Christian Eduardo Galarza Morales

Quantile Regression for Mixed-Effects Models

Regressão Quantílica para Modelos de Efeitos Mistos

CAMPINAS
2015

UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Matemática, Estatística
e Computação Científica

CHRISTIAN EDUARDO GALARZA MORALES

QUANTILE REGRESSION FOR MIXED-EFFECTS MODELS

REGRESSÃO QUANTÍLICA PARA MODELOS DE EFEITOS MISTOS

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

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em estatística.

Orientador: Víctor Hugo Lachos Dávila

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO DEFENDIDA PELO ALUNO CHRISTIAN EDUARDO GALARZA MORALES, E ORIENTADA PELO PROF. DR. VÍCTOR HUGO LACHOS DÁVILA.

Assinatura do Orientador

CAMPINAS
2015

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

G131q

Galarza Morales, Christian Eduardo, 1988-
Quantile regression for mixed-effects models / Christian Eduardo Galarza Morales. – Campinas, SP : [s.n.], 2015.

Orientador: Víctor Hugo Lachos Dávila.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

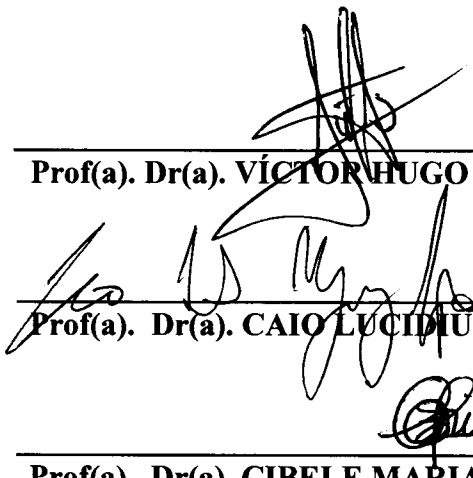
1. Modelos lineares (Estatística). 2. Modelos não lineares (Estatística). 3. Estimativa de parâmetro. 4. Análise de regressão. 5. Aproximação estocástica. I. Lachos Dávila, Víctor Hugo,1973-. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Regressão quantílica para modelos de efeitos mistos
Palavras-chave em inglês:
Linear models (Statistics)
Nonlinear models (Statistics)
Parameter estimation
Regression analysis
Stochastic approximation
Área de concentração: Estatística
Titulação: Mestre em Estatística
Banca examinadora:
Víctor Hugo Lachos Dávila [Orientador]
Caio Lucidius Naberezny Azevedo
Cibele Maria Russo Noveli
Data de defesa: 16-03-2015
Programa de Pós-Graduação: Estatística

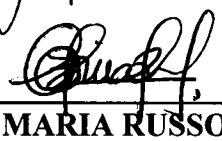
Dissertação de Mestrado defendida em 16 de março de 2015 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.



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

Prof(a). Dr(a). CAIO LUCIDIUS NABEREZNY AZEVEDO



Prof(a). Dr(a). CIBELE MARIA RUSSO NOVELI

Abstract

Longitudinal data are frequently analyzed using normal mixed effects models. Moreover, the traditional estimation methods are based on mean regression, which leads to non-robust parameter estimation for non-normal error distributions. Compared to the conventional mean regression approach, quantile regression (QR) can characterize the entire conditional distribution of the outcome variable and is more robust to the presence of outliers and misspecification of the error distribution. This thesis develops a likelihood-based approach to analyzing QR models for correlated continuous longitudinal data via the asymmetric Laplace distribution (ALD). Exploiting the nice hierarchical representation of the ALD, our classical approach follows the stochastic Approximation of the EM (SAEM) algorithm for deriving exact maximum likelihood (ML) estimates of the fixed-effects and variance components in linear and nonlinear mixed effects models. We evaluate the finite sample performance of the algorithm and the asymptotic properties of the ML estimates through empirical experiments and applications to four real life datasets. The proposed SAEM algorithms are implemented in the `R` packages `qrLMM()` and `qrNLMM()` respectively.

Keywords: Asymmetric Laplace distribution, Mixed Effects Models, Quantile regression, SAEM algorithm, Stochastic Approximations.

Resumo

Os dados longitudinais são frequentemente analisados usando modelos de efeitos mistos normais. Além disso, os métodos de estimação tradicionais baseiam-se em regressão na média da distribuição considerada, o que leva a estimação de parâmetros não robusta quando a distribuição do erro não é normal. Em comparação com a abordagem de regressão na média convencional, a regressão quantílica (RQ) pode caracterizar toda a distribuição condicional da variável de resposta e é mais robusta na presença de outliers e especificações erradas da distribuição do erro. Esta tese desenvolve uma abordagem baseada em verossimilhança para analisar modelos de RQ para dados longitudinais contínuos correlacionados através da distribuição Laplace assimétrica (DLA). Explorando a conveniente representação hierárquica da DLA, a nossa abordagem clássica segue a aproximação estocástica do algoritmo EM (SAEM) para derivar estimativas de máxima verossimilhança (MV) exatas dos efeitos fixos e componentes de variância em modelos lineares e não lineares

de efeitos mistos. Nós avaliamos o desempenho do algoritmo em amostras finitas e as propriedades assintóticas das estimativas de MV através de experimentos empíricos e aplicações para quatro conjuntos de dados reais. Os algoritmos SAEM propostos são implementados nos pacotes do `R` `qrLMM()` e `qrNLMM()` respectivamente.

Keywords: Distribuição Laplace assimétrica, Modelos de Efeitos Mistos, Regressão quantílica, algoritmo SAEM, Aproximações estocásticas.

Contents

A mis dos grandes amores,
Esther & Silvia.

Acknowledgement

At first and foremost I would like to thank God for having been my strength and biggest support throughout the course of this master. Undoubtedly, God is faithful.

I would like to express my gratitude to everyone who has supported to me during these long two years, in a very especial way to my lovely fiancée Silvia and to my parents, Víctor and Esther. This achievement is ours.

I am really grateful to CAPES/CNPq – IEL Nacional - Brazil for the economic support during this master. I would also like to thank to the Secretaría de Educación Superior, Ciencia, Tecnología e Innovación, SENESCYT - Ecuador, for its grateful help, without what, I would not be able to develop this thesis successfully. It would have been impossible to reach my goal without their help.

I also like to express my warm thanks to Mr. Víctor Hugo Lachos, my thesis supervisor, for his aspiring guidance, invaluably constructive criticism and friendship. Thanks to Mr. Gaudencio Zurita for his constant motivation in order to do my postgraduate studies. Thanks for teaching me that a talent should not be lost.

I shall never forget my entire close friends and teachers during my two academic years here; sincerely, they have made this journey easier. Finally I wanted to thank to Brazil, an open arms country whose wealth is not in their lands but in its people.

List of Figures

List of Tables

Chapter 1

Introduction

Mixed-effects models (MEM) are frequently used to analyze grouped/clustered data (such as longitudinal data, repeated measures, and multilevel data) because of their potential to handle within-subject correlations that characterizes grouped data (J. C. Pinheiro and D. M. Bates, 2000). In adittion, nonlinear mixed-efects models (NLMM) handle with nonlinearities in the relationship between the observed response and the covariates and random effects. In general, linear models can lead to highly accurate prediction model if the number of covariates increases, e.g. increasing the order of a polynomial model. Unfortunately, too many parameters will make interpretation difficult and good prediction are obtained just within the observed range of the data. On the other hand, nonlinear models incorporate theoretical considerations of the models unlike linear models that are based in the relationship between the response and the covariates. Nonlinear models are also flexible and often mechanistic, based on biological, chemical and physics mechanisms (among others). They lead to a natural modeling using a known family of nonlinear functions providing desirable characteristics such as asymptotes, a unique maximum value, monotonicity, positive range, etc.

Majority of these MEMs model and estimate covariate effects on the response through a mean regression, controlling for between-cluster heterogeneity via normally-distributed cluster-specific random effects and random errors. However, this centrality-based inferential framework is often inadequate when the conditional distribution of the response (conditional on the random terms) is skewed, multimodal, or affected by atypical observations. In contrast, conditional quantile regression (QR) methods (Roger Koenker, 2004, 2005) quantifying the entire conditional distribution of the outcome variable were developed that can provide assessment of covariate effects at any arbitrary quantiles of the outcome. In addition, QR methods do not impose any distribution assumption on the error, except requiring that the error term has a zero conditional quantile such as the asymmetric Laplace distribution (ALD). There are another zero-quantile families of distributions as detailed in Wichitaksorn et al. (2014). Because of its popularity and the flexibility that it provides, standard QR methods are implementable via available software packages, for example, the `R` package `quantreg`.

Although QR was initially developed under a univariate framework, the abundance of clustered data in recent times lead to its extensions into mixed modeling framework via either the

distribution-free route (Fu and Y.-G. Wang, 2012; Galvao Jr, 2011; Galvao and Montes-Rojas, 2010; Lipsitz et al., 1997), or the traditional likelihood-based route mostly using the ALD (Geraci and Bottai, 2007, 2014; Yuan and Yin, 2010). Among the ALD-based models, Geraci and Bottai (2007) proposed a Monte Carlo EM (MCEM)-based conditional QR model for continuous responses with a subject-specific random (univariate) intercept to account for within-subject dependence in the context of longitudinal data. However, due to the limitations of a simple random intercept model to account for the between-cluster heterogeneity, Geraci and Bottai (2014) extended their previous Geraci and Bottai (2007) model to a general linear quantile mixed effects regression model (QR-LMM) with multiple random effects (both intercepts and slopes). However, instead of going the MCEM route, the estimation of the fixed effects and the covariance components were implemented by using an efficient combination of Gaussian quadrature approximations and non-smooth optimization algorithms. Yuan and Yin (2010) applied the version of QR of Geraci and Bottai (2007) to linear mixed effects models (LMM) for longitudinal measurements with missing data. In other hand, J. Wang (2012) considered QR-NLMMs from a Bayesian perspective. Although the literature on QR-LMM is now substantial and some results on QR-NLMMs have recently appeared in the literature, to the best of our knowledge, there seem to be no studies on exact inference for QR-NLMMs from a likelihood based perspective.

In this work, we proceed via a robust parametric ALD-based QR for linear and nonlinear MEM, where the full likelihood-based implementation follows a stochastic version of the EM algorithm (SAEM) proposed by Delyon et al. (1999), for maximum likelihood (ML) estimation in contrast to the approximations proposed by Geraci and Bottai (2014). The SAEM algorithm has been proved to be more computationally efficient than the classical MCEM algorithm due to the recycling of simulations from one iteration to the next in the smoothing phase of the algorithm. Moreover, as pointed out by Meza et al. (2012) the SAEM algorithm, unlike the MCEM, converges even in a typically small simulation size. Recently, Kuhn and Lavielle (2005) showed that the SAEM algorithm is very efficient in computing the ML estimates in mixed effects models. Our empirical results using the SAEM are more efficient than the proposition of Geraci and Bottai (2014) for simulated data for LMMs and a likelihood-based model for NLMMs in contrast to the bayesian modeling proposed by J. Wang (2012). The methods developed are readily implementable via the `R` packages `qrLMM()` and `qrNLMM()`, illustrating four real data sets.

The rest of the thesis proceeds as follows. Chapter 1 presents some preliminaries, in particular the connection between QR and ALD and an outline of the EM and SAEM algorithms. Chapter 2 and 3 develops the MCEM and the SAEM algorithms for a general LMM and NLMM respectively, both outlining the likelihood estimation and standard errors, simulation studies to analyze the finite sample performance of our proposed methods, application of the SAEM method to two longitudinal datasets and conclusions. Chapter 4 presents some general concluding remarks as technical production resulting of this thesis and some conclusions sketching some future research directions.

1.1 Preliminaries

In this section, we provide some useful results on the ALD and QR, and introduce the EM and SAEM algorithms for ML estimation.

1.1.1 Connection between QR and ALD

Following K. Yu and Moyeed (2001), a random variable Y is distributed as an ALD with location parameter μ, scale parameter $\sigma > 0$ and skewness parameter $p \in (0,1)$, if its probability density function (pdf) is given by

$$f(y|\mu,\sigma,p) = \frac{p(1-p)}{\sigma} \exp\left\{-\rho_p\left(\frac{y-\mu}{\sigma}\right)\right\}, \tag{1.1.1}$$

where $\rho_p(.)$ is the check (or loss) function defined by $\rho_p(u) = u(p - \mathbb{I}\{u < 0\})$, with $\mathbb{I}\{.\}$ the usual indicator function. This distribution is denoted by $ALD(\mu,\sigma,p)$. It is easy to see that $W = \rho_p(\frac{Y-\mu}{\sigma})$ follows an exponential distribution with mean 1. Figure 1.1 plots the ALD illustrating how the the skewness changes with altering choices for p. For example, when $p = 0.1$, most of the mass is concentrated around the right tail, while for $p = 0.5$, both tails of the ALD have equal mass and the distribution resemble the more common double exponential distribution. In contrast to the normal distribution with a quadratic term in the exponent, the ALD is linear in the exponent term. This results in a more peaked mode for the ALD together with thicker tails. On the contrary, the normal distribution has heavier tails compared to the ALD. Is worth mentioning that we have implemented in the `R` package `ald()`, the probability density function, distribution function, quantile function, random number generator function, likelihood function, moment generating function and MLE for a given sample for the ALD defined above.

The ALD abides by the following stochastic representation (Kotz et al., 2001; Kuzobowski and Podgorski, 2000). Let $U \sim \exp(\sigma)$ and $Z \sim N(0,1)$ be two independent random variables. Then, $Y \sim ALD(\mu,\sigma,p)$ can be represented as

$$Y \stackrel{d}{=} \mu + \vartheta_p U + \tau_p\sqrt{\sigma U}Z, \tag{1.1.2}$$

where $\vartheta_p = \frac{1-2p}{p(1-p)}$ and $\tau_p^2 = \frac{2}{p(1-p)}$, where $\stackrel{d}{=}$ denotes equality in distribution. This representation is useful in obtaining the moment generating function (mgf), and formulating the estimation algorithm. From (1.1.2), the hierarchical representation of the ALD is given as

$$\begin{aligned} Y|U = u &\sim N(\mu + \vartheta_p u, \tau_p^2\sigma u), \\ U &\sim \exp(\sigma). \end{aligned} \tag{1.1.3}$$

This representation will be useful for the implementation of the EM algorithm. Moreover, since $Y|U = u \sim N(\mu + \vartheta_p u, \tau_p^2\sigma u)$, one can easily derive the pdf of Y, given by

$$f(y|\mu,\sigma,p) = \frac{1}{\sqrt{2\pi}}\frac{1}{\tau_p\sigma^{\frac{3}{2}}} \exp\big(\frac{\delta(y)}{\gamma}\big)A(y), \tag{1.1.4}$$

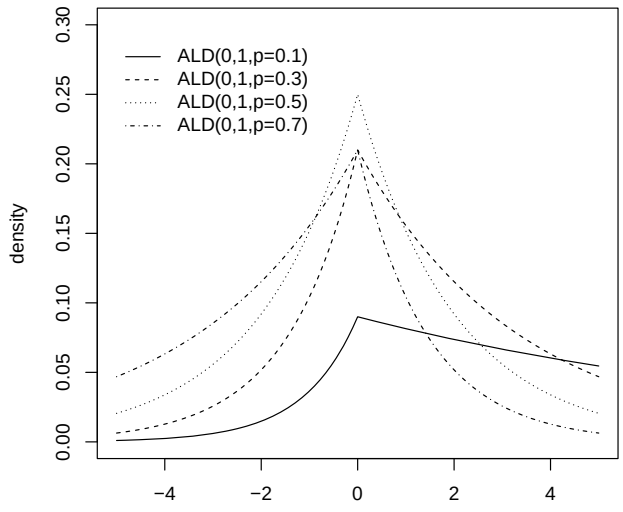


Figure 1.1: Standard asymmetric Laplace density

where $\delta(y) = \frac{|y-\mu|}{\tau_p\sqrt{\sigma}}$, $\gamma = \sqrt{\frac{1}{\sigma}(2 + \frac{\vartheta_p^2}{\tau_p^2})} = \frac{\tau_p}{2\sqrt{\sigma}}$ and $A(y) = 2\Big(\frac{\delta(y)}{\gamma}\Big)^{1/2} K_{1/2}(\delta(y)\gamma)$, with $K_\nu(.)$, the modified Bessel function of the third kind. It easy to see that that the conditional distribution of U, given $Y = y$, is $U|(Y = y) \sim GIG(\frac{1}{2}, \delta, \gamma)$, where $GIG(\nu, a, b)$ represents the Generalized Inverse Gaussian (GIG) distribution (Barndorff-Nielsen and Shephard, 2001) with the pdf

$$h(u|\nu, a, b) = \frac{(b/a)^\nu}{2K_\nu(ab)} u^{\nu-1} \exp\Big\{ -\frac{1}{2}(a^2/u + b^2 u)\Big\},\ u > 0, \quad \nu \in \mathbb{R},\ a, b > 0.$$

The moments of U can be expressed as

$$E[U^k] = \Big(\frac{a}{b}\Big)^k \frac{K_{\nu+k}(ab)}{K_\nu(ab)}, k \in \mathbb{R} \tag{1.1.5}$$

Some useful properties of the Bessel function of the third kind $K_\lambda(u)$ are: (i) $K_\nu(u) = K_{-\nu}(u)$; (ii) $K_{\nu+1}(u) = \frac{2\nu}{u} K_\nu(u) + K_{\nu-1}(u)$; (iii) for non-negative integer r, $K_{r+1/2}(u) = \sqrt{\frac{\pi}{2u}} \exp(-u)$ $\sum_{k=0}^{r} \frac{(r+k)!(2u)^{-k}}{(r-k)!k!}$. A special case is $K_{1/2}(u) = \sqrt{\frac{\pi}{2u}} \exp(-u)$.

1.1.2 The EM and SAEM algorithms

In models with missing data, the EM algorithm (Dempster et al., 1977) has established itself as the most popular tool for obtaining the ML estimates of the model parameters. This iterative algorithm maximizes the complete log-likelihood function $\ell_c(\boldsymbol{\theta}; \mathbf{y}_{\text{com}})$ at each step, converging

quickly to a stationary point of the observed likelihood ($\ell(\boldsymbol{\theta}; \mathbf{y}_{\text{obs}})$) under mild regularity conditions (Vaida, 2005; C. J. Wu, 1983). The EM algorithm proceeds in two simple steps:

E-Step: Replace the observed likelihood by the complete likelihood and compute its conditional expectation $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \mathrm{E}\left\{\ell_c(\boldsymbol{\theta}; \mathbf{y}_{\text{com}})|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_{\text{obs}}\right\}$, where $\widehat{\boldsymbol{\theta}}^{(k)}$ is the estimate of $\boldsymbol{\theta}$ at the k-th iteration;

M-Step: Maximize $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$ obtaining $\widehat{\boldsymbol{\theta}}^{(k+1)}$.

However, in some applications of the EM algorithm, the E-step cannot be obtained analytically and has to be calculated using simulations. Wei and Tanner (1990) proposed the Monte Carlo EM (MCEM) algorithm in which the E-step is replaced by a Monte Carlo approximation based on a large number of independent simulations of the missing data. This simple solution is infact computationally expensive, given the need to generate a large number of independent simulations of the missing data for a good approximation. Thus, in order to reduce the amount of required simulations compared to the MCEM algorithm, the SAEM algorithm proposed by Delyon et al. (1999) replaces the E-step of the EM algorithm by a stochastic approximation procedure, while the Maximization step remains unchanged. Besides having good theoretical properties, the SAEM estimates the population parameters accurately, converging to the global maxima of the ML estimates under quite general conditions (Allassonnière et al., 2010; Delyon et al., 1999; Kuhn and Lavielle, 2004).

At each iteration, the SAEM algorithm successively simulates missing data with the conditional distribution, and updates the unknown parameters of the model. Thus, at iteration k, the SAEM algorithm proceeds as follows:

E-Step:

- Simulation: Draw $(\mathbf{q}^{(\ell,k)})$, $\ell = 1, \ldots, m$ from the conditional distribution $f(\mathbf{q}|\boldsymbol{\theta}^{(k-1)}, \mathbf{y}_i)$.
- Stochastic Approximation: Update the $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ function as

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) \approx Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k-1)}) + \delta_k \left[\frac{1}{m}\sum_{\ell=1}^{m} \ell_c(\boldsymbol{\theta}\,; \mathbf{y}_{\text{obs}}, \mathbf{q}^{(\ell,k)}|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_{\text{obs}}) - Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k-1)})\right] \quad (1.1.6)$$

M-Step:

- Maximization: Update $\widehat{\boldsymbol{\theta}}^{(k)}$ as $\widehat{\boldsymbol{\theta}}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$,

where $(\mathbf{q}^{(\ell,k)})$ is a sample from the missing values and δ_k is a smoothness parameter (Kuhn and Lavielle, 2004), i.e., a decreasing sequence of positive numbers such that $\sum_{k=1}^{\infty} \delta_k = \infty$ and $\sum_{k=1}^{\infty} \delta_k^2 < \infty$. Note that, for the SAEM algorithm, the E-Step coincides with the MCEM algorithm, however a small number of simulations m (suggested to be $m \leq 20$) is necessary. This is

possible because unlike the traditional EM algorithm and its variants, the SAEM algorithm uses not only the current simulation of the missing data at the iteration k denoted by $(\mathbf{q}^{(\ell,k)})$, $\ell = 1, \ldots, m$ but some or all previous simulations, where this 'memory' property is set by the smoothing parameter δ_k.

Note, in equation (1.1.6), if the smoothing parameter δ_k is equal to 1 for all k, the SAEM algorithm will have 'no memory', and will be equivalent to the MCEM algorithm. The SAEM with no memory will converge quickly (convergence in distribution) to a solution neighbourhood, however when the algorithm has memory, it will converge slowly (almost sure convergence) to the ML solution. We suggested the following choice of the smoothing parameter given as

$$\delta_k = \begin{cases} 1, & \text{for} \quad 1 \leq k \leq cW \\ \frac{1}{k - cW}, & \text{for} \quad cW + 1 \leq k \leq W \end{cases}$$

where W is the maximum number of iterations, and c a cut point $(0 \leq c \leq 1)$ which determines the percentage of initial iterations with no memory. For example, if $c = 0$ the algorithm will have memory for all iterations, and hence will converge slowly to the ML estimates. If $c = 1$, the algorithm will have no memory, and so will converge quickly to a solution neighborhood. For the first case, W would need to be large in order to achieve the ML estimates. For the second one, the algorithm will output a Markov Chain where after applying a *burn in* and *thin*, the mean of the chain observations can be a reasonable estimate.

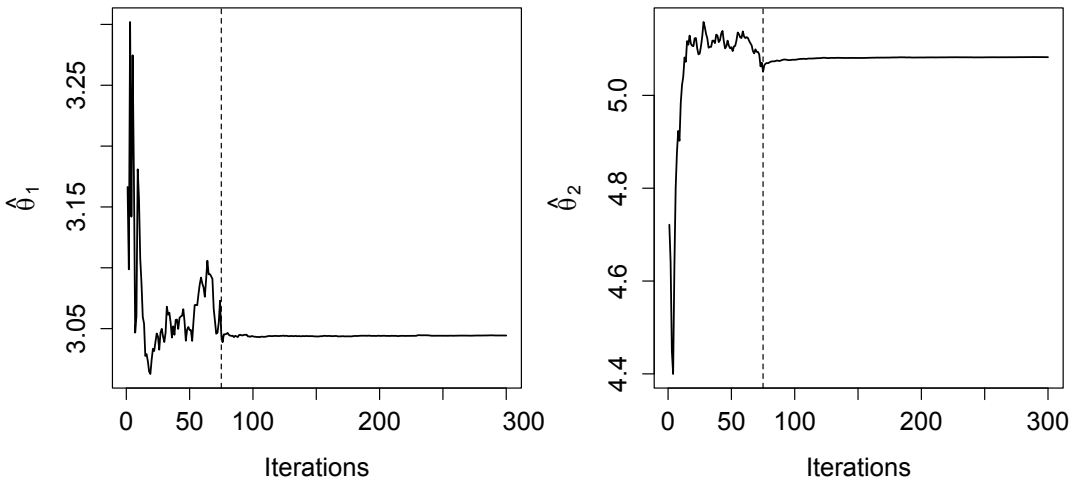


Figure 1.2: Graphical approach for assessing convergence. Sequences $\hat{\theta}_i$ for $i = 1, 2$ of estimatives for the fixed effects of a Linear Mixed Model with $W = 300$ and $c = 1/4$.

A number between 0 and 1 $(0 < c < 1)$ will assure an initial convergence in distribution to a solution neighbourhood for the first cW iterations and an almost sure convergence for the rest of

the iterations. Hence, this combination will leads us to a fast algorithm with good estimates. To implement SAEM, the user must fix several constants matching the number of total iterations W and the cut point c that defines the starting of the smoothing step of the SAEM algorithm, however those parameters will vary depending of the model and the data. To determinate those constants, a graphical approach (see Figure 1.2) is recommended to monitor the convergence of the estimates for all the parameters, and, if possible, to monitor the difference (relative difference) between two successive evaluations of the log-likelihood $\ell(\boldsymbol{\theta}|\mathbf{y}_{obs})$, given by $||\ell(\boldsymbol{\theta}^{(k+1)}|\mathbf{y}_{obs}) - \ell(\boldsymbol{\theta}^{(k)}|\mathbf{y}_{obs})||$ or $||\ell(\boldsymbol{\theta}^{(k+1)}|\mathbf{y}_{obs})/\ell(\boldsymbol{\theta}^{(k)}|\mathbf{y}_{obs}) - 1||$, respectively.

Chapter 2

Quantile Regression for Linear Mixed Models

This chapter develops a likelihood-based approach to analyzing quantile regression (QR) models for continuous longitudinal data via the asymmetric Laplace distribution (ALD). Compared to the conventional mean regression approach, QR can characterize the entire conditional distribution of the outcome variable and is more robust to the presence of outliers and misspecification of the error distribution. Exploiting the nice hierarchical representation of the ALD, our classical approach follows the Stochastic Approximation of the EM (SAEM) algorithm for deriving exact maximum likelihood estimates of the fixed-effects and variance components. We evaluate the finite sample performance of the algorithm and the asymptotic properties of the ML estimates through empirical experiments and applications to two real datasets. Our empirical results clearly indicate that the SAEM estimates outperforms the estimates obtained via the combination of Gaussian quadrature and non-smooth optimization routines of the Geraci (2014)'s approach in terms of standard errors and mean square error. The proposed SAEM algorithm is implemented in the `R` package `qrLMM()`.

2.1 Introduction

Linear mixed-effects models (LMM) are frequently used to analyze grouped/clustered data (such as longitudinal data, repeated measures, and multilevel data) because of their potential to handle within-subject correlations that characterizes grouped data (J. C. Pinheiro and D. M. Bates, 2000). Majority of these LMMs model and estimate covariate effects on the response through a mean regression, controlling for between-cluster heterogeneity via normally-distributed cluster-specific random effects and random errors. However, this centrality-based inferential framework is often inadequate when the conditional distribution of the response (conditional on the random terms) is skewed, multimodal, or affected by atypical observations. In contrast, conditional quantile regression (QR) methods (Roger Koenker, 2004, 2005) quantifying the entire conditional distribution of the outcome variable were developed that can provide assessment of covariate effects at any arbitrary quantiles of the outcome. In addition, QR methods do not impose any distribution

assumption on the error, except requiring that the error term has a zero conditional quantile.

Although QR was initially developed under a univariate framework, the abundance of clustered data in recent times leads to its extensions into mixed modeling framework (classical, or Bayesian) via either the distribution-free route (Fu and Y.-G. Wang, 2012; Galvao Jr, 2011; Galvao and Montes-Rojas, 2010; Lipsitz et al., 1997), or the traditional likelihood-based route mostly using the ALD (Geraci and Bottai, 2007, 2014; Yuan and Yin, 2010). Among the ALD-based models, Geraci and Bottai (2007) proposed a Monte Carlo EM (MCEM)-based conditional QR model for continuous responses with a subject-specific random (univariate) intercept to account for within-subject dependence in the context of longitudinal data. However, due to the limitations of a simple random intercept model to account for the between-cluster heterogeneity, Geraci and Bottai (2014) extended their previous Geraci and Bottai (2007) model to a general linear quantile mixed effects regression model (QR-LMM) with multiple random effects (both intercepts and slopes). However, instead of going the MCEM route, the estimation of the fixed effects and the covariance components were implemented using an efficient combination of Gaussian quadrature approximations and non-smooth optimization algorithms.

Although the literature on QR-LMM is now substantial, there are no studies conducting exact inferences in the context of QR-LMM from a likelihood-based perspective. In this paper, we proceed to achieve that via a robust parametric ALD-based QR-LMM, where the full likelihood-based implementation follows a stochastic version of the EM algorithm (SAEM), proposed by Delyon et al. (1999), for maximum likelihood (ML) estimation in contrast to the approximations proposed by Geraci and Bottai (2014). The SAEM algorithm has been proved to be more computationally efficient than the classical MCEM algorithm due to the recycling of simulations from one iteration to the next in the smoothing phase of the algorithm. Moreover, as pointed out by Meza et al. (2012) the SAEM algorithm, unlike the MCEM, converges even in a typically small simulation size. Recently, Kuhn and Lavielle (2005) showed that the SAEM algorithm is very efficient in computing the ML estimates in mixed effects models. Our empirical results using the SAEM are more efficient than the proposition of Geraci and Bottai (2014) for simulated data. Furthermore, application of our method to two longitudinal datasets is illustrated via the `R` package `qrLMM()`.

2.2 QR for linear mixed models and algorithms

We consider the following general LMM $y_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + \mathbf{z}_{ij}\mathbf{b}_i + \epsilon_{ij},\ i = 1, \ldots, n,\ j = 1, \ldots, n_i$, where y_{ij} is the jth measurement of a continuous random variable for the ith subject, $\mathbf{x}_{ij}^\top$ are row vectors of a known design matrix of dimension $N \times k$ corresponding to the $k \times 1$ vector of population-averaged fixed effects $\boldsymbol{\beta}$, $\mathbf{z}_{ij}$ is a $q \times 1$ design matrix corresponding to the $q \times 1$ vector of random effects $\mathbf{b}_i$, and ϵ_{ij} the independent and identically distributed random errors. We define pth quantile function of the response y_{ij} as

$$Q_p(y_{ij}|\mathbf{x}_{ij}, \mathbf{b}_i) = \mathbf{x}_{ij}^\top \boldsymbol{\beta}_p + \mathbf{z}_{ij}\mathbf{b}_i. \tag{2.2.1}$$

where Q_p denotes the inverse of the unknown distribution function F, $\boldsymbol{\beta}_p$ is the regression coefficient corresponding to the pth quantile. As seen in (2.2.1) the pth quantile is equal to a mixed linear

predictor. The random effects $\mathbf{b}_i$ are distributed as $\mathbf{b}_i \overset{\text{iid}}{\sim} \mathrm{N}_q(\mathbf{0}, \boldsymbol{\Psi})$, where the dispersion matrix $\boldsymbol{\Psi} = \boldsymbol{\Psi}(\boldsymbol{\alpha})$ depends on unknown and reduced parameters $\boldsymbol{\alpha}$ (the distinct elements of $\boldsymbol{\Psi}$), and the errors $\epsilon_{ij} \sim ALD(0, \sigma)$. Then, $y_{ij}|\mathbf{b}_i$ independently follows as ALD with the density given by

$$f(y_{ij}|\boldsymbol{\beta}_p, \mathbf{b}_i, \sigma) = \frac{p(1-p)}{\sigma} \exp\left\{-\rho_p\left(\frac{y_{ij} - \mathbf{x}_{ij}^\top \boldsymbol{\beta}_p - \mathbf{z}_{ij}\mathbf{b}_i}{\sigma}\right)\right\}, \tag{2.2.2}$$

Using a MCEM algorithm, a QR-LMM with random intercepts ($q = 1$) was proposed by Geraci and Bottai (2007). More recently, Geraci and Bottai (2014) extended that setup to accommodate multiple random effects where the estimation of fixed effects and covariance matrix of the random effects were accomplished via a combination of Gaussian quadrature approximations and non-smooth optimization algorithms. Here, we consider a more general correlated random effects framework with general dispersion matrix $\boldsymbol{\Psi} = \boldsymbol{\Psi}(\boldsymbol{\alpha})$.

2.2.1 A MCEM algorithm

First, we develop a MCEM algorithm for ML estimation of the parameters in the QR-LMM. The model exhibits a flexible hierarchical representation, which is useful in deriving the theoretical properties. From (1.1.3), the QR-LMM defined in (2.2.1)-(2.2.2), can be represented in a hierarchical form as:

$$\begin{aligned} \mathbf{y}_i|\mathbf{b}_i, \mathbf{u}_i &\sim N_{n_i}\left(\mathbf{x}_i^\top \boldsymbol{\beta}_p + \mathbf{z}_i\mathbf{b}_i + \vartheta_p \mathbf{u}_i, \sigma\tau_p^2 \mathbf{D}_i\right), \\ \mathbf{b}_i &\sim N_q\,(\mathbf{0}, \boldsymbol{\Psi}), \\ \mathbf{u}_i &\sim \prod_{j=1}^{n_i} \exp(\sigma), \end{aligned} \tag{2.2.3}$$

for $i = 1, \ldots, n$, where ϑ_p and τ_p^2 are as in (1.1.2); $\mathbf{D}_i$ represents a diagonal matrix that contains the vector of missing values $\mathbf{u}_i = (u_{i1}, \ldots, u_{in_i})^\top$ and $\exp(\sigma)$ denotes the exponential distribution with mean σ. Let $\mathbf{y}_{ic} = (\mathbf{y}_i^\top, \mathbf{b}_i^\top, \mathbf{u}_i^\top)^\top$, with $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$, $\mathbf{b}_i = (b_{i1}, \ldots, b_{iq})^\top$, $\mathbf{u}_i = (u_{i1}, \ldots, u_{in_i})^\top$and let $\boldsymbol{\theta}^{(k)} = (\boldsymbol{\beta}_p^{(k)\top}, \sigma^{(k)}, \boldsymbol{\alpha}^{(k)\top})^\top$, the estimate of θ at the k-th iteration. Since $\mathbf{b}_i$ and $\mathbf{u}_i$ are independent for all $i = 1, \ldots, n$, it follows from (1.1.3) that the complete-data log-likelihood function is of the form

$$\ell_c(\boldsymbol{\theta};\, \mathbf{y}_c) = \sum_{i=1}^{n} \ell_c(\boldsymbol{\theta};\, \mathbf{y}_{ic}),$$

where

$$\begin{aligned} \ell_c(\boldsymbol{\theta};\, \mathbf{y}_{ic}) &= \text{constant} - \frac{3}{2}\, n_i log\sigma - \frac{1}{2}\, log|\boldsymbol{\Psi}| - \frac{1}{2}\, \mathbf{b}_i^\top \boldsymbol{\Psi}^{-1}\mathbf{b}_i - \frac{1}{\sigma}\, \mathbf{u}_i^\top \mathbf{1}_{n_i} \\ &- \frac{1}{2\sigma\tau_p^2}\, (\mathbf{y}_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p - \mathbf{z}_i\mathbf{b}_i - \vartheta_p\mathbf{u}_i)^\top \mathbf{D}_i^{-1} (\mathbf{y}_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p - \mathbf{z}_i\mathbf{b}_i - \vartheta_p\mathbf{u}_i). \end{aligned} \tag{2.2.4}$$

Given the current estimate $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$, the E-step calculates the function

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

where

$$\begin{aligned} Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) &= \mathrm{E}\left\{\ell_c(\boldsymbol{\theta};\, \mathbf{y}_{ic})|\boldsymbol{\theta}^{(k)}, \mathbf{y}\right\} \\ &\propto -\frac{3}{2}\, n_i log\sigma - \frac{1}{2\sigma\tau_p^2}\Big[(\mathbf{y}_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p)^\top \widehat{\mathbf{D}_i^{-1}}^{(k)} (\mathbf{y}_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p) \\ &\quad -2(\mathbf{y}_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p)(\widehat{\mathbf{D}_i^{-1}\mathbf{z}\mathbf{b}})_i^{(k)} + \mathrm{tr}\left\{\mathbf{z}_i(\widehat{\mathbf{b}\mathbf{b}^\top\mathbf{z}\mathbf{D}_i^{-1}})_i^{(k)}\right\} \\ &\quad -2\vartheta_p(\mathbf{y}_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p)^\top \mathbf{1}_{n_i} + 2\vartheta_p(\mathbf{z}\widehat{\mathbf{b}}^{(k)})_i^\top \mathbf{1}_{n_i} + \frac{\tau_p^4}{4}\widehat{\mathbf{u}_i}^{(k)\top}\mathbf{1}_{n_i}\Big] \\ &\quad -\frac{1}{2}\, log|\boldsymbol{\Psi}| - \frac{1}{2}\,\mathrm{tr}\left\{\widehat{(\mathbf{b}\mathbf{b}^\top)_i}^{(k)}\boldsymbol{\Psi}^{-1}\right\}, \end{aligned} \tag{2.2.5}$$

where $\mathrm{tr}(\mathbf{A})$ indicates the trace of matrix $\mathbf{A}$ and $\mathbf{1}_p$ is the vector of ones of dimension p. The calculation of these function requires expressions for

$$\begin{aligned} \widehat{\mathbf{b}_i}^{(k)} &= \mathrm{E}\left\{\mathbf{b}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i\right\}, & \widehat{\mathbf{u}_i}^{(k)} &= \mathrm{E}\left\{\mathbf{u}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i\right\}, \\ \widehat{(\mathbf{b}\mathbf{b}^\top)_i}^{(k)} &= \mathrm{E}\left\{\mathbf{b}_i\mathbf{b}_i^\top|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i\right\}, & \widehat{\mathbf{D}_i^{-1}}^{(k)} &= \mathrm{E}\left\{\mathbf{D}_i^{-1}|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i\right\}, \\ (\widehat{\mathbf{b}\mathbf{b}^\top\mathbf{z}\mathbf{D}^{-1}})_i^{(k)} &= \mathrm{E}\left\{\mathbf{b}_i\mathbf{b}_i^\top\mathbf{z}_i^\top\mathbf{D}_i^{-1}|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i\right\}, & (\widehat{\mathbf{D}^{-1}\mathbf{z}\mathbf{b}})_i^{(k)} &= \mathrm{E}\left\{\mathbf{D}_i^{-1}\mathbf{z}_i\mathbf{b}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i\right\}, \end{aligned}$$

which do not have closed forms. Since the joint distribution of the missing data $(\mathbf{b}_i^{(k)}, \mathbf{u}_i^{(k)})$ is unknown and the conditional expectations cannot be computed analytically, for any function $g(.)$, the MCEM algorithm approximates the conditional expectations above by their Monte Carlo approximations

$$\mathrm{E}[\, g\,(\mathbf{b}_i, \mathbf{u}_i)\,|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i] \approx \frac{1}{m}\sum_{\ell=1}^{m} g(\mathbf{b}_i^{(\ell,k)}, \mathbf{u}_i^{(\ell,k)}), \tag{2.2.6}$$

which depend of the simulations of the two latent (missing) variables $\mathbf{b}_i^{(k)}$ and $\mathbf{u}_i^{(k)}$ from the conditional joint density $f(\mathbf{b}_i, \mathbf{u}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i)$. A Gibbs Sampler can be easily implemented as shown in Appendix A.4 given that the two full conditional distributions $f(\mathbf{b}_i|\boldsymbol{\theta}^{(k)}, \mathbf{u}_i, \mathbf{y}_i)$ and $f(\mathbf{u}_i|\boldsymbol{\theta}^{(k)}, \mathbf{b}_i, \mathbf{y}_i)$ are known. However, using known properties of conditional expectations, the expected value in (2.2.6) can be more accurately approximated as

$$\begin{aligned} \mathrm{E}_{\mathbf{b}_i,\mathbf{u}_i}[\, g(\mathbf{b}_i, \mathbf{u}_i)|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i] &= \mathrm{E}_{\mathbf{b}_i}[\, \mathrm{E}_{\mathbf{u}_i}[\, g(\mathbf{b}_i, \mathbf{u}_i)|\boldsymbol{\theta}^{(k)}, \mathbf{b}_i, \mathbf{y}_i]|\mathbf{y}_i\,] \\ &\approx \frac{1}{m}\sum_{\ell=1}^{m} \mathrm{E}_{\mathbf{u}_i}[\, g(\mathbf{b}_i^{(\ell,k)}, \mathbf{u}_i)|\boldsymbol{\theta}^{(k)}, \mathbf{b}_i^{(\ell,k)}, \mathbf{y}_i], \end{aligned} \tag{2.2.7}$$

where $\mathbf{b}^{(\ell,k)}$ is a sample from the conditional density $f(\mathbf{b}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i)$. Note that (2.2.7) is a more accurate approximation once it only depends of one MC approximation, instead two as needed in (2.2.6).

Now, to drawn random samples from the full conditional distribution $f(\mathbf{u}_i|\mathbf{y}_i, \mathbf{b}_i)$, first note that the vector $\mathbf{u}_i|\mathbf{y}_i, \mathbf{b}_i$ can be written as $\mathbf{u}_i|\mathbf{y}_i, \mathbf{b}_i = [\ u_{i1}|y_{i1}, \mathbf{b}_i,\ \ u_{i2}|y_{i2}, \mathbf{b}_i,\ \ \cdots\ \ , u_{in_i}|y_{in_i}, \mathbf{b}_i\ ]^\top$, since $u_{ij}|\ y_{ij}, \mathbf{b}_i$ is independent of $u_{ik}|\ y_{ik}, \mathbf{b}_i$, for all $j, k = 1, 2, \ldots, n_i$ and $j \neq k$. Thus, the distribution of $f(u_{ij}|y_{ij}, \mathbf{b}_i)$ is proportional to

$$f(u_{ij}|y_{ij}, \mathbf{b}_i) \propto \phi(y_{ij}|\mathbf{x}_{ij}^\top\boldsymbol{\beta}_p + \mathbf{z}_{ij}^\top\mathbf{b}_i + \vartheta_p u_{ij},\ \sigma\tau_p^2 u_{ij}) \times \exp(\sigma),$$

which, from Subsection 2.1, leads to $u_{ij}|y_{ij}, \mathbf{b}_i \sim GIG(\frac{1}{2}, \chi_{ij}, \psi)$, where χ_{ij} and ψ are given by

$$\chi_{ij} = \frac{|y_{ij} - \mathbf{x}_{ij}^\top\boldsymbol{\beta}_p - \mathbf{z}_{ij}^\top\mathbf{b}_i|}{\tau_p\sqrt{\sigma}} \quad \text{and} \quad \psi = \frac{\tau_p}{2\sqrt{\sigma}} \tag{2.2.8}$$

From (1.1.5), and after generating samples from $f(\mathbf{b}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i)$ (see Subsection 2.2.3), the conditional expectation $\mathrm{E}_{\mathbf{u}_i}[\cdot|\boldsymbol{\theta}, \mathbf{b}_i, \mathbf{y}_i]$ in (2.2.7) can be computed analytically. Finally, the proposed MCEM algorithm for estimating the parameters of the QR-LMM can be summarized as follows:

MC E-step: Given $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$, for $i = 1, \ldots, n$;

- **Simulation Step:** For $\ell = 1, \ldots, m$, draw $\mathbf{b}_i^{(\ell,k)}$ from $f(\mathbf{b}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i)$, as described later in Subsection 2.2.3.
- **Monte Carlo approximation:** Using (1.1.5) and the simulated sample above, evaluate

$$\mathrm{E}[\,g\,(\mathbf{b}_i, \mathbf{u}_i)\,|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i] \approx \frac{1}{m}\sum_{\ell=1}^{m} \mathrm{E}_{\mathbf{u}_i}[\,g(\mathbf{b}_i^{(\ell,k)}, \mathbf{u}_i)|\boldsymbol{\theta}^{(k)}, \mathbf{b}_i^{(\ell,k)}, \mathbf{y}_i].$$

M-step: Update $\widehat{\boldsymbol{\theta}}^{(k)}$ by maximizing $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) \approx \frac{1}{m}\sum_{l=1}^{m}\sum_{i=1}^{n}\ell_c(\theta;\, \mathbf{y}_i, \mathbf{b}_i^{(l,k)}, \mathbf{u}_i)$ over $\widehat{\boldsymbol{\theta}}^{(k)}$, which leads to the following estimates:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_p^{(k+1)} &= \left[\sum_{i=1}^{n}\left\{\frac{1}{m}\sum_{\ell=1}^{m}\mathbf{x}_i\mathcal{E}(\mathbf{D}_i^{-1})^{(\ell,k)}\mathbf{x}_i^\top\right\}\right]^{-1} \times \\
&\quad \left[\sum_{i=1}^{n}\left\{\frac{1}{m}\sum_{\ell=1}^{m}\left[\mathbf{x}_i\mathcal{E}(\mathbf{D}_i^{-1})^{(\ell,k)}\left[\mathbf{y}_i - \mathbf{z}_i^\top\mathbf{b}_i^{(\ell,k)} - \vartheta_p\mathcal{E}(\mathbf{u}_i)^{(\ell,k)}\right]\right]\right\}\right], \\
\widehat{\sigma}^{(k+1)} &= \frac{1}{3N\tau_p^2}\sum_{i=1}^{n}\left\{\frac{1}{m}\sum_{\ell=1}^{m}\left[(\mathbf{y}_i - \mathbf{x}_i^\top\boldsymbol{\beta}_p^{(k+1)} - \mathbf{z}_i\mathbf{b}_i^{(\ell,k)})^\top\mathcal{E}(\mathbf{D}^{-1})^{(\ell,k)}(\mathbf{y}_i - \mathbf{x}_i^\top\boldsymbol{\beta}_p^{(k+1)} - \mathbf{z}_i\mathbf{b}_i^{(\ell,k)})\right.\right. \\
&\quad \left.\left. -2\vartheta_p(\mathbf{y}_i - \mathbf{x}_i^\top\boldsymbol{\beta}_p^{(k+1)} - \mathbf{z}_i\mathbf{b}_i^{(\ell,k)})^\top\mathbf{1}_{n_i} + \frac{\tau_p^4}{4}\mathcal{E}(\mathbf{u}_i)^{(\ell,k)\top}\mathbf{1}_{n_i}\right]\right\}, \\
\widehat{\boldsymbol{\Psi}}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{m}\sum_{\ell=1}^{m}\mathbf{b}_i^{(\ell,k)}\mathbf{b}_i^{(\ell,k)^\top}\right],
\end{aligned}$$

where $N = \sum_{i=1}^{n} n_i$ and expressions $\mathcal{E}(\mathbf{u}_i)^{(\ell,k)}$ and $\mathcal{E}(\mathbf{D}_i^{-1})^{(\ell,k)}$ are defined in Appendix A.2. Note that for the MC E-step, we need to draw samples $\mathbf{b}_i^{(\ell,k)}$, $\ell = 1, \ldots, m$, from $f(\mathbf{b}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i)$, where m is the number of Monte Carlo simulations to be used, a number suggested to be large enough. A simulation method to draw samples from $f(\mathbf{b}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i)$, is described in Subsection 2.2.3.

2.2.2 A SAEM algorithm

As mentioned in Subsection 1.1.2, the SAEM circumvents the cumbersome problem of simulating a large number of missing values at every iteration, leading to a faster and efficient solution than the MCEM. In summary, the SAEM algorithm proceeds as follows:

E-step: Given $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$ for $i = 1, \ldots, n$;

- **Simulation step:** Draw $\mathbf{b}_i^{(\ell,k)}$, $\ell = 1, \ldots, m$, from $f(\mathbf{b}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i)$, for $m \leq 20$.
- **Stochastic approximation:** Update the MC approximations for the conditional expectations by their stochastic approximations, given by

$$
\begin{aligned}
S_{1,i}^{(k)} &= S_{1,i}^{(k-1)} + \delta_k \left[\frac{1}{m} \sum_{\ell=1}^{m} [\mathbf{x}_i \mathcal{E}(\mathbf{D}_i^{-1})^{(\ell,k)} \mathbf{x}_i^\top] - S_{1,i}^{(k-1)} \right], \\
S_{2,i}^{(k)} &= S_{2,i}^{(k-1)} + \delta_k \left[\frac{1}{m} \sum_{\ell=1}^{m} \left[\mathbf{x}_i \mathcal{E}(\mathbf{D}_i^{-1})^{(\ell,k)} \left[\mathbf{y}_i - \mathbf{z}_i^\top \mathbf{b}_i^{(\ell,k)} - \vartheta_p \mathcal{E}(\mathbf{u}_i)^{(\ell,k)} \right] \right] - S_{2,i}^{(k-1)} \right], \\
S_{3,i}^{(k)} &= S_{3,i}^{(k-1)} + \delta_k \left[\frac{1}{m} \sum_{\ell=1}^{m} \left[(\mathbf{y}_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p^{(k+1)} - \mathbf{z}_i \mathbf{b}_i^{(\ell,k)})^\top \mathcal{E}(\mathbf{D}^{-1})^{(\ell,k)} (\mathbf{y}_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p^{(k+1)} - \mathbf{z}_i \mathbf{b}_i^{(\ell,k)}) \right. \right. \\
&\qquad \left. \left. - 2\vartheta_p (\mathbf{y}_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p^{(k+1)} - \mathbf{z}_i \mathbf{b}_i^{(\ell,k)})^\top \mathbf{1}_{n_i} + \frac{\tau_p^4}{4} \mathcal{E}(\mathbf{u}_i)^{(\ell,k)\top} \mathbf{1}_{n_i} \right] - S_{3,i}^{(k-1)} \right], \\
S_{4,i}^{(k)} &= S_{4,i}^{(k-1)} + \delta_k \left[\frac{1}{m} \sum_{\ell=1}^{m} [\mathbf{b}_i^{(\ell,k)} \mathbf{b}_i^{(\ell,k)^\top}] - S_{4,i}^{(k-1)} \right].
\end{aligned}
$$

M-step: Update $\widehat{\boldsymbol{\theta}}^{(k)}$ by maximizing $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ over $\widehat{\boldsymbol{\theta}}^{(k)}$, which leads to the following expressions:

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_p^{(k+1)} &= \left[\sum_{i=1}^{n} S_{1,i}^{(k)} \right]^{-1} \sum_{i=1}^{n} S_{2,i}^{(k)}, \\
\widehat{\sigma}^{(k+1)} &= \frac{1}{3N\tau_p^2} \sum_{i=1}^{n} S_{3,i}^{(k)}, \\
\widehat{\Psi}^{(k+1)} &= \frac{1}{n} \sum_{i=1}^{n} S_{4,i}^{(k)}.
\end{aligned} \tag{2.2.9}
$$

Given a set of suitable initial values $\widehat{\boldsymbol{\theta}}^{(0)}$ (as detailed Appendix A.1), the SAEM iterates till convergence at iteration k if

$$
\max_i \left\{ \frac{|\widehat{\theta}_i^{(k+1)} - \widehat{\theta}_i^{(k)}|}{|\widehat{\theta}_i^{(k)}| + \delta_1} \right\} < \delta_2 \tag{2.2.10}
$$

is satisfied for three consecutive times where δ_1 and δ_2 are some small values pre established. The consecutive evalution of (2.2.10) avoids a fake convergence produced by an unlucky Monte Carlo

simulation. Based on Searle et al. (1992) pag. 269, we use $\delta_1 = 0.001$ and $\delta_2 = 0.0001$ as suggested by several researchers.

The proposed criterion above will need an extreme large number of iterations (more than usual) in order to detect convergence for parameters that are close to the boundary of the parametric space. In this case for variance components, a parameter value close to zero will inflate the ratio in (2.2.10) and the convergence will not be attained even though the likelihood was maximized with few iterations. As proposed by Booth and Hobert (1999) we use also a second convergence criteria besides to the first one, defined by

$$\max_i \left\{ \frac{|\widehat{\theta}_i^{(k+1)} - \widehat{\theta}_i^{(k)}|}{\sqrt{\widehat{\text{var}}(\theta_i^{(k)})} + \delta_1} \right\} < \delta_2, \tag{2.2.11}$$

where (2.2.11) evaluates the parameter estimates changes relative to their standard errors leading to a convergence detection even for bounded parameters. Also the values δ_1 and δ_2 are some small values pre established and not necessarily equal to the one for (2.2.10). Based on simulation we suggest to fix $\delta_1 = 0.0001$ and to test different values for δ_2 between 0.0001 and 0.0005 when smaller means more accuracy. We use $\delta_1 = 0.0001$ and $\delta_2 = 0.0002$ by default which assures us a high accuracy. This stopping criteria is similar to the one proposed by D. M. Bates and Watts (1981) for nonlinear least squares.

2.2.3 Missing data simulation method

In order to draw samples from $f(\mathbf{b}_i|\mathbf{y}_i, \boldsymbol{\theta})$, we utilize the Metropolis-Hastings (MH) algorithm (Hastings, 1970; Metropolis et al., 1953), a MCMC algorithm for obtaining a sequence of random samples from a probability distribution for which direct sampling is not possible. The MH algorithm proceeds as follows:

Given $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$, for $i = 1, \ldots, n$;

1. Start with an initial value $\mathbf{b}_i^{(0,k)}$.
2. Draw $\mathbf{b}_i^* \sim h(\mathbf{b}_i^*|\mathbf{b}_i^{(\ell-1,k)})$ from a proposal distribution with the same support as the objective distribution $f(\mathbf{b}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i)$.
3. Generate $U \sim U(0,1)$.
4. If $U > \min\left\{1\,,\ \frac{f\left(\mathbf{b}_i^*|\boldsymbol{\theta}^{(k)},\mathbf{y}_i\right)h\left(\mathbf{b}_i^{(0,k)}|\mathbf{b}_i^*\right)}{f\left(\mathbf{b}_i^{(0,k)}|\boldsymbol{\theta}^{(k)},\mathbf{y}_i\right)h\left(\mathbf{b}_i^*|\mathbf{b}_i^{(0,k)}\right)}\right\}$, return to the step 2, else $\mathbf{b}_i^{(\ell,k)} = \mathbf{b}_i^*$
5. Repeat steps 2-4 until m samples $(\mathbf{b}_i^{(1,k)}, \mathbf{b}_i^{(2,k)}, \ldots, \mathbf{b}_i^{(m,k)})$ are drawn from $\mathbf{b}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i$.

Note that the marginal distribution $f\left(\mathbf{b}_i|\mathbf{y}_i, \boldsymbol{\theta}\right)$ (omitting $\boldsymbol{\theta}$) can be represented as

$$f\left(\mathbf{b}_i|\mathbf{y}_i\right) \propto f\left(\mathbf{y}_i|\mathbf{b}_i\right) \times f\left(\mathbf{b}_i\right),$$

where $\mathbf{b}_i \sim N_q(\mathbf{0}, \boldsymbol{\Psi})$ and $f(\mathbf{y}_i|\mathbf{b}_i) = \prod_{j=1}^{n_i} f(y_{ij}|\mathbf{b}_i)$, with $y_{ij}|\mathbf{b}_i \sim ALD\left(\mathbf{x}_{ij}^\top \boldsymbol{\beta}_p + \mathbf{z}_{ij}\mathbf{b}_i, \sigma, p\right)$. Since the objective function is a product of two distributions (with both support lying in $\mathbb{R}$), a suitable choice for the proposal density is a multivariate normal distribution with the mean and variance-covariance matrix that are the stochastic approximations of the conditional expectation $\mathrm{E}(\mathbf{b}_i^{(k-1)}|\mathbf{y}_i)$ and the conditional variance $\mathrm{Var}(\mathbf{b}_i^{(k-1)}|\mathbf{y}_i)$ respectively, obtained from the last iteration of the SAEM algorithm. This candidate (with possible information about the shape of the target distribution) leads to better acceptance rate, and consequently a faster algorithm. The resulting chain $\mathbf{b}_i^{(1,k)}, \mathbf{b}_i^{(2,k)}, \ldots, \mathbf{b}_i^{(m,k)}$ is a MCMC sample from the marginal conditional distribution $f(\mathbf{b}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i)$. Due the dependent nature of these MCMC samples, at least 10 MC simulations are suggested.

2.3 Estimation of the likelihood and standard errors

2.3.1 Likelihood Estimation

Usual model selection criteria are based in the observed likelihood function. Then, given the observed data, the likelihood function $\ell_o(\boldsymbol{\theta}|\mathbf{y})$ of the model defined in (1.1.3) is given by

$$\ell_o(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^{n} \log\, f(\mathbf{y}_i|\boldsymbol{\theta})) = \sum_{i=1}^{n} \log \int_{\mathbb{R}^q} f(\mathbf{y}_i|\mathbf{b}_i;\boldsymbol{\theta})\, f(\mathbf{b}_i;\boldsymbol{\theta})\, d\mathbf{b}_i, \tag{2.3.1}$$

where the integral can be expressed as an expectation with respect to $\mathbf{b}_i$, i.e., $E_{\mathbf{b}_i}[f(\mathbf{y}_i|\mathbf{b}_i;\theta)]$. The evaluation of this integral is not available analytically and is often replaced by its MC approximation involving a large number of simulations. However, alternative importance sampling (IS) procedures might require a smaller number of simulations than the typical MC procedure. Following Meza et al. (2012), we can compute this integral using an IS scheme for any continuous distribution $\widehat{f}(\mathbf{b}_i;\boldsymbol{\theta})$ of $\mathbf{b}_i$ having the same support as $f(\mathbf{b}_i;\theta)$. Rewriting (2.3.1) as

$$\ell_o(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^{n} \log \int_{\mathbb{R}^q} f(\mathbf{y}_i|\mathbf{b}_i;\boldsymbol{\theta})\, \frac{f(\mathbf{b}_i;\boldsymbol{\theta})}{\widehat{f}(\mathbf{b}_i;\boldsymbol{\theta})}\, \widehat{f}(\mathbf{b}_i;\boldsymbol{\theta})\, d\mathbf{b}_i.$$

we can express it as an expectation with respect to $\mathbf{b}_i^*$, where $\mathbf{b}_i^* \sim \widehat{f}(\mathbf{b}_i^*;\theta)$. Thus, the likelihood function can now be expressed as

$$\ell_o(\boldsymbol{\theta}|\mathbf{y}) \approx \sum_{i=1}^{n} \log \left\{ \frac{1}{m} \sum_{\ell=1}^{m} \left[\prod_{j=1}^{n_i} [f(y_{ij}|\mathbf{b}_i^{*(\ell)};\boldsymbol{\theta})]\, \frac{f(\mathbf{b}_i^{*(\ell)};\boldsymbol{\theta})}{\widehat{f}(\mathbf{b}_i^{*(\ell)};\boldsymbol{\theta})} \right] \right\}, \tag{2.3.2}$$

where $\{\mathbf{b}_i^{*(\ell)}\}$, $l = 1, \ldots, m$, is a MC sample from $\widehat{f}(\mathbf{b}_i^*;\boldsymbol{\theta})$, and $f(\mathbf{y}_i|\mathbf{b}_i^{*(\ell)};\boldsymbol{\theta})$ is expressed as $\prod_{j=1}^{n_i} f(y_{ij}|\mathbf{b}_i^{*(\ell)};\boldsymbol{\theta})$ due to independence. An efficient choice for $\widehat{f}(\mathbf{b}_i^{*(\ell)};\theta)$ is $f(\mathbf{b}_i|\mathbf{y}_i)$. Therefore, we use the same proposal distribution discussed in Subsection 2.2.3, and generate samples $\mathbf{b}_i^{*(\ell)} \sim N_q(\widehat{\boldsymbol{\mu}}_{\mathbf{b}_i}, \widehat{\boldsymbol{\Sigma}}_{\mathbf{b}_i})$, where $\widehat{\boldsymbol{\mu}}_{\mathbf{b}_i} = \mathrm{E}(\mathbf{b}_i^{(w)}|\mathbf{y}_i)$ and $\widehat{\boldsymbol{\Sigma}}_{\mathbf{b}_i} = \mathrm{Var}(\mathbf{b}_i^{(w)}|\mathbf{y}_i)$, which are estimated empirically during the last few iterations of the SAEM at convergence.

2.3.2 Standard error approximation

Louis' missing information principle (Louis, 1982) relates the score function of the incomplete data log-likelihood with the complete data log-likelihood through the conditional expectation $\boldsymbol{\nabla}_o(\boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{\theta}}[\boldsymbol{\nabla}_c(\boldsymbol{\theta}; \mathbf{Y}_{com}|\mathbf{Y}_{obs})]$, where $\boldsymbol{\nabla}_o(\theta) = \partial\ell_o(\boldsymbol{\theta}; \mathbf{Y}_{obs})/\partial\theta$ and $\boldsymbol{\nabla}_c(\boldsymbol{\theta}) = \partial\ell_c(\theta; \mathbf{Y}_{com})/\partial\boldsymbol{\theta}$ are the score functions for the incomplete and complete data, respectively. As defined in Meilijson (1989), the empirical information matrix can be computed as

$$\mathbf{I}_e(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^{n} \mathbf{s}(\mathbf{y}_i|\boldsymbol{\theta})\, \mathbf{s}^\top(\mathbf{y}_i|\boldsymbol{\theta}) - \frac{1}{n}\, \mathbf{S}(\mathbf{y}|\boldsymbol{\theta})\, \mathbf{S}^\top(\mathbf{y}|\boldsymbol{\theta}), \tag{2.3.3}$$

where $\mathbf{S}(\mathbf{y}|\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathbf{s}(\mathbf{y}_i|\boldsymbol{\theta})$, with $\mathbf{s}(\mathbf{y}_i|\boldsymbol{\theta})$ the empirical score function for the i-th individual. Replacing $\boldsymbol{\theta}$ by its ML estimator $\hat{\boldsymbol{\theta}}$ and considering $\boldsymbol{\nabla}_o(\hat{\boldsymbol{\theta}}) = \mathbf{0}$, equation (2.3.3) takes the simple form

$$\mathbf{I}_e(\widehat{\boldsymbol{\theta}}|\mathbf{y}) = \sum_{i=1}^{n} \mathbf{s}(\mathbf{y}_i|\widehat{\boldsymbol{\theta}})\, \mathbf{s}^\top(\mathbf{y}_i|\widehat{\boldsymbol{\theta}}). \tag{2.3.4}$$

At the kth iteration, the empirical score function for the i-th subject can be computed as

$$\mathrm{s}(\mathbf{y}_i|\boldsymbol{\theta})^{(k)} = \mathbf{s}(\mathbf{y}_i|\boldsymbol{\theta})^{(k-1)} + \delta_k \left[\frac{1}{m}\sum_{\ell=1}^{m} \mathbf{s}(\mathbf{y}_i, \mathbf{q}^{(k,\ell)}; \boldsymbol{\theta}^{(k)}) - \mathbf{s}(\mathbf{y}_i|\boldsymbol{\theta})^{(k-1)}\right], \tag{2.3.5}$$

where $\mathbf{q}^{(\ell,k)}$, $\ell = 1, \ldots, m$, are the simulated missing values drawn from the conditional distribution $f(\cdot|\boldsymbol{\theta}^{(k-1)}, \mathbf{y}_i)$. Thus, at iteration k, the observed information matrix can be approximated as $\mathbf{I}_e(\boldsymbol{\theta}|\mathbf{y})^{(k)} = \sum_{i=1}^{n} \mathbf{s}(\mathbf{y}_i|\boldsymbol{\theta})^{(k)}\, \mathbf{s}^\top(\mathbf{y}_i|\boldsymbol{\theta})^{(k)}$, such that at convergence, $\mathbf{I}_e^{-1}(\widehat{\boldsymbol{\theta}}|\mathbf{y}) = (\mathbf{I}_e(\boldsymbol{\theta}|\mathbf{y})|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}})^{-1}$ is an estimate of the covariance matrix of the parameter estimates. Expressions for the elements of the score vector with respect to $\boldsymbol{\theta}$ are given in Appendix A.3.

2.4 Simulation studies

In this section, the finite sample performance of the proposed algorithm and its performance comparison with the Geraci and Bottai (2014) method is evaluated via simulation studies. These computational procedures were implemented using the `R` software (R Core Team, 2014). In particular, we consider the following linear mixed model:

$$y_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + \mathbf{z}_{ij}\mathbf{b}_i + \epsilon_{ij},\ i = 1, \ldots, n,\ j = 1, \ldots, 3, \tag{2.4.1}$$

where the goal is to estimate the fixed effects parameters $\boldsymbol{\beta}$ for a grid of percentiles $p = \{0.05, 0.10, 0.50, 0.90, 0.95\}$. We simulated a 3×3 design matrix $\mathbf{x}_{ij}^\top$ for the fixed effects $\boldsymbol{\beta}$, where the first column corresponds to the intercept and the other columns generated from a $N_2(\mathbf{0}, \mathbf{I}_2)$ density, for all $i = 1, \ldots, n$. We also simulated a 3×2 design matrix associated with the random effects, with the columns distributed as $N_2(\mathbf{0}, \mathbf{I}_2)$. The fixed effects parameters were chosen randomly as $\beta_1 = 0.8$, $\beta_2 = 0.5$ and $\beta_3 = 1$, $\sigma = 0.20$, and the matrix $\boldsymbol{\Psi}$ with elements $\Psi_{11} = 0.8$, $\Psi_{12} = 0.5$ and $\Psi_{22} = 1$. For varying sample sizes of $n = 50$, 100, 200 and 300, we generate 100 data samples for

each scenario. In addition, we also choose $m = 20$, $c = 0.2$ and $W = 500$ in order to assure a quick distribution convergence for the first 20% of iterations and a quite a.s. convergence.

For all scenarios, we compute the square root of the mean square error (RMSE), the bias (Bias) and the Monte carlo standard deviation (MC-Sd) for each parameter over the 100 replicates. They are defined as

$$\text{MC-Sd}(\widehat{\theta}_i) = \sqrt{\frac{1}{99}\sum_{j=1}^{100}(\widehat{\theta}_i^{(j)} - \overline{\widehat{\theta}_i})^2} \qquad \text{and} \qquad \text{Bias}(\widehat{\theta}_i) = \overline{\widehat{\theta}_i} - \theta_i \tag{2.4.2}$$

where $\text{RMSE}(\widehat{\theta}_i) = \sqrt{\text{MC-Sd}^2(\widehat{\theta}_i) + \text{Bias}^2(\widehat{\theta}_i)}$, the Monte carlo mean $\overline{\widehat{\theta}_i} = \frac{1}{100}\sum_{j=1}^{100}\widehat{\theta}_i^{(j)}$ (MC Mean) and ${\theta_i}^{(j)}$ is the estimate of θ_i from the j-th sample, $j = 1 \ldots 100$. In addition, we also computed the average of the standard deviations (IM-Sd) obtained via the observed information matrix derived in Subsection 4.2 and the 95% coverage probability (MC-CP) as $\text{CP}(\hat{\theta}_i) = \frac{1}{100}\sum_{j=1}^{100} I(\theta_i \in [\hat{\theta}_{i,LCL}, \hat{\theta}_{i,UCL}])$, where I is the indicator function such that θ_i lies in the interval $[\hat{\theta}_{i,LCL}, \hat{\theta}_{i,UCL}]$, with $\hat{\theta}_{i,LCL}$ and $\hat{\theta}_{i,UCL}$ as the estimated lower and upper bounds of the 95% CIs, respectively.

The results are summarized in Figure 2.1. We observe that the *Bias*, *SD* and *RMSE* for the regression parameters β_1 and β_2 tends to approach zero with increasing sample size (n), re-

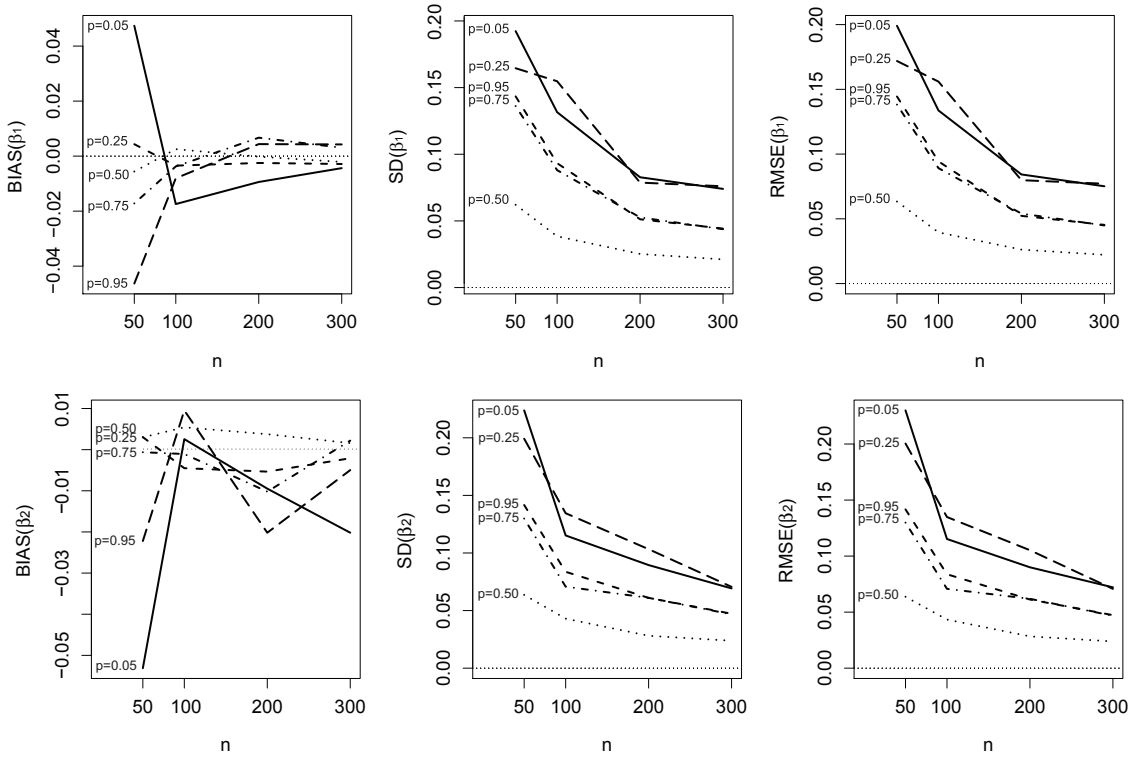


Figure 2.1: Bias, standard deviation and RMSE for β_1 (upper panel) and β_2 (lower panel) for varying sample sizes over the quantiles $p = 0.05$, 0.10, 0.50, 0.90, 0.95.

vealing that the ML estimates obtained via the proposed SAEM algorithm are conformable to the expected asymptotic properties. In addition, Table 2.2 presents the IM-Sd, MC-Sd and MC-CP for β_1 and β_2 across various quantiles. The estimates of MC-Sd and IM-Sd are very close, hence we can infer that the asymptotic approximation of the parameter standard errors are reliable. Furthermore, as expected, we observe that the MC-CP remains lower for extreme quantiles.

Finally, we compare the performance of SAEM algorithm with the approximate method proposed by Geraci (2014). The Geraci's algorithm can be implemented using the `R` package `lqmm()`. The results are presented in Table 2.1 and Figure A.1 in Appendix A. We observe that the RMSE from the proposed SAEM algorithm are lower than Geraci method across all scenarios, with the differences considerably higher for the extreme quantiles. Finally, Figure A.2 (see Appendix A) compares the differences in SD between the two methods for fixed effects β_1 and β_2 at specified quantiles reveals that the SD are mostly smaller for the SAEM method. Thus, we conclude that the SAEM algorithm produces more precise estimates.

RMSE									
		β_0		β_1		β_2		σ	
Quantile (%)	n	SAEM	Geraci	SAEM	Geraci	SAEM	Geraci	SAEM	Geraci
5	50	0.249	0.622	0.199	0.311	0.230	0.296	0.024	0.046
	100	0.209	0.496	0.134	0.180	0.115	0.165	0.017	0.037
	200	0.195	0.303	0.084	0.099	0.090	0.137	0.017	0.029
	300	0.163	0.345	0.075	0.100	0.072	0.101	0.012	0.031
10	50	0.159	0.382	0.144	0.187	0.142	0.201	0.023	0.048
	100	0.112	0.355	0.094	0.117	0.084	0.130	0.019	0.048
	200	0.082	0.231	0.052	0.087	0.061	0.081	0.017	0.036
	300	0.073	0.223	0.045	0.072	0.047	0.076	0.011	0.034
50	50	0.063	0.107	0.063	0.090	0.064	0.102	0.025	0.174
	100	0.042	0.052	0.040	0.056	0.043	0.070	0.021	0.196
	200	0.027	0.053	0.026	0.048	0.028	0.039	0.016	0.164
	300	0.024	0.034	0.022	0.022	0.024	0.040	0.012	0.180
90	50	0.160	0.389	0.138	0.159	0.130	0.177	0.025	0.050
	100	0.102	0.394	0.089	0.100	0.071	0.126	0.019	0.051
	200	0.085	0.240	0.054	0.097	0.062	0.078	0.014	0.038
	300	0.065	0.276	0.045	0.066	0.047	0.064	0.011	0.038
95	50	0.255	0.552	0.172	0.255	0.200	0.243	0.020	0.040
	100	0.233	0.470	0.156	0.169	0.135	0.161	0.020	0.036
	200	0.146	0.423	0.080	0.160	0.105	0.106	0.015	0.038
	300	0.157	0.468	0.077	0.113	0.071	0.061	0.014	0.036

Table 2.1: Simulation 1: Root Mean Squared Error (RMSE) for the fixed effects β_0, β_1, β_2 and the nuisance parameter σ, obtained after fitting the QRLMM and the Geraci (2014) model to simulated data under various settings of quantiles and sample sizes.

	β_1			β_2		
Quantile (%)	MC-Sd	IM-Sd	MC-CP	MC-Sd	IM-Sd	MC-CP
5	0.073	0.060	90	0.067	0.059	90
10	0.045	0.044	95	0.047	0.044	96
50	0.022	0.024	97	0.024	0.025	96
90	0.045	0.045	92	0.047	0.044	96
95	0.060	0.056	88	0.071	0.056	83

Table 2.2: Monte Carlo standard deviation (MC-Sd), mean standard deviation (IM-Sd) and Monte Carlo coverage probability (MC-CP) estimates of the fixed effects β_1 and β_2 from fitting the QR-LMM under various quantiles for sample size $n = 100$.

2.5 Applications

In this section, we illustrate the application of our method to two interesting longitudinal datasets from the literature via our developed `R` package `qrLMM`, currently available for free download from the `R CRAN` (Comprehensive R Archive Network).

2.5.1 Cholesterol data

The Framingham cholesterol study generated a benchmark dataset (D. Zhang and M. Davidian, 2001) for longitudinal analysis to examine the role of serum cholesterol as a risk factor for the evolution of cardiovascular disease. We analyze this dataset with the aim of explaining the full conditional distribution of the serum cholesterol as a function of a set of covariates of interest via modelling a grid of response quantiles. We fit a LMM model to the data as specified by

$$Y_{ij} = \beta_0 + \beta_1 \text{gender}_i + \beta_2 \text{age}_i + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij}, \tag{2.5.1}$$

where Y_{ij} is the cholesterol level (divided by 100) at the jth time point for the ith subject, $t_{ij} = (\tau - 5)/10$ where τ is the time measured in years from the start of the study, age denotes the subject's baseline age, gender is the dichotomous gender (0=female, 1=male), b_{0i} and b_{1i} the normal random intercept and slope, respectively, for subject i, and ϵ_{ij} the measurement error term assumed ALD, for 200 randomly selected subjects.

After fitting the QR-LMM over the grid $p = \{0.05, 0.10, \ldots, 0.95\}$, we present a graphical summary of the results in Figure 2.2. The figure displays the 95% confidence band for the fixed effects parameters $\beta_0, \beta_1, \beta_2$, and for the nuisance parameter σ. The solid lines represent the $Q_{0.025}$ and $Q_{0.975}$ percentiles, obtained from the estimated standard errors defined in Subsection 2.3.2. The figure reveals that the effect of gender and age become more prominent with increasing conditional quantiles (p). In addition, although age exhibits a positive influence on the cholesterol level across all quantiles, the confidence band for gender includes 0 across all quantiles, and hence its effect is non-significant. The estimated nuisance parameter σ is symmetric about $p = 0.5$, taking its maximum value at that point and decreasing for the extreme quantiles. Figure A.3 (see Appendix A) plots the fitted regression lines for the quantiles $0.10, 0.25, 0.50$(mean), 0.75 and 0.90

by gender. From this figure, it is clear how the extreme quantiles capture the full data variability and detect some atypical observations. The intercept of the quantile functions look very similar for both panels because of the non-significance of gender.

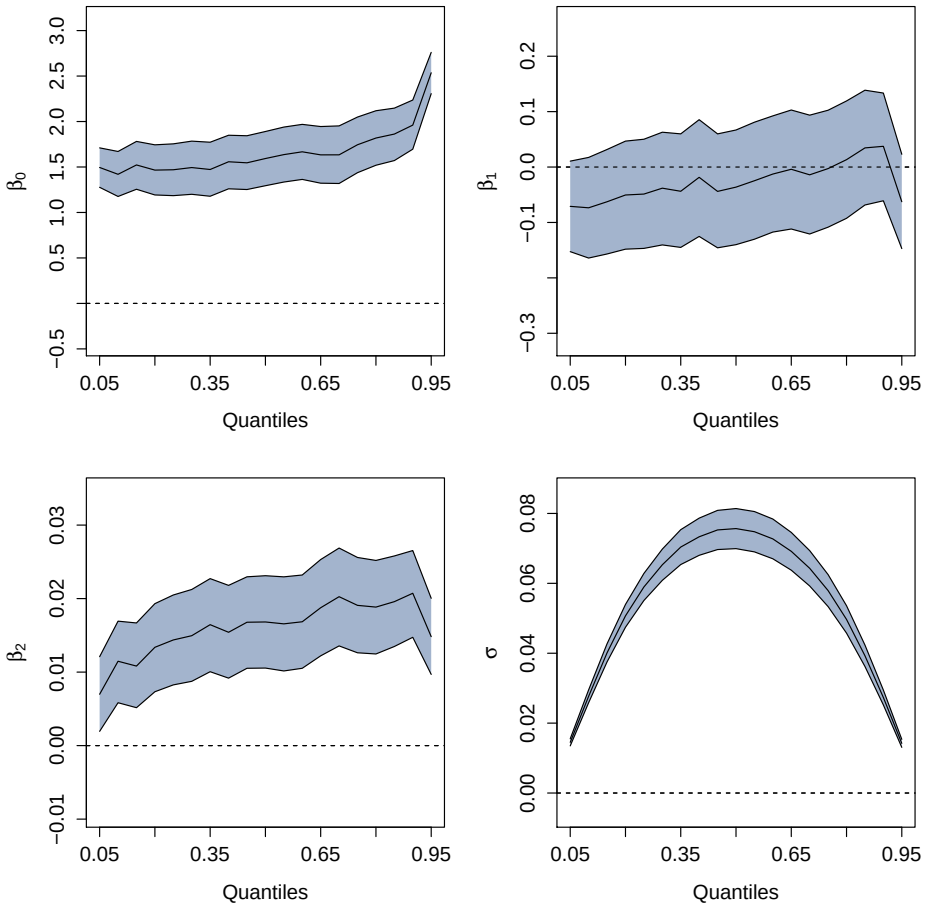


Figure 2.2: Point estimates (center solid line) and 95% confidence intervals for model parameters after fitting the QR-LMM using the `qrLMM` package to the Cholesterol data across various quantiles. The interpolated curves are spline-smoothed.

2.5.2 Orthodontic distance growth data

A second application was developed using a data set form a longitudinal orthodontic study (J. C. Pinheiro, Liu, et al., 2001; Potthoff and Roy, 1964) performed at the University of North Carolina Dental School. Here, researchers measured the distance between the pituitary and the pterygomaxillary fissure (two points that are easily identified on x-ray exposures of the side of the head) for 27 children (16 boys and 11 girls) every two years from age 8 until age 14. Similar to

Application 1, we fit the following LMM to the data:

$$Y_{ij} = \beta_0 + \beta_1 \text{gender}_i + \beta_2 t_{ij} + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij}, \tag{2.5.2}$$

where Y_{ij} is the distance between the pituitary and the pterygomaxillary fissure (in mm) at the jth time for the ith child, t_{ij} is the child's age at time j taking values 8, 10, 12, and 14 years, gender is a dichotomous variable (0=female, 1=male) for child i and ϵ_{ij} the random measurement error term. Initial exploratory plots for 10 random children in the left panel of Figure A.4 in Appendix A) suggest an increasing distance with respect to age. The individual profiles by gender (right panel) show differences between distances for boys and girls (distance for boys greater than those for girls), and hence we could expect a significant gender effect.
Once again, after fitting the QR-LMM over the grid $p = \{0.05, 0.10, \ldots, 0.95\}$, the point estimates

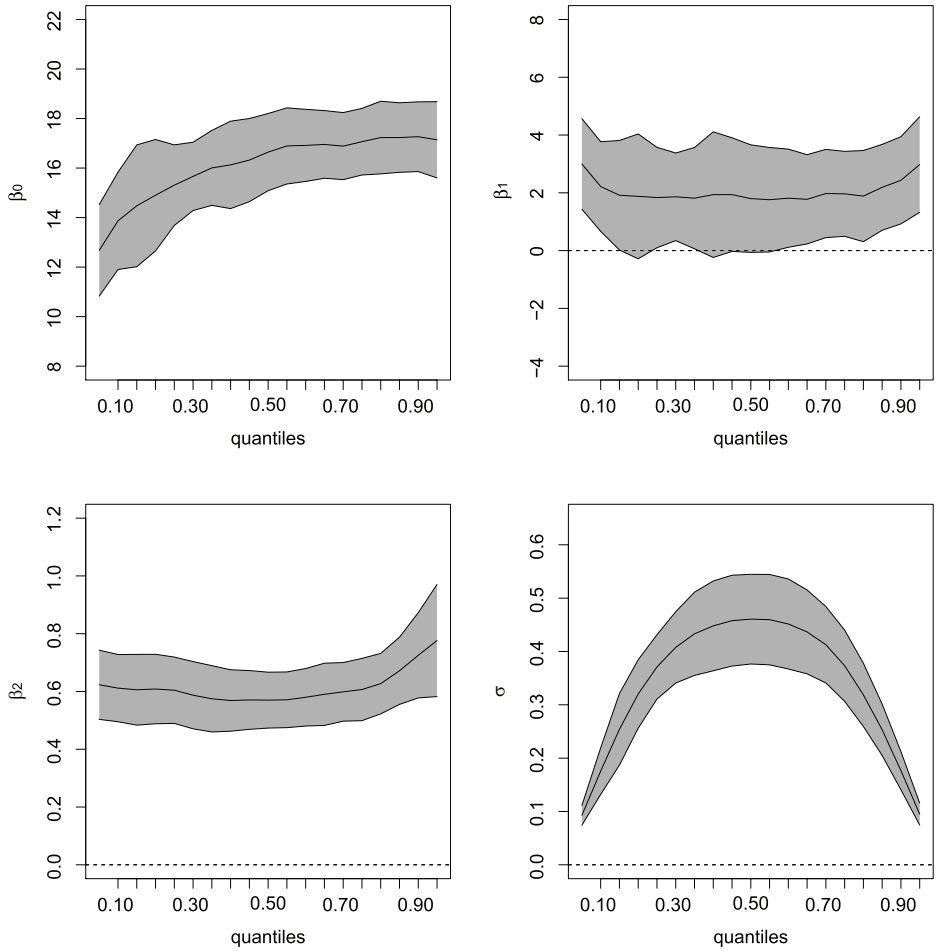


Figure 2.3: Point estimates (center solid line) and 95% confidence intervals for model parameters after fitting the QR-LMM using the `qrLMM` package to the orthodontic growth distance data across various quantiles. The interpolated curves are spline-smoothed.

and associated 95% confidence bands for model parameters are presented in Figure 2.3. From the figure, we infer that the effect of gender and age are significant across all quantiles, with their effect increasing for higher conditional quantiles. Effect of age is always positive across all quantiles, with a higher effect at the two extremes. σ behaves the same as in Application 1. Figure A.5 (in Appendix A) plots the fitted regression lines for the quantiles $0.10, 0.25, 0.50, 0.75$ and 0.90, overlayed with the individual profiles (gray solid lines), by gender. These fits capture the variability of the individual profiles, and also differ by gender due to its significance in the model. The `R` package also produces graphical summaries of point estimates and confidence intervals (95% by default) across various quantiles, as presented in Figures 2.2 and 2.3. Trace plots showing convergence of these estimates are presented in Figure A.6 in Appendix A. For example, for the 75th quantile, we can confirm that the convergence parameters for the SAEM algorithm ($M = 10,\ c = 0.25$ and $W = 300$) has been set adequately leading to a quick convergence in distribution within the first 75 iterations, and then converging almost surely to a local maxima in a total of 300 iterations. Sample output from the `qrLMM` package is provided in Appendix A.6.

2.6 Conclusions

In this work, we developed a likelihood-based inference for QR-LMM with the likelihood function based on the ALD. The ALD presents a convenient framework for the implementation of the SAEM algorithm leading to the exact ML estimation of the parameters. The methodology is illustrated via application to two longitudinal clinical datasets. We believe this work is the first attempt for exact ML estimation in the context of QR-LMMs, and thus provides an improvement over the Geraci and Bottai (2014) method. The methods developed are readily implementable via the `R` package `qrLMM()`.

Although the QR-LMM considered here has shown great flexibility to quantify the entire conditional distribution of the outcome variable, its robustness against outliers can be seriously affected by the presence of skewness and thick-tails. Recently, V. H. Lachos et al. (2010) proposed a remedy to accommodate these using scale mixtures of skew-normal distributions in the random effects. We conjecture that methodology can be transferred to the QR-LMM framework, and should yield satisfactory results at the expense of additional complexity in implementation. An in-depth investigation of such extension is beyond the scope of the present work, but certainly an interesting topic for future research.

Chapter 3

Quantile Regression for Nonlinear Mixed Models

Longitudinal data are frequently analyzed using normal mixed effects models. Moreover, the traditional estimation methods are based on mean regression, which leads to non-robust parameter estimation for non-normal error distributions. Compared to the conventional mean regression approach, quantile regression (QR) can characterize the entire conditional distribution of the outcome variable and is more robust to the presence of outliers and misspecification of the error distribution. This work develops a likelihood-based approach to analyzing QR models for correlated continuous longitudinal data via the asymmetric Laplace distribution (ALD). Exploiting the nice hierarchical representation of the ALD, our classical approach follows the Stochastic Approximation of the EM (SAEM) algorithm for deriving exact maximum likelihood estimates of the fixed-effects and variance components in nonlinear mixed effects models (NLMMs). We evaluate the finite sample performance of the algorithm and the asymptotic properties of the ML estimates through empirical experiments and applications to two real life datasets. The proposed SAEM algorithm is implemented in the `R` package `qrNLMM`.

3.1 Introduction

Nonlinear mixed-effects models (NLMMs) are frequently used to analyze grouped, clustered, longitudinal and multilevel data because of their potential to handle, on one hand, nonlinearities in the relationship between the observed response and the covariates and random effects, and on the other hand, to take into account within and between-subject correlations presented in this type of data (J. C. Pinheiro and D. M. Bates, 2000; L. Wu, 2010). Majority of these NLMMs estimate covariate effects on the response through a mean regression, controlling for between-cluster heterogeneity via normally-distributed cluster-specific random effects and random errors. However, this centrality-based inferential framework is often inadequate when the conditional distribution of the response (conditional on the random terms) is skewed, multimodal, or affected by atypical observations. In contrast, conditional quantile regression (QR) methods (Roger Koenker, 2004,

2005) quantifying the entire conditional distribution of the outcome variable were developed that can provide assessment of covariate effects at any arbitrary quantiles of the outcome. In addition, QR methods do not impose any specific distribution assumption on the error, except requiring that the error term has a zero conditional quantile such as the ALD.

Although QR was initially developed under a univariate framework, the abundance of clustered data in recent times lead to its extensions into mixed modeling framework via either the distribution-free route (Fu and Y.-G. Wang, 2012; Galvao Jr, 2011; Galvao and Montes-Rojas, 2010; Lipsitz et al., 1997), or the traditional likelihood-based route mostly using the ALD (Geraci and Bottai, 2007, 2014; Yuan and Yin, 2010). Among the ALD-based models, Geraci and Bottai (2007) proposed a Monte Carlo EM (MCEM)-based conditional QR model for continuous responses with a subject-specific random (univariate) intercept to account for within-subject dependence in the context of longitudinal data. However, due to the limitations of a simple random intercept model to account for the between-cluster heterogeneity, Geraci and Bottai (2014) extended their previous Geraci and Bottai (2007) model to a general linear quantile mixed effects regression model (QR-LMM) with multiple random effects (both intercepts and slopes). However, instead of going the MCEM route, the estimation of the fixed effects and the covariance components were implemented using an efficient combination of Gaussian quadrature approximations and non-smooth optimization algorithms. Yuan and Yin (2010) applied the version of QR of Geraci and Bottai (2007) to linear mixed effects models for longitudinal measurements with missing data. J. Wang (2012) considered QR-NLMMs from a Bayesian perspective and shown that QR-NLMMs may be a better measure of centrality for skewed or multimodal data and more robust against nonnormality of the distribution of random errors than the mean regression estimator. Although some results on QR-NLMMs have recently appeared in the literature, to the best of our knowledge, there seem to be no studies on exact inference for QR-NLMMs from a likelihood based perspective.

In this work, we proceed to achieve that via a robust parametric ALD-based QR-NLMMs, where the full likelihood-based implementation follows a stochastic version of the EM algorithm (SAEM), proposed by Delyon et al. (1999), for maximum likelihood (ML) estimation in contrast to the bayesian work proposed by J. Wang (2012) for QR-NLMMs. The SAEM algorithm has been proved to be more computationally efficient than the classical MCEM algorithm due to the recycling of simulations from one iteration to the next in the smoothing phase of the algorithm. Moreover, as pointed out by Meza et al. (2012) the SAEM algorithm, unlike the MCEM, converges even in a typically small simulation size. Recently, Kuhn and Lavielle (2005) showed that the SAEM algorithm is very efficient in computing the ML estimates in mixed effects models. Our empirical results shows that the ML estimates based on the SAEM algorithm do provide good asymptotic properties. Furthermore, application of our method to two longitudinal datasets is illustrated via the `R` package `qrNLMM()`.

3.2 QR for nonlinear mixed models and algorithms

We proposed the following general mixed-effects model. Let $\mathbf{y}_i = (y_{i1}, ..., y_{in_i})^\top$ denote the continuous response for subject i and let $\eta = (\eta(\phi_i, x_{i1}), ..., \eta(\phi_i, x_{in_i}))^\top$ represents a nonlinear differentiable function of vector-valued mixed-effects random parameters ϕ_i of dimension r and a matrix of covariates $\mathbf{x}_i$ of dimensions $n_i \times r$. We define the NLMM as

$$\mathbf{y}_i = \eta(\boldsymbol{\phi}_i, \mathbf{x}_i) + \boldsymbol{\epsilon}_i, \qquad \phi_i = \mathbf{A}_i\boldsymbol{\beta}_p + \mathbf{B}_i\mathbf{b}_i, \tag{3.2.1}$$

where $\mathbf{A}_i$ and $\mathbf{B}_i$ are design matrices of dimensions $r \times d$ and $r \times q$, respectively, possibly depending on elements of $\mathbf{x}_i$ and incorporating time varying covariates in fixed or random effects, $\boldsymbol{\beta}_p$ is the regression coefficient corresponding to the pth quantile, $\mathbf{b}_i$ is a q-dimensional random effects vector associated to the i-th subject and and ϵ_i the independent and identically distributed vector of random errors. We define pth quantile function of the response y_{ij} as

$$Q_p(y_{ij}|\mathbf{x}_{ij}, \mathbf{b}_i) = \eta(\phi_i, x_{ij}) = \eta(\mathbf{A}_i\boldsymbol{\beta}_p + \mathbf{B}_i\mathbf{b}_i, x_{ij}). \tag{3.2.2}$$

where Q_p denotes the inverse of the unknown distribution function F, the random effects $\mathbf{b}_i$ are distributed as $\mathbf{b}_i \overset{\text{iid}}{\sim} \mathrm{N}_q(\mathbf{0}, \boldsymbol{\Psi})$, where the dispersion matrix $\boldsymbol{\Psi} = \boldsymbol{\Psi}(\boldsymbol{\alpha})$ depends on unknown and reduced parameters $\boldsymbol{\alpha}$, and the errors are distributed as $\epsilon_{ij} \overset{\text{iid}}{\sim} ALD(0, \sigma)$ and both uncorrelated. Then, $y_{ij}|\mathbf{b}_i$ independently follows as ALD with the density given by

$$f(y_{ij}|\boldsymbol{\beta}_p, \mathbf{b}_i, \sigma) = \frac{p(1-p)}{\sigma} \exp\left\{-\rho_p\left(\frac{y_{ij} - \eta(\mathbf{A}_i\boldsymbol{\beta}_p + \mathbf{B}_i\mathbf{b}_i, \mathbf{x}_{ij})}{\sigma}\right)\right\}. \tag{3.2.3}$$

First, we develop a MCEM algorithm for ML estimation of the parameters in the QR-NLMM. The model exhibits a flexible hierarchical representation, which is useful in deriving the theoretical properties. From (1.1.3), the QR-NLMM defined in (3.2.2)-(3.2.3), can be represented in a hierarchical form as:

$$\begin{aligned} \mathbf{y}_i|\mathbf{b}_i, \mathbf{u}_i &\sim N_{n_i}\left(\boldsymbol{\eta}(\mathbf{A}_i\boldsymbol{\beta}_p + \mathbf{B}_i\mathbf{b}_i, \mathbf{x}_i) + \vartheta_p\mathbf{u}_i, \sigma\tau_p^2\mathbf{D}_i\right), \\ \mathbf{b}_i &\sim N_q\,(\mathbf{0}, \boldsymbol{\Psi}), \\ \mathbf{u}_i &\sim \prod_{j=1}^{n_i} \exp(\sigma), \end{aligned} \tag{3.2.4}$$

for $i = 1, \ldots, n$, where ϑ_p and τ_p^2 are as in (1.1.2); $\mathbf{D}_i$ represents a diagonal matrix that contains the vector of missing values $\mathbf{u}_i = (u_{i1}, \ldots, u_{in_i})^\top$ and $\exp(\sigma)$ denotes the exponential distribution with mean σ. Let $\mathbf{y}_{ic} = (\mathbf{y}_i^\top, \mathbf{b}_i^\top, \mathbf{u}_i^\top)^\top$, with $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$, $\mathbf{b}_i = (b_{i1}, \ldots, b_{iq})^\top$, $\mathbf{u}_i = (u_{i1}, \ldots, u_{in_i})^\top$and let $\boldsymbol{\theta}^{(k)} = (\boldsymbol{\beta}_p^{(k)\top}, \sigma^{(k)}, \boldsymbol{\alpha}^{(k)\top})^\top$, the estimate of θ at the k-th iteration. Since $\mathbf{b}_i$ and $\mathbf{u}_i$ are independent for all $i = 1, \ldots, n$, it follows from (1.1.3) that the complete-data log-likelihood function is of the form

$$\ell_c(\boldsymbol{\theta};\, \mathbf{y}_c) = \sum_{i=1}^{n} \ell_c(\boldsymbol{\theta};\, \mathbf{y}_{ic}),$$

where

$$\ell_c(\boldsymbol{\theta};\, \mathbf{y}_{ic}) = \text{constant} - \frac{3}{2}\, n_i log\sigma - \frac{1}{2}\, log|\boldsymbol{\Psi}| - \frac{1}{2}\, \mathbf{b}_i^\top \boldsymbol{\Psi}^{-1}\mathbf{b}_i - \frac{1}{\sigma}\, \mathbf{u}_i^\top \mathbf{1}_{n_i}$$
$$- \frac{1}{2\sigma\tau_p^2}\, (\mathbf{y}_i - \boldsymbol{\eta}(\mathbf{A}_i\boldsymbol{\beta}_p + \mathbf{B}_i\mathbf{b}_i, \mathbf{x}_i) - \vartheta_p\mathbf{u}_i)^\top \mathbf{D}_i^{-1}(\mathbf{y}_i - \boldsymbol{\eta}(\mathbf{A}_i\boldsymbol{\beta}_p + \mathbf{B}_i\mathbf{b}_i, \mathbf{x}_i) - \vartheta_p\mathbf{u}_i). \quad (3.2.5)$$

Since $\mathbf{A}_i$, $\mathbf{B}_i$ and $\mathbf{x}_i$ are known matrices, we will simplify the notation by writing $\boldsymbol{\eta}(\boldsymbol{\beta}_p, \mathbf{b}_i)$ to represent $\boldsymbol{\eta}(\boldsymbol{\phi}_i, \mathbf{x}_i) = \boldsymbol{\eta}(\mathbf{A}_i\boldsymbol{\beta}_p + \mathbf{B}_i\mathbf{b}_i, \mathbf{x}_i)$. Given the current estimate $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$, the E-step calculates the function

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \textstyle\sum_{i=1}^n Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}),$$

where

$$Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \mathrm{E}\left\{\ell_c(\boldsymbol{\theta};\, \mathbf{y}_{ic})|\boldsymbol{\theta}^{(k)}, \mathbf{y}\right\} \quad (3.2.6)$$
$$\propto -\frac{3}{2}\, n_i log\sigma - \frac{1}{2}\, log|\boldsymbol{\Psi}| - \frac{1}{2}\, \mathrm{tr}\Big\{\widehat{(\mathbf{b}\mathbf{b}^\top)_i}^{(k)} \boldsymbol{\Psi}^{-1}\Big\} - \frac{1}{2\sigma\tau_p^2}\Big[\mathbf{y}_i^\top \widehat{\mathbf{D}_i^{-1}}^{(k)} \mathbf{y}_i$$
$$- 2\vartheta_p \mathbf{y}_i^\top \mathbf{1}_{n_i} + \frac{\tau_p^4}{4}\widehat{\mathbf{u}}_i^{(k)\top}\mathbf{1}_{n_i} - 2\mathbf{y}_i^\top \widehat{(\mathbf{D}^{-1}\boldsymbol{\eta})}_i^{(k)} + 2\vartheta_p \mathbf{1}_{n_i}^\top \widehat{\boldsymbol{\eta}}_i^{(k)} + \widehat{\boldsymbol{\eta}_i^\top \mathbf{D}_i^{-1}\boldsymbol{\eta}_i}^{(k)}\Big]$$

where $\boldsymbol{\eta}_i = \boldsymbol{\eta}(\mathbf{A}_i\boldsymbol{\beta}_p + \mathbf{B}_i\mathbf{b}_i, \mathbf{x}_i)$ for simplicity, $\mathrm{tr}(\mathbf{A})$ indicates the trace of matrix $\mathbf{A}$ and $\mathbf{1}_p$ is the vector of ones of dimension p. The calculation of these function requires expressions for

$$\widehat{\boldsymbol{\eta}}_i^{(k)} = \mathrm{E}\left\{\boldsymbol{\eta}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i\right\}, \qquad \widehat{\mathbf{u}}_i^{(k)} = \mathrm{E}\left\{\mathbf{u}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i\right\},$$
$$\widehat{(\mathbf{b}\mathbf{b}^\top)_i}^{(k)} = \mathrm{E}\left\{\mathbf{b}_i\mathbf{b}_i^\top|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i\right\}, \qquad \widehat{\mathbf{D}_i^{-1}}^{(k)} = \mathrm{E}\left\{\mathbf{D}_i^{-1}|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i\right\},$$
$$\widehat{(\mathbf{D}^{-1}\boldsymbol{\eta})}_i^{(k)} = \mathrm{E}\left\{\mathbf{D}_i^{-1}\boldsymbol{\eta}_i^{(k)}|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i\right\}, \qquad \widehat{(\boldsymbol{\eta}^\top\mathbf{D}^{-1}\boldsymbol{\eta})}_i^{(k)} = \mathrm{E}\left\{\boldsymbol{\eta}_i^\top\mathbf{D}_i^{-1}\boldsymbol{\eta}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i\right\},$$

which do not have closed forms. Since the joint distribution of the missing data $(\mathbf{b}_i^{(k)}, \mathbf{u}_i^{(k)})$ is unknown and the conditional expectations cannot be computed analytically, for any function $g(.)$, the MCEM algorithm approximates the conditional expectations above by their Monte Carlo approximations

$$\mathrm{E}[\, g\,(\mathbf{b}_i, \mathbf{u}_i)\,|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i] \approx \frac{1}{m}\sum_{\ell=1}^m g(\mathbf{b}_i^{(\ell,k)}, \mathbf{u}_i^{(\ell,k)}), \quad (3.2.7)$$

which depend of the simulations of the two latent (missing) variables $\mathbf{b}_i^{(k)}$ and $\mathbf{u}_i^{(k)}$ from the conditional joint density $f(\mathbf{b}_i, \mathbf{u}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i)$. Using known properties of conditional expectations, the expected value in (3.2.7) can be more accurately approximated as

$$\mathrm{E}_{\mathbf{b}_i,\mathbf{u}_i}[\, g(\mathbf{b}_i, \mathbf{u}_i)|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i] = \mathrm{E}_{\mathbf{b}_i}[\, \mathrm{E}_{\mathbf{u}_i}[\, g(\mathbf{b}_i, \mathbf{u}_i)|\boldsymbol{\theta}^{(k)}, \mathbf{b}_i, \mathbf{y}_i]|\mathbf{y}_i\,]$$
$$\approx \frac{1}{m}\sum_{\ell=1}^m \mathrm{E}_{\mathbf{u}_i}[\, g(\mathbf{b}_i^{(\ell,k)}, \mathbf{u}_i)|\boldsymbol{\theta}^{(k)}, \mathbf{b}_i^{(\ell,k)}, \mathbf{y}_i], \quad (3.2.8)$$

where $\mathbf{b}^{(\ell,k)}$ is a sample from the conditional density $f(\mathbf{b}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i)$. Note that (3.2.8) is a more accurate approximation once it only depends of one MC approximation, instead two as needed in (3.2.7).

Now, to drawn random samples from the full conditional distribution $f(\mathbf{u}_i|\mathbf{y}_i, \mathbf{b}_i)$, first note that the vector $\mathbf{u}_i|\mathbf{y}_i, \mathbf{b}_i$ can be written as $\mathbf{u}_i|\mathbf{y}_i, \mathbf{b}_i = [\ u_{i1}|y_{i1}, \mathbf{b}_i, \quad u_{i2}|y_{i2}, \mathbf{b}_i, \quad \cdots \quad , u_{in_i}|y_{in_i}, \mathbf{b}_i\ ]^\top$, since $u_{ij}|\, y_{ij}, \mathbf{b}_i$ is independent of $u_{ik}|\, y_{ik}, \mathbf{b}_i$, for all $j, k = 1, 2, \ldots, n_i$ and $j \neq k$. Thus, the distribution of $f(u_{ij}|y_{ij}, \mathbf{b}_i)$ is proportional to

$$f(u_{ij}|y_{ij}, \mathbf{b}_i) \propto \phi(y_{ij}|\eta_{ij}(\boldsymbol{\beta}_p, \mathbf{b}_i) + \vartheta_p u_{ij},\, \sigma\tau_p^2 u_{ij}) \times \exp(\sigma),$$

which, from Subsection 2.1, leads to $u_{ij}|y_{ij}, \mathbf{b}_i \sim GIG(\,\frac{1}{2}, \chi_{ij}, \psi)$, where χ_{ij} and ψ are given by

$$\chi_{ij} = \frac{|y_{ij} - \eta_{ij}(\boldsymbol{\beta}_p, \mathbf{b}_i)|}{\tau_p\sqrt{\sigma}} \quad \text{and} \quad \psi = \frac{\tau_p}{2\sqrt{\sigma}} \tag{3.2.9}$$

From (1.1.5), and after generating samples from $f(\mathbf{b}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i)$ (see Subsection 3.2.2), the conditional expectation $\mathrm{E}_{\mathbf{u}_i}[\cdot|\boldsymbol{\theta}, \mathbf{b}_i, \mathbf{y}_i]$ in (3.2.8) can be computed analytically. Finally, the proposed MCEM algorithm for estimating the parameters of the QR-NLMM can be summarized as follows:

MC E-step: Given $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$, for $i = 1, \ldots, n$;

- **Simulation Step:** For $\ell = 1, \ldots, m$, draw $\mathbf{b}_i^{(\ell,k)}$ from $f(\mathbf{b}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i)$, as described later in Subsection 3.2.2.
- **Monte Carlo approximation:** Using (1.1.5) and the simulated sample above, evaluate

$$\mathrm{E}[\, g\,(\mathbf{b}_i, \mathbf{u}_i)\,|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i] \approx \frac{1}{m}\sum_{\ell=1}^{m} \mathrm{E}_{\mathbf{u}_i}[\, g(\mathbf{b}_i^{(\ell,k)}, \mathbf{u}_i)|\boldsymbol{\theta}^{(k)}, \mathbf{b}_i^{(\ell,k)}, \mathbf{y}_i].$$

M-step: Update $\widehat{\boldsymbol{\theta}}^{(k)}$ by maximizing $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) \approx \frac{1}{m}\sum_{l=1}^{m}\sum_{i=1}^{n} \ell_c(\theta;\, \mathbf{y}_i, \mathbf{b}_i^{(l,k)}, \mathbf{u}_i)$ over $\widehat{\boldsymbol{\theta}}^{(k)}$, which leads to the following estimates:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}_p}^{(k+1)} &= \widehat{\boldsymbol{\beta}_p}^{(k)} + \left[\sum_{i=1}^{n}\left\{\frac{1}{m}\sum_{\ell=1}^{m} \mathbf{J}_i^{(k)^\top}\mathcal{E}(\mathbf{D}_i^{-1})^{(\ell,k)}\mathbf{J}_i^{(k)}\right\}\right]^{-1} \times \\
&\qquad \left[\sum_{i=1}^{n}\left\{\frac{1}{m}\sum_{\ell=1}^{m}\left[2\mathbf{J}_i^{(k)^\top}\mathcal{E}(\mathbf{D}_i^{-1})^{(\ell,k)}\left[\mathbf{y}_i - \boldsymbol{\eta}(\widehat{\boldsymbol{\beta}_p}^{(k)}, \mathbf{b}_i^{(\ell,k)}) - \vartheta_p\mathcal{E}(\mathbf{u}_i)^{(\ell,k)}\right]\right]\right\}\right], \\
\widehat{\sigma}^{(k+1)} &= \frac{1}{3N\tau_p^2}\sum_{i=1}^{n}\left\{\frac{1}{m}\sum_{\ell=1}^{m}\left[(\mathbf{y}_i - \boldsymbol{\eta}(\widehat{\boldsymbol{\beta}_p}^{(k+1)}, \mathbf{b}_i^{(\ell,k)}))^\top\mathcal{E}(\mathbf{D}^{-1})^{(\ell,k)}(\mathbf{y}_i\boldsymbol{\eta}(\widehat{\boldsymbol{\beta}_p}^{(k+1)}, \mathbf{b}_i^{(\ell,k)}))\right.\right. \\
&\qquad \left.\left. -2\vartheta_p(\mathbf{y}_i\boldsymbol{\eta}(\widehat{\boldsymbol{\beta}_p}^{(k+1)}, \mathbf{b}_i^{(\ell,k)}))^\top\mathbf{1}_{n_i} + \frac{\tau_p^4}{4}\mathcal{E}(\mathbf{u}_i)^{(\ell,k)^\top}\mathbf{1}_{n_i}\right]\right\} \text{ and} \\
\widehat{\boldsymbol{\Psi}}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{m}\sum_{\ell=1}^{m}\mathbf{b}_i^{(\ell,k)}\mathbf{b}_i^{(\ell,k)^\top}\right],
\end{aligned}$$

where $\mathbf{J}_i = \partial\boldsymbol{\eta}(\boldsymbol{\beta}_p, \mathbf{b}_i)/\partial\boldsymbol{\beta}_p^\top$, $N = \sum_{i=1}^n n_i$ and expressions $\mathcal{E}(\mathbf{u}_i)^{(\ell,k)}$ and $\mathcal{E}(\mathbf{D}_i^{-1})^{(\ell,k)}$ are defined in Appendix B.2. Note that for the MC E-step, we need to draw samples $\mathbf{b}_i^{(\ell,k)}$, $\ell = 1, \ldots, m$, from $f(\mathbf{b}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i)$, where m is the number of Monte Carlo simulations to be used, a number suggested to be large enough. A simulation method to draw samples from $f(\mathbf{b}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i)$, is described in Subsection 3.2.2.

3.2.1 A SAEM algorithm

As mentioned in Subsection 1.1.2, the SAEM circumvents the cumbersome problem of simulating a large number of missing values at every iteration, leading to a more efficient solution than the MCEM. In summary, the SAEM algorithm proceeds as follows:

E-step: Given $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$ for $i = 1, \ldots, n$;

- **Stochastic approximation:** Update the MC approximations for the conditional expectations by their stochastic approximations, given by

$$
\begin{aligned}
S_{1,i}^{(k)} &= S_{1,i}^{(k-1)} + \delta_k \left[\frac{1}{m} \sum_{\ell=1}^{m} \mathbf{J}_i^{(k)^\top} \mathcal{E}(\mathbf{D}_i^{-1})^{(\ell,k)} \mathbf{J}_i^{(k)} - S_{1,i}^{(k-1)} \right], \\
S_{2,i}^{(k)} &= S_{2,i}^{(k-1)} + \delta_k \left[\frac{1}{m} \sum_{\ell=1}^{m} \left[2\mathbf{J}_i^{(k)^\top} \mathcal{E}(\mathbf{D}_i^{-1})^{(\ell,k)} \left[\mathbf{y}_i - \boldsymbol{\eta}(\widehat{\boldsymbol{\beta}_p}^{(k)}, \mathbf{b}_i^{(\ell,k)}) - \vartheta_p \mathcal{E}(\mathbf{u}_i)^{(\ell,k)} \right] \right] - S_{2,i}^{(k-1)} \right], \\
S_{3,i}^{(k)} &= S_{3,i}^{(k-1)} + \delta_k \left\{ \frac{1}{m} \sum_{\ell=1}^{m} \left[(\mathbf{y}_i - \boldsymbol{\eta}(\widehat{\boldsymbol{\beta}_p}^{(k+1)}, \mathbf{b}_i^{(\ell,k)}))^\top \mathcal{E}(\mathbf{D}^{-1})^{(\ell,k)} (\mathbf{y}_i - \boldsymbol{\eta}(\widehat{\boldsymbol{\beta}_p}^{(k+1)}, \mathbf{b}_i^{(\ell,k)})) \right. \right. \\
&\qquad \left. \left. -2\vartheta_p (\mathbf{y}_i - \boldsymbol{\eta}(\widehat{\boldsymbol{\beta}_p}^{(k+1)}, \mathbf{b}_i^{(\ell,k)}))^\top \mathbf{1}_{n_i} + \frac{\tau_p^4}{4} \mathcal{E}(\mathbf{u}_i)^{(\ell,k)^\top} \mathbf{1}_{n_i} \right] - S_{3,i}^{(k-1)} \right\} \quad \text{and} \\
S_{4,i}^{(k)} &= S_{4,i}^{(k-1)} + \delta_k \left[\frac{1}{m} \sum_{\ell=1}^{m} [\mathbf{b}_i^{(\ell,k)} \mathbf{b}_i^{(\ell,k)^\top}] - S_{4,i}^{(k-1)} \right].
\end{aligned}
$$

M-step: Update $\widehat{\boldsymbol{\theta}}^{(k)}$ by maximizing $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ over $\widehat{\boldsymbol{\theta}}^{(k)}$, which leads to the following expressions:

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}_p}^{(k+1)} &= \widehat{\boldsymbol{\beta}_p}^{(k)} + \left[\sum_{i=1}^{n} S_{1,i}^{(k)} \right]^{-1} \sum_{i=1}^{n} S_{2,i}^{(k)}, \\
\widehat{\sigma}^{(k+1)} &= \frac{1}{3N\tau_p^2} \sum_{i=1}^{n} S_{3,i}^{(k)}, \\
\widehat{\Psi}^{(k+1)} &= \frac{1}{n} \sum_{i=1}^{n} S_{4,i}^{(k)}.
\end{aligned} \tag{3.2.10}
$$

Note that the $\widehat{\boldsymbol{\beta}_p}^{(k+1)}$ term is updated iteratively at each stage via Newton-Raphson since it does not have a closed-form expression. Finally, given a set of suitable initial values $\widehat{\boldsymbol{\theta}}^{(0)}$ (see Appendix

B), the SAEM iterates till convergence at iteration k using the same stopping criterion defined in Chapter 2 for the Linear Model. Likelihood and Standard Error Estimation are also computed using importance sampling and the empirical information matrix as in subsection 2.3.

3.2.2 Missing data simulation method

In order to draw samples from $f(\mathbf{b}_i|\mathbf{y}_i, \boldsymbol{\theta})$, we utilize the Metropolis-Hastings (MH) algorithm (Hastings, 1970; Metropolis et al., 1953), a MCMC algorithm for obtaining a sequence of random samples from a probability distribution for which direct sampling is not possible. The MH algorithm proceeds as follows:

Given $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$, for $i = 1, \ldots, n$;

1. Start with an initial value $\mathbf{b}_i^{(0,k)}$.
2. Draw $\mathbf{b}_i^* \sim h(\mathbf{b}_i^*|\mathbf{b}_i^{(\ell-1,k)})$ from a proposal distribution with the same support as the objective distribution $f(\mathbf{b}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i)$.
3. Generate $U \sim U(0,1)$.
4. If $U > \min\left\{1\,,\ \frac{f\left(\mathbf{b}_i^*|\boldsymbol{\theta}^{(k)},\mathbf{y}_i\right)h\left(\mathbf{b}_i^{(0,k)}|\mathbf{b}_i^*\right)}{f\left(\mathbf{b}_i^{(0,k)}|\boldsymbol{\theta}^{(k)},\mathbf{y}_i\right)h\left(\mathbf{b}_i^*|\mathbf{b}_i^{(0,k)}\right)}\right\}$, return to the step 2, else $\mathbf{b}_i^{(\ell,k)} = \mathbf{b}_i^*$
5. Repeat steps 2-4 until m samples $(\mathbf{b}_i^{(1,k)}, \mathbf{b}_i^{(2,k)}, \ldots, \mathbf{b}_i^{(m,k)})$ are drawn from $\mathbf{b}_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}_i$.

Note that the marginal distribution $f\left(\mathbf{b}_i|\mathbf{y}_i, \boldsymbol{\theta}\right)$ (omitting $\boldsymbol{\theta}$) can be represented as

$$f\left(\mathbf{b}_i|\mathbf{y}_i\right) \propto f\left(\mathbf{y}_i|\mathbf{b}_i\right) \times f\left(\mathbf{b}_i\right),$$

where $\mathbf{b}_i \sim N_q(\mathbf{0}, \boldsymbol{\Psi})$ and $f\left(\mathbf{y}_i|\mathbf{b}_i\right) = \prod_{j=1}^{n_i} f\left(y_{ij}|\mathbf{b}_i\right)$, with $y_{ij}|\mathbf{b}_i \sim ALD(\eta(\mathbf{A}_i\boldsymbol{\beta}_p + \mathbf{B}_i\mathbf{b}_i, \mathbf{x}_{ij}), \sigma, p)$. Since the objective function is a product of two distributions (with both support lying in $\mathbb{R}$), a suitable choice for the proposal density is a multivariate normal distribution with the mean and variance-covariance matrix that are the stochastic approximations of the conditional expectation $\mathrm{E}(\mathbf{b}_i^{(k-1)}|\mathbf{y}_i)$ and the conditional variance $\mathrm{Var}(\mathbf{b}_i^{(k-1)}|\mathbf{y}_i)$ respectively, obtained from the last iteration of the SAEM algorithm. This candidate (with possible information about the shape of the target distribution) leads to better acceptance rate, and consequently a faster algorithm. The resulting chain $\mathbf{b}_i^{(1,k)}, \mathbf{b}_i^{(2,k)}, \ldots, \mathbf{b}_i^{(m,k)}$ is a MCMC sample from the marginal conditional distribution $f(\mathbf{b}_i|\boldsymbol{\grave{}}^{(k)}, \mathbf{y}_i)$. Due the dependent nature of these MCMC samples, at least 10 MC simulations are suggested.

3.3 Simulated data

In order to examine the performance of the proposed method, here we present some simulation studies. The first simulation study shows that the ML estimates based on the SAEM algorithm do provide good asymptotic properties. The second study investigates the consequences for population inferences when the normality assumption is inappropriate. We used heavy tailed distribution for the random error term in order to test the robustness of the proposed method in terms of parameter recovery.

3.3.1 Asymptotic properties

As in J. Pinheiro and D. Bates (1995), we performed the first simulation study with the following three parameter nonlinear growth-curve logistic model:

$$y_{ij} = \frac{\beta_1 + b_{1i}}{1 + \exp\left(-[t_{ij} - \beta_2]/\beta_3\right)} + \epsilon_{ij},\ i = 1, \ldots, n,\ \ j = 1, \ldots, 10, \tag{3.3.1}$$

where $t_{ij} = 100, 267, 433, 600, 767, 933, 1100, 1267, 1433, 1600$ for all i. The goal is to estimate the fixed effects parameters β's for a grid of percentiles $p = \{0.50, 0.75, 0.95\}$. A random effects b_{1i} was added to the first growth parameter β_1 and its effect over the growth-curve is shown in Figure 3.3. Parameters interpretation for this model is going to be discussed in the Application Section. The random effects b_{1i} and the error $\boldsymbol{\epsilon}_i = (\epsilon_{i1} \ldots, \epsilon_{i10})^\top$ are non-correlated been $b_{1i} \overset{\text{iid}}{\sim} N(0, \sigma_b^2)$ and $\epsilon_{ij} \overset{\text{iid}}{\sim} ALD(0, \sigma_e, p)$. We set $\boldsymbol{\beta}_p = (\beta_1, \beta_2, \beta_3)^\top = (200, 700, 350)^\top$, $\sigma_e = 0.5$, $\sigma_b^2 = 10$. Using the notation in (3.2.1) the matrices $\mathbf{A}_i$ and $\mathbf{B}_i$ are given by $\mathbf{I}_3$ and $(1, 0, 0)^\top$ respectively. For varying sample sizes of $n = 25$, 50, 100 and 200, we generate 100 data samples for each scenario. In addition, we also choose $m = 20$, $c = 0.25$ and $W = 500$ for the SAEM convergence parameters.

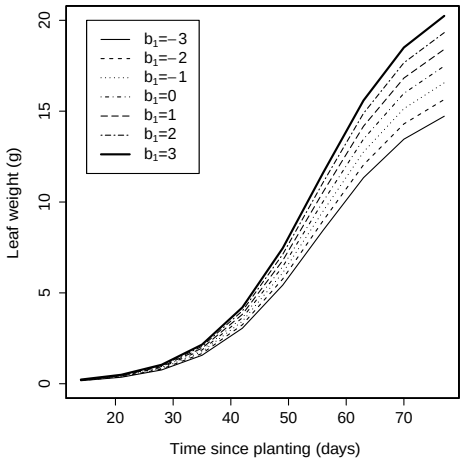


Figure 3.1: Illustration of the effect of including the random effect $b1_i$ in the first parameter of the nonlinear growth-curve logistic model.

For all scenarios, we compute the square root of the mean square error (RMSE), the bias (Bias) and the Monte carlo standard deviation (MC-Sd) for each parameter over the 100 replicates. They are defined as

$$\text{MC-Sd}(\widehat{\theta}_i) = \sqrt{\frac{1}{99}\sum_{j=1}^{100}(\widehat{\theta}_i^{(j)} - \overline{\widehat{\theta}_i})^2} \qquad \text{and} \qquad \text{Bias}(\widehat{\theta}_i) = \overline{\widehat{\theta}_i} - \theta_i \tag{3.3.2}$$

where $\text{RMSE}(\widehat{\theta}_i) = \sqrt{\text{MC-Sd}^2(\widehat{\theta}_i) + \text{Bias}^2(\widehat{\theta}_i)}$, the Monte carlo mean $\overline{\widehat{\theta}_i} = \frac{1}{100}\sum_{j=1}^{100}\widehat{\theta}_i^{(j)}$ (MC Mean) and ${\theta_i}^{(j)}$ is the estimate of θ_i from the j-th sample, $j = 1 \ldots 100$.

Based on Figure 3.2, for the bias we can see patterns of convergence to zero when n increases for both parameters. The values of MC-Sd and RMSE decrease monotonically when n is increased where it is evident that for extreme quantiles estimating, the standard deviation is much higher while for quantiles $q = 50$ and $q = 75$ are asymptotically equal. The worst scenario seems to happen while estimating extreme quantiles and maybe a sample size greater than 200 is needed to obtain a reasonably reduction of bias and SD. However, as a general rule, we can say that bias and MSE tend to approach to zero when the sample size is increasing, indicating that the approximates ML estimates based on the proposed SAEM algorithm do provide good asymptotic properties. The parameter β_1 has been discarded in the graphical analysis because it varies along

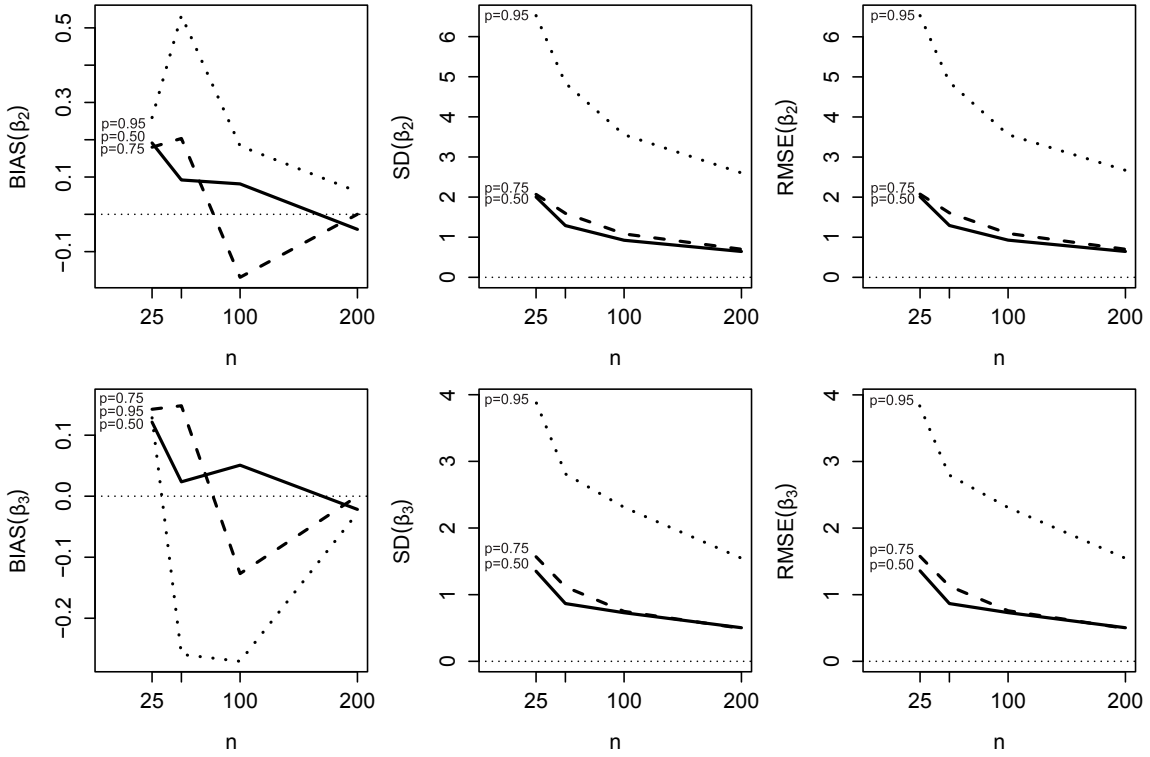


Figure 3.2: Bias, Standard Deviation and RMSE for β_1 (upper panel) and β_2 (lower panel) for varying sample sizes over the quantiles $p = 0.50, 0.90, 0.95$.

quantiles so its bias too as seen in Table 3.1. This parameter represents the asymptotic growth so this parameter is highly susceptible to the quantile to be estimated, however it also provides good asymptotic properties for its standard deviation. Table 3.1 also show an excellent recovery for the nuisance parameter σ_e, small standard deviations and good asymptotic properties in terms of bias and SD.

		β_1		β_2		β_3		σ_e	
Quantile (%)	n	MC Mean	MC-Sd	MC Mean	MC-Sd	MC Mean	MC-Sd	MC Mean	MC-Sd
50	25	199.75	(2.35)	700.19	(2.00)	350.13	(1.35)	0.503	(0.035)
	50	199.79	(1.69)	700.09	(1.29)	350.03	(0.86)	0.498	(0.021)
	100	200.16	(1.15)	700.08	(0.92)	350.06	(0.72)	0.497	(0.017)
	200	200.03	(0.75)	699.96	(0.64)	349.98	(0.50)	0.499	(0.012)
75	25	203.77	(2.50)	700.18	(2.07)	350.15	(1.56)	0.499	(0.035)
	50	203.90	(1.81)	700.20	(1.60)	350.16	(1.11)	0.495	(0.025)
	100	204.20	(1.31)	699.83	(1.08)	349.88	(0.74)	0.499	(0.017)
	200	204.34	(0.92)	700.00	(0.70)	350.01	(0.49)	0.498	(0.011)
95	25	201.15	(2.79)	700.26	(6.52)	350.14	(3.92)	0.506	(0.035)
	50	201.77	(2.15)	700.53	(4.84)	349.74	(2.83)	0.508	(0.024)
	100	201.94	(1.56)	700.18	(3.55)	349.73	(2.32)	0.505	(0.015)
	200	202.11	(1.08)	700.06	(2.60)	349.98	(1.54)	0.502	(0.012)

Table 3.1: Simulation 1: Results based on 100 simulated samples. Monte carlo mean and standard deviation (MC Mean and MC-Sd) for the fixed effects β_1. β_2. β_3 and the nuisance parameter σ_e. obtained after fitting the QR-NLMM model under different settings of quantiles and sample sizes.

3.3.2 Robustness study

The goal of this simulation study is to asses the robustness or bias incurred when one assumes a normal distribution for random effects and the actual distribution belongs to a heavy tailed distributions. The use of heavy tailed distributions for the random effects will let us to simulate the presence of outliers leading us to test adequately the performance of the proposed method in terms of robustness. The design of this simulation study is as in the previous subsection but for a set of quantiles $\{0.50, 0.75\}$ and a fixed sample size $n = 50$ we are going to simulate 100 Monte Carlo samples generating the random effect term from a Student-t distribution with $\nu = 4$ degrees of freedom and from a Normal Contaminated distribution ($\nu_1 = 0.1, \nu_2 = \{0.1, 0.2, 0, 3\}$), i.e., with three scenarios of contamination, 10%, 20% and 30%. All simulations are created by using the same values of $\boldsymbol{\beta}_p = (200, 700, 350)^\top$, nuisance parameter $\sigma_e = 0.5$ and scale parameter $\sigma_b^2 = 10$ for the respectively random effect distribution.

From Table 3.2 we can see that the proposed model is really robust even for worst scenarios of contamination. The parameter recovery is highly accurate even for the non-centered quantile 0.75. For quantile 0.75, the β_1 parameter estimate tends to increase for higher levels of contamination. As expected, the MC-Sd and consequently the RMSE increase in the presence of outliers. As a general

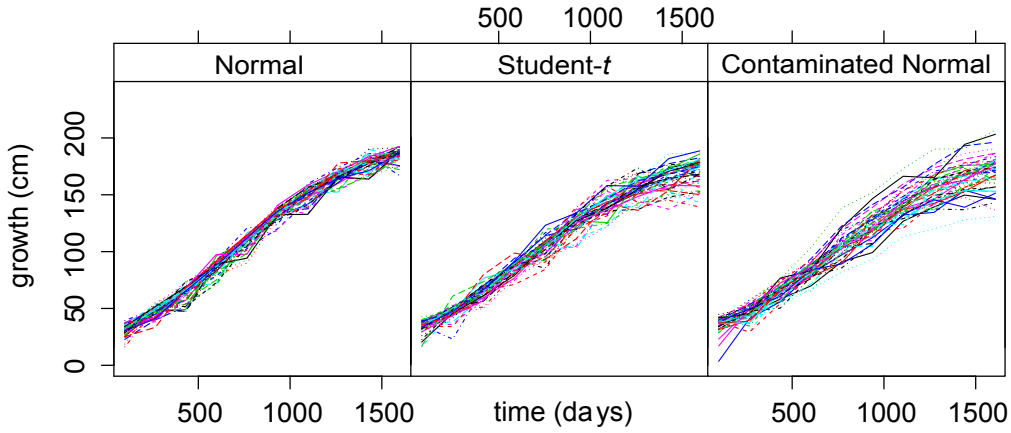


Figure 3.3: Illustration of 50 simulated curves from the growth-curve logistic model using different distributions for the random effect term. From left to right panel, the random effects has been generated from a Normal, a Student t_4 and a Contaminated Normal($\nu_1 = 0.1$,$\nu_2 = 0.1$), all with location parameter $\mu = 0$ and scale parameter $\sigma_b^2 = 10$.

Fit		Quantile 50%				Quantile 75%			
		β_1 (200)	β_2 (700)	β_3 (350)	σ_e (0.5)	β_1 (200)	β_2 (700)	β_3 (350)	σ_e (0.5)
Student-t_4	MC Mean	200.22	700.00	349.99	0.501	204.43	700.39	350.18	0.501
	Bias	0.22	0.00	-0.01	0.001	4.43	0.39	0.18	0.001
	MC-Sd	(1.98)	(1.28)	(0.98)	(0.024)	(2.17)	(1.69)	(1.09)	(0.024)
	RMSE	1.99	1.28	0.98	0.024	4.93	1.74	1.11	0.024
Contamination									
10%	MC Mean	199.87	700.10	349.9	0.499	205.02	700.18	350.05	0.501
	Bias	-0.13	0.10	-0.1	-0.001	5.02	0.18	0.05	0.001
	MC-Sd	(1.90)	(1.26)	(0.88)	(0.024)	(1.92)	(1.80)	(1.16)	(0.024)
	RMSE	1.90	1.27	0.88	0.024	5.38	1.81	1.16	0.024
20%	MC Mean	200.05	699.91	350.08	0.497	205.35	700.20	350.11	0.496
	Bias	0.05	-0.09	0.08	-0.003	5.35	0.20	0.11	-0.004
	MC-Sd	(1.96)	(1.28)	(0.90)	(0.024)	(2.00)	(1.55)	(1.19)	(0.023)
	RMSE	1.96	1.28	0.90	0.024	5.71	1.56	1.20	0.023
30%	MC Mean	200.16	700.06	350.07	0.496	206.63	699.91	350.01	0.497
	Bias	0.16	0.06	0.07	-0.004	6.63	-0.09	0.01	-0.003
	MC-Sd	(2.10)	(1.05)	(0.93)	(0.024)	(2.60)	(1.60)	(1.06)	(0.022)
	RMSE	2.11	1.05	0.93	0.024	7.13	1.60	1.06	0.023

Table 3.2: Simulation 2: Results based on 100 simulated samples. MC Mean, Bias, MC-Sd and RMSE for the fixed effects β_1, β_2, β_3 and the nuisance parameter σ_e obtained after fitting the QR-NLMM for quantiles 0.50 and 0.75 using different distribution settings for the random effects.

rule, we can conclude that the proposed model is robust in presence of outliers or misspecification of the random effect distribution.

3.4 Illustrative examples

In this section, we illustrate the application of our method to two interesting longitudinal datasets from the literature.

3.4.1 Growth curve: Soybean data

For the first application, we are going to consider the Soybean genotypes data analyzed by Marie Davidian and Giltinan (1995) and J. C. Pinheiro and D. M. Bates (2000), a longitudinal experiment consisting of measuring along time the leaf weight (in g) as a measure of growth of two kinds of Soybean genotype plants to be compared, a commercial variety, Forrest (F), and an experimental strain, Plan Introduction #416937 (P). The samples were taken approximately weekly during 8 to 10 weeks. For three consecutive years, 1988, 1989 and 1990, the plants were planted in 16 plots (8 per each genotype) and the mean leaf weight of six randomly selected plants was measured.

We use the three parameter logistic model in (3.3.1) introducing a random effect term for each parameter and a dichotomic covariate as

$$y_{ij} = \frac{\varphi_{1i}}{1 + \exp\left(-[t_{ij} - \varphi_{2i}]/\varphi_{3i}\right)} + \epsilon_{ij}, \;\; i = 1, \ldots, 412, \;\; j = 1, \ldots, n_i, \tag{3.4.1}$$

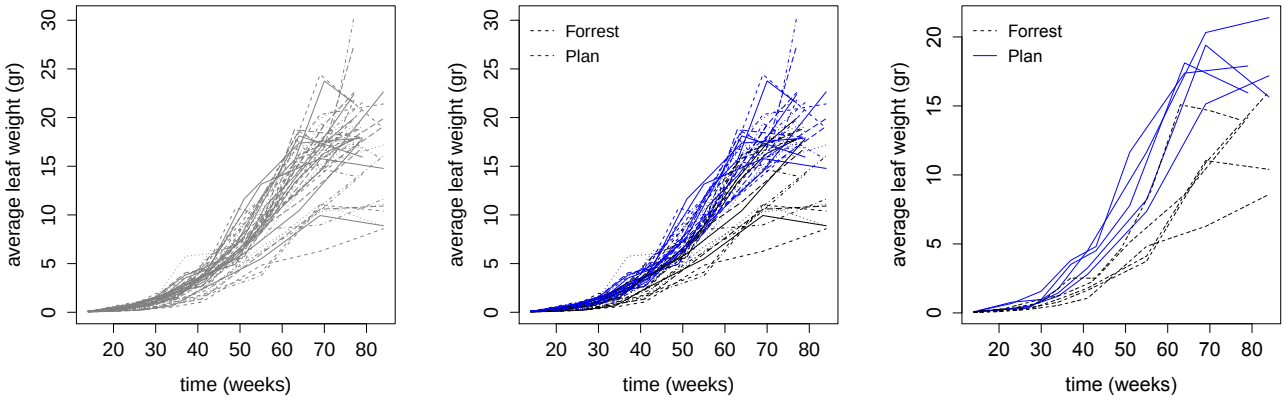


Figure 3.4: Soybean data: (a) Leaf weight profiles versus time. (b) Leaf weight profiles versus time by genotype. (c) Ten randomly selected leaf weight profiles versus time been five per each genotype.

where,

$$\begin{aligned}\varphi_{1i} =& \beta_1 + \beta_4 gen_i + b_{1i}\\ \varphi_{2i} =& \beta_2 + b_{2i}\\ \varphi_{3i} =& \beta_3 + b_{3i}.\end{aligned}$$

The observed value y_{ij} represents mean weight of leaves (in g) from six randomly selected soybean plants in the ith plot, after t_{ij} days of been planted; gen_i is a dichotomic variable for the genotype of plant i (0=forrest, 1=plan Introduction) and ϵ_{ij} is the measurement error for the 412 plants. Let be $\boldsymbol{\beta}_p = (\beta_1, \beta_2, \beta_3, \beta_4)^\top$ and $\mathbf{b}_i = (b_{1i}, b_{2i}, b_{3i})^\top$ the fixed and random effects vector respectively. Then the matrices $\mathbf{A}_i$ and $\mathbf{B}_i$ are defined as

$$\mathbf{A}_i = \begin{pmatrix} 1 & 0 & 0 & gen_i \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \qquad \text{and} \qquad \mathbf{B}_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.4.2}$$

The three parameter interpretation are 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 $0.7311 = 1/(1 + e^{-1})$ of its asymptotic weight, respectively. Due the goal of comparing the final (asymptotic) growth of the two kind of Soybeans, the dichotomic covariate gen_i was incorporated in the first component of the growth function, then the fourth fixed effect β_4 will represent the difference (in g) of the asymptotic leaf weight between the plan introduction type and the forrest one (control). As seen in middle and right panel of figure 3.4, it appears to exist a significance difference between the experimental and control Soybean so we expect a positive non zero β_4 estimate for most of quantiles.

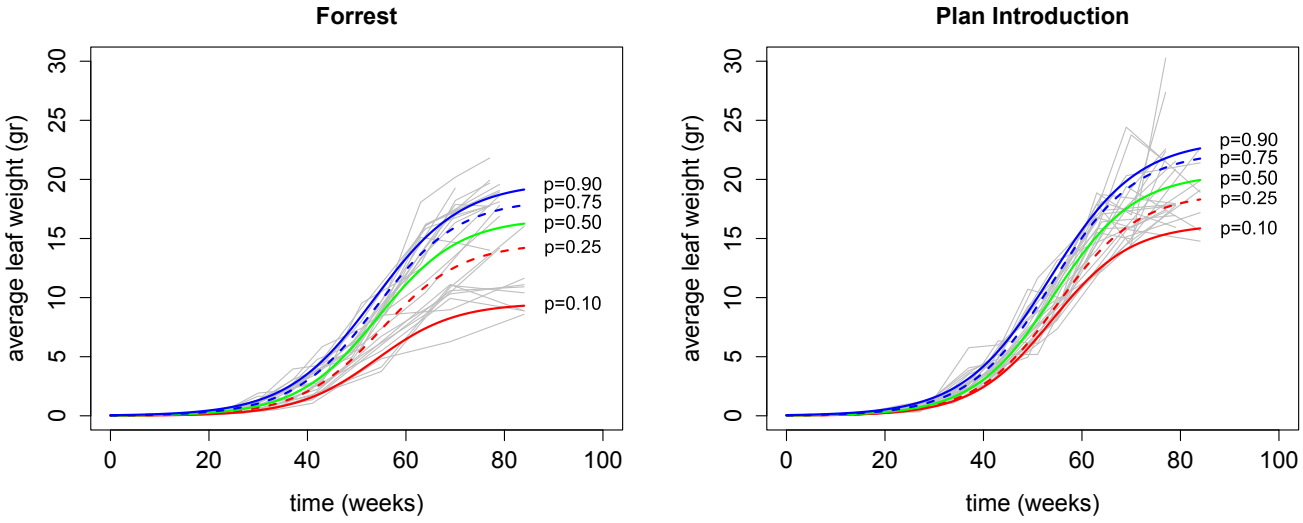


Figure 3.5: Fitted quantile regression for several quantiles for the Soybean data by genotype.

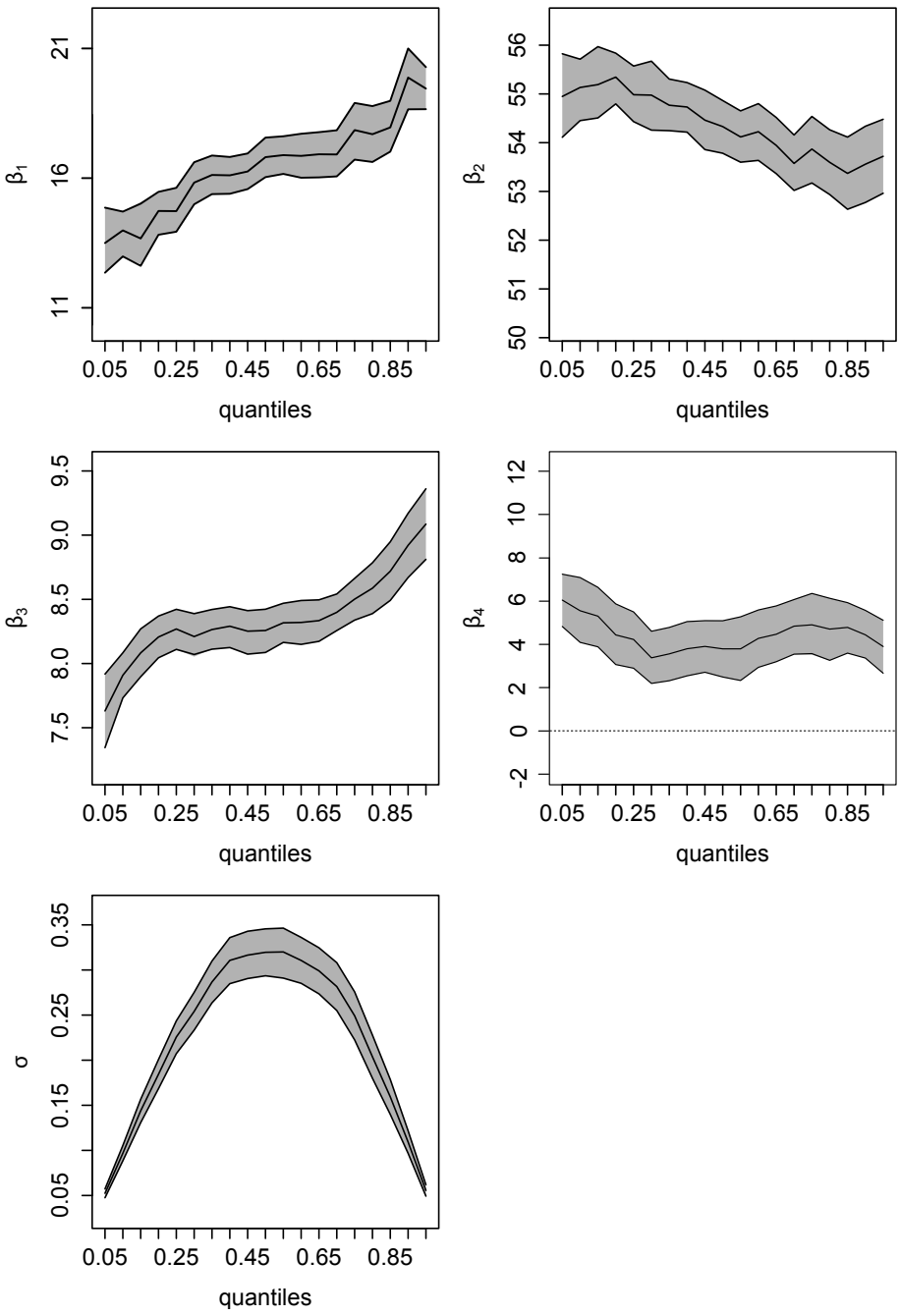


Figure 3.6: Point estimates (center solid line) and 95% confidence intervals for model parameters after fitting the QR to the Soybean data across various quantiles. The interpolated curves are spline-smoothed.

Figure 3.5 shows the fitted regression lines for quantiles 0.10, 0.25, 0.50, 0.75 and 0.90 by genotype. From this figure we can see clear how the extreme quantiles estimation functions captures the full data variability and evidences some atypical observations, specially for the plan introduction genotype. Quantile functions (for same quantile value) looks really different for each genotype due the significance of β_4 over the model as seen in Figure 3.6.

After fitting the quantile regression over the grid $p = \{0.05, 0.10, ..., 0.95\}$, we show a graphical summary of the obtained results in Figure 3.6. It shows a 95% confidence band for the fixed effect parameters β_1, β_2, β_3, β_4 and for the nuisance parameter σ where the solid lines are the $Q_{0.025}$ percentile and $Q_{0.975}$ percentile obtained through the estimation of the standard errors based on the empirical information matrix. We can see that the effect of the genotype results significant for all the quantile profile and the difference varies with respect to the conditional quantile been more significant for lower quantiles. We assessed the convergence of the fixed effect estimates, variance components of the random effects and nuisance parameters using graphical criteria as shown in Figure B.1 in Appendix B.4.

This can be corroborated in Figure 3.5 where the difference between the 0.10 estimated quantile functions for different genotypes is greater than for other quantiles. Using the information provided by the 95th percentile, we infer that the Soybean plants that grew more have a mean leaf weight around 19.35 grams for the Forrest genotype and 23.25 grams for the plan introduction one, then the asymptotic difference for the two genotypes is around 4 grams. The behavior of the estimate of the nuisance parameter σ is symmetric with respect to $p = 0.50$, taking its maximum value and variability on it and both decreasing for extreme quantiles. This behavior is because the variance within subjects depends of the quantile to be estimated, been proportional to the asymmetry of the error term then for extreme quantiles the nuisance parameter should be reduced. Sample output from the `qrNLMM` package is provided in Appendix B.6.

3.4.2 HIV viral load study

The data set belongs to a clinical trial (ACTG 315) studied in previous works by Lang Wu (2002) and Lachos et al. (2013). In this study, we analyze the HIV viral load of 46 HIV-1 infected patients under antiretroviral treatment (protease inhibitor and reverse transcriptase inhibitor drugs). The viral load and some other covariates were mesured several times days after the start of treatment being 4 and 10 the minimum and maximum number of measures per patient respectively. Lang Wu (2002) found that the only significant covariate for modelling the virus load was the CD4 therefore the other covariates even though they could be incorporated to the model for instance they are going to be discarded. Figure 3.7 shows the profile of viral load in log10 scale and CD4 cell count/100 per cubic millimeter versus time (in days/100) for six randomly selected patients. We can see that appear to exist some relationship between the viral load and the CD4 cell count and it seems to be inversely proportional, i.e., high CD4 cell count leads to lower levels of viral load. This is because the CD4 cells (also called T-cells) alert the immune system to invasion of viruses and/or bacteria so lower CD4 count means a weaker immune system. Normal counts of CD4 cells are from 500-1000 cells per cubic millimeter whereas fewer counts than 200 cells/mm^3 will be a high qualification to diagnose AIDS. We can evidence the mentioned before in the right

panel of Figure 3.7 where the three patients who have less than 200 CD4 cells/mm^3 (delimited by the horizontal dashed line in 0.02) are the ones with higher levels of viral load.

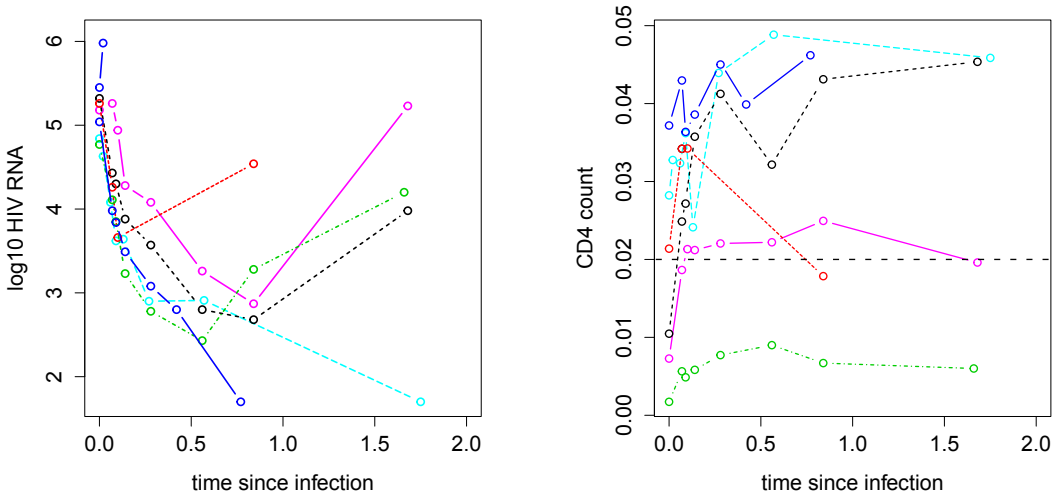


Figure 3.7: ACTG 315 data. Profiles of viral load (response) in log10 scale and CD4 cell count (in cells/100mm^3) for six randomly selected patients.

In order to fit the nonlinear data we will use the bi-exponential nonlinear model proposed by Lang Wu (2002) and also used by Lachos et al. (2013), a nonlinear model widely used for modeling viral load. The proposed bi-exponential NLME model is given by:

$$y_{ij} = \log_{10}\left(e^{(\varphi_{1i}-\varphi_{2i}t_{ij})} + e^{(\varphi_{3i}-\varphi_{4i}t_{ij})}\right) + \epsilon_{ij}, \qquad i = 1,\ldots,46, \;\; j = 1,\ldots,n_i, \tag{3.4.3}$$

with

$$\begin{aligned} \varphi_{1i} &= \beta_1 + b_{1i} & \varphi_{2i} &= \beta_2 + b_{2i} \\ \varphi_{3i} &= \beta_3 + b_{3i} & \varphi_{4ij} &= \beta_4 + \beta_5 CD4_{ij} + b_{4i}, \end{aligned}$$

where the observed value y_{ij} represents the log-10 transformation of the viral load for the ith patient at time j, $CD4_{ij}$ is the CD4 cell count (in cells/100mm^3) for the ith patient at time j and ϵ_{ij} is the measurement error for the 46 patients. Let be $\boldsymbol{\beta}_p = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^\top$ and $\mathbf{b}_i = (b_{1i}, b_{2i}, b_{3i}, b_{4i})^\top$ the fixed and random effects vector respectively and $\mathbf{CD4}_i = (CD4_{i1}, \ldots, CD4_{in_i})^\top$. Then the matrices $\mathbf{A}_i$ and $\mathbf{B}_i$ are defined as

$$\mathbf{A}_i = \begin{pmatrix} \mathbf{I}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_i} & \mathbf{CD4}_i \end{pmatrix} \qquad \text{and} \qquad \mathbf{B}_i = \begin{pmatrix} \mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_i} \end{pmatrix}. \tag{3.4.4}$$

The parameters φ_{2i} and φ_{4i} are the two-phase viral decay rates, which represent the minimum turnover rates of productively infected cells and that of latently or long-lived infected cells if therapy was successful, respectively. For more details about the model in (3.4.3) (see Zvi Grossman et al. (1999) and Perelson et al. (1997)).

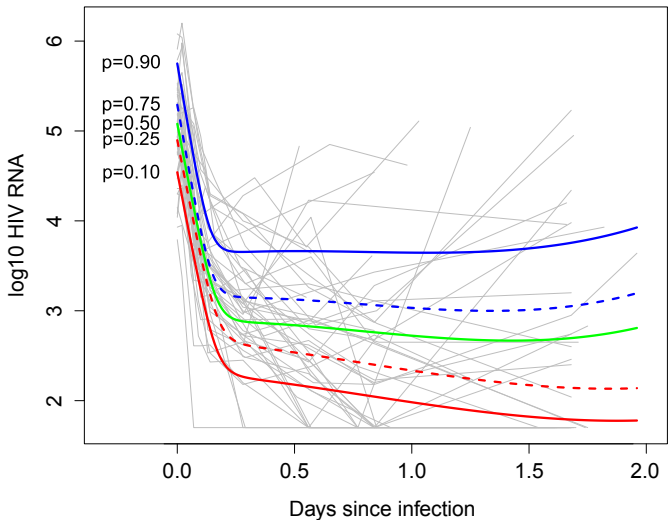


Figure 3.8: ACTG 315 data: Fitted quantile regression functions overlayed for the HIV data.

Figure 3.8 shows the fitted regression lines for quantiles 0.10, 0.25, 0.50, 0.75 and 0.90 for the HIV data. In order to plot, first, we fixed the CD4 covariate using the predicted sequence from a linear regression (including a quadratic term) for explaning the CD4 cell count with respect to time. We can see how quantile estimated functions follow the data behaviour satisfactorily and turn easily to estimate a specific viral load quantile at any time of the experiment. Extreme quantile functions bound the most of the observed profiles and evidence possible influential observations.

The results after fitting QR over the grid of quantiles $p = \{0.05, 0.10, ..., 0.95\}$ are shown in figure 3.9. The convergence of estimates for all parameters were also assessed using the graphical criteria in Figure B.2 in Appendix B.4. Based on Figure 3.9, we have found that the first phase viral decay rate is positive and its effect tends to increase proportionally along quantiles. For the second phase viral decay rate we have that this second rate is positive correlated with the CD4 count and therefore with the therapy time. Then, more days of treatment implies a higher CD4 cell count and therefore a higher second phase viral decay. The CD4 cell process for this model has a different behavior than for the expansion phase (Huang and Dagne, 2011). The significance of the CD4 covariate increases positively with respect to quantiles (until quantile $p = 0.60$ approximately) and then its effect becomes constant for greater quantiles. The behavior of the estimate of the nuisance parameter σ is the same as in Application 1.

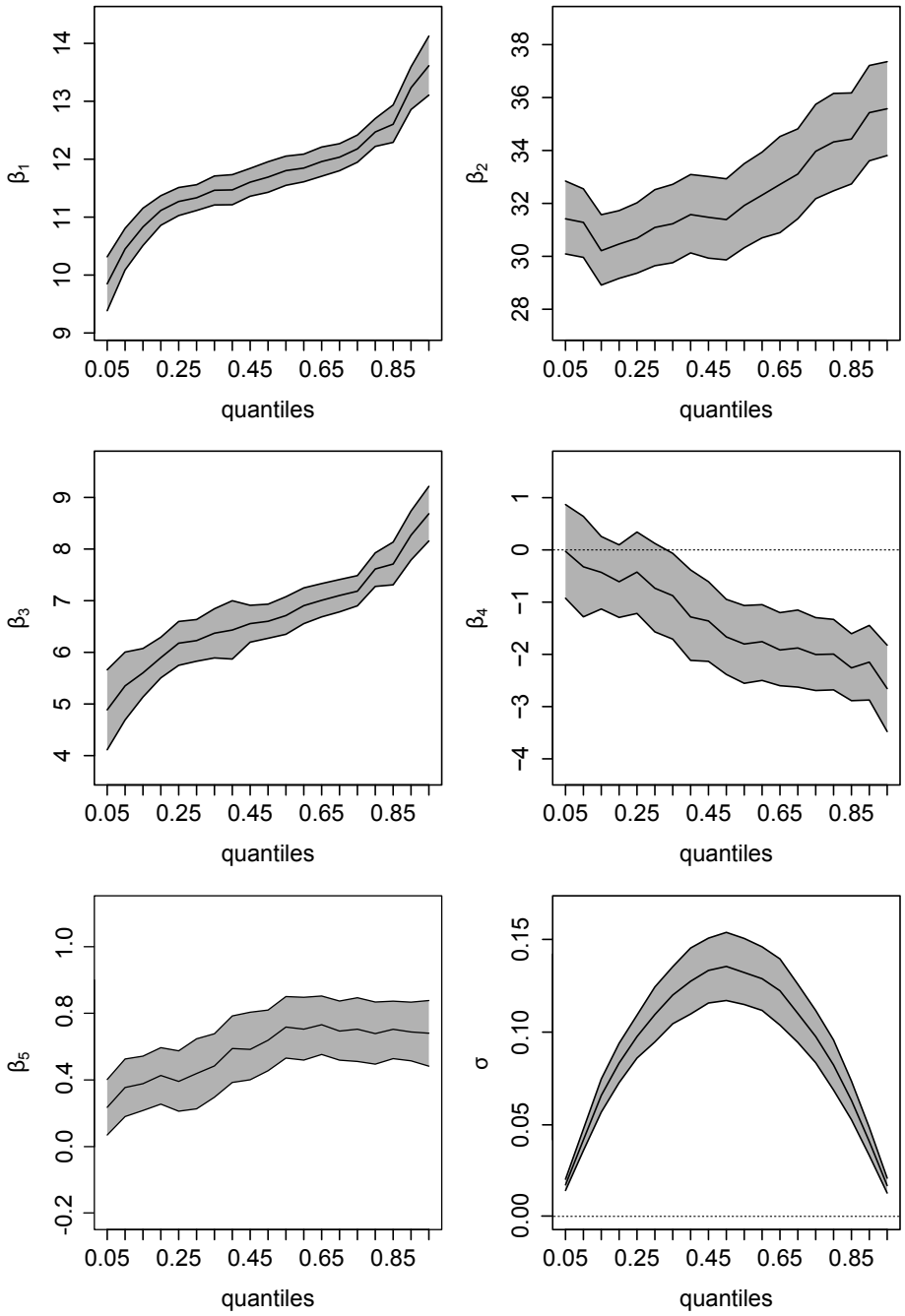


Figure 3.9: ACTG 315 data: Point estimates (center solid line) and 95% confidence intervals for model parameters after fitting the QR-NLMM to the HIV data across various quantiles. The interpolated curves are spline-smoothed.

3.5 Conclusions

In this work, we investigate quantile regression under non-linear mixed effects models from a likelihood-based perspective. The AL distribution and SAEM algorithm are combined efficiently to propose an exact ML estimation method, in contrast to the approximated method proposed by Geraci and Bottai, 2014. We evaluate the robustness of estimates, as well as the finite sample performance of the algorithm and the asymptotic properties of the ML estimates through empirical experiments. To the best of our knowledge, we consider that this work is the first attempt for exact ML estimation in the context of QR-NLME models. The methods developed can be readily implemented inside `R` through package `qrNLMM()`, making our approach quite powerful and accessible to practitioners.

We apply our method to a two data set from longitudinal studies, obtained interesting results from the point of view of quantile estimation. Moreover, in the two applications considered, similar conclusions to the previous analysis of these data sets have been obtained as in Marie Davidian and Giltinan (1995) and Lachos et al. (2013).

There are a large number of possible extensions of the current work. For modelling both skewness and heavy tails in the random effects, the use of scale mixtures of skew-normal (SMSN) distributions (V. H. Lachos et al., 2010) is a feasible choice. Also, HIV viral loads studies include covariates (CD4 cell counts) that often comes with substantial measurement errors (Lang Wu, 2002). How to incorporate measurement error in covariates within our robust framework can also be part of future research. An in-depth investigation of such extensions is beyond the scope of the present work, but certainly an interesting topic for future research.

Chapter 4

Concluding remarks

4.1 Technical production

In this section, we describe the technical production as result of this thesis.

4.1.1 Submitted papers

- "*Quantile Regression for Linear Mixed Models: A Stochastic Approximation EM approach*"
  Journal: Statistica Sinica

- "*Likelihood-Based Inference For Quantile Regression Under Nonlinear Mixed Effects Models*"
  Journal: Statistics and Computing

4.1.2 `R` packages

`ald`*: The Asymmetric Laplace Distribution*

It provides the probability density function, distribution function, quantile function, random number generator function, likelihood function, moment generating function and MLE for a given sample, all this for the three parameter Asymmetric Laplace Distribution with pdf (1.1.1) as defined in R. Koenker and Machado (1999) and K. Yu and Moyeed (2001) useful for quantile regression. It is available to download for free in the website: *http://cran.r-project.org/web/packages/ald/index.html*.

Description

The principal functions of this package are the `p,d,q,r` functions, that computes the probability density function, distribution function, quantile function and random number generator function. They can be run as:

R code

```
dALD(y, mu = 0, sigma = 1, p = 0.5)
pALD(q, mu = 0, sigma = 1, p = 0.5, lower.tail = TRUE)
qALD(prob, mu = 0, sigma = 1, p = 0.5, lower.tail = TRUE)
rALD(n, mu = 0, sigma = 1, p = 0.5)
```

Arguments

`y,q`	vector of quantiles.
`prob`	vector of probabilities.
`n`	number of observations.
`mu`	location parameter.
`sigma`	scale parameter.
`p`	skewness parameter.
`lower.tail`	logical; if `TRUE` (default), probabilities are $P[X \leq x]$ otherwise, $P[X > x]$.

Details

If `mu`, `sigma` or `p` are not specified they assume the default values of 0, 1 and 0.5, respectively, belonging to the Symmetric Standard Laplace Distribution denoted by $ALD(0, 1, 0.5)$. The scale parameter `sigma` must be positive and non zero. The skew parameter `p` must be between zero and one ($0 < p < 1$). The numerical arguments other than `n` are recycled to the length of the result.

Value

`dALD` gives the density, `pALD` gives the distribution function, `qALD` gives the quantile function, and `rALD` generates a random sample. The length of the result is determined by `n` for `rALD`, and is the maximum of the lengths of the numerical arguments for the other functions `dALD`, `pALD` and `qALD`.

Examples

R code

```
dens = dALD(y=seq(-40,80,0.5),mu=50,sigma=3,p=0.75)
sample = rALD(n=10000,mu=50,sigma=3,p=0.75)
hist(sample,breaks = 70,freq = FALSE,ylim=c(0,max(dens)),main="")
title(main="Histogram and True density")
lines(seq(-40,80,0.5),dens,col="red",lwd=2)
```

`qrLMM`: *Quantile Regression for Linear Mixed-Effects Models*

Quantile regression (QR) for Linear Mixed-Effects Models via the asymmetric Laplace distribution (ALD). It uses the Stochastic Approximation of the EM (SAEM) algorithm for deriving exact maximum likelihood estimates and full inference results for the fixed-effects and variance components. It also provides graphical summaries for assessing the algorithm convergence and fitting results. It is available to download for free in the website: *http://cran.r-project.org/web/packages/qrLMM/index.html*.

Description

The principal function of this package is the `QRLMM` function that performs a quantile regression for a LMEM using the Stochastic-Approximation of the EM (SAEM) algorithm for an unique or a set of quantiles. It can be run as:

`R` code

```
QRLMM(y,x,z,nj,p=0.5,precision=0.0001,MaxIter=300,M=10,cp=0.25,
beta=NA,sigma=NA,Psi=NA,show.convergence=TRUE,CI=95)
```

Arguments

`y`	the response vector of dimension N where N is the total of observations.
`x`	design matrix for the fixed effects of dimension $N \times d$ where d represents the number of fixed effects including the intercept, if considered.
`z`	design matrix for the random effects of dimension $N \times q$ where q represents the number of random effects.
`groups`	factor of dimension N specifying the partitions of the data over which the random effects vary.
`p`	unique quantile or a set of quantiles related to the quantile regression.
`precision`	the convergence maximum error.
`MaxIter`	the maximum number of iterations of the SAEM algorithm. Default = 300.
`M`	Number of Monte Carlo simulations used by the SAEM Algorithm. Default = 10. For more accuracy we suggest to use `M=20`.
`cp`	cut point ($0 \leq cp \leq 1$) which determines the percentage of initial iterations with no memory.

`beta`	fixed effects vector of initial parameters, if desired.
`sigma`	dispersion initial parameter for the error term, if desired.
`Psi`	Variance-covariance random effects matrix of initial parameters, if desired.
`show.convergence`	if `TRUE`, it will show a graphical summary for the convergence of the estimates of all parameters for each quantile in order to assess the convergence.
`CI`	Confidence to be used for the Confidence Interval when a grid of quantiles is provided. Default=95.

Details

If the initial parameters are not provided, by default, the fixed effects parameter β and dispersion parameter σ will be the MLE for an ALD (obviating the random term) as detailed in Appendix A.1.

When a grid of quantiles is provided, a graphical summary with point estimates and confidence intervals for model parameters is shown (e.g. see Figure 2.2) and also a graphical summary for the convergence of these estimates (e.g. see figure B.2) for each quantile, if `show.convergence = TRUE`. Also, the result will be a list of the same dimension where each element corresponds to each quantile as detailed above.

If the convergence graphical summary shows that convergence has not be attained, it's suggested to increase `M` to 20, to increase the total number of iterations `MaxIter` to 500 or both. This program uses progress bars that will close when the algorithm ends. They must not be closed before, if not the algorithm will stop.

Value

The function returns a list with two objects:

`conv` A two elements list with the matrices `teta` and `se` containing the point estimates and standard error estimate for all parameters along all iterations.

The second element of the list is `res`, a list of 12 elements detailed as:

`iter`	number of iterations.
`criteria`	attained criteria value.
`beta`	fixed effects estimates.
`sigma`	scale parameter estimate for the error term.
`Psi`	Random effects variance-covariance estimate matrix.
`SE`	Standard Error estimates.

`table`	Table containing the inference for the fixed effects parameters.
`loglik`	Log-likelihood value.
`AIC`	Akaike information criterion.
`BIC`	Bayesian information criterion.
`HQ`	Hannan-Quinn information criterion.
`time`	processing time.

Examples

`R` code

```
#Using the Orthodontic distance growth data

data(Orthodont)
attach(Orthodont)

y  = distance #response
x  = cbind(1,c(rep(0,64),rep(1,44)),age) #design matrix for fixed effects
z  = cbind(1,age) #design matrix for random effects
groups = Subject

QRLMM(y,x,z,groups,MaxIter=500) #a median regression
QRLMM(y,x,z,groups,p = c(0.25,0.50,0.75),MaxIter=300,M=10) #a quartile regression
```

qrNLMM: *Quantile Regression for Nonlinear Mixed-Effects Models*

Quantile regression (QR) for Nonlinear Mixed-Effects Models via the asymmetric Laplace distribution (ALD). It uses the Stochastic Approximation of the EM (SAEM) algorithm for deriving exact maximum likelihood estimates and full inference results for the fixed-effects and variance components. It also provides graphical summaries for assessing the algorithm convergence and fitting results. It is available to download for free in the website: *http://cran.r-project.org/web/packages/qrNLMM/index.html*.

Description

The principal function of this package is the `QRNLMM` function that performs a quantile regression for a NLMEM using the Stochastic-Approximation of the EM (SAEM) algorithm for an unique or a set of quantiles. It can be run as:

```
QRNLMM(y,x,groups,initial,exprNL,covar=NA,p=0.5,precision=0.0001,MaxIter=500,
M=20,cp=0.25,beta=NA,sigma=NA,Psi=NA,show.convergence=TRUE,CI=95)
```

Arguments

`y`	the response vector of dimension N where N is the total of observations.
`x`	vector of longitudinal (repeated measures) covariate of dimension N. For example: Time, location, etc.
`groups`	factor of dimension N specifying the partitions of the data over which the random effects vary.
`initial`	an numeric vector, or list of initial estimates for the fixed effects. It must be provide adequately (see details section) in order to ensure a proper convergence.
`exprNL`	expression containing the proposed nonlinear function. It can be of class `character` or `expression`. It must have a defined structure defined in the details section in order to be correctly read by the derivate `R` function `deriv`.
`covar`	a vector of dimension N containing a second covariate.
`p`	unique quantile or a set of quantiles related to the quantile regression.
`precision`	the convergence maximum error.
`MaxIter`	the maximum number of iterations of the SAEM algorithm. Default = 300.
`M`	Number of Monte Carlo simulations used by the SAEM Algorithm. Default = 10. For more accuracy we suggest to use `M=20`.
`cp`	cut point ($0 \leq cp \leq 1$) which determines the percentage of initial iterations with no memory.
`beta`	fixed effects vector of initial parameters, if desired.
`sigma`	dispersion initial parameter for the error term, if desired.
`Psi`	Variance-covariance random effects matrix of initial parameters, if desired.
`show.convergence`	if `TRUE`, it will show a graphical summary for the convergence of the estimates of all parameters for each quantile in order to assess the convergence.
`CI`	Confidence to be used for the Confidence Interval when a grid of quantiles is provided. Default=95.

Details

This algorithm performs the SAEM algorithm proposed by Delyon et al. (1999), deriving exact maximum likelihood estimates of the fixed-effects and variance components. For the moment just two covariates are allowed, the longitudinal (repeated measures) covariate `x` and a second covariate `covar`.

About initial values: Estimation for fixed effects parameters envolves a Newton-Raphson step. In adition, NL models are highly sensitive to initial values. So, we suggest to set of intial values quite good, this based in the parameter interpretation of the proposed NL function.

About the nonlinear expression: For the NL expression `exprNL` just the variables `x`, `covar`, `fixed` and `random` can be defined. Both `x` and `covar` represent the covariates defined above. The fixed effects must be declared as `fixed[1]`, `fixed[2]`,..., `fixed[d]` representing the first, second and dth fixed effect. Exactly the same for the random effects where the term `fixed` should be replace for `random`.

For instance, if we use the exponential nonlinear function with two parameters, each parameter represented by a fixed and a random effect, this will be defined by

$$y_{ij} = (\beta_1 + b_1) \exp^{-(\beta_2 + b_2)x_{ij}}$$

and the `exprNL` should be a character or and expression defined by

```
exprNL = "(fixed[1]+random[1])*exp(-(fixed[2]+random[2])*x)"
```

or

```
exprNL = expression((fixed[1]+random[1])*exp(-(fixed[2]+random[2])*x)).
```

If we are interested in adding a second covariate in order to explain one of the parameters, the covariate `covar` must be included in the model. For example, for the nonlinear function

$$y_{ij} = (\beta_1 + \beta_3 covar_{ij} + b_1) \exp^{-(\beta_2 + b_2)x_{ij}}$$

the `exprNL` should be

```
exprNL = "(fixed[1]+fixed[3]*covar+random[1])*exp(-(fixed[2]+random[2])*x)"
```

or

```
exprNL = expression((fixed[1]+fixed[3]*covar+random[1])*exp(-(fixed[2]+random[2])*x)).
```

Note that the mathematical function `exp` was used. For derivating the `deriv` R function recognizes in the `exprNL` expression the arithmetic operators `+`, `-`, `*`, `/` and `^`, and the single-variable functions `exp`, `log`, `sin`, `cos`, `tan`, `sinh`, `cosh`, `sqrt`, `pnorm`, `dnorm`, `asin`, `acos`, `atan`, `gamma`,

`lgamma`, `digamma` and `trigamma`, as well as `psigamma` for one or two arguments (but derivative only with respect to the first).

General details: When a grid of quantiles is provided, a graphical summary with point estimates and confidence intervals for model parameters is shown (e.g. see Figure 3.6) and also a graphical summary for the convergence of these estimates (e.g. see Figure B.1) for each quantile, if `show.convergence = TRUE`. Also, the result will be a list of the same dimension where each element corresponds to each quantile as detailed above.

Value

The function returns a list with two objects:

`conv` A two elements list with the matrices `teta` and `se` containing the point estimates and standard error estimate for all parameters along all iterations.

The second element of the list is `res`, a list of 12 elements detailed as:

`iter`	number of iterations.
`criteria`	attained criteria value.
`nlmodel`	the proposed nonlinear function.
`beta`	fixed effects estimates.
`sigma`	scale parameter estimate for the error term.
`Psi`	Random effects variance-covariance estimate matrix.
`SE`	Standard Error estimates.
`table`	Table containing the inference for the fixed effects parameters.
`loglik`	Log-likelihood value.
`AIC`	Akaike information criterion.
`BIC`	Bayesian information criterion.
`HQ`	Hannan-Quinn information criterion.
`time`	processing time.

Examples

`R` code

```
#Using the Soybean data
data(Soybean)
```

```
attach(Soybean)

#A full model (no covariate)
y = weight #response
x = Time #time

#Expression for the three parameter logistic curve
exprNL = expression((fixed[1]+random[1])/
(1 + exp(((fixed[2]+random[2])- x)/(fixed[3]+random[3]))))

#Initial values for fixed effects
initial = c(max(y),0.6*max(y),0.73*max(y))

#A median regression (by default)
median_reg = QRNLMM(y,x,Plot,initial,exprNL)

#Assing the fit
fxd = median_reg$res$beta
nlmodel = median_reg$res$nlmodel
seqc = seq(min(x),max(x),length.out = 500)
group.plot(x = Time,y = weight,groups = Plot,type="l",main="Soybean profiles",
xlab="time (days)",ylab="mean leaf weight (gr)",col="gray")
lines(seqc,nlmodel(x = seqc,fixed = fxd,random = rep(0,3)),lwd=2,col="blue")
```

4.2 Conclusions

In this work, we developed a likelihood-based inference for QR in mixed-effects models from a likelihood-based perspective based on the ALD. The stochastic representation of the ALD leads to a simple implementation of EM-type algorithms where a SAEM algorithm has been developed since it leads quickly to exact ML estimation of the parameters than traditional MCEM. The ALD and SAEM algorithm are combined efficiently to propose an exact ML estimation method outperforming the approximated method proposed by Geraci and Bottai (2014) for linear mixed-effects models in terms of standard errors and mean square error.

For the nonlinear model we evaluated the robustness of estimates, as well as the finite sample performance of the algorithm and the asymptotic properties of the ML estimates through empirical experiments. We found that the model is robust to misspecification of the random effects leading to quite good ML estimates. To the best of our knowledge, we consider that this work is the first attempt for exact ML estimation in the context of QR in mixed-effect models. We apply our method to a four data set from longitudinal studies, obtained interesting results from the point of view of quantile estimation as robust estimation, identification of possible outliers profiles and full conditional distribution estimation. The methods developed are readily implementable via the `R`

packages `qrLMM()` and `qrNLMM()`, making our approach quite powerful and accessible to practitioners.

4.3 Future research

There are a large number of possible extensions of the current work. For modelling both skewness and heavy tails in the random effects, the use of scale mixtures of skew-normal (SMSN) distributions (V. H. Lachos et al., 2010) is a feasible choice. It is also possible to proposed a semi parametric QR, including a semi-parametric structure for modeling the conditional quantiles extending the work in D. Zhang and M. Davidian (2001) or using a spatial covariance structure for spatially correlated data. We conjecture this QR model can be also transferred for describing multivariate responses in mixed-effects models. Other zero-quantile families of distributions can be used for modeling the error term as proposed in Wichitaksorn et al. (2014).

Experimental studies include covariates that often comes with substantial measurement errors (Lang Wu, 2002). How to incorporate measurement error in covariates within our robust framework can also be part of future research. An in-depth investigation of such extensions is beyond the scope of the present work, but certainly an interesting topic for future research.

Bibliography

Allassonnière, Stéphanie, Estelle Kuhn, Alain Trouvé, et al. (2010). "Construction of Bayesian deformable models via a stochastic approximation algorithm: a Convergence study". In: *Bernoulli* 16.3, pp. 641–678.

Barndorff-Nielsen, Ole E and Neil Shephard (2001). "Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.2, pp. 167–241.

Bates, Douglas M and Donald G Watts (1981). "A Relative Off set Orthogonality Convergence Criterion for Nonlinear least Squares". In: *Technometrics* 23.2, pp. 179–183.

Booth, J. G. and J. P. Hobert (1999). "Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61.1, pp. 265–285. ISSN: 1467-9868. DOI: 10.1111/1467-9868.00176. URL: http://dx.doi.org/10.1111/1467-9868.00176.

Davidian, Marie and David M Giltinan (1995). *Nonlinear Models for Repeated Measurement Data*. Vol. 62. CRC Press.

Delyon, Bernard, Marc Lavielle, and Eric Moulines (1999). "Convergence of a stochastic approximation version of the EM algorithm". In: *Annals of Statistics* 8, pp. 94–128.

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

Fu, Liya and You-Gan Wang (2012). "Quantile regression for longitudinal data with a working correlation model". In: *Computational Statistics & Data Analysis* 56.8, pp. 2526–2538.

Galvao Jr, Antonio F (2011). "Quantile regression for dynamic panel data with fixed effects". In: *Journal of Econometrics* 164.1, pp. 142–157.

Galvao, Antonio F and Gabriel V Montes-Rojas (2010). "Penalized quantile regression for dynamic panel data". In: *Journal of Statistical Planning and Inference* 140.11, pp. 3476–3497.

Geraci, Marco (2014). "Linear Quantile Mixed Models: The lqmm Package for Laplace Quantile Regression". In: *Journal of Statistical Software* 57.13, pp. 1–29. URL: http://www.jstatsoft.org/v57/i13/.

Geraci, Marco and Matteo Bottai (2007). "Quantile regression for longitudinal data using the asymmetric Laplace distribution". In: *Biostatistics* 8.1, pp. 140–154.

Geraci, Marco and Matteo Bottai (2014). "Linear quantile mixed models". In: *Statistics and Computing* 24.3, pp. 461–479.

Grossman, Zvi et al. (1999). "Ongoing HIV dissemination during HAART". In: *Nature medicine* 5.10, pp. 1099–1104.

Hastings, W Keith (1970). "Monte Carlo sampling methods using Markov chains and their applications". In: *Biometrika* 57.1, pp. 97–109.

Huang, Yangxin and Getachew Dagne (2011). "A Bayesian Approach to Joint Mixed-Effects Models with a Skew-Normal Distribution and Measurement Errors in Covariates". In: *Biometrics* 67.1, pp. 260–269.

Koenker, R. and J.A.F. Machado (1999). "Goodness of fit and related inference processes for quantile regression". In: *Journal of the american statistical association* 94.448, pp. 1296–1310.

Koenker, Roger (2004). "Quantile regression for longitudinal data". In: *Journal of Multivariate Analysis* 91.1, pp. 74–89.

Koenker, Roger (2005). *Quantile Regression*. New York, NY: Cambridge University Press.

Kotz, S., T.J. Kozubowski, and K. Podgorski (2001). *The Laplace distribution and generalizations: A revisit with applications to communications, economics, engineering, and finance*. Birkhauser.

Kuhn, Estelle and Marc Lavielle (2004). "Coupling a stochastic approximation version of EM with an MCMC procedure". In: *ESAIM: Probability and Statistics* 8, pp. 115–131.

Kuhn, Estelle and Marc Lavielle (2005). "Maximum likelihood estimation in nonlinear mixed effects models". In: *Computational Statistics & Data Analysis* 49.4, pp. 1020–1038.

Kuzobowski, T. J. and K. Podgorski (2000). "A Multivariate and Asymmetric Generalization of Laplace Distribution". In: *Computational Statistics* 15(4), pp. 531–540.

Lachos, V. H., P. Ghosh, and R. B. Arellano-Valle (2010). "Likelihood based Inference for Skew–Normal Independent Linear Mixed Models". In: *Statistica Sinica* 20.1, pp. 303–322.

Lachos, Victor H, Luis M Castro, and Dipak K Dey (2013). "Bayesian inference in nonlinear mixed-effects models using normal independent distributions". In: *Computational Statistics & Data Analysis* 64, pp. 237–252.

Lavielle, Marc (2014). *Mixed Effects Models for the Population Approach*. Boca Raton, FL: Chapman and Hall/CRC.

Lipsitz, Stuart R et al. (1997). "Quantile Regression Methods for Longitudinal Data with Drop-outs: Application to CD4 Cell Counts of Patients Infected with the Human Immunodeficiency Virus". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 46.4, pp. 463–476.

Louis, Thomas A (1982). "Finding the observed information matrix when using the EM algorithm". In: *Journal of the Royal Statistical Society - Series B (Methodological)* 44.2, pp. 226–233.

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

Metropolis, Nicholas et al. (1953). "Equation of State Calculations by Fast Computing Machines". In: *Journal of Chemical Physics* 21, pp. 1087–1092.

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

Perelson, Alan S et al. (1997). "Decay characteristics of HIV-1-infected compartments during combination therapy". In:

Pinheiro, J.C. and D.M. Bates (1995). "Approximations to the Log-likelihood Function in the Nonlinear Mixed Effects Model". In: *Journal of Computational and Graphical Statistics* 4, pp. 12–35.

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

Pinheiro, José C, Chuanhai Liu, and Ying Nian Wu (2001). "Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution". In: *Journal of Computational and Graphical Statistics* 10.2, pp. 249–276.

Potthoff, Richard F and SN Roy (1964). "A generalized multivariate analysis of variance model useful especially for growth curve problems". In: *Biometrika* 51.3-4, pp. 313–326.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: `http://www.R-project.org`.

Searle, Shayle R, G Casella, and CE McCulloch (1992). *Variance components, 1992*.

Vaida, Florin (2005). "Parameter convergence for EM and MM algorithms". In: *Statistica Sinica* 15.3, pp. 831–840.

Wang, J. (2012). "Bayesian quantile regression for parametric nonlinear mixed effects models". In: *Statistical Methods and Applications* 21, pp. 279–295.

Wei, Greg CG and Martin A Tanner (1990). "A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms". In: *Journal of the American Statistical Association* 85.411, pp. 699–704.

Wichitaksorn, Nuttanan, ST Choy, and Richard Gerlach (2014). "A generalized class of skew distributions and associated robust quantile regression models". In: *Canadian Journal of Statistics* 42.4, pp. 579–596.

Wu, CF Jeff (1983). "On the convergence properties of the EM algorithm". In: *The Annals of statistics*, pp. 95–103.

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

Wu, Lang (2002). "A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to AIDS studies". In: *Journal of the American Statistical association* 97.460, pp. 955–964.

Yu, K. and R.A. Moyeed (2001). "Bayesian quantile regression". In: *Statistics & Probability Letters* 54.4, pp. 437–447.

Yu, Keming and Jin Zhang (2005). "A three-parameter asymmetric Laplace distribution and its extension". In: *Communications in Statistics - Theory and Methods* 34.9-10, pp. 1867–1879.

Yuan, Ying and Guosheng Yin (2010). "Bayesian quantile regression for longitudinal studies with nonignorable missing data". In: *Biometrics* 66.1, pp. 105–114.

Zhang, D. and M. Davidian (2001). "Linear mixed models with flexible distributions of random effects for longitudinal data". In: *Biometrics* 57.3, pp. 795–802.

Appendix A

Supplementary Material for QR in Linear Mixed-Effect Models

A.1 Specification of initial values

It is well known that a smart choice of the initial values of ML estimates can assure a fast convergence of an algorithm to the global maxima solution for the respective likelihood. Obviating the random effects term, let $\mathbf{y}_i \sim ALD(\mathbf{x}_i^\top \boldsymbol{\beta}_p, \sigma, p)$. Next, considering the MLEs of $\boldsymbol{\beta}_p$ and σ as defined in Keming Yu and J. Zhang (2005) for this model, we follow the steps below for the QR-LMM implementation:

1. Compute an initial value $\widehat{\boldsymbol{\beta}}_p^{(0)}$ as

$$\widehat{\boldsymbol{\beta}}_p^{(0)} = \arg\min_{\boldsymbol{\beta}_p \in \mathbb{R}^k} \sum_{i=1}^{n} \rho_p(\mathbf{y}_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p).$$

2. Using the initial value for $\widehat{\boldsymbol{\beta}}_p^{(0)}$ obtained above, compute $\widehat{\sigma}^{(0)}$ as

$$\widehat{\sigma}^{(0)} = \frac{1}{n} \sum_{i=1}^{n} \rho_p(\mathbf{y}_i - \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}_p^{(0)}).$$

3. Use a $q \times q$ identity matrix $\mathbf{I}_{q\times q}$ for the the initial value $\boldsymbol{\Psi}^{(0)}$.

A.2 Computing the conditional expectations

Due the independence between $u_{ij}|\, y_{ij}, \mathbf{b}_i$ and $u_{ik}|\, y_{ik}, \mathbf{b}_i$, for all $j, k = 1, 2, \ldots, n_i$ and $j \neq k$, we can write $\mathbf{u}_i|\mathbf{y}_i, \mathbf{b}_i = [\ u_{i1}|y_{i1}, \mathbf{b}_i \quad u_{i2}|y_{i2}, \mathbf{b}_i \quad \cdots \quad u_{in_i}|y_{in_i}, \mathbf{b}_i\ ]^\top$. Using this fact, we are able to compute the conditional expectations $\mathcal{E}(\mathbf{u}_i)$ and $\mathcal{E}(\mathbf{D}_i^{-1})$ in the following way. Using matrix

expectation properties, we define these expectations as

$$\mathcal{E}(\mathbf{u}_i) = [\mathcal{E}(u_{i1})\ \mathcal{E}(u_{i1})\ \cdots\ \mathcal{E}(u_{in_i})]^\top \tag{A.2.1}$$

and

$$\mathcal{E}(\mathbf{D}_i^{-1}) = \text{diag}(\mathcal{E}(\mathbf{u}_i^{-1})) = \begin{bmatrix} \mathcal{E}(u_{i1}^{-1}) & 0 & ... & 0 \\ 0 & \mathcal{E}(u_{i2}^{-1}) & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & \mathcal{E}(u_{in_i}^{-1}) \end{bmatrix}. \tag{A.2.2}$$

We already have $u_{ij}|y_{ij}, \mathbf{b}_i \sim GIG(\frac{1}{2}, \chi_{ij}, \psi)$, where χ_{ij} and ψ are defined in (14). Then, using (5), we compute the moments involved in the equations above as $\mathcal{E}(u_{ij}) = \frac{\chi_{ij}}{\psi}(1 + \frac{1}{\chi_{ij}\psi})$ and $\mathcal{E}(u_{ij}^{-1}) = \frac{\psi}{\chi_{ij}}$. Thus, for iteration k of the algorithm and for the ℓth Monte Carlo realization, we can compute $\mathcal{E}(\mathbf{u}_i)^{(\ell,k)}$ and $\mathcal{E}(\mathbf{D}_i^{-1})^{(\ell,k)}$ using equations (A.2.1)-(A.2.2) where

$$\mathcal{E}(u_{ij})^{(\ell,k)} = \frac{2|y_{ij} - \mathbf{x}_{ij}^\top \boldsymbol{\beta}_p^{(k)} - \mathbf{z}_{ij}^\top \mathbf{b}_i^{(\ell,k)}| + 4\sigma^{(k)}}{\tau_p^2} \quad \text{and} \quad \mathcal{E}(u_{ij}^{-1})^{(\ell,k)} = \frac{\tau_p^2}{2|y_{ij} - \mathbf{x}_{ij}^\top \boldsymbol{\beta}_p^{(k)} - \mathbf{z}_{ij}^\top \mathbf{b}_i^{(\ell,k)}|}.$$

A.3 The empirical information matrix

In light of (10), the complete log-likelihood function can be rewritten as

$$\ell_{ci}(\boldsymbol{\theta}) = -\frac{3}{2} n_i \log\sigma - \frac{1}{2\sigma\tau_p^2}\zeta_i^\top \mathbf{D}_i^{-1}\zeta_i - \frac{1}{2}\log|\boldsymbol{\Psi}| - \frac{1}{2}\mathbf{b}_i^\top \boldsymbol{\Psi}^{-1}\mathbf{b}_i - \frac{1}{\sigma}\mathbf{u}_i^\top \mathbf{1}_{n_i} \tag{A.3.1}$$

where $\zeta_i = \mathbf{y}_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p - \mathbf{z}_i\mathbf{b}_i - \vartheta_p \mathbf{u}_i$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}_p^\top, \sigma, \boldsymbol{\alpha}^\top)^\top$. Taking partial derivatives with respect to $\boldsymbol{\theta}$, we have the following score functions:

$$\frac{\partial \ell_{ci}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_p} = \frac{\partial \zeta_i}{\partial \boldsymbol{\beta}_p}\frac{\partial \ell_{ci}(\boldsymbol{\theta})}{\partial \zeta_i} = \frac{1}{\sigma\tau_p^2}\mathbf{x}_i \mathbf{D}_i^{-1}\zeta_i,$$

and

$$\frac{\partial \ell_{ci}(\boldsymbol{\theta})}{\partial \sigma} = -\frac{3n_i}{2}\frac{1}{\sigma} + \frac{1}{2\sigma^2\tau_p^2}\zeta_i^\top \mathbf{D}_i^{-1}\zeta_i + \frac{1}{\sigma^2}\mathbf{u}_i^\top \mathbf{1}_{n_i}.$$

Let $\boldsymbol{\alpha}$ be the vector of reduced parameters from $\boldsymbol{\Psi}$, the dispersion matrix for $\mathbf{b}_i$. Using the trace properties and differentiating the complete log-likelihood function, we have that

$$\begin{aligned} \frac{\partial \ell_{ci}(\boldsymbol{\theta})}{\partial \boldsymbol{\Psi}} &= \frac{\partial}{\partial \boldsymbol{\Psi}}\left[-\frac{n}{2} log|\boldsymbol{\Psi}| - \frac{1}{2}\text{tr}\{\boldsymbol{\Psi}^{-1}\mathbf{b}_i\mathbf{b}_i^\top\}\right] \\ &= -\frac{1}{2}\text{tr}\{\boldsymbol{\Psi}^{-1}\} + \frac{1}{2}\text{tr}\{\boldsymbol{\Psi}^{-1}\boldsymbol{\Psi}^{-1}\mathbf{b}_i\mathbf{b}_i^\top\} \\ &= \frac{1}{2}\text{tr}\{\boldsymbol{\Psi}^{-1}(\mathbf{b}_i\mathbf{b}_i^\top - \boldsymbol{\Psi})\boldsymbol{\Psi}^{-1}\} \end{aligned}$$

Next, taking derivatives with respect to a specific α_j from $\boldsymbol{\alpha}$ based on the chain rule, we have

$$\begin{aligned}\frac{\partial \ell_{ci}(\boldsymbol{\theta})}{\partial \alpha_j} &= \frac{\partial \boldsymbol{\Psi}}{\partial \alpha_j}\frac{\partial \ell_{ci}(\boldsymbol{\theta})}{\partial \boldsymbol{\Psi}} \\ &= \frac{\partial \boldsymbol{\Psi}}{\partial \alpha_j}\frac{1}{2}\operatorname{tr}\{\boldsymbol{\Psi}^{-1}(\mathbf{b}_i\mathbf{b}_i^\top - \boldsymbol{\Psi})\boldsymbol{\Psi}^{-1}\}. \end{aligned} \tag{A.3.2}$$

where, using the fact that $\operatorname{tr}\{\mathbf{ABCD}\} = (\operatorname{vec}(\mathbf{A}^\top))^\top(\mathbf{D}^\top \otimes \mathbf{B})(\operatorname{vec}(\mathbf{C}))$, (A.3.2) can be rewritten as

$$\frac{\partial \ell_{ci}(\boldsymbol{\theta})}{\partial \alpha_j} = (\operatorname{vec}(\tfrac{\partial \boldsymbol{\Psi}}{\partial \alpha_j}^\top))^\top \frac{1}{2}(\boldsymbol{\Psi}^{-1} \otimes \boldsymbol{\Psi}^{-1})(\operatorname{vec}(\mathbf{b}_i\mathbf{b}_i^\top - \boldsymbol{\Psi})). \tag{A.3.3}$$

Let $\mathcal{D}_q$ be the elimination matrix (Lavielle, 2014) that transforms the vectorized $\boldsymbol{\Psi}$ (written as $\operatorname{vec}(\boldsymbol{\Psi})$) into its half-vectorized form $\operatorname{vech}(\boldsymbol{\Psi})$, such that $\mathcal{D}_q\operatorname{vec}(\boldsymbol{\Psi}) = \operatorname{vech}(\boldsymbol{\Psi})$. Using the fact that for all $j = 1, \ldots, \frac{1}{2}q(q+1)$, the vector $(\operatorname{vec}(\frac{\partial \boldsymbol{\Psi}}{\partial \alpha_j})^\top)^\top$ corresponds to the jth row of the elimination matrix $\mathcal{D}_q$, we can generalize the derivative in (A.3.3) for the vector of parameters $\boldsymbol{\alpha}$ as

$$\frac{\partial \ell_{ci}(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}} = \frac{1}{2}\mathcal{D}_q(\boldsymbol{\Psi}^{-1} \otimes \boldsymbol{\Psi}^{-1})(\operatorname{vec}(\mathbf{b}_i\mathbf{b}_i^\top - \boldsymbol{\Psi})).$$

Finally, at each iteration, we can compute the empirical information matrix (19) by approximating the score for the observed log-likelihood using the stochastic approximation given in (20).

A.4 A Gibbs Sampler Algorithm

In order to draw a sample from $f(\mathbf{b}_i, \mathbf{u}_i|\mathbf{y}_i)$ we can use a the Gibbs Sampler, an Markov chain Monte Carlo (MCMC) algorithm proposed by (Casella and George, 1992) for obtaining a sequence of observations which are approximated from the joint probability distribution of two or several random variables just using their full conditional distributions. Computing the full conditional distributions $f(\mathbf{b}_i|\mathbf{u}_i, \mathbf{y}_i)$ and $f(\mathbf{u}_i|\mathbf{b}_i, \mathbf{y}_i)$, we have for the first one that

$$\begin{aligned} f(\mathbf{b}_i|\mathbf{y}_i, \mathbf{u}_i) &\propto f(\mathbf{y}_i|\mathbf{b}_i, \mathbf{u}_i)\, f(\mathbf{b}_i), \\ &\propto \phi_{n_i}\left(\mathbf{y}_i|\, \mathbf{X}_i^\top\boldsymbol{\beta}_p + \mathbf{Z}_i\mathbf{b}_i + \vartheta_p\mathbf{u}_i, \sigma\tau_p^2 D(\mathbf{u}_i)\right) \times \phi_q(\mathbf{b}_i|\mathbf{0}, \boldsymbol{\Psi}) \end{aligned} \tag{A.4.1}$$

so we have a product of multivariate normal densities which solution is based in the next lemma:

Lemma 1. *Simplifying the notation above it follows that*

$$\phi_n(\boldsymbol{y}|\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b}, \boldsymbol{\Omega})\phi_q(\boldsymbol{b}|\mathbf{0}, \boldsymbol{\Psi}) = \phi_n(\boldsymbol{y}|\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})\phi_q(\boldsymbol{b}|\boldsymbol{\mu}_1(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \boldsymbol{\Lambda}) \tag{A.4.2}$$

where

$$\boldsymbol{\mu}_1 = \boldsymbol{\Lambda}\mathbf{Z}^T\boldsymbol{\Omega}^{-1}, \quad \boldsymbol{\Sigma} = \boldsymbol{\Omega} + \mathbf{Z}\boldsymbol{\Psi}\mathbf{Z}^T, \quad \boldsymbol{\Lambda} = (\boldsymbol{\Psi}^{-1} + \mathbf{Z}^T\boldsymbol{\Omega}^{-1}\mathbf{Z})^{-1}. \tag{A.4.3}$$

Due the equation (A.4.2) from the lemma 2 it leads us to

$$f(\mathbf{b}_i|\mathbf{y}_i,\mathbf{u}_i) \propto \phi_{n_i}\left(\mathbf{y}_i|\ \mathbf{X}_i^\top\boldsymbol{\beta}_p+\vartheta_p\mathbf{u}_i,\sigma\tau_p^2D(\mathbf{u}_i)+\mathbf{Z}_i\Psi\mathbf{Z}_i^\top\right)\times \phi_q\left(\mathbf{b}_i|\ \Lambda_iZ_i^\top\left(\sigma\tau_p^2D(\mathbf{u}_i)\right)^{-1}\left(\mathbf{y}_i-\mathbf{X}_i^\top\boldsymbol{\beta}_p-\vartheta_p\mathbf{u}_i\right),\Lambda_i\right)$$

where $\Lambda_i=\left(\Psi^{-1}+\sigma\tau_p^2\mathbf{Z}_i^\top D(\mathbf{u}_i)\mathbf{Z}_i\right)^{-1}$. Then dropping the first term of the product by proportionality it's easy to see that $\mathbf{b}_i|\mathbf{y}_i,\mathbf{u}_i\sim N_q\left(\Lambda_i\mathbf{Z}_i^\top\left(\sigma\tau_p^2D(\mathbf{u}_i)\right)^{-1}\left(\mathbf{y}_i-\mathbf{X}_i^\top\boldsymbol{\beta}_p-\vartheta_p\mathbf{u}_i\right),\Lambda_i\right)$.

On other hand, for the full conditional distribution $f(\mathbf{u}_i|\mathbf{y}_i,\mathbf{b}_i)$ note that the vector $\mathbf{u}_i|\mathbf{y}_i,\mathbf{b}_i$ can be constructed as $\mathbf{u}_i|\mathbf{y}_i,\mathbf{b}_i=\left[\ u_{i1}|y_{i1},\mathbf{b}_i\quad u_{i2}|y_{i2},\mathbf{b}_i\quad\cdots\quad u_{in_i}|y_{in_i},\mathbf{b}_i\ \right]^\top$ given that $u_{ij}|\ y_{ij},\mathbf{b}_i\perp u_{ik}|\ y_{ik},\mathbf{b}_i$ for all $j,k=1,2,\ldots,n_i$ and $j\neq k$. So, the univariate distribution of the $f(u_{ij}|y_{ij},\mathbf{b}_i)$ is proportional to the product of $f(y_{ij}|\mathbf{b}_i,u_{ij})$ and $f(u_{ij})$, a Normal and a Exponential distribution, that is

$$f(u_{ij}|y_{ij},\mathbf{b}_i)\propto\phi(y_{ij}|\mathbf{X}_{ij}^\top\boldsymbol{\beta}_p+\mathbf{Z}_{ij}^\top\mathbf{b}_i+\vartheta_pu_{ij},\ \sigma\tau_p^2u_{ij})\times G_{U_{ij}}(1,\sigma),$$

then the Lemma 1 leads us that $u_{ij}|y_{ij},\mathbf{b}_i\sim GIG(\frac{1}{2},\chi_{ij},\psi)$, where $\chi_{ij}=\frac{|y_{ij}-\mathbf{X}_{ij}^\top\boldsymbol{\beta}_p-\mathbf{Z}_{ij}^\top\mathbf{b}_i|}{\tau_p\sqrt{\sigma}}$ and $\psi=\frac{\tau_p}{2\sqrt{\sigma}}$.

In resume, the Gibbs Sampler proceeds as follow:

Given $\boldsymbol{\theta}=\boldsymbol{\theta}^{(k)}$ for $i=1,\ldots,n$;

(1) Start with suitable initial values $(\mathbf{b}_i^{(0,k)},\mathbf{u}_i^{(0,k)})$

(2) Draw $\mathbf{b}_i^{(1,k)}|\mathbf{y}_i,\mathbf{u}_i^{(0,k)}\sim N_q\left(\Lambda_i^{(k)}\mathbf{Z}_i^\top\left(\sigma^{(k)}\tau_p^2D(\mathbf{u}_i^{(0,k)})\right)^{-1}\left(\mathbf{y}_i-\mathbf{X}_i^\top\boldsymbol{\beta}_p^{(k)}-\vartheta_p\mathbf{u}_i^{(0,k)}\right),\Lambda_i^{(k)}\right)$

(3) Draw $u_{ij}^{(1,k)}|y_{ij},\mathbf{b}_i^{(1,k)}\sim GIG\left(\frac{1}{2},\frac{\left|y_{ij}-\mathbf{X}_{ij}^\top\boldsymbol{\beta}_p^{(k)}-\mathbf{Z}_{ij}^\top\mathbf{b}_i^{(1,k)}\right|}{\tau_p\sqrt{\sigma^{(k)}}},\frac{\tau_p}{2\sqrt{\sigma^{(k)}}}\right)$ for all $j=1,2,\ldots,n_i$

(4) Construct $\mathbf{u}_i^{(1,k)}|\mathbf{y}_i,\mathbf{b}_i^{(1,k)}$ as $\left[\ u_{i1}^{(1,k)}|y_{i1},\mathbf{b}_i^{(1,k)}\quad u_{i2}^{(1,k)}|y_{i2},\mathbf{b}_i^{(1,k)}\quad\cdots\quad u_{in_i}^{(1,k)}|y_{in_i},\mathbf{b}_i^{(1,k)}\ \right]^\top$

(5) Repeat the steps 2-4 until draw m samples $\left(\mathbf{b}_i^{(1,k)},\mathbf{u}_i^{(1,k)}\right),\left(\mathbf{b}_i^{(2,k)},\mathbf{u}_i^{(2,k)}\right),\ldots,\left(\mathbf{b}_i^{(m,k)},\mathbf{u}_i^{(m,k)}\right)$ from $\mathbf{b}_i,\mathbf{u}_i|\boldsymbol{\theta}^{(k)},\mathbf{y}_i$.

Note that for a given a iteration k and for all $i=1,\ldots,n$, drawing from the conditional distribution of the vector $\mathbf{u}_i^{(l,k)}|\mathbf{y}_i,\mathbf{b}_i^{(l,k)}$implies to draw from the univariate conditional distributions $u_{ij}^{(k)}|y_{ij},\mathbf{b}_i^{(k)}$ for all $j=1,2,\ldots,n_i$, so this construction results in a heavy computational algorithm. Also, given that the Gibbs Sampler is a Markov Chain we must apply a burn in and a thin, in order to discard the first observations before convergence and to avoid the correlation between nearby observations.

A.5 Figures

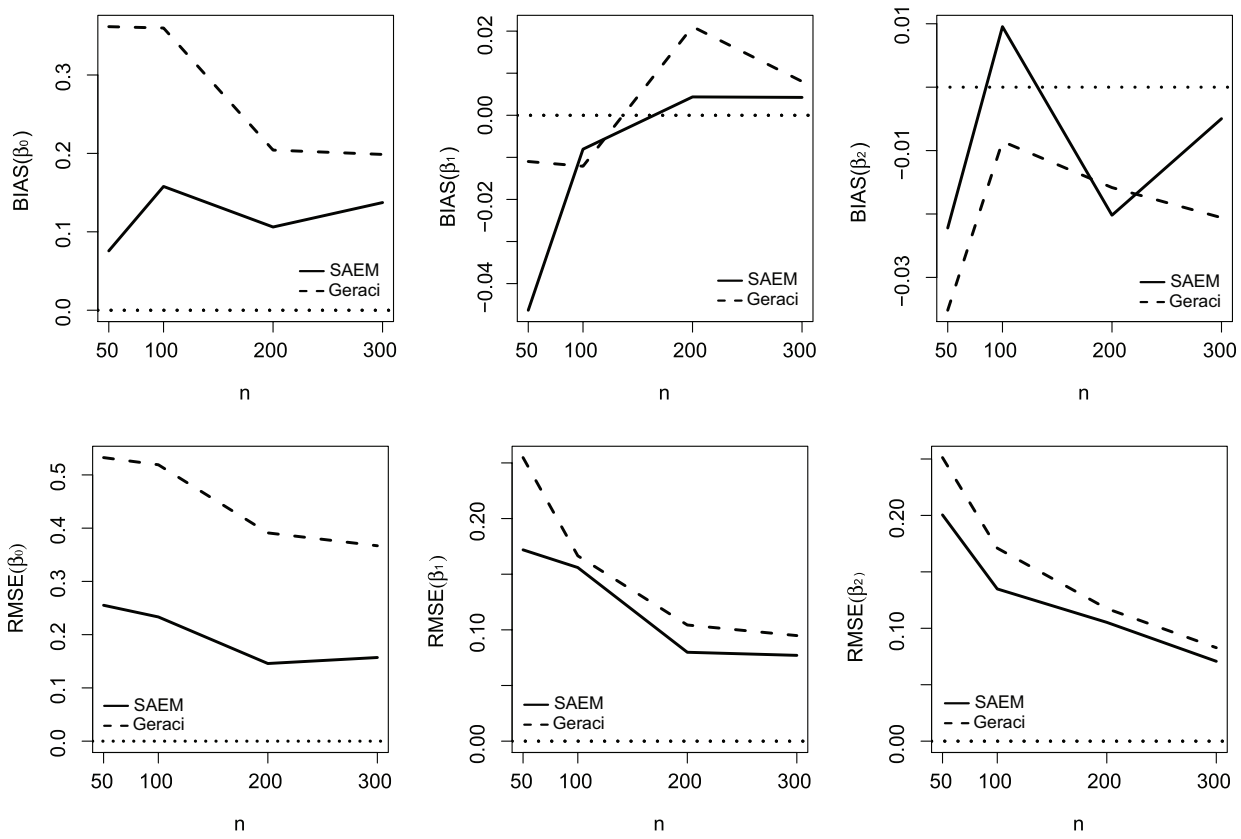


Figure A.1: Comparison of the Bias (upper row) and RMSE (lower row) at the 95-th quantile from fitting the QR-LMM and the Geraci, 2014 model for the fixed effects β_0, β_1 and β_2.

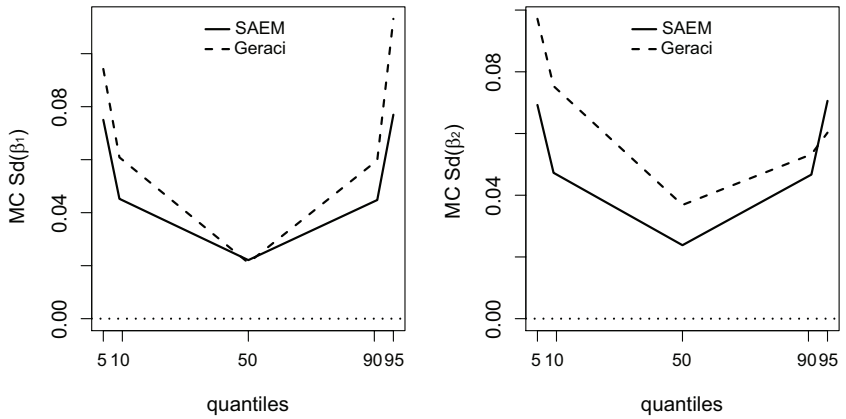


Figure A.2: Comparison of the Monte Carlo standard deviation for the estimatives of β_1 and β_2 obtained by the SAEM procedure and the Geraci (2014) algorithm for the set of quantiles 5, 10, 50, 90 and 95.

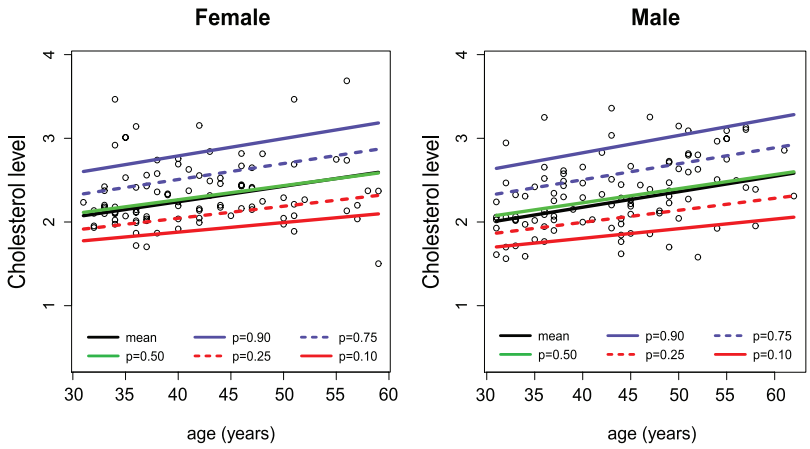


Figure A.3: Fitted mean regression overlayed with five different quantile regression lines for the Cholesterol data, by gender.

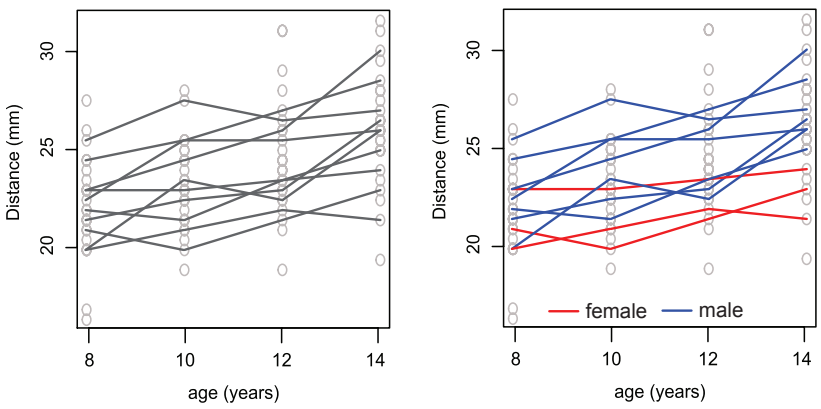


Figure A.4: Orthodontic distance growth data: Individual profiles for 10 random children (Panel a); Individual profiles for the same children, by gender (Panel b).

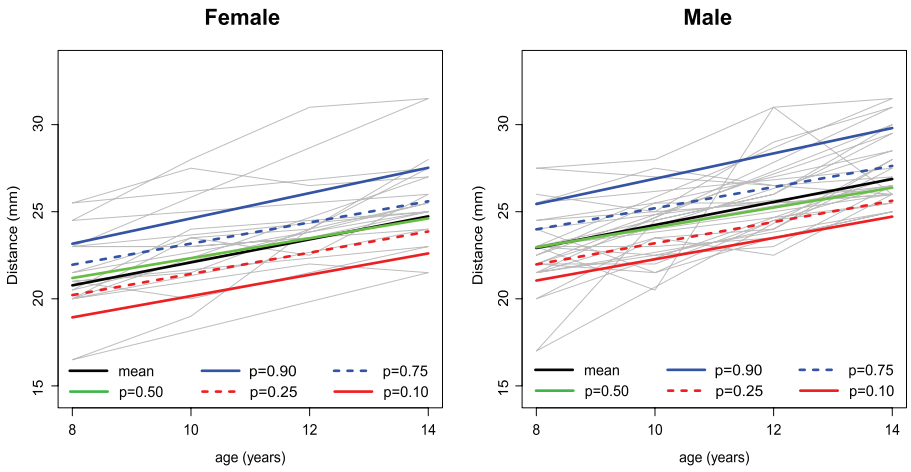


Figure A.5: Fitted mean regression overlayed with five different quantile regression lines for the Orthodontic distance growth data, by gender.

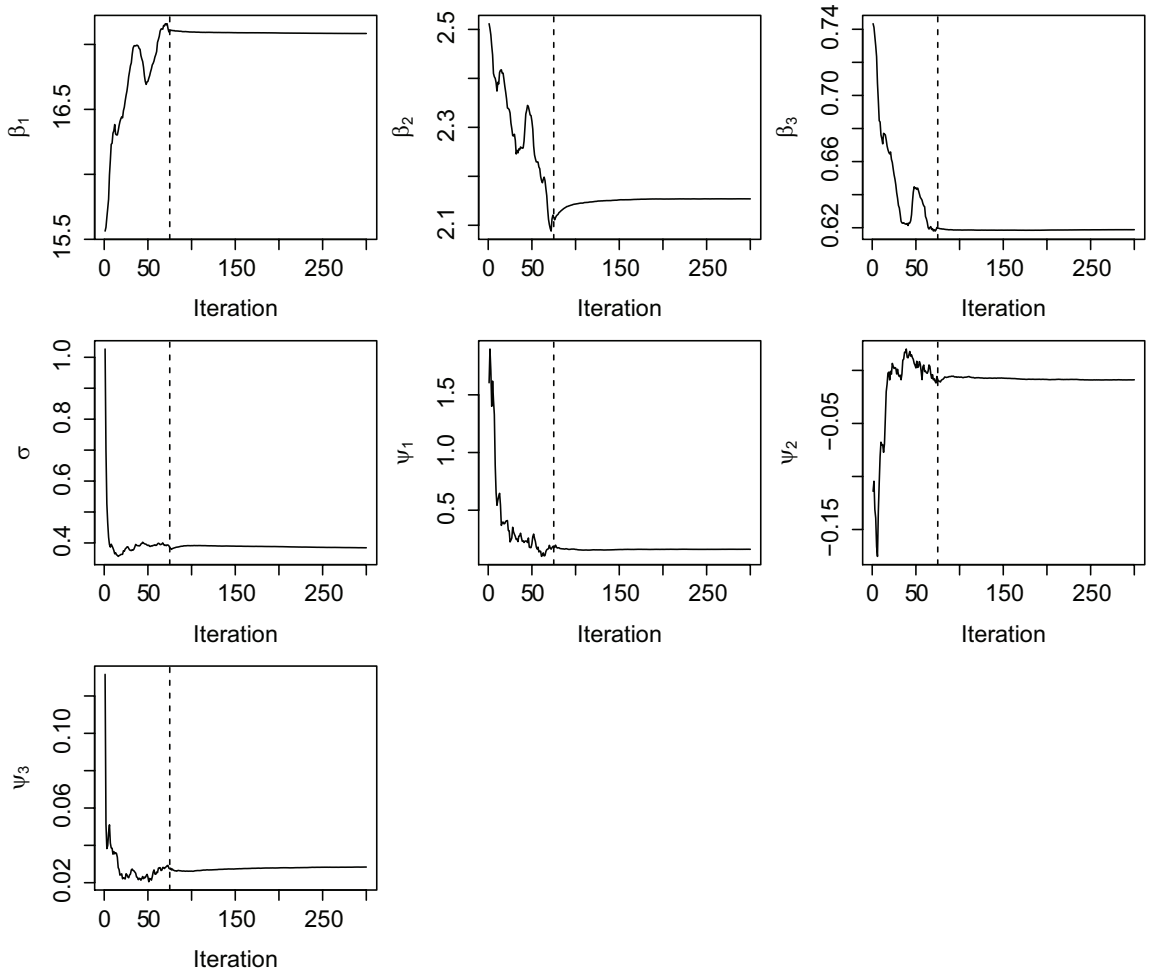


Figure A.6: Graphical summary of convergence for the fixed effect parameters, variance components of the random effects, and nuisance parameters, generated from the `qrLMM` package for the orthodontic distance growth data. The vertical dashed line delimits the beginning of the almost sure convergence, as defined by the cut-point parameter c.

A.6 Sample output from R package qrLMM()

```
------------------------------------------------
Quantile Regression for Linear Mixed Models
------------------------------------------------
Quantile = 0.75
Subjects = 27 ; Observations = 108 ; Balanced = 4

-----------
Estimates
-----------
- Fixed effects

Estimate Std. Error  z value Pr(>|z|)
beta 1 17.08405    0.53524 31.91831        0
beta 2  2.15393    0.36929  5.83265        0
beta 3  0.61882    0.05807 10.65643        0

sigma = 0.38439

Random effects Variance-covariance matrix
      b1       b2
b1  0.16106 -0.00887
b2 -0.00887  0.02839

------------------------
Model selection criteria
------------------------
Loglik     AIC     BIC     HQ
Value -216.454 446.907 465.682 454.52

-------
Details
-------
Convergence reached? = FALSE
Iterations = 300 / 300
Criteria = 0.00381
MC sample = 10
Cut point = 0.25
Processing time = 7.590584 mins
```

Appendix B

Supplementary Material for QR in Nonlinear Mixed-Effect Models

B.1 Specification of initial values

It is well known that a smart choice of the initial values for the ML estimates can assure a fast convergence of an algorithm to the global maxima solution. Obviating the random effects term, i.e., $\mathbf{b}_i = \mathbf{0}$, let $\mathbf{y}_i \sim ALD(\boldsymbol{\eta}(\boldsymbol{\beta}_p, \mathbf{0}), \sigma, p)$. Next, considering the MLEs for $\boldsymbol{\beta}_p$ and σ as defined in Keming Yu and J. Zhang (2005) for this model, we follow the steps below for the QR-LMM implementation:

1. Compute an initial value $\widehat{\boldsymbol{\beta}}_p^{(0)}$ as

$$\widehat{\boldsymbol{\beta}}_p^{(0)} = \arg\min_{\boldsymbol{\beta}_p \in \mathbb{R}^k} \sum_{i=1}^{n} \rho_p(\mathbf{y}_i - \boldsymbol{\eta}(\boldsymbol{\beta}_p, \mathbf{0})).$$

2. Using the initial value for $\widehat{\boldsymbol{\beta}}_p^{(0)}$ obtained above, compute $\widehat{\sigma}^{(0)}$ as

$$\widehat{\sigma}^{(0)} = \frac{1}{n} \sum_{i=1}^{n} \rho_p(\mathbf{y}_i - \boldsymbol{\eta}(\boldsymbol{\beta}_p, \mathbf{0})).$$

3. Use a $q \times q$ identity matrix $\mathbf{I}_{q\times q}$ for the the initial value $\boldsymbol{\Psi}^{(0)}$.

B.2 Computing the conditional expectations

Due the independence between $u_{ij}|\, y_{ij}, \mathbf{b}_i$ and $u_{ik}|\, y_{ik}, \mathbf{b}_i$, for all $j, k = 1, 2, \ldots, n_i$ and $j \neq k$, we can write $\mathbf{u}_i|\mathbf{y}_i, \mathbf{b}_i = [\ u_{i1}|y_{i1}, \mathbf{b}_i \quad u_{i2}|y_{i2}, \mathbf{b}_i \quad \cdots \quad u_{in_i}|y_{in_i}, \mathbf{b}_i\ ]^\top$. Using this fact, we are able to compute the conditional expectations $\mathcal{E}(\mathbf{u}_i)$ and $\mathcal{E}(\mathbf{D}_i^{-1})$ in the following way. Using matrix

expectation properties, we define these expectations as

$$\mathcal{E}(\mathbf{u}_i) = [\mathcal{E}(u_{i1})\ \mathcal{E}(u_{i1})\ \cdots\ \mathcal{E}(u_{in_i})]^\top \tag{B.2.1}$$

and

$$\mathcal{E}(\mathbf{D}_i^{-1}) = \text{diag}(\mathcal{E}(\mathbf{u}_i^{-1})) = \begin{bmatrix} \mathcal{E}(u_{i1}^{-1}) & 0 & \ldots & 0 \\ 0 & \mathcal{E}(u_{i2}^{-1}) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \mathcal{E}(u_{in_i}^{-1}) \end{bmatrix}. \tag{B.2.2}$$

We already have $u_{ij}|y_{ij}, \mathbf{b}_i \sim GIG(\frac{1}{2}, \chi_{ij}, \psi)$ where χ_{ij} and ψ are defined in (3.2.9). Then, using (1.1.5), we compute the moments involved in the equations above as $\mathcal{E}(u_{ij}) = \frac{\chi_{ij}}{\psi}(1 + \frac{1}{\chi_{ij}\psi})$ and $\mathcal{E}(u_{ij}^{-1}) = \frac{\psi}{\chi_{ij}}$. Thus, for iteration k of the algorithm and for the ℓth Monte Carlo realization, we can compute $\mathcal{E}(\mathbf{u}_i)^{(\ell,k)}$ and $\mathcal{E}(\mathbf{D}_i^{-1})^{(\ell,k)}$ using equations (B.2.1)-(B.2.2) where

$$\mathcal{E}(u_{ij})^{(\ell,k)} = \frac{2|y_{ij} - \eta_{ij}(\boldsymbol{\beta}_p^{(k)}, \mathbf{b}_i^{(\ell,k)})| + 4\sigma^{(k)}}{\tau_p^2} \qquad \text{and} \quad \mathcal{E}(u_{ij}^{-1})^{(\ell,k)} = \frac{\tau_p^2}{2|y_{ij} - \eta_{ij}(\boldsymbol{\beta}_p^{(k)}, \mathbf{b}_i^{(\ell,k)})|}.$$

B.3 The empirical information matrix

In light of (3.2.5), the complete log-likelihood function can be rewritten as

$$\ell_{ci}(\boldsymbol{\theta}) \ = \ -\frac{3}{2} n_i \log\sigma - \frac{1}{2\sigma\tau_p^2}\zeta_i^\top \mathbf{D}_i^{-1}\zeta_i - \frac{1}{2}\log|\boldsymbol{\Psi}| - \frac{1}{2}\mathbf{b}_i^\top \boldsymbol{\Psi}^{-1}\mathbf{b}_i - \frac{1}{\sigma}\mathbf{u}_i^\top \mathbf{1}_{n_i} \tag{B.3.1}$$

where $\zeta_i = \mathbf{y}_i - \boldsymbol{\eta}(\boldsymbol{\beta}_p, \mathbf{b}_i) - \vartheta_p \mathbf{u}_i$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}_p^\top, \sigma, \boldsymbol{\alpha}^\top)^\top$. Differentiating with respect to $\boldsymbol{\theta}$, we have the following score functions:

$$\frac{\partial \ell_{ci}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_p} \ = \ \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\beta}_p}\frac{\partial \zeta_i}{\partial \boldsymbol{\eta}}\frac{\partial \ell_{ci}(\boldsymbol{\theta})}{\partial \zeta_i} = \frac{1}{\sigma\tau_p^2}\mathbf{J}_\mathbf{i}^\top \mathbf{D}_i^{-1}\zeta_i,$$

with $\mathbf{J_i}$ defined in subsection 3.2. and

$$\frac{\partial \ell_{ci}(\boldsymbol{\theta})}{\partial \sigma} \ = \ -\frac{3n_i}{2}\frac{1}{\sigma} + \frac{1}{2\sigma^2\tau_p^2}\zeta_i^\top \mathbf{D}_i^{-1}\zeta_i + \frac{1}{\sigma^2}\mathbf{u}_i^\top \mathbf{1}_{n_i}.$$

Let $\boldsymbol{\alpha}$ be the vector of reduced parameters from $\boldsymbol{\Psi}$, the dispersion matrix for $\mathbf{b}_i$. Using the trace properties and differentiating the complete log-likelihood function, we have that

$$\begin{aligned} \frac{\partial \ell_{ci}(\boldsymbol{\theta})}{\partial \boldsymbol{\Psi}} \ &= \ \frac{\partial}{\partial \boldsymbol{\Psi}}\left[-\frac{n}{2} log|\boldsymbol{\Psi}| - \frac{1}{2}\text{tr}\{\boldsymbol{\Psi}^{-1}\mathbf{b}_i\mathbf{b}_i^\top\}\right] \\ &= \ -\frac{1}{2}\text{tr}\{\boldsymbol{\Psi}^{-1}\} + \frac{1}{2}\text{tr}\{\boldsymbol{\Psi}^{-1}\boldsymbol{\Psi}^{-1}\mathbf{b}_i\mathbf{b}_i^\top\} \\ &= \ \frac{1}{2}\text{tr}\{\boldsymbol{\Psi}^{-1}(\mathbf{b}_i\mathbf{b}_i^\top - \boldsymbol{\Psi})\boldsymbol{\Psi}^{-1}\} \end{aligned}$$

Next, taking derivatives with respect to a specific α_j from $\boldsymbol{\alpha}$ based on the chain rule, we have

$$\begin{aligned}\frac{\partial \ell_{ci}(\boldsymbol{\theta})}{\partial \alpha_j} &= \frac{\partial \boldsymbol{\Psi}}{\partial \alpha_j}\frac{\partial \ell_{ci}(\boldsymbol{\theta})}{\partial \boldsymbol{\Psi}} \\ &= \frac{\partial \boldsymbol{\Psi}}{\partial \alpha_j}\frac{1}{2}\operatorname{tr}\{\boldsymbol{\Psi}^{-1}(\mathbf{b}_i\mathbf{b}_i^\top - \boldsymbol{\Psi})\boldsymbol{\Psi}^{-1}\}. \end{aligned} \tag{B.3.2}$$

where, using the fact that $\operatorname{tr}\{\mathbf{ABCD}\} = (\operatorname{vec}(\mathbf{A}^\top))^\top(\mathbf{D}^\top \otimes \mathbf{B})(\operatorname{vec}(\mathbf{C}))$, (B.3.2) can be rewritten as

$$\frac{\partial \ell_{ci}(\boldsymbol{\theta})}{\partial \alpha_j} = (\operatorname{vec}(\tfrac{\partial \boldsymbol{\Psi}}{\partial \alpha_j}^\top))^\top \frac{1}{2}(\boldsymbol{\Psi}^{-1} \otimes \boldsymbol{\Psi}^{-1})(\operatorname{vec}(\mathbf{b}_i\mathbf{b}_i^\top - \boldsymbol{\Psi})). \tag{B.3.3}$$

Let $\mathcal{D}_q$ be the elimination matrix (Lavielle, 2014) that transforms the vectorized $\boldsymbol{\Psi}$ (written as $\operatorname{vec}(\boldsymbol{\Psi})$) into its half-vectorized form $\operatorname{vech}(\boldsymbol{\Psi})$, such that $\mathcal{D}_q\operatorname{vec}(\boldsymbol{\Psi}) = \operatorname{vech}(\boldsymbol{\Psi})$. Using the fact that for all $j = 1, \ldots, \frac{1}{2}q(q+1)$, the vector $(\operatorname{vec}(\frac{\partial \boldsymbol{\Psi}}{\partial \alpha_j})^\top)^\top$ corresponds to the jth row of the elimination matrix $\mathcal{D}_q$, we can generalize the derivative in (B.3.3) for the vector of parameters $\boldsymbol{\alpha}$ as

$$\frac{\partial \ell_{ci}(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}} = \frac{1}{2}\mathcal{D}_q(\boldsymbol{\Psi}^{-1} \otimes \boldsymbol{\Psi}^{-1})(\operatorname{vec}(\mathbf{b}_i\mathbf{b}_i^\top - \boldsymbol{\Psi})).$$

Finally, at each iteration, we can compute the empirical information matrix (2.3.4) by approximating the score for the observed log-likelihood by the stochastic approximation given in (2.3.5).

B.4 Figures

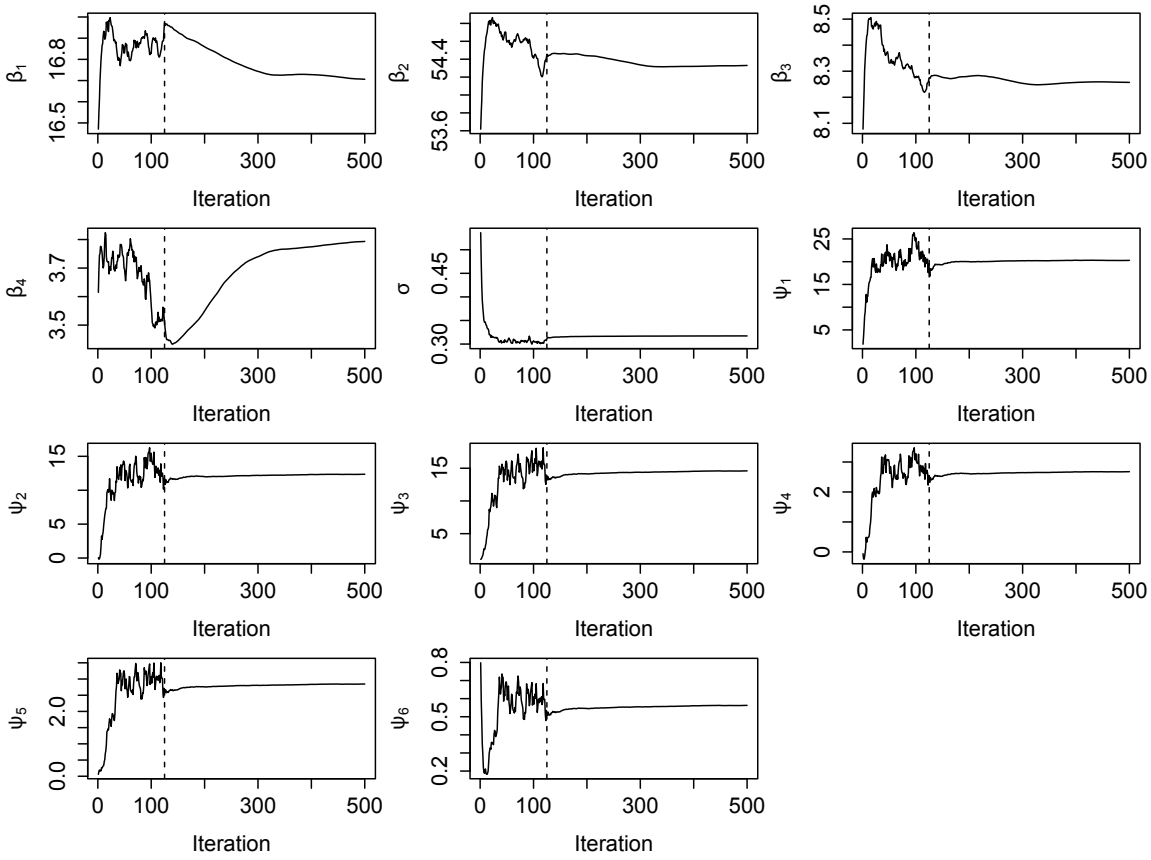


Figure B.1: Graphical summary for the convergence of the fixed effect estimates, variance components of the random effects, and nuisance parameters performing a median regression for the Soybean data. The vertical dashed line delimits the beginning of the almost sure convergence as defined by the cut-point parameter $c = 0.25$.

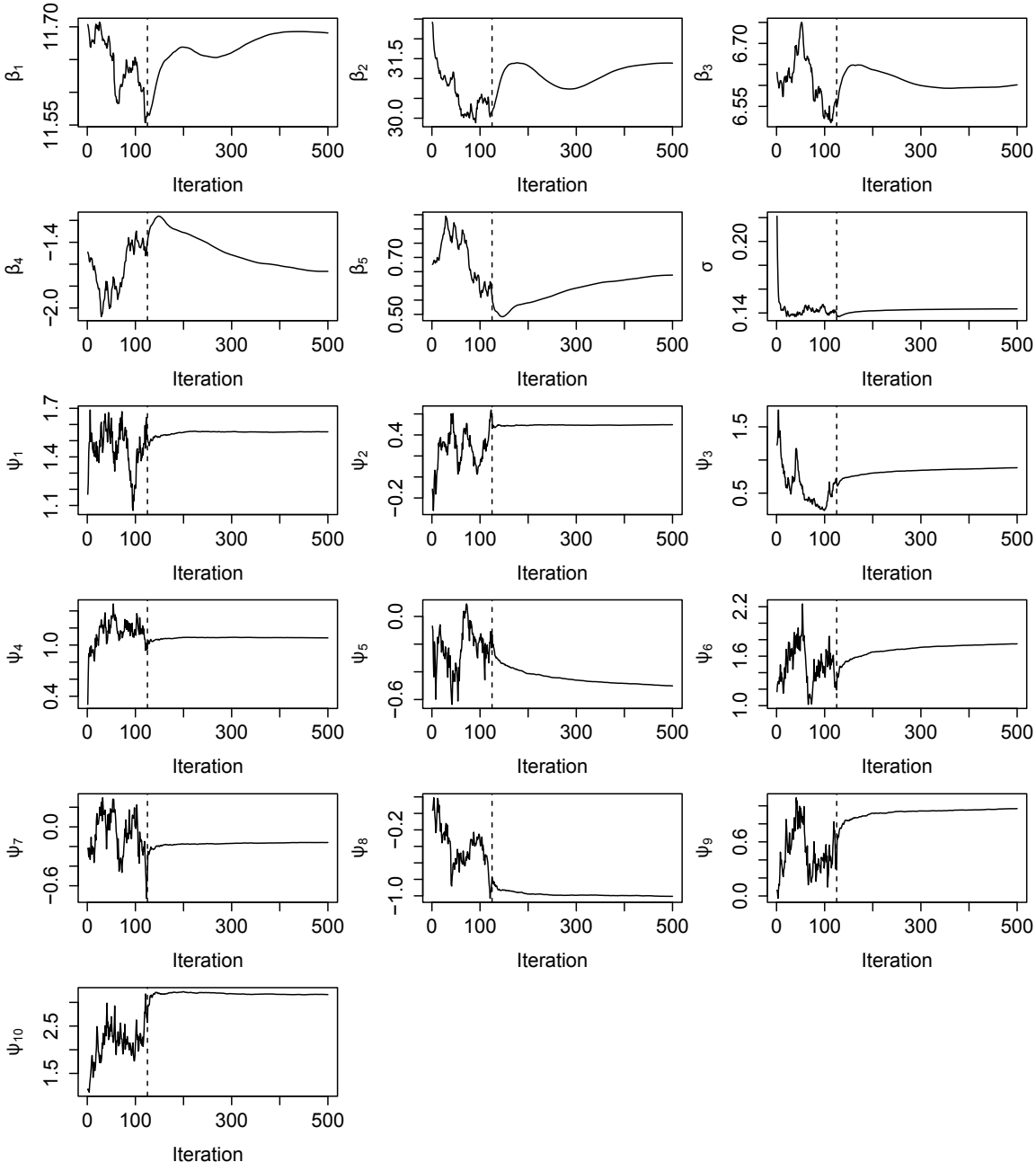


Figure B.2: Graphical summary for the convergence of the fixed effect estimates, variance components of the random effects, and nuisance parameters performing a median regression for the HIV data. The vertical dashed line delimits the beginning of the almost sure convergence as defined by the cut-point parameter $c = 0.25$.

B.5 Sample output from R package qrNLMM()

```
------------------------------------------------
Quantile Regression for Nonlinear Mixed Model
------------------------------------------------
Quantile = 0.5
Subjects = 48 ; Observations = 412

- Nonlinear function
function(x,fixed,random,covar=NA){
resp = (fixed[1] + random[1])/(1 + exp(((fixed[2] +
random[2]) - x)/(fixed[3] + random[3])))
return(resp)}

-----------
Estimates
-----------
- Fixed effects
Estimate Std. Error   z value Pr(>|z|)
beta 1 18.80029    0.53098  35.40704        0
beta 2 54.47930    0.29571 184.23015        0
beta 3  8.25797    0.09198  89.78489        0

sigma = 0.31569

Random effects Variance-Covariance Matrix matrix
      b1       b2      b3
b1 24.36687 12.27297 3.24721
b2 12.27297 15.15890 3.09129
b3  3.24721  3.09129 0.67193

------------------------
Model selection criteria
------------------------
Loglik      AIC      BIC       HQ
Value -622.899 1265.798 1306.008 1281.703

-------
Details
-------
Convergence reached? = FALSE
Iterations = 300 / 300
Criteria = 0.00058
MC sample = 20
Cut point = 0.25
Processing time = 22.83885 mins
```

Appendix C

Licença

C.1 Sobre a licença dessa obra

A licença Creative Commons Atribuição-CompartilhaIgual 3.0 Não Adaptada utilizada nessa obra diz que:

1. Você tem a liberdade de:
    - Compartilhar — copiar, distribuir e transmitir a obra;
    - Remixar — criar obras derivadas;
    - fazer uso comercial da obra.
2. Sob as seguintes condições:
    - Atribuição — Você deve creditar a obra da forma especificada pelo autor ou licenciante (mas não de maneira que sugira que estes concedem qualquer aval a você ou ao seu uso da obra).
    - Compartilhamento pela mesma licença — Se você alterar, transformar ou criar em cima desta obra, você poderá distribuir a obra resultante apenas sob a mesma licença, ou sob uma licença similar à presente.