UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Estimation methods in heavy-tailed nonlinear mixed-effects models

José Clelto Barros Gomes

Tese de Doutorado do Programa Interinstitucional de Pós-Graduação em Estatística (PIPGEs)

ICMC USP
SÃO CARLOS

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA
UFSCar-USP

José Clelto Barros Gomes

Estimation methods in heavy-tailed nonlinear mixed-effects models

Doctoral dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP and to the Department of Statistics – DEs-UFSCar, in partial fulfillment of the requirements for the degree of the Doctorate Interagency Program Graduate in Statistics. *FINAL VERSION*

Concentration Area: Statistics

Advisor: Profa. Dra. Cibele Maria Russo Novelli

USP – São Carlos
March 2020

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

Gomes, José Clelto Barros
G633e Estimation methods in heavy-tailed nonlinear mixed effects models / José Clelto Barros Gomes; orientadora Cibele Maria Russo Novelli. -- São Carlos, 2019.
101 p.

Tese (Doutorado - Programa Interinstitucional de Pós-graduação em Estatística) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2019.

1. nonlinear models. 2. mixed-effects models. 3. estimation methods. 4. ML. 5. REML. I. Russo Novelli, Cibele Maria, orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA
UFSCar-USP

José Clelto Barros Gomes

Métodos de estimação em modelos de efeitos mistos não lineares de caudas pesadas

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Doutor em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

Área de Concentração: Estatística

Orientadora: Profa. Dra. Cibele Maria Russo Novelli

USP – São Carlos
Março de 2020

To my mum Antônia (in memoriam).

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Tese de Doutorado do candidato José Clelto Barros Gomes, realizada em 05/12/2019:

Profa. Dra. Cibele Maria Russo Novelli
USP

Prof. Dr. Felipe Alberto Osório Salgado
UTFSM

Profa. Dra. Camila Borelli Zeller
UFJF

Prof. Dr. Cristian Marcelo Villegas Lobos
ESALQ/USP

Prof. Dr. Jorge Luis Bazán Guzmán
USP

ACKNOWLEDGEMENTS

First of all, I would like to thank God for His wonderful creation and for allowing me to come here.

I would also like to thank my advisor, Prof. Dr. Cibele Maria Russo Novelli, for her patience, encouragement, professionalism and support while preparing of this work.

I would like to thank my family, Antônia (mum, in memoriam), Antônio (dad), Antônio, Antônia, Cleia, Max and Maria Eduarda for their support during all the years of study.

I gratefully acknowledge UFAM, particularly the Department of Statistics, for permitting and supporting my leave so that I could do my PhD.

I also acknowledge Dr. Geert Molenberghs, Dr. Geert Verbeke from KU Leuven for the seven months of my sandwich PhD exchange in Leuven, Belgium.

I would like to thank my lecturers at PIPGEs for teaching me.

I would also like to thank my friends Cleber, Daiane, Danila, Jeremias, José Mir, Roberta, Themis and everyone else who contributed to this work in any way.

This research was carried out using the computational resources of the Center for Mathematical Sciences Applied to Industry (CeMEAI) funded by FAPESP (grant 2013/07375-0).

This research was developed using HPC resources provided by the Information Technology Superintendence at the University of Sao Paulo.

I would also like to thank the committee members Professor Dr. Felipe Alberto Osorio Salgado (UTFSM), Professor Dr. Camila Borelli Zeller (UFJF), Professor Dr. Cristian Marcelo Villegas Lobos (ESALQ/USP) and Professor Dr. Jorge Luis Bazan (ICMC/USP).

Finally, I would like to thank the Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM) for their financial support for this project.

"Eu quase que nada não sei,
mas desconfio de muita coisa."
(Guimarães Rosa)

RESUMO

GOMES, J. C. B. **Métodos de estimação em modelos de efeitos mistos não lineares de caudas pesadas**. 2020. 112 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2020.

A estimação de parâmetros em modelos não lineares com efeitos mistos é muitas vezes desafiadora. Neste trabalho, propomos a comparação de alguns de métodos de estimação nesses modelos sob o enfoque frequentista. Em um primeiro momento, propomos um estimador de máxima verossimilhança em um esquema de estimação exata contra o estimador de máxima versossimilhança em um modelo linearizado pela expansão de Taylor, o que é frequentemente utilizado na literatura. No primeiro cenário usamos o algoritmo MCEM. Em um segundo momento, visando diminuir o viés para estimativas das componentes de variância, propomos um estimador de máxima verossimilhança restrita também dentro de um esquema de estimação exata, baseada na integração da função de verossimilhança em relação aos efeitos fixos. Esse estimador é comparado com o de máxima verossimilhança. Neste caso, usamos o algoritmo SAEM, para os dois métodos de estimação. Assume-se para os erros e efeitos aleatórios algumas distribuições simétricas multivariadas de escala de misturas de distribuições normais, que compõem a classe de distribuições de caudas pesadas, a saber: normal, t e slash. Por último propomos um modelo não linear mais flexível, em que não é assumida uma forma linear para a inclusão dos efeitos aleatórios. Em todos os casos utilizamos dados reais e estudos de simulação para avaliar as propriedades dos estimadores.

Palavras-chave: modelos não-lineares, modelos mistos, dados correlacionados, métodos de estimação, máxima verossimilhança, máxima verossimilhança restrita .

ABSTRACT

GOMES, J. C. B. **Estimation methods in heavy-tailed nonlinear mixed-effects models**. 2020. 112 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2020.

Parameter estimation in nonlinear mixed-effects models is often challenging. In this thesis, a comparison of estimation methods for these models is proposed under a frequentist approach. In the first study, a comparison of maximum likelihood estimates under an exact method via Monte Carlo expectation-maximization (MCEM) and an approximate method based on a Taylor expansion, frequently used in the literature, is provided. In a second study, a restricted maximum likelihood estimation method is proposed, aiming to decrease the bias for the variance components estimates, based on the integration of the likelihood function on the fixed-effects, also in an exact likelihood context. These estimates are compared to the maximum likelihood ones. For the latter comparison, stochastic approximation of expectation-maximization (SAEM) algorithms are considered. The random effects and errors are assumed to follow multivariate symmetric distributions, namely the scale mixture of normal distributions, which include the normal, t and slash distributions. Finally, a general nonlinear mixed-effects model is proposed, where no linear relation is assumed in the random effects structure. In all the proposals, real data sets and simulation studies are used to illustrate the estimates' properties.

Keywords: nonlinear models, mixed-effects models, correlated data, estimation methods, maximum likelihood, restricted maximum likelihood..

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AIC	Akaike information criterion
BIC	Bayesian information criterion
HVC	high vocal center
LME	linear mixed-effects
MCEM	Monte-Carlo expectation-maximization
MCMC	Monte Carlo via Markov chain
ML	maximum likelihood
MRI	magnetic resonance imaging
MSE	mean squared error
NLME	nonlinear mixed-effects
PNLS	penalized nonlinear least squares
RA	robustus arcopallii
REML	restricted maximum likelihood
RMSE	square root of the mean squared error
SAEM	stochastic-approximation expectation-maximization
SI	signal intensity
SMN	scale mixture of normal

CONTENTS

CHAPTER

1

INTRODUCTION

Nonlinear mixed-effects models (NLME) aim to model the nonlinear relationship between a response variable and covariates. Mixed-effects models are usually proposed for problems with correlated data, such as longitudinal or clustered data, with applications in several areas such as epidemiology, pharmacokinetics, economics and agriculture. An introduction to the theory of nonlinear mixed-effects models can be found in Davidian and Giltinan (2003) and in Vonesh and Chinchilli (1997) and a large number of applications can be found in Demidenko (2013) and Pinheiro and Bates (2000). Although nonlinearity often occurs in both fixed and random effects, some authors assume that the random effects can be added linearly to the model (see, for instance, Russo, Paula and Aoki (2009)). However, even when the random effects are included nonlinearly in the model, they are often linearly connected to a set of covariates.

Nonlinear models usually bring a great challenge in parameter estimation; frequently the likelihood does not have an analytic solution (PINHEIRO; BATES, 2000) which leads to using numerical methods. There are several numerical approximation methods useful for approximation, namely: linearization methods, such as first-order Taylor-series expansion; integral approximation methods which use Laplace approximation methods based on the second-order Taylor-series expansion, Gaussian quadrature and importance sampling; expectation-maximization (EM) algorithms.

Approximation methods often face estimation problems when intraindividual measures are small and the variability of random effects is large (Davidian and Giltinan (1995); Pinheiro and Bates (1995), Lindstrom and Bates (1990)). To circumvent these problems, exact methods such as Monte Carlo EM methods are proposed by Wei and Tanner (1990), Walker (1996) and Wang (2007). These methods are efficient but require considerable computation time. Delyon, Lavielle and Moulines (1999) proposed a stochastic version of the EM algorithm. Later, Kuhn and Lavielle (2005) showed that the SAEM is very efficient for calculating ML estimates (see also in Meza, Jaffrézic and Foulley (2007) and Meza,

Osorio and De La Cruz (2012)).

The model by Meza, Osorio and De La Cruz (2012), for example, is too restrictive in the sense that the model assumes a linear predictor for the subject-specific parameter vector in nonlinear function. Here, we propose to remove this condition for subject-specific parameter vector in such a way that it is not necessary to specify it.

A Bayesian approach to NLME models was proposed by Bennett and Wakefield (1993), Wakefield (1996) and Wakefield *et al.* (1994). These authors studied a three-stage model and Monte Carlo via Markov Chains (MCMC). In the first stage they specified the conditional density of the observations given the random effects and the fixed effects. In the second stage, they specified the density of the random effects, and in the third stage, an a prior density for the parameters. The posterior density of the random effects can be obtained using an MCMC algorithm. The most popular MCMC methods are Gibbs sampler and Metropolis-Hastings. MCMC has enormous potential in dealing with mixed-effect models with high-dimensional random effects.

In the Bayesian context, many authors have been investing in estimating models with mixed effects, linear and non-linear. Chen (2012) studied mixed models, with asymmetric elliptical distributions whereas Kazemi *et al.* (2013) considers a class of asymmetric elliptical distributions in mixed multivariate linear models.

Lachos, Castro and Dey (2013) considered independent normal distributions for errors and random effects in the Bayesian context, namely: normal, t, slash and contaminated normal distribution. The authors also studied diagnostics based on the case deletion and data perturbation, model selection criteria and presented an application in data from patients infected with AIDS/HIV.

De La Cruz (2014) studied an application of the theophylline data set in the estimation of NLME, considering the same class of distributions addressed in Lachos, Castro and Dey (2013) using Bayesian methodology.

1.1 Preliminary results

As preliminary results related to this thesis, one can cite:

- Gomes, J. C. B.; Russo, C.M. REML in nonlinear mixed-effects models with heavy-tailed distributions. In: 32nd International Workshop on Statistical Modelling, 2017, Groningen. Proceedings of the 32nd International Workshop on Statistical Modelling, v.1. p.184-189.

- Gomes, J. C. B.; Russo, C. M. Estimador de máxima verossimilhança restrita em modelos não-lineares com efeitos mistos. In: XV Escola de modelos de Regressão,

2017, Goiânia. XV Escola de modelos de Regressão, 2017. (*In Portuguese*)

- Russo, C. M, Gomes, J. C. B., Lachos, V. H., Aoki, R. and Paula, G. A. Fast inference for robust nonlinear mixed–effects models. Submitted to Journal of Applied Statistics, 2018.

It is important to comment that the participation in the 32nd IWSM was granted by the Statistical Modelling Society, as follows.

- IWSM Travel Grant from Statistical Modelling Society, to participate in the 32nd International Workshop on Statistical Modelling in Groningen, Netherlands.

Other production developed in the period of the PhD course but not included as chapters of this thesis is

- Gomes, J. C. B.; Fossaluza, V. ; Russo, C. M. . Modelling cortisol data in Brazilian children using Bayesian linear mixed-effects models. In: 60th ISI World Statistics Congress (WSC), 2015, Rio de Janeiro. Proceedings: 60th ISI World Statistics Congress (WSC), 2015.

1.2 Aim and outline of this thesis

The main objective of this thesis is to study estimation methods in nonlinear mixed-effects models with heavy-tailed distributions assumed for the random components, namely the random effects and errors. In particular, for the variance components, restricted maximum likelihood and maximum likelihood methods are considered, using expectation-maximization and variants as stochastic-approximation expectation-maximization (SAEM) and Monte-Carlo expectation-maximization (MCEM) methods.

In Chapter 2, the motivating examples for the development of the nonlinear mixed-effects (NLME) models are presented. The first one is related to growth curve models, for which a three-parameter logistic model will be used. The second one is related to pharmacokinetics of theophylline, for which a first compartment model will be considered. The third motivating example deals with measurements of signal intensity in the songbird brain.

Chapter 3 is devoted to preliminary definitions of scale mixture of normal distributions, as well as specific distributions in this class and the criteria used for models selection, that are AIC and BIC.

Chapter 4 aims to compare two simple methods for maximum likelihood estimation in nonlinear mixed-effects models, namely the exact method, which is based on a Monte

Carlo EM-type algorithm and an approximated method based on the first-order Taylor expansion of the nonlinear function. The approximated method is the most common approach used for nonlinear mixed-effects models in the literature. A Monte Carlo simulation study is presented to compare both methods according to bias and mean-squared error.

In Chapter 5, the main interest lies in comparing a restricted maximum likelihood (REML) estimation to maximum likelihood (ML) estimation methods, aiming to verify mainly if the REML produces less-biased estimates for the variance components in NLMEMs. A very interesting result is shown via a simulation study: maximum likelihood methods may underestimate variance components. In some scenarios, more than 50% of obtained estimates for specific variance components are lower than the theoretical value used to generate the data.

Chapter 6 brings a new proposal for general nonlinear mixed-effects models, where the random effects may be included nonlinearly to the model, without needing a linear structure to link the random effects to the nonlinear function. Scale mixture of normal distributions are also considered for the distribution of the response given the random effects and a stochastic-approximation expectation maximization algorithm (SAEM) is used. The songbird data is considered for numerical illustration and global influence is performed to assess the effect of excluding observations in the estimating process.

Finally, in Chapter 7 some discussion and future research ideas are provided.

CHAPTER

2

MOTIVATING EXAMPLES

In this chapter, three motivating examples of nonlinear correlated data are presented. The first dataset deals with a growth curve problem and it was discussed in Pinheiro and Bates (1995). The second dataset appears frequently in nonlinear mixed-effects model studies and describes the pharmacokinetic absorption and elimination of the anti-asthmatic theophylline. Authors that have already worked on this application with nonlinear mixed-effects models are Pinheiro and Bates (2000) and Meza, Osorio and De La Cruz (2012). The third motivating example was discussed by Molenberghs and Verbeke (2005), chap. 20, and Serroyen *et al.* (2009) and deals with measurements of signal intensity in different areas of song control system in the songbird brain.

2.1 Growth curve problem

The three-parameter logistic model is frequently used to model growth curve data where the mean of a response variate Y is related to a covariate T (frequently the time) according to the nonlinear function g as follows:

$$\mathrm{E}(Y) = g(\beta_1, \beta_2, \beta_3, T) = \frac{\beta_1}{1 + \exp\{-[T - \beta_2]/\beta_3\}}.$$

One example is the growth of soybean plants (see, for instance, Pinheiro and Bates (2000)), where Y is the average leaf weight per plant (in g) and T is the time after planting (in days). The parameters β_1, β_2 and β_3 have physical interpretations according to the response variable, in the example where Y represents the leaf weight, then the parameters β_1, β_2 and β_3 represent the asymptotic leaf weight, the time at which the leaf reaches half of its asymptotic weight and the time elapsed between the leaf reaching half and $1/(1+e^{-1}) \approx 3/4$ of its asymptotic weight, respectively. The observed dataset, unbalanced, is presented in Figure 1, where the points indicate the measurements and line segments illustrate subsequent measurements taken in the same plot. The leaf weight

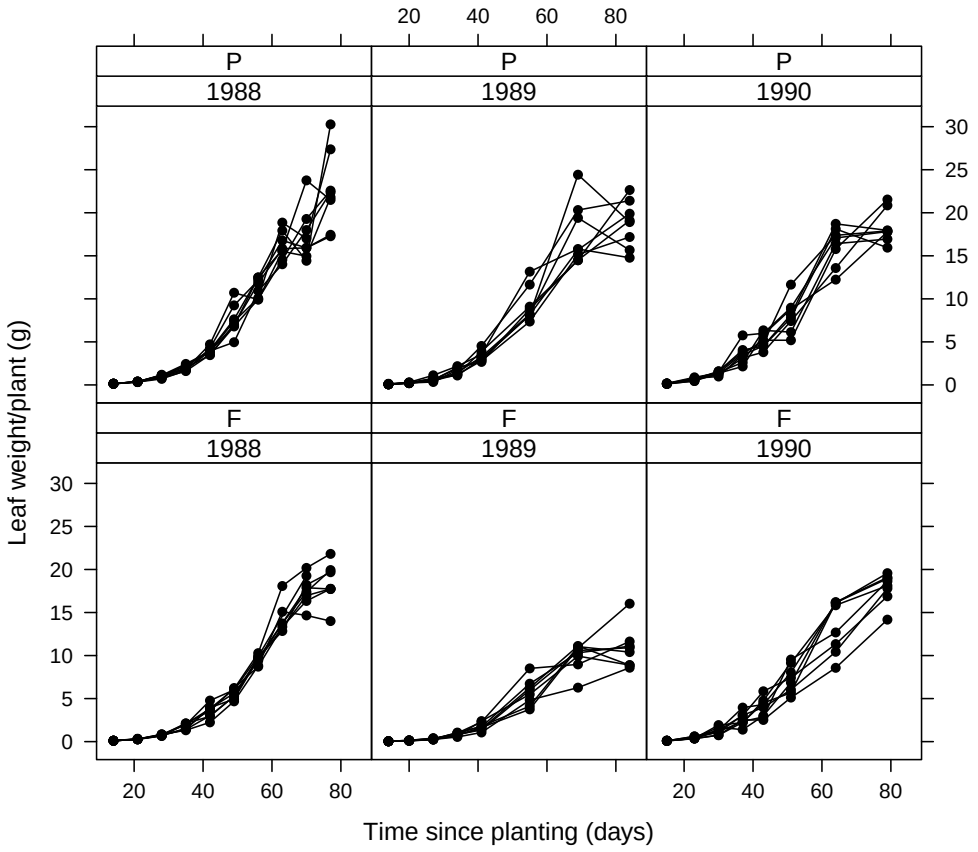


Figure 1 – Leaf weight versus the time for the soybean plant dataset.

measurements come from two plant varieties, P (*plant introduction*) and F (*forrest*), in 1988, 1989 and 1990. This dataset is available in the computational package R under the name Soybean {nlme}. Further details on the dataset will be discussed in Subsection 4.5.1.

2.2 Pharmacokinetic problem

Another nonlinear problem considered here is the pharmacokinetics of a substance in the body. This type of problem involves the absorption and elimination of a substance, and it is usual to model the mean concentration of the substance, Y, by using the nonlinear function g of time, T, and dose, D, as follows

$$\mathrm{E}(Y) = g(lK_\mathrm{e}, lK_\mathrm{a}, lC_\mathrm{l}, T, D) = D\exp(lK_\mathrm{e} + lK_\mathrm{a} - lC_\mathrm{l})\frac{[\exp(-e^{lK_\mathrm{e}}T) - \exp(-e^{lK_\mathrm{a}}T)]}{e^{lK_\mathrm{a}} - e^{lK_\mathrm{e}}}.$$

An example is described by Pinheiro and Bates (2000) on the anti-asthmatic agent theophylline, where the serum concentration of the substance, Y (in mg/L), was measured eleven times (in h) after administering a D dose (in mg/kg) in each of the 12 patients.

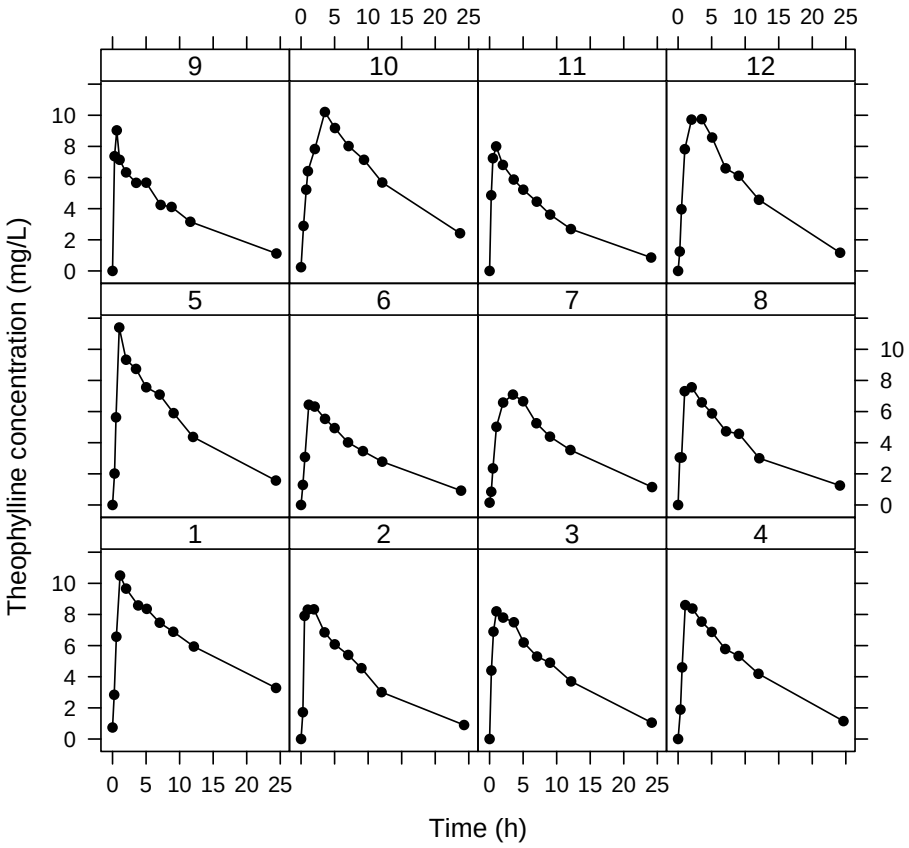


Figure 2 – Theophylline dataset.

The nonlinearity of the data can be observed in Figure 2. This dataset is available in the R program under the name *Theoph {datasets}*.

It is usual to call this nonlinear model a first order compartment model, with the following interpretation for the parameters: lK_{a} represents the logarithm of the substance absorption rate, lK_{e} is the logarithm of the substance elimination rate and lC_{l} represents the logarithm of plasma clearance. In this case, the fixed effects parameter vector is $\boldsymbol{\beta} = (lK_{\mathrm{a}}, lk_{\mathrm{e}}, lC_{\mathrm{l}})^{\top}$. For more details, refer to Section 4.5.2.

2.3 Songbird - RA data

The songbird data were initially presented by Van der Linden *et al.* (2002) and Van Meir *et al.* (2004), who established a novel in vivo magnetic resonance imaging (MRI) approach to discern the functional characteristics of specific neuronal populations in a strongly connected brain circuitry, the so-called song control system in the songbird brain. The high vocal center (HVC), one of the major nuclei in this circuit, contains interneurons and two distinct types of neurons projecting respectively to the so-called nucleus robustus arcopallii (RA) or to area X, according to Molenberghs and Verbeke

(2005), Chap. 20. A representation of the song control nuclei in the songbird brain can be found in Figure 3. Serroyen *et al.* (2009) analyzed this dataset with a nonlinear mixed-effects model assuming normal distribution. After selecting the model, they settled on the following model for MRI signal intensity (SI) in RA.

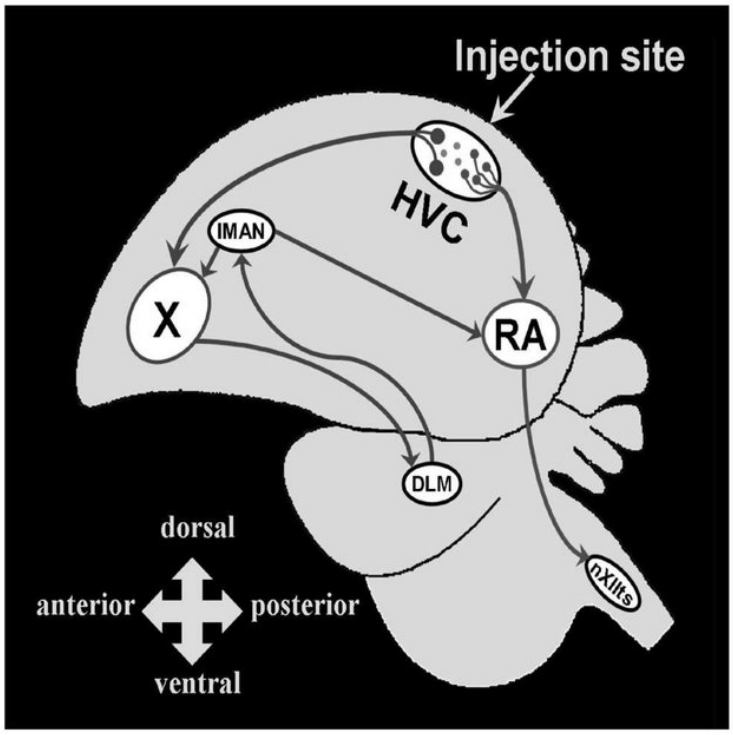


Figure 3 – Schematic representation of song control nuclei in the songbird brain, figure from Serroyen *et al.* (2009).

The model to be considered is

$$E(Y) = g(\phi_0, \tau_0, \eta_0, \eta_1, T, G) \quad = \quad \frac{(\phi_0) T^{\eta_0+\eta_1 G}}{(\tau_0)^{\eta_0+\eta_1 G} + T^{\eta_0+\eta_1 G}},$$

where Y is the measurement of MRI signal intensity for the bird, G is an indicator for group membership, where 1 is for testosterone treated birds and 0 otherwise, and T is the measurement time. The fixed effects are described as: ϕ_0 is the maximal signal intensity, sometimes called SI_{max}, for an untreated bird; τ_0 is the time required to reach 50% of this maximum (T_{50}); η_0 and $\eta_0 + \eta_1$ govern the shape of the curve. In this case, the fixed effects parameter vector is $\boldsymbol{\beta} = (\phi_0, \eta_0, \eta_1, \tau_0)$.

Figures 4 and 5 show the data for songbird HVC and RA.

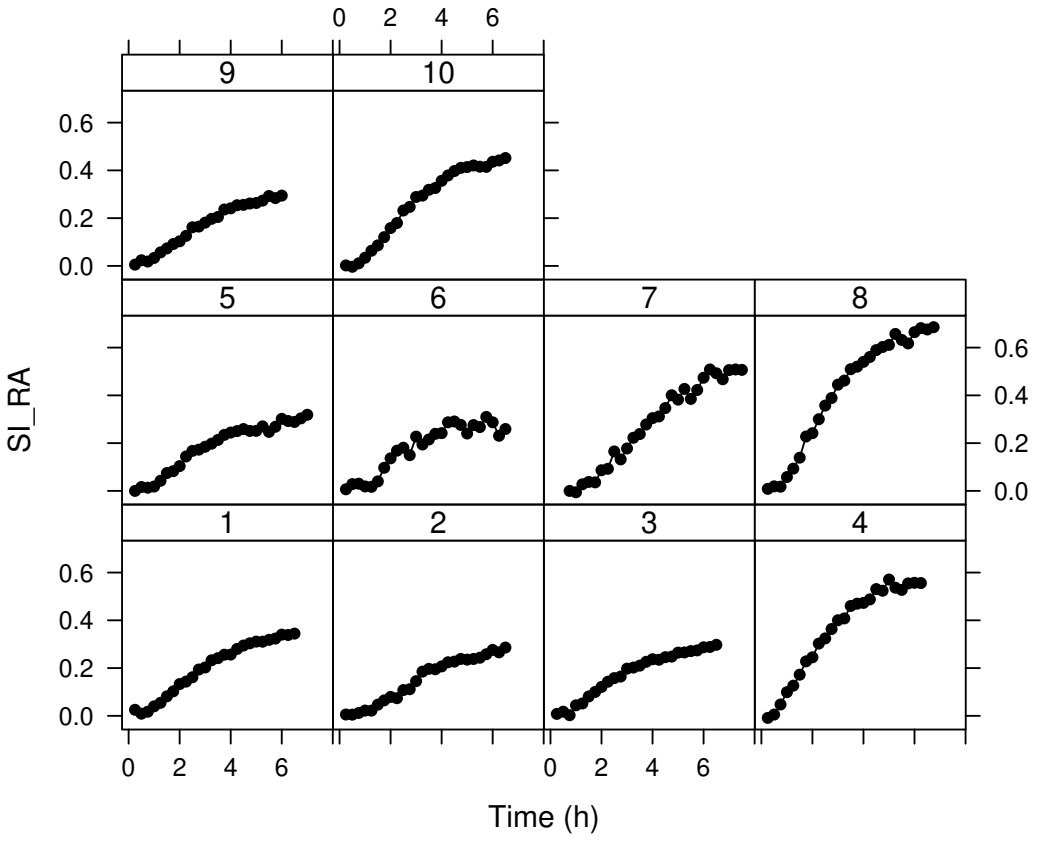


Figure 4 – Songbird data for SI in RA for each individual bird.

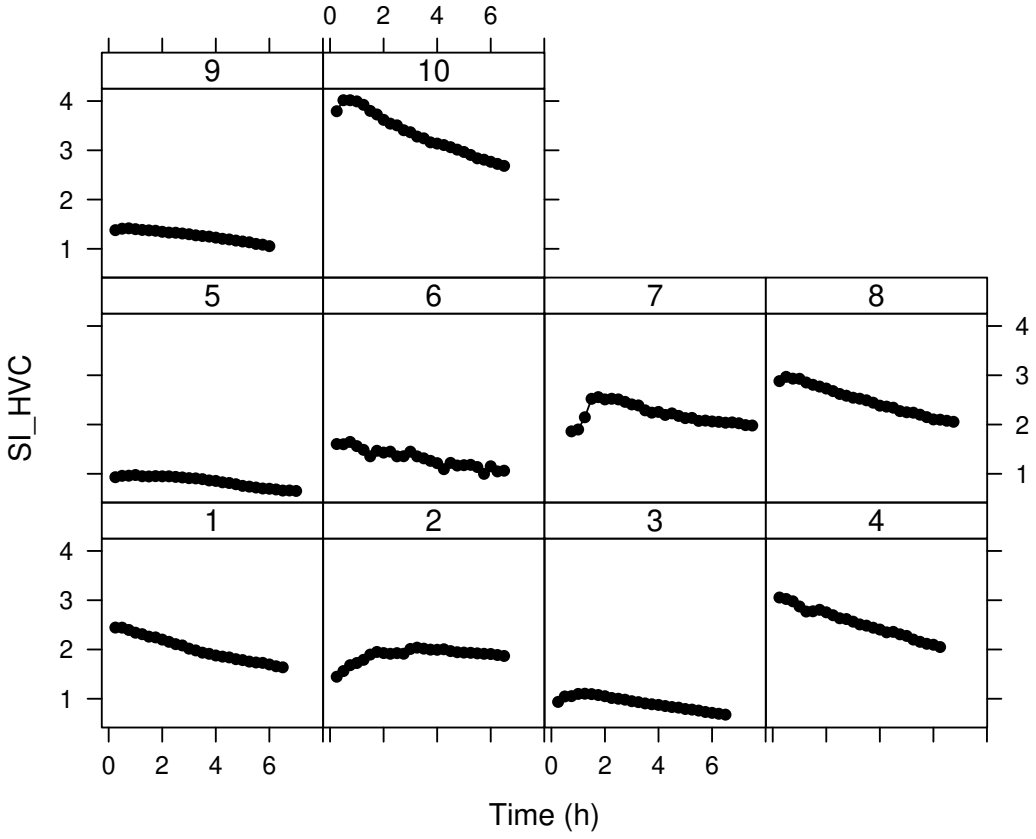


Figure 5 – Songbird data for SI in HVC for each individual bird.

CHAPTER

3

PRELIMINARIES

In this chapter, some preliminary definitions about scale mixture of normal distributions are presented, as well as some basic properties and particular distributions.

3.1 Scale mixture of Normal distributions

In this section, we present a subclass of elliptical contoured distributions, the scale mixture of normal (SMN) distributions. This subclass of distribution includes heavy-tailed multivariate distributions, for instance, normal, t, slash and normal contaminated are presented by Lange and Sinsheimer (1993), where the ideas of stochastic representation were investigated in Andrews and Mallows (1974).

Let $\mathbf{Y}$ be a m-dimensional random vector with a scale mixture of normal distribution. It has a stochastic representation given by

$$\mathbf{Y} = \mu + U^{-1/2}\mathbf{Z}, \tag{3.1}$$

where $\mathbf{Z}$ has a multivariate normal distribution with mean vector 0 and variance-covariance matrix Σ and U is a positive random variable independent of $\mathbf{Z}$. We denote the mixing cumulative distribution function of U by $H(u;\nu)$, satisfying $H(0)=0$, where ν is a scalar or vector valued parameter. Note that the conditional distribution of $\mathbf{Y}$ given $U=u$ is normal with mean vector μ and covariance matrix $u^{-1}\Sigma$. The probability density function (pdf) of $\mathbf{Y}$ is given by

$$\begin{aligned} f(\mathbf{y}) &= \int_0^\infty \phi_m(\mathbf{y}|\mu, u^{-1}\Sigma)\mathrm{d}H(u;\nu) \\ &= \frac{1}{(2\pi)^{m/2}|\Sigma|^{1/2}} \int_0^\infty u^{m/2} \exp\left\{-\frac{1}{2}u(\mathbf{y}-\mu)^\top\Sigma^{-1}(\mathbf{y}-\mu)\right\}\mathrm{d}H(u;\nu), \end{aligned} \tag{3.2}$$

where $\phi_m(\cdot|\mu,\Sigma)$ denotes an m-dimensional normal probability density function with mean vector μ and covariance matrix Σ. For a random vector with distribution in (3.2), we

say that $\mathbf{Y} \sim \text{SMN}_m(\mu, \Sigma; H)$. The form of the SMN_m distribution is determined by the distribution of U and $H(u;\nu)$ is called the mixture distribution. However when U is assigned 1 we have the normal distribution.

Using iterated expectations, we calculate the expectation and the covariance of $\mathbf{Y}$, if they exist, as respectively:

(i) $\text{E}(\mathbf{Y}) = \text{E}[\text{E}(\mathbf{Y}|U)] = \mu$;

(ii) $\text{Cov}(\mathbf{Y}) = \text{E}[\text{Cov}(\mathbf{Y}|U)] + \text{Cov}[\text{E}(\mathbf{Y}|U)] = \text{E}(U^{-1})\Sigma$.

The SMN subclass has been used as an important tool in robust estimation. Among these subclasses we can cite the normal, slash and t distribution, described as follows:

3.1.1 Multivariate t distribution

Consider that U has a gamma distribution with shape and rate parameters $\nu/2$ and $\nu/2$, respectively with pdf

$$h(u;\nu) = \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} u^{(\nu/2)-1} \exp\left\{-\frac{\nu}{2}u\right\},\ u > 0,\ \nu > 0.$$

Now, replacing the gamma distribution in equation (3.2) the multivariate t distribution is recovered with ν degree of freedom with pdf

$$f(\mathbf{y}) = \frac{\Gamma(\frac{m+\nu}{2})}{\Gamma\left(\frac{\nu}{2}\right)\pi^{m/2}} \nu^{-m/2} |\Sigma|^{-1/2} \left(1 + \frac{d}{\nu}\right)^{-\frac{m+\nu}{2}},\ \mathbf{y} \in \mathbb{R}^m, \tag{3.3}$$

$$\tag{3.4}$$

where $d = (\mathbf{y}-\mu)^\top \Sigma^{-1}(\mathbf{y}-\mu)$ is the Mahalanobis distance. Later we will see this distance as a very useful tool for evaluating possible outliers. The limiting case, as ν goes to infinity, the distribution becomes a multivariate normal.

3.1.2 Multivariate slash distribution

The multivariate slash distribution can be generated from equation (3.2), assuming to the U variable a Beta distribution with parameter ν and parameter one, $\text{Beta}(\nu, 1)$ with a probability density function given by

$$h(u) \quad = \quad \nu u^{\nu-1},\ 0 \le u \le 1,\ \nu > 0. \tag{3.5}$$

From this, the density of the slash is given by

$$f(\mathbf{y}) \quad = \quad |2\pi\Sigma|^{-1/2} \nu \int_0^1 x^{\nu+m/2-1} e^{-xd/2} \mathrm{d}x \tag{3.6}$$

$$= \quad \begin{cases} \frac{\nu|\Sigma|^{-1/2} 2^{\nu+m/2}}{(2\pi)^{m/2} d^{\nu+m/2}} \gamma(\nu+m/2;\ d/2), & \mathbf{y} \neq \mu, \\ \frac{|\Sigma|^{-1/2} 2\nu}{(2\pi)^{m/2}(2\nu+m)}, & \mathbf{y} = \mu, \end{cases} \tag{3.7}$$

where $\gamma(b;c) = \int_0^c x^{b-1}e^{-x}dx$ is the incomplete gamma function and $d = (\mathbf{y}-\mu)^T\Sigma^{-1}(\mathbf{y}-\mu)$. We can say that $Y \sim \text{slash}(\mu,\Sigma,\nu)$, representing the location parameter $\mu \in \mathbb{R}^m$, positive definite scatter matrix parameter Σ and degree of freedom ν. Here, the limiting case, as ν goes to infinity, the distribution becomes a multivariate normal.

3.2 Graphical representation of distributions

Figure 6 shows the effect of degrees of freedom on the t and slash distribution tails compared to the normal distribution. Note that the gain from working with t and slash distributions is on the tails of these distributions, with heavier tails compared to the normal distribution.

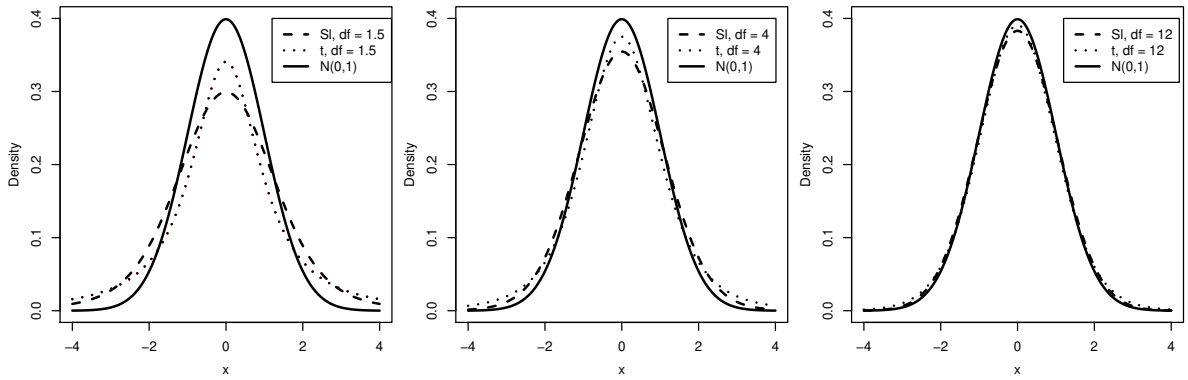


Figure 6 – Univariate representations of slash, t distributions with degrees of freedom 1.5, 4 and 12 and the standard normal distribution.

We provided in Figure 7 the density contours of the bivariate standard normal, slash and t-distributions with $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\Sigma = \begin{pmatrix} 0.5 & 0.3 \\ 0.3 & 1 \end{pmatrix}$. Both t and slash distributions with $df = 4$ degrees of freedom.

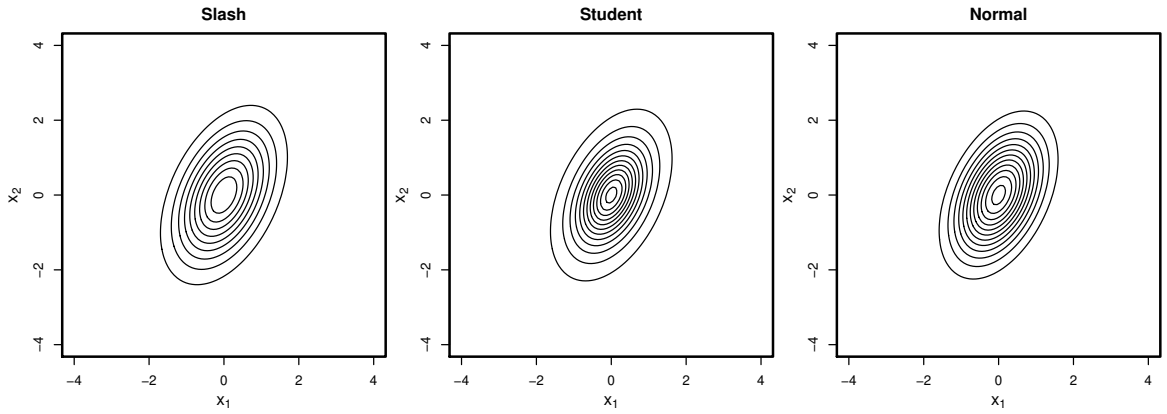


Figure 7 – Panels show the contour plot of the bivariate standard normal distribution, the contour plot of the bivariate t-distribution and the contour plot of the bivariate slash distribution. Both t and slash distributions with μ, Σ and $df = 4$ degrees of freedom.

3.3 Selection of models

The selection of models in mixed-effects models is usually performed by the Akaike information criterion (AIC) from Akaike (1973) and Bayesian information criterion (BIC) from Schwarz (1978), respectively

$$\text{AIC} = -2\log(L) + 2p \tag{3.8}$$

and

$$\text{BIC} = -2\log(L) + \log(n)p, \tag{3.9}$$

where L is the likelihood function, p is the total number of parameters to be estimated and n the number of subjects. The best choice of the model is given by the lowest value of AIC or BIC. It is also important to note that the BIC model uses the number of individuals in its formula, and not the total observations of all individuals. This is due to a modification by Raftery (1995).

CHAPTER

4

EXACT AND APPROXIMATE ESTIMATION METHODS IN NLME MODELS

In this chapter, a comparison of exact and approximate methods for estimating parameters in nonlinear mixed-effects models is revisited. Initially proposed in Russo (2010) and presented in Russo *et al.* (2018), a scale mixture of normal distributions is considered for nonlinear mixed-effects models. An exact likelihood-based method is developed and an approximated method, as usually considered in the literature, is considered for comparison.

4.1 Robust nonlinear mixed-effects models

Suppose that $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$ is a vector of observed continuous multivariate responses with $\mathbf{y}_i$ a $(n_i \times 1)$ vector containing the observations for the experimental unit i, $i = 1, \ldots, n$, such that

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

in which $\mathbf{X}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{in_i})^\top$ is a matrix of explanatory variables for the i-th unit, $\mathbf{b}_i$ is a $(q \times 1)$ vector of random effects, $\boldsymbol{\varepsilon}_i$ is an $(n_i \times 1)$ vector of random errors values for $i = 1, \ldots, n$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a $(p \times 1)$ location vector and $\mathbf{A}_i$ is a full rank $(q \times p)$ matrix of known constants. This nonlinear model was considered by Lee and Xu (2004), for instance, under normality. In this work, we assume that

$$\begin{pmatrix} \boldsymbol{\varepsilon}_i \\ \mathbf{b}_i \end{pmatrix} \overset{ind.}{\sim} \mathrm{S}MN_{n_i+q}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \Sigma_i & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{pmatrix}; H\right), \tag{4.2}$$

where $\mathbf{D}$ and Σ_i are positive-definite dispersion matrices. We assume that $\mathbf{D} = \mathbf{D}(\boldsymbol{\tau}) = \mathrm{diag}(\boldsymbol{\tau})$ is a diagonal matrix and denote its elements by $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_q)^\top$. The matrix Σ_i with dimension $(n_i \times n_i)$ is typically dependent upon i through its dimension, and

it is considered, for example, $\Sigma_i = \sigma^2 \mathbf{I}_{n_i}$ for $i = 1, \ldots, n$ and $\sigma > 0$ a scalar. Since $\mathbf{A}_i$ and $\mathbf{X}_i$ are known matrices, we simplify the notation by writing $\mathbf{g}(\boldsymbol{\beta}, \mathbf{b}_i)$ to represent $\mathbf{g}(\boldsymbol{\varphi}_i, \mathbf{X}_i) = \mathbf{g}(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{b}_i, \mathbf{X}_i)$. Finally, as it was indicated in the previous section, $H = H(\cdot, \nu)$ is the cumulative distribution function (cdf) generator that determines the specific SMN model that was assumed.

Table 1 shows characterizations based on the scale mixture of multivariate normal distributions for slash (MSl) and t-distribution (Mt).

Table 1 – Characterization of some SMN distributions.

Distribution	$\kappa(u)$	U	Density function $f(\mathbf{y})$
$\mathrm{Mt}_m(\mu, \Sigma, \nu)$	$\frac{1}{u}$	$\mathrm{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$, $u > 0, \nu > 0$	$\frac{\Gamma(\frac{m+\nu}{2})}{\Gamma\left(\frac{\nu}{2}\right)\pi^{m/2}} \nu^{-m/2} \lvert\Sigma\rvert^{-1/2} \left(1 + \frac{d}{\nu}\right)^{-\frac{m+\nu}{2}}$
$\mathrm{MSl}_m(\mu, \Sigma, \nu)$	$\frac{1}{u}$	$\mathrm{Beta}(\nu, 1)$, $0 < u < 1, \nu > 0$	$\nu \int_0^1 u^{\nu-1} \phi_m(\mathbf{y} \mid \mu, u^{-1}\Sigma) du$

with $d = (\mathbf{y} - \mu)^\top \Sigma^{-1} (\mathbf{y} - \mu)$

`Remarks:`

(i) From (4.2) , it follows that marginally

$$\boldsymbol{\varepsilon}_i \overset{ind.}{\sim} \mathrm{SMN}_{n_i}(\mathbf{0}, \Sigma_i; H) \qquad \text{and} \qquad \mathbf{b}_i \overset{iid.}{\sim} \mathrm{SMN}_q(\mathbf{0}, \mathbf{D}; H) \;\; i = 1, \ldots, n. \tag{4.3}$$

(ii) Since for each $i = 1, \ldots, n$, $\mathbf{b}_i$ and $\boldsymbol{\varepsilon}_i$ are indexed by the same scale mixing factor U_i, they are not independent in general. The independence corresponds to the case where $U_i = 1, i = 1, \ldots, n$, so that the SMN–NLME model reduces to the normal NLME model as defined in Walker (1996). However, conditional on U_i, $\mathbf{b}_i$ and $\boldsymbol{\varepsilon}_i$ are independent for each $i = 1, \ldots, n$, which implies that $\mathbf{b}_i$ and $\boldsymbol{\varepsilon}_i$ are not correlated, once $\mathrm{Cov}(\mathbf{b}_i, \boldsymbol{\varepsilon}_i) = \mathrm{E}[\mathbf{b}_i \boldsymbol{\varepsilon}_i^\top] = \mathrm{E}_{U_i}[\mathrm{E}[\mathbf{b}_i \boldsymbol{\varepsilon}_i^\top | U_i]] = \mathbf{0}$. Therefore, an attractive and convenient way to specify (4.1) and (4.2) is the following hierarchical representation:

$$\mathbf{b}_i | U_i = u_i \overset{ind.}{\sim} \mathrm{N}_q(\mathbf{0}, \kappa(u_i)\mathbf{D}) \text{ and } \boldsymbol{\varepsilon}_i | U_i = u_i \overset{ind.}{\sim} \mathrm{N}_{n_i}(\mathbf{0}, \kappa(u_i)\Sigma_i), i = 1, \ldots, n,$$

$\mathbf{b}_i | U_i = u_i$ and $\boldsymbol{\varepsilon}_i | U_i = u_i$ are independent, where $U_i \overset{iid}{\sim} h(u_i, \nu)$, and $\kappa(\cdot)$ is the weight function, $i = 1, \ldots, n$.

(iii) Aiming to choose between the different fitted models, we use the Akaike information criterion (AIC) and Bayesian information criterion (BIC), which also provide an alternative to select parameter ν from the scale mixture of normal distributions. To obtain an approximation for the log-likelihood, the importance sampling method is considered, following the suggestion of Meza, Osorio and De La Cruz (2012). Further discussion about fixing or estimating the extra parameters may be found in

the literature (for more details, see Lucas (1997)), taking into account the possible sensitivity added by unbounded behaviour of the influence and change-of-variance functions of the location parameter. However, the methodology proposed here could be easily adapted for the case in which ν is estimated.

(iv) The hierarchical representation (three–stage) to the NLMEM defined in (4.1) and (4.2) is given by

$$\mathbf{y}_i|\mathbf{b}_i, U_i = u_i \overset{ind.}{\sim} \mathrm{N}_{n_i}(\mathbf{g}(\boldsymbol{\beta}, \mathbf{b}_i), \kappa(u_i)\Sigma_i), \tag{4.4}$$

$$\mathbf{b}_i|U_i = u_i \overset{ind.}{\sim} \mathrm{N}_q(0, \kappa(u_i)\mathbf{D}), \quad \text{and} \tag{4.5}$$

$$U_i \overset{iid.}{\sim} h(u_i, \nu). \tag{4.6}$$

Classical inference of the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\tau}^\top, \sigma^2)^\top$, is based on the marginal distribution of $\mathbf{y}_i$, particularly by the frequentist approach, and the maximum likelihood estimates (ML) of the parameters can be obtained from the joint distribution

$$f(\mathbf{y}_1, \ldots, \mathbf{y}_n) = \prod_{i=1}^{n} \int \int \phi_{n_i}(\mathbf{y}_i|\mathbf{g}(\boldsymbol{\beta}, \mathbf{b}_i), \kappa(u_i)\Sigma_i)\phi_q(\mathbf{b}_i|0, \kappa(u_i)\mathbf{D})\mathrm{d}\mathbf{b}_i \mathrm{d}H(u_i, \nu), \tag{4.7}$$

which generally does not have a closed form expression because the model function is not linear in the random effect. In the next section, we propose a Monte Carlo EM algorithm that facilitates the likelihood inference and also an approximate method based on iterative approximations to the linear mixed-effects model.

4.2 Maximum likelihood estimation

4.2.1 Monte Carlo EM method

Initially proposed by Dempster, Laird and Rubin (1977), the EM algorithm represents an efficient tool to obtain the maximum likelihood estimates in problems with incomplete data, for instance. By augmenting the observed data with a "missing" quantity, it is worth using this iterative procedure when the maximization of the complete data likelihood is easier than the original data. Basically, the process consists of repeating the steps of expectation and maximization of the complete data likelihood until the convergence is achieved. These steps are known in the literature as the E-step and M-step, respectively.

In the $(t+1)$th iteration, the E-Step usually consists of calculating the expectation of the log-likelihood of the full data $\ell_c(\boldsymbol{\theta})$ conditional on the observed data $\mathbf{y}$ in the current estimate of the parameters $\boldsymbol{\theta}^{(t)}$,

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = E[\ell_c(\boldsymbol{\theta})|\mathbf{y}; \boldsymbol{\theta}^{(t)}].$$

The M-step consists of finding $\theta^{(t+1)}$ which maximizes $Q(\theta|\theta^{(t)})$, such that

$$Q(\theta^{(t+1)}|\theta^{(t)}) \geq Q(\theta|\theta^{(t)}).$$

Since it is not always straightforward to obtain the E-step expressions explicitly, additional tools may be required to estimate the expected values numerically. In particular, the Monte Carlo EM algorithm (MCEM) has been used for mixed-effects models by treating the random effects as latent data (see, for instance, Wei and Tanner (1990) and Walker (1996)) and is used here to obtain the maximum likelihood estimates for the NLMEM parameters. Modified versions of the EM algorithm for NLMEM are presented in the literature, as for instance the SEM or SAEM algorithms (Kuhn and Lavielle (2005)). However, previous papers have shown that the results considering these different methods are very close to each other, therefore in this work we only consider the MCEM method.

Considering the model defined in (4.1-4.2), the unobserved random effect $\mathbf{b}_i$ and the scale factor U_i are considered as missing data, so that the "complete data" is given by $\{(\mathbf{y}_i, \mathbf{b}_i, u_i),\ i = 1, \ldots, n\}$. Therefore, the complete–data log-likelihood for all individuals can be written as

$$\begin{aligned} \ell_c(\theta) &= \sum_{i=1}^{n} \ell(\theta; \mathbf{y}_i, \mathbf{b}_i, u_i) \\ &= \sum_{i=1}^{n} \{\log f(\mathbf{y}_i|\mathbf{b}_i, u_i, \boldsymbol{\beta}, \sigma^2) + \log f(\mathbf{b}_i|u_i, \tau) + \log h(u_i, \nu)\}, \end{aligned}$$

where τ is the parameter vector of the scale matrix $\mathbf{D}$ and $\Sigma_i = \sigma^2 \mathbf{I}_{n_i}$.

Let $\theta^{(t)}$ be the parameter estimates from the t-th EM iteration. The Expectation step ("E step") from individual i at the $(t+1)$th iteration can be written as

$$\begin{aligned} Q_i(\theta|\theta^{(t)}) &= E[\ell(\theta; \mathbf{y}_i, \mathbf{b}_i, u_i)|\mathbf{y}_i, \theta^{(t)}] \\ &= \int\int \Big\{ \log \phi_{n_i}(\mathbf{y}_i|\mathbf{g}(\boldsymbol{\beta}, \mathbf{b}_i), \kappa(u_i)\sigma^2 \mathbf{I}_{n_i}) + \log \phi_q(\mathbf{b}_i|\mathbf{0}, \kappa(u_i)\mathbf{D}) \\ &\quad + \log h(u_i, \nu) \Big\} \times f(u_i, \mathbf{b}_i|y_i, \theta^{(t)}) du_i d\mathbf{b}_i. \end{aligned} \tag{4.8}$$

It is well known that the foregoing integral does not have a closed form in general and the evaluation of the integral by numerical quadrature is in general unfeasible except for simple cases. However, note that expression (4.8) is an expectation with respect to $f(u_i, \mathbf{b}_i|y_i, \theta^{(t)})$, and therefore it may be evaluated by using the Monte Carlo EM algorithm by Wei and Tanner (1990), as discussed by Wu (2004). Specifically, we may use the Gibbs sampler with a Metropolis-Hastings step (Gilks, Richardson and Spiegelhalter (1996)) to generate samples from $[u_i, \mathbf{b}_i|\mathbf{y}_i, \theta^{(t)}]$ by sampling from the full conditionals $[u_i|\mathbf{b}_i, \mathbf{y}_i, \theta^{(t)}]$

Table 2 – Full conditional distributions of $(u_i|\mathbf{b}_i,\mathbf{y}_i,\boldsymbol{\theta})$.

Distribution	$f(u_i\|\mathbf{b}_i,\mathbf{y}_i,\boldsymbol{\theta})$
$\text{Mt}_m(\boldsymbol{\mu},\Sigma,\nu)$	Gamma$(\frac{\nu+n_i+q}{2},\frac{\nu+G}{2})$, with $u_i>0$ and $\nu>0$
$\text{MSl}_m(\boldsymbol{\mu},\Sigma,\nu)$	TGamma$(\nu+\frac{n_i+q}{2},\frac{G}{2},1)$, with $0<u_i<1$ and $\nu>0$

with $G=\left[\mathbf{b}_i^\top\mathbf{D}^{-1}\mathbf{b}_i+\dfrac{||\mathbf{y}_i-\mathbf{g}(\boldsymbol{\beta},\mathbf{b}_i)||^2}{\sigma^2}\right]$

and $[\mathbf{b}_i|u_i,\mathbf{y}_i,\boldsymbol{\theta}^{(t)}]$. Note that

$$\begin{aligned} f(u_i|\mathbf{b}_i,\mathbf{y}_i,\boldsymbol{\theta}^{(t)}) &\propto h(u_i,\nu)\phi_{n_i}(\mathbf{y}_i|\mathbf{g}(\boldsymbol{\beta},\mathbf{b}_i),\kappa(u_i)\sigma^2\mathbf{I}_{n_i})\phi_q(\mathbf{b}_i|0,\kappa(u_i)\mathbf{D}) \\ &\propto h(u_i,\nu)\kappa^{-(n_i+q)/2}(u_i)\exp\left\{-\frac{1}{2}\kappa^{-1}(u_i)\left[\mathbf{b}_i^\top\mathbf{D}^{-1}\mathbf{b}_i\right.\right. \\ &\quad \left.\left.+\frac{||\mathbf{y}_i-\mathbf{g}(\boldsymbol{\beta},\mathbf{b}_i)||^2}{\sigma^2}\right]\right\} \end{aligned} \tag{4.9}$$

and

$$f(\mathbf{b}_i|u_i,\mathbf{y}_i,\boldsymbol{\theta}^{(t)}) \propto \phi_{n_i}(\mathbf{y}_i|\mathbf{g}(\boldsymbol{\beta},\mathbf{b}_i),\kappa(u_i)\sigma^2\mathbf{I}_{n_i})\phi_q(\mathbf{b}_i|0,\kappa(u_i)\mathbf{D}). \tag{4.10}$$

Monte Carlo samples can be generated from these full conditionals by using rejection sampling methods, in this case the Gibbs sampler is considered (see, for instance, Gelfand *et al.* (1990)). It can be seen that the full conditional distribution of u_i under the t model follows a gamma distribution and under the slash case it follows a truncated gamma (TGamma) distribution (see Table 2), in which the notation tgamma(a,b,t) represents a random variable with gamma(a,b) distribution with right truncation at the value t. Meanwhile, the full conditional distribution of $\mathbf{b}_i$ does not have a closed form as it appears inside the nonlinear function $\mathbf{g}(\boldsymbol{\beta},\mathbf{b}_i)$, and one alternative would be to implement the Metropolis-Hastings algorithm (see, for instance, Gilks, Richardson and Spiegelhalter (1996)) to obtain samples from $\mathbf{b}_i$. Notice that if $\mathbf{g}(\boldsymbol{\beta},\mathbf{b}_i)$ is linear with respect to $\mathbf{b}_i$, for instance $\boldsymbol{\varphi}_i=\mathbf{A}_i\boldsymbol{\beta}+\mathbf{b}_i$, then depending on the distribution of u_i the full conditional distribution of $\mathbf{b}_i$ could also be written in a closed form expression.

For individual i, let $\{(\mathbf{b}_i^{(1)},u_i^{(1)}),\ldots,(\mathbf{b}_i^{(M)},u_i^{(M)})\}$ denote a random sample of size M generated from $[u_i,\mathbf{b}_i|\mathbf{y}_i,\boldsymbol{\theta}^{(t)}]$ then the E step at the $(t+1)$th EM iteration can be written as

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) &= \sum_{i=1}^n Q_i(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})=\sum_{i=1}^n\left[\frac{1}{M}\sum_{j=1}^M \ell(\boldsymbol{\theta};\mathbf{y}_i,\mathbf{b}_i^{(j)},u_i^{(j)})\right] \\ &\propto \sum_{i=1}^n\sum_{j=1}^M\frac{1}{M}\left[-\frac{n_i}{2}\log\sigma^2-\frac{\kappa^{-1}(u_i^{(j)})}{2\sigma^2}||\mathbf{y}_i-\mathbf{g}(\boldsymbol{\beta},\mathbf{b}_i^{(j)})||^2\right] \\ &\quad +\sum_{i=1}^n\sum_{j=1}^M\frac{1}{M}\left[-\frac{1}{2}\log|\mathbf{D}|-\frac{\kappa^{-1}(u_i^{(j)})}{2}\mathbf{b}_i^{(j)\top}\mathbf{D}^{-1}\mathbf{b}_i^{(j)}\right]. \end{aligned}$$

The Maximization step ("M step") of the Monte Carlo EM algorithm (MCEM) maximizes $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ to produce an updated estimate $\boldsymbol{\theta}^{(t+1)}$, and therefore it is like a complete-data maximization. From $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ it can be easily seen that the unique solution is given by

$$\widehat{\sigma}^{2(t+1)} = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{M}\frac{1}{M}\left[\kappa^{-1}(u_i^{(j)})||\mathbf{y}_i - \mathbf{g}(\boldsymbol{\beta}^{(t)}, \mathbf{b}_i^{(j)})||^2\right], \; N = \sum_{i=1}^{n} n_i, \tag{4.11}$$

$$\widehat{\mathbf{D}}^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{M}\frac{1}{M}\left[\kappa^{-1}(u_i^{(j)})\mathrm{diag}(\mathbf{b}_i^{(j)}\mathbf{b}_i^{(j)\top})\right], \tag{4.12}$$

To update $\boldsymbol{\beta}$, Newton-Raphson algorithm iterations are used, which are given by

$$\widehat{\boldsymbol{\beta}}^{(t+1)} = \widehat{\boldsymbol{\beta}}^{(t)} + \left(\sum_{i=1}^{n}\sum_{j=1}^{M}\frac{\mathbf{J}_i^{(t)\top}\mathbf{J}_i^{(t)}}{\kappa(u_i^{(j)})}\right)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{M}\frac{\mathbf{J}_i^{(t)\top}[\mathbf{y}_i - \mathbf{g}(\boldsymbol{\beta}^{(t)}, \mathbf{b}_i^{(j)})]}{\kappa(u_i^{(j)})} \tag{4.13}$$

where $\mathbf{J}_i = \partial\mathbf{g}(\boldsymbol{\beta}, \mathbf{b}_i)/\partial\boldsymbol{\beta}^{\top}$.

In summary, the MCEM algorithm takes the following steps:

Algorithm 1 Monte Carlo EM for the obtention of maximum likelihood estimates.

1: Start with initial values $M = M_0$, $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, $\tau = \tau_0$ and $\sigma^2 = \sigma_0^2$;
2: Sample $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(m)}$ and $\mathbf{b}^{(1)}, \ldots, \mathbf{b}^{(m)}$ from (4.9) and (4.10), respectively;
3: M step:
   a. Compute $\hat{\sigma}^{2(t+1)}$ and $\hat{\mathbf{D}}^{(t+1)}$ according to (4.11) and (4.12), respectively;
   b. Obtain $\hat{\boldsymbol{\beta}}^{(t+1)}$ according to (4.13);
4: Repeat steps 2 and 3 until convergence.

We used the stopping criterion described in Wang (2007) to diagnose convergence. In particular, the criterion for the algorithm is given by

$$\max_i\left(\frac{|\theta_i^{(t+1)} - \theta_i^{(t)}|}{|\theta_i^{(t)}|}\right) < \delta,$$

where δ is a small fixed constant.

4.2.2 An approximate method to maximum likelihood estimation

In this section, we discuss using an approximate method proposed by Wu (2004) in the normal case, which represents an alternative to the MCEM method and avoid some challenges found in the MCEM approach, such as the difficulty of convergence, for instance. As stated by Wu (2004), the approximate method may involve less computational effort than the MCEM method, specially when the dimension of the random effects vector is

high. The main advantages of the approximate method are that sampling the random effects in the E step is not necessary and the explicit M step expressions may be obtained trivially. On the other hand, the approximation of the model may also provide additional errors to the problem.

The most frequently approximate methods found in the literature are based on Taylor expansions of the nonlinear function or Laplace approximations. Here we consider a method which is similar to the one used by Wolfinger (1993) and consists of iteratively solving the linear mixed-effects (LME) model and proceeds in the standard way to estimate the parameters.

First, we rewrite the SMN–NLME model (4.1) and (4.2) as a single equation by combining the first two stages

$$y_{ij} = g_{ij}(\boldsymbol{\beta}, \mathbf{b}_i) + \varepsilon_{ij} \text{ for } i = 1, \ldots, n \text{ and } j = 1, \ldots, n_i.$$

Denote the current estimates of $(\boldsymbol{\beta}, \mathbf{b}_i)$ by $(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{b}}_i)$. Taking the first-order Taylor expansion of g_{ij} around the current parameter estimate $\widehat{\boldsymbol{\beta}}$ and the random effect estimates $\widehat{\mathbf{b}}_i$, the approximate method consists of iteratively solving the LME response model

$$\widetilde{\mathbf{y}}_i = \mathbf{W}_i\boldsymbol{\beta} + \mathbf{T}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \tag{4.14}$$

where $\widetilde{\mathbf{y}}_i = \mathbf{y}_i - \mathbf{g}_i(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{b}}_i) + \mathbf{W}_i\widehat{\boldsymbol{\beta}} + \mathbf{T}_i\widehat{\mathbf{b}}_i$ with $\mathbf{g}_i = (g_{i1}, \ldots, g_{in_i})^\top$, $\mathbf{W}_i = (\mathbf{W}_{i1}^\top, \ldots, \mathbf{W}_{in_i}^\top)^\top$, $\mathbf{T}_i = (\mathbf{T}_{i1}^\top, \ldots, \mathbf{T}_{in_i}^\top)^\top$ and $\widetilde{\mathbf{y}}_i = (\widetilde{y}_{i1}, \ldots, \widetilde{y}_{in_i})^\top$, in which

$$\mathbf{W}_{ij} = \left.\frac{\partial g_{ij}(\boldsymbol{\beta}, \widehat{\mathbf{b}}_i)}{\partial \boldsymbol{\beta}^\top}\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} \text{ and } \mathbf{T}_{ij} = \left.\frac{\partial g_{ij}(\widehat{\boldsymbol{\beta}}, \mathbf{b}_i)}{\partial \mathbf{b}_i^\top}\right|_{\mathbf{b}_i=\widehat{\mathbf{b}}_i} \text{ for } i = 1, \ldots, n, \text{ and } j = 1, \ldots, n_i.$$

Note that the dimensions of $\mathbf{W}_i$ and $\mathbf{T}_i$ are $(n_i \times p)$ and $(n_i \times q)$, respectively.

Now we combine the LME response model (4.14) with (4.2). Therefore, by standard arguments of scale mixture of normal distributions and matrix algebra, it is not difficult to see that

$$\mathbf{b}_i | \widetilde{\mathbf{y}}_i, \widehat{\theta}, u_i \sim \mathrm{N}_q(\widetilde{\mathbf{b}}_i, \kappa(u_i)\widetilde{\Sigma}_i), \tag{4.15}$$

with $\widetilde{\Sigma}_i = (\widehat{\mathbf{D}}^{-1} + \widehat{\sigma}^{-2}\mathbf{T}_i{}^\top\mathbf{T}_i)^{-1}$ and $\widetilde{\mathbf{b}}_i = \widehat{\sigma}^{-2}\widetilde{\Sigma}_i\mathbf{T}_i^\top(\widetilde{\mathbf{y}}_i - \mathbf{W}_i\widehat{\boldsymbol{\beta}})$.

After some algebra, we can then integrate out $\mathbf{b}_i$ and u_i from (4.8) and obtain the following E step

$$Q(\theta|\theta^{(t)}) = \sum_{i=1}^{n} Q_i(\theta|\theta^{(t)}),$$

where

$$\begin{aligned} Q_i(\theta|\theta^{(t)}) &= E[\ell(\theta; \widetilde{\mathbf{y}}_i, \mathbf{b}_i, u_i)|\widetilde{\mathbf{y}}_i, \theta^{(t)}] \\ &\propto -\frac{n_i}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\left[\widetilde{u}_i(\widetilde{\mathbf{y}}_i - \mathbf{W}_i\boldsymbol{\beta} - \mathbf{T}_i\widetilde{\mathbf{b}}_i)^\top(\widetilde{\mathbf{y}}_i - \mathbf{W}_i\boldsymbol{\beta} - \mathbf{T}_i\widetilde{\mathbf{b}}_i) + \mathrm{tr}(\widetilde{\Sigma}_i\mathbf{T}_i{}^\top\mathbf{T}_i)\right] \\ &\quad -\frac{1}{2}\log|\mathbf{D}| - \frac{1}{2}\mathrm{tr}\left((\widetilde{\Sigma}_i + \widetilde{u}_i\widetilde{\mathbf{b}}_i\widetilde{\mathbf{b}}_i^\top)\mathbf{D}^{-1}\right), \end{aligned}$$

with $\widetilde{u}_i = E[\kappa^{-1}(u_i)|\widetilde{\mathbf{y}}_i, \boldsymbol{\theta}^{(t)}]$.

Finally, we can update the parameter estimates as follows:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= \left(\sum_{i=1}^{n} \widetilde{u}_i \mathbf{W}_i^\top \mathbf{W}_i\right)^{-1} \left[\sum_{i=1}^{n} \widetilde{u}_i \mathbf{W}_i^\top (\widetilde{y}_i - \mathbf{T}_i \widetilde{\mathbf{b}}_i)\right], \\
\widehat{\sigma}^2 &= \frac{1}{N}\sum_{i=1}^{n} \left[\widetilde{u}_i (\widetilde{\mathbf{y}}_i - \mathbf{W}_i\widehat{\boldsymbol{\beta}} - \mathbf{T}_i\widetilde{\mathbf{b}}_i)^\top (\widetilde{\mathbf{y}}_i - \mathbf{W}_i\widehat{\boldsymbol{\beta}} - \mathbf{T}_i\widetilde{\mathbf{b}}_i) + \mathrm{tr}(\widetilde{\Sigma}_i \mathbf{T}_i{}^\top \mathbf{T}_i)\right], \\
\widehat{\mathbf{D}} &= \frac{1}{n}\sum_{i=1}^{n} \left(\widetilde{\Sigma}_i + \widetilde{u}_i \widetilde{\mathbf{b}}_i \widetilde{\mathbf{b}}_i^\top\right).
\end{aligned}$$

It can be shown that $\widetilde{\mathbf{y}}_i \sim \mathrm{S}MN_{n_i}(\mathbf{W}_i\boldsymbol{\beta}, \mathbf{T}_i\mathbf{D}\mathbf{T}_i^\top + \sigma^2\mathbf{I}_{n_i}; H)$ and the values of $\widetilde{u}_i = E[\kappa^{-1}(u_i)|\widetilde{\mathbf{y}}_i, \boldsymbol{\theta}^{(t)}]$ for some distributions are given in Table 3. It is important to note in the same table that the maximum likelihood estimates bring a type of robustness to the model as the iterative processes encompass terms to control the influence of large Mahalanobis distances. Further discussion about the robustness in heavy-tailed mixed-effects models can be found in Paula, Medeiros and Vilca-Labra (2009), Osorio, Paula and Galea (2007) and Russo, Aoki and Paula (2012) for example.

Table 3 – Characterization of $\widetilde{u}_i = E[\kappa^{-1}(u_i)|\widetilde{\mathbf{y}}_i, \boldsymbol{\theta}^{(t)}]$ for some distributions

Distribution	$\widetilde{u}_i$
$\mathrm{Mt}_m(\boldsymbol{\mu}, \Sigma, \nu)$	$\dfrac{\nu + n_i}{\nu + d_i}$
$\mathrm{MSl}_m(\boldsymbol{\mu}, \Sigma, \nu)$	$\dfrac{2\nu + n_i}{d_i} \dfrac{P_1(n_i/2 + \nu + 1, d_i/2)}{P_1(n_i/2 + \nu, d_i/2)}$

with $d_i = (\mathbf{y}_i - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{y}_i - \boldsymbol{\mu})$ and $P_x(a,b) = \dfrac{b^a}{\Gamma(a)} \displaystyle\int_0^x s^{a-1} e^{-bs} ds$ is the cdf of a random variable with distribution gamma(a,b).

4.3 Standard error estimates

Following the expressions developed in Louis (1982) and used by Tan, Tian and Fang (2007), the observed information matrix may be written as

$$-\mathrm{E}\left\{\frac{\partial^2 \ell(\boldsymbol{\theta}|\mathbf{y}, \mathbf{b}, \mathbf{u})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right\}\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} - \mathrm{Var}\left\{\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{y}, \mathbf{b}, \mathbf{u})}{\partial\boldsymbol{\theta}}\right\}\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}, \tag{4.16}$$

in which $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$ and $\mathbf{u} = (u_1, \ldots, u_n)^\top$ and the expectation and variance are computed with respect to $f(u_i, \mathbf{b}_i|\mathbf{y}_i, \boldsymbol{\theta}^{(t)})$. Since the expressions in equation (4.16) may not be easily evaluated analytically, one alternative is to obtain estimates of these quantities by using the samples generated in the Monte Carlo method, and the standard errors are obtained from the square roots of the diagonal elements of the inverse of the estimated information matrix.

In the next section, we present a simulation study to compare the two methodologies.

4.4 Simulation study

A Monte Carlo simulation study was conducted to compare the MCEM and approximate methods considering the normal, t and slash distributions. In each of the scenarios considered, 2000 data sets were generated according to the growth curves model or to the pharmacokinetic model. In general, in order to evaluate parameter estimates, for each parameter we present the sample mean, the bias and mean squared error (MSE), respectively

$$\text{Mean} = \frac{1}{2000}\sum_{l=1}^{2000}\hat{\theta}_l,$$

$$\text{Bias} = \frac{1}{2000}\sum_{l=1}^{2000}\hat{\theta}_l - \theta$$

and

$$\text{MSE}(\hat{\theta}) = \frac{1}{2000}\sum_{l=1}^{2000}(\theta - \hat{\theta}_l)^2.$$

For the fixed-effects parameters, theoretical values were fixed close to the maximum likelihood estimates since they provide an interpretation to the physical phenomenon in the two data sets. For the variance components, we have considered some variations, as follows.

- Growth curves model: the theoretical fixed-effects parameters were taken as $(\beta_1,\beta_2,\beta_3)^\top = (19,55,9)^\top$ for the three scenarios and for the theoretical values for the variance components were

  (1) $(\tau_1,\tau_2,\tau_3,\sigma^2)_1^\top = (16,6,0.1,1)^\top$;

  (2) $(\tau_1,\tau_2,\tau_3,\sigma^2)_2^\top = (10,10,1,1)^\top$; and

  (3) $(\tau_1,\tau_2,\tau_3,\sigma^2)_3^\top = (16,6,0.1,5)^\top$.

- Pharmacokinetic model: for the fixed-effects parameters, theoretical values were taken as $(lK_\text{e}, lK_\text{a}, lC_1)^\top = (-2.5, 0.5, -3)^\top$ for the three scenarios and for the theoretical values for the variance components were

  (1) $(\tau_1,\tau_2,\tau_3,\sigma^2)_1^\top = (0.02,0.05,0.5,0.5)^\top$;

  (2) $(\tau_1,\tau_2,\tau_3,\sigma^2)_2^\top = (0.05,0.05,0.05,3)^\top$; and

  (3) $(\tau_1,\tau_2,\tau_3,\sigma^2)_3^\top = (0.02,0.05,0.5,1)^\top$.

The results presented in Tables 4 and 5 were obtained considering the growth curves model in the first and second scenarios. In all the scenarios, $n = 48$ and the sizes

of the groups were the same as in soybean application. The normal distributions, t with $\nu \in \{4,5,6,7,8\}$ and slash with $\nu \in \{4,5,6,7,8\}$ were considered and we present here the results for normal, t with $\nu = 4$ and slash with $\nu = 4$. No significant differences were observed for the other cases. The obtained results in the third scenario, where $(\tau^\top, \sigma^2)_3^\top = (16,6,0.1,5)^\top$, were similar to the other ones and are omitted here. In general, the methods are equivalent, but in some cases the estimates presented a slightly larger bias under the MCEM method. Regarding the MSE, in most cases the MCEM method also performed worse than the approximate method.

Table 4 – Simulation results for growth curves model with theoretical fixed-effects parameters $(\beta_1,\beta_2,\beta_3)^\top = (19,55,9)^\top$ and theoretical variance components $(\tau^\top, \sigma^2)_1^\top = (16,6,0.1,1)^\top$.

		MCEM			Approximate		
Dist.	Par.	Mean	Bias	MSE	Mean	Bias	MSE
Normal	β_1	19.560	0.560	0.543	19.536	0.536	0.592
	β_2	54.997	-0.003	0.279	54.659	-0.341	0.384
	β_3	9.580	0.580	0.396	9.381	0.381	0.200
	τ_1	15.965	-0.035	13.516	15.624	-0.376	12.963
	τ_2	5.928	-0.072	3.807	5.783	-0.217	3.577
	τ_3	0.252	0.152	0.114	0.213	0.113	0.079
	σ^2	0.858	-0.142	0.025	0.860	-0.140	0.025
t_4	β_1	19.680	0.680	0.723	19.619	0.619	0.733
	β_2	55.018	0.018	0.333	54.620	-0.380	0.465
	β_3	9.647	0.647	0.493	9.409	0.409	0.232
	τ_1	15.952	-0.048	18.575	15.517	-0.483	17.314
	τ_2	6.735	0.735	6.354	6.613	0.613	5.724
	τ_3	0.317	0.217	0.189	0.230	0.130	0.098
	σ^2	0.841	-0.159	0.040	0.846	-0.154	0.039
$Slash_4$	β_1	19.762	0.762	0.892	19.727	0.727	0.929
	β_2	55.003	0.003	0.337	54.557	-0.443	0.514
	β_3	9.736	0.736	0.621	9.476	0.476	0.296
	τ_1	15.812	-0.188	13.702	15.392	-0.608	12.951
	τ_2	6.100	0.100	4.133	5.925	-0.075	3.736
	τ_3	0.277	0.177	0.141	0.223	0.123	0.089
	σ^2	0.844	-0.156	0.031	0.848	-0.152	0.030

Table 5 – Simulation results for growth curves model with theoretical fixed-effects parameters $(\beta_1,\beta_2,\beta_3)^\top = (19,55,9)^\top$ and theoretical variance components $(\tau^\top,\sigma^2)_2^\top = (10,10,1,1)^\top$.

		MCEM			Approximate		
Dist.	Par.	Mean	Bias	MSE	Mean	Bias	MSE
Normal	β_1	19.473	0.473	0.384	19.239	0.239	0.216
	β_2	54.854	-0.146	0.395	54.409	-0.591	0.699
	β_3	9.828	0.828	0.766	9.551	0.551	0.378
	τ_1	9.947	-0.053	5.462	9.907	-0.093	5.315
	τ_2	10.052	0.052	8.265	9.804	-0.196	7.520
	τ_3	1.139	0.139	0.345	0.949	-0.051	0.308
	σ^2	0.858	-0.142	0.026	0.865	-0.135	0.024
t_4	β_1	19.512	0.512	0.439	19.257	0.257	0.263
	β_2	54.863	-0.137	0.433	54.318	-0.682	0.863
	β_3	9.917	0.917	0.951	9.586	0.586	0.437
	τ_1	10.062	0.062	7.456	9.878	-0.122	6.964
	τ_2	11.280	1.280	15.993	11.216	1.216	14.773
	τ_3	1.349	0.349	0.625	0.986	-0.014	0.372
	σ^2	0.846	-0.154	0.039	0.856	-0.144	0.036
$Slash_4$	β_1	19.616	0.616	0.571	19.329	0.329	0.315
	β_2	54.812	-0.188	0.525	54.208	-0.792	1.073
	β_3	10.052	1.052	1.215	9.691	0.691	0.572
	τ_1	9.903	-0.097	5.950	9.817	-0.183	5.703
	τ_2	10.172	0.172	9.618	9.843	-0.157	8.419
	τ_3	1.143	0.143	0.356	0.912	-0.088	0.308
	σ^2	0.850	-0.150	0.030	0.858	-0.142	0.028

The results presented in Tables 6 were obtained considering the pharmacokinetic model and the theoretical values $(\tau^\top,\sigma^2)_1^\top = (0.02,0.05,0.5,0.5)^\top$. In the three scenarios, $n = 12$ and the sizes of the groups are $n_i = 11,\ i = 1,\ldots,12$, the same as in the theophylline application. The obtained results for the other cases were similar to the results in Table 6 and they are omitted here.

As the results of the simulation study, we highlight that the MCEM method produced slightly larger bias in some of the cases but as a general result, this method does not seem to produce important deviations from the results obtained in the approximate method.

Table 6 – Simulation results for the pharmacokinetic model with theoretical fixed-effects parameters $(lK_e, lK_a, lC_l)^\top = (-2.5, 0.5, -3)^\top$ and theoretical variance components $(\tau^\top, \sigma^2)_1^\top = (0.02, 0.05, 0.5, 0.5)^\top$.

		MCEM			Approximate		
Dist.	Par.	Mean	Bias	MSE	Mean	Bias	MSE
Normal	lK_e	-2.524	-0.024	0.006	-2.517	-0.017	0.006
	lK_a	0.506	0.006	0.009	0.499	-0.001	0.009
	lC_l	-3.018	-0.018	0.004	-3.014	-0.014	0.006
	τ_1	0.019	-0.001	0.000	0.018	-0.002	0.000
	τ_2	0.044	-0.006	0.001	0.043	-0.007	0.001
	τ_3	0.321	-0.179	0.048	0.314	-0.186	0.050
	σ^2	0.481	-0.019	0.005	0.482	-0.018	0.005
t_4	lK_e	-2.532	-0.032	0.008	-2.524	-0.024	0.007
	lK_a	0.509	0.009	0.010	0.500	0.000	0.010
	lC_l	-3.029	-0.029	0.006	-3.026	-0.026	0.008
	τ_1	0.020	0.000	0.001	0.019	-0.001	0.001
	τ_2	0.049	-0.001	0.002	0.047	-0.003	0.002
	τ_3	0.277	-0.223	0.065	0.269	-0.231	0.068
	σ^2	0.503	0.003	0.019	0.505	0.005	0.019
$Slash_4$	lK_e	-2.535	-0.035	0.009	-2.526	-0.026	0.008
	lK_a	0.510	0.010	0.011	0.501	0.001	0.011
	lC_l	-3.030	-0.030	0.006	-3.026	-0.026	0.008
	τ_1	0.020	0.000	0.001	0.018	-0.002	0.001
	τ_2	0.045	-0.005	0.002	0.044	-0.006	0.001
	τ_3	0.272	-0.228	0.063	0.265	-0.235	0.066
	σ^2	0.480	-0.020	0.006	0.482	-0.018	0.006

4.5 Application

Two real applications are discussed in this section; the first one about a growth curve problem and the second one about a pharmacokinetic problem.

4.5.1 Growth curve data

Considering the growth soybean data set analyzed by Pinheiro and Bates (2000), Chap.6, and Davidian and Giltinan (1995), Chap.1, it is usual to consider a mixed-effects model with the random effect in the three fixed-effects parameters, which leads to the model

$$y_{ij} = \frac{\varphi_{1i}}{1+\exp\{-[x_{ij}-\varphi_{2i}]/\varphi_{3i}\}} + \varepsilon_{ij},\ j = 1,\ldots,n_i,\ i = 1,\ldots,n, \tag{4.17}$$

where $\varphi_{1i} = \beta_1 + b_{1i}$, $\varphi_{2i} = \beta_2 + b_{2i}$, $\varphi_{3i} = \beta_3 + b_{3i}$ and n_i assumes the values 8, 9 or 10 depending on the value of $i \in \{1,\ldots,n = 48\}$.

The measurements of leaf weights were taken within approximately weekly intervals after planting, over three years, 1988, 1989 and 1990, and two genotypes, P (*plant*

introduction) and F (*forrest*). The observed value y_{ij} represents the jth mean weight (in g) of leafs from soybean plants in the ith plot, after t days of being planted, where for each of the 6 year-genotype combination there were 8 plots. In this case, β_1, β_2 and β_3 represent the asymptotic leaf weight, the time at which the leaf reaches half of its asymptotic weight and the time elapsed between the leaf reaching half and $1/(1+e^{-1})$ of its asymptotic weight, respectively.

The maximum likelihood estimates of the parameters obtained by the MCEM and the approximate method with standard errors considering the normal, t and slash distribution are given in Table 7. In all the cases, the same distribution was considered for the random effects and error. Parameter ν was selected from a range of integer values, according to the lower AIC (BIC) achieved. The parameter estimates and the values of the asymptotic standard errors of the parameters are close considering these two methodologies, but it is worth noting that the approximate method is much faster than the MCEM method. The fitted profiles under the t model with 4 degrees of freedom for one plant randomly chosen for each combination of variety and year are presented in Figure 8. The model seems to deliver an adequate fit to the data set.

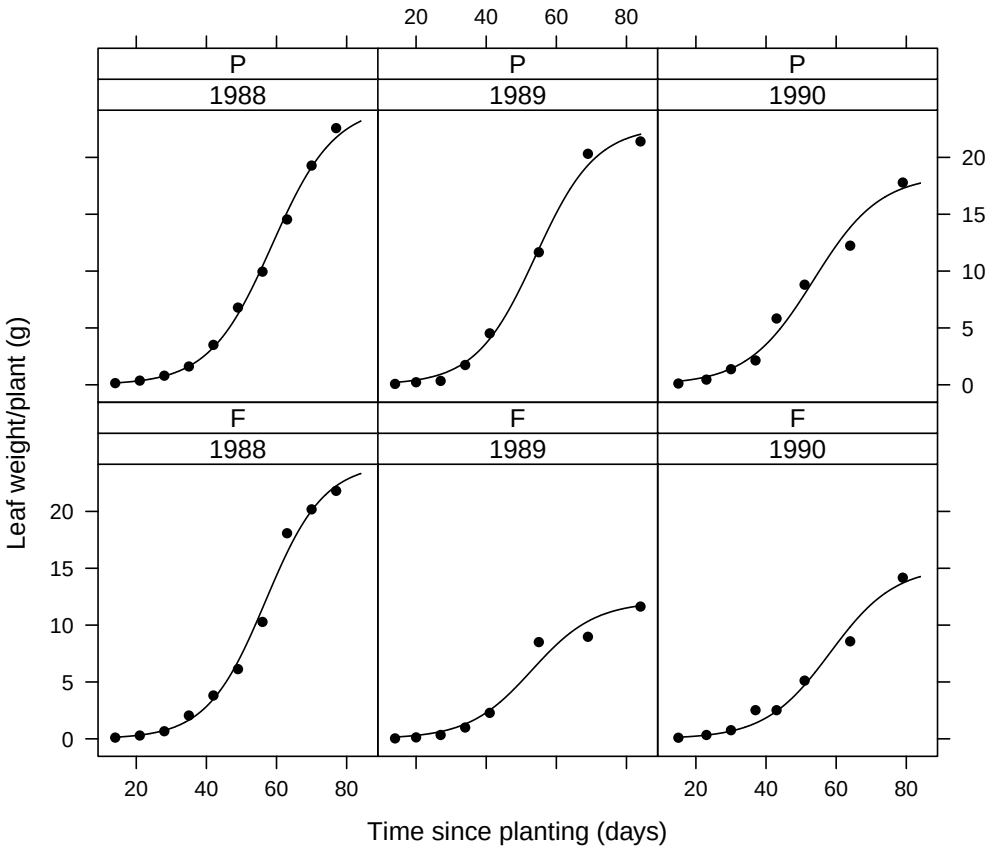


Figure 8 – Fitted profiles (lines) for randomly chosen plants under the t model with 4 degrees of freedom for the growth curves problem. The dots represent the observed values.

Table 7 – Maximum likelihood estimates of the parameters for the soybean plant growth curve problem using the MCEM and the approximate method.

	Normal				t_4				$Slash_4$			
	MCEM		Approximate		MCEM		Approximate		MCEM		Approximate	
	Estim.	SE	Estim.	SE	Estim.	SE	Estim.	SE	Estim.	SE	Estim.	SE
β_1	19.180	0.260	18.958	0.232	19.384	0.177	19.251	0.161	18.940	0.214	18.818	0.185
β_2	55.475	0.375	55.135	0.340	56.121	0.257	55.639	0.236	55.382	0.305	55.028	0.277
β_3	8.783	0.241	8.535	0.225	8.871	0.167	8.618	0.155	8.712	0.200	8.481	0.185
σ^2	1.407	0.101	1.404	0.098	0.627	0.049	0.628	0.044	0.701	0.053	0.698	0.049
τ_1	17.321	2.053	16.684	1.746	15.211	1.818	14.908	1.489	12.797	1.517	12.494	1.257
τ_2	6.685	0.788	6.587	0.472	8.030	0.950	8.226	0.646	6.572	0.776	6.629	0.516
τ_3	0.105	0.012	0.136	0.001	0.481	0.057	0.541	0.012	0.286	0.034	0.363	0.007
Log-lik	-755.270		-755.659		-707.753		-708.036		-725.499		-725.525	
AIC	1524.539		1525.319		1429.506		1430.071		1464.998		1465.051	
BIC	1537.638		1538.416		1442.604		1443.17		1478.096		1478.148	

4.5.2 Pharmacokinetic data

In the experiment described by Pinheiro and Bates (2000) on the agent theophylline, serum concentration (in mg/L) of the substance was measured at eleven times (in h) after administering of D dose (in mg/kg) in each of the twelve patients. First, nonlinear mixed-effects models with three random effects were considered, namely

$$y_{ij} = D_{ij}\exp(\varphi_{1i}+\varphi_{2i}-\varphi_{3i})\frac{[\exp(-e^{\varphi_{1i}}T_{ij})-\exp(-e^{\varphi_{2i}}T_{ij})]}{e^{\varphi_{2i}}-e^{\varphi_{1i}}}+\varepsilon_{ij},$$

where $i=1,\ldots,n$ and $j=1,\ldots,n_i$, $\varphi_{1i}=lK_{\mathrm{e}}+b_{1i}$, $\varphi_{2i}=lK_{\mathrm{a}}+b_{2i}$ and $\varphi_{3i}=lC_{\mathrm{l}}+b_{3i}$, where lK_{e}, lK_{a} and lC_{l} are the fixed-effects and b_{1i}, b_{2i} and b_{3i} are the random effects.

The estimates and standard errors for the parameters in the normal selected models, t with $\nu=4$ and the slash with $\nu=4$ are presented in Table 8 and the fitted profiles for the chosen model (t with 4 degrees of freedom) are illustrated in Figure 9. Note that the model delivers an adequate fit for most of the individuals.

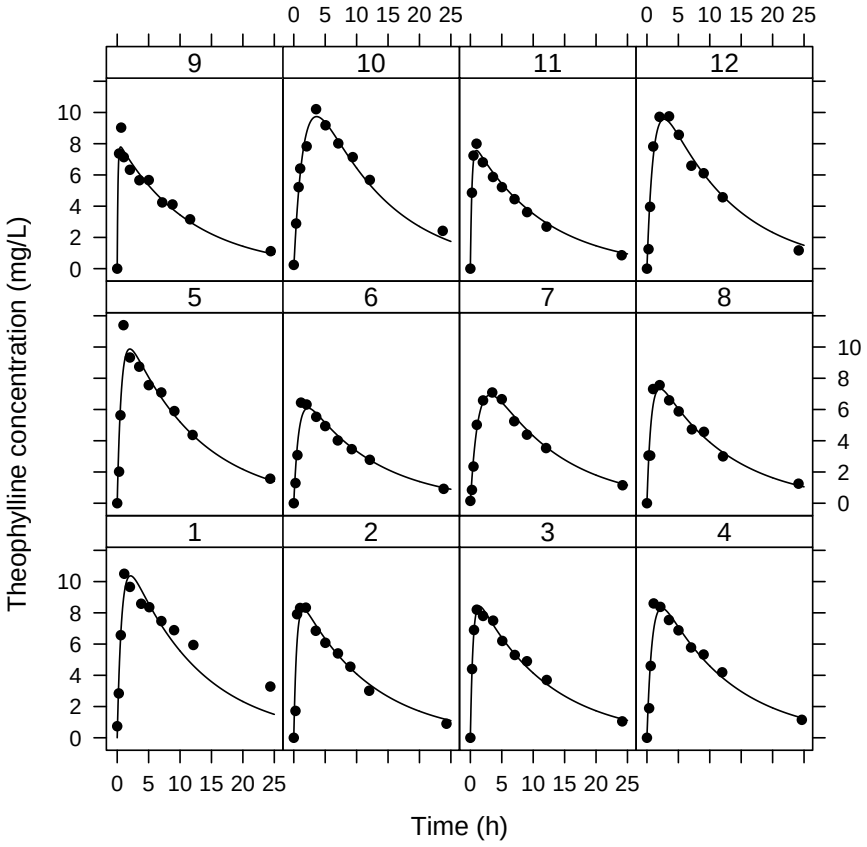


Figure 9 – Fitted profiles (lines) under the t model with 4 degrees of freedom for the pharmacokinetic problem.

4.5.2.1 Computational aspects

For the estimation process based on MCEM, samples of size $M \geq 10000$ of the full conditionals of $[u_i|\mathbf{b}_i,\mathbf{y}_i,\boldsymbol{\theta}^{(t)}]$ and $[\mathbf{b}_i|u_i,\mathbf{y}_i,\boldsymbol{\theta}^{(t)}]$ were generated by using the Monte Carlo

Table 8 – Maximum likelihood estimates of the parameters for the theophylline data set using the MCEM and approximate method.

	Normal				t_4				$Slash_4$			
	MCEM		Approximate		MCEM		Approximate		MCEM		Approximate	
	Estim.	SE	Estim.	SE	Estim.	SE	Estim.	SE	Estim.	SE	Estim.	SE
lK_e	-2.458	0.047	-2.455	0.050	-2.436	0.036	-2.432	0.039	-2.443	0.043	-2.438	0.046
lK_a	0.473	0.055	0.464	0.051	0.484	0.044	0.459	0.041	0.478	0.051	0.450	0.047
lC_l	-3.224	0.033	-3.227	0.034	-3.168	0.026	-3.167	0.026	-3.201	0.031	-3.203	0.031
σ^2	0.503	0.063	0.503	0.062	0.295	0.040	0.299	0.037	0.322	0.042	0.325	0.040
τ_1	0.000	<0.01	0.000	<0.01	0.000	<0.01	0.000	<0.01	0.000	<0.01	0.000	<0.01
τ_2	0.422	0.012	0.414	0.159	0.493	0.009	0.468	0.183	0.374	0.008	0.352	0.146
τ_3	0.028	0.012	0.028	0.002	0.022	0.009	0.022	0.001	0.020	0.008	0.020	0.001
Log-lik	-177.658		-177.970		-169.740		-170.209		-174.649		-174.670	
AIC	369.316		369.940		353.480		354.419		363.298		363.340	
BIC	372.710		373.334		356.874		357.812		366.692		366.734	

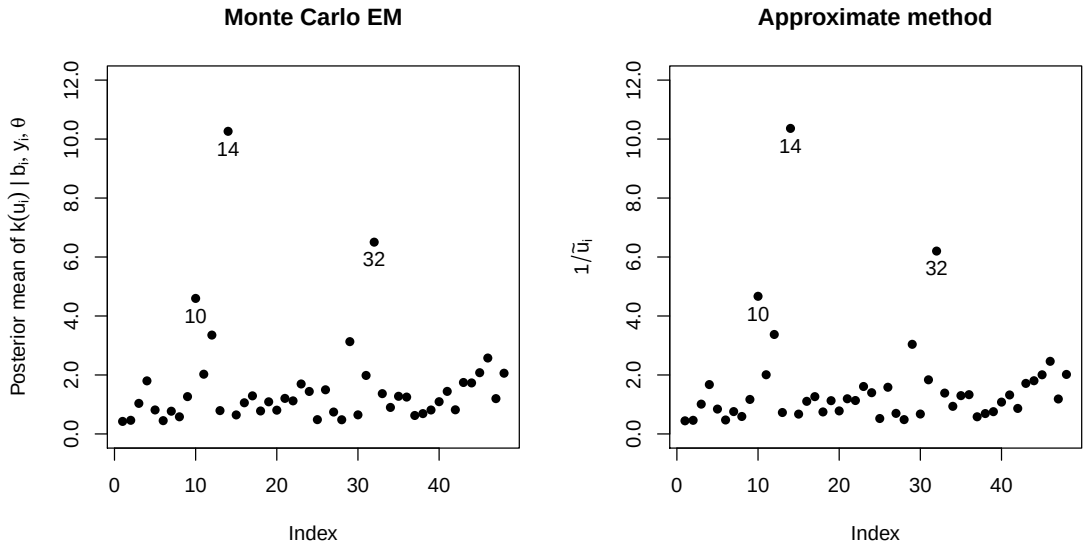
EM method, with Metropolis-Hastings within Gibbs algorithm for the E-step. Initially we made $M = 10000$ and for the following iterations, we increment M with 1000 more until reaching convergence. Four parallel runs were generated in each case; the first 80% were discarded and, with a spacing of size 100, four samples with the remaining elements were used. The convergence was monitored by using the ANOVA diagnostic method proposed by Gelman and Rubin (1992), observing that the estimated potential scale reduction factor $\widehat{R}$ was smaller than 1.1 in all the cases.

An important discussion concerns the computational aspects of the two estimation methods. Although the Monte Carlo EM may deliver slightly better estimates to the parameters, it is significantly more expensive than the approximate method, as it requires sampling from the distribution of $[u_i, \mathbf{b}_i|\mathbf{y}_i, \boldsymbol{\theta}^{(t)}]$. This computational effort increases with the inclusion of more random effects, the number of individuals and depends on initial values. On the other hand, the approximate method provides sufficiently good estimates for the parameters in few seconds. For the pharmacokinetic application, for instance, the MCEM method takes around one hour (3600s) and the approximate method takes around 2s to reach convergence. Moreover, it is worth noting that the approximate method can be easily implemented. The routines were implemented in Ox (Doornik (2009)) and run in a DELL PowerEdge 1950 server, with 2 Xeon 5430 with 2.66 GHz and 16 GB of RAM. The figures were produced using R (R Core Team (2016)).

4.5.2.2 Robustness aspects

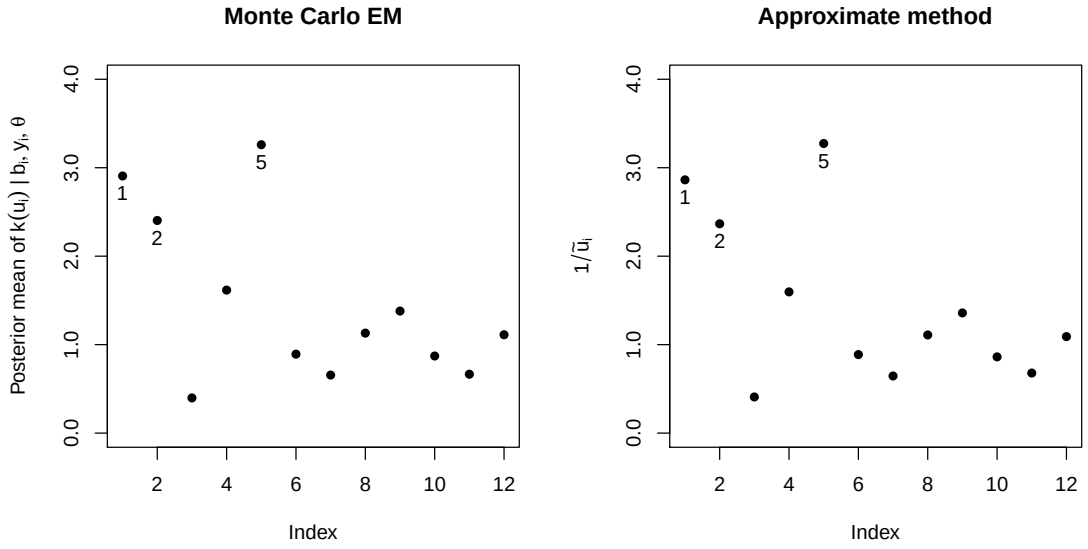
It is widely discussed in the literature that heavy-tailed distributions may deliver robust estimates for the parameters (see, for instance, Paula, Medeiros and Vilca-Labra (2009), Osorio, Paula and Galea (2007), Russo, Aoki and Paula (2012)). For models with scale mixture of normal distributions, this robustness is due to u_i, included in the model according to equations (4.4-4.6) through the function $\kappa(u_i)$. It is worth noting that $\kappa(u_i)$ appears as weights on the expressions of the maximization step for the Monte Carlo EM estimation. For the approximate method, the weights are defined by $\widetilde{u}_i$ in the iterative procedure expressions, and the robustness may be achieved due to the characterization of this quantity, according to Table 3.

For the presented models, the observations with bigger Mahalanobis distance receive the lowest weights in the iterative procedure. Thus, a possible procedure to identify outlying observations would be the graphs of the posterior mean of $\kappa(u_i)|\mathbf{b}_i, \mathbf{y}_i, \boldsymbol{\theta}$ for the Monte Carlo EM method and $\widetilde{u}_i$ for the approximate method, as shown in Figures 10 and 11 for the growth curves and pharmacokinetic applications considering the t with 4 degrees of freedom. Both estimation methods lead to the identification of the same observations. For the growth curves problem, observations 10, 14 and 32 are identified as outliers and for the pharmacokinetic application, individuals 1, 2 and 5 are pointed out



(a) (b)

Figure 10 – Posterior means of $\kappa(u_i)$ given $\mathbf{b}_i, \mathbf{y}_i, \boldsymbol{\theta}$ (MCEM) and $1/\widetilde{u}_i$ (approximate method) under the t model with 4 degrees of freedom for the soybean application.



(a) (b)

Figure 11 – Posterior means of $\kappa(u_i)$ given $\mathbf{b}_i, \mathbf{y}_i, \boldsymbol{\theta}$ (MCEM) and $1/\widetilde{u}_i$ (approximate method) under the t model with 4 degrees of freedom for the theophylline application.

as outlying observations.

In the growth curve application, the three experimental units identified (Figure 10) present a non-expected behaviour, as can be observed in Figure 12. Although the leaf weight measurements are expected to grow between two subsequent observing times, the opposite situation occurs for some experimental units including the three identified, namely observations 10, 14 and 32. When compared to the other data collected in 1988

from variety P, observation 10 has the smallest of the measurements and the second largest measurement of that group. Observation 14 has the biggest of the measurements when compared to the data collected in 1988 from variety P and that measurement has a large distance of the previous point. Observation 32 has a high growth curve when compared with the data collected in 1989 from variety P. For that observation, the $7th$ and $8th$ measurements are the biggest and the smallest of the group, respectively, indicating an unexpected decrease in the mean leaf weight.

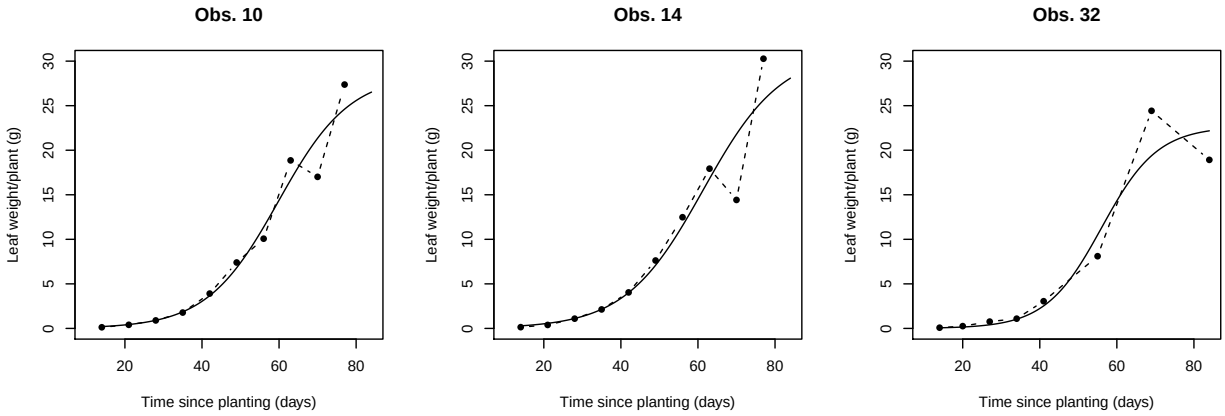


Figure 12 – Leaf weight measurements (dots) and fitted profiles (lines) for the observations identified in the robustness analysis.

For the theophylline application, patient 1 presented a slower substance elimination than predicted by the model (see Figure 9). Moreover, the highest dose of theophylline was administered to individual 5, who also presented the highest substance concentration among all the individuals. It is important to observe that the substance doses administered were 4.02, 4.40, 4.53, 4.40, 5.86, 4.00, 4.95, 4.53, 3.10, 5.50, 4.92 and 5.30 for the twelve subjects.

The case deletion diagnostics was performed and indicated a larger variation in the estimates under a normal model than under the t and slash models for the fixed-effects parameters, which confirms the robustness of heavy-tailed models. For the variance components, in some cases, the t and slash distributions led to larger variations.

4.6 Discussion

The assumption of the scale mixture of normal distributions for the joint distribution of the random effects and errors in nonlinear mixed-effects models represent an important tool to fit nonlinear correlated data as it may provide robust estimates to the involved parameters. In this work, we compare two approaches to obtain the maxi-

mum likelihood estimates in these models, and perform a simulation study to compare the MCEM and approximate method. It was observed that in general the approximate method may perform well when compared to the MCEM method and it is computationally efficient. Although there are no important differences in the bias of the estimates of the parameters related to the fixed effects and to the variance of the random errors, there is a significant gain in the computational time when the approximate method is applied. In conclusion, we recommend using the approximate method to reach reasonably good estimates for the parameters in nonlinear mixed-effects models, and if the researcher aims to obtain more accurate estimates, the approximate method can also provide fairly good initial values to the MCEM algorithm.

CHAPTER

5

RESTRICTED MAXIMUM LIKELIHOOD ESTIMATION IN NLME UNDER SMN

In this chapter, a formulation for frequentist estimating methods are presented, namely the maximum likelihood (ML) and restricted maximum likelihood (REML), for obtaining estimates for fixed effects and variance parameters in the NLME model using the SAEM algorithm. In Section 5.1, a brief introduction is provided, in Section 5.2, a motivating Monte Carlo simulation study is presented using the nlme R package for a first order compartment model and for a growth curve model. In Section 5.3, a REML formulation is presented using a scale mixture of normal distributions in the heavy tails distribution class, such as t and slash.

5.1 Introduction

Estimation methods usually considered for linear mixed-effects models cannot always be applied for nonlinear mixed-effects models. The most used procedures for these models are based on linearizations of the likelihood function through the Taylor series. The main concern of these approximated methods is the possibility of producing inconsistent estimates, mainly when the number of intragroup observations are not sufficiently large. On the other hand, working with the original likelihood function in NLME may require stochastic methods based on Monte Carlo via Markov chain (MCMC) procedures.

Many studies on NLME are based on the maximum likelihood estimates, as seen in Meza, Osorio and De La Cruz (2012) and Pinheiro and Bates (2000). However, some authors, such as Pinheiro and Bates (1995), Meza, Jaffrézic and Foulley (2007) and Arribas-Gil *et al.* (2014) claim that the variance components ML estimates might present a larger bias, since they do not consider the degrees of freedom lost in the estimation of the fixed effects.

The idea of REML is to provide bias correction through the maximization of residual contrasts. However, the original formulation of REML proposed by Patterson and Thompson (1971) to estimate the variance components via the likelihood maximization subject to a set of errors contrasts, in the context of unbalanced linear models, do not extend to cases beyond these models as centered error contrasts do not make sense for nonlinear models.

Alternatively, other proposals for the obtention of REML estimates were developed by Liao and Lipsitz (2002), Harville (1974) and Meza, Jaffrézic and Foulley (2007). The first two authors proposed, in the context of generalized linear mixed-effects models, the bias correction on the profile score function of the variance components, although this algorithm showed to be extremely slow, whereas the latter proposed the obtention of REML estimates by integrating the likelihood function on the fixed effects. This integration may be solved by Gaussian quadrature or stochastic methods. However, for NLME, it is not straightforward to apply the Gaussian quadrature method due to the dimensionality growth.

5.2 REML and ML with *nlme* package

The *nlme* package is based on the linearization method proposed by Lindstrom and Bates (1990) and also presented in Pinheiro and Bates (2000). A brief description of the method is provided as follows.

We rewrite the nonlinear mixed-effects model from Chapter 4 below. This is the mixed-effects model that we will use here,

$$\begin{aligned} \mathbf{y}_i &= \mathbf{g}(\boldsymbol{\varphi}_i, \mathbf{X}_i) + \boldsymbol{\varepsilon}_i,\ i = 1, \ldots, n, \\ \boldsymbol{\varphi}_i &= \mathbf{A}_i\boldsymbol{\beta} + \mathbf{b}_i. \end{aligned} \tag{5.1}$$

For the model (5.1), writing the random effects variance-covariance matrix in terms of a precision factor, Δ, such that $\mathbf{D}^{-1} = \sigma^2\Delta^\top\Delta$, the probability density function of $\mathbf{y}$ may be written as

$$p(\mathbf{y}; \boldsymbol{\beta}, \sigma^2, \Delta) = \frac{|\Delta|^n}{(2\pi\sigma^2)^{(N+nq)/2}} \prod_{i=1}^{n} \int \exp\left\{ \frac{\|\mathbf{y}_i - g_i(\boldsymbol{\beta}, \mathbf{b}_i)\|^2 + \|\Delta\mathbf{b}_i\|^2}{-2\sigma^2} \right\} \mathrm{d}\mathbf{b}_i. \tag{5.2}$$

The estimating algorithm described by Lindstrom and Bates (1990) alternates between two steps, a step with the obtention of penalized nonlinear least squares (PNLS) and the other with the obtention of linear mixed-effects. In the PNLS step, an estimate for Δ is fixed, then the random effects are predicted as $\mathbf{b}_i$ and the fixed effects are estimated $\boldsymbol{\beta}$ by minimizing

$$\sum_{i=1}^{n} \left[\|\mathbf{y}_i - g_i(\boldsymbol{\beta}, \mathbf{b}_i)\|^2 + \|\Delta\mathbf{b}_i\|^2 \right]. \tag{5.3}$$

In the LME step, matrix Δ is updated based in the first order Taylor expansion of the g function around the current estimates of β and $\mathbf{b}_i$, denoted by $\widehat{\beta}^{(w)}$ and $\widehat{\mathbf{b}}_i^{(w)}$, respectively. Let us define

$$\widehat{w}_i^{(w)} = \mathbf{y}_i - g_i(\widehat{\boldsymbol{\beta}}^{(w)}, \widehat{\mathbf{b}}_i^{(w)}) + \widehat{\mathbf{X}}_i^{(w)}\widehat{\boldsymbol{\beta}}^{(w)} + \widehat{\mathbf{Z}}_i^{(w)}\widehat{\mathbf{b}}_i^{(w)}, \tag{5.4}$$

where

$$\widehat{\mathbf{X}}_i^{(w)} = \left.\frac{\partial g_i}{\partial \boldsymbol{\beta}^\top}\right|_{\widehat{\boldsymbol{\beta}}^{(w)}, \widehat{\mathbf{b}}_i^{(w)}} \text{ and } \widehat{\mathbf{Z}}_i^{(w)} = \left.\frac{\partial g_i}{\partial \mathbf{b}_i^\top}\right|_{\widehat{\boldsymbol{\beta}}^{(w)}, \widehat{\mathbf{b}}_i^{(w)}}. \tag{5.5}$$

Thus, the approximate log-likelihood function to estimate Δ is given by

$$\begin{aligned} \ell_{LME}(\boldsymbol{\beta}, \sigma^2, \Delta; \mathbf{y}) &= -\tfrac{N}{2}\log(2\pi\sigma^2) - \tfrac{1}{2}\sum_{i=1}^{n}\Big\{ \log|\Sigma_i(\Delta)| \\ &\quad + \sigma^{-2}\left[w_i^{(w)} - \widehat{\mathbf{X}}_i^{(w)}\boldsymbol{\beta}\right]^\top \Sigma_i^{-1}(\Delta)\left[w_i^{(w)} - \widehat{\mathbf{X}}_i^{(w)}\boldsymbol{\beta}\right]\Big\}, \end{aligned} \tag{5.6}$$

where $\Sigma_i(\Delta) = \boldsymbol{I} + \widehat{\mathbf{Z}}_i^{(w)}\Delta^{-1}(\Delta^{-1})^\top \widehat{\mathbf{Z}}_i^{(w)^\top}$.

Similar to what is done for linear mixed-effects models, one can obtain values for $\boldsymbol{\beta}$ and σ^2, which maximizes (5.6), as a function of Δ and works with the profile likelihood function of Δ. For more details, see for instance, Pinheiro and Bates (2000).

To obtain the restricted maximum likelihood estimates for Δ, one can replace the log-likelihood function of step LME on the restricted log-likelihood, given by

$$\begin{aligned} \ell_{REML}(\sigma^2, \Delta|\mathbf{y}) &= \ell_{LME}\left(\widehat{\boldsymbol{\beta}}(\Delta), \sigma^2, \Delta|\mathbf{y}\right) \\ &\quad - \tfrac{1}{2}\sum_{i=1}^{n}\log\left|\sigma^{-2}\widehat{\mathbf{X}}_i^{(w)^\top}\Sigma_i^{-1}(\Delta)\widehat{\mathbf{X}}_i^{(w)}\right|. \end{aligned} \tag{5.7}$$

Notice that the restricted likelihood contains $\ell_{LME}\left(\widehat{\boldsymbol{\beta}}(\Delta), \sigma^2, \Delta|\mathbf{y}\right)$, the log-likelihood in (5.6) with $\widehat{\boldsymbol{\beta}}$ in the place of $\boldsymbol{\beta}$. The algorithm alternates between the PNLS and LME steps until convergence is achieved, according to some specific criterion.

A Monte Carlo simulation study with 2000 samples is performed, as follows. Let $\hat{\theta}_l$ be the estimate obtained from of the lth simulated sample and θ the true parameter value. In the equations shown in (5.8), the mean indicates the average of the estimates $\hat{\theta}_l$; bias denotes the parameter empirical bias, the difference between the mean and the theoretical value; square root of the mean squared error (RMSE) which gives the standard

deviation of the model predicting error, considered as a standard measure of efficiency,

$$\begin{aligned} \text{Mean} &= \frac{1}{2000}\sum_{l=1}^{2000}\hat{\theta}_l, \\ \text{RMSE} &= \sqrt{\frac{\sum_{l=1}^{2000}(\hat{\theta}_l-\theta)^2}{2000}} \quad \text{and} \\ \text{Bias} &= \frac{1}{2000}\sum_{l=1}^{2000}\hat{\theta}_l-\theta. \end{aligned} \tag{5.8}$$

The same proposal in the equation (5.8) is used for the fixed-effects and variance component parameters to be estimated in the next two subsections.

5.2.1 First order compartment model

The model used for the simulation study is the same as the one used for the *theophylline* data set from Pinheiro and Bates (2000). As in the data set, the sample sizes considered are $n = 12$ and $n_i = 11$, $i = 1,\ldots,12$, $j = 1,\ldots,11$. The nonlinear mixed-effects model is given by

$$\begin{aligned} y_{ij} &= \frac{\text{Dose}_{ij}\times\exp(-(\beta_1+b_{i1})+(\beta_2+b_{i2})+\beta_3)}{\exp(\beta_2+b_{i2})-\exp(\beta_3)}\times \\ &\quad \{\exp[-\exp(\beta_2+b_{i2})\ t_{ij}]-\exp[-\exp(\beta_2+b_{i2})\ t_{ij}]\}+\varepsilon_{ij}, \end{aligned} \tag{5.9}$$

where y_{ij} is the observed concentration (mg/L) of the ith subject in the time j (hours), $i = 1,\ldots,12$. Dose_i is the dose level for the ith subject, the fixed effects $\boldsymbol{\beta} = (\beta_1,\beta_2,\beta_3)^\top = (-3.2,\ 0.5,\ -2.5)^\top$ and the random effects $\mathbf{b}_i = (b_{i1},b_{i2})^\top \sim N_2(0,\mathbf{D})$ are assumed to be independent and identically distributed (i.i.d.), the random errors $\varepsilon_{ij} \sim N(0,\sigma^2)$ also i.i.d and independent of $\mathbf{b}_i$.

Similar to what was done by Pinheiro and Bates (1995), Wolfinger and Lin (1997) and Zhou (2009), two situations are considered for the variance components, which we called large and small variances. For the small variances, the values were assumed

$$\mathbf{D} = \begin{bmatrix} D_{11} & 0 \\ 0 & D_{22} \end{bmatrix} = \begin{bmatrix} 0.04 & 0 \\ 0 & 0.15 \end{bmatrix} \tag{5.10}$$

and $\sigma^2 = 0.1$, and for the large variances, the values of small variances were multiplied by 5

$$\mathbf{D} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.75 \end{bmatrix} \tag{5.11}$$

and $\sigma^2 = 0.5$.

In Figure 13, an example of the simulated first order compartment curves can be observed, assuming small variances (left) and large variances (right).

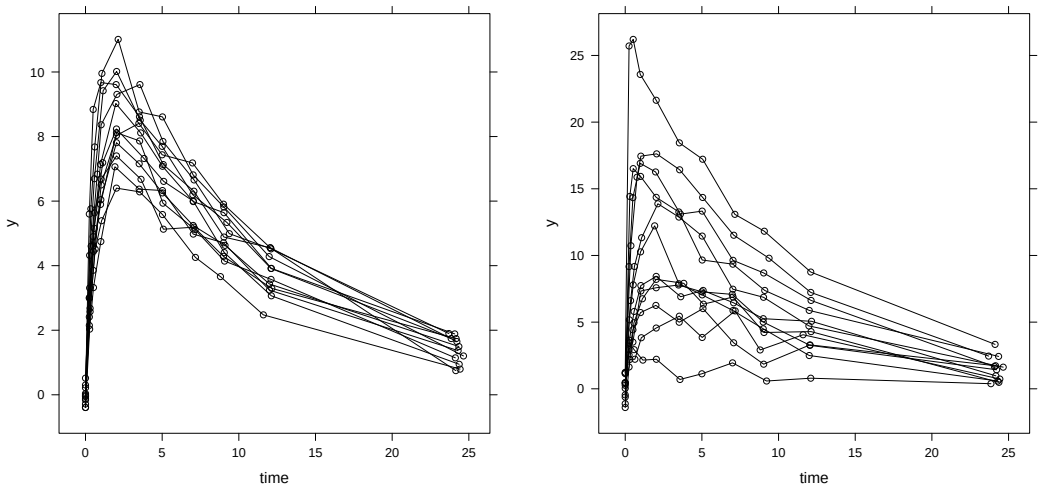


Figure 13 – Examples of first order compartment curves simulated for small and large variance components.

In Tables 9 and 10, the fixed-effects and variance component estimates can be seen in the first order compartment model for small and large variances, respectively. The estimates for the fixed effects do not differ much between REML and ML, since these estimates are obtained from the same procedure, varying in the function *nlme* from package *nlme* between REML and ML. There may be some rounding error between the two estimates. In this package, REML has a greater effect than ML on the variance components. We recall that the goal of REML is to verify bias reduction in the estimates of variance components.

Two scenarios can be observed, considering small variances (Table 9) and large variances (Table 10), where, in an absolute value, REML presents a smaller bias for the variance components D_{11} and D_{22}, whereas for the σ^2 parameter, the biases are equivalent to the first and second cases, where bias for ML estimates are smaller. Both methods led to similar RMSE and standard deviation. It is worth noting that ML estimates tend to underestimate the theoretical values of variance components.

Table 9 – REML and ML estimates' means, biases, standard deviations (sd) and square roots of the mean square error (RMSE) for the first-order compartment in the scenario with small variances .

Parameters		Mean	Bias	sd	RMSE
$\beta_1 = -3.2$					
	REML	-3.203	-0.003	0.059	0.059
	ML	-3.203	-0.003	0.059	0.059
$\beta_2 = 0.5$					
	REML	0.497	-0.003	0.116	0.116
	ML	0.497	-0.003	0.116	0.116
$\beta_3 = -2.5$					
	REML	-2.399	0.001	0.021	0.021
	ML	-2.399	0.001	0.021	0.021
$D_{11} = 0.04$					
	REML	0.038	-0.002	0.016	0,016
	ML	0.037	-0.003	0.016	0,016
$D_{22} = 0.15$					
	REML	0.141	-0.009	0.064	0.064
	ML	0.138	-0.012	0.062	0.063
$\sigma^2 = 0.1$					
	REML	0.101	0.001	0.014	0.014
	ML	0.099	-0.001	0.013	0.014

5.2.2 Logistic model

In this section, the three-parameter logistic model is considered with two random effects. Simulations with this model were studied by Pinheiro and Bates (1995), then by Wolfinger and Lin (1997) and Zhou (2009).

For the jth observation of the ith subject, with $i = 1,\ldots,15$ and $j = 1,\ldots,10$, a NLME model is given by

$$y_{ij} = \frac{\beta_1 + b_{i1}}{1 + \exp\{-[t_{ij} - (\beta_2 + b_{i2})]/\beta_3\}} + \varepsilon_{ij}, \tag{5.12}$$

where the time points are $t_{ij} \in \{100, 267, 433, 600, 767, 933, 1100, 1267, 1433, 1600\}$, the random effects $\mathbf{b}_i \sim N_2(0, \mathbf{D})$ are assumed to be i.i.d, the random errors $\varepsilon_{ij} \sim N(0, \sigma^2)$ are supposed to be independent of $\mathbf{b}_i$. We assume that $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^\top = (200, 700, 350)^\top$.

For the variance components, two situations were assumed, small and large variances. The small variances assumed the values

$$\mathbf{D} = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} = \begin{bmatrix} 40 & 0 \\ 0 & 250 \end{bmatrix} \tag{5.13}$$

and $\sigma^2 = 10$, and for large elements of $\mathbf{D}$, we considered the values of the small $\mathbf{D}$ multiplied by 10

$$\mathbf{D} = \begin{bmatrix} 400 & 0 \\ 0 & 2500 \end{bmatrix} \tag{5.14}$$

Table 10 – REML and ML estimates' means, biases, standard deviations and square roots of the mean square error for the first-order compartment in the scenario with large variances.

Parameters		Mean	Bias	sd	RMSE
$\beta_1 = -3.2$					
	REML	-3.206	-0.006	0.133	0.133
	ML	-3.206	-0.006	0.133	0.133
$\beta_2 = 0.5$					
	REML	0.468	-0.032	0.249	0,251
	ML	0.468	-0.032	0.249	0,251
$\beta_3 = -2.5$					
	REML	-2.395	0.005	0.042	0.043
	ML	-2.395	0.005	0.042	0.043
$D_{11} = 0.2$					
	REML	0.188	-0.012	0.086	0.086
	ML	0.184	-0.016	0.084	0.085
$D_{22} = 0.75$					
	REML	0.667	-0.083	0.298	0.309
	ML	0.652	-0.098	0.291	0.307
$\sigma^2 = 0.5$					
	REML	0.508	0.008	0.068	0.069
	ML	0.496	-0.004	0.067	0.067

and $\sigma^2 = 10$.

In Figure 14, examples of simulated logistic curves can be observed, assuming small and large variances.

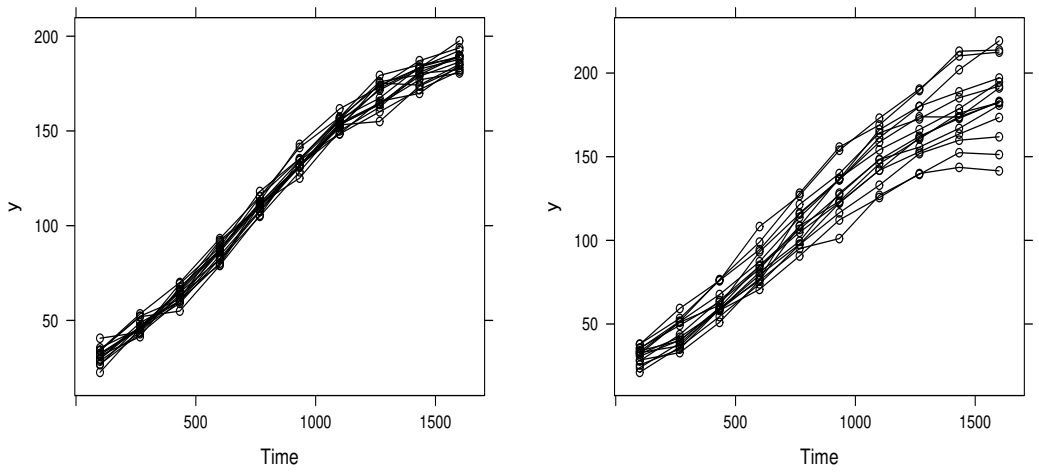


Figure 14 – Examples of simulated logistic curves for two variance component sizes, small D (left) and large D (right).

In Tables 11 and 12, the fixed-effects and variance component estimates can be observed, considering small and large elements of D, respectively. It can be observed that,

considering small variances (Table 11) and large variances (Table 12), where in all the cases the variance component estimates present, in absolute value, smaller biases for the REML method, although the RMSEs are smaller in the ML method. It is worth noting that the ML tends to underestimate the real parameter values.

Table 11 – REML and ML estimates' means, biases and square roots of the mean square error for the logistic model in the scenario with small variances.

Parameters		Mean	Bias	RMSE
$\beta_1 = 200$				
	REML	199.8476	-0.1524	2.0433
	ML	199.8476	-0.1524	2.0432
$\beta_2 = 700$				
	REML	699.0669	-0.9331	7.3304
	ML	699.0668	-0.9332	7.3305
$\beta_3 = 350$				
	REML	349.4032	-0.5968	4.6469
	ML	349.4033	-0.5967	4.6471
$D_{11} = 40$				
	REML	37.8627	-2.1373	17.0137
	ML	37.1062	-2.8938	16.7932
$D_{22} = 250$				
	REML	241.6320	-8.3680	169.0272
	ML	236.7354	-13.2646	165.9916
$\sigma^2 = 10$				
	REML	10.0963	0.0963	1.2794
	ML	9.8945	-0.1055	1.2547

5.3 Restricted maximum likelihood estimation

The restricted maximum likelihood estimation (REML) can be formulated by at least two different ways; the first one assuming error contrasts and the other applying a Bayesian approach, which is the approach considered in this thesis. When error contrasts are considered, REML estimates are obtained by maximizing the likelihood of $\boldsymbol{\theta}$ based, not on $\mathbf{y}$ but on any set of full rank values, $\mathbf{u}^\top\mathbf{y}$, such that $E(\mathbf{u}^\top\mathbf{y}) = 0$. For this approach, details can be found in Searle, Casella and McCulloch (2009).

Now, let us show details of the construction of the REML estimation criterion by the Bayesian formulation. One definition that provides computational convenience by Laird and Ware (1982) is to write the likelihood function as

$$\ell(\boldsymbol{\sigma}^2, \mathbf{D}|\mathbf{y}) = \int \ell(\boldsymbol{\beta}, \boldsymbol{\sigma}^2, \mathbf{D}|\mathbf{y}) \mathrm{d}\boldsymbol{\beta}$$

which, in a Bayesian parallel consists of assuming a uniform prior distribution for the fixed effects $\boldsymbol{\beta}$ and integrating them out in the likelihood function.

Table 12 – REML and ML estimates' means, biases and square roots of the mean square error for the logistic model in the scenario with large variances.

Parameters		Mean	Bias	RMSE
$\beta_1 = 200$				
	REML	199.7517	-0.2483	5.1521
	ML	199.7517	-0.2483	5.1521
$\beta_2 = 700$				
	REML	698.5224	-1.4776	14.2591
	ML	698.5224	-1.4776	14.2591
$\beta_3 = 350$				
	REML	349.0707	-0.9293	4.6217
	ML	349.0707	-0.9293	4.6217
$D_{11} = 400$				
	REML	380.4476	-19.5524	145.2181
	ML	372.8387	-27.1613	143.6098
$D_{22} = 2500$				
	REML	2355.8713	-144.1287	998.9081
	ML	2308.7537	-191.2463	987.3851
$\sigma^2 = 10$				
	REML	10.0995	0.0995	1.3071
	ML	9.8975	-0.1025	1.2813

It is important to observe that this method does not consist of a full Bayesian approach, but refers to the Bayesian technique to estimating fixed-effect parameter $\boldsymbol{\beta}$.

Zhou (2009) developed REML for simple and multilevel nonlinear mixed-effects models, using the EM algorithm with fully exponential Laplace approximation. In the context of linear mixed-effects models, Dempster, Laird and Rubin (1977) and Laird and Ware (1982) showed that the integration on fixed effects to obtain REML estimates can be developed by using the EM algorithm. In fact, using a Bayesian model formulation, the REML estimates can be obtained by considering fixed effects as part of the missing data using a normal distribution with a very large variance. They are integrated with other random effects through the EM algorithm.

The hierarchical representation of the NLME defined in (5.1) is given by,

$$\begin{aligned} \mathbf{Y}_i|\mathbf{b}_i, u_i &\overset{\text{ind}}{\sim} \text{SMN}_{n_i}(\mathbf{g}(\boldsymbol{\varphi}_i, \mathbf{X}_i), u_i^{-1}\sigma^2\mathbf{I}_{n_i}),\ i = 1, \ldots, n, \\ \mathbf{b}_i|v_i &\overset{\text{ind}}{\sim} \text{SMN}_q(0, v_i^{-1}\mathbf{D}), \\ u_i &\overset{\text{ind}}{\sim} H_1(\boldsymbol{\nu}) \text{ and } v_i \overset{\text{ind}}{\sim} H_2(\boldsymbol{\omega}), \end{aligned} \tag{5.15}$$

where u_i and v_i are random weights; $\boldsymbol{\nu}$ and $\boldsymbol{\omega}$ are scalar or vector parameters indexing the mixture distribution. $\mathbf{D} = \mathbf{D}(\boldsymbol{\tau}) = \text{diag}(\boldsymbol{\tau})$ is a positive-definite dispersion matrix with its elements given by $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_q)^\top$.

In this section, fixed effects are considered random, as described by Foulley and

Quaas (1995), adopting a noninformative prior, $\pi(\boldsymbol{\beta})$, proportional to a constant. Meza, Jaffrézic and Foulley (2007) considered $\boldsymbol{\beta}$ normally distributed with infinite variance.

Now, for obtaining REML estimates, consider that the parameters vector $\boldsymbol{\theta}$ become $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\sigma}^2, \boldsymbol{\tau}^\top)^\top$ and the vector of unobserved quantities is denoted by $\boldsymbol{z} = (\boldsymbol{\beta}^\top, \mathbf{b}^\top, \mathbf{u}^\top, \mathbf{v}^\top)^\top$. As the vector $\mathbf{z}$ contains unobserved quantities, the maximum likelihood estimation is based on the marginal distribution of $\mathbf{y}$. Assuming a noninformative prior for the fixed effects $\boldsymbol{\beta}$, the REML estimation for the variance-covariance components $\boldsymbol{\tau}$ and $\boldsymbol{\sigma}^2$ can be obtained by integrating the function on the fixed-effects, as well as integrating it into $\mathbf{b}$, $\mathbf{u}$ and $\mathbf{v}$, which is given by

$$\begin{aligned} f(\mathbf{y};\boldsymbol{\tau},\boldsymbol{\sigma}^2) &= \prod_{i=1}^{n} \int_0^\infty \int_{\mathbb{R}^q} \int_{\mathbb{R}^p} \phi_{n_i}(\mathbf{y}_i | g(\boldsymbol{\beta},\mathbf{b}_i), u_i^{-1}\Sigma_i)\phi_q(\mathbf{b}_i|0, v_i^{-1}\mathbf{D}) \\ &\quad \times \phi_p(\boldsymbol{\beta}) \mathrm{d}\mathbf{b}_i \, \mathrm{d}\boldsymbol{\beta} \, \mathrm{d}H_1(v) \, \mathrm{d}H_2(\omega). \end{aligned} \tag{5.16}$$

The integral in (5.16) usually does not have an analytic expression as g is a nonlinear function on the fixed and random effects. Here, to solve this integral we use Stochastic Approximation Expectation-Maximization (SAEM) algorithm which makes the exact likelihood inference easier.

Originally proposed by Patterson and Thompson (1971) in linear mixed-effects models, REML has become widely used to estimate the variance components as it presents itself as an alternative estimation that reduces the bias of the estimates, unlike ML. There are several ways to calculate the REML, however, still in the linear context, the most used method in the literature is the one where REML is obtained from the likelihood function of a linear combination of observations, called error contrasts. Another way is to consider the fixed-effects parameters as random, such that the variance component estimates are obtained from the likelihood of the observed data after integrating out the fixed-effects and random effects, as considered by Harville (1974). Foulley and Quaas (1995) used the EM algorithm and a flat prior for the fixed effects parameters.

In nonlinear mixed-effects models, the linearization technique is frequently used to calculate the REML as it is done in linear mixed models as error contrasts do not apply to the model. Here we calculate the REML by integrating the fixed effect and we will assign it a flat prior, combining with the algorithm SAEM of the same way described in Meza, Jaffrézic and Foulley (2007) and more recently Arribas-Gil *et al.* (2014). Note that it is not necessary to use likelihood approximation methods.

Let us denote $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$ the observed data, $\mathbf{b} = (\mathbf{b}_1, \ldots, \mathbf{b}_n)^\top$ the random effects vector and $\mathbf{u} = (u_1, \ldots, u_n)^\top$, $\mathbf{v} = (v_1, \ldots, v_n)^\top$ the scalar factor. Then consider $\mathbf{z} = (\boldsymbol{\beta}^\top, \mathbf{b}^\top, \mathbf{u}^\top, \mathbf{v}^\top)^\top$ as missing data. The complete log-likelihood function associated with the complete data $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{z}^\top)^\top$, considering $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\sigma}^2, \boldsymbol{\tau}^\top)^\top$ as the vector of variance

components, given

$$\begin{aligned} l_c(\tilde{\theta};\mathbf{y}_c) &= \sum_{i=1}^{n} \log l(\tilde{\theta};\mathbf{y}_i,\mathbf{z}_i) \\ &= \sum_{i=1}^{n} [\log p(\mathbf{y}_i|\boldsymbol{\beta},\mathbf{b}_i,u_i;\sigma^2) + \log p(\mathbf{b}_i|\boldsymbol{\beta},v_i;\tau) + \log p(\boldsymbol{\beta}) \\ &\quad + \log h_1(u_i;\nu) + \log h_2(v_i;\omega)]. \end{aligned} \tag{5.17}$$

We can write the complete log-likelihood as

$$\begin{aligned} l_c(\tilde{\theta};\mathbf{y}_c) &= -\frac{1}{2}\left\{ N\log\sigma^2 + n\log|\mathbf{D}| + \frac{1}{\sigma^2}\sum_{i=1}^{n} u_i\|\mathbf{y}_i - \mathbf{g}(\boldsymbol{\beta},\mathbf{b}_i)\|^2 \right. \\ &\quad \left. + \sum_{i=1}^{n} v_i \mathbf{b}_i^\top \mathbf{D}^{-1}\mathbf{b}_i \right\} + C, \end{aligned} \tag{5.18}$$

where $N = \sum_{i=1}^{n} n_i$ and C is a constant that does not depend on θ and the Q-function in this case is given by

$$\begin{aligned} Q(\tilde{\theta}|\tilde{\theta}^{(k)}) &= E[l_c(\tilde{\theta};\mathbf{y}_c)|\mathbf{y},\tilde{\theta}^{(k)}] \\ &= -\frac{1}{2}\left\{ N\log\sigma^2 + n\log|\mathbf{D}| + \frac{1}{\sigma^2}\sum_{i=1}^{n} E\left[u_i\|\mathbf{y}_i - \mathbf{g}(\boldsymbol{\beta},\mathbf{b}_i)\|^2|\mathbf{y},\tilde{\theta}^{(k)}\right] \right. \\ &\quad \left. + \sum_{i=1}^{n} E\left[v_i\mathbf{b}_i^\top\mathbf{D}^{-1}\mathbf{b}_i|\mathbf{y},\tilde{\theta}^{(k)}\right] \right\} + C. \end{aligned} \tag{5.19}$$

The SAEM is useful for fitting models that belong to the exponential family, as it can be seen, for instance, in (MEZA; OSORIO; De La Cruz, 2012).

The $(k+1)$th iteration of the SAEM algorithm consists of three steps, one simulation, stochastic approximation and the maximization step. The steps of the SAEM algorithm to accomplish the REML are as follows:

In the first step (simulation), we generate M values of the missing data vector

$$\mathbf{z}^{(k+1,m)} = (\boldsymbol{\beta}^{(k+1,m)}, \mathbf{b}^{(k+1,m)}, \mathbf{u}^{(k+1,m)}, \mathbf{v}^{(k+1,m)})$$

from the conditional distribution $p(\cdot|\mathbf{y};\tilde{\theta}^{(k)})$.

The E step cannot always generate data directly from the conditional distribution $p(\boldsymbol{\beta},\mathbf{b},\mathbf{u},\mathbf{v}|\mathbf{y};\tilde{\theta})$, then Monte Carlo Markov Chain methods are useful to simulate values of each element of $\mathbf{z}$. The elements $\boldsymbol{\beta}^{(k+1,m)}$ and $\mathbf{b}^{(k+1,m)}$ are considered Markov Chains with transition kernels $\boldsymbol{\Pi}_{\tilde{\theta}_{(k)}}$ generating from the transition probability (KUHN; LAVIELLE, 2004).

The full conditional distribution of $\boldsymbol{\beta}$ is given by

$$p(\boldsymbol{\beta}|\mathbf{y}_i,\mathbf{b}_i,u_i,v_i;\tilde{\theta}) \propto \phi_{n_i}(\mathbf{y}_i|\mathbf{g}(\boldsymbol{\beta},\mathbf{b}_i), u_i^{-1}\sigma^2\mathbf{I}_{n_i})p(\boldsymbol{\beta}) \tag{5.20}$$

$$p(\mathbf{b}_i|\mathbf{y}_i,\boldsymbol{\beta},u_i,v_i;\tilde{\theta}) \propto \phi_{n_i}(\mathbf{y}_i|\mathbf{g}(\boldsymbol{\beta},\mathbf{b}_i),u_i^{-1}\sigma^2\mathbf{I}_{n_i})\phi_q(\mathbf{b}_i|\mathbf{0},v_i^{-1}\mathbf{D}) \propto \exp\left\{-\frac{1}{2}[\frac{1}{\sigma^2}u_i\|\mathbf{y}_i-\mathbf{g}(\boldsymbol{\beta},\mathbf{b}_i)\|^2+v_i\mathbf{b}_i^\top\mathbf{D}^{-1}\mathbf{b}_i]\right\}. \quad (5.21)$$

The full conditional distribution of u_i is given by

$$p(u_i|\mathbf{y}_i,\boldsymbol{\beta},\mathbf{b}_i,v_i;\tilde{\theta}) \propto \phi_{n_i}(\mathbf{y}|\mathbf{g}(\boldsymbol{\beta},\mathbf{b}_i),u_i^{-1}\sigma^2\mathbf{I}_{n_i})\phi_q(\mathbf{b}_i|\mathbf{0},v_i^{-1}\mathbf{D})h_1(u_i;\nu). \quad (5.22)$$

We can explicit it, considering each situation: when within-subject error is the multivariate t-distribution, (5.22) corresponds to a Gamma distribution

$$p(u_i|\mathbf{y}_i,\boldsymbol{\beta},\mathbf{b}_i,v_i;\tilde{\theta}) = \text{Gamma}\left(\frac{n_i+\nu}{2},\frac{\sigma^{-2}\|\mathbf{y}_i-\mathbf{g}(\boldsymbol{\beta},\mathbf{b}_i)\|^2+\nu}{2}\right) \quad (5.23)$$

and when the within-subject error is assumed multivariate slash distribution, (5.22) corresponds to a truncated Gamma distribution

$$p(u_i|\mathbf{y}_i,\boldsymbol{\beta},\mathbf{b}_i,v_i;\tilde{\theta}) = \text{Truncated-Gamma}\left(\frac{n_i}{2}+\nu,\frac{\sigma^{-2}\|\mathbf{y}_i-\mathbf{g}(\boldsymbol{\beta},\mathbf{b}_i)\|^2}{2},T\right), \quad (5.24)$$

where we do $T=1$, representing the point which the right truncation occurs.

Similarly to the full conditional distribution in (5.22), the full conditional distribution of v_i is given by

$$p(v_i|\mathbf{y}_i,\boldsymbol{\beta},\mathbf{b}_i,u_i;\tilde{\theta}) \propto \phi_{n_i}(\mathbf{y}|\mathbf{g}(\boldsymbol{\beta},\mathbf{b}_i),u_i^{-1}\sigma^2\mathbf{I}_{n_i})\phi_q(\mathbf{b}_i|\mathbf{0},v_i^{-1}\mathbf{D})h_2(v_i;\omega) \quad (5.25)$$

and considering each case of distribution for random effects we have: when the random effects assume multivariate t-distribution

$$p(v_i|\mathbf{y}_i,\boldsymbol{\beta},\mathbf{b}_i,u_i;\tilde{\theta}) = \text{Gamma}\left(\frac{q+\omega}{2},\frac{\mathbf{b}_i^\top\mathbf{D}^{-1}\mathbf{b}_i+\omega}{2}\right) \quad (5.26)$$

and when the random effects are assumed multivariate slash distribution, (5.22) corresponds to a truncated Gamma distribution

$$p(v_i|\mathbf{y}_i,\boldsymbol{\beta},\mathbf{b}_i,u_i;\tilde{\theta}) = \text{Truncated-Gamma}\left(\frac{q}{2}+\omega,\frac{\mathbf{b}_i^\top\mathbf{D}^{-1}\mathbf{b}_i}{2},T\right), \quad (5.27)$$

where we make $T=1$, representing the point in which the right truncation occurs.

In the second step, the stochastic approximation step, we update $Q(\tilde{\theta}|\tilde{\theta}^{(k+1)})$ according to

$$Q(\tilde{\theta}|\tilde{\theta}^{(k+1)}) = Q(\tilde{\theta}|\tilde{\theta}^{(k)})+\gamma_{k+1}\left[\frac{1}{M}\sum_{m=1}^{M}l_c(\tilde{\theta};\mathbf{y},\mathbf{z}^{(k+1,m)})-Q(\tilde{\theta}|\tilde{\theta}^{(k)})\right], \quad (5.28)$$

where γ_{k+1} is the decreasing sequence to zero of positive numbers in such a way which makes the convergence faster to parameter $\tilde{\theta}$. The convergence of SAEM to a maximum

local of the likelihood is proved by Delyon, Lavielle and Moulines (1999) under fairly general conditions. We follow the strategy adopted by Kuhn and Lavielle (2005), where the following is pointed out: 1 for the first K iterations and $(k-K)^{-1}$ from iteration $K+1$ for, as follows:

$$\gamma_k = \begin{cases} 1, & \text{if } 1 \leq k \leq K \\ \frac{1}{(k-K)}, & \text{if } k \geq K+1. \end{cases} \tag{5.29}$$

The value of K is set as in Jank (2006). First, the algorithm is run with $\gamma_k = 1$. After finding the K value, the simulation runs again. The number of simulated samples is considered until convergence is reached, according to the graph of the estimates.

When the complete data likelihood in (5.18) belongs to the exponential family, we can extract minimal sufficient statistics $S(\mathbf{y}, \mathbf{z})$ such that step (5.28) is set as:

$$s^{(k)} = s^{(k)} + \gamma_{k+1}\left[\frac{1}{M}\sum_{m=1}^{M} S(\mathbf{y}, \mathbf{z}^{(k+1,m)}) - s^{(k)}\right].$$

Then, the minimal sufficient statistics for variance components, $\mathbf{D}$ and σ^2 are respectively given by

$$s_1^{(k+1)} = s_1^{(k)} + \gamma_{k+1}\left[\frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{n} u_i^{(k+1,m)}(\mathbf{b}_i^{(k+1,m)}\mathbf{b}_i^{(k+1,m)\top}) - s_1^{(k)}\right] \tag{5.30}$$

and

$$s_2^{(k+1)} = s_2^{(k)} + \gamma_{k+1}\left[\frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{n} v_i^{(k+1,m)}\|\mathbf{y}_i - \mathbf{g}(\boldsymbol{\beta}^{(k+1,m)}, \mathbf{b}_i^{(k+1,m)})\|^2 - s_2^{(k)}\right]. \tag{5.31}$$

The last step is the maximization step where $\tilde{\theta}^{(k)}$ is updated in a general form by

$$\tilde{\theta}^{(k+1)} = \arg\max Q(\tilde{\theta}|\tilde{\theta}^{(k)}), \tag{5.32}$$

and each parameter estimator is represented by

$$\widehat{\mathbf{D}}^{(k+1)} = \frac{1}{n}s_1^{(k+1)} \text{ and } \widehat{\sigma}^{2(k+1)} = \frac{1}{N}s_2^{(k+1)}. \tag{5.33}$$

It is still possible to compute estimates for fixed-effects parameters. As the fixed effects have a prior distribution, then we can update its estimate from the expectation of the conditional distribution of fixed effects given the observations $\mathbf{y}$ and under the REML of the variance components,

$$\widehat{\boldsymbol{\beta}} = E(\boldsymbol{\beta}|\mathbf{y}; \tilde{\theta}). \tag{5.34}$$

In practice, estimates of $\boldsymbol{\beta}$ are computed from the elements of the MCMC simulation.

In summary, the SAEM algorithm is given by

Algorithm 2 SAEM for the obtention of restricted maximum likelihood estimates.

1: Start with initial values $\tau = \tau_0$ and $\sigma^2 = \sigma_0^2$;
2: **Simulation step:** Draw samples of $\boldsymbol{\beta}^{(1)},\ldots,\boldsymbol{\beta}^{(m)}$; $\mathbf{b}^{(1)},\ldots,\mathbf{b}^{(m)}$; $\mathbf{u}^{(1)},\ldots,\mathbf{u}^{(m)}$ and $\mathbf{v}^{(1)},\ldots,\mathbf{v}^{(m)}$ from (5.20), (5.21), (5.22) and (5.25) respectively;
3: **Stochastic Approximation:** Compute the values in (5.30) and (5.31);
4: **Maximization step:**
   a. Compute $\hat{\sigma}^{2(k+1)}$ and $\hat{\mathbf{D}}^{(k+1)}$ according to (5.33);
   b. Obtain $\hat{\boldsymbol{\beta}}^{(k+1)}$ according to (5.34);
5: Repeat steps 2 and 4 until convergence.

5.4 Simulation study of REML using the SAEM algorithm

In this section, we present a comparison of our proposed SAEM-REML and SAEM-ML by simulation studies. A Monte Carlo simulation study was conducted to evaluate the REML and ML estimators using the SAEM algorithm. The SAEM algorithm was performed in R.

Again the model used for the simulation study is the first-order compartment model used in the theophylline dataset from Pinheiro and Bates (2000) which is available in the R *nlme* package. The NLME model is given below with two random effects,

$$y_{ij} = Dose_{ij}\exp(lK_{\mathrm{a}} + lK_{\mathrm{e}} - lC_{\mathrm{l}})\frac{\{\exp[-\exp(lK_{\mathrm{e}})t_{ij}] - \exp[-\exp(lK_{\mathrm{a}})t_{ij}]\}}{[\exp(lK_{\mathrm{a}}) - \exp(lK_{\mathrm{e}})]} + \varepsilon_{ij},$$

where y_{ij} is the observed concentration (mg/L) of the ith subject in the time j (hours), $i = 1,\ldots,n$ and $j = 1,\ldots,n_i$. Dose$_i$ is the dose level for the ith subject, with $lC_{\mathrm{l}} = \beta_1 + b_{i1}$; $lK_{\mathrm{a}} = \beta_2 + b_{i2}$; $lK_{\mathrm{e}} = \beta_3$ the fixed-effects $\boldsymbol{\beta} = (\beta_1,\beta_2,\beta_3)^\top = (-3.0,\ 0.5,\ -2.0)^\top$ and the random effects $\mathbf{b}_i = (b_{i1},b_{i2})^\top$.

We consider symmetric distributions of a scale mixture of normal distributions in the class of heavy-tailed distributions for error and random effects, $\varepsilon_{ij} \sim SMN(0,\sigma^2,H_1)$ and $\mathbf{b}_i \sim SMN_2(0,\mathbf{D};H_2)$. The slash and t-distributions are assumed with four degrees of freedom each one in all scenarios presented.

The study consists of five scenarios, where in each scenario we use theoretical fixed-effects parameters $(lC_{\mathrm{l}},lK_{\mathrm{a}},lK_{\mathrm{e}})^\top = (-3.0,0.5,-2.0)^\top$ and for theoretical variance components the following values are used

(1) $(\tau_1,\tau_2,\sigma^2)_1^\top = (0.04,0.15,0.1)^\top$, with $n = 12$ and $n_i = 11$;

(2) $(\tau_1,\tau_2,\sigma^2)_2^\top = (0.2,0.75,0.1)^\top$, with $n = 12$ and $n_i = 11$;

(3) $(\tau_1,\tau_2,\sigma^2)_3^\top = (0.04,0.15,0.1)^\top$, with $n = 20$ and $n_i = 20$;

(4) $(\tau_1,\tau_2,\sigma^2)_4^\top = (0.04,0.15,0.1)^\top$, with $n = 5$ and $n_i = 7$;

(5) $(\tau_1, \tau_2, \sigma^2)_5^\top = (0.04, 0.15, 0.5)^\top$, with $n = 12$ and $n_i = 11$.

From now on, we describe the scenarios considered and we detail how the SAEM was used in the simulation process. Monte Carlo simulation studies were performed from 3000 artificial data sets.

The SAEM algorithm handles a number of Markov chains in each iteration. The Markov chains number were fixed at 20 for the parameter estimates by the ML method and we fixed at 30 chains for the REML method. The smoothing parameter of the SAEM algorithm was chosen as described in Meza, Osorio and De La Cruz (2012), in both cases we use $\gamma_k = 1$, for $1 \leq k \leq 300$ and $\gamma_k = 1/(k-500)$, for $301 \leq k \leq 500$. The number 301 was chosen after a first run of the algorithm, considering the smoothing parameter equal to 1 and observing from which iteration the estimates converged.

Considering $\hat{\theta}_l$, a parameter estimate at the lth simulation and θ for the theoretical value of the parameter, $\text{Mean} = \sum_{l=1}^{3000} \hat{\theta}_l/3000$, $\text{RMSE} = \sqrt{\sum_{l=1}^{3000}(\hat{\theta}_l - \theta)^2/3000}$ and $\text{Bias} = \frac{1}{3000}\sum_{l=1}^{3000} \hat{\theta}_l - \theta$.

Regarding the five scenarios from the simulation study, we have that the scenarios 1, 2 and 5 have the same configuration as the theophylline data, which corresponds to 12 individuals with 11 observations. Moreover, scenarios 2, and 5 can be found in the appendix. Scenario 3 can also be found in the appendix, Also in the appendix is the scenario 3, we increased both the number of individuals and number of observations to 20. Finally, we have scenario 4, where we use a smaller number of both individuals and observations, 5 and 7, respectively. In the next section, we have comments on scenario 1.

5.4.1 Scenario One

In the first scenario we use theoretical fixed-effects parameters $(lC_l, lK_a, lK_e)^\top = (-3.0, 0.5, -2.0)^\top$ and theoretical variance components $(\tau_1, \tau_2, \sigma^2)_1^\top = (0.04, 0.15, 0.1)^\top$, and the number of subjects is $n = 12$ and the number of observations is $n_i = 11$. Table 13 shows the simulation results for fixed-effects and variance components.

When errors and random effects assume normal distribution, it can be observed that numerically the absolute bias of the lK_l parameter is greater for the ML estimator and the RMSE are equal; lK_a has a lower absolute bias in ML but the RMSE is higher than the RMSE in the REML; lK_e presents the same bias in both estimation methods, with lower RMSE in REML. For the variance components, it can be observed that the bias is lower for all components τ_1, τ_2 and σ^2 in the REML, however the RMSE are lower in the ML method. When errors and random effects assume t distribution with 4 degrees of freedom, it can be observed for parameter lK_l that numerically the absolute bias and RMSE are higher in the ML estimator than for the REML; lK_a has a lower absolute bias, but the RMSE is higher in ML than for REML; lK_e presents the same bias in the two

estimation methods, ML and REML, with the RMSE being higher in ML. For the variance components τ_1, τ_2 and σ^2, it can be observed that the biases are numerically smaller for the REML than for the ML, however the RMSE are smaller in the ML. Finally, when both errors and random effects assume a slash distribution with 4 degrees of freedom, it can be observed that for lK_l, numerically, the absolute bias and RMSE are greater in the ML method; lK_a presents the same bias in both estimation methods, but the RMSE is higher in ML; lK_e has a lower bias in the ML method but its RMSE is greater than in the REML estimation method. Regarding the variance components τ_1, τ_2 and σ^2, we can see that all biases are lower in the REML method, however the RMSE are lower for the ML estimator.

Table 13 – Simulation results for the pharmacokinetic model with theoretical fixed-effects parameters $(lC_l, lK_a, lK_e)^\top = (-3.0, 0.5, -2.0)^\top$ and theoretical variance components $(\tau^\top, \sigma^2)_l^\top = (0.04, 0.15, 0.1)^\top$.

		ML			REML		
Dist.	Par.	Mean	Bias	RMSE	Mean	Bias	RMSE
Normal	lC_l	-3.0009	-0.0009	0.0589	-3.0003	-0.0003	0.0589
	lK_a	0.5002	0.0002	0.1172	0.4996	-0.0004	0.1162
	lK_e	-2.5003	-0.0003	0.0288	-2.5003	-0.0003	0.0286
	τ_1	0.0368	-0.0032	0.0168	0.0402	0.0002	0.0183
	τ_2	0.1384	-0.0116	0.0637	0.1516	0.0016	0.0684
	σ^2	0.0987	-0.0013	0.0138	0.0995	-0.0005	0.0139
t_4	lC_l	-3.0067	-0.0067	0.0647	-3.0035	-0.0035	0.0634
	lK_a	0.5017	0.0017	0.1364	0.5043	0.0043	0.1337
	lK_e	-2.5014	-0.0014	0.0307	-2.5014	-0.0014	0.0304
	τ_1	0.0360	-0.0040	0.0204	0.0404	0.0004	0.0222
	τ_2	0.1313	-0.0187	0.0781	0.1485	-0.0015	0.0850
	σ^2	0.0990	-0.0010	0.0254	0.0996	-0.0004	0.0256
$Slash_4$	lC_l	-3.0065	-0.0065	0.0675	-3.0035	-0.0035	0.0664
	lK_a	0.4974	-0.0026	0.1324	0.4974	-0.0026	0.1321
	lK_e	-2.5017	-0.0017	0.0324	-2.5020	-0.0020	0.0322
	τ_1	0.0371	-0.0029	0.0179	0.0406	0.0006	0.0194
	τ_2	0.1401	-0.0099	0.0678	0.1544	0.0044	0.0738
	σ^2	0.0996	-0.0004	0.0158	0.1002	0.0002	0.0160

In Figure 15, the letter (a) represents boxplots for variance components when errors and random effects assume normal distribution; letter (b) represents boxplots for variance components when errors and random effects assume t distribution with degree of freedom 4; letter (a) represents boxplots for variance components when errors and random effects assume slash distribution, which also has 4 degrees of freedom. The horizontal line represents the theoretical parameter in each case. It can be seen that the ML method underestimates the variance components. This underestimation is more intense in parameters τ_1 and τ_2 for all distributions. The worst case can be observed in the figures in line (b), where at least 50% of the estimates are lower than the theoretical value used for the simulation. REML provides, in these cases, better estimates for the parameters τ_1 and τ_2.

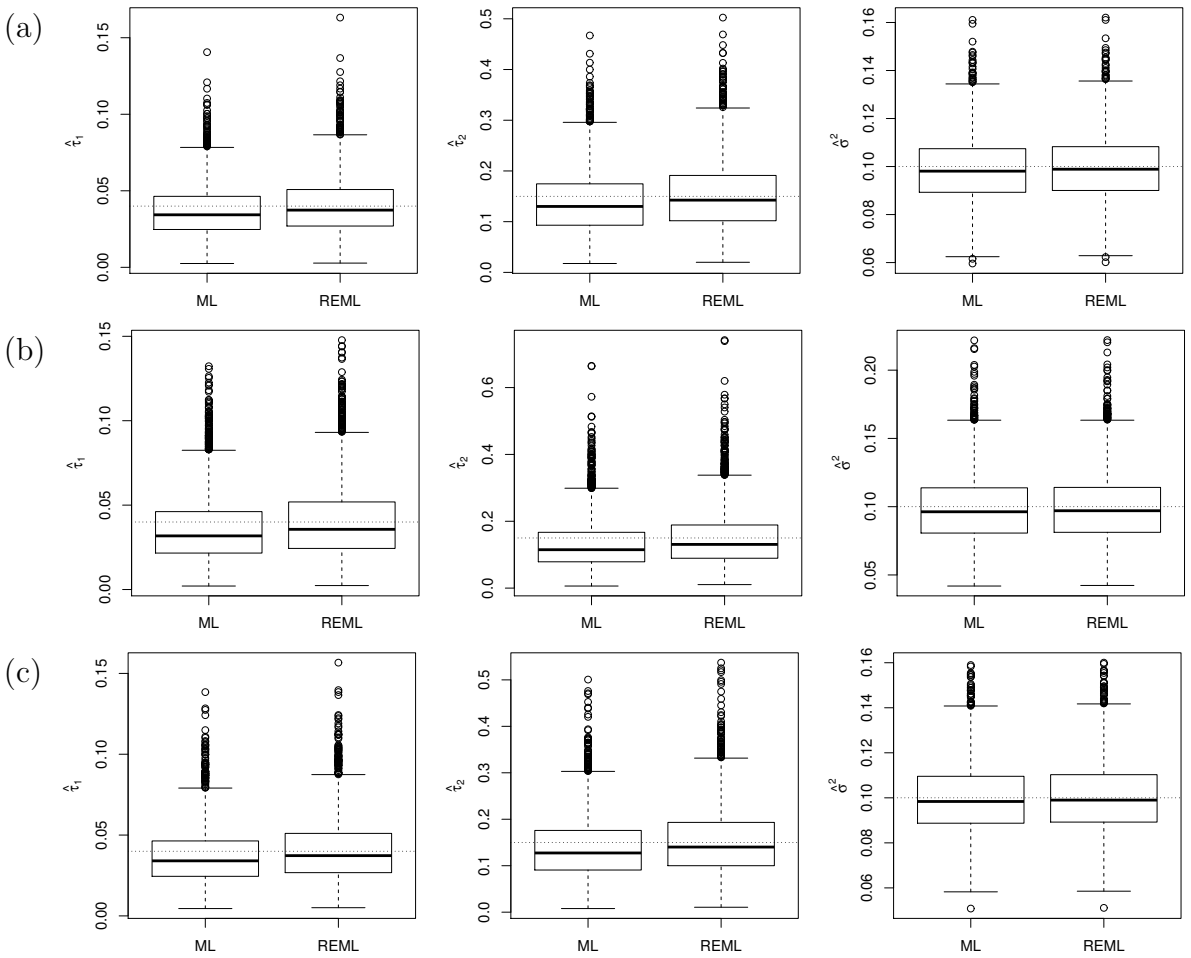


Figure 15 – Boxplots of the variance component estimates obtained from the simulation study. Panels in line (a) correspond to the normal distribution, panels in line (b) represent the t_4 distribution and in line (c) $slash_4$ distribution. The horizontal dotted lines are the theoretical values.

5.5 Application

5.5.1 Theophylline data

Here we use theophylline data as an illustration for REML and ML estimation methods. The nonlinear function is slightly different from that described in Chapter 4. Here we assume two random effects. Thus, the way of displaying the fixed effects is slightly different.

Table 14 shows the parameter estimates and standard errors of REML and ML for the theophylline data set using SAEM algorithm. We consider the same distributions for both errors and random effects. The degrees of freedom for the t and slash distributions were fixed.

Figures 16 and 17 show the SAEM iterations of the estimates by the REML and ML methods, respectively in model $t_{3.5}$ / t_4. Note that there was convergence in all cases and that did not take long to happen.

Table 14 – Restricted Maximum likelihood and Maximum likelihood estimates of the parameters for the theophylline data set.

		REML	SE	ML	SE
Normal / Normal	lC_l	-3.21539	0.04722	-3.22293	0.09783
	lK_a	0.47662	0.20236	0.54101	0.46812
	lK_e	-2.45389	0.05992	-2.45885	0.05198
	τ_1	0.03050	0.00951	0.02816	0.00874
	τ_2	0.48223	0.15317	0.44156	0.14939
	σ^2	0.50606	0.06921	0.50063	0.06839
		REML	SE	ML	SE
$t_{3.5}$ / t_4	lC_l	-3.19259	0.05722	-3.18037	0.05589
	lK_a	0.37114	0.23242	0.31371	0.22543
	lK_e	-2.43740	0.05302	-2.43354	0.04318
	τ_1	0.02171	0.00371	0.01657	0.00310
	τ_2	0.43803	0.04358	0.28037	0.03142
	σ^2	0.27938	0.07954	0.27361	0.07346
		REML	SE	ML	SE
$Slash_{2.5}$ / $Slash_3$	lC_l	-3.20929	0.06162	-3.20832	0.06981
	lK_a	0.42541	0.23156	0.48424	0.22078
	lK_e	-2.44107	0.05403	-2.43988	0.04898
	τ_1	0.01963	0.00480	0.01822	0.00713
	τ_2	0.31760	0.08571	0.30380	0.12140
	σ^2	0.26502	0.05370	0.26397	0.05388

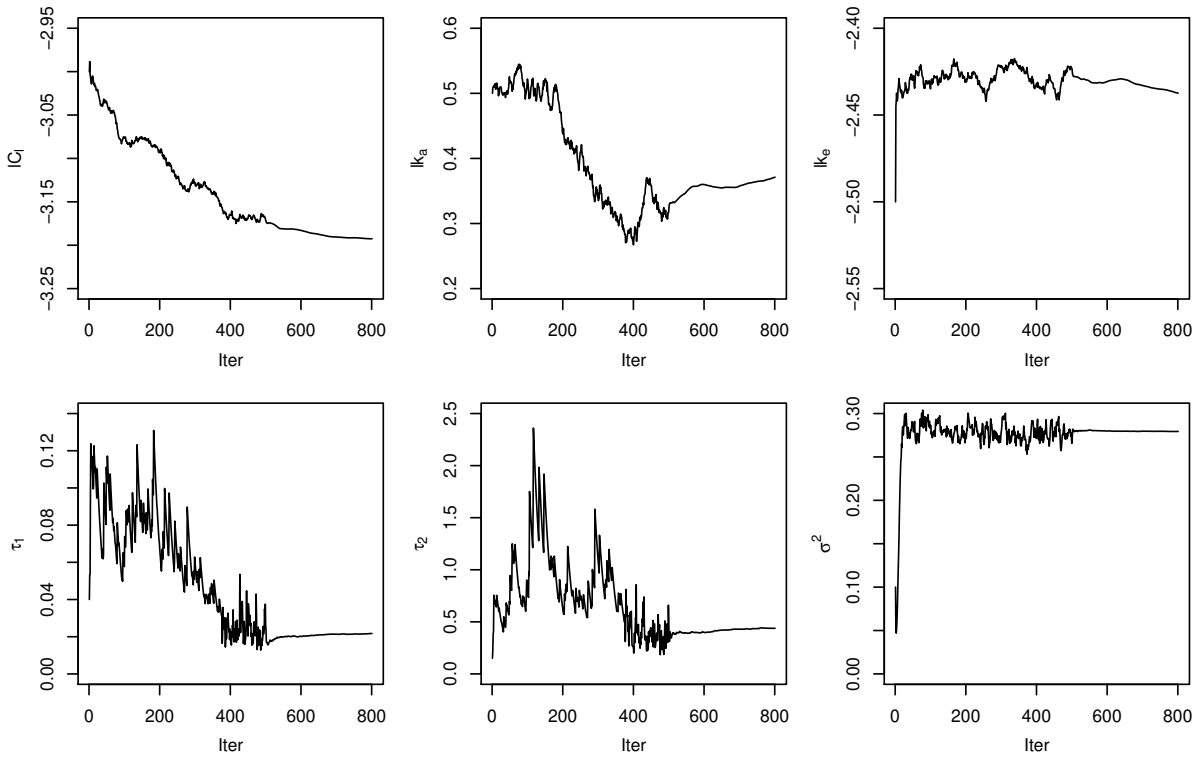


Figure 16 – Iterations of the SAEM algorithm of REML estimates for the theophylline model for $t_{3.5}$ / t_4.

5.6 Discussion

In this chapter, we study the ML and REML estimators in NLME models. The purpose of the REML estimator is to estimate the variance components as the ML estimators tend to underestimate the parameters while the REML generally obtains the least bias. At first, we analysed the gain in the REML bias in the NLME model implemented *nlme* package of the R program proposed by Pinheiro and Bates (2000). The estimation method of the fixed-effects parameter and random-effects parameters of the NLME model of the R package is based on the method that linearizes the non-linear model by the Taylor approximation and considers the normal distribution for errors and random effects. We did a simulation study with 2000 sets of artificial data, assuming distribution, with the characteristics of a first order compartment model and also for a growth curve model. There was a gain in terms of bias for the REML estimators of the variance components against the ML estimators in both the first order compartment model and the curve model growth. It is worth mentioning that in the growth curve model, the absolute biases of the variance components were not close to zero in comparison, even though REML still had less bias. In the second step, we propose a Harville (1974)-based REML estimator where fixed effects are integrated into the function likelihood of complete data. In this case, we propose the class of mixtures of normal distribution scale symmetric to errors and random effects, more specifically we use the slash, t heavy tails distributions and compare it to

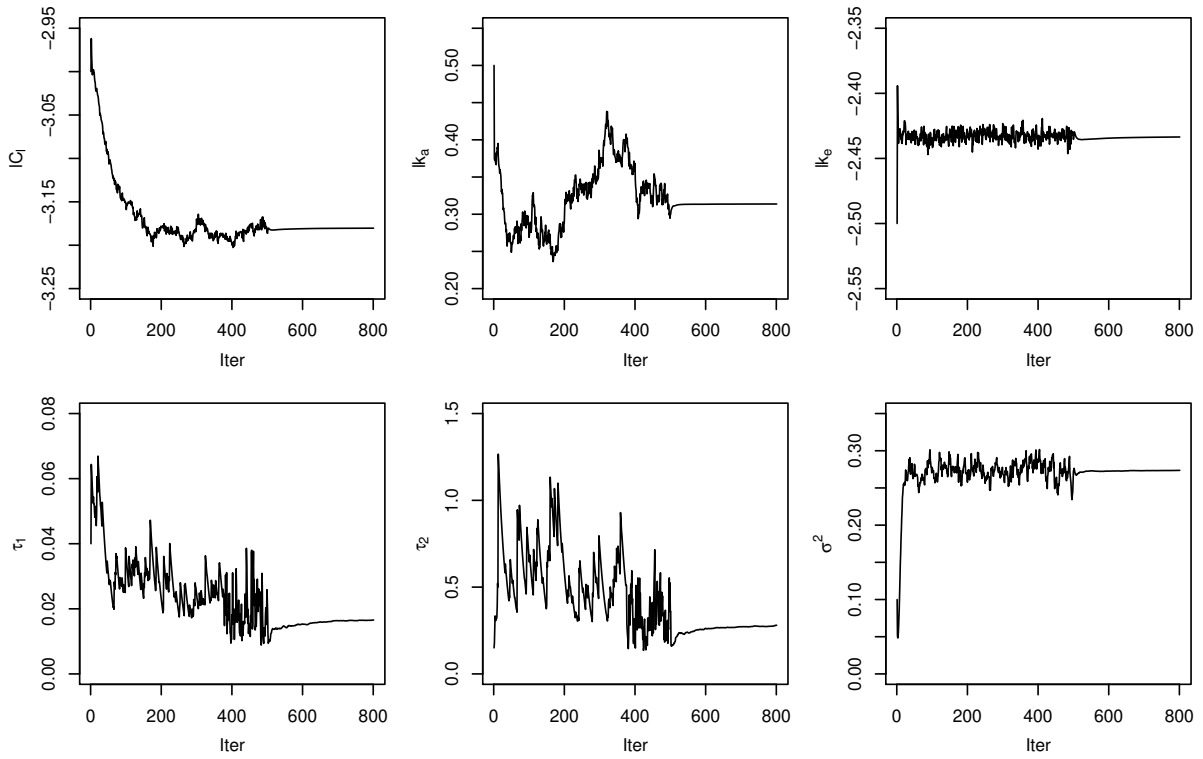


Figure 17 – Iterations of the SAEM algorithm of ML estimates for the theophylline model $t_{3.5}$ / t_4.

normal. To the best of our knowledge, there are no studies in the literature about the REML estimator in NLME models for this class of distributions following the Harville approach. Meza, Jaffrézic and Foulley (2007) studied only to the usual normal distribution. Moreover, to compare the performance of the REML estimator, we calculated the bias and root of the mean square error in a Monte Carlo simulation study for 3000 artificial dataset. The parameter estimation was obtained using the version of the stochastic approximation of the EM algorithm proposed by Kuhn and Lavielle (2005), known as SAEM. This algorithm, according to reports by Meza, Osorio and De La Cruz (2012), has the advantage of being faster than the MCEM algorithm, because it uses few Markov chains in each iteration. With the data from the theopylline application, the SAEM algorithm obtained the parameter estimates in around 130 seconds (ML) and around 100 seconds (REML) when the distribution was normal for errors and random effects; also around 130 seconds (ML) and around 133 seconds for slash distribution in both errors and random effects; and around 127 seconds (ML) and around 105 seconds (REML) for t distribution in both errors and random effects. The routine were implemented in R Core Team (2016) and run in a DELL Inspiron 14 5448 with Intel Core i7 – 5500 CPU 2.40 GHz and 8 GB of RAM.

CHAPTER

6

GENERAL NONLINEAR MIXED-EFFECTS MODEL

This chapter brings a new proposal of general nonlinear mixed-effects models, where the random effects, which do not necessarily follow a normal distribution are not included in a linear structure as usual in the literature.

6.1 Introduction

Nonlinear mixed-effects models are widely used to model correlated data where there is generally a strong appeal for longitudinal data, but in general the use of repeated measure is mandatory. Several authors have developed works in this area, some present both the linear mixed-effects model and the nonlinear mixed-effects models. Some works that we can highlight are Molenberghs and Verbeke (2005), (RUSSO; LESAFFRE; PAULA, 2012), Lachos, Castro and Dey (2013), De La Cruz (2014) and Meza, Osorio and De La Cruz (2012).

More recently, Pereira and Russo (2019) developed asymmetric nonlinear mixed-effects models as opposed to symmetric ones. They worked with the SMN distribution class in order to detect atypical observations. In the nonlinear model of their work, random effects enter linearly and they use the EM-type and Newton-Raphson algorithms to estimate the parameters.

The model by Meza, Osorio and De La Cruz (2012) is too restrictive, as the model assumes a linear predictor for the a subject-specific parameter vector in the nonlinear function. Here, we propose to replace this condition for subject-specific parameter vector in such a way that it is not necessary to specify it.

In this chapter, we work on the obtention of ML estimates for fixed-effects and

variance components in the context of the NLME model, where the error distributions and random effects do not follow a normal distribution. Here we assume symmetric heavy-tailed distributions in the scale mixture of normal distribution class for error and random effects. These types of distributions are appropriate to accommodate outlying observations, as it can be seen, for instance, in Savalli, Paula and Cysneiros (2006), Russo, Paula and Aoki (2009), Staudenmayer, Lake and Wand (2009), Meza, Osorio and De La Cruz (2012) in a frequentist context and Rosa, Padovani and Gianola (2003), Lachos, Castro and Dey (2013), De La Cruz (2014) in a Bayesian approach. Here we assume outliers within subjects, which are evaluated by within-subject error.

6.2 The SMN nonlinear mixed-effect model

The NLME can be written as in Molenberghs and Verbeke (2005), Chap. 20, by

$$Y_i = f_i(X_i, \boldsymbol{\beta}, Z_i, \mathbf{b}_i) + \boldsymbol{\varepsilon}_i, \ i = 1, \ldots, n, \tag{6.1}$$

where Y_i $(n_i \times 1)$ denotes the response vector for subject i, X_i and Z_i are design matrices, where in the longitudinal study usually the time should be entered into either X_i or Z_i or both. n is the number of subjects, n_i the number of observations of subject i and $\boldsymbol{\varepsilon}_i$ the within-subject errors, independent across subjects and are assumed to have normal distribution $N_{n_i}(0, \Sigma_i)$. Furthermore, function f is allowed to be non-linear, $\boldsymbol{\beta}$ $(p \times 1)$ denotes the vector of fixed-effects parameters, $\mathbf{b}_i$ $(q \times 1)$ denotes the vector of random effects, assumed independent and with normal distribution $N_q(0, \Gamma_i)$. We assume that the $\boldsymbol{\varepsilon}_i$ and $\mathbf{b}_i$ are mutually independent.

The most common assumption for the distribution of errors and random effects in NLME models is multivariate normal, which may not be the most appropriate choice in cases of heavy-tailed data or the presence of outliers. Here we will use the SMN distributions for random effects as well, allowing the distributions for errors and random effects to be assumed differently. From (6.1), the model takes the form of

$$\begin{aligned} \mathbf{Y}_i | \mathbf{b}_i &\overset{\text{ind}}{\sim} \text{SMN}_{n_i}(f_i(X_i, \boldsymbol{\beta}, Z_i, \mathbf{b}_i), \Sigma_i; H_1) \\ \mathbf{b}_i &\overset{\text{ind}}{\sim} \text{SMN}_q(0, \Gamma; H_2), \end{aligned} \tag{6.2}$$

where Σ_i and Γ are positive-definite dispersion matrices. Here we assume $\Gamma = \Gamma(\boldsymbol{\tau})$ with elements $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_P)^\top$ its elements and $\Sigma_i = \sigma^2 \mathbf{I}_{n_i}$, with $\sigma^2 > 0$ a scalar and $\mathbf{I}_r$ $(r \times r)$ the identity matrix .

For $i = 1, \ldots, n$, the hierarchical representation of the NLME defined in (6.2) is

given by

$$\begin{aligned} \mathbf{Y}_i|\mathbf{b}_i,u_i &\overset{\text{ind}}{\sim} \text{SMN}_{n_i}(f_i(\mathbf{X}_i,\boldsymbol{\beta},\mathbf{Z}_i,\mathbf{b}_i),u_i^{-1}\sigma^2\mathbf{I}_{n_i}),\\ \mathbf{b}_i|v_i &\overset{\text{ind}}{\sim} \text{SMN}_q(0,v_i^{-1}\Gamma),\\ u_i &\overset{\text{ind}}{\sim} H_1(\nu) \text{ and } v_i \overset{\text{ind}}{\sim} H_2(\omega), \end{aligned} \tag{6.3}$$

where u_i and v_i are random weights; ν and ω are scalar or vector parameters indexing the mixture distribution.

The SMN subclass has been used as an important tool in robust estimation. Among these subclasses we can cite the t, slash, normal and contaminated normal distributions and the Laplace or double exponential. In this chapter, the t and slash distributions are considered to illustrate the proposed methodology. In fact, the representation of the multivariate t distribution occurs when in (6.2) $\mathbf{Y}\sim \text{Mt}(\mu,\Sigma;\nu)$, $\nu>0$ and $u\sim\text{Gamma}(\nu/2,\nu/2)$ while that for the multivariate slash distribution is denoted by $\mathbf{Y}\sim\text{MSl}(\mu,\Sigma;\nu)$, $\nu>0$ and $u\sim\text{Beta}(\nu,1)$, $\nu>0$. For both, the t and slash distributions, ν represents the degrees of freedom and this parameter controls the kurtosis of the distribution. It is interesting to note that when $\nu\to\infty$ the normal distribution is recovered. The joint distribution is represented as

$$f(\mathbf{y};\boldsymbol{\beta},\tau,\sigma^2)=\prod_{i=1}^{n}\int_0^\infty\int_{R^q}\phi_{n_i}(\mathbf{Y}_i;f_i(X_i,\boldsymbol{\beta},Z_i,\mathbf{b}_i),u_i^{-1}\sigma^2\mathbf{I}_{n_i})\phi_q(\mathbf{b}_i;0,v_i^{-1}\Gamma)\text{d}\mathbf{b}_i\text{d}H_1(\nu)\text{d}H_2(\omega), \tag{6.4}$$

where $\phi_n(\cdot|\mu,\Sigma)$ denotes the n-dimensional normal probability density function with location and scale parameters μ and Σ, respectively.

6.3 Maximum likelihood estimation

Let us denote by $\mathbf{y}=(\mathbf{y}_1^\top,\ldots,\mathbf{y}_n^\top)^\top$ the observed data, $\mathbf{b}=(\mathbf{b}_1,\ldots,\mathbf{b}_n)^\top$ the random effects vector and $u=(u_1,\ldots,u_n)^\top$ and $v=(v_1,\ldots,v_n)^\top$, the scalar factors. Then consider $\mathbf{z}=(\mathbf{b}^\top,u^\top,v^\top)^\top$ as missing data. The complete log-likelihood function associated with the complete data $\mathbf{y}_c=(\mathbf{y}^\top,\mathbf{z}^\top)^\top$, considering $\theta=(\boldsymbol{\beta}^\top,\sigma^2,\tau^\top)^\top$ as parameter vector given by

$$\begin{aligned} l_c(\theta;\mathbf{y}_c) &= \sum_{i=1}^{n}\log l(\theta;\mathbf{y}_{i_c})\\ &= \sum_{i=1}^{n}[\log p(\mathbf{y}_i|\mathbf{b}_i,u_i;\boldsymbol{\beta},\sigma^2)+\log p(\mathbf{b}_i|v_i;\tau)+\log h_1(u_i;\nu)+\log h_2(v_i;\omega)]. \end{aligned}$$

We can write the complete log-likelihood as

$$\begin{aligned} l_c(\theta;\mathbf{y}_c) &= -\frac{1}{2}\left\{N\log\sigma^2 + n\log|\Gamma| + \frac{1}{\sigma^2}\sum_{i=1}^{n} u_i\|\mathbf{y}_i - f(\mathbf{X}_i,\boldsymbol{\beta},\mathbf{Z}_i,\mathbf{b}_i)\|^2 \right. \\ &\quad \left. + \sum_{i=1}^{n} v_i\mathbf{b}_i^\top\Gamma^{-1}\mathbf{b}_i)\right\} + C, \end{aligned} \tag{6.5}$$

where $N = \sum_{i=1}^{n} n_i$ and C is a constant that does not depend on θ and the Q-function in this case is given by

$$\begin{aligned} Q(\theta|\theta^{(k)}) &= E[l_c(\theta;\mathbf{y}_c)|\mathbf{y},\theta^{(k)}] \\ &= -\frac{1}{2}\left\{N\log\sigma^2 + n\log|\Gamma| + \frac{1}{\sigma^2}\sum_{i=1}^{n} E\left[u_i\|\mathbf{y}_i - f(\mathbf{X}_i,\boldsymbol{\beta},\mathbf{Z}_i,\mathbf{b}_i)\|^2|\mathbf{y},\theta^{(k)}\right]\right. \\ &\quad \left. + \sum_{i=1}^{n} E\left[v_i\mathbf{b}_i^\top\Gamma^{-1}\mathbf{b}_i|\mathbf{y},\theta^{(k)}\right]\right\} + C. \end{aligned} \tag{6.6}$$

The SAEM is useful for fitting models that belong to the exponential family. However it is discussed by Meza, Osorio and De La Cruz (2012) that when the degrees of freedom regarding the distribution of weights are known, the likelihood belongs to the exponential family.

The $(k+1)$th iteration of the SAEM algorithm is consists of three steps: the simulation, stochastic approximation and maximization steps. The steps of the SAEM algorithm to accomplish the ML are as follows:

In the first step, the simulation step, we generate M values of the missing data vector $\mathbf{z}^{(k+1,m)} = (\mathbf{b}^{(k+1,m)}, u^{(k+1,m)}, v^{(k+1,m)})$ from the conditional distribution $p(\cdot|\mathbf{y};\theta^{(k)})$. This step cannot always generate data directly from the conditional distribution $p(\mathbf{b},u,v|\mathbf{y};\theta)$. Then, Monte Carlo Markov Chain methods are useful to simulate values of each element of $\mathbf{z}$. The elements and $\mathbf{b}^{(k+1,m)}$ are considered Markov chains with transition kernels $\Pi_{\theta_{(k)}}$ generated from the transition probability, Kuhn and Lavielle (2004)

$$\begin{aligned} p(\mathbf{b}_i|\mathbf{y}_i,u_i,v_i;\theta) &\propto \phi_{n_i}(\mathbf{y}_i|f(\mathbf{X}_i,\boldsymbol{\beta},\mathbf{Z}_i,\mathbf{b}_i),u_i^{-1}\sigma^2\mathbf{I}_{n_i})\phi_q(\mathbf{b}_i|0,v_i^{-1}\Gamma) \\ &\propto \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma^2}u_i\|\mathbf{y}_i - f(\mathbf{X}_i,\boldsymbol{\beta},\mathbf{Z}_i,\mathbf{b}_i)\|^2 + v_i\mathbf{b}^\top\Gamma^{-1}\mathbf{b}_i\right]\right\}. \end{aligned} \tag{6.7}$$

The full conditional distribution of u_i is given by:

$$p(u_i|\mathbf{y}_i,\mathbf{b}_i,v_i;\theta) \propto \phi_{n_i}(\mathbf{y}|f(\mathbf{X}_i,\boldsymbol{\beta},\mathbf{Z}_i,\mathbf{b}_i),u_i^{-1}\sigma^2\mathbf{I}_{n_i})\phi_q(\mathbf{b}_i|0,v_i^{-1}\Gamma)h_1(u_i;\nu). \tag{6.8}$$

We can explicitate it, considering each situation: when the within-subject error is assumed to follow a multivariate t-distribution, (6.8) corresponds to a gamma distribution

$$p(u_i|\mathbf{y}_i,\mathbf{b}_i,v_i;\theta) = \text{Gamma}\left(\frac{n_i+\nu}{2}, \frac{\sigma^{-2}\|\mathbf{y}_i - f\|^2 + \nu}{2}\right) \tag{6.9}$$

and when the within-subject error is assumed to follow a multivariate slash distribution, (6.8) corresponds to a truncated gamma distribution

$$p(u_i|\mathbf{y}_i,\mathbf{b}_i,v_i;\boldsymbol{\theta}) = \text{Truncated-Gamma}\left(\frac{n_i}{2}+\nu, \frac{\boldsymbol{\sigma}^{-2}\|\mathbf{y}_i - f\|^2}{2}, T\right), \tag{6.10}$$

where we set $T=1$, representing the point at which the right truncation occurs. Similar calculations can be made for a variable v, following the steps in the equations (6.8) to (6.10).

In the second step, i.e, the stochastic approximation step, we update $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k+1)})$ according to

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k+1)}) \;=\; Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) + \gamma_{k+1}\left[\frac{1}{M}\sum_{m=1}^{M} l_c(\boldsymbol{\theta};\mathbf{y},\mathbf{z}^{(k+1,m)}) - Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})\right], \tag{6.11}$$

where γ_{k+1} is a smoothing parameter, a sequence decreasing to zero, of positive numbers, chosen to improve the convergence to parameter $\boldsymbol{\theta}$. The convergence of SAEM to a local maximum of the likelihood is proved by Delyon, Lavielle and Moulines (1999) under fairly general conditions. Here, the choice of the smoothing parameter is the same adopted by Kuhn and Lavielle (2005), given as $\gamma_k = 1$, for $1 \leq k \leq K$; $\gamma_k = \frac{1}{(k-K)}$, if $k \geq K+1$. The value of K is set as in Jank (2006).

When the complete data likelihood in (6.5) belongs to the exponential family, we can extract minimal sufficient statistics $S(\mathbf{y},\mathbf{z})$ such that in (6.11) is set to:

$$s^{(k)} \;=\; s^{(k-1)} + \gamma_k\left[\frac{1}{M}\sum_{m=1}^{M} S(\mathbf{y},\mathbf{z}^{(k,m)}) - s^{(k-1)}\right].$$

Then, the minimal sufficient statistics for variance components $\boldsymbol{\Gamma}$ and $\boldsymbol{\sigma}^2$ are given by, respectively:

$$s_{1,i}^{(k)} \;=\; s_{1,i}^{(k-1)} + \gamma_k\left[\frac{1}{M}\sum_{j=1}^{M} u_i^{(k,m)} J_{n_i}^{(k)\top} J_{n_i}^{(k)} - s_{1,i}^{(k-1)}\right]$$

and

$$s_{2,i}^{(k)} \;=\; s_{2,i}^{(k-1)} + \gamma_k\left[\frac{1}{M}\sum_{j=1}^{M} J_{n_i}^{(k)\top} u_i^{(k,m)}[\mathbf{Y}_i - f(\boldsymbol{\beta}^{(k)},\mathbf{b}_i^{(j)})] - s_{2,i}^{(k-1)},\right]$$

where $J_{n_i} = \partial f(\boldsymbol{\beta},\mathbf{b}_i)/\partial\boldsymbol{\beta}^\top$. Thus,

$$s_{3,i}^{(k)} \;=\; s_{3,i}^{(k-1)} + \gamma_k\left[\frac{1}{M}\sum_{m=1}^{M} v_i^{(k,m)}\mathbf{b}_i^{(k,m)}\mathbf{b}_i^{(k,m)T} - s_{3,i}^{(k-1)}\right],$$

and

$$s_{4,i}^{(k)} \;=\; s_{4,i}^{(k-1)} + \gamma_k\left[\frac{1}{M}\sum_{m=1}^{M} u_i^{(k,m)}\|\mathbf{y}_i - f(\boldsymbol{\beta}^{(k,m)},\mathbf{b}_i^{(k,m)})\|^2 - s_{4,i}^{(k-1)}\right].$$

The last step is the maximization step where $\theta^{(k)}$ is updated in a general form by $\theta^{(k+1)} = \arg\max Q(\theta|\theta^{(k)})$, and each parameter estimator is represented by

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^{(k+1)} &= \widehat{\boldsymbol{\beta}}^{(k)} + \left(\sum_{i=1}^{n} s_{1,i}^{(k)}\right)^{-1} \left(\sum_{i=1}^{n} s_{2,i}^{(k)}\right), \\
\widehat{\Gamma}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^{n} s_{3,i}^{(k)} \quad \text{and} \\
\widehat{\sigma}^{2(k+1)} &= \frac{1}{N}\sum_{i=1}^{n} s_{4,i}^{(k)}.
\end{aligned}$$

In M-step a Newton-Raphson method was used to obtain fixed-effects parameter estimates $\widehat{\beta}^{(k+1)}$.

6.3.1 Estimation of the likelihood function

For the likelihood estimation, we used the scheme of importance sampling considering the likelihood function of the observed data $\mathbf{y}$

$$l(\theta;\mathbf{y}) = \int f(;\theta)d\mathbf{b} = \int f(\mathbf{y}|\mathbf{b};\theta)f(\mathbf{b};\theta)d\mathbf{b}, \tag{6.12}$$

where $f(\mathbf{y}|\mathbf{b})$ and $f(\mathbf{b};\theta)$ are the densities of the models.

The scheme of importance sampling consists of rewriting the equation (6.12) as

$$l(\theta;\mathbf{y}) = \int f(\mathbf{y}|\mathbf{b};\theta)\frac{f(\mathbf{b};\theta)}{f^*(\mathbf{b};\theta)}f^*(\mathbf{b};\theta)\mathrm{d}\mathbf{b} \tag{6.13}$$

where f^* is any continuous distribution. Thus, we can generate $\mathbf{b}^* \sim f^*(b^*;\theta)$ and in a Monte Carlo scheme substitute $\mathbf{b}^*$ in the function

$$l(\theta;\mathbf{y}) = \frac{1}{M}\sum_{l=1}^{M} f(\mathbf{y}|\mathbf{b}_l;\theta)\frac{f(\mathbf{b}_l;\theta)}{f^*(\mathbf{b}_l;\theta)}. \tag{6.14}$$

6.4 Application

In this section, we present an application in a songbird data set.

6.4.1 Songbird - RA data

The songbird data were initially presented by Van der Linden *et al.* (2002) and Van Meir *et al.* (2004), who established a novel in vivo magnetic resonance imaging (MRI) approach to discern the functional characteristics of specific neuronal populations in a strongly connected brain circuitry, the so-called song control system in the songbird brain. The high vocal center (HVC), one of the major nuclei in this circuit, contains interneurons and two distinct types of neurons projecting respectively to the so-called

nucleus robustus arcopallii (RA) or to area X. Molenberghs and Verbeke (2005), chap. 20, analysed this data set by way of an NLME assuming normal distribution. After selecting the model, they settled on the following model for signal intensity (SI) in RA.

$$SI_{ij}(RA) = \frac{(\phi_0 + f_i)T_{ij}^{\eta_0+\eta_1 G_i}}{(\tau_0 + t_i)^{\eta_0+\eta_1 G_i} + T_{ij}^{\eta_0+\eta_1 G_i}} + \varepsilon_{ij} \tag{6.15}$$

where, $SI_{ij}(RA)$ is the measurement at time j for bird i, G_i is an indicator for group membership (1 for testosterone treated birds and 0 otherwise), and T_{ij} is the measurement time. The fixed effects are described as: ϕ_0 is the maximal signal intensity, sometimes called SI_{max}, for an untreated bird; τ_0 is the time required to reach 50% of this maximum (T_{50}); η_0 and $\eta_0 + \eta_1$ govern the shape of the curve.

The vector (f_i, t_i) is a bird-specific vector of random effects, assumed to follow a bivariate normal distribution with mean 0 and covariance matrix Γ. Here, we consider ε_{ij} random error following SMN distributions. Table 15 shows the parameter estimations for the model in (6.15).

Although many structures for variance-covariance matrix could be assumed for the random effects, in this analysis we use a diagonal variance-covariance matrix.

The metric used to identify possible outliers, assuming SMN distributions, was the usual Mahalanobis distance Russo, Paula and Aoki (2009). Regarding the distance of Mahalanobis we rewrote it as,

$$MD_1 = \frac{\|\mathbf{y}_i - f_i(X_i, \boldsymbol{\beta}, Z_i, \mathbf{b}_i)\|^2}{\sigma^2}$$

when we consider identifying atypical points due to within-subject error, $\boldsymbol{\varepsilon}-$outliers and

$$MD_2 = \mathbf{b}_i^\top \Gamma^{-1} \mathbf{b}_i$$

to consider identifying atypical points due to random effects, $\mathbf{b}-$outliers. In the case where the random effects and errors are normally distributed, MD_1 follows a chi-squared distribution with n_i degrees of freedom and MD_2 follows a chi-squared distribution with q degrees of freedom. In the t and slash cases, the Mahalanobis distance distribution does not have an expression that refers to specific known distribution, but the expression in terms of cumulative distributions functions can be seen in Lange and Sinsheimer (1993).

As seen previously, a great Mahalanobis distance indicates possible outlying or even influence observations. These points are related to lower weights in SMN distributions, as it can be noted in Table 16.

Considering the mentioned distances and the benchmarks corresponding to the 97,5%-quantiles, according to Figure 18, assuming normal distribution in errors and random effects, for the random effects observation 6 stands out the benchmark. For the error

term, two observations stand above the benchmark, cases 6 and 7. Figure 19, a normal Q-Q plot can be observed. Notice one outstanding point in the first plot. In the second plot, two more outstanding points are observed, possibly influential points. Given this, we use other t and slash distributions for random effects.

In Figures 20–27 one can observe Mahalanobis distances assuming normal, t, and slash combinations for errors and random effects.

For the degrees of freedom choice in the SMN distributions, we considered a grid of predefined values and select the one that leads to the low AIC (Burnham and Anderson (2002)), as discussed by Russo, Paula and Aoki (2009) and Lucas (1997).

The model parameters were estimated via maximum likelihood method using SAEM algorithm, and standard errors using Kuhn and Lavielle (2005).

To select the most adequate model, among the proposed ones, AIC and BIC were used. Both criteria are based on the minimization of Kullback-Leibler distance Akaike (1973) and Schwarz (1978). The model with the lowest AIC is the model assuming slash$_2$ / t_4 distribution. Note that the model with $t_{2.5}$ / $t_{3.5}$ has the second lowest AIC. Now, relating these choices to the respective Mahalanobis distance, we can see that the slash$_2$ / t_4 model still has an out point for the random error unlike the $t_{2.5}$ / $t_{3.5}$ model which has no points outside.

Table 15 – Maximum likelihood estimates for the songbird data: SI-RA, using multivariate normal (N), slash (Sl) and t distribution.

Par.	N / N	$Sl_{2.5}$ / Sl_2	$t_{2.5}$ / $t_{3.5}$
ϕ_0	0.42859 (0.02391)	0.42621 (0.00018)	0.37293 (0.01681)
η_0	2.18154 (0.08593)	2.18015 (0.01942)	2.18548 (0.15735)
η_1	0.43263 (0.10042)	0.33332 (0.00313)	0.30579 (0.64117)
τ_0	2.87180 (0.15725)	2.89418 (0.00138)	2.75117 (0.02536)
d_{11}	0.02061 (0.00807)	0.01406 (0.00485)	0.00946 (0.00801)
d_{22}	0.25282 (0.11069)	0.15552 (0.05828)	0.07163 (0.03347)
σ^2	0.00019 (0.00002)	0.00007 (0.00003)	0.00015 (0.00009)
LogLik	700.1577	725.1677	726.6103
AIC	-1386.315	-1436.335	-1439.221
BIC	-1384.197	-1434.217	-1437.102
Par.	N / Sl_2	N / $t_{3.5}$	$Sl_{3.5}$ / N
ϕ_0	0.43345 (0.00474)	0.40891 (0.01753)	0.41053 (0.16478)
η_0	2.17925 (0.06067)	2.15301 (0.66766)	2.18521 (0.10784)
η_1	0.32712 (0.15368)	0.35443 (0.65726)	0.42788 (0.63953)
τ_0	2.90442 (0.01456)	2.92751 (0.01993)	2.84357 (0.14552)
d_{11}	0.02145 (0.00764)	0.02303 (0.00219)	0.01651 (0.00165)
d_{22}	0.26745 (0.10139)	0.26766 (0.01306)	0.18152 (0.01142)
σ^2	0.00006 (0.00001)	0.00011 (0.00004)	0.00019 (0.00005)
LogLik	725.4588	726.4636	700.0004
AIC	-1436.918	-1438.927	-1386.001
BIC	-1434.8	-1436.809	-1383.883
Par.	t_3 / N	Sl_2 / t_4	t_3 / Sl_2
ϕ_0	0.42124 (0.00871)	0.39394 (0.01630)	0.40906 (0.03105)
η_0	2.19140 (0.08836)	2.15482 (0.71667)	2.17737 (0.43392)
η_1	0.40790 (0.15432)	0.34607 (0.20199)	0.32332 (0.26923)
τ_0	2.87325 (0.03194)	2.87353 (0.00906)	2.90700 (0.03041)
d_{11}	0.01240 (0.00356)	0.01331 (0.00441)	0.01254 (0.00478)
d_{22}	0.12992 (0.04006)	0.12534 (0.01810)	0.12810 (0.04327)
σ^2	0.00019 (0.00002)	0.00010 (0.00003)	0.00007 (0.00009)
LogLik	700.2355	726.9277	725.6607
AIC	-1386.471	-1439.855	-1437.321
BIC	-1384.353	-1437.737	-1435.203

Table 16 – Fitted weights considering $Slash_2$ and t_4 models in Songbird-RA data.

Bird	1	2	3	4	5	6	7	8	9	10
Random effects	0.76	0.70	0.72	0.61	0.74	0.67	**0.27**	0.53	0.74	0.75
Residual errors	1.83	0.89	1.93	0.58	0.82	**0.20**	**0.29**	0.66	1.42	1.50

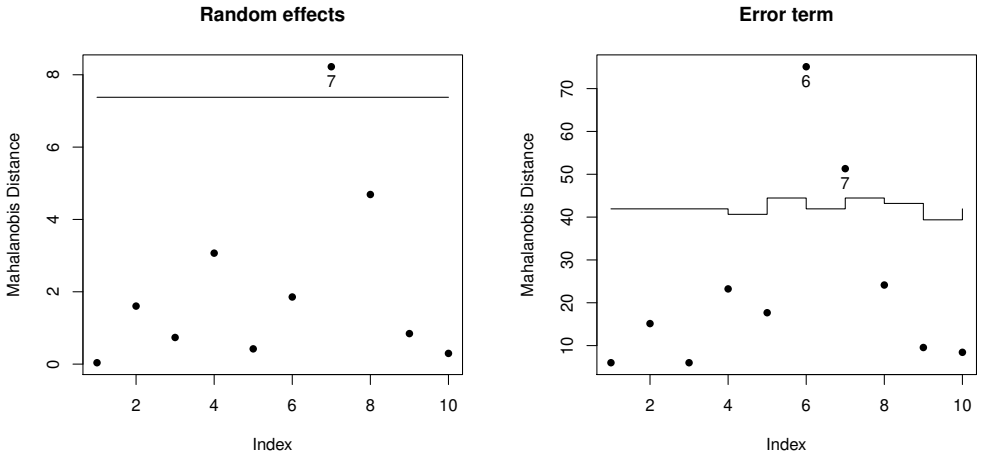


Figure 18 – Songbird SI-RA. Mahalanobis distance for random effects and error term, considering Normal/Normal Model.

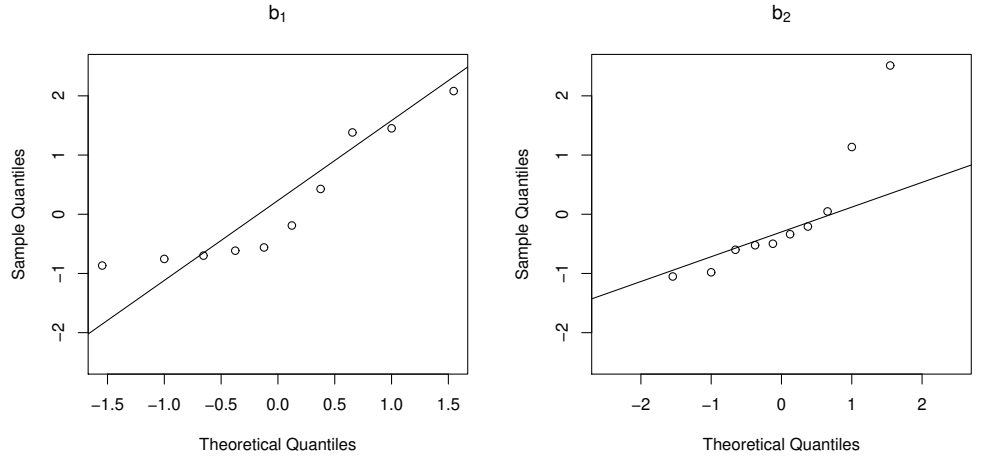


Figure 19 – Songbird SI-RA. Normal Q-Q plot for random effects b_i. Normal/Normal.

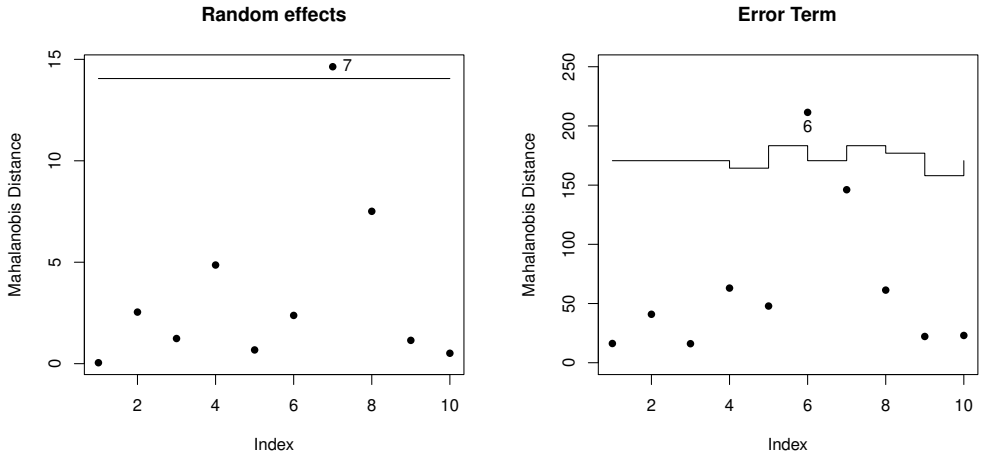


Figure 20 – Songbird SI-RA. Mahalanobis distance for random effects and error term, considering $\text{Slash}_{2.5}/\text{Slash}_2$.

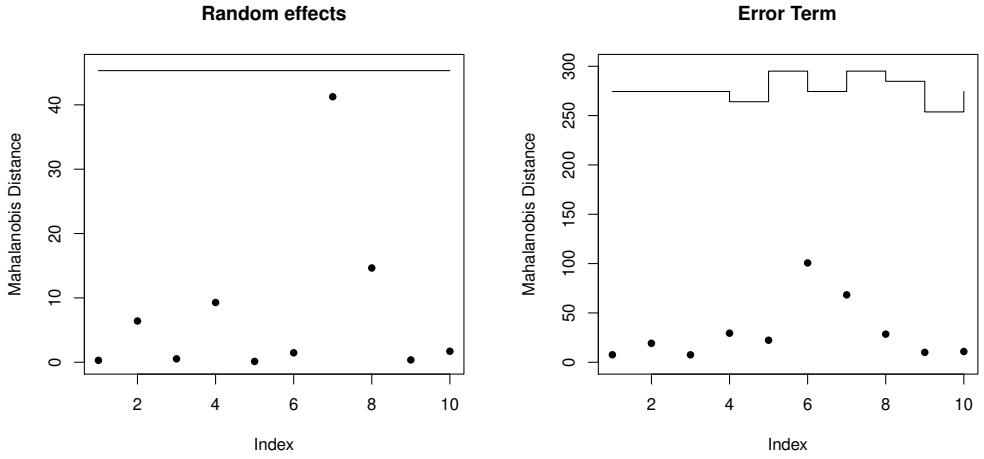


Figure 21 – Songbird SI-RA. Mahalanobis distance for random effects and error term, considering $t_{2.5}/t_{3.5}$.

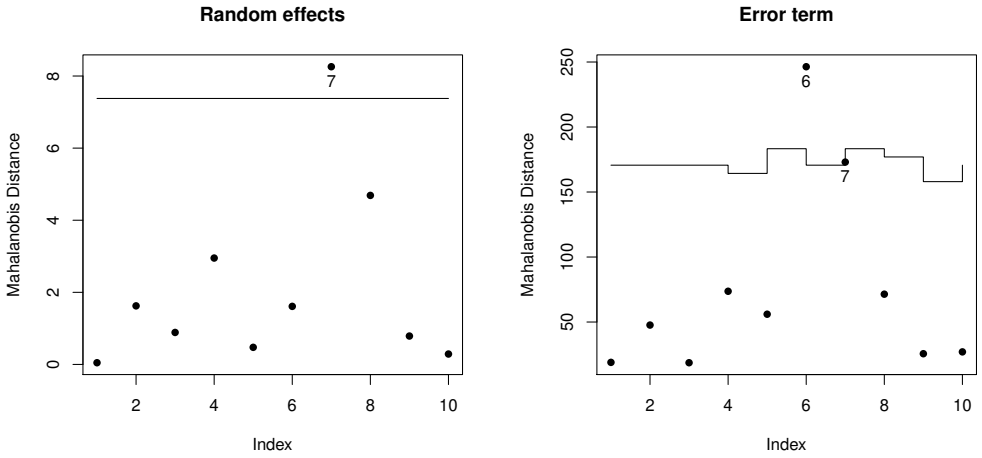


Figure 22 – Songbird SI-RA. Mahalanobis distance for random effects and error term, considering Normal/Slash$_2$.

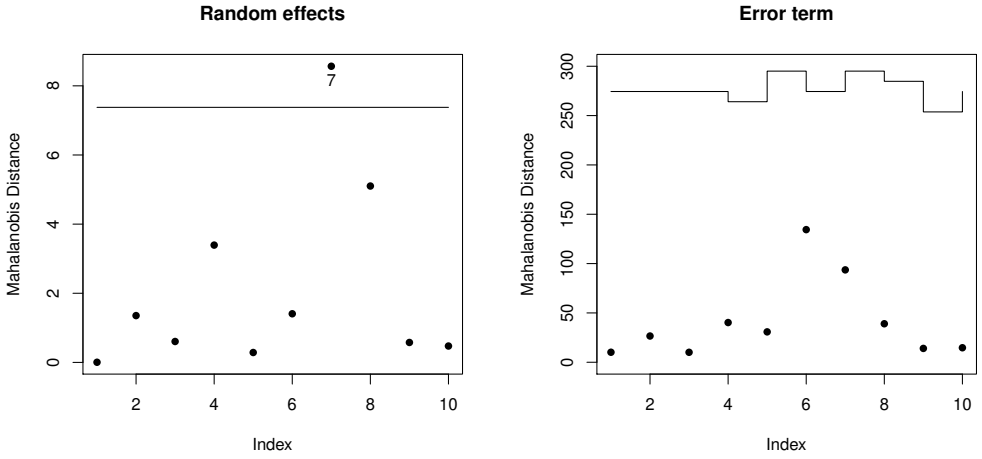


Figure 23 – Songbird SI-RA. Mahalanobis distance for random effects and error term, considering Normal/$t_{3.5}$.

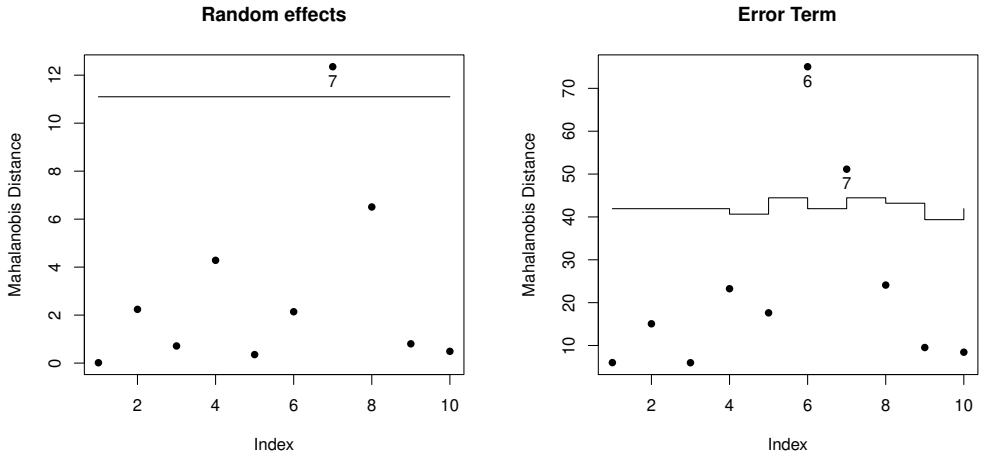


Figure 24 – Songbird SI-RA. Mahalanobis distance for random effects and error term, considering $\text{Slash}_{3.5}$/Normal.

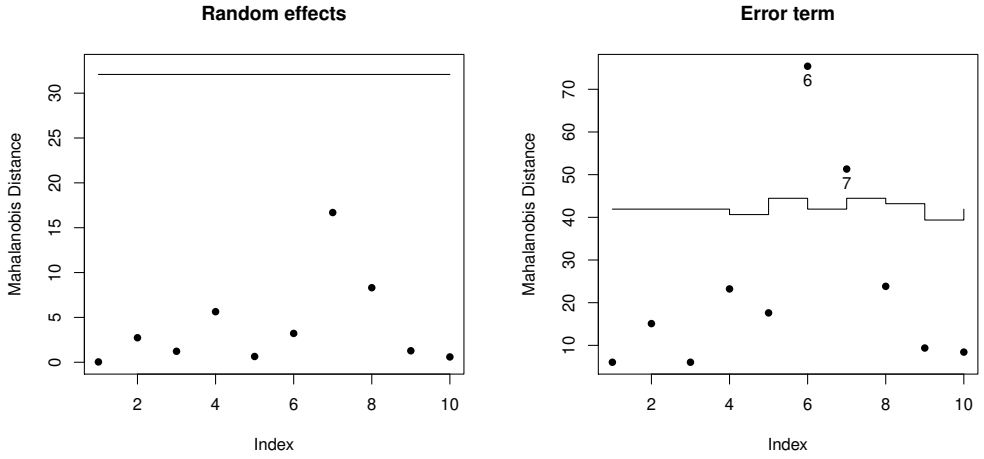


Figure 25 – Songbird SI-RA. Mahalanobis distance for random effects and error term, considering t_3/Normal.

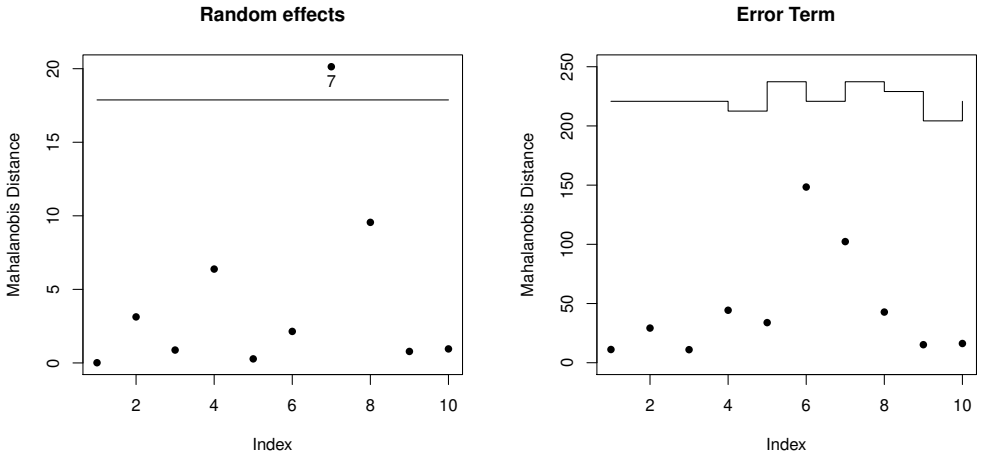


Figure 26 – Songbird SI-RA. Mahalanobis distance for random effects and error term, considering $\text{Slash}_2/\text{t}_4$.

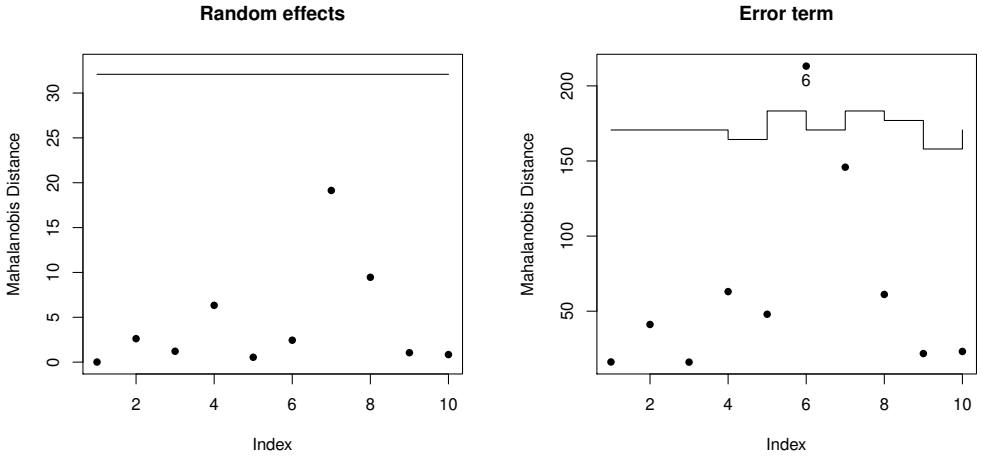


Figure 27 – Songbird SI-RA. Mahalanobis distance for random effects and error term, considering $\text{t}_3/\text{Slash}_2$.

6.5 Standard error estimation via bootstrap simulation

The bootstrap simulation method introduced by Efron (1979), consists of resampling the data with replacement a large number of times. There is a number of versions of bootstrap in parametric and nonparametric versions, each one with its particularities. In this thesis, a nonparametric version of bootstrap is used. In an adaptation of bootstrap, the notation for this work is as follows:

1. Organize the observed values in pairs, $(\mathbf{y}_i, \mathbf{X}_i), i = 1, \ldots, n$ and resample from these pairs, with replacements, into n new pairs. Repeat this a large number of times (B), such that the new pairs are $(\mathbf{y}^*_{i,1}, \mathbf{X}^*_{i,1}), \ldots, (\mathbf{y}^*_{i,B}, \mathbf{X}^*_{i,B}),\ i = 1, \ldots, n.$

2. Now estimate the fixed-effects and variance component parameters for each of the bootstrap samples, resulting in B estimates for each parameter, $\hat{\theta}^{*1}, \hat{\theta}^{*2}, \ldots, \hat{\theta}^{*B}$.

3. Estimate the parameters using equation (6.16) and find the standard error using equation (6.17).

The measures used to evaluate the fixed and random effects after bootstraping are the mean, the standard error and the percentile confidence interval.

The mean is computed as the average estimates from the resampled data

$$\overline{\hat{\theta}^*} = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}^{*b} \tag{6.16}$$

and the standard error is given by

$$\hat{se}(\hat{\theta}) = \left[\frac{\sum_{i=1}^{B} \left(\hat{\theta}^{*b} - \overline{\hat{\theta}^*} \right)^2}{B-1} \right]^{1/2}, \tag{6.17}$$

in which $\overline{\hat{\theta}^*} = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}^{*b}$.

The confidence interval based on the percentile method is used. This method takes into account the estimate distributions form. After generating the bootstrap estimates, they are ordered and the lower and upper percentiles are used to determine the percentile confidence limits. In a percentile confidence interval with 95% confidence level, limits are given by

$$(\hat{\theta}^*[0.025];\ \hat{\theta}^*[0.975]). \tag{6.18}$$

Notice that the lower limit corresponds to the 2.5% percentile and the upper limit to the 97.5% percentile. More details about the bootstrapping and percentile confidence interval may be found in Efron and Hastie (2016).

Table 17 shows summary statistics, the bootstrap mean, standard error (se) from equation (17) and percentile confidence interval with a 95% confidence level from equation (6.18). The same scenarios seen in Table 15 were used for the bootstrap simulations, considering normal, t and slash distribution. It can be observed that the parameters are significant in the normal distribution scenario as the confidence interval does not contain zero. In the scenarios for the slash and t distributions, a confidence interval for the parameter η_1 contains zero, showing to be statistically non-significant. This parameter represents the group effect.

Table 17 – Results of bootstrap simulation for songbird data, considering normal (N), t and slash (Sl) distributions.

Dist.	Par.	Mean	se	2.5%	97.5%
N/N	ϕ_0	0.42610	0.03382	0.36066	0.49479
	η_0	2.19670	0.08528	2.02953	2.39740
	η_1	0.42440	0.17521	0.07485	0.77414
	τ_0	2.87900	0.20068	2.55402	3.31096
	d_{11}	0.02140	0.00865	0.00538	0.04017
	d_{22}	0.24620	0.14886	0.01348	0.51878
	σ^2	0.00020	0.00005	0.00009	0.00030
$t_{2.5}$ / $t_{3.5}$	ϕ_0	0.39770	0.01603	0.37094	0.43458
	η_0	2.17360	0.07801	2.06454	2.36179
	η_1	0.33550	0.17076	-0.04195	0.73579
	τ_0	2.89900	0.17944	2.61660	3.29534
	d_{11}	0.02610	0.01178	0.00531	0.05159
	d_{22}	0.25820	0.16046	0.01017	0.56041
	σ^2	0.00010	0.00003	0.00007	0.00017
$Sl_{2.5}$ / Sl_2	ϕ_0	0.43310	0.03269	0.36785	0.49919
	η_0	2.18880	0.08068	2.06125	2.37838
	η_1	0.32870	0.16463	-0.03602	0.69720
	τ_0	2.90560	0.19145	2.60807	3.31878
	d_{11}	0.02180	0.00826	0.00530	0.03899
	d_{22}	0.25850	0.16113	0.00965	0.56459
	σ^2	0.00010	0.00002	0.00004	0.00011

6.6 Global Influence

The global influence was performed to evaluate the changes in the estimates when each observation was excluded from the data set. It can be observed in Table 18 that the larger percentual changes are caused in the variance components, namely d_{22}, when subjects 7 and 2 were individually excluded, and on d_{11} when subjects 8 and 4 were excluded.

Table 18 – Percentual changes in the estimates after excluding of individuals considering Slash_2 and t_4 models fitted to Songbird data.

Subject excluded	Dist.	ϕ_0	η_0	η_1	τ_0	d_{11}	d_{22}	σ^2
1	N; N	2.87	0.96	-5.64	0.18	7.23	12.43	5.26
	$t_{2.5}$; $t_{3.5}$	5.57	0.72	-3.89	2.47	44.19	75.15	-6.67
	$\text{Sl}_{2.5}$; Sl_2	-0.27	1.28	-7.18	-2.16	13.44	7.37	14.29
	Sl_2; t_4	-4.09	2.00	-15.26	-0.50	27.42	21.73	20.00
2	N; N	2.29	-1.21	5.91	-1.43	4.46	-5.98	0.00
	$t_{2.5}$; $t_{3.5}$	1.78	-2.46	18.83	-1.12	-3.81	-50.55	-26.67
	$\text{Sl}_{2.5}$; Sl_2	3.27	-2.46	14.99	-1.42	7.33	-19.79	-14.29
	Sl_2; t_4	2.74	-1.50	10.89	-2.36	0.75	-45.17	40.00
3	N; N	4.66	0.51	0.60	1.00	-0.49	7.45	5.26
	$t_{2.5}$; $t_{3.5}$	5.58	1.38	-13.42	1.16	18.71	27.64	-6.67
	$\text{Sl}_{2.5}$; Sl_2	4.07	1.48	-8.04	-0.73	2.70	-1.34	14.29
	Sl_2; t_4	3.50	2.10	-11.13	-1.44	-1.65	-12.04	30.00
4	N; N	-3.59	0.64	-2.74	1.67	-8.83	1.90	-5.26
	$t_{2.5}$; $t_{3.5}$	3.97	-0.87	8.97	1.84	-31.92	3.04	-40.00
	$\text{Sl}_{2.5}$; Sl_2	-1.47	-0.44	1.15	1.62	-12.87	-0.30	0.00
	Sl_2; t_4	-1.73	0.58	-2.11	0.99	-31.86	-9.71	30.00
5	N; N	-0.53	-0.05	-3.49	0.52	8.25	13.28	0.00
	$t_{2.5}$; $t_{3.5}$	9.31	-1.84	14.70	5.18	31.40	80.87	-26.67
	$\text{Sl}_{2.5}$; Sl_2	-0.21	-1.28	1.69	1.00	14.65	24.71	-14.29
	Sl_2; t_4	-1.74	-0.59	3.68	0.29	19.08	21.35	30.00
6	N; N	7.80	0.04	-13.17	4.87	-1.36	8.77	-26.32
	$t_{2.5}$; $t_{3.5}$	7.84	-1.52	8.83	4.74	0.42	5.24	-40.00
	$\text{Sl}_{2.5}$; Sl_2	6.48	-0.76	1.68	5.44	2.77	23.58	-14.29
	Sl_2; t_4	6.23	0.08	-4.50	1.13	1.50	1.91	0.00
7	N; N	-4.95	2.21	-31.91	-6.63	-3.20	-74.04	-15.79
	$t_{2.5}$; $t_{3.5}$	0.41	-0.31	-12.53	-0.21	-11.10	-53.43	-26.67
	$\text{Sl}_{2.5}$; Sl_2	-4.47	2.06	-32.44	-7.69	-2.63	-73.79	0.00
	Sl_2; t_4	-1.61	1.56	-23.02	-6.31	-3.98	-67.15	0.00
8	N; N	-4.79	0.29	26.54	-1.55	-36.00	1.67	-5.26
	$t_{2.5}$; $t_{3.5}$	1.62	-1.14	16.44	2.74	-33.83	49.31	-26.67
	$\text{Sl}_{2.5}$; Sl_2	-7.19	0.10	3.54	-1.33	-42.74	-14.38	-14.29
	Sl_2; t_4	-0.09	0.46	5.22	-0.44	-46.36	-3.81	20.00
9	N; N	1.44	-0.49	14.36	1.60	-0.19	6.30	5.26
	$t_{2.5}$; $t_{3.5}$	5.49	-0.69	29.68	4.93	15.75	71.56	-13.33
	$\text{Sl}_{2.5}$; Sl_2	4.76	-0.87	26.90	1.98	1.21	15.60	-14.29
	Sl_2; t_4	2.33	0.19	22.01	-0.23	7.29	20.72	50.00
10	N; N	2.39	-1.01	9.48	3.90	5.82	15.86	5.26
	$t_{2.5}$; $t_{3.5}$	6.98	-1.19	0.54	2.36	23.26	69.65	-26.67
	$\text{Sl}_{2.5}$; Sl_2	1.78	-0.16	-6.43	2.91	12.30	23.46	14.29
	Sl_2; t_4	2.76	-0.05	-11.25	-1.81	1.43	-0.94	10.00

6.7 Discussion

In this chapter, we implemented a general nonlinear mixed-effects model where random effects enter the model in a more flexible way. In the literature, for example, in

Pinheiro and Bates (2000) and Demidenko (2013), the so-called subject-specific vector, which accommodates fixed and random effects, assumes a linear function. Here we propose to make the entry of random effects more flexible so that the subject-specific vector can assume a non-linear function. The procedure for applying the fixed and random effects parameters was the maximum likelihood method using the SAEM algorithm. We adopted the methodology to estimate the NLME model parameters for songbird data studied by Molenberghs and Verbeke (2005). We used the selection criteria of AIC and BIC models in various distribution combinations for errors and random effects. The advantage of using the SMN distribution is that it gives different weights to observations, so that a low weight corresponds to a high Mahalanobis distance. The AIC (-1439.855) or BIC (-1437.737) information criteria suggest selecting the Sl_2 / t_4 model (random effects/errors), but it is still an indicator of the Mahalanobis distance corresponding to bird 7 which outstands the benchmark (see Figure 23), but the $t_{2.5}/t_{3.5}$ scenario (see Figure 21) records a slightly larger AIC (BIC) and does not show points from the Mahalanobis distance which outstands the benchmark. It still obtains standard errors through the non-parametric bootstrap simulation for scenarios N/N, $\mathrm{t}_{2.5}$/ $\mathrm{t}_{3.5}$ e $\mathrm{Sl}_{2.5}$/ Sl_2.

CHAPTER

7

DISCUSSION AND FUTURE RESEARCH

In this thesis, some estimation methods are proposed to obtain parameters' estimates in nonlinear mixed-effects models.

First, in Chapter 4, a comparison between a MCEM method and a Taylor-expansion approximate method is made to obtain maximum likelihood estimates. This study showed that the approximate method is much faster than Monte Carlo EM and can provide fairly good initial estimates for the second method.

A second study, in Chapter 5, was devoted to compare REML to ML estimates using the nlme R package and a stochastic-approximation EM (SAEM) method, to verify if the REML produces more accurate or less-biased estimates for the variance components in the NLMEMs. This study showed that in some cases the ML method underestimates the variance components.

The third research topic, in Chapter 6, consisted of proposing a general nonlinear mixed-effects model where no linear structure for random effects are considered, with random components following a scale mixture of normal distributions. This model is estimated via the SAEM algorithm and showed a superiority of heavy-tailed models over the normal distribution.

As future research, a more intensive diagnostic study is intended to be done, by considering local influence, for instance, Cook (1986) and Poon and Poon (2001). Other data sets will also be considered, to illustrate the general nonlinear mixed-models proposed in Chapter 6.

REML and the general mixed-effects models will be studied for skewed distributions in the scale-mixture of normal distributions.

BIBLIOGRAPHY

AKAIKE, H. Maximum likelihood identification of gaussian autoregressive moving average models. **Biometrika**, Oxford University Press, v. 60, n. 2, p. 255–265, 1973. Citations on pages 38 and 86.

ANDREWS, D. F.; MALLOWS, C. L. Scale mixtures of normal distributions. **Journal of the Royal Statistical Society, Series B (Statistical Methodology)**, v. 36, p. 99–102, 1974. Citation on page 35.

ARRIBAS-GIL, A.; BERTIN, K.; MEZA, C.; RIVOIRARD, V. Lasso-type estimators for semiparametric nonlinear mixed-effects models estimation. **Statistics and Computing**, Springer, v. 24, n. 3, p. 443–460, 2014. Citations on pages 59 and 68.

BENNETT, J. E.; WAKEFIELD, J. C. Markov chain Monte Carlo for nonlinear hierarchical models. **Technical Report TR-93-11, Statistics Section, Imperial College.**, 1993. Citation on page 26.

BURNHAM, K.; ANDERSON, D. **Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach**. [S.l.]: Springer Verlag, 2002. Citation on page 86.

CHEN, R. Bayesian inference on mixed-effects models with skewed distributions for hiv longitudinal data. **University of South Florida**, 2012. Citation on page 26.

COOK, D. Assessment of local influence. **Journal of the Royal Statistical Society - Series B**, v. 48, p. 133–169, 1986. Citation on page 97.

DAVIDIAN, M.; GILTINAN, D. M. **Nonlinear Models for Repeated Measurement Data**. [S.l.]: London: Chapman and Hall, 1995. Citations on pages 25 and 50.

_____. Nonlinear models for repeated measurement data: an overview and update. **Journal of Agricultural, Biological, and Environmental Statistics**, Springer, v. 8, n. 4, p. 387–419, 2003. Citation on page 25.

De La Cruz, R. Bayesian analysis for nonlinear mixed-effects models under heavy-tailed distributions. **Pharmaceutical statistics**, Wiley Online Library, v. 13, n. 1, p. 81–93, 2014. Citations on pages 26, 79, and 80.

DELYON, B.; LAVIELLE, M.; MOULINES, E. Convergence of a stochastic approximation version of the EM algorithm. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 27, n. 1, p. 94–128, 1999. Citations on pages 25, 71, and 83.

DEMIDENKO, E. **Mixed models: theory and applications with R**. [S.l.]: John Wiley & Sons, 2013. Citations on pages 25 and 96.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximun likelihood from incomplete data via the EM algorithm (with discussion). **Journal of the Royal Statistical Society,**

Series B (Statistical Methodology), v. 39, p. 1–38, 1977. Citations on pages 41 and 67.

DOORNIK, J. A. An object-oriented matrix programming language ox 6. Timberlake Consultants Ltd, 2009. Citation on page 55.

EFRON, B. Bootstrap methods: another look at the jackknife. **The Annals of Statistics**, v. 7, p. 1–26, 1979. Citation on page 93.

EFRON, B.; HASTIE, T. **Computer age statistical inference**. [S.l.]: Cambridge University Press, 2016. Citation on page 93.

FOULLEY, J. L.; QUAAS, R. Heterogeneous variances in gaussian linear mixed models. **Genetics Selection Evolution**, EDP Sciences, v. 27, n. 3, p. 211–228, 1995. Citation on page 68.

GELFAND, A.; HILLS, S.; RACINE-POON, A.; SMITH, A. Illustration of Bayesian inference in normal data models using Gibbs sampling. **Journal of the American Statistical Association**, v. 85, p. 972–985, 1990. Citation on page 43.

GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. **Statistical science**, JSTOR, p. 457–472, 1992. Citation on page 55.

GILKS, W. R.; RICHARDSON, S.; SPIEGELHALTER, D. J. **Markov Chain Monte Carlo In Practice**. [S.l.]: London: Chapman and Hall, 1996. Citations on pages 42 and 43.

HARVILLE, D. A. Bayesian inference for variance components using only error contrasts. **Biometrika**, Biometrika Trust, v. 61, n. 2, p. 383–385, 1974. Citations on pages 60, 68, and 77.

JANK, W. Implementing and diagnosing the stochastic approximation EM algorithm. **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 15, n. 4, p. 803–829, 2006. Citations on pages 71 and 83.

KAZEMI, I.; MAHDIYEH, Z.; MANSOURIAN, M.; PARK, J. J. Bayesian analysis of multivariate mixed models for a prospective cohort study using skew-elliptical distributions. **Biometrical Journal**, Wiley Online Library, v. 55, n. 4, p. 495–508, 2013. Citation on page 26.

KUHN, E.; LAVIELLE, M. Coupling a stochastic approximation version of EM with an MCMC procedure. **ESAIM: Probability and Statistics**, EDP Sciences, v. 8, p. 115–131, 2004. Citation on page 69.

_____. Maximum likelihood estimation in nonlinear mixed effects models. **Computational Statistics & Data Analysis**, v. 49, p. 1020–1038, 2005. Citations on pages 25, 42, 71, 78, 83, and 86.

LACHOS, V. H.; CASTRO, L. M.; DEY, D. K. Bayesian inference in nonlinear mixed-effects models using normal independent distributions. **Computational Statistics & Data Analysis**, Elsevier, v. 64, p. 237–252, 2013. Citations on pages 26, 79, and 80.

LAIRD, N. M.; WARE, J. H. Random-effects models for longitudinal data. **Biometrics**, JSTOR, p. 963–974, 1982. Citations on pages 66 and 67.

LANGE, K.; SINSHEIMER, J. S. Normal/independent distributions and their applications in robust regression. **Journal of Computational and Graphical Statistics**, v. 2, p. 175–198, 1993. Citations on pages 35 and 85.

LEE, S. Y.; XU, L. Influence analyses of nonlinear mixed-effects models. **Computational Statistics & Data Analysis**, v. 45, p. 321–341, 2004. Citation on page 39.

LIAO, J.; LIPSITZ, S. R. A type of restricted maximum likelihood estimator of variance components in generalised linear mixed models. **Biometrika**, Biometrika Trust, v. 89, n. 2, p. 401–409, 2002. Citation on page 60.

LINDSTROM, M. J.; BATES, D. M. Nonlinear mixed effects models for repeated measures data. **Biometrics**, v. 46, p. 673–687, 1990. Citations on pages 25 and 60.

LOUIS, T. A. Finding the observed information matrix when using the EM algorithm. **Journal of the Royal Statistical Society. Series B (Methodological)**, JSTOR, p. 226–233, 1982. Citation on page 46.

LUCAS, A. Robustness of the student-t based m-estimator. **Communications in Statistics - Theory and Methods**, v. 26, p. 1165–1182, 1997. Citations on pages 41 and 86.

MEZA, C.; JAFFRÉZIC, F.; FOULLEY, J.-L. REML estimation of variance parameters in nonlinear mixed effects models using the SAEM algorithm. **Biometrical Journal**, Wiley Online Library, v. 49, n. 6, p. 876–888, 2007. Citations on pages 25, 59, 60, 68, and 78.

MEZA, C.; OSORIO, F.; De La Cruz, R. Estimation in nonlinear mixed-effects models using heavy-tailed distributions. **Statistics and Computing**, Springer, v. 22, p. 121–139, 2012. Citations on pages 26, 29, 40, 59, 69, 73, 78, 79, 80, and 82.

MOLENBERGHS, G.; VERBEKE, G. **Models for discrete longitudinal data**. [S.l.]: New York, Springer, 2005. Citations on pages 29, 32, 79, 80, 85, and 96.

OSORIO, F.; PAULA, G. A.; GALEA, M. Assessment of local influence in elliptical linear models with longitudinal structure. **Computational Statistics & Data Analysis**, v. 51, p. 4354–4368, 2007. Citations on pages 46 and 55.

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, Biometrika Trust, v. 58, n. 3, p. 545–554, 1971. Citations on pages 60 and 68.

PAULA, G. A.; MEDEIROS, M.; VILCA-LABRA, F. E. Influence diagnostics for linear models with first-order autoregressive elliptical errors. **Statistics and Probability Letters**, 79, p. 339–346, 2009. Citations on pages 46 and 55.

PEREIRA, M. A. A.; RUSSO, C. M. Nonlinear mixed-effects models with scale mixture of skew-normal distributions. **Journal of Applied Statistics**, Taylor & Francis, v. 46, n. 9, p. 1602–1620, 2019. Citation on page 79.

PINHEIRO, J. C.; BATES, D. M. Approximations to the log likelihood function in the nonlinear mixed-effects model. **Journal of Computational and Graphical Statistics**, v. 4, p. 12–35, 1995. Citations on pages 25, 29, 59, 62, and 64.

_____. **Mixed-Effects Models in S and S-Plus**. [S.l.]: Springer, 2000. Citations on pages 25, 29, 30, 50, 53, 59, 60, 61, 62, 72, 77, and 96.

POON, W. Y.; POON, Y. S. Conditional local influence in case-weights linear regression. **Brit J Math Stat Psy**, v. 54, p. 177–191, 2001. Citation on page 97.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2016. Available: <https://www.R-project.org/>. Citations on pages 55 and 78.

RAFTERY, A. E. Bayesian model selection in social research. **Sociological Methodology**, v. 25, p. 111–163, 1995. Citation on page 38.

ROSA, G.; PADOVANI, C. R.; GIANOLA, D. Robust linear mixed models with normal/independent distributions and bayesian MCMC implementation. **Biometrical Journal: Journal of Mathematical Methods in Biosciences**, Wiley Online Library, v. 45, n. 5, p. 573–590, 2003. Citation on page 80.

RUSSO, C. M. **Modelos não lineares elípticos para dados correlacionados**. Phd Thesis (PhD Thesis) — IME - USP, 2010. Citation on page 39.

RUSSO, C. M.; AOKI, R.; PAULA, G. A. Assessment of variance components in nonlinear mixed-effects elliptical models. **TEST (Madrid)**, v. 21, p. 519–545, 2012. Citations on pages 46 and 55.

RUSSO, C. M.; GOMES, J. C. B.; LACHOS, V. H.; AOKI, R.; PAULA, G. A. Fast inference for robust nonlinear mixed–effects models. **Submitted to Journal of Applied Statistics**, 2018. Citation on page 39.

RUSSO, C. M.; LESAFFRE, E.; PAULA, G. A. A penalized elliptical mixture partially nonlinear mixed effects model. **In: Proceedings of the 27th International Workshop on Statistical Modelling, Prague, Czech Republic.**, p. 711–715, 2012. Citation on page 79.

RUSSO, C. M.; PAULA, G. A.; AOKI, R. Influence diagnostics in nonlinear mixed-effects elliptical models. **Computational Statistics and Data Analysis**, v. 53, n. 12, p. 4143–4156, 2009. Citations on pages 25, 80, 85, and 86.

SAVALLI, C.; PAULA, G. A.; CYSNEIROS, F. J. A. Assessment of variance components in elliptical linear mixed models. **Statistical Modelling**, v. 6, p. 59–76, 2006. Citation on page 80.

SCHWARZ, G. Estimating the dimension of a model. **The annals of statistics**, Institute of Mathematical Statistics, v. 6, n. 2, p. 461–464, 1978. Citations on pages 38 and 86.

SEARLE, S. R.; CASELLA, G.; MCCULLOCH, C. E. **Variance components**. [S.l.]: John Wiley & Sons, 2009. Citation on page 66.

SERROYEN, J.; MOLENBERGHS, G.; VERBEKE, G.; DAVIDIAN, M. Nonlinear models for longitudinal data. **The American Statistician**, Taylor & Francis, v. 63, n. 4, p. 378–388, 2009. Citations on pages 17, 29, and 32.

STAUDENMAYER, J.; LAKE, E.; WAND, M. Robustness for general design mixed models using the t-distribution. **Statistical Modelling**, SAGE Publications Sage India: New Delhi, India, v. 9, n. 3, p. 235–255, 2009. Citation on page 80.

TAN, M.; TIAN, G. L.; FANG, H. B. An efficient MCEM algorithm for fitting generalized linear mixed models for correlated binary data. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 77, n. 11, p. 929–943, 2007. Citation on page 46.

Van der Linden, A.; VERHOYE, M.; Van Meir, V.; TINDEMANS, I.; EENS, M.; ABSIL, P.; BALTHAZART, J. In vivo manganese-enhanced magnetic resonance imaging reveals connections and functional properties of the songbird vocal control system. **Neuroscience**, Elsevier, v. 112, n. 2, p. 467–474, 2002. Citations on pages 31 and 84.

Van Meir, V.; VERHOYE, M.; ABSIL, P.; EENS, M.; BALTHAZART, J.; Van der Linden, A. Differential effects of testosterone on neuronal populations and their connections in a sensorimotor brain nucleus controlling song production in songbirds: a manganese enhanced-magnetic resonance imaging study. **Neuroimage**, Elsevier, v. 21, n. 3, p. 914–923, 2004. Citations on pages 31 and 84.

VONESH, E. F.; CHINCHILLI, V. M. **Linear and Nonlinear Models for the Analysis of Repeated Measurements.** [S.l.]: Marcel Dekker. New York, NY, USA, 1997. Citation on page 25.

WAKEFIELD, J. The bayesian analysis of population pharmacokinetic models. **Journal of the American Statistical Association**, Taylor & Francis, v. 91, n. 433, p. 62–75, 1996. Citation on page 26.

WAKEFIELD, J.; SMITH, A.; RACINE-POON, A.; GELFAND, A. Bayesian analysis of linear and non-linear population models by using the gibbs sampler. **Applied Statistics**, JSTOR, p. 201–221, 1994. Citation on page 26.

WALKER, S. An EM algorithm for nonlinear random effects models. **Biometrics**, v. 52, p. 934–944, 1996. Citations on pages 25, 40, and 42.

WANG, J. EM algorithms for nonlinear mixed effects models. **Computational Statistics & Data Analysis**, v. 51, p. 3244–3256, 2007. Citations on pages 25 and 44.

WEI, G. C. G.; TANNER, M. A. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithm. **Journal of the American Statistical Association**, v. 85, p. 699–704, 1990. Citations on pages 25 and 42.

WOLFINGER, R. Laplace's approximation for nonlinear mixed models. **Biometrika**, v. 80, p. 791–795, 1993. Citation on page 45.

WOLFINGER, R. D.; LIN, X. Two Taylor-series approximation methods for nonlinear mixed models. **Computational Statistics & Data Analysis**, Elsevier, v. 25, n. 4, p. 465–490, 1997. Citations on pages 62 and 64.

WU, L. Exact and approximate inferences for nonlinear mixed-effects models with missing covariates. **Journal of the American Statistical Association**, v. 99, p. 700–709, 2004. Citations on pages 42 and 44.

ZHOU, M. **Fully exponential Laplace approximation EM algorithm for nonlinear mixed effects models**. Phd Thesis (PhD Thesis) — Graduate College, University of Nebraska, 2009. Citations on pages 62, 64, and 67.

APPENDIX

A

OTHER SCENARIOS FOR THE REML

This appendix shows a summary of the statistics and figures from other scenarios in the simulation study. In scenario 2, we used:

A.1 Scenario Two

Table 19 – Simulation results for the pharmacokinetic model with theoretical fixed-effects parameters $(lC_{\mathrm{l}}, lK_{\mathrm{a}}, lK_{\mathrm{e}})^{\top} = (-3.0, 0.5, -2)^{\top}$ and theoretical variance components$(\tau^{\top}, \sigma^2)_{\mathrm{l}}^{\top} = (0.2, 0.75, 0.1)^{\top}$.

		ML			REML		
Dist.	Par.	Mean	Bias	RMSE	Mean	Bias	RMSE
Normal	lC_{l}	-2.9908	0.0092	0.0471	-2.9660	0.0340	0.0702
	lK_{a}	0.5118	0.0118	0.1838	0.5139	0.0139	0.1322
	lK_{e}	-2.5017	-0.0017	0.0266	-2.5007	-0.0007	0.0278
	τ_1	0.1797	-0.0203	0.0781	0.2296	0.0296	0.0956
	τ_2	0.6977	-0.0523	0.3258	0.7613	0.0113	0.3423
	σ^2	0.0982	-0.0018	0.0139	0.0995	-0.0005	0.0141
t_4	lC_{l}	-3.0037	-0.0037	0.0441	-2.9816	0.0184	0.0690
	lK_{a}	0.5133	0.0133	0.1966	0.5222	0.0222	0.1554
	lK_{e}	-2.5029	-0.0029	0.0279	-2.5035	-0.0035	0.0285
	τ_1	0.1506	-0.0494	0.0879	0.1771	-0.0229	0.0776
	τ_2	0.6058	-0.1442	0.3555	0.6797	-0.0703	0.3508
	σ^2	0.0930	-0.0070	0.0233	0.0941	-0.0059	0.0230
$Slash_4$	lC_{l}	-2.9984	0.0016	0.0514	-2.9813	0.0187	0.0695
	lK_{a}	0.5140	0.0140	0.2051	0.5193	0.0193	0.1650
	lK_{e}	-2.5037	-0.0037	0.0303	-2.5044	-0.0044	0.0301
	τ_1	0.1674	-0.0326	0.0792	0.1979	-0.0021	0.0782
	τ_2	0.6780	-0.0720	0.3370	0.7449	-0.0051	0.3496
	σ^2	0.0981	-0.0019	0.0154	0.0990	-0.0010	0.0154

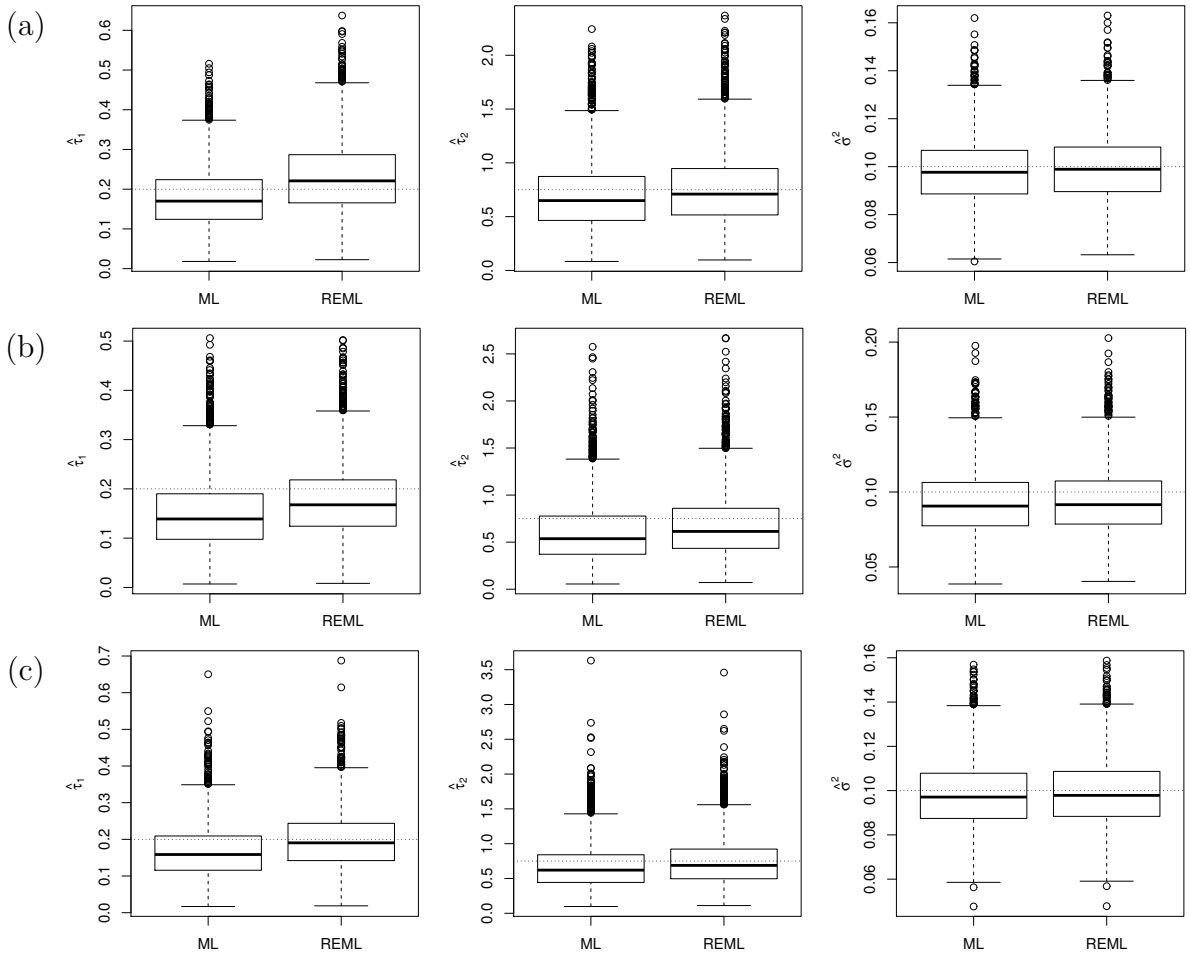


Figure 28 – Boxplots of the variance component estimates obtained from the simulation study. Panels in line (a) correspond to normal distribution, panels in line (b) represents t_4 distribution and in line (c) the slash$_4$ distribution. The horizontal dotted lines are the theoretical values.

A.2 Scenario Three

Table 20 – Simulation results for the pharmacokinetic model with theoretical fixed-effects parameters $(lCl, lKa, lKe)^\top = (-3.0, 0.5, -2.0)^\top$ and theoretical variance components $(\tau^\top, \sigma^2)_1^\top = (0.04, 0.15, 0.1)^\top$.

		ML			REML		
Dist.	Par.	Mean	Bias	RMSE	Mean	Bias	RMSE
Normal	lCl	-2.9977	0.0023	0.0350	-2.9870	0.0130	0.0387
	lKa	0.4980	-0.0020	0.0895	0.4988	-0.0012	0.0817
	lKe	-2.5001	-0.0001	0.0148	-2.5001	-0.0001	0.0149
	τ_1	0.0381	-0.0019	0.0126	0.0443	0.0043	0.0174
	τ_2	0.1427	-0.0073	0.0493	0.1507	0.0007	0.0514
	σ^2	0.0997	-0.0003	0.0072	0.0999	-0.0001	0.0072
t_4	lCl	-3.0035	-0.0035	0.0327	-2.9864	0.0136	0.0380
	lKa	0.5017	0.0017	0.1016	0.5040	0.0040	0.0832
	lKe	-2.5007	-0.0007	0.0152	-2.5030	-0.0030	0.0181
	τ_1	0.0358	-0.0042	0.0155	0.0469	0.0069	0.0204
	τ_2	0.1326	-0.0174	0.0596	0.1474	-0.0026	0.0627
	σ^2	0.0988	-0.0012	0.0181	0.0999	-0.0001	0.0179
$Slash_4$	lCl	-3.0002	-0.0002	0.0384	-2.9861	0.0139	0.0407
	lKa	0.4982	-0.0018	0.1049	0.5006	0.0006	0.0915
	lKe	-2.5007	-0.0007	0.0167	-2.5016	-0.0016	0.0169
	τ_1	0.0378	-0.0022	0.0135	0.0453	0.0053	0.0182
	τ_2	0.1428	-0.0072	0.0520	0.1529	0.0029	0.0552
	σ^2	0.0998	-0.0002	0.0093	0.1000	0.0000	0.0093

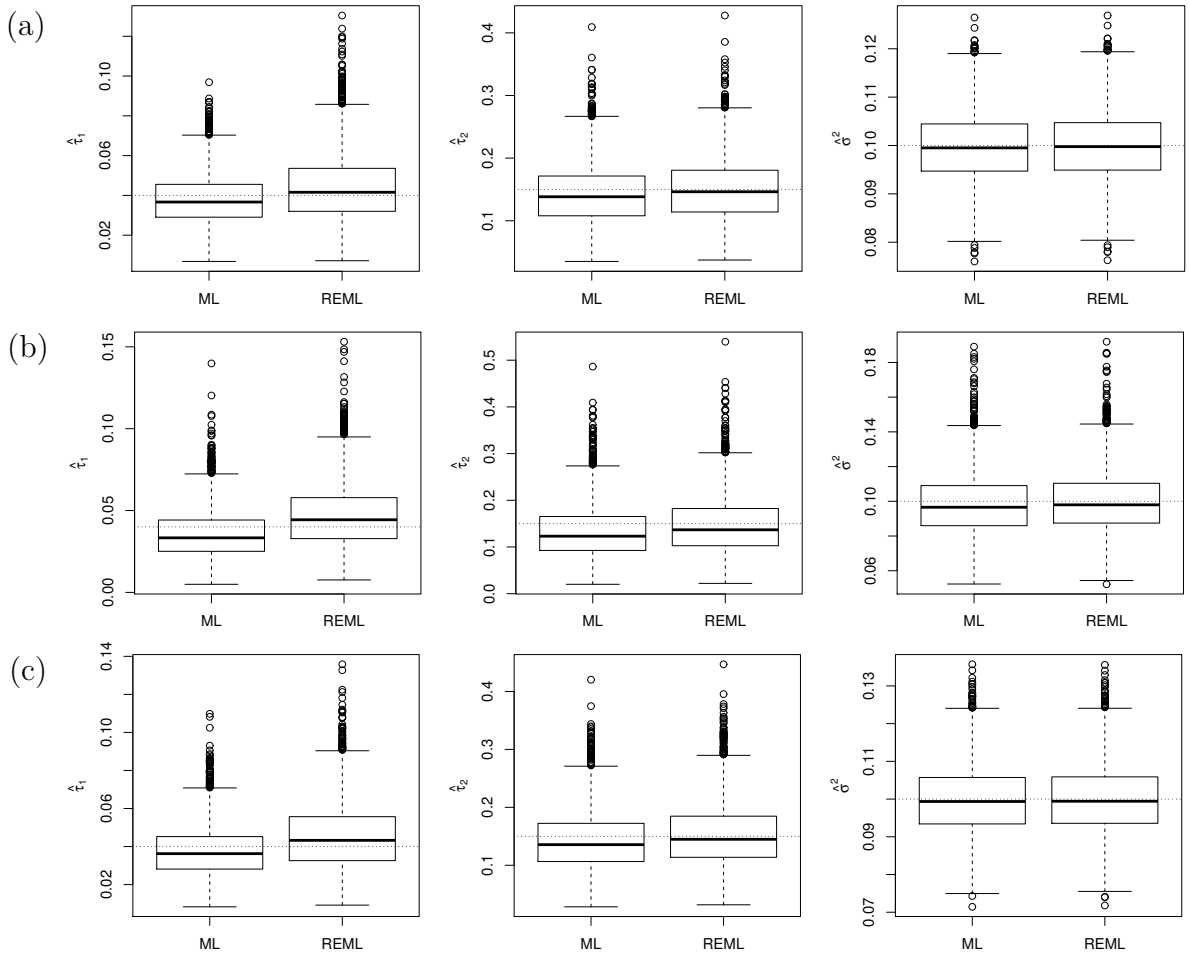


Figure 29 – Boxplots of the variance component estimates obtained from the simulation study. Panels in line (a) correspond to normal distribution, panels in line (b) represent t_4 distribution and in line (c) the slash$_4$ distribution. The horizontal dotted lines are the theoretical values.

A.3 Scenario Four

Table 21 – Simulation results for the pharmacokinetic model with theoretical fixed-effects parameters $(lCl, lKa, lKe)^\top = (-3.0, 0.5, -2.0)^\top$ and theoretical variance components $(\tau^\top, \sigma^2)_1^\top = (0.04, 0.15, 0.1)^\top$.

		ML			REML		
Dist.	Par.	Mean	Bias	RMSE	Mean	Bias	RMSE
Normal	lCl	-3.0015	-0.0015	0.0925	-2.9997	0.0003	0.0925
	lKa	0.5056	0.0056	0.1915	0.5066	0.0066	0.1912
	lKe	-2.5021	-0.0021	0.0511	-2.5017	-0.0017	0.0512
	τ_1	0.0317	-0.1683	0.1700	0.0398	-0.0002	0.0298
	τ_2	0.1205	-0.6295	0.6378	0.1550	0.0050	0.1330
	σ^2	0.0962	-0.0038	0.0280	0.0997	-0.0003	0.0288
t_4	lCl	-3.0053	-0.0053	0.0991	-3.0009	-0.0009	0.1022
	lKa	0.4979	-0.0021	0.2180	0.5007	0.0007	0.2172
	lKe	-2.5066	-0.0066	0.0639	-2.5046	-0.0046	0.0578
	τ_1	0.0339	-0.0061	0.0338	0.0445	0.0045	0.0419
	τ_2	0.1260	-0.0240	0.1488	0.1756	0.0256	0.2111
	σ^2	0.1030	0.0030	0.0523	0.1069	0.0069	0.0540
$Slash_4$	lCl	-3.0054	-0.0054	0.1055	-3.0025	-0.0025	0.1057
	lKa	0.5089	0.0089	0.2260	0.5128	0.0128	0.2283
	lKe	-2.5059	-0.0059	0.0590	-2.5056	-0.0056	0.0590
	τ_1	0.0331	-0.0069	0.0272	0.0414	0.0014	0.0328
	τ_2	0.1267	-0.0233	0.1209	0.1661	0.0161	0.1621
	σ^2	0.0968	-0.0032	0.0312	0.1004	0.0004	0.0323

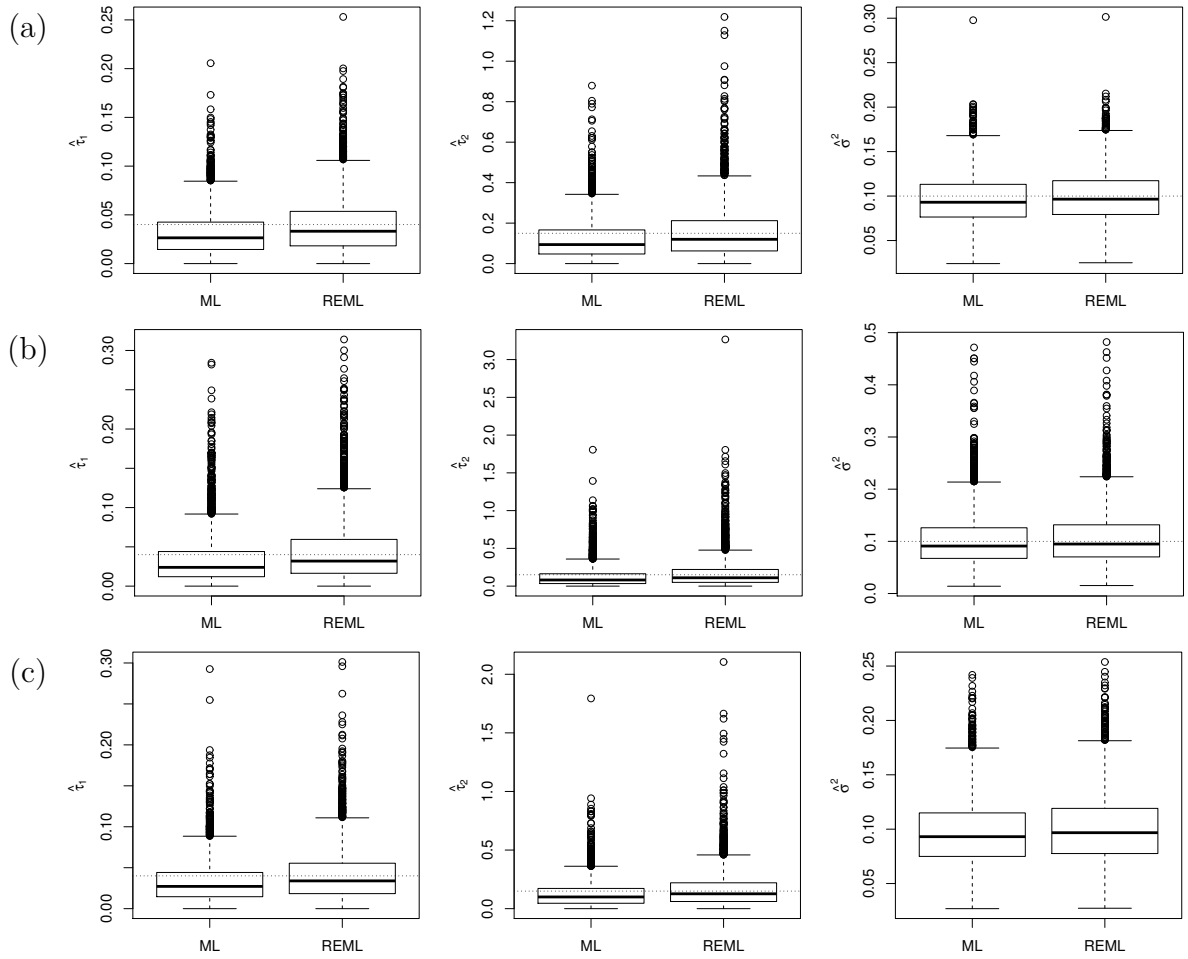


Figure 30 – Boxplots of the variance component estimates obtained from the simulation study. Panels in line (a) correspond to normal distribution, panels in line (b) represent t_4 distribution and in line (c) the slash$_4$ distribution. The horizontal dotted lines are the theoretical values.

A.4 Scenario Five

Table 22 – Simulation results for the pharmacokinetic model with theoretical fixed-effects parameters $(lCl, lKa, lKe)^\top = (-3.0, 0.5, -2.0)^\top$ and theoretical variance components $(\tau^\top, \sigma^2)_1^\top = (0.04, 0.15, 0.5)^\top$.

		ML			REML		
Dist.	Par.	Mean	Bias	RMSE	Mean	Bias	RMSE
Normal	lCl	-3.0214	-0.0214	0.0739	-3.0212	-0.0212	0.0728
	lKa	0.5135	0.0135	0.1356	0.5145	0.0145	0.1359
	lKe	-2.5179	-0.0179	0.0665	-2.5176	-0.0176	0.0654
	τ_1	0.0359	-0.0041	0.0175	0.0394	-0.0006	0.0186
	τ_2	0.1353	-0.0147	0.0789	0.1516	0.0016	0.0851
	σ^2	0.4827	-0.0173	0.0703	0.4873	-0.0127	0.0700
t_4	lCl	-3.0377	-0.0377	0.0831	-3.0341	-0.0341	0.0790
	lKa	0.5224	0.0224	0.1450	0.5216	0.0216	0.1431
	lKe	-2.5249	-0.0249	0.0712	-2.5214	-0.0214	0.0675
	τ_1	0.0305	-0.0095	0.0206	0.0341	-0.0059	0.0211
	τ_2	0.1040	-0.0460	0.0899	0.1205	-0.0295	0.0926
	σ^2	0.4512	-0.0488	0.1209	0.4563	-0.0437	0.1201
$Slash_4$	lCl	-3.0388	-0.0388	0.0881	-3.0364	-0.0364	0.0854
	lKa	0.5225	0.0225	0.1548	0.5224	0.0224	0.1536
	lKe	-2.5294	-0.0294	0.0771	-2.5272	-0.0272	0.0743
	τ_1	0.0354	-0.0046	0.0180	0.0389	-0.0011	0.0191
	τ_2	0.1320	-0.0180	0.0830	0.1493	-0.0007	0.0898
	σ^2	0.4807	-0.0193	0.0783	0.4852	-0.0148	0.0779

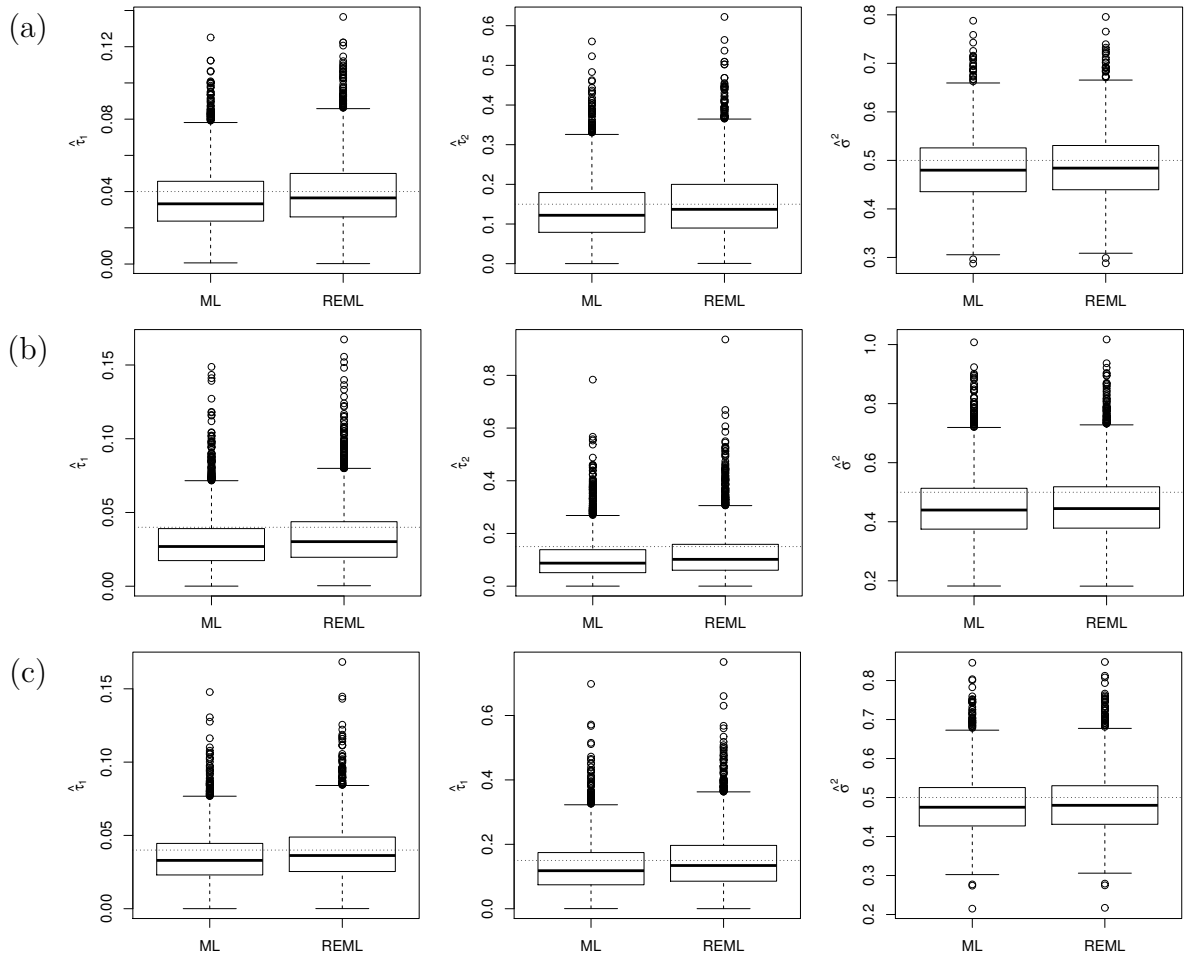


Figure 31 – Boxplots of the variance component estimates obtained from the simulation study. Panels in line (a) correspond to normal distribution, panels in line (b) represent t_4 distribution and in line (c) the slash$_4$ distribution. The horizontal dotted lines are the theoretical values.


ICMC
USP
SÃO CARLOS