{\mathbf{s}}_i(\boldsymbol{\alpha}) &= \left(\widehat{\mathbf{s}}_i(\alpha_1), \ldots, \widehat{\mathbf{s}}_i(\alpha_{q(q+1)/2})\right)^\top, \\
\widehat{\mathbf{s}}_i(\boldsymbol{\phi}) &= \left(\widehat{\mathbf{s}}_i(\phi_1), \widehat{\mathbf{s}}_i(\phi_2)\right)^\top, \\
\widehat{\mathbf{s}}_i(\boldsymbol{\lambda}) &= \left(\widehat{\mathbf{s}}_i(\lambda_1), \ldots, \widehat{\mathbf{s}}_i(\lambda_q)\right)^\top,
\end{aligned}$$

with

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

$$
\begin{aligned}
\widehat{\mathbf{s}}_i(\alpha_r) &= -\frac{1}{2}\mathrm{tr}\left(\widehat{\mathbf{\Gamma}}^{-1}\dot{\mathbf{\Gamma}}_{\alpha_r}\right) + \frac{1}{2}\left\{\mathrm{tr}\left(\widehat{\mathbf{\Gamma}}^{-1}\dot{\mathbf{\Gamma}}_{\alpha_r}\widehat{\mathbf{\Gamma}}^{-1}\widehat{\mathbf{b}_i\mathbf{b}_i^\top}\right) + \widehat{t_i\mathbf{b}_i^\top}\left(\widehat{\mathbf{\Gamma}}^{-1}\dot{\mathbf{F}}_r\widehat{\boldsymbol{\delta}} - \widehat{\mathbf{\Gamma}}^{-1}\dot{\mathbf{\Gamma}}_{\alpha_r}\widehat{\mathbf{\Gamma}}^{-1}\widehat{\mathbf{\Delta}}\right)\right.\\
&+ \left(\widehat{\mathbf{\Gamma}}^{-1}\dot{\mathbf{F}}_r\widehat{\boldsymbol{\delta}} - \widehat{\mathbf{\Gamma}}^{-1}\dot{\mathbf{\Gamma}}_{\alpha_r}\widehat{\mathbf{\Gamma}}^{-1}\widehat{\mathbf{\Delta}}\right)^\top\widehat{t_i\mathbf{b}_i} - \widehat{t_i^2}\left[\widehat{\mathbf{\Delta}}^\top\left(\widehat{\mathbf{\Gamma}}^{-1}\dot{\mathbf{F}}_r\widehat{\boldsymbol{\delta}} - \widehat{\mathbf{\Gamma}}^{-1}\dot{\mathbf{\Gamma}}_{\alpha_r}\widehat{\mathbf{\Gamma}}^{-1}\widehat{\mathbf{\Delta}}\right)\right.\\
&+ \left.\left.\widehat{\boldsymbol{\delta}}^\top\dot{\mathbf{F}}_r\widehat{\mathbf{\Gamma}}^{-1}\widehat{\mathbf{\Delta}}\right]\right\},\\
\widehat{\mathbf{s}}_i(\phi_s) &= \frac{1}{2\widehat{\sigma}^2}\left[\mathrm{tr}\left(\widehat{\mathbf{y}_i\mathbf{y}_i^\top} - 2\widehat{\mathbf{y}_i\mathbf{b}_i^\top}\mathbf{Z}_i^\top + \mathbf{Z}_i\widehat{\mathbf{b}_i\mathbf{b}_i^\top}\mathbf{Z}_i^\top - 2\left(\widehat{\mathbf{y}}_i - \mathbf{Z}_i\widehat{\mathbf{b}}_i\right)\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_i^\top\right.\right.\\
&+ \left.\left.\mathbf{X}_i\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_i^\top\right)\widehat{\mathbf{E}}_i^{-1}\dot{\mathbf{E}}_i^s\widehat{\mathbf{E}}_i^{-1}\right] - \frac{1}{2}\mathrm{tr}\left(\widehat{\mathbf{E}}_i^{-1}\dot{\mathbf{E}}_i^s\right),\\
\widehat{\mathbf{s}}_i(\lambda_t) &= -\frac{1}{2}\mathrm{tr}\left(\widehat{\mathbf{\Gamma}}^{-1}\dot{\mathbf{\Gamma}}_{\lambda_t}\right) + \frac{1}{2}\left\{\mathrm{tr}\left(\widehat{\mathbf{\Gamma}}^{-1}\dot{\mathbf{\Gamma}}_{\lambda_t}\widehat{\mathbf{\Gamma}}^{-1}\widehat{\mathbf{b}_i\mathbf{b}_i^\top}\right) + \widehat{t_i\mathbf{b}_i^\top}\left(\widehat{\mathbf{\Gamma}}^{-1}\dot{\mathbf{\Delta}}_{\lambda_t} - \widehat{\mathbf{\Gamma}}^{-1}\dot{\mathbf{\Gamma}}_{\lambda_t}\widehat{\mathbf{\Gamma}}^{-1}\widehat{\mathbf{\Delta}}\right)\right.\\
&+ \left(\widehat{\mathbf{\Gamma}}^{-1}\dot{\mathbf{\Delta}}_{\lambda_t} - \widehat{\mathbf{\Gamma}}^{-1}\dot{\mathbf{\Gamma}}_{\lambda_t}\widehat{\mathbf{\Gamma}}^{-1}\widehat{\mathbf{\Delta}}\right)^\top\widehat{t_i\mathbf{b}_i}\\
&- \left.\widehat{t_i^2}\left[\widehat{\mathbf{\Delta}}^\top\left(\widehat{\mathbf{\Gamma}}^{-1}\dot{\mathbf{\Delta}}_{\lambda_t} - \widehat{\mathbf{\Gamma}}^{-1}\dot{\mathbf{\Gamma}}_{\lambda_t}\widehat{\mathbf{\Gamma}}^{-1}\widehat{\mathbf{\Delta}}\right) + \dot{\mathbf{\Delta}}_{\lambda_t}^\top\widehat{\mathbf{\Gamma}}^{-1}\widehat{\mathbf{\Delta}}\right]\right\},
\end{aligned}
$$

where

$$
\begin{aligned}
\dot{\mathbf{\Gamma}}_{\alpha_r} &= \left.\frac{\partial\mathbf{\Gamma}}{\partial\alpha_r}\right|_{\boldsymbol{\alpha}=\widehat{\boldsymbol{\alpha}}} = \mathbf{F}\dot{\mathbf{F}}_r + \dot{\mathbf{F}}_r\mathbf{F} - \mathbf{F}\boldsymbol{\delta}\boldsymbol{\delta}^\top\dot{\mathbf{F}}_r - \dot{\mathbf{F}}_r\boldsymbol{\delta}\boldsymbol{\delta}^\top\mathbf{F},\\
\dot{\mathbf{F}}_r &= \left.\frac{\partial\mathbf{F}}{\partial\alpha_r}\right|_{\boldsymbol{\alpha}=\widehat{\boldsymbol{\alpha}}},\ r = 1,\ldots,q(q+1)/2,\\
\dot{\mathbf{\Gamma}}_{\lambda_t} &= \left.\frac{\partial\mathbf{\Gamma}}{\partial\lambda_t}\right|_{\boldsymbol{\lambda}=\widehat{\boldsymbol{\lambda}}} = -\mathbf{F}\left(\frac{\dot{\boldsymbol{\lambda}}_{2t}}{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}} - \frac{2\lambda_t\boldsymbol{\lambda}\boldsymbol{\lambda}^\top}{(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^2}\right)\mathbf{F},\\
\dot{\mathbf{\Delta}}_{\lambda_t} &= \left.\frac{\partial\mathbf{\Delta}}{\partial\lambda_t}\right|_{\boldsymbol{\lambda}=\widehat{\boldsymbol{\lambda}}} = \mathbf{F}\left(\frac{\dot{\boldsymbol{\lambda}}_t}{(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}} - \frac{\lambda_t\boldsymbol{\lambda}}{(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{3/2}}\right),\\
\dot{\boldsymbol{\lambda}}_t &= \left.\frac{\partial\boldsymbol{\lambda}}{\partial\lambda_t}\right|_{\boldsymbol{\lambda}=\widehat{\boldsymbol{\lambda}}},\quad \dot{\boldsymbol{\lambda}}_{2t} = \left.\frac{\partial\boldsymbol{\lambda}\boldsymbol{\lambda}^\top}{\partial\lambda_t}\right|_{\boldsymbol{\lambda}=\widehat{\boldsymbol{\lambda}}},\quad t = 1,\ldots,q,\\
\dot{\mathbf{E}}_i^s &= \left.\frac{\partial\mathbf{E}_i}{\partial\phi_s}\right|_{\boldsymbol{\phi}=\widehat{\boldsymbol{\phi}}},\ s = 1,2.
\end{aligned}
$$

For the DEC structure we have that

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

4.3.5 Estimation of the random effects

In this section, we consider an empirical Bayes inference for the random effects that is useful for interpreting the subject-specific variability. From (4.3)-(4.4), it implies that $\mathbf{Y}_i|\mathbf{b}_i \sim \mathrm{N}_{n_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \mathbf{\Omega}_i)$ and $\mathbf{b}_i \sim \mathrm{SN}_q(c\mathbf{\Delta}, \mathbf{D}, \boldsymbol{\lambda})$. The conditional distribution of $\mathbf{b}_i$ given $\mathbf{Y}_i$ belong to the extended skew-normal (ESN), and its pdf is

$$
\begin{aligned}
f(\mathbf{b}_i|\mathbf{Y}_i) &= \frac{f(\mathbf{Y}_i|\mathbf{b}_i)f(\mathbf{b}_i)}{\int f(\mathbf{Y}_i|\mathbf{b}_i)f(\mathbf{b}_i)\mathrm{d}\mathbf{b}_i}\\
&= \frac{\phi_q\left(\mathbf{b}_i; c\mathbf{\Delta} + \mathbf{D}\mathbf{Z}_i^\top\mathbf{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_ic\mathbf{\Delta}), \mathbf{\Lambda}_i\right)\Phi_1\left(\boldsymbol{\lambda}^\top\mathbf{D}^{-1/2}(\mathbf{b}_i - c\mathbf{\Delta})\right)}{\Phi_1\left(\bar{\boldsymbol{\lambda}}_i^\top\mathbf{\Sigma}_i^{-1/2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_ic\mathbf{\Delta})\right)},
\end{aligned}
$$

i.e.,

$$\begin{aligned}\mathbf{b}_i|\mathbf{Y}_i &\sim \mathrm{ESN}_q(c\boldsymbol{\Delta} + \mathbf{D}\mathbf{Z}_i^\top\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i c\boldsymbol{\Delta}), \boldsymbol{\Lambda}_i, \boldsymbol{\Lambda}_i^{1/2}\boldsymbol{\zeta},\\ &\quad \boldsymbol{\zeta}^\top\mathbf{D}\mathbf{Z}_i^\top\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i c\boldsymbol{\Delta})).\end{aligned}$$

Thus, from Lemma 1, it follows that

$$\begin{aligned}\mathbb{E}\left[\mathbf{b}_i\middle|\mathbf{Y}_i = \mathbf{y}_i, \boldsymbol{\theta}\right] &= c\boldsymbol{\Delta} + \mathbf{D}\mathbf{Z}_i^\top\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i c\boldsymbol{\Delta})\\ &+ \frac{\boldsymbol{\Lambda}_i\boldsymbol{\zeta}}{\sqrt{1+\boldsymbol{\zeta}^\top\boldsymbol{\Lambda}_i\boldsymbol{\zeta}}}\mathrm{W}_\Phi\left(\bar{\boldsymbol{\lambda}}_i\boldsymbol{\Sigma}_i^{-1/2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i c\boldsymbol{\Delta})\right).\end{aligned}$$

The minimum mean-squared error (MSE) estimator of $\mathbf{b}_i$ obtained by the conditional mean of $\mathbf{b}_i$ given $\mathbf{V}_i$ and $\mathbf{C}_i$ is

$$\begin{aligned}\widehat{\mathbf{b}}_i(\boldsymbol{\theta}) &= \mathbb{E}\left[\mathbf{b}_i|\mathbf{V}_i, \mathbf{C}_i\right] = \mathbb{E}\left[\mathbb{E}(\mathbf{b}_i|\mathbf{Y}_i, \boldsymbol{\theta})|\mathbf{V}_i, \mathbf{C}_i\right]\\ &= c\boldsymbol{\Delta} + \mathbf{D}\mathbf{Z}_i^\top\boldsymbol{\Sigma}_i^{-1}(\widehat{\mathbf{y}}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i c\boldsymbol{\Delta}) + \frac{\boldsymbol{\Lambda}_i\boldsymbol{\zeta}}{\sqrt{1+\boldsymbol{\zeta}^\top\boldsymbol{\Lambda}_i\boldsymbol{\zeta}}}\widehat{\kappa}_i,\end{aligned}$$

where $\widehat{\mathbf{y}}_i = \mathbb{E}[\mathbf{Y}_i\big|\mathbf{V}_i, \mathbf{C}_i]$ and $\widehat{\kappa}_i = \mathbb{E}[\mathrm{W}_\Phi(\cdot)\big|\mathbf{V}_i, \mathbf{C}_i]$ depend on the censoring pattern of subject i (see Subsection 4.3.3).

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

4.3.6 Prediction of future observations

The prediction problem for longitudinal data is also of great importance in a number of practical applications. Rao *et al.* (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-SNLMEC model, we use the approach proposed by Wang (2013). Thus, let $\mathbf{y}_{i,\mathrm{obs}}$ be an observed response vector of dimension $n_{i,\mathrm{obs}} \times 1$ for a new subject i over the first portion of time and $\mathbf{y}_{i,\mathrm{pred}}$ be the corresponding $n_{i,\mathrm{pred}} \times 1$ response vector over the future portion of time. Moreover, let $\mathbf{X}_i^* = (\mathbf{X}_{i,\mathrm{obs}}, \mathbf{X}_{i,\mathrm{pred}})$ and $\mathbf{Z}_i^* = (\mathbf{Z}_{i,\mathrm{obs}}, \mathbf{Z}_{i,\mathrm{pred}})$ denote the $(n_{i,\mathrm{obs}} + n_{i,\mathrm{pred}}) \times p$ and $(n_{i,\mathrm{obs}} + n_{i,\mathrm{pred}}) \times q$ design matrices corresponding to $\bar{\mathbf{y}}_i = (\mathbf{y}_{i,\mathrm{obs}}^\top, \mathbf{y}_{i,\mathrm{pred}}^\top)$.

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

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

where $\boldsymbol{\Sigma}_i^* = \begin{pmatrix} \boldsymbol{\Sigma}_i^{\text{obs*,obs*}} & \boldsymbol{\Sigma}_i^{\text{obs*,pred}} \\ \boldsymbol{\Sigma}_i^{\text{pred,obs*}} & \boldsymbol{\Sigma}_i^{\text{pred,pred}} \end{pmatrix}$, $\bar{\boldsymbol{\lambda}}_i^* = \dfrac{\boldsymbol{\Sigma}_i^{*-1/2}\mathbf{Z}_i^*\mathbf{D}\boldsymbol{\zeta}}{\sqrt{1+\boldsymbol{\zeta}^\top\boldsymbol{\Lambda}_i^*\boldsymbol{\zeta}}}$. As mentioned in Wang (2013), the best linear predictor of $\mathbf{y}_{i,\text{pred}}$ with respect to the minimum mean squared error (MSE) criterion is the conditional expectation of $\mathbf{y}_{i,\text{pred}}$ given $\mathbf{y}_{i,\text{obs*}}$, which, from Proposition 5, is given by

$$\widehat{\mathbf{y}}_{i,\text{pred}}(\boldsymbol{\theta}) = \boldsymbol{\mu}^* + \mathrm{W}_\Phi\left(\frac{\tau^*}{\sqrt{1+\mathbf{v}_{2i}^\top\mathbf{S}_i\mathbf{v}_{2i}}}\right)\frac{\mathbf{S}_i\mathbf{v}_{2i}}{\sqrt{1+\mathbf{v}_{2i}^\top\mathbf{S}_i\mathbf{v}_{2i}}}, \tag{4.10}$$

where $\mathbf{v}_i = (\mathbf{v}_{1i}^\top, \mathbf{v}_{2i}^\top)^\top = \boldsymbol{\Sigma}_i^{*-1/2}\bar{\boldsymbol{\lambda}}_i^*$, $\mathbf{S}_i = \boldsymbol{\Sigma}_i^{\text{pred,pred}} - \boldsymbol{\Sigma}_i^{\text{pred,obs*}}(\boldsymbol{\Sigma}_i^{\text{obs*,obs*}})^{-1}\boldsymbol{\Sigma}_i^{\text{obs*,pred}}$, $\boldsymbol{\mu}^* = \mathbf{X}_{i,\text{pred}}\boldsymbol{\beta} + \mathbf{Z}_{i,\text{pred}}c\boldsymbol{\Delta} + \boldsymbol{\Sigma}_i^{\text{pred,obs*}}(\boldsymbol{\Sigma}_i^{\text{obs*,obs*}})^{-1}(\mathbf{y}_{i,\text{obs*}} - \mathbf{X}_{i,\text{obs*}}\boldsymbol{\beta} - \mathbf{Z}_{i,\text{obs*}}c\boldsymbol{\Delta})$,
$\tau^* = \left(\mathbf{v}_{1i} + (\boldsymbol{\Sigma}_i^{\text{obs*,obs*}})^{-1}\boldsymbol{\Sigma}_i^{\text{obs*,pred}}\mathbf{v}_{2i}\right)^\top\left(\mathbf{y}_{i,\text{obs*}} - \mathbf{X}_{i,\text{obs*}}\boldsymbol{\beta} - \mathbf{Z}_{i,\text{obs*}}c\boldsymbol{\Delta}\right)$.

Therefore, $\mathbf{y}_{i,\text{pred}}$ can be estimated directly by substituting $\widehat{\boldsymbol{\theta}}$ into (4.10), leading to $\widehat{\mathbf{y}_{i,\text{pred}}} = \widehat{\mathbf{y}}_{i,\text{pred}}(\widehat{\boldsymbol{\theta}})$.

4.4 The Nonlinear case

Extending the notation of the previous section and ignoring censoring, we first propose the following general mixed-effects model in which the random terms are assumed to follow a multivariate skew-normal distribution (SN-NLME).

Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$ denote the response vector for subject i and $\mathbf{f}_i(\mathbf{X}_i, \psi_i) = (f(\mathbf{X}_{i1}, \psi_i), \ldots, f(\mathbf{X}_{in_i}, \psi_i))^\top$ be a nonlinear vector-valued differentiable function of the random parameter ψ_i and covariate vector $\mathbf{X}_i$. The SN-NLME can then be expressed as

$$\mathbf{y}_i = f(\mathbf{X}_i, \psi_i) + \epsilon_i, \quad \psi_i = \mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \quad i = 1, \ldots, n, \tag{4.11}$$

where the joint distribution of $(\mathbf{b}_i, \boldsymbol{\epsilon}_i)$ is as in (4.4), $\boldsymbol{\beta}$ is a p-vector of fixed population parameters, $\mathbf{b}_i$ is a q-vector of random effects associated with subject i, $\mathbf{A}_i$ and $\mathbf{B}_i$ are know design matrices of dimensions $r \times p$ and $r \times q$ for the fixed and random effects, respectively.

As mentioned by Vaida & Liu (2009), the linearization (L) procedure to obtain the approximate MLE of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \boldsymbol{\alpha}^\top, \boldsymbol{\phi}^\top, \boldsymbol{\lambda}^\top)^\top$ involves taking the first-order Taylor expansion of f 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)

$$\widetilde{\mathbf{y}}_i = \widetilde{\mathbf{X}}_i\boldsymbol{\beta} + \widetilde{\mathbf{Z}}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n, \tag{4.12}$$

where $\widetilde{\mathbf{y}}_i = \mathbf{y}_i - f(\mathbf{A}_i\widetilde{\boldsymbol{\beta}} + \mathbf{B}_i\widetilde{\mathbf{b}}_i, \mathbf{X}_i) - \widetilde{\mathbf{X}}_i\widetilde{\boldsymbol{\beta}} - \widetilde{\mathbf{Z}}_i\widetilde{\mathbf{b}}_i$ and

$$\widetilde{\mathbf{X}}_i = \frac{\partial f(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i)}{\partial\boldsymbol{\beta}^\top}\Big|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}, \qquad \widetilde{\mathbf{Z}}_i = \frac{\partial f(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i)}{\partial\mathbf{b}_i^\top}\Big|_{\mathbf{b}_i=\widetilde{\mathbf{b}}_i}.$$

Therefore, for the censored response the linearized model (4.12) is an LME with censored data, with same structure as (4.3), which is then solved as indicated in the previous section. The model matrix for (4.12) depends on the current parameter value, and needs to be recalculated at each iteration. The algorithm iterates between L-,E-, and M-steps until convergence.

4.4.1 Starting values

It is well known that maximum likelihood estimation in nonlinear mixed models may face some computational hurdles, in the sense that the method may not give maximum global solutions if the starting values are far from the real parameter values. Thus, the choice of starting values for the EM algorithm in the nonlinear context plays a big role in parameter estimation. To circumvent such a limitation, a convenient way is to initialize the EM algorithm with a variety of initial values that are representatives of the parameter space. In practice, a default procedure for obtaining reasonable initial values is summarized below.

- Compute $\widehat{\beta}^{(0)}$, $\widehat{\sigma^2}^{(0)}$, $\widehat{\mathbf{D}}^{(0)}$ and $\widehat{\mathbf{b}}^{(0)}$ using the NLME model through the library `nlme()` in R software, for instance.
- The initial value for the skewness parameter $\boldsymbol{\lambda}$ is obtained in the following way: Let $\widehat{\rho}_l$ be the sample skewness coefficient of the lth column of $\widehat{\mathbf{b}}^{(0)}$, obtained under normality. Then, we let $\widehat{\lambda}_l^{(0)} = 3 \times sign(\widehat{\rho}_l)$, $l = 1, \ldots, q$.
- The initial values for ϕ, depending on the structure, are simply chosen to give a condition of nearly uncorrelated errors.

Even though these procedures look reasonable for computing the starting points, the tradition in practice is to try several initial values for the EM algorithm, in order to get the highest likelihood value.

4.5 Simulation studies

In order to study the performance of our proposed model and algorithm, we present two simulation studies. The first simulation study is to show that the parameter estimates based on the EM algorithm of the SN-NLMEC models provides good asymptotic properties. The goal of the second simulation study is to compare the behavior and performance of the NLMEC in the presence of asymmetry. Lastly, we present a third simulation study in which attention is focused on comparing the predictive abilities of the proposed SN-NLMEC model. The computational procedures were implemented using the R software (R Core Team, 2020).

The simulation study was based on the model proposed for the AIEDRP data discussed in Section 4.6.2. We considered a similar logistic model (4.16) (see Application 2) with random set-points value α_{1i} and random decreases from the maximum HIV RNA α_{2i}, as follows

$$y_{ij} = \alpha_{1i} + \frac{\alpha_{2i}}{1 + \exp\left((t_{ij} - \alpha_3)/\alpha_4\right)} + \epsilon_{ij}, \tag{4.13}$$

with $i = 1, \ldots, n$, $j = 1, \ldots, 10$, $\alpha_{1i} = \exp(\beta_1 + b_{1i})$, $\alpha_{2i} = \exp(\beta_2 + b_{2i})$, $\beta_k = \log(\alpha_k)$, $k = 3, 4$, $(b_{0i}, b_{1i}) \overset{\text{iid.}}{\sim} \mathrm{SN}_2(c\boldsymbol{\Delta}, \mathbf{D}, \boldsymbol{\lambda})$, $\boldsymbol{\epsilon}_i \overset{\text{iid.}}{\sim} \mathrm{N}_{10}(\mathbf{0}, \boldsymbol{\Omega}_i)$, such that $\boldsymbol{\Omega}_i = \sigma^2\mathbf{E}_i$. The parameters in the simulations were chosen similarly to the estimated values based on the original data using SN-NLMEC under uncorrelated (UNC) structure : $\boldsymbol{\beta} = (1.53, 0.71, 3.51, 1.78)^\top$, $\sigma^2 = 0.22$, $\mathbf{D}$ with elements $\alpha_{11} = 0.09$, $\alpha_{12} = \alpha_{21} = -0.16$, $\alpha_{22} = 0.43$, and $\boldsymbol{\lambda} = (-5, 3)^\top$.

4.5.1 Simulation study 1

The first simulation study examines the finite sample behavior of ML estimates obtained through our proposed EM algorithm. The parameter settings are identical to those given above. For this simulation, the samples sizes were fixed as $n = 50, 150, 300, 450$ and 600 and the correlation structure of the error term was a continuous-time AR(1) model with $\phi_1 = 0.7$. For each sample size, 500 samples from the SN-NLMEC model with 10% and 20% of censoring proportion were generated.

For this purpose, we analyzed the absolute bias (Bias) and the mean square error (MSE) of the ML estimates obtained from the SN-NLMEC model for five different sample sizes. These measures are defined by

$$\mathrm{Bias}(\theta_i) = \frac{1}{500}\sum_{j=1}^{500} |\widehat{\theta}_i^{(j)} - \theta_i| \quad \text{and} \quad \mathrm{MSE}(\theta_i) = \frac{1}{500}\sum_{j=1}^{500} (\widehat{\theta}_i^{(j)} - \theta_i)^2, \tag{4.14}$$

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

Figures 16, 17 and 18 show that the Bias and the MSE of the parameter estimates of $\boldsymbol{\theta}$, σ^2, $\boldsymbol{\alpha}$, ϕ_1 and $\boldsymbol{\lambda}$ tends to zero as the sample size increases. For the parameter $\boldsymbol{\lambda}$, we noticed that the Bias and MSE values are a little high. This can be explained by the wide estimation range of the parameter - see Figure 19. We can also notice from Figure 19 that the variation decreases as the sample size increases. In conclusion, the results provide empirical evidence about the consistency of the ML estimates of the SN-NLMEC model, even considering the linearization procedure described in Section 4.4.

4.5.2 Simulation study 2

For the second study, we simulated 500 datasets from the SN-NLMEC model (4.13) and we considered 10% and 20% of the observations in each dataset were censored with samples sizes $n = 50$ and 150. Once the simulated datasets were generated, we fitted

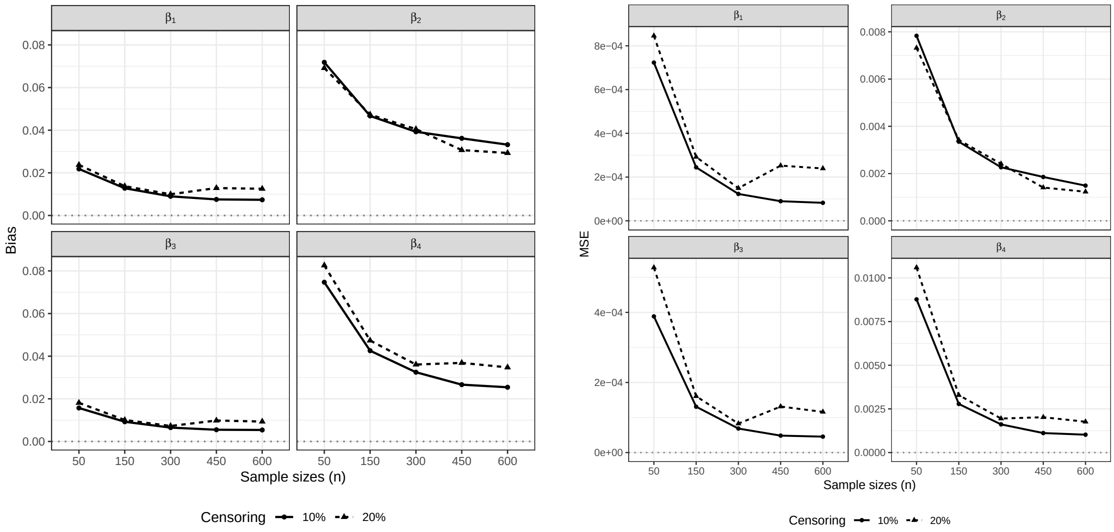


Figure 16 – **Simulation study 1**. Bias and MSE of $\boldsymbol{\beta}$ estimates under the AR(1) model for different sample sizes.

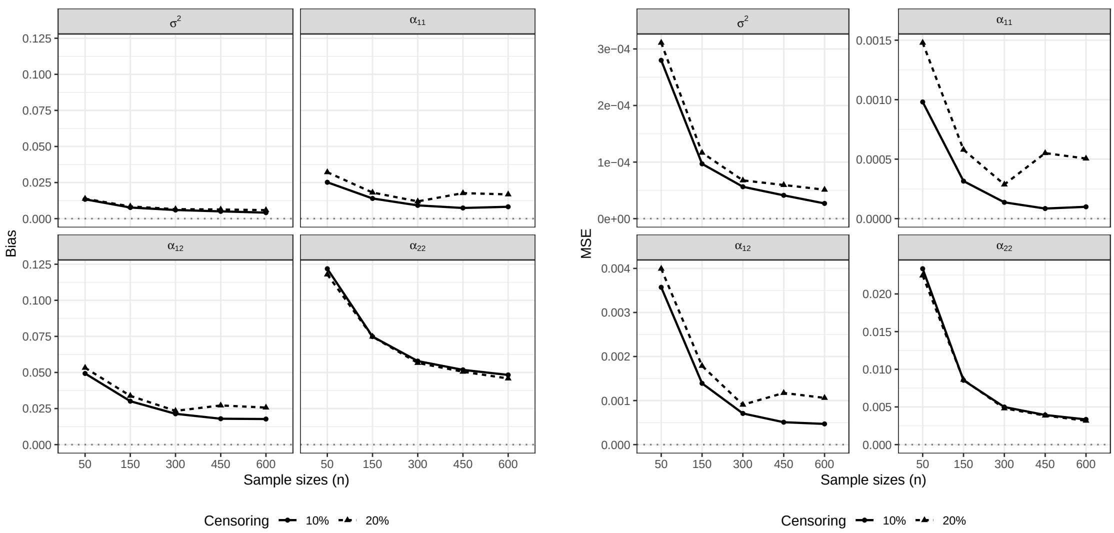


Figure 17 – **Simulation study 1**. Bias and MSE of σ^2 and $\boldsymbol{\alpha}$ estimates under the AR(1) model for different sample sizes.

the SN-NLMEC model under the uncorrelated (UNC) structure and the NLMEC model using the `nlmmcl()` function provided by Vaida & Liu (2009). The model selection criteria (AIC and BIC) as well as the estimates of the model parameters were stored for each simulation. We evaluate the models by comparing the estimates of the parameters with their true values based on the absolute bias (4.14). The simulation results are summarized in Table 8. We see that when we fitted the NLMEC model to asymmetric data, the $\boldsymbol{\alpha}$ parameter estimates are the most affected. The α_{11} and α_{22} components are underestimated and the α_{12} component is overestimated. For the parameters $\boldsymbol{\beta}$ and σ^2 both models give similar estimates, but the SN-NLMEC produces smaller bias with relation to $\boldsymbol{\beta}$.

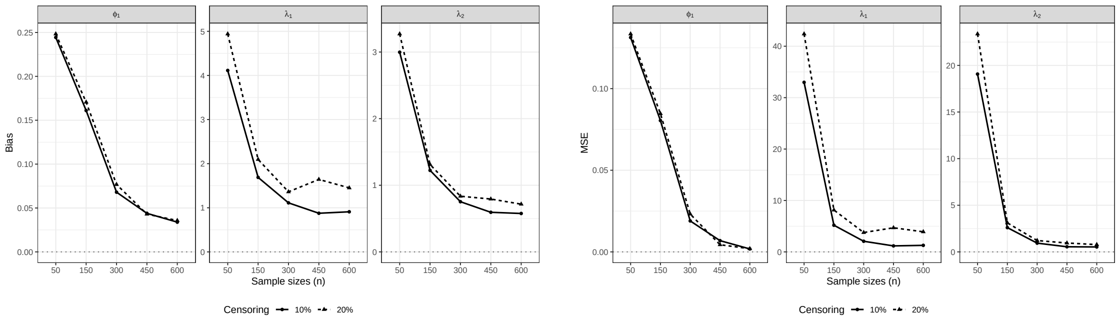


Figure 18 – **Simulation study 1**. Bias and MSE of ϕ_1 and $\boldsymbol{\lambda}$ estimates under the AR(1) model for different sample sizes.

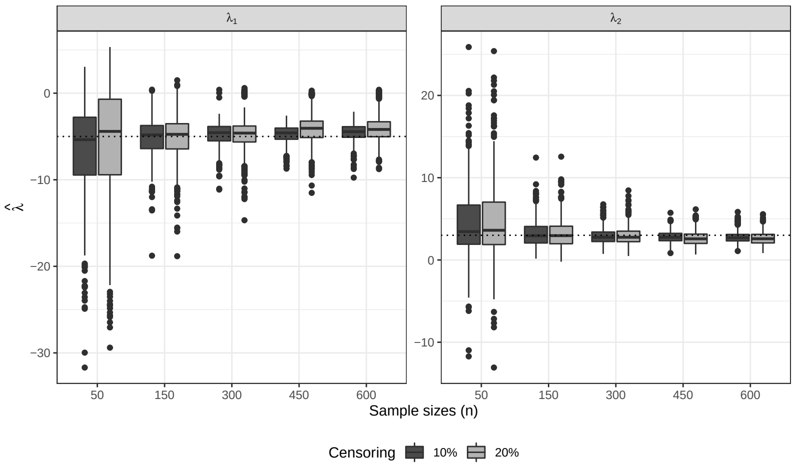


Figure 19 – **Simulation study 1**. Boxplots of the $\boldsymbol{\lambda}$ estimates under the AR(1) model for different sample sizes. Dotted lines indicate the true parameter value.

4.5.3 Simulation study 3

The third simulation study analyzes the performance of the prediction of future values described in Subsection 4.3.6. For this purpose, we use the pseudo-cross-validation approach to assess their predictive performances. This approach of comparing forecasts with the corresponding actual values. We generated 500 datasets of size $n = 100$ and $n = 200$ under the AR(1) structure with parameter $\phi_1 = 0.7$, considering two different settings of censoring proportions, 5% and 15%. Then, we drop out the last two measurements y_{i9}, y_{i10} on the ith individual, we compute the ML estimates using the remaining data as the sample and the prediction of $\mathbf{y}_i = (y_{i9}, y_{i10})$, denoted by $\widehat{\mathbf{y}}_i = (\widehat{y}_{i9}, \widehat{y}_{i10})$, is made.

To evaluate prediction accuracies, we considered two measures of accuracy, namely the MARE (Mean Absolute Relative Error) and MSRE (Mean Square Relative Error). These measures are given by

$$\text{MARE} = \frac{1}{2n}\sum_{ij}\left|\frac{y_{ij}-\widehat{y}_{ij}}{y_{ij}}\right| \quad \text{and} \quad \text{MSRE} = \frac{1}{2n}\sum_{ij}\left(\frac{y_{ij}-\widehat{y}_{ij}}{y_{ij}}\right)^2,$$

Table 8 – **Simulation study 2**. Simulation results based on 500 simulated samples.

Censoring 10%														
			Parameters								Criteria		
Distribution			β_1	β_2	β_3	β_4	σ^2	α_{11}	α_{12}	α_{22}	loglik	AIC	BIC
$n = 50$	SN	MC Mean	1.5284	0.7500	3.5062	1.7622	0.2214	0.0778	-0.1428	0.4190	-449.9404	919.8807	962.0268
		Bias	0.0217	0.0687	0.0151	0.0735	0.0131	0.0243	0.0478	0.1215			
	N	MC Mean	1.5306	0.7497	3.5037	1.7539	0.2209	0.0309	-0.0616	0.2516	-459.1068	934.2135	967.9304
		Bias	0.0221	0.0693	0.0154	0.0748	0.0131	0.0591	0.0984	0.1791			
$n = 150$	SN	MC Mean	1.5325	0.7443	3.5073	1.7653	0.2208	0.0829	-0.1461	0.3985	-1353.274	2726.549	2779.681
		Bias	0.0128	0.0457	0.0090	0.0428	0.0075	0.0138	0.0298	0.0750			
	N	MC Mean	1.5341	0.7457	3.5044	1.7555	0.2201	0.0306	-0.0617	0.2556	-1380.36	2776.72	2819.226
		Bias	0.0132	0.0466	0.0099	0.0456	0.0075	0.0594	0.0983	0.1744			
Censoring 20%													
			Parameters								Criteria		
Distribution			β_1	β_2	β_3	β_4	σ^2	α_{11}	α_{12}	α_{22}	loglik	AIC	BIC
$n = 50$	SN	MC Mean	1.5323	0.7449	3.5040	1.7548	0.2214	0.0721	-0.1370	0.4260	-427.8434	875.6868	917.8329
		Bias	0.0223	0.0670	0.0163	0.0783	0.0140	0.0285	0.0485	0.1191			
	N	MC Mean	1.5348	0.7467	3.5003	1.7469	0.2211	0.0286	-0.0582	0.2489	-457.5571	931.1142	964.831
		Bias	0.0232	0.0687	0.0171	0.0800	0.0140	0.0614	0.1018	0.1817			
$n = 150$	SN	MC Mean	1.5336	0.7442	3.5062	1.7615	0.2212	0.0800	-0.1430	0.4027	-1287.648	2595.295	2648.428
		Bias	0.0141	0.0464	0.0097	0.0473	0.0082	0.0180	0.0336	0.0753			
	N	MC Mean	1.5380	0.7434	3.5014	1.7509	0.2204	0.0285	-0.0588	0.2537	-1376.61	2769.219	2811.725
		Bias	0.0149	0.0456	0.0116	0.0505	0.0081	0.0615	0.1012	0.1763			

where y_{ij} is the original value and $\hat{y}_{ij}$ is the predicted value, for $i = 1, \ldots, n$ and $j = 1, 2$.

Table 9 shows the comparison between the predicted values and real ones under the SN-NLMEC model considring AR(1) structure. One can see from these results that the SN-NLMEC predictor performs encouragingly well in both cases since the values of the two measures are close to zero and, as expected, these values increase as the censoring level increase.

Table 9 – **Simulation study 3**. Evaluation of the prediction accuracy for the SN-NLMEC model.

	Censoring 5%		Censoring 15%	
n	MARE	MSRE	MARE	MSRE
100	0.1106	0.0233	0.1521	0.043
200	0.1103	0.0229	0.1522	0.0421

4.6 Illustrative examples

4.6.1 A5055 data

This section illustrates the performance of the proposed methods with the analysis of A5055 data.

As was mentioned in the Introduction, the dataset consists of 44 patients were randomized in one of two regimens and plasma HIV-1 RNA (viral load) was measured (copies/mL) in blood samples collected irregularly on study days 0, 7, 14, 28, 56, 84, 112, 140, and 168 of follow-up. The nucleic acid sequence-based amplification assay (NASBA) was used to measure plasma HIV-1 RNA, with a lower limit of quantification of 50 copies/mL, and there were 106 out of 316 (around 33.54%) RNA viral load measurements below the detection limit, so there was left censoring. A series of potentially explanatory variables was collected at the same time. For the data analysis, we consider only the covariate $\mathrm{CD4}^+$ cell counts. The number of measurements per subject varied from 1 to 9. Figure 15b shows the longitudinal trajectories of RNA viral load (in log-base-10 scale) across days for patients. It can be noted that the viral load trajectory is complex and is substantially different across individuals.

This data was previously analyzed by Lachos *et al.* (2019) using the scale mixtures of normal distribution (SMN) in the multivariate censored linear mixed effect (MLMEC) model. Wang *et al.* (2018) analyzed this data using multivariate t linear mixed-effects models with censored observations. In Figure 20 we can see the scatter plot of the estimates of random effects obtained by fitting a LMEC model using `R` package `lmec` (Vaida & Liu, 2012) and the boxplots of estimates. The plots reveal subject-specific behaving somewhat asymmetrically. Therefore, an assumption of symmetric distribution for random effects is not very realistic for the A5055 data set.

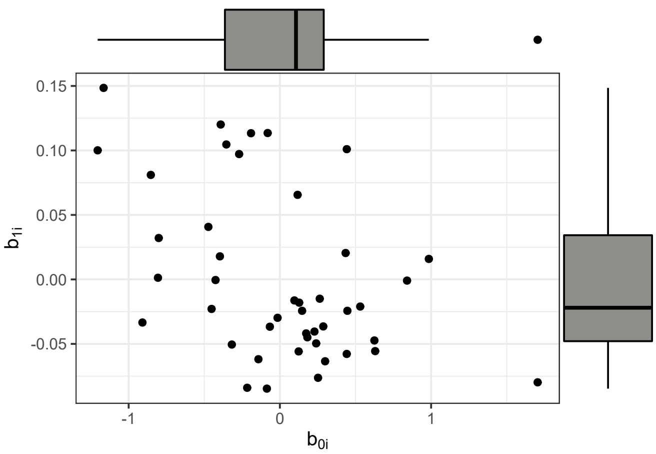


Figure 20 – **A5055 data**. Scatter plot of estimated random effects for LMEC together with a summary boxplot of the marginal densities.

In this section, we revisit the A5055 data with the aim providing additional inferences for the use of SN-LMEC. The model considered for modeling the A5055 data is given by

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 \sqrt{t_{ij}} + \beta_3 \mathrm{CD4}^+_{ij} + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij}, \tag{4.15}$$

where y_{ij} is $\log_{10}(\mathrm{RNA})$ for subject i measured roughly at day_{ij}, $t_{ij} = \mathrm{day}_{ij}/7$ (week), $\sqrt{t_{ij}}$ is the square root at time t_{ij}, $\mathrm{CD4}^+_{ij}$ indicates the standardized version of CD4 cell count for subject i at time t_{ij}, and b_{0i}, b_{1i} are the random intercept and random slope, respectively for the ith subject. The ML estimates were obtained using the EM algorithm describes in Subsection 4.3.3.

The values of loglikelihood, AIC and BIC for the four considered models are presented in Table 10. It also presents the ML estimates of the parameters of interest under the different correlation structure. It can be noted that for each criterion a structure was selected as the best, that is, for the AIC the DEC structure was selected and for BIC the AR(1). Note that the parameter estimates for both models are close. It can be observed that the estimated values of ϕ_1 and ϕ_2 under the DEC model are 0.9 and 1.3917 respectively, close to the estimated values for AR(1), $\hat{\phi}_1 = 0.8539$ and $\phi_2 = 1$. We can also notice that in the AR (1) model the sign of the estimated values for λ are in accordance with the asymmetry of the random effects estimated in Figure 20. Based on these observations and the criteria, the most parsimonious model is obtained using the continuous-time autoregressive of order 1 correlation (AR(1)). Note that, for the AR(1) model, the estimate of β_1 reveal that RNA viral loads change over time. In other words, the mean viral load ($\mathbb{E}[y_{ij}]$) at time zero with 300 CD4 cells count is 3.7339 $\log_{10}$ RNA, after 66 days it is 2.0677 $\log_{10}$ RNA, keeping CD4 cells count fixed. From the negative estimate of β_3 indicates that per unit increase in CD4 cells count may a decrease of $\log_{10}$ RNA by an average of 0.4925 in infected patients. Figure 21 (left panel) shows some individual profiles (in log10 scale) for HIV viral load and estimated trajectories for the SN-LMEC model under AR(1) structure.

Table 10 – **A5055 data**. Parameter estimates of the SN-LMEC model for A5055 data under different correlation structures. The SE values are estimated as mentioned in Section 4.3.4.

	AR(1)		CS		DEC		UNC	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_0	3.5718	0.2567	3.6327	0.2708	3.5563	0.2290	3.6358	0.2505
β_1	0.1226	0.0278	0.1379	0.0220	0.1190	0.0277	0.1393	0.0209
β_2	-0.9192	0.1295	-0.9955	0.1162	-0.9067	0.1267	-0.9949	0.1143
β_3	-0.4925	0.1611	-0.4875	0.1348	-0.4731	0.1461	-0.4773	0.1292
σ^2	0.7754		0.5299		0.7587		0.4283	
α_{11}	0.027		0.3962		0.0470		0.6803	
α_{12}	-0.006		-0.0130		-0.0013		-0.0082	
α_{22}	0.008		0.0080		0.0079		0.0067	
ϕ_1	0.8539		0.1887		0.9			
ϕ_2	1		0		1.3917			
λ_1	-0.8068		2.2924		0.7548		3.0991	
λ_2	5.0123		3.8218		4.4416		2.6436	
loglik	-303.3264		-324.7245		-301.3819		-324.6679	
AIC	628.6529		671.449		**626.7638**		669.3359	
BIC	**669.9661**		712.7622		671.8327		706.8933	

We are also interested in investigating the performance of the prediction for future values described in Subsection 4.3.6. We exclude the last two measurements of each individual in the datasets with more than 6 (inclusive) observations (total 39 individuals) and we compute the predicted values under SN-LMEC model under AR(1) correlation structure. Figure 21 (right panel) shows the comparison between the estimated, the predicted values and the real ones, indicating the good performance of the SN-LMEC in term of prediction.

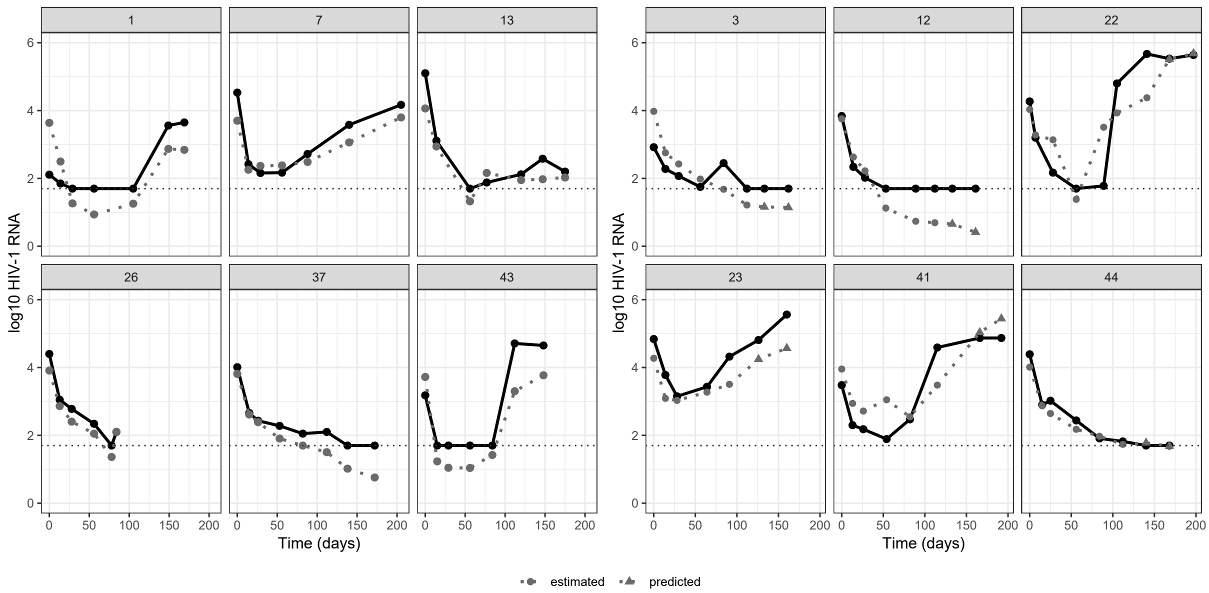
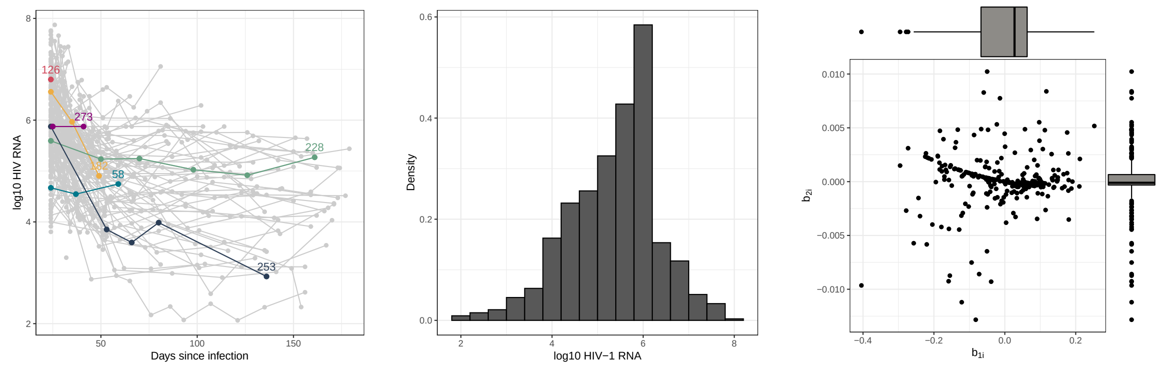


Figure 21 – **A5055 data**. (left panel) Viral loads in $\log_{10}$ scale (black, solid line) for 6 random subjects and estimated trajectories for the SN-LMEC model under AR(1) structure. (right panel) Evaluation of the prediction performance for 6 random subjects, considering the SN-LMEC model under AR(1) structure.

4.6.2 AIEDRP data

The AIEDRP data set consists of longitudinal HIV RNA measurements taken on 320 subjects from the Acute Infection and Early Disease Research Program (AIEDRP), a large multicenter observational established to develop and evaluate data from studies of patients with acute or recent HIV infection. In contrast with A5055 data, some observations here are right-censored, since during the acute stage of infection the large HIV RNA observations may lay above the limit of quantification of the assay. The subjects had between 1 and 14 observations: 129 had one, 82 had two, and 109 had three or more observations. Of the 830 recorded observations, 185 (22%) were above the limit of quantification of the assay (see Vaida & Liu (2009), for more details). The individual profiles are shown in Figure 22a. Figure 22c shows a scatter plot of the estimates of random effects obtained by fitting a NLMEC model given in (4.16) using the `nlmmcl()` function provided by Vaida & Liu (2009). A visual inspection of this figure reveals that there is considerable asymmetry among the estimated random effects. The sample skewness for the estimated random effects b_{1i} and b_{2i} are -0.6054 and -0.9413, respectively, revealing that they are moderately to highly asymmetric. Therefore, this reflects the appropriateness of using a bivariate SN distribution for random effects.



(a) (b) (c)

Figure 22 – **AIEDRP data**. (a) Individual profiles (in $\log_{10}$ scale) for HIV viral load at different follow-up times. (b) Histogram for HIV viral load (in $\log_{10}$ scale). (c) Scatter plot of estimated random effects for NLMEC together with a summary boxplot of the marginal densities.

This dataset was also analyzed by Vaida & Liu (2009) and Matos *et al.* (2013b) using the N-NLMEC and t-NLMEC models, respectively. Therefore, in our analysis we consider a right-censored five-parameter NLME model (inverted S-shaped curve) as Vaida & Liu (2009) and Matos *et al.* (2013b):

$$y_{ij} = \alpha_{1i} + \frac{\alpha_2}{1 + \exp\left((t_{ij} - \alpha_3)/\alpha_4\right)} + \alpha_{5i}(t_{ij} - 50) + \epsilon_{ij}, \tag{4.16}$$

where y_{ij} is the $\log_{10}$HIV RNA for subject i at time t_{ij}. The parameters α_{1i} and α_2 represent the subject-specific set-point values and the decrease from the maximum HIV RNA. The location parameter α_3 indicates the time point at which half of the change in HIV-1 RNA is attained, α_4 is a scale parameter modeling the rate of decline and α_{5i} allows increasing the HIV-1 RNA trajectory after day 50. To force the parameters to be positive we reparameterized the model to $\beta_{1i} = \log(\alpha_{1i}) = \beta_1 + b_{1i}$, $\beta_k = \log(\alpha_k)$, $k = 2, 3, 4$, and $\alpha_{5i} = \beta_5 + b_{2i}$. Also, $(b_{1i}, b_{2i}) \overset{\text{iid.}}{\sim} \mathrm{SN}_2(c\boldsymbol{\Delta}, \mathbf{D}, \boldsymbol{\lambda})$ are the random effects for the ith subject. The ML estimates were obtained using the EM algorithm described in Section 4.4.

As in Subsection 4.6.1, the correlation structures UNC, DEC, AR(1) and CS are considered. Table 11 summarizes the values of loglikelihood, AIC and BIC for all considered models. It can be noted that the values of loglikelihood for the AR(1) and DEC models are close. This is explained because the estimated values of ϕ_1 and ϕ_2 under the DEC model are 0.8214 and 1.2348 respectively, close to the estimated values for AR(1), $\hat{\phi}_1 = 0.7824$ and $\phi_2 = 1$. Based on this observation and the criteria, the most parsimonious model is obtained using the continuous-time autoregressive of order 1 correlation (AR(1)).

The ML estimates of $\boldsymbol{\theta}$ and the corresponding standard errors are presented in Table 11. We 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 4.4794 $\log_{10}$ units. Figure 23 (left panel) shows some individual profiles (in $\log 10$ scale) for HIV viral load at different follow-up times. Table 12 shows the results obtained when adjust a nonlinear mixed effects (NLMEC) model for normal distribution to the data (Matos *et al.*, 2016). As expected, the criteria for the N-NLMEC model are higher than those for the SN-NLMEC, since the estimated values of the skewness parameter $\boldsymbol{\lambda}$ are large. Thus, the SN-NLMEC model seems to be more appropriate than the normal counterpart for this dataset.

Prediction performance is an important measure of model adequacy. To check the prediction performance of the HIV viral load, we considered the following approach: we exclude the last two measurements of each individual in the datasets with more than 6 (inclusive) observations (total of 36 individuals), refit the model based on the remaining data, obtain the new estimates, and we compute the predicted values under SN-NLMEC model under AR(1) correlation structure. Figure 23 (right panel) shows the comparison between the estimated, the predicted values and the real ones. Once again, this figure indicates a good performance of the SN-LMEC model in terms of prediction.

Table 11 – **AIEDRP data**. Parameter estimates of the SN-LMEC model for AIEDRP data. The SE values are estimated as mentioned in Section 4.3.4.

	AR(1)		CS		DEC		UNC	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_1	1.5979	0.0146	1.5858	0.0164	1.5969	0.0144	1.5929	0.0132
β_2	-0.1190	0.1424	0.0336	0.1348	-0.1137	0.1422	-0.0798	0.1219
β_3	3.5432	0.0316	3.5734	0.0291	3.5477	0.0321	3.5516	0.0262
β_4	1.0756	0.3531	1.2661	0.3099	1.0859	0.3620	0.9997	0.3346
β_5	-0.0036	0.0020	-0.0028	0.0025	-0.0035	0.0020	-0.0034	0.0020
σ^2	0.284		0.4555		0.2785		0.2553	
α_{11}	0.0367		0.0147		0.0370		0.03625	
α_{12}	0.0006		0.0006		0.0006		0.00071	
α_{22}	0.00003		0.0001		0.00004		0.00005	
ϕ_1	0.7824		0.4535		0.8214			
ϕ_2	1		0		1.2348			
λ_1	-3.7492		-3.0373		-3.8078		-4.1573	
λ_2	-1.3682		-2.5609		-1.4096		-1.8388	
loglik	-668.4097		-680.2904		-667.7367		-681.7877	
AIC	**1360.82**		1384.581		1361.474		1385.575	
BIC	**1417.477**		1441.238		1422.852		1437.511	

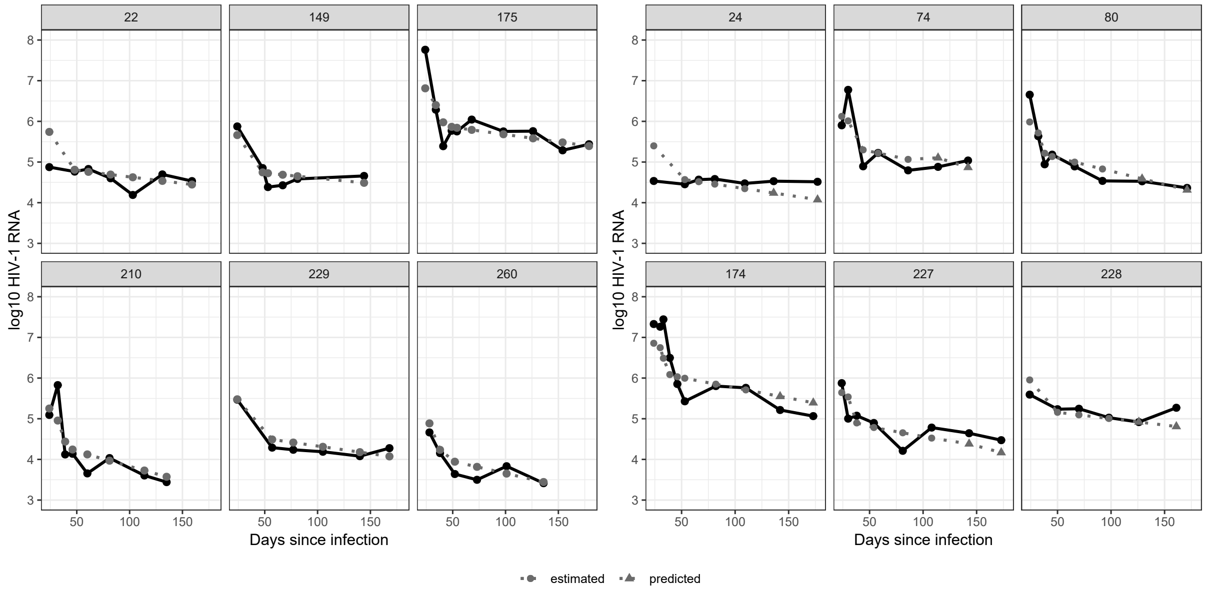


Figure 23 – **AIEDRP data**. (left panel) Viral loads in $\log_{10}$ scale (black, solid line) for 6 random subjects and estimated trajectories (gray, dotted line) for the SN-NLMEC model under AR(1) structure. (right panel) Evaluation of the prediction performance for 6 random subjects, considering the SN-NLMEC model under AR(1) structure.

Table 12 – **AIEDRP data.** Model selection criterion for the NLMEC model under different correlation structures (Matos *et al.*, 2016).

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

4.7 Conclusions

In this chapter we have proposed an approach to a linear and nonlinear mixed model with censored responses where the random effects are assumed to have a multivariate skew-normal distribution. We adopted a DEC structure as proposed by Muñoz *et al.* (1992) to model the autocorrelation existing among irregularly observed measures. The proposed model generalizes previous proposals, such as, the SN-LME model proposed by Arellano-Valle *et al.* (2005) (see also, Lin & Lee, 2008) and in the context of censored data, the N-LMEC/NLMEC model proposed by Vaida & Liu (2009) (see also, Matos *et al.*, 2016), which are restricted to a left or right censored problem. We developed a computationally tractable EM algorithm for carrying out ML estimation. The algorithm has a closed-form expression for the E-step, based on formulas for the mean and variance of the truncated extended multivariate skew-normal distribution (Galarza *et al.*, 2019). The computation procedures for the estimation of random effects and the prediction of future responses are easy to implement once the ML estimates are obtained. Several simulation studies were performed, indicating that under the skew-normal distribution assumption, there is a gain efficiency and accuracy in estimating certain parameters when the normality assumption does not hold. Furthermore, the proposed methods were applied on two AIDS studies, providing support for the usefulness and effectiveness of our proposal. The `R` codes are available upon request.

Although the SN-LMEC/NLMEC models showed flexibility to model asymmetric data, they can be seriously affected by the presence of outliers. A natural generalization of our method is to extend by considering the skew-t distribution (Azzalini & Capitanio, 2003) or the multivariate skew-elliptical distribution (Branco & Dey, 2001).

Chapter 5

Concluding remarks

In this thesis, from a classical perspective, we discuss some approaches for flexible modeling of censored longitudinal data, motivated by data sets from AIDS clinical trials. This work is a generalization of the works presented by Lachos *et al.* (2010), Matos *et al.* (2013b), Matos *et al.* (2016), Ibacache-Pulgar *et al.* (2012).

In Chapter 2, we proposed a semiparametric mixed model to analyze censored longitudinal data, called SMEC. This model takes into account the autocorrelation existing among irregularly observed measures and it is possible to model the effects of the covariates that contribute in a parametric and nonparametric way on the response variable. A robust alternative for modeling censored longitudinal data with tails heavier than normal is presented in Chapter 3, called t-SMEC. Since, skewness of the HIV viral load is still noticeable even after transformation, it is important to use a distribution that can allow us to relax the normal assumption. In Chapter 4, we developed a linear and nonlinear mixed model with censored responses where the random effects are assumed to follow a multivariate skew-normal distribution.

The EM algorithm (Dempster *et al.*, 1977) was developed to obtain the maximum likelihood estimates for the parameters of the models. This methodology was applied and tested on four clinical trials data, as well as on simulated data in order to show how our procedures can be used to evaluate censored models and obtain robust estimates for the parameters.

5.1 Future research

Several research works can be derived and/or directed from the results of this work.

The first work perspective is related to the development of a method of local influence to detect influential observations and evaluate the sensitivity of the estimates in the models. In the semiparametric context, Ibacache-Pulgar *et al.* (2012) proposed influence diagnostics for elliptical semiparametric mixed models. Following Ibacache-Pulgar *et al.* (2012), we are developing an local influence analysis for the model described in Chapter 2, see Appendix B. Following Montenegro *et al.* (2009), we also extend the local influence analysis to the model proposed in Chapter 4.

The second perspective for future research is to extend the models in Chapters 2 and 3 (SMEC and t-SMEC) to semiparametric additive mixed models. In HIV/AIDS clinical trials, it is often interesting to investigate differences between treatments for decreasing a patient's viral load. For example, Hammer *et al.* (2002) assess whether adding a second protease inhibitor (PI) improves antiviral efficacy of a 4-drug combination in patients with virologic failure while taking a PI-containing regimen. With an additive mixed model, we can consider a nonparametric function for each treatment and evaluate the differences between them.

Finally, a third research perspective is to consider distributions with asymmetry and heavy tails for SMEC models, such as the multivariate skew-normal distribution, the skew-t distribution, and the multivariate skew-elliptical distribution, to accommodate the censoring, skewness and heaviness in the tails of the distribution, simultaneously. Recently, Castro *et al.* (2019) proposed a Bayesian semiparametric approach considering skew-normal distribution for modeling the random effects. Therefore, we can propose a frequentist approach to these models.

In summary, since censored modeling is a promising area many issues still remain for filling the gap to modeling censored data. We plan to investigate these issues in our future research.

Bibliography

Acosta, E. P., Wu, H., Hammer, S. M., Yu, S., Kuritzkes, D. R., Walawander, A., Eron, J. J., Fichtenbaum, C. J., Pettinelli, C., Neath, D. *et al.* (2004). Comparison of two indinavir/ritonavir regimens in the treatment of HIV-infected individuals. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, **37**(3), 1358–1366. pages 24

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Cont.*, **19**, 716–723. pages 39, 64

Arellano-Valle, R. B. & Bolfarine, H. (1995). On some characterizations of the t-distribution. *Statistics & Probability Letters*, **25**(1), 79–85. pages 52, 55

Arellano-Valle, R. B., Bolfarine, H. & Lachos, V. (2005). Skew-normal linear mixed models. *Journal of Data Science*, **3**, 415–438. pages 98

Arellano-Valle, R. B., Branco, M. D. & Genton, M. G. (2006). A unified view on skewed distributions arising from selections. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, **34**(4), 581–601. pages 75

Arribas-Gil, A., De la Cruz, R., Lebarbier, E. & Meza, C. (2015). Classification of longitudinal data through a semiparametric mixed-effects model based on Lasso-type estimators. *Biometrics*, **71**(2), 333–343. pages 19

Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, **32**(2), 159–188. pages 75

Azzalini, A. & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**(2), 367–389. pages 99

Azzalini, A. & Genton, M. (2008). Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review*, **76**, 1490–1507. pages 49

Azzalini, A. & Valle, A. D. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**(4), 715–726. pages 49, 74

Bandyopadhyay, D., Lachos, V. H., Castro, L. M. & Dey, D. K. (2012). Skew-normal/independent linear mixed models for censored responses with applications to HIV viral loads. *Biometrical Journal*, **54**(3), 405–425. pages 20, 74

Bandyopadhyay, D., Castro, L. M., Lachos, V. H. & Pinheiro, H. P. (2015). Robust Joint Non-linear Mixed-Effects Models and Diagnostics for Censored HIV Viral Loads with CD4 Measurement Error. *Journal of Agricultural, Biological, and Environmental Statistics*, **20**(1), 121–139. pages 20

Branco, M. D. & Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, **79**, 99–113. pages 20, 74, 99

Castro, L. M., Wang, W.-L., Lachos, V. H., Inácio de Carvalho, V. & Bayes, C. L. (2019). Bayesian semiparametric modeling for HIV longitudinal data with censoring and skewness. *Statistical Methods in Medical Research*, **28**(5), 1457–1476. pages 20, 101

Cook, R. (1977). Detection of influential observation in linear regression. *Technometrics*, pages 15–18. pages 115

Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society, Series B,*, **48**, 133–169. pages 115, 118

Cook, R. D. & Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman & Hall/CRC, Boca Raton, FL. pages 116

Davidian, M. & Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data*. Routledge. pages 27, 50, 64, 73

Dempster, A., Laird, N. & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B,*, **39**, 1–38. pages 21, 57, 80, 100

Diggle, P. (2002). *Analysis of longitudinal data*. Oxford University Press, USA. ISBN 0198524846. pages 27, 50

Galarza, C., Matos, L., Dey, D. & Lachos, V. (2019). On moments of folded and truncated multivariate extended skew-normal distributions. Technical Report ID 19-14, University of Connecticut. pages 76, 77, 83, 98

Galarza, C. E., Kan, R. & Lachos, V. H. (2020). *MomTrunc: Moments of Folded and Doubly Truncated Multivariate Distributions*. R package version 5.69. pages 34, 60, 62, 83

Galea-Rojas, M. (1995). *Calibraçao comparativa estrutural e funcional*. Ph.D. thesis, Tese de Doutorado em Estatıstica, Universidade de Sao Paulo, Sao Paulo. pages 64

Green, P. J. (1987). Penalized likelihood for general semi-parametric regression models. *International Statistical Review/Revue Internationale de Statistique*, **55**, 245–259. pages 32, 34, 37, 48, 56, 62, 72

Green, P. J. (1990). On use of the em for penalized likelihood estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, **52**(3), 443–452. pages 57

Green, P. J. & Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: a roughness penalty approach*. CRC Press. pages 28, 34, 38, 56, 62

Hammer, S. M., Vaida, F., Bennett, K. K., Holohan, M. K., Sheiner, L., Eron, J. J., Wheat, L. J., Mitsuyasu, R. T., Gulick, R. M., Valentine, F. T. *et al.* (2002). Dual vs single protease inhibitor therapy following antiretroviral treatment failure: a randomized trial. *JAMA*, **288**(2), 169–180. pages 101

Helsel, D. R. (2011). *Statistics for censored environmental data using Minitab and R*, volume 77. John Wiley & Sons. pages 51

HO, H. & LIN, T. (2010). Robust linear mixed models using the skew t distribution with application to schizophrenia data. *Biometrical journal*, **52**(4), 449–469. pages 74

Hughes, J. (1999). Mixed effects models with censored data with application to HIV RNA levels. *Biometrics*, **55**, 625–629. pages 19, 27, 36, 51, 74

Ibacache-Pulgar, G., Paula, G. A. & Galea, M. (2012). Influence diagnostics for elliptical semiparametric mixed models. *Statistical Modelling*, **12**(2), 165–193. pages 20, 51, 100, 101

Ibacache-Pulgar, G., Paula, G. A. & Cysneiros, F. J. A. (2013). Semiparametric additive models under symmetric distributions. *TEST*, **22**(1), 103–121. pages 31, 34, 56, 72

Jacqmin-Gadda, H., Thiebaut, R., Chene, G. & Commenges, D. (2000). Analysis of left-censored longitudinal data with application to viral load in HIV infection. *Biostatistics*, **1**, 355–368. pages 19, 74

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994). *Continuous univariate distributions*. Wiley New York. pages 64

Kan, R. & Robotti, C. (2017). On moments of folded and truncated multivariate normal distributions. *Journal of Computational and Graphical Statistics*, **26**(4), 930–934. pages 34

Kohn, R., Ansley, C. F. & Tharm, D. (1991). The performance of cross-validation and maximum likelihood estimators of spline smoothing parameters. *Journal of the American Statistical Association*, **86**(416), 1042–1050. pages 37, 62

Kotzin, B. L., Kuritzkes, D. R., Connick, E., Lederman, M. M., Sevin, A. D., Spritzler, J., Rousseau, F., St. Clair, M., Martinez, A., Roe, J. D., Fox, L., Leonard, J. M., Chiozzi, M. H., Landay, A. & Kessler, H. (2000). Immune Reconstitution in the First Year of Potent Antiretroviral Therapy and Its Relationship to Virologic Response. *The Journal of Infectious Diseases*, **181**(1), 358–363. pages 23

Lachos, V., Bandyopadhyay, D. & Dey, D. (2011). Linear and nonlinear mixed-effects models for censored HIV viral loads using normal/independent distributions. *Biometrics*, **55**, 1304–1318. pages 20, 51

Lachos, V. H., Ghosh, P. & Arellano-Valle, R. B. (2010). Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica*, **20**(1), 303–322. pages 20, 49, 73, 78, 81, 100

Lachos, V. H., A. Matos, L., Castro, L. M. & Chen, M.-H. (2019). Flexible longitudinal linear mixed models for multiple censored responses data. *Statistics in Medicine*, **38**(6), 1074–1102. pages 28, 93

Laird, N. M. & Ware, J. (1982). Random effects models for longitudinal data. *Biometrics*, **38**, 963–974. pages 18, 73

Landay, A., Martinez, A., Kotzin, B. L., Kuritzkes, D. R., Connick, E., Rousseau, F., Kessler, H., Roe, J. D., Spritzler, J., Leonard, J. M., Fox, L., Heath Chiozzi, M., St. Clair, M., Wade, M. & Lederman, M. M. (1998). Immunologic Responses Associated with 12 Weeks of Combination Antiretroviral Therapy Consisting of Zidovudine, Lamivudine, and Ritonavir: Results of AIDS Clinical Trials Group Protocol 315. *The Journal of Infectious Diseases*, **178**(1), 70–79. pages 23

Lange, K. L. & Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *J. Comput. Graph. Stat*, **2**, 175–198. pages 20

Lange, K. L., Little, R. & Taylor, J. (1989). Robust statistical modeling using t distribution. *Journal of the American Statistical Association*, **84**, 881–896. pages 60, 64

Lee, S. Y. & Xu, L. (2004). R influence analysis of nonlinear mixed-effects models. *Computational Statistics & Data Analysis*, **45**, 321–341. pages 119

Lin, T. I. & Lee, J. C. (2008). Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. *Statistics in Medicine*, **27**(9), 1490–1507. pages 98

Liu, C. & Rubin, D. B. (1994). The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika*, **80**, 267–278. pages 22, 57

Louis, T. A. (1982). Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, **44**(2), 226–233. pages 36, 60, 84

Matos, L. A., Lachos, V. H., Balakrishnan, N. & Labra, F. V. (2013a). Influence diagnostics in linear and nonlinear mixed-effects models with censored data. *Computational Statistics & Data Analysis*, **57**(1), 450 – 464. pages 27, 32, 34, 38, 56

Matos, L. A., Prates, M. O., Chen, M.-H. & Lachos, V. H. (2013b). Likelihood-based inference for mixed-effects models with censored response using the multivariate-t distribution. *Statistica Sinica*, **23**, 1323–1345. pages 20, 28, 51, 53, 69, 72, 74, 96, 100

Matos, L. A., Castro, L. M. & Lachos, V. H. (2016). Censored mixed-effects models for irregularly observed repeated measures with applications to HIV viral loads. *TEST*, **25**(4), 627–653. pages 13, 31, 48, 74, 97, 98, 100, 112

Meng, X. & Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, **81**, 633–648. pages 22, 32

Meza, C., Osorio, F. & De la Cruz, R. (2012). Estimation in nonlinear mixed-effects models using heavy-tailed distributions. *Statistics and Computing*, **22**(1), 121–139. pages 51

Montenegro, L. C., Lachos, V. H. & Bolfarine, H. (2009). Local influence analysis for skew-normal linear mixed models. *Communications in Statistics—Theory and Methods*, **38**(4), 484–496. pages 101

Moulton, L. & Halsey, N. (1995). A mixture model with detection limits for regression analyses of antibody response to vaccine. *Biometrics*, **51**(4), 1570–1578. pages 19

Muñoz, A., Carey, V., Schouten, J. P., Segal, M. & Rosner, B. (1992). A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics*, **48**(3), 733–742. pages 22, 29, 30, 54, 78, 98

Palarea-Albaladejo, J. & Martin-Fernandez, J. (2013). Values below detection limit in compositional chemical data. *Analytica Chimica Acta*, **764**, 32–43. pages 51

Pinheiro, J. & Bates, D. (2006). *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media. pages 27, 50

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & R Core Team (2020). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-144. pages 35, 59

Pinheiro, J. C., Liu, C. & Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics*, **10**(2), 249–276. pages 19, 27, 50

Poon, W. & Poon, Y. (1999). Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **61**, 51–61. pages 118, 119

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. pages 39, 49, 65, 72, 88

Rao, C. R. *et al.* (1987). Prediction of future observations in growth curve models. *Statistical Science*, **2**(4), 434–447. pages 86

Reiss, P. T. & Ogden, R. T. (2009). Smoothing parameter selection for a class of semiparametric linear models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **71**(2), 505–523. pages 37

Rice, J. A. & Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, **57**(1), 253–259. pages 28

Ruppert, D., Wand, M. P. & Carroll, R. J. (2003). *Semiparametric regression*. Number 12. Cambridge university press. pages 19

Saitoh, A., Foca, M., Viani, R., Heffernan-Vacca, S., Vaida, F., Lujan-Zilbermann, J., Emmanuel, P., Deville, J. & Spector, S. (2008). Clinical outcomes after an unstructured treatment interruption in children and adolescents with perinatally acquired HIV infection. *Pediatrics*, **121**, 513–521. pages 68

Schwarz, G. *et al.* (1978). Estimating the dimension of a model. *The annals of Statistics*, **6**(2), 461–464. pages 64

Segal, M. R., Bacchetti, P. & Jewell, N. P. (1994). Variances for maximum penalized likelihood estimates obtained via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, **56**(2), 345–352. pages 36, 60

Singh, A. & Nocerino, J. (2002). Robust estimation of mean and variance using environmental data sets with below detection limit observations. *Chemometrics and Intelligent Laboratory Systems*, **60**(1), 69 – 86. Fourth International Conference on Environ metrics and Chemometrics held in Las Vegas, NV, USA, 18-20 September 2000. pages 19

Speed, T. (1991). [that blup is a good thing: The estimation of random effects]: Comment. *Statistical Science*, **6**(1), 42–44. pages 37

Stevens, R. A., Lempicki, R. A., Natarajan, V., Higgins, J., Adelsberger, J. W. & Metcalf, J. A. (2006). General immunologic evaluation of patients with human immunodeficiency virus infection. In *Manual of Molecular and Clinical Laboratory Immunology, 7th Edition*, pages 847–861. American Society of Microbiology. pages 24

Szczesniak, R., Li, D. & Raouf, S. (2015). Semiparametric mixed models for medical monitoring data: An overview. *Journal of Biometrics & Biostatistics*, **6**(2). pages 19

Tibshirani, R. J. (1990). *Generalized additive models*. London: Chapman and Hall. pages 38

Vaida, F. & Liu, L. (2009). Fast implementation for normal mixed effects models with censored response. *Journal of Computational and Graphical Statistics*, **18**, 797–817. pages 19, 27, 31, 32, 34, 36, 48, 69, 74, 78, 79, 87, 90, 95, 96, 98

Vaida, F. & Liu, L. (2012). *lmec: Linear Mixed-Effects Models with Censored Responses*. R package version 1.0. pages 68, 93

Verbeke, G. & Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, **91**(433), 217–221. pages 73

Vock, D. M., Davidian, M., Tsiatis, A. A. & Muir, A. J. (2011). Mixed model analysis of censored longitudinal data with flexible random-effects density. *Biostatistics*, **13**(1), 61–73. pages 28

Wahba, G. (1985). A comparison of gcv and gml for choosing the smoothing parameter in the generalized spline smoothing problem. *The Annals of Statistics*, **13**(4), 1378–1402. pages 37

Wahba, G. & Wold, S. (1975). A completely automatic french curve: fitting spline functions by cross validation. *Communications in Statistics*, **4**(1), 1–17. pages 62

Wang, W.-L. (2013). Multivariate t linear mixed models for irregularly observed multiple repeated measures with missing outcomes. *Biometrical Journal*, **55**(4), 554–571. pages 86, 87

Wang, W.-L., Lin, T.-I. & Lachos, V. H. (2018). Extending multivariate-t linear mixed models for multiple longitudinal data with censored responses and heavy tails. *Statistical Methods in Medical Research*, **27**(1), 48–64. pages 93

Wang, Y. (1998a). Mixed effects smoothing spline analysis of variance. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **60**(1), 159–174. pages 28, 31, 37

Wang, Y. (1998b). Smoothing spline models with correlated random errors. *Journal of the American Statistical Association*, **93**(441), 341–348. pages 62

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, **99**(467), 673–686. pages 28

Wu, H. (2005). Statistical methods for HIV dynamic studies in AIDS clinical trials. *Statistical Methods in Medical Research*, **14**, 171–192. pages 18

Wu, H. & Ding, A. (1999). Population HIV-1 dynamics in vivo: applicable models and inferential tools for virological data from AIDS clinical trials. *Biometrics*, **55**(2), 410–418. pages 74

Wu, H., Ding, A. & De Gruttola, V. (1998). Estimation of HIV dynamic parameters. *Statistics in Medicine*, **17**(21), 2463–2485. pages 18

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. pages 46

Wu, L. (2010). *Mixed Effects Models for Complex Data*. Chapman & Hall/CRC, Boca Raton, FL. pages 27

Zeger, S. L. & Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics*, **50**, 689–699. pages 19, 28, 50, 62

Zhang, D., Lin, X., Raz, J. & Sowers, M. (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association*, **93**(442), 710–719. pages 28, 37, 50, 62

Zhu, H. & Lee, S. (2001). Local influence for incomplete-data models. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **63**, 111–126. pages 115, 117, 118, 119

Zhu, H., Lee, S., Wei, B. & Zhou, J. (2001). Case-deletion measures for models with incomplete data. *Biometrika*, **88**, 727–737. pages 115, 116, 117

APPENDIX A

Chapter 2: Additional simulation study results

A.1 Simulation Study 3

This simulation verifies the behavior of the proposed model for different sizes of the time dimensions.

For this simulation study, we considered a DEC-SMEC model as in (2.16). The left censoring proportion was fixed at 15% and sample size at $n = 200$. We generated 500 samples of the DEC-SMEC model considering an AR(1) structure with parameter $\phi_1 = 0.6$. For this study, we chose the function $f(t_{ij}) = \cos(\pi\sqrt{t_{ij}})$ with t_{ij} in three scenarios:

- Scenario 1: $t_{ij} = (2, 4, 6, 8, 10, 12)$;
- Scenario 2: $t_{ij} = (2, 3, 6, 9, 10, 12)$;
- Scenario 3: $t_{ij} = (1, 5, 9, 13, 17, 21, 25, 29)$.

To evaluate each scenario we computed the measures described in the Simulation Study 2 (Section 2.5.2). Table 13 summarize some results of the parameter estimates for each scenario, respectively, and in Figure 24 we show the 500 estimated curves with the average estimates curves for each scenario. From the Table and Figure, we notice that when the spacing between the times is large, the non-parametric function is overestimated. In addition, we can note that although the estimates of the non-parametric function are not close to the true function in Scenario 3 (Table 13), the estimates of the parametric components provided good estimates. However, when the dimension of the number of times is small (Scenarios 1 and 2), the components of variance are not well estimated.

Table 13 – **Simulation study 3**. Summary statistics based on 500 simulated AR(1) samples for Scenario 1, 2 and 3.

Scenario 1					
Parameter	MC M	MAE	MC SE	MC SD	CP
$\beta_1(2)$	1.9980	0.0615	0.0778	0.0778	94.8%
$\beta_2(-1.5)$	-1.4972	0.0213	0.0259	0.0265	94.8%
$f_1(2) = -0.2663$	-0.2732	0.0851	0.1225	0.1063	98.2%
$f_2(4) = 1$	0.9878	0.1233	0.1710	0.1549	97.2%
$f_3(6) = 0.1580$	0.1560	0.1665	0.2224	0.2075	97%
$f_4(8) = -0.8582$	-0.8329	0.2139	0.2752	0.2677	95.4%
$f_5(10) = -0.8728$	-0.8082	0.2599	0.3289	0.3226	95.6%
$f_6(12) = -0.1125$	-0.0008	0.3104	0.3826	0.3770	94.6%
σ^2 (0.55)	0.6789	0.1289			
α_{11} (0.25)	0.4269	0.1809			
α_{12} (0.1)	0.2487	0.1487			
α_{22} (0.2)	0.1492	0.0509			
ϕ_1 (0.6)	0.6803	0.0805			
Scenario 2					
Parameter	MC M	MAE	MC SE	MC SD	CP
$\beta_1(2)$	1.9989	0.0556	0.0702	0.0699	95.8%
$\beta_2(-1.5)$	-1.4986	0.0191	0.0237	0.0241	94.8%
$f_1(2) = -0.2663$	-0.2679	0.0835	0.1142	0.1045	96.4%
$f_2(3) = 0.6661$	0.6593	0.1016	0.1393	0.1287	96.6%
$f_3(6) = 0.1580$	0.1561	0.1667	0.2201	0.2075	96.8%
$f_4(9) = -1$	-0.9763	0.2381	0.3052	0.2976	96%
$f_5(10) = -0.8728$	-0.8383	0.2618	0.3339	0.3264	96.4%
$f_6(12) = -0.1125$	-0.0519	0.3126	0.3916	0.3870	95.4%
σ^2 (0.55)	0.6205	0.0917			
α_{11} (0.25)	0.3308	0.1282			
α_{12} (0.1)	0.1829	0.0921			
α_{22} (0.2)	0.1711	0.0378			
ϕ_1 (0.6)	0.6419	0.0625			
Scenario 3					
Parameter	MC M	MAE	MC SE	MC SD	CP
$\beta_1(2)$	1.9930	0.0617	0.0744	0.0763	93.8%
$\beta_2(-1.5)$	-1.5013	0.0209	0.0251	0.0260	93.8%
$f_1(1) = -1$	-0.9065	0.1077	0.0861	0.0881	82%
$f_2(5) = 0.7374$	1.0505	0.3210	0.1895	0.1900	62.6%
$f_3(9) = -1$	-0.4544	0.5567	0.3109	0.3047	59.6%
$f_4(13) = 0.3256$	1.0960	0.7859	0.4355	0.4290	57.4%
$f_5(17) = 0.9261$	1.9201	1.0122	0.5614	0.5446	56.8%
$f_6(21) = -0.2565$	0.9670	1.2426	0.6881	0.6676	57.4%
$f_7(25) = -1$	0.4507	1.4737	0.8148	0.7875	57%
$f_8(29) = -0.3530$	1.3255	1.7045	0.9414	0.9062	57.2%
σ^2 (0.55)	0.5495	0.0270			
α_{11} (0.25)	0.2541	0.0581			
α_{12} (0.1)	0.1044	0.0235			
α_{22} (0.2)	0.2011	0.0173			
ϕ_1 (0.6)	0.5917	0.0475			

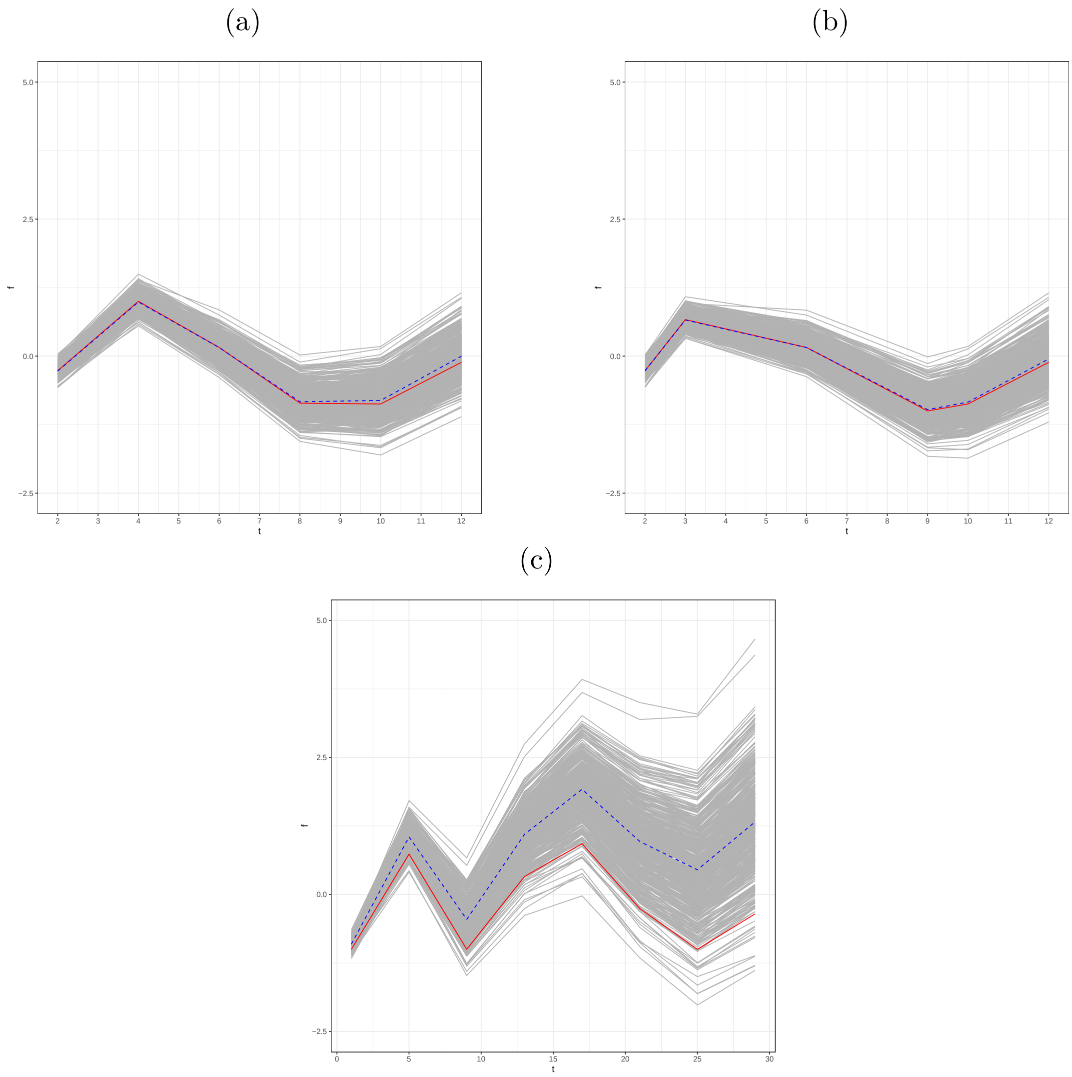


Figure 24 – **Simulation study 3**. Graphs of the nonparametric components with 500 replications. Estimated curves (gray lines), true curves (red lines) and the average estimates curves (blue lines). (a) $t_{ij} = (2, 4, 6, 8, 10, 12)$ (b) $t_{ij} = (2, 3, 6, 9, 10, 12)$ (c) $t_{ij} = (1, 5, 9, 13, 17, 21, 25, 29)$

A.2 Simulation Study 4

The purpose of this simulation study is to evaluate the benefits of the proposed model when compared to the existing literature.

For this simulation, we consider a N-NLMEC model (Matos *et al.*, 2016) as follows

$$y_{ij} = \lambda_{1i} + \frac{\lambda_{2i}}{1 + \exp((t_{ij} - \lambda_3)/\lambda_4)} + \epsilon_{ij}, \tag{A.1}$$

with $i = 1, \ldots, 100$, $j = 1, \ldots, 10$, $\lambda_{1i} = \exp(\beta_1 + b_{1i})$, $\lambda_{2i} = \exp(\beta_2 + b_{2i})$, $\beta_k = \log(\lambda_k)$, $k = 3, 4$, $(b_{1i}, b_{2i}) \overset{\text{ind.}}{\sim} \mathrm{N}_2(\mathbf{0}, \mathbf{D})$, and $\epsilon_{ij} \overset{\text{ind.}}{\sim} \mathrm{N}_{n_i}(\mathbf{0}, \sigma^2\mathbf{I})$. 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 $\alpha_{11} = 0.0025, \alpha_{12} = -0.001, \alpha_{22} = 0.01$, $t_{ij} = (0, 10, 20, 30, 40, 50, 60, 70, 80, 90)$.

We simulated 500 datasets from model (A.1) considering 10% of left-censored observation. Once the simulated datasets were generated, we fitted the proposed model and the N-NLMEC model. To compare the performance of the fitted, we considered two empirical discrepancy measures, namely the MAE (mean absolute error) and MSE (mean square error). For each dataset generated, we calculate the MAE and the MSE. These measures are given by

$$\mathrm{MAE}^{(l)} = \frac{1}{N}\sum_{i,j} |y_{ij}^{(l)} - \widehat{y_{ij}}^{(l)}| \quad \text{and} \quad \mathrm{MSE}^{(l)} = \frac{1}{N}\sum_{i,j} (y_{ij}^{(l)} - \widehat{y_{ij}}^{(l)})^2,$$

$l = 1, \ldots, 500$, and $N = \sum_{i=1}^{n} n_i$.

Table 14 – **Simulation study 4**. Average of the MAE and MSE for SMEC and N-NLMEC model.

	MAE		MSE	
Model	Mean	SD	Mean	SD
SMEC	0.5367	0.0166	0.4517	0.0274
N-NLMEC	0.5296	0.0160	0.4393	0.0239

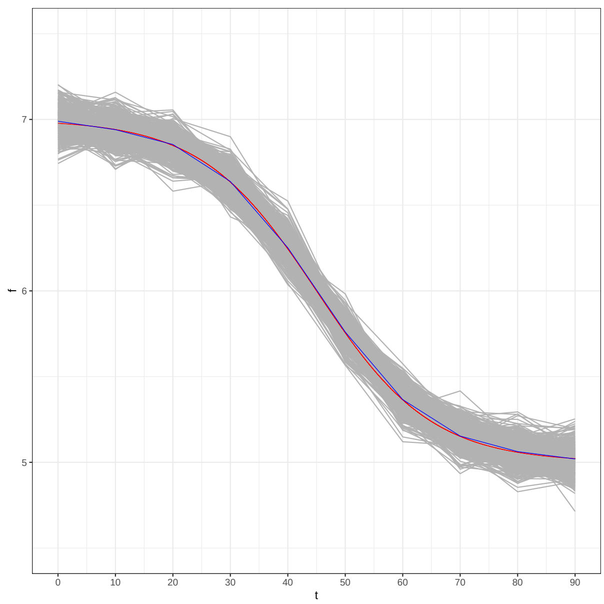


Figure 25 – **Simulation study 4**. Graphs of the nonparametric components with 500 replications. Estimated curves (gray lines), true curve (red line) and the average estimates curve (blue line).

From Table 14 we can observe that the values of MAE and MSE are close and also the fit of the SMEC model is close to the fit of the N-NLMEC model, which is the true model.

To investigate the accuracy of estimating the nonlinear function (A.1), the true shape of this function is plotted in Figure 25 with the 500 estimated curves. We can note that the shape of the average estimates of $f(t_{ij})$ is very close to the true function. We note that the nonparametric part captures well the nonlinear function.

APPENDIX B

Influence diagnostics semiparametric mixed-effects models with censored data

Influence diagnostic techniques are used to identify anomalous observations that impact on model fitting or statistical inference for the assumed statistical model. There are primarily two approaches for detecting influential observations. The case-deletion approach (Cook, 1977) is the most popular one for identifying influential observations. To assess the impact of influential observations on parameter estimates some metrics have been used for measuring the distance between $\widehat{\boldsymbol{\theta}}_{[i]}$ and $\widehat{\boldsymbol{\theta}}$, such as the likelihood distance and Cook's distance. The second approach is a general statistical technique used to assess the stability of the estimation outputs with respect to the model inputs (Cook, 1986).

Below we describe two of the main procedures to determine the influence of outlying observations. We consider diagnostic measures suitable for models with incomplete data, based on the MPL estimation using the penalized EM algorithm. First, we present the approach of case deletion using the generalized Cook distance (Zhu *et al.*, 2001). Subsequently, we develop the diagnostic using the local influence method proposed by Zhu & Lee (2001). All methods are described for the DEC-SMEC model (Chapter 2) and the notation is in accordance as well.

B.1 Case-deletion measures

Case-deletion is a common approach for studying the effects of dropping the ith case from the data set. In the following, a quantity with a subscript "$[i]$" denotes the original quantity with the ith case deleted; for example, $\mathbf{y}_{\text{com}[i]}$, denotes the complete-data with the ith case deleted. The penalized log-likelihood function of $\boldsymbol{\theta}$, based on the data with the ith case deleted, is then denoted by $\ell_{p_c}(\boldsymbol{\theta}|\mathbf{y}_{\text{com}[i]})$. Let $\widehat{\boldsymbol{\theta}}_{[i]} = (\widehat{\boldsymbol{\beta}}^{\top}_{[i]}, \widehat{\mathbf{f}}^{\top}_{[i]}, \widehat{\sigma}^2{}_{[i]}, \widehat{\boldsymbol{\alpha}}^{\top}_{[i]}, \widehat{\boldsymbol{\phi}}^{\top}_{[i]})^{\top}$

be the maximizer of the function $Q_{p_{[i]}}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \mathbb{E}\left[\ell_{p_c}(\boldsymbol{\theta}|\mathbf{y}_{\text{com}[i]})|\mathbf{V},\mathbf{C},\widehat{\boldsymbol{\theta}}\right]$, where $\widehat{\boldsymbol{\theta}}$ is the MPL estimate of $\boldsymbol{\theta}$. To assess the influence of the ith case on the MPL estimate $\widehat{\boldsymbol{\theta}}$, we compare the difference between $\widehat{\boldsymbol{\theta}}_{[i]}$ and $\widehat{\boldsymbol{\theta}}$. If the deletion of a case seriously influences the estimates, more attention should be paid to that case. Hence, if $\widehat{\boldsymbol{\theta}}_{[i]}$ is far from $\widehat{\boldsymbol{\theta}}$ in some sense, then the ith case is regarded as influential. As $\widehat{\boldsymbol{\theta}}_{[i]}$ is needed for every case, the required computational effort may become quite heavy, especially when the sample size is large. Hence, the following one-step pseudo approximation $\widehat{\boldsymbol{\theta}}^1_{[i]}$ is used to reduce the computational effort (see Cook & Weisberg, 1982; Zhu *et al.*, 2001):

$$\widehat{\boldsymbol{\theta}}^1_{[i]} = \widehat{\boldsymbol{\theta}} + \left\{-\ddot{Q}_p(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\right\}^{-1}\dot{Q}_{p_{[i]}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \tag{B.1}$$

where $\ddot{Q}_p(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \dfrac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ and $\dot{Q}_{p_{[i]}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \dfrac{\partial Q_{p_{[i]}}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ represent the Hessian matrix and the individual score vector, respectively.

Thus, $\dot{Q}_{p_{[i]}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \left(\dot{Q}^{\beta}_{p_{[i]}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \dot{Q}^{\mathbf{f}}_{p_{[i]}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \dot{Q}^{\sigma^2}_{p_{[i]}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \dot{Q}^{\alpha}_{p_{[i]}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \dot{Q}^{\phi}_{p_{[i]}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\right)$, with its elements given by

$$\begin{aligned}
\dot{Q}^{\beta}_{p_{[i]}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= \left\{\frac{\partial Q_{p_{[i]}}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}}\right\}\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = \frac{1}{\widehat{\sigma}^2}\sum_{j\neq i}\left\{\mathbf{X}_j^\top\widehat{\mathbf{E}}_j^{-1}\left(\widehat{\mathbf{y}}_j - \widehat{\boldsymbol{\mu}}_j - \mathbf{Z}_j\widehat{\mathbf{b}}_j\right)\right\},\\
\dot{Q}^{\mathbf{f}}_{p_{[i]}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= \left\{\frac{\partial Q_{p_{[i]}}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\mathbf{f}}\right\}\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = \frac{1}{\widehat{\sigma}^2}\sum_{j\neq i}\left\{\mathbf{N}_j^\top\widehat{\mathbf{E}}_j^{-1}\left(\widehat{\mathbf{y}}_j - \widehat{\boldsymbol{\mu}}_j - \mathbf{Z}_j\widehat{\mathbf{b}}_j\right)\right\} - \lambda\mathbf{K}\mathbf{f},\\
\dot{Q}^{\sigma^2}_{p_{[i]}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= \left\{\frac{\partial Q_{p_{[i]}}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\sigma^2}\right\}\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = -\frac{1}{2\widehat{\sigma}^2}\sum_{j\neq i}\left\{n_j - \frac{1}{\widehat{\sigma}^2}\left[\widehat{a}_j - 2\widehat{\boldsymbol{\mu}}_j^\top\widehat{\mathbf{E}}_j^{-1}(\widehat{\mathbf{y}}_j - \mathbf{Z}_j\widehat{\mathbf{b}}_j)\right.\right.\\
&+ \left.\left.\widehat{\boldsymbol{\mu}}_j^\top\widehat{\mathbf{E}}_j^{-1}\widehat{\boldsymbol{\mu}}_j\right]\right\},\\
\dot{Q}^{\alpha}_{p_{[i]}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= \left\{\frac{\partial Q_{p_{[i]}}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\alpha}}\right\}\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}},\\
\dot{Q}^{\phi}_{p_{[i]}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= \left\{\frac{\partial Q_{p_{[i]}}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\phi}}\right\}\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}},
\end{aligned}$$

where $\widehat{a}_j = \text{tr}\left(\widehat{\mathbf{y}_j\mathbf{y}_j^\top}\widehat{\mathbf{E}}_j^{-1} - 2\widehat{\mathbf{y}_j\mathbf{b}_j^\top}\mathbf{Z}_j^\top\widehat{\mathbf{E}}_j^{-1} + \widehat{\mathbf{b}_j\mathbf{b}_j^\top}\mathbf{Z}_j^\top\widehat{\mathbf{E}}_j^{-1}\mathbf{Z}_j\right)$, $\widehat{\boldsymbol{\mu}}_j = \mathbf{X}_j\widehat{\boldsymbol{\beta}} + \mathbf{N}_j\widehat{\mathbf{f}}$, and the elements of $\dot{Q}^{\alpha}_{p_{[i]}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ and $\dot{Q}^{\phi}_{p_{[i]}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})$ are of the form

$$\begin{aligned}
\dot{Q}^{\alpha_u}_{p_{[i]}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= -\frac{1}{2}\sum_{j\neq i}\text{tr}\left(\widehat{\mathbf{D}}^{-1}\dot{\mathbf{D}}^u - \widehat{\mathbf{D}}^{-1}\dot{\mathbf{D}}^u\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{b}_j\mathbf{b}_j^\top}\right),\\
\dot{Q}^{\phi_s}_{p_{[i]}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= -\frac{1}{2}\sum_{j\neq i}\text{tr}\left(\widehat{\mathbf{E}}_j^{-1}\dot{\mathbf{E}}_j^s\right) + \frac{1}{2\widehat{\sigma}^2}\sum_{j\neq i}\left\{\text{tr}\left[\left(\widehat{\mathbf{y}_j\mathbf{y}_j^\top} - 2\widehat{\mathbf{y}_j\mathbf{b}_j^\top}\mathbf{Z}_j^\top + \widehat{\mathbf{b}_j\mathbf{b}_j^\top}\mathbf{Z}_j^\top\mathbf{Z}_j\right)\mathbf{A}_j(s)\right]\right.\\
&- \left.2\widehat{\boldsymbol{\mu}}_j^\top\mathbf{A}_j(s)(\widehat{\mathbf{y}}_j - \mathbf{Z}_j\widehat{\mathbf{b}}_j) + \widehat{\boldsymbol{\mu}}_j^\top\mathbf{A}_j(s)\widehat{\boldsymbol{\mu}}_j\right\},
\end{aligned}$$

where $\dot{\mathbf{D}}^u = \dfrac{\partial\mathbf{D}}{\partial\alpha_u}\Big|_{\boldsymbol{\alpha}=\widehat{\boldsymbol{\alpha}}}$, $u = 1,\ldots,\dim(\boldsymbol{\alpha})$; and $\mathbf{A}_j(s) = \widehat{\mathbf{E}}_j^{-1}\dot{\mathbf{E}}_j^s\widehat{\mathbf{E}}_j^{-1}$, $\dot{\mathbf{E}}_j^s = \dfrac{\partial\mathbf{E}_j}{\partial\phi_s}\Big|_{\boldsymbol{\phi}=\widehat{\boldsymbol{\phi}}}$, $s = 1, 2$.

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

$$
\begin{aligned}
\frac{\partial^2 Q_{p_i}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} &= -\frac{1}{\sigma^2}\mathbf{X}_i^\top\mathbf{E}_i^{-1}\mathbf{X}_i,\\
\frac{\partial^2 Q_{p_i}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\mathbf{f}^\top} &= -\frac{1}{\sigma^2}\mathbf{X}_i^\top\mathbf{E}_i^{-1}\mathbf{N}_i,\\
\frac{\partial^2 Q_{p_i}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\sigma^2} &= -\frac{1}{\sigma^4}\mathbf{X}_i^\top\mathbf{E}_i^{-1}\left(\widehat{\mathbf{y}}_i - \boldsymbol{\mu}_i - \mathbf{Z}_i\widehat{\mathbf{b}}_i\right),\\
\frac{\partial^2 Q_{p_i}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\phi_s} &= -\frac{1}{\sigma^2}\mathbf{X}_i^\top\mathbf{A}_i(s)\left(\widehat{\mathbf{y}}_i - \boldsymbol{\mu}_i - \mathbf{Z}_i\widehat{\mathbf{b}}_i\right),\\
\frac{\partial^2 Q_{p_i}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\mathbf{f}\partial\mathbf{f}^\top} &= -\frac{1}{\sigma^2}\mathbf{N}_i^\top\mathbf{E}_i^{-1}\mathbf{N}_i - \frac{\lambda}{n}\mathbf{K},\\
\frac{\partial^2 Q_{p_i}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\mathbf{f}\partial\sigma^2} &= -\frac{1}{\sigma^4}\mathbf{N}_i^\top\mathbf{E}_i^{-1}\left(\widehat{\mathbf{y}}_i - \boldsymbol{\mu}_i - \mathbf{Z}_i\widehat{\mathbf{b}}_i\right),\\
\frac{\partial^2 Q_{p_i}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\mathbf{f}\partial\phi_s} &= -\frac{1}{\sigma^2}\mathbf{N}_i^\top\mathbf{A}_i(s)\left(\widehat{\mathbf{y}}_i - \boldsymbol{\mu}_i - \mathbf{Z}_i\widehat{\mathbf{b}}_i\right),\\
\frac{\partial^2 Q_{p_i}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\sigma^2} &= \frac{n_i}{2\sigma^4} - \frac{1}{\sigma^6}\left[\widehat{a}_i - 2\boldsymbol{\mu}_i^\top\mathbf{E}_i^{-1}(\widehat{\mathbf{y}}_i - \mathbf{Z}_i\widehat{\mathbf{b}}_i) + \boldsymbol{\mu}_i^\top\mathbf{E}_i^{-1}\boldsymbol{\mu}_i\right],\\
\frac{\partial^2 Q_{p_i}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\phi_s} &= -\frac{1}{2\sigma^4}\left\{\mathrm{tr}\left[\left(\widehat{\mathbf{y}_i\mathbf{y}_i^\top} - 2\widehat{\mathbf{y}_i\mathbf{b}_i^\top}\mathbf{Z}_i^\top + \widehat{\mathbf{b}_i\mathbf{b}_i^\top}\mathbf{Z}_i^\top\mathbf{Z}_i\right)\mathbf{A}_i(s)\right]\right.\\
&+ \left. 2\boldsymbol{\mu}_i^\top\mathbf{A}_i(s)(\widehat{\mathbf{y}}_i - \mathbf{Z}_i\widehat{\mathbf{b}}_i) - \boldsymbol{\mu}_i^\top\mathbf{A}_i(s)\boldsymbol{\mu}_i\right\},\\
\frac{\partial^2 Q_{p_i}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\alpha_u\partial\alpha_v} &= \frac{1}{2}\mathrm{tr}\left(\mathbf{B}(u,v)\right) - \frac{1}{2}\mathrm{tr}\left(\mathbf{C}(u,v)\widehat{\mathbf{b}_i\mathbf{b}_i^\top}\right),\\
\frac{\partial^2 Q_{p_i}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\phi_s\partial\phi_t} &= \frac{1}{2\sigma^2}\left\{\mathrm{tr}\left[\left(\widehat{\mathbf{y}_i\mathbf{y}_i^\top} - 2\widehat{\mathbf{y}_i\mathbf{b}_i^\top}\mathbf{Z}_i^\top + \widehat{\mathbf{b}_i\mathbf{b}_i^\top}\mathbf{Z}_i^\top\mathbf{Z}_i\right)\mathbf{A}_i(s,t)\right] - \boldsymbol{\mu}_i^\top\mathbf{A}_i(s,t)\boldsymbol{\mu}_i\right.\\
&+ \left. 2\boldsymbol{\mu}_i^\top\mathbf{A}_i(s,t)(\widehat{\mathbf{y}}_i - \mathbf{Z}_i\widehat{\mathbf{b}}_i)\right\} - \frac{1}{2}\mathrm{tr}\left(\mathbf{A}_i(t)\dot{\mathbf{E}}_i^s + \mathbf{E}_i^{-1}\ddot{\mathbf{E}}_i^{st}\right),\\
\frac{\partial^2 Q_{p_i}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\alpha_r} &= \mathbf{0},\quad \frac{\partial^2 Q_{p_i}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\mathbf{f}\partial\alpha_r} = \mathbf{0},\quad \frac{\partial^2 Q_{p_i}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\alpha_r} = 0,\quad \frac{\partial^2 Q_{p_i}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\alpha_r\partial\phi_s} = 0,
\end{aligned}
$$

where $\mathbf{B}(u,v) = \mathbf{D}^{-1}[\dot{\mathbf{D}}^v\mathbf{D}^{-1}\dot{\mathbf{D}}^u - \ddot{\mathbf{D}}^{uv}]$ and $\mathbf{C}(u,v) = \mathbf{D}^{-1}[\dot{\mathbf{D}}^v\mathbf{D}^{-1}\dot{\mathbf{D}}^u - \ddot{\mathbf{D}}^{uv} + \dot{\mathbf{D}}^u\mathbf{D}^{-1}\dot{\mathbf{D}}^v]\mathbf{D}^{-1}$ with $\ddot{\mathbf{D}}^{uv} = \frac{\partial^2\mathbf{D}}{\partial\alpha_u\partial\alpha_v}$, $u,v = 1,\ldots,\dim(\boldsymbol{\alpha})$, and $\mathbf{A}_i(s,t) = \frac{\partial\mathbf{A}_i(s)}{\partial\phi_t}$, $s,t = 1,2$.

Zhu *et al.* (2001) proposed the generalized Cook distance for models with incomplete data defined by

$$GD_i = (\widehat{\boldsymbol{\theta}}_{[i]} - \widehat{\boldsymbol{\theta}})^\top\{-\ddot{Q}_p(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\}(\widehat{\boldsymbol{\theta}}_{[i]} - \widehat{\boldsymbol{\theta}}),\quad i = 1,\ldots,n. \tag{B.2}$$

Now, upon substituting (B.1) into (B.2), we obtain the approximation

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

B.2 Local influence

In this subsection, we derive the normal curvature of the local influence (Cook, 1986) for some common perturbation schemes either in the model or in the data. We will consider the case-weight, scale matrix perturbation schemes, and response perturbation schemes, for this purpose.

Consider a perturbation vector $\boldsymbol{\omega} = (\omega_1, ..., \omega_g)^\top$ varying in an open region $\boldsymbol{\Omega} \subset \mathbb{R}^g$. Let $\ell_{p_c}(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{y}_{\text{com}})$ be the complete-data-penalized log-likelihood to the perturbed model. We assume that there is a $\boldsymbol{\omega}_0$ in $\boldsymbol{\Omega}$ such that $\ell_{p_c}(\boldsymbol{\theta}, \boldsymbol{\omega}_0|\mathbf{y}_{\text{com}}) = \ell_{p_c}(\boldsymbol{\theta}|\mathbf{y}_{\text{com}})$ for all $\boldsymbol{\theta}$. Let $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ denote the maximum of the function $Q_p(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \mathbb{E}[\ell_{p_c}(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{y}_{\text{com}})|\mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}]$. The influence graph is then defined as $\boldsymbol{\alpha}(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, f_Q(\boldsymbol{\omega}))^\top$, where $f_Q(\boldsymbol{\omega})$ is the Q-displacement function defined as

$$f_Q(\boldsymbol{\omega}) = 2\left[Q_p\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right) - Q_p\left(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})|\widehat{\boldsymbol{\theta}}\right)\right].$$

Following the approach of Cook (1986) and Zhu & Lee (2001), the normal curvature $C_{f_Q,\mathbf{d}}$ of $\boldsymbol{\alpha}(\boldsymbol{\omega})$ at $\boldsymbol{\omega}_0$ in the direction of some unit vector $\mathbf{d}$ can be used to summarize the local behavior of the Q-displacement function. It can be shown that

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

where $\ddot{Q}_p(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \dfrac{\partial^2 Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ and $\boldsymbol{\Delta}_{\boldsymbol{\omega}} = \dfrac{\partial^2 Q_p(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}$.

Following the same procedure as in Cook (1986), the quantity $-\ddot{Q}_{\boldsymbol{\omega}_0}$ is useful for detecting influential observations. From the spectral decomposition of a symmetric matrix $-2\ddot{Q}_{\boldsymbol{\omega}_0} = \sum_{k=1}^{g} \zeta_k \boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k^\top$, where $\{(\zeta_k, \boldsymbol{\varepsilon}_k), k = 1, \ldots, g\}$ are eigenvalue–eigenvector pairs of $-2\ddot{Q}_{\boldsymbol{\omega}_0}$ with $\zeta_1 \geqslant \ldots \geqslant \zeta_r > \zeta_{r+1} = \ldots = 0$ and orthonormal eigenvectors $\{\boldsymbol{\varepsilon}_k, k = 1, \ldots, g\}$, Zhu & Lee (2001) proposed to inspect all eigenvectors corresponding to nonzero eigenvalues for capturing more information. Following the work of Zhu & Lee (2001), we consider the following aggregated contribution vector of all eigenvectors that correspond to nonzero eigenvalues. Let $\tilde{\zeta}_k = \zeta_k/(\zeta_1 + \ldots + \zeta_r)$, $\boldsymbol{\varepsilon}_k^2 = (\boldsymbol{\varepsilon}_{k1}^2, \ldots, \boldsymbol{\varepsilon}_{kg}^2)^\top$ and $M(0) = \sum_{k=1}^{r} \tilde{\zeta}_k \boldsymbol{\varepsilon}_k^2$. The lth component of $M(0)$, $M(0)_l$, is equal to $\sum_{k=1}^{r} \tilde{\zeta}_k \boldsymbol{\varepsilon}_{kl}^2$. The assessment of influential cases is based on the visual inspection of the $\{M(0)_l, l = 1, \ldots, g\}$ plotted against the index l. The lth case may be regarded as influential if $M(0)_l$ is larger than the benchmark value.

The inconvenience in the use of the normal curvature is in deciding about the influence of the observations, since $C_{f_Q,\mathbf{d}}(\boldsymbol{\theta})$ may assume any value and it is not invariant under a uniform change of scale. Based on the work of Poon & Poon (1999) in using a conformal normal curvature, Zhu & Lee (2001) considered 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 0. Zhu & Lee (2001) then showed that for all l, $M(0)_l = B_{f_Q,\mathbf{d}_l}$. We can, therefore, obtain $M(0)_l$ via $B_{f_Q,\mathbf{d}_l}$.

So far, there is no general rule to judge how large is the influence of a specific case in the data. Let $\overline{M}(0)$ and $SM(0)$ denote, respectively, the mean and standard error of $\{M(0)_l : l = 1,\ldots,g\}$, where $\overline{M(0)} = 1/g$. Poon & Poon (1999) proposed to use $2\overline{M}(0)$ as a benchmark for $M(0)$. But, we may use different functions of $M(0)$. For instance, Zhu & Lee (2001) proposed to use $\overline{M}(0) + 2SM(0)$ as a benchmark to take into account the variance of $\{M(0)_l : l = 1,\ldots,g\}$ as well. According to Lee & Xu (2004), the exact choice of the function of $\overline{M}(0)$ as the benchmark is subjective. Lee & Xu (2004) also proposed to use $\overline{M}(0) + c^*SM(0)$, where c^* is a selected constant, and depending on the specific application, c^* may be chosen suitably.

B.2.1 Pertubation schemes

Now, in this subsection, we will evaluate the $\boldsymbol{\Delta}$ matrix under the following perturbation schemes for DEC-SMEC models. *Case-weight* made for detecting observations with outstanding contribution on the log-likelihood function and that may exercise high influence on the maximum likelihood estimates. *Scale perturbation* made on the scale matrix $\boldsymbol{\Sigma}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top + \boldsymbol{\Omega}_i$. It also can be made on either σ^2 or $\mathbf{D}$ which may reveal individuals that are most influential, in the sense, of the likelihood displacement on the scale structure. Finally, *perturbation of response variables* made on the response values, which may indicate observations with large influence on the MPL.

For each perturbation scheme, one has the partitioned form

$$\boldsymbol{\Delta}_{\boldsymbol{\omega}_0} = (\boldsymbol{\Delta}_\beta^\top, \boldsymbol{\Delta}_{\mathbf{f}}^\top, \boldsymbol{\Delta}_{\sigma^2}^\top, \boldsymbol{\Delta}_\alpha^\top, \boldsymbol{\Delta}_\phi^\top)^\top,$$

where $\boldsymbol{\Delta}_\beta = \dfrac{\partial^2 Q_p(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^\top}\Big|_{\boldsymbol{\omega}_0} \in \mathbb{R}^{p\times g}$, $\boldsymbol{\Delta}_{\mathbf{f}} = \dfrac{\partial^2 Q_p(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\mathbf{f}\partial\boldsymbol{\omega}^\top}\Big|_{\boldsymbol{\omega}_0} \in \mathbb{R}^{r\times g}$, $\boldsymbol{\Delta}_{\sigma^2} = \dfrac{\partial^2 Q_p(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\boldsymbol{\omega}^\top}\Big|_{\boldsymbol{\omega}_0} \in \mathbb{R}^{1\times g}$, and $\boldsymbol{\Delta}_\alpha = (\boldsymbol{\Delta}_{\alpha_1}^\top,\ldots,\boldsymbol{\Delta}_{\alpha_{q*}}^\top)^\top$, $\boldsymbol{\Delta}_\phi = (\boldsymbol{\Delta}_{\phi_1}^\top, \boldsymbol{\Delta}_{\phi_2}^\top)$, with $\boldsymbol{\Delta}_{\alpha_u} = \dfrac{\partial^2 Q_p(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\alpha_u\partial\boldsymbol{\omega}^\top}\Big|_{\boldsymbol{\omega}_0} \in \mathbb{R}^{1\times g}$, $u = 1,\ldots,\dim(\boldsymbol{\alpha})$, $\boldsymbol{\Delta}_{\phi_s} = \dfrac{\partial^2 Q_p(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\phi_s\partial\boldsymbol{\omega}^\top}\Big|_{\boldsymbol{\omega}_0} \in \mathbb{R}^{1\times g}$, $s = 1,2$ and g being the dimensions of the perturbation vector $\boldsymbol{\omega}$.

Case weight perturbation

First, 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, represented by writing

$$Q_p(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \mathbb{E}\left[\ell_{p_c}(\boldsymbol{\theta},\boldsymbol{\omega}|\mathbf{y}_{\text{com}})|\mathbf{V},\mathbf{C},\widehat{\boldsymbol{\theta}}\right] = \sum_{i=1}^n \omega_i \mathbb{E}\left[\ell_{p_i}(\boldsymbol{\theta}|\mathbf{y}_{\text{com}})|\mathbf{V},\mathbf{C},\widehat{\boldsymbol{\theta}}\right] = \sum_{i=1}^n \omega_i Q_{p_i}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}).$$

Here $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ is a $n \times 1$ vector and $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$. For this perturbation scheme, we find

$$
\begin{aligned}
\boldsymbol{\Delta}_\beta &= \frac{1}{\sigma^2}\left[\mathbf{X}_1^\top \mathbf{E}_1^{-1}\mathcal{E}_1, \ldots, \mathbf{X}_n^\top \mathbf{E}_n^{-1}\mathcal{E}_n\right], \\
\boldsymbol{\Delta}_{\mathbf{f}} &= \frac{1}{\sigma^2}\left[\mathbf{N}_1^\top \mathbf{E}_1^{-1}\mathcal{E}_1 - \frac{\lambda}{n}\mathbf{K}\mathbf{f}, \ldots, \mathbf{N}_n^\top \mathbf{E}_n^{-1}\mathcal{E}_n - \frac{\lambda}{n}\mathbf{K}\mathbf{f}\right], \\
\boldsymbol{\Delta}_{\sigma^2} &= \left[-\frac{n_1}{2\sigma^2} + \frac{m_1}{2\sigma^4}, \ldots, -\frac{n_n}{2\sigma^2} + \frac{m_n}{2\sigma^4}\right], \\
\boldsymbol{\Delta}_{\alpha_u} &= \left[\frac{\partial Q_{p_1}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \alpha_u}, \ldots, \frac{\partial Q_{p_n}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \alpha_u}\right], \quad u = 1, \ldots, \dim(\boldsymbol{\alpha}), \\
\boldsymbol{\Delta}_{\phi_s} &= \left[\frac{\partial Q_{p_1}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \phi_s}, \ldots, \frac{\partial Q_{p_n}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \phi_s}\right], \quad s = 1, 2,
\end{aligned}
$$

where $\mathcal{E}_i = (\widehat{\mathbf{y}}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{N}_i\mathbf{f} - \mathbf{Z}_i\widehat{\mathbf{b}}_i)$, $m_i = \left(\widehat{a}_i - 2\boldsymbol{\mu}_i^\top \mathbf{E}_i^{-1}(\widehat{\mathbf{y}}_i - \mathbf{Z}_i\widehat{\mathbf{b}}_i) + \boldsymbol{\mu}_i^\top \mathbf{E}_i^{-1}\boldsymbol{\mu}_i\right)$,

$$
\frac{\partial Q_{p_i}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \alpha_u} = -\frac{1}{2}\mathrm{tr}\left(\mathbf{D}^{-1}\dot{\mathbf{D}}^u - \mathbf{D}^{-1}\dot{\mathbf{D}}^u\mathbf{D}^{-1}\widehat{\mathbf{b}_i\mathbf{b}_i^\top}\right), \quad \text{and}
$$

$$
\begin{aligned}
\frac{\partial Q_{p_i}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \phi_s} &= -\frac{1}{2}\mathrm{tr}\left(\mathbf{E}_i^{-1}\dot{\mathbf{E}}_i^s\right) + \frac{1}{2\sigma^2}\left\{\mathrm{tr}\left[\left(\widehat{\mathbf{y}_i\mathbf{y}_i^\top} - 2\widehat{\mathbf{y}_i\mathbf{b}_i^\top}\mathbf{Z}_i^\top + \widehat{\mathbf{b}_i\mathbf{b}_i^\top}\mathbf{Z}_i^\top\mathbf{Z}_i\right)\mathbf{A}_i(s)\right]\right. \\
&- \left. 2\boldsymbol{\mu}_i^\top\mathbf{A}_i(s)(\widehat{\mathbf{y}}_i - \mathbf{Z}_i\widehat{\mathbf{b}}_i) + \boldsymbol{\mu}_i^\top\mathbf{A}_i(s)\boldsymbol{\mu}_i\right\}.
\end{aligned}
$$

Scale matrix perturbation

To study the effects of departures from the assumption regarding the scale matrix, we consider the perturbations $\mathbf{D}(\omega_i) = \omega_i^{-1}\mathbf{D}$ or $\sigma^2(\omega_i) = \omega_i^{-1}\sigma^2$, for $i = 1, \ldots, n$. Under this perturbation scheme, the non-perturbed model is obtained when $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$. Moreover, the perturbed Q-function is as in (2.7), switching $\mathbf{D}(\omega_i)$ and $\sigma^2(\omega_i)$ with $\mathbf{D}$ and σ^2, respectively. The matrix $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$ has its elements as follows:

- Perturbation on $\mathbf{D}$: $\boldsymbol{\Delta}_\beta = \mathbf{0}$, $\boldsymbol{\Delta}_{\mathbf{f}} = \mathbf{0}$, $\boldsymbol{\Delta}_{\sigma^2} = \mathbf{0}$, $\boldsymbol{\Delta}_\phi = \mathbf{0}$ and

$$
\boldsymbol{\Delta}_{\alpha_u} = \frac{1}{2}\left[\mathrm{tr}\left(\widehat{\mathbf{D}}^{-1}\dot{\mathbf{D}}^u\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{b}_1\mathbf{b}_1^\top}\right), \ldots, \mathrm{tr}\left(\widehat{\mathbf{D}}^{-1}\dot{\mathbf{D}}^u\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{b}_n\mathbf{b}_n^\top}\right)\right], \quad u = 1, \ldots, \dim(\boldsymbol{\alpha}).
$$

- Perturbation on σ^2:

$$
\begin{aligned}
\boldsymbol{\Delta}_\beta &= \frac{1}{\sigma^2}\left[\mathbf{X}_1^\top \mathbf{E}_1^{-1}\mathcal{E}_1, \ldots, \mathbf{X}_n^\top \mathbf{E}_n^{-1}\mathcal{E}_n\right], \\
\boldsymbol{\Delta}_{\mathbf{f}} &= \frac{1}{\sigma^2}\left[\mathbf{N}_1^\top \mathbf{E}_1^{-1}\mathcal{E}_1 - \frac{\lambda}{n}\mathbf{K}\mathbf{f}, \ldots, \mathbf{N}_n^\top \mathbf{E}_n^{-1}\mathcal{E}_n - \frac{\lambda}{n}\mathbf{K}\mathbf{f}\right], \\
\boldsymbol{\Delta}_{\sigma^2} &= \frac{1}{2\sigma^4}[m_1, \ldots, m_n], \\
\boldsymbol{\Delta}_\alpha &= \mathbf{0}, \\
\boldsymbol{\Delta}_{\phi_s} &= \frac{1}{2\sigma^2}\left[c_1, \ldots, c_n\right],
\end{aligned}
$$

where $c_i = \text{tr}\left[\left(\widehat{\mathbf{y}_i\mathbf{y}_i^\top} - 2\widehat{\mathbf{y}_i\mathbf{b}_i^\top}\mathbf{Z}_i^\top + \widehat{\mathbf{b}_i\mathbf{b}_i^\top}\mathbf{Z}_i^\top\mathbf{Z}_i\right)\mathbf{A}_i(s)\right] - 2\boldsymbol{\mu}_i^\top\mathbf{A}_i(s)(\widehat{\mathbf{y}}_i - \mathbf{Z}_i\widehat{\mathbf{b}}_i) + \boldsymbol{\mu}_i^\top\mathbf{A}_i(s)\boldsymbol{\mu}_i$.

Response perturbation

A perturbation of the response variables V_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, n_i$, can be introduced by replacing V_{ij} by $V_{ij}(\omega) = V_{ij} + \omega_i g_{ij}$, where g_{ij} is a known constant. Hence, for the DEC-SMEC model, the perturbed response is obtained as

$$C_{ij} = \begin{cases} 1 & \text{if} \quad V_{1ij} \leqslant y_{ij}(\omega) \leqslant V_{2ij}, \\ 0 & \text{if} \quad y_{ij}(\omega) = V_{0i}, \end{cases}$$

where $y_{ij}(\omega) = y_{ij} - \omega_i g_{ij}$. Again, the perturbed Q-function follows (2.7) with $\widehat{\mathbf{y}}_i$, $\widehat{\mathbf{y}_i\mathbf{y}_i^\top}$ and $\widehat{\mathbf{y}_i\mathbf{b}_i^\top}$ replaced by $\widehat{\mathbf{y}}_{i\omega} = \widehat{\mathbf{y}}_i - \omega_i\mathbf{g}_i$, $\widehat{\mathbf{y}_{i\omega}\mathbf{y}_{i\omega}^\top} = \widehat{\mathbf{y}_i\mathbf{y}_i^\top} - \omega_i(\widehat{\mathbf{y}}_i\mathbf{g}_i^\top + \mathbf{g}_i\widehat{\mathbf{y}}_i^\top) + \omega_i^2\mathbf{g}_i\mathbf{g}_i^\top$ and $\widehat{\mathbf{y}_{i\omega}\mathbf{b}_{i\omega}^\top} = \widehat{\mathbf{y}_i\mathbf{b}_i^\top} - \omega_i\mathbf{g}_i\widehat{\mathbf{b}}_i^\top$, respectively, with $\mathbf{g}_i = (g_{i1}, \ldots, g_{in_i})^\top$. Under this perturbation scheme the vector ω_0, representing no perturbation, is given by $\omega_0 = \mathbf{0}$ and $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$ has the following elements:

$$\begin{aligned}
\boldsymbol{\Delta}_\beta &= -\frac{1}{\sigma^2}\left[\mathbf{X}_1^\top\mathbf{E}_1^{-1}\mathbf{g}_1, \ldots, \mathbf{X}_n^\top\mathbf{E}_n^{-1}\mathbf{g}_n\right], \\
\boldsymbol{\Delta}_\mathbf{f} &= -\frac{1}{\sigma^2}\left[\mathbf{N}_1^\top\mathbf{E}_1^{-1}\mathbf{g}_1, \ldots, \mathbf{N}_n^\top\mathbf{E}_n^{-1}\mathbf{g}_n\right], \\
\boldsymbol{\Delta}_{\sigma^2} &= -\frac{1}{\sigma^4}\left[\mathcal{E}_1^\top\mathbf{E}_1^{-1}\mathbf{g}_1, \ldots, \mathcal{E}_n^\top\mathbf{E}_n^{-1}\mathbf{g}_n\right], \\
\boldsymbol{\Delta}_\alpha &= \mathbf{0}, \\
\boldsymbol{\Delta}_{\phi_s} &= -\frac{1}{\sigma^2}\left[\mathcal{E}_1^{-1}\mathbf{A}_1(s)\mathbf{g}_1, \ldots, \mathcal{E}_n^{-1}\mathbf{A}_n(s)\mathbf{g}_n\right].
\end{aligned}$$