


Quantile regression for nonlinear mixed effects models: a likelihood based perspective

Christian E. Galarza^{1,2} · Luis M. Castro³  · Francisco Louzada⁴ · Victor H. Lachos⁵

Received: 1 May 2016 / Revised: 3 February 2018 / Published online: 24 February 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

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 under non-normal error distribution. However, at least in principle, quantile regression (QR) is more robust in the presence of outliers/influential observations and misspecification of the error distributions when compared to the conventional mean regression approach. In this context, this paper develops a likelihood-based approach for estimating QR models with correlated continuous longitudinal data using the asymmetric Laplace distribution. Our approach relies on the stochastic approximation of the EM algorithm (SAEM algorithm), obtaining maximum likelihood estimates of the fixed effects and variance components in the case of nonlinear mixed effects (NLME) models. We evaluate the finite sample performance of the SAEM algorithm and asymptotic properties of the ML estimates through simulation experiments. Moreover, two real life datasets are used to illustrate our proposed method using the `qrNLMM` package from R.

✉ Luis M. Castro
mcastro@mat.puc.cl

- 1 Departamento de Matemáticas, Escuela Superior Politécnica del Litoral, ESPOL, Guayaquil, Ecuador
- 2 Department of Statistics, Universidade Estadual de Campinas, São Paulo, Brazil
- 3 Department of Statistics, Pontificia Universidad Católica de Chile, Casilla 306, Correo 22, Santiago, Chile
- 4 Department of Applied Mathematics and Statistics, Universidade de São Paulo, São Carlos, Brazil
- 5 Department of Statistics, University of Connecticut, Storrs 06269, USA

Keywords Asymmetric Laplace distribution · Nonlinear mixed effects models · Quantile regression · SAEM algorithm

1 Introduction

Nonlinear mixed-effects (NLME) models are frequently used to analyze grouped, clustered, longitudinal and multilevel data, among other situations. This is because this type of model allows dealing with nonlinear relationships between the observed response and the covariates and/or random effects, and at the same time, takes into account within- and between-subject correlations in the statistical modeling of the observed data. In general, NLME models arise as a consequence of the mathematical modeling of biological, chemical and physical phenomena, using known families of nonlinear functions with attractive properties such as the presence of asymptote, uniqueness of maximum value, monotonicity and positive range (Pinheiro and Bates 2000; Davidian and Giltinan 2003; Wu 2010). Although most of the current NLME model research is focused on estimating of the conditional mean of the response given some covariates, sometimes the estimation of this quantity lacks meaning, especially when the conditional distribution of the response (given the covariates) is asymmetric, multimodal or simply is severely affected by atypical observations (outliers). In this case, the use of conditional quantile regression (QR) methods (Koenker 2004, 2005) becomes an appropriate strategy for describing the conditional distribution of the outcome variable given the covariates. One of the advantages of using QR methods is that they do not impose any distribution assumption on the error term, except that this term must have a conditional quantile equal to zero. Moreover, and from a practical viewpoint, standard QR methods are already implemented in statistical software such R in its package `quantreg()` (R Core Team 2017).

QR methods were initially developed under a univariate framework (see for example Galarza et al. 2017) but, nowadays, the abundance of correlated data in real-life applications has generated the study of several extensions of QR methods based on mixed models. Some of these extensions consider the distribution-free approach (Lipsitz et al. 1997; Galvao and Montes-Rojas 2010; Galvao 2011; Fu and Wang 2012; Andriyana et al. 2016), while others consider the traditional likelihood-based approach using the asymmetric Laplace (AL) distribution (see for example, Geraci and Bottai 2007; Yuan and Yin 2010; Geraci and Bottai 2014) and Aghamohammadi and Mohammadi (2017). In this context, Geraci and Bottai (2007) proposed a Monte Carlo EM (MCEM) algorithm for the QR model considering continuous responses with a subject-specific univariate random intercept. Recently, Geraci and Bottai (2014) extended their previous work by considering a general QR linear mixed effects (QR-LME) model with multiple random effects. In that work, the authors considered the estimation of the fixed effects and the covariance components through an efficient combination of Gaussian quadrature approximations and nonsmooth optimization algorithms. On the other hand, Yuan and Yin (2010) extend the QR model proposed by Geraci and Bottai (2007) to the case of linear mixed effects (LME) models for longitudinal measurements with missing data. In the nonlinear case, Wang (2012) considered a QR-NLME model from a Bayesian perspective, showing that this model may be a better alternative

than the mean regression estimation in the presence of asymmetric and multimodal data.

Although some results based on QR-NLME models have recently appeared in the statistical literature, to the best of our knowledge, there are no studies and contributions considering an exact inference for QR-NLME models from a likelihood-based perspective. For that reason, the aim of this paper is to propose a QR-NLME model using the AL distribution and considering a full likelihood-based inference via the implementation of the stochastic version of the EM algorithm, *a.k.a.* SAEM algorithm, proposed by Delyon et al. (1999) for the maximum likelihood (ML) estimation. The SAEM algorithm has been proved to be more computationally efficient algorithm than the classic MCEM algorithm due to the recycling of simulations from one iteration to the next in the smoothing phase. Moreover, as pointed out by Meza et al. (2012), the SAEM algorithm, unlike the MCEM, converges even with a small simulation size. In this context, Kuhn and Lavielle (2005) showed that the SAEM algorithm is very efficient for computing the ML estimates in mixed effects models. It is important to stress that, the empirical results show that the ML estimates based on our proposed SAEM algorithm provide good asymptotic properties. Moreover, the application of our method is conducted by using our recent R package, called, `qrNLMM` ().

The paper is organized as follows. Section 2 presents some preliminary results. Particularly, we present some properties of the AL distribution and an outline of the SAEM algorithm. Section 3 provides the MCEM and SAEM algorithms for a general NLME model, while Sect. 4 outlines the likelihood-based estimation and standard errors of the parameter estimates under the proposed model. Section 5 presents the results of simulation studies conducted to analyze the performance of the proposed method with respect to the asymptotic properties of the ML estimates and the consequences of the misspecification of the random effects distribution. The analysis of two longitudinal datasets is presented in Sect. 6. Finally, Sect. 7 closes the paper by presenting a brief plan of our future work.

2 Preliminaries

In this section, we provide some useful results related to the AL distribution. We also present some background on the SAEM algorithm for ML estimation.

2.1 AL distribution

A random variable Y has an AL distribution (see Kozubowski and Nadarajah 2010) with location parameter μ , scale parameter $\sigma > 0$ and skewness parameter $p \in (0, 1)$, if its probability density function (*pdf*) is given by

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

where $\rho_p(\cdot)$ is the check (or loss) function defined by $\rho_p(u) = u(p - \mathbb{I}\{u < 0\})$, with $\mathbb{I}\{\cdot\}$ the indicator function. This distribution is denoted by $\text{AL}(\mu, \sigma, p)$.

The random variable $Y \sim \text{AL}(\mu, \sigma, p)$ can be represented hierarchically as

$$\begin{aligned} Y | U = u &\sim N(\mu + \vartheta_p u, \tau_p^2 \sigma u), \\ U &\sim \text{Exp}(\sigma), \end{aligned} \tag{1}$$

where $\text{Exp}(\sigma)$ denotes the exponential distribution with mean σ , ϑ_p and τ_p^2 are constants such that $\vartheta_p = \frac{1-2p}{p(1-p)}$ and $\tau_p^2 = \frac{2}{p(1-p)}$. Note that $U | Y \sim \text{GIG}(1/2, \delta, \gamma)$, where $\text{GIG}(v, a, b)$ denotes the Generalized Inverse Gaussian (GIG) distribution (Barndorff-Nielsen and Shephard 2001) with *pdf* given by

$$h(z | v, a, b) = \frac{(b/a)^v}{2K_v(ab)} z^{v-1} \exp \left\{ -\frac{1}{2}(a^2/z + b^2z) \right\},$$

with $u > 0, v \in \mathbb{R}, a, b > 0$, where $K_v(\cdot)$ is the modified Bessel function of the third kind. The moments of the GIG distribution are given by

$$E[Z^k] = \left(\frac{a}{b}\right)^k \frac{K_{v+k}(ab)}{K_v(ab)}, \quad k \in \mathbb{R}. \tag{2}$$

2.2 The SAEM algorithm

The SAEM algorithm proposed by Delyon et al. (1999) replaces the E-step of the EM algorithm (Dempster et al. 1977) by a stochastic approximation procedure. Besides having good theoretical properties, the SAEM algorithm estimates the population parameters accurately, converging to the global maxima of the ML estimates under quite general conditions (Allasonnière et al. 2010; Delyon et al. 1999; Kuhn and Lavielle 2004).

Let θ be the vector of the parameter of interest, \mathbf{y} be the observed data and \mathbf{q} be the missing data and/or latent variables. At each iteration, the SAEM algorithm successively simulates missing data/latent variables using their conditional distributions, updating the model parameters. Thus, at iteration k , the SAEM algorithm proceeds as follows:

E-Step:

- Simulation: draw a sample $(\mathbf{q}^{(\ell,k)})$, $\ell = 1, \dots, m$ from the conditional distribution $f(\mathbf{q} | \theta^{(k-1)}, \mathbf{y})$.
- Stochastic approximation: update $Q(\theta | \hat{\theta}^{(k)})$ as

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

where m is the number of simulations 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 $\sum_{k=1}^{\infty} \delta_k^2 < \infty$.

M-Step:

- Maximization: update $\widehat{\boldsymbol{\theta}}^{(k)}$ as $\widehat{\boldsymbol{\theta}}^{(k+1)} = \operatorname{argmax}_{\boldsymbol{\theta}} Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)})$.

Note that, although the E-Step is similar in the SAEM and MCEM algorithms, a small number of simulations m (for practical situations, $m \leq 20$ is suggested) is necessary in the first one. 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/latent variables at the iteration k but some or all the previous simulations. In fact, this ‘memory’ property is set by the smoothing parameter δ_k . In our case, we suggested the following choice of the smoothing parameter given as

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

where W is the maximum number of iterations, and c is a cutoff point ($0 \leq c \leq 1$) which determines the percentage of initial iterations with no memory.

3 The QR nonlinear mixed model

3.1 The model

In this section, we propose the following general mixed-effects model. Let $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^\top$ be the continuous response of subject i and $\boldsymbol{\eta} = (\eta(\boldsymbol{\phi}_i, x_{i1}), \dots, \eta(\boldsymbol{\phi}_i, x_{in_i}))^\top$ a nonlinear differentiable function of vector-valued random parameters $\boldsymbol{\phi}_i$ of dimension r . Moreover, let \mathbf{x}_i be a matrix of covariates of dimension $n_i \times r$. The NLME model is defined as

$$\begin{aligned} \mathbf{y}_i &= \boldsymbol{\eta}(\boldsymbol{\phi}_i, \mathbf{x}_i) + \boldsymbol{\epsilon}_i, \\ \boldsymbol{\phi}_i &= \mathbf{A}_i \boldsymbol{\beta}_p + \mathbf{B}_i \mathbf{b}_i, \end{aligned} \tag{3}$$

where \mathbf{A}_i and \mathbf{B}_i are design matrices (fixed) 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 p -th quantile, $\mathbf{b}_i = (b_{i1}, \dots, b_{iq})^\top$ is a q -dimensional random effects vector associated with the i -th subject and $\boldsymbol{\epsilon}_i$ is the independent and identically distributed vector of random errors. We define the p -th quantile function of the response y_{ij} as

$$\begin{aligned} Q_p(y_{ij} \mid x_{ij}, \mathbf{b}_i) &= \eta(\boldsymbol{\phi}_i, x_{ij}) \\ &= \eta(\mathbf{A}_i \boldsymbol{\beta}_p + \mathbf{B}_i \mathbf{b}_i, x_{ij}), \end{aligned} \tag{4}$$

where Q_p denotes the inverse of the unknown distribution function F . In this setting, the random effects \mathbf{b}_i are independent and identically distributed (*i.i.d*) as $N_q(\mathbf{0}, \boldsymbol{\Psi})$, where the dispersion matrix $\boldsymbol{\Psi} = \boldsymbol{\Psi}(\boldsymbol{\alpha})$ depends on unknown and reduced parameters $\boldsymbol{\alpha}$. The error terms are distributed as $\epsilon_{ij} \stackrel{\text{iid}}{\sim} \text{AL}(0, \sigma, p)$, being uncorrelated with the

random effects. Then, conditionally on \mathbf{b}_i , the observed responses for subject i , *i.e.*, y_{ij} for $j = 1, \dots, n_i$, are independent following an AL distribution with location, scale and skewness parameters given by $\eta(\mathbf{A}_i\boldsymbol{\beta}_p + \mathbf{B}_i\mathbf{b}_i, \mathbf{x}_{ij})$, σ and p , respectively.

The main reason for considering the AL distribution in our approach is that this distribution proves useful as a unifying bridge between the likelihood inference and the inference about QR estimation and seems to be very promising within those models, since it clearly represents a suitable error law for the least absolute estimator (Geraci and Bottai 2007). In addition, and from the Bayesian standpoint, Yu and Moyeed (2001) have showed that irrespective of the original distribution of the data, the use of the AL distribution is a very natural and effective way to model QR. Recently, Sriram et al. (2013) studied the posterior consistency of model parameters for a QR model when the AL distribution is misspecified, concluding that the use of the AL distribution works for a variety of possibilities of the true likelihood, including location models, scale models and location-scale models.

3.2 The MCEM algorithm

In this section we develop a MCEM algorithm for ML estimation of the parameters in the QR-NLME model. This model has a flexible hierarchical representation, which is useful for deriving interesting theoretical properties. From (1), we have that the QR-NLME model defined in (3)–(4), can be represented as

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

for $i = 1, \dots, n$, where ϑ_p and τ_p^2 are given as in (1) and \mathbf{D}_i represents a diagonal matrix that contains the vector of latent variables $\mathbf{u}_i = (u_{i1}, \dots, u_{in_i})^\top$. Let $\mathbf{y}_{ic} = (\mathbf{y}_i^\top, \mathbf{b}_i^\top, \mathbf{u}_i^\top)^\top$ and let $\boldsymbol{\theta}^{(k)} = (\boldsymbol{\beta}_p^{(k)\top}, \sigma^{(k)}, \boldsymbol{\alpha}^{(k)\top})^\top$, the estimate of $\boldsymbol{\theta}$ at the k -th iteration. For simplicity, we denote $\boldsymbol{\eta}_i = \eta(\mathbf{A}_i\boldsymbol{\beta}_p + \mathbf{B}_i\mathbf{b}_i, \mathbf{x}_i)$. Since \mathbf{b}_i and \mathbf{u}_i are independent for all $i = 1, \dots, n$, it follows from (1) that the complete-data log-likelihood function is given by $\ell_c(\boldsymbol{\theta}; \mathbf{y}_c) = \sum_{i=1}^n \ell_c(\boldsymbol{\theta}; \mathbf{y}_{ic})$, where

$$\begin{aligned} \ell_c(\boldsymbol{\theta}; \mathbf{y}_{ic}) &= K - \frac{3}{2}n_i \log \sigma - \frac{1}{2} \log |\boldsymbol{\Psi}| - \frac{1}{2} \mathbf{b}_i^\top \boldsymbol{\Psi}^{-1} \mathbf{b}_i - \frac{1}{\sigma} \mathbf{u}_i^\top \mathbf{1}_{n_i} \\ &\quad - \frac{1}{2\sigma\tau_p^2} (\mathbf{y}_i - \boldsymbol{\eta}_i - \vartheta_p\mathbf{u}_i)^\top \mathbf{D}_i^{-1} (\mathbf{y}_i - \boldsymbol{\eta}_i - \vartheta_p\mathbf{u}_i), \end{aligned} \tag{5}$$

where K denotes a constant that does not depend on the parameter of interest and $\mathbf{1}_p$ is a vector of ones of dimension p . Given the current estimate $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$, the E-step calculates the function $Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^n Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)})$, where

$$\begin{aligned}
 Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) &= E \left\{ \ell_c(\boldsymbol{\theta}; \mathbf{y}_{i_c}) \mid \boldsymbol{\theta}^{(k)}, \mathbf{y} \right\} \propto -\frac{3}{2}n_i \log \sigma - \frac{1}{2} \log |\boldsymbol{\Psi}| \\
 &\quad - \frac{1}{2} \text{tr} \left\{ (\widehat{\mathbf{b}\mathbf{b}^\top})_i^{(k)} \boldsymbol{\Psi}^{-1} \right\} - \frac{1}{2\sigma\tau_p^2} \left[\mathbf{y}_i^\top \widehat{\mathbf{D}}_i^{-1(k)} \mathbf{y}_i - 2\vartheta_p \mathbf{y}_i^\top \mathbf{1}_{n_i} \right. \\
 &\quad \left. + \frac{\tau_p^4}{4} \widehat{\mathbf{u}}_i^{(k)\top} \mathbf{1}_{n_i} - 2\mathbf{y}_i^\top (\widehat{\mathbf{D}}^{-1}\boldsymbol{\eta})_i^{(k)} + 2\vartheta_p \mathbf{1}_{n_i}^\top \widehat{\boldsymbol{\eta}}_i^{(k)} + (\boldsymbol{\eta}^\top \widehat{\mathbf{D}}^{-1}\boldsymbol{\eta})_i^{(k)} \right],
 \end{aligned}$$

where $\text{tr}(\mathbf{A})$ denotes the trace of matrix \mathbf{A} . The calculation of this function requires the expressions $\widehat{\boldsymbol{\eta}}_i^{(k)} = E \{ \boldsymbol{\eta}_i \mid \boldsymbol{\theta}^{(k)}, \mathbf{y}_i \}$, $\widehat{\mathbf{u}}_i^{(k)} = E \{ \mathbf{u}_i \mid \boldsymbol{\theta}^{(k)}, \mathbf{y}_i \}$, $(\widehat{\mathbf{b}\mathbf{b}^\top})_i^{(k)} = E \{ \mathbf{b}_i \mathbf{b}_i^\top \mid \boldsymbol{\theta}^{(k)}, \mathbf{y}_i \}$, $\widehat{\mathbf{D}}_i^{-1(k)} = E \{ \mathbf{D}_i^{-1} \mid \boldsymbol{\theta}^{(k)}, \mathbf{y}_i \}$, $(\widehat{\mathbf{D}}^{-1}\boldsymbol{\eta})_i^{(k)} = E \{ \mathbf{D}_i^{-1} \boldsymbol{\eta}_i^{(k)} \mid \boldsymbol{\theta}^{(k)}, \mathbf{y}_i \}$, $(\boldsymbol{\eta}^\top \widehat{\mathbf{D}}^{-1}\boldsymbol{\eta})_i^{(k)} = E \{ \boldsymbol{\eta}_i^\top \mathbf{D}_i^{-1} \boldsymbol{\eta}_i \mid \boldsymbol{\theta}^{(k)}, \mathbf{y}_i \}$, which do not have closed forms. Since the joint distribution of the latent variables $(\mathbf{b}_i^{(k)}, \mathbf{u}_i^{(k)})$ is unknown and the conditional expectations cannot be computed analytically, for any function $g(\cdot)$, the MCEM algorithm approximates these expectations using a Monte Carlo approximation given by

$$E[g(\mathbf{b}_i, \mathbf{u}_i) \mid \boldsymbol{\theta}^{(k)}, \mathbf{y}_i] \approx \frac{1}{m} \sum_{\ell=1}^m g(\mathbf{b}_i^{(\ell,k)}, \mathbf{u}_i^{(\ell,k)}), \tag{6}$$

which depends of the simulations of the two latent variables $\mathbf{b}_i^{(k)}$ and $\mathbf{u}_i^{(k)}$ from the conditional density $f(\mathbf{b}_i, \mathbf{u}_i \mid \boldsymbol{\theta}^{(k)}, \mathbf{y}_i)$. Since $E[g(\mathbf{b}_i, \mathbf{u}_i) \mid \boldsymbol{\theta}^{(k)}, \mathbf{y}_i] = E[E[g(\mathbf{b}_i, \mathbf{u}_i) \mid \boldsymbol{\theta}^{(k)}, \mathbf{b}_i, \mathbf{y}_i] \mid \mathbf{y}_i]$, the expected value given in (6) can be more accurately approximated by

$$\frac{1}{m} \sum_{\ell=1}^m E[g(\mathbf{b}_i^{(\ell,k)}, \mathbf{u}_i) \mid \boldsymbol{\theta}^{(k)}, \mathbf{b}_i^{(\ell,k)}, \mathbf{y}_i], \tag{7}$$

where $\mathbf{b}_i^{(\ell,k)}$ is generated from $f(\mathbf{b}_i \mid \boldsymbol{\theta}^{(k)}, \mathbf{y}_i)$. Note that this approximation is more accurate since it only depends on one Monte Carlo approximation, instead of two (as is needed in (6)).

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

$$\phi(y_{ij} \mid \eta_{ij}(\boldsymbol{\beta}_p, \mathbf{b}_i) + \vartheta_p u_{ij}, \sigma\tau_p^2 u_{ij}) \times \text{Exp}(\sigma),$$

which, from Sect. 2.1, leads to $u_{ij} \mid y_{ij}, \mathbf{b}_i \sim \text{GIG}(\frac{1}{2}, \chi_{ij}, \psi)$, where χ_{ij} and ψ are given by $\chi_{ij} = |y_{ij} - \eta_{ij}|/\tau_p\sqrt{\sigma}$ and $\psi = \tau_p/2\sqrt{\sigma}$, and η_{ij} is the component j of the vector $\boldsymbol{\eta}_i$. From (2), and after generating samples from $f(\mathbf{b}_i \mid \boldsymbol{\theta}^{(k)}, \mathbf{y}_i)$ (see Sect. 3.4), the conditional expectation $E[\cdot \mid \boldsymbol{\theta}, \mathbf{b}_i, \mathbf{y}_i]$ in (7) can be computed analytically. Finally, the proposed MCEM algorithm to estimate the parameters of the QR-NLME model can be summarized as follows:

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

- Simulation step: For $\ell = 1, \dots, m$, generate $\mathbf{b}_i^{(\ell,k)}$ from $f(\mathbf{b}_i \mid \theta^{(k)}, \mathbf{y}_i)$, as described next in Sect. 3.4.
- Monte Carlo approximation: Using (2) and $\mathbf{b}_i^{(\ell,k)}$, for $\ell = 1, \dots, m$, evaluate $E[g(\mathbf{b}_i, \mathbf{u}_i) \mid \theta^{(k)}, \mathbf{y}_i]$ using the approximation proposed in (7).

M-step: Update $\widehat{\theta}^{(k)}$ by maximising $Q(\theta \mid \widehat{\theta}^{(k)}) \approx \frac{1}{m} \sum_{\ell=1}^m \sum_{i=1}^n \ell_c(\theta; \mathbf{y}_i, \mathbf{b}_i^{(\ell,k)}, \mathbf{u}_i)$ over θ , leading the following estimates:

$$\widehat{\beta}_p^{(k+1)} = \widehat{\beta}_p^{(k)} + \left[\sum_{i=1}^n \left\{ \frac{1}{m} \sum_{\ell=1}^m \mathbf{J}_i^{(k)\top} \mathcal{E}(\mathbf{D}_i^{-1})^{(\ell,k)} \mathbf{J}_i^{(k)} \right\} \right]^{-1} \times$$

$$\left[\sum_{i=1}^n \left\{ \frac{1}{m} \sum_{\ell=1}^m \left[2\mathbf{J}_i^{(k)\top} \mathcal{E}(\mathbf{D}_i^{-1})^{(\ell,k)} \left[\mathbf{y}_i - \eta(\widehat{\beta}_p^{(k)}, \mathbf{b}_i^{(\ell,k)}) - \vartheta_p \mathcal{E}(\mathbf{u}_i)^{(\ell,k)} \right] \right] \right\} \right],$$

$$\widehat{\sigma}^{(k+1)} = \frac{1}{3N\tau_p^2} \sum_{i=1}^n \left\{ \frac{1}{m} \sum_{\ell=1}^m \left[(\mathbf{y}_i - \eta(\widehat{\beta}_p^{(k+1)}, \mathbf{b}_i^{(\ell,k)}))^\top \mathcal{E}(\mathbf{D}^{-1})^{(\ell,k)} (\mathbf{y}_i \eta(\widehat{\beta}_p^{(k+1)}, \mathbf{b}_i^{(\ell,k)})) \right. \right.$$

$$\left. \left. - 2\vartheta_p (\mathbf{y}_i \eta(\widehat{\beta}_p^{(k+1)}, \mathbf{b}_i^{(\ell,k)}))^\top \mathbf{1}_{n_i} + \frac{\tau_p^4}{4} \mathcal{E}(\mathbf{u}_i)^{(\ell,k)\top} \mathbf{1}_{n_i} \right] \right\},$$

$$\widehat{\psi}^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{m} \sum_{\ell=1}^m \mathbf{b}_i^{(\ell,k)} \mathbf{b}_i^{(\ell,k)\top} \right],$$

where $\mathbf{J}_i = \partial \eta(\beta_p, \mathbf{b}_i) / \partial \beta_p^\top$, $N = \sum_{i=1}^n n_i$ and expressions $\mathcal{E}(\mathbf{u}_i)^{(\ell,k)}$ and $\mathcal{E}(\mathbf{D}_i^{-1})^{(\ell,k)}$ are defined in Appendix A2.

Note that, for the MC E-step, we need to generate $\mathbf{b}_i^{(\ell,k)}$, $\ell = 1, \dots, m$, from $f(\mathbf{b}_i \mid \theta^{(k)}, \mathbf{y}_i)$ (see Sect. 3.4), where m is the number of Monte Carlo simulations to be used (a number suggested to be large enough).

3.3 The SAEM algorithm

As was mentioned in Sect. 2.2, the SAEM circumvents the problem of simulating a large number of latent values at each iteration, leading to a faster and efficient solution to the MCEM algorithm. In summary, the SAEM algorithm proceeds as follows:

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

- Stochastic approximation: Update the Monte Carlo approximations using stochastic ones, given by

$$S_{1,i}^{(k)} = S_{1,i}^{(k-1)} + \delta_k \left[\frac{1}{m} \sum_{\ell=1}^m \mathbf{J}_i^{(k)\top} \mathcal{E}(\mathbf{D}_i^{-1})^{(\ell,k)} \mathbf{J}_i^{(k)} - S_{1,i}^{(k-1)} \right],$$

$$S_{2,i}^{(k)} = S_{2,i}^{(k-1)} + \delta_k \left[\frac{1}{m} \sum_{\ell=1}^m \left[2\mathbf{J}_i^{(k)\top} \mathcal{E}(\mathbf{D}_i^{-1})^{(\ell,k)} \left[\mathbf{y}_i - \eta(\widehat{\beta}_p^{(k)}, \mathbf{b}_i^{(\ell,k)}) \right] \right. \right.$$

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

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

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

provided that the inverse of the matrix $S_{1,i}^{(k)}$ exists. Given a set of suitable initial values $\widehat{\boldsymbol{\theta}}^{(0)}$ (a simple strategy for specifying $\widehat{\boldsymbol{\theta}}^{(0)}$ is described in Appendix A1), the SAEM iterates until convergence at iteration k , if $\max_i \left\{ \frac{|\widehat{\theta}_i^{(k+1)} - \widehat{\theta}_i^{(k)}|}{|\widehat{\theta}_i^{(k)}| + \delta_1} \right\} < \delta_2$, where δ_1 and δ_2 are pre-established small values. As suggested on page 269 in Searle et al. (1992), we used $\delta_1 = 0.001$ and $\delta_2 = 0.0001$. In addition, we also used a second convergence criterion defined by $\max_i \left\{ \frac{|\widehat{\theta}_i^{(k+1)} - \widehat{\theta}_i^{(k)}|}{\sqrt{\widehat{\text{var}}(\widehat{\theta}_i^{(k)}) + \delta_1}} \right\} < \delta_2$, as proposed by Booth and Hobert (1999).

3.4 Simulation of random effects

In order to generate samples from $f(\mathbf{b}_i \mid \mathbf{y}_i)$, we use the Metropolis-Hastings (MH) algorithm (Metropolis et al. 1953; Hastings 1970), noting that the conditional distribution $f(\mathbf{b}_i \mid \mathbf{y}_i)$ can be represented as $f(\mathbf{b}_i \mid \mathbf{y}_i) \propto f(\mathbf{y}_i \mid \mathbf{b}_i) f(\mathbf{b}_i)$, where $\mathbf{b}_i \sim N_q(\mathbf{0}, \Psi)$ and $f(\mathbf{y}_i \mid \mathbf{b}_i) = \prod_{j=1}^{n_i} f(y_{ij} \mid \mathbf{b}_i)$, with $y_{ij} \mid \mathbf{b}_i \sim AL(\boldsymbol{\eta}(\mathbf{A}_i \boldsymbol{\beta}_p + \mathbf{B}_i \mathbf{b}_i, \mathbf{x}_{ij}), \sigma, p)$. Since this function is a product of two distributions

(with support in \mathbb{R}^q), a suitable choice for the proposed density is a multivariate normal distribution with mean and variance-covariance matrix given by $\widehat{\boldsymbol{\mu}}_{\mathbf{b}_i} = E(\mathbf{b}_i^{(k-1)} \mid \mathbf{y}_i)$ and $\widehat{\boldsymbol{\Sigma}}_{\mathbf{b}_i} = \text{Var}(\mathbf{b}_i^{(k-1)} \mid \mathbf{y}_i)$ respectively. These quantities are obtained from the last iteration of the SAEM algorithm. Note that this candidate generates a high acceptance rate, making the algorithm fast.

4 Estimation of the likelihood and standard errors

4.1 Likelihood estimation

Given an observed data, the likelihood function $\ell_o(\boldsymbol{\theta} \mid \mathbf{y})$ of the model defined in (3)-(4) is given by

$$\sum_{i=1}^n \log \int_{\mathbb{R}^q} f(\mathbf{y}_i \mid \mathbf{b}_i; \boldsymbol{\theta}) f(\mathbf{b}_i; \boldsymbol{\theta}) d\mathbf{b}_i, \tag{8}$$

where the integral can be expressed as an expectation with respect to \mathbf{b}_i , *i.e.*, $E[f(\mathbf{y}_i \mid \mathbf{b}_i; \boldsymbol{\theta})]$. The evaluation of this integral is not possible analytically, so it is often replaced by its Monte Carlo approximation involving a large number of simulations. However, alternative importance sampling (IS) procedures might require a smaller number of simulations than the typical Monte Carlo procedure. Following Meza et al. (2012), we 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 $f(\mathbf{b}_i; \boldsymbol{\theta})$. Re-writing (8) as

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

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

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

where $\{\mathbf{b}_i^{*(\ell)}\}, l = 1, \dots, m$, is a Monte Carlo sample from $\widehat{f}(\mathbf{b}_i^*; \boldsymbol{\theta})$, and $f(\mathbf{y}_i \mid \mathbf{b}_i^{*(\ell)}; \boldsymbol{\theta})$ is expressed as $\prod_{j=1}^{n_i} f(y_{ij} \mid \mathbf{b}_i^{*(\ell)}; \boldsymbol{\theta})$ due to the conditional independence assumption. An efficient choice for $\widehat{f}(\mathbf{b}_i^{*(\ell)}; \boldsymbol{\theta})$ is $f(\mathbf{b}_i \mid \mathbf{y}_i)$. Therefore, we use the same proposed distribution discussed in Sect. 3.4, generating $\mathbf{b}_i^{*(\ell)} \sim N_q(\widehat{\boldsymbol{\mu}}_{\mathbf{b}_i}, \widehat{\boldsymbol{\Sigma}}_{\mathbf{b}_i})$.

4.2 Standard error approximation

At the k th iteration, the empirical score function $s(\mathbf{y}_i \mid \boldsymbol{\theta})^{(k)}$ for the i -th subject can be computed as

$$s(\mathbf{y}_i | \boldsymbol{\theta})^{(k-1)} + \delta_k \left[\frac{1}{m} \sum_{\ell=1}^m s(\mathbf{y}_i, \mathbf{q}^{(k,\ell)}; \boldsymbol{\theta}^{(k)}) - s(\mathbf{y}_i | \boldsymbol{\theta})^{(k-1)} \right], \tag{9}$$

where $\mathbf{q}^{(\ell,k)}, \ell = 1, \dots, m$, are the simulated values of the conditional distribution $f(\cdot | \boldsymbol{\theta}^{(k-1)}, \mathbf{y}_i)$. Using Louis’s method (Louis 1982), the observed information matrix at iteration k , can be approximated as $\mathbf{I}_e(\boldsymbol{\theta} | \mathbf{y})^{(k)} = \sum_{i=1}^n s(\mathbf{y}_i | \boldsymbol{\theta})^{(k)} s^\top(\mathbf{y}_i | \boldsymbol{\theta})^{(k)}$. Expressions for the elements of the score vector are given in Appendix A3.

5 Simulated data

In order to examine the performance of the proposed method, we conduct some simulation studies. The first simulation study shows that the ML estimates based on the SAEM algorithm provide good asymptotic properties. The second study investigates the consequences on population inferences when the normality assumption of the random effects is not considered. To do that, we used a heavy tailed distribution for the random effect term, testing the robustness of the proposed method in terms of the parameter estimation.

5.1 Finite sample properties

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

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

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 Gaussian random effect b_{1i} , for $i = 1, \dots, n$ is added to the first growth parameter β_1 and its effect on the growth-curve is shown in Fig. 1.

Parameter interpretation of this model is discussed in Sect. 6. The random effect b_{1i} and the error term $\epsilon_i = (\epsilon_{i1} \dots, \epsilon_{i10})^\top$ are non-correlated. In fact, $b_{1i} \stackrel{iid}{\sim} N(0, \sigma_b^2)$ and $\epsilon_{ij} \stackrel{iid}{\sim} AL(0, \sigma_e, p)$. We set $\boldsymbol{\beta}_p = (\beta_1, \beta_2, \beta_3)^\top = (200, 700, 350)^\top, \sigma_e = 0.5$ and $\sigma_b^2 = 10$. Using the notation in (3), the matrices \mathbf{A}_i and \mathbf{B}_i are given by \mathbf{I}_3 and $(1, 0, 0)^\top$ respectively. For different sample sizes, say, $n = 25, 50, 100$ and 200 , we generate 100 datasets for each scenario. In addition, we choose $m = 20, c = 0.25$ and $W = 500$ for the SAEM algorithm convergence parameters. For all scenarios, we compute the square root of the mean square error (RMSE), bias (Bias) and Monte Carlo standard deviation (Sd) for each parameter over the 100 replicates. These quantities are defined as $RMSE(\hat{\theta}_i) = \sqrt{Sd^2(\hat{\theta}_i) + Bias^2(\hat{\theta}_i)}$, $Bias(\hat{\theta}_i) = \bar{\hat{\theta}}_i - \theta_i$ and $Sd(\hat{\theta}_i) = \sqrt{\frac{1}{99} \sum_{j=1}^{100} (\hat{\theta}_i^{(j)} - \bar{\hat{\theta}}_i)^2}$, where $\bar{\hat{\theta}}_i = \frac{1}{100} \sum_{j=1}^{100} \hat{\theta}_i^{(j)}$ is the Monte Carlo mean (Mean) with $\hat{\theta}_i^{(j)}$ the estimate of θ_i from the j -th sample, $j = 1 \dots 100$. Based on Fig. 2, we conclude that the bias in the estimation of fixed effects converges to zero when

Fig. 1 Effect of including a random effect b_1 in the first parameter of the nonlinear growth-curve logistic model

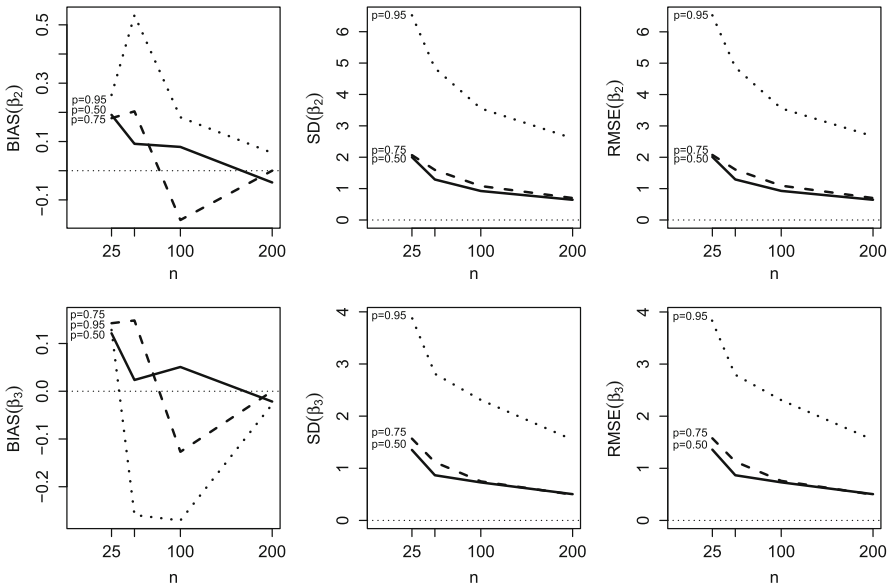
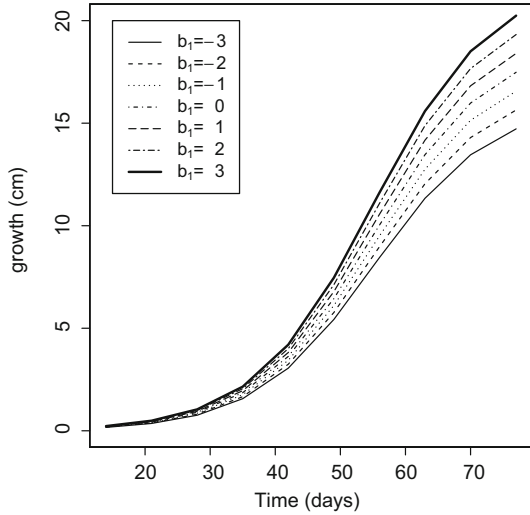


Fig. 2 Bias, Sd and RMSE for β_2 (upper panel) and β_3 (lower panel) for different sample sizes over the quantiles $p = \{0.50, 0.75, 0.95\}$

n increases, *i.e.* the estimators are asymptotically unbiased for the parameters. The values of Sd and RMSE decrease monotonically when n increases, indicating that the estimators are consistent. Note that for quantile $p = 0.95$ (an extreme quantile), the standard deviation is much higher than quantiles $p = 0.50$ and $p = 0.75$. As an overall conclusion, we can say that, in general, the proposed SAEM algorithm provides good asymptotic properties for the ML estimates.

Table 1 Simulation 1: Monte Carlo mean and RMSE for fixed effects β and scale parameter σ_e obtained after fitting the QR-NLME model under different settings of quantiles and sample sizes

Quantile	n	β_1		β_2		β_3		σ_e	
		Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
50%	25	199.75	1.554	700.19	1.428	350.13	1.170	0.503	0.186
	50	199.79	1.315	700.09	1.139	350.03	0.931	0.498	0.145
	100	200.16	1.085	700.08	0.965	350.06	0.856	0.497	0.131
	200	200.02	0.866	699.96	0.803	349.98	0.710	0.499	0.111
75%	25	203.77	4.086	700.18	1.449	350.15	1.262	0.499	0.186
	50	203.90	4.125	700.20	1.279	350.16	1.069	0.495	0.157
	100	204.19	4.347	699.83	1.054	349.88	0.874	0.499	0.131
	200	204.34	4.445	700.00	0.836	350.01	0.703	0.498	0.106
95%	25	201.15	2.029	700.26	2.567	350.14	1.986	0.506	0.186
	50	201.77	2.299	700.53	2.263	349.74	1.704	0.508	0.156
	100	201.93	2.303	700.18	1.894	349.73	1.549	0.505	0.123
	200	202.11	2.341	700.06	1.714	349.98	1.241	0.502	0.109

The results are based on 100 simulated datasets

Table 2 Simulation 1: Sd, Monte Carlo mean of the approximate standard error obtained through the information-based method described in Sect. 4.2 (IM SE) and Monte Carlo coverage probability at 95% (CP %) obtained after fitting the QR-NLME model for quantiles $p = 0.50$ and $p = 0.75$ and sample of size $n = 50$

Q	β_2			β_3			σ_e		
	Sd	IM-SE	CP %	Sd	IM-SE	CP %	Sd	IM-SE	CP %
50%	0.644	0.559	92	0.504	0.477	93	0.012	0.012	95
75%	0.699	0.608	92	0.495	0.516	95	0.011	0.012	96

The results are based on 100 simulated samples

Table 1 shows the estimation of the model parameters. Note that the intercept (β_1) increases along the quantiles. This is a natural property of the intercept in QR models. As a consequence, the bias and RMSE increase significantly for extreme quantiles. In general, and for all the model parameters, good asymptotic properties in terms of bias and RMSE are observed.

As was suggested by a reviewer, Table 2 shows the coverage probability and the theoretical standard errors (obtained using the approximation proposed in Sect. 4.2) for the QR-NLME model (10) considering, for computational time reasons, a sample of size $n = 50$ and quantiles $p = 0.50$ and $p = 0.75$. From these results, one can see that the empirical (Sd) and theoretical (IM-SE) standard errors are similar. Moreover, the Monte Carlo coverage probability (CP %) is high, showing the good behavior of the obtained estimates. We have omitted the results for the intercept β_1 , because this parameter varies the quantile, capturing the location of the different quantiles of the conditional distribution. This evidence indicates that the AL assumption does not generate an important bias in the results.

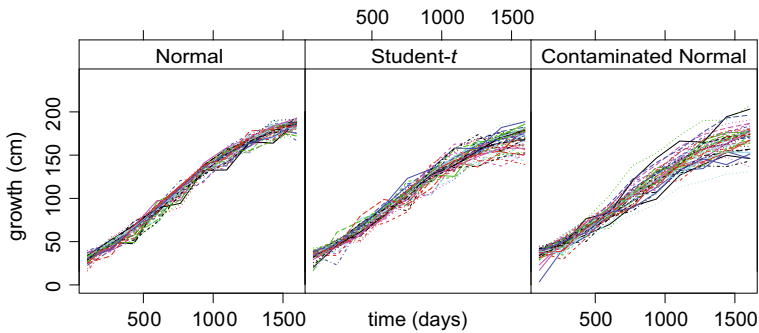


Fig. 3 50 simulated curves from the growth-curve logistic model using different distributions for the random effect term: normal (right), Student's- t with $\nu = 4$ (center), contaminated normal with $\nu_1 = 0.1$ and $\nu_2 = 0.1$ (left). In all cases the location and scale parameters are $\mu = 0$ and $\sigma_b^2 = 10$, respectively

5.2 Robustness study

The aim of this simulation scheme is to study the behavior of the parameter estimates when the distribution of random effects is misspecified. We consider a similar simulation scheme as in the previous subsection, but considering a set of quantiles $p = \{0.50, 0.75\}$ and a fixed sample size $n = 50$. We consider 100 Monte Carlo samples, generating the random effect term from (a) a Student's- t distribution with $\nu = 4$ degrees of freedom (t_4) and from (b) a contaminated normal distribution with parameters $\nu_1 = 0.1$ and $\nu_2 = \{0.1, 0.2, 0, 3\}$, *i.e.*, three scenarios of contamination, say, 10% (C 10%), 20% (C 20%) and 30% (C 30%). We set the value of parameters as follows: $\beta_p = (200, 700, 350)^\top$, $\sigma_e = 0.5$ and $\sigma_b^2 = 10$.

Figure 3 shows the simulated curves from the growth-curve logistic model assuming different probability distributions for the random effect term, say, normal, Student's- t with $\nu = 4$ and contaminated normal with $\nu_1 = 0.1$ and $\nu_2 = 0.1$.

From Table 3 we can see that the proposed model is robust even when the level of contamination is high. For quantile $p = 0.75$, naturally, the parameter β_1 tends to increase for higher levels of contamination. As expected, the Sd and RMSE increase when the distribution of the random effects is heavy-tailed.

6 Illustrative examples

In this section, we illustrate the application of our method by analyzing two longitudinal datasets.

6.1 Growth curve: Soybean data

For the first application, we consider the soybean genotype data analyzed previously by Davidian and Giltinan (1995) and Pinheiro and Bates (2000). The experiment consists of measuring (along time) the leaf weight (in grams) as a measure of growth of two soybean genotypes, namely, a commercial variety called Forrest (F) and an

Table 3 Simulation 2: Mean, Bias, Sd and RMSE for the fixed effects β and scale parameter σ_e obtained after fitting the QR-NLME model for quantiles $p = 0.50$ and $p = 0.75$ using four different distribution settings for random effects

Model	Quantile 50%				Quantile 75%				
	β_1	β_2	β_3	σ_e	β_1	β_2	β_3	σ_e	
	(200)	(700)	(350)	(0.5)	(200)	(700)	(350)	(0.5)	
t_4	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
	Sd	1.41	1.13	0.99	0.024	1.47	1.30	1.04	0.024
	RMSE	1.42	1.13	0.99	0.024	4.67	1.36	1.06	0.024
C 10%	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
	Sd	1.38	1.12	0.94	0.024	1.39	1.34	1.08	0.024
	RMSE	1.38	1.13	0.94	0.024	5.21	1.35	1.08	0.024
C 20%	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
	Sd	1.40	1.13	0.95	0.024	1.41	1.24	1.09	0.023
	RMSE	1.40	1.13	0.95	0.024	5.53	1.26	1.10	0.023
C 30%	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
	Sd	1.45	1.02	0.96	0.024	1.61	1.26	1.03	0.022
	RMSE	1.46	1.03	0.97	0.024	6.82	1.27	1.03	0.023

The results are based on 100 simulated samples

experimental strain called Plan Introduction #416937 (P). The samples were taken approximately weekly during 8 to 10 weeks. Plants were planted for three consecutive years (1988, 1989 and 1990) in 16 plots (8 per genotype) and the mean leaf weight of six randomly selected plants was measured.

We use the three parameter logistic model in (10), introducing a random effect term for each parameter and a dichotomic covariate (*gen*) as

$$y_{ij} = \frac{\varphi_{1i}}{1 + \exp(-[t_{ij} - \varphi_{2i}]/\varphi_{3i})} + \epsilon_{ij}, \quad i = 1, \dots, 412, \quad j = 1, \dots, n_i,$$

where, $\varphi_{1i} = \beta_1 + \beta_4 gen_i + b_{1i}$, $\varphi_{2i} = \beta_2 + b_{2i}$ and $\varphi_{3i} = \beta_3 + b_{3i}$. The observed value y_{ij} represents the mean weight of leaves (in grams) from six randomly selected soybean plants in the i -th plot, t_{ij} days after planting; gen_i is a dichotomic variable indicating the genotype of plant i (0=Forrest, 1= Introduction) and ϵ_{ij} is the measurement error term. Moreover, $\beta_p = (\beta_1, \beta_2, \beta_3, \beta_4)^T$ and $\mathbf{b}_i = (b_{1i}, b_{2i}, b_{3i})^T$ are the fixed and random effects vectors respectively.

The matrices \mathbf{A}_i and \mathbf{B}_i are defined as

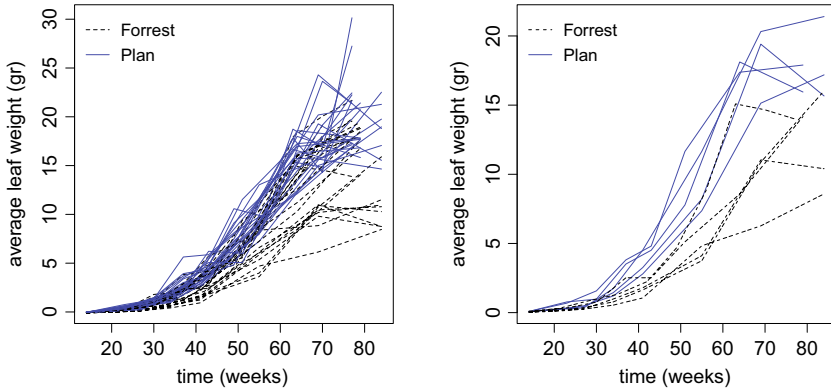


Fig. 4 Soybean data: Leaf weight profiles versus time by genotype (left panel). Ten randomly selected leaf weight profiles versus time (right panel)

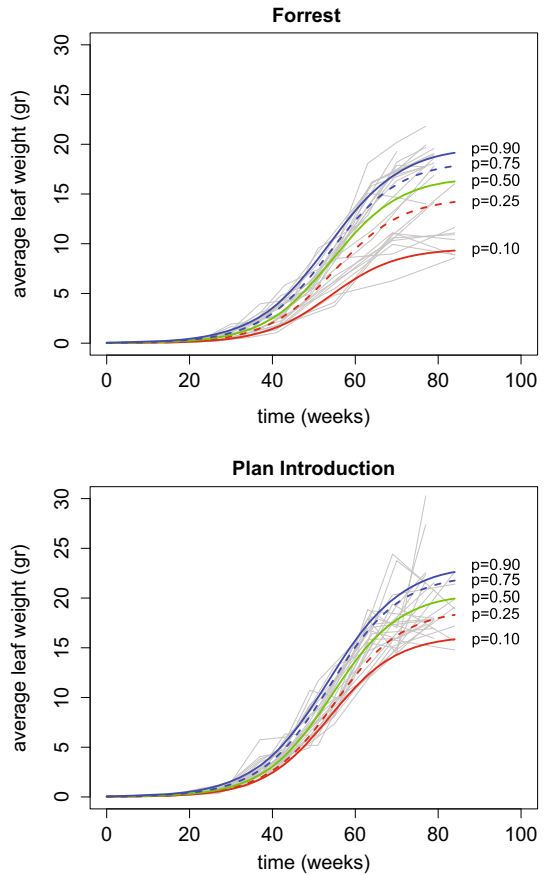
$$\mathbf{A}_i = \begin{pmatrix} 1 & 0 & 0 & gen_i \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{B}_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The interpretation of parameters is as follows: φ_1 is the asymptotic leaf weight, φ_2 is the time at which the leaf reaches half its asymptotic weight and φ_3 is the time elapsed between the leaf reaching half and $0.7311 = 1/(1 + e^{-1})$ of its asymptotic weight. Since the aim of the study is to compare the final (asymptotic) growth of the two kinds of soybeans, the covariate gen_i was incorporated in the first component of the growth function. Therefore, the coefficient β_4 represents the difference (in grams) of the asymptotic leaf weight between the introduction plan and Forrest type (control). Figure 4 shows the leaf weight profiles.

Figure 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 how the extreme quantile estimation functions capture the full data variability, detecting some atypical observations (particularly for the Introduction genotype).

Figure 8 in Appendix A4 shows a summary of the obtained results. It can be seen that the effect of the genotype is significant for all the quantile profiles. Moreover, the difference varies with respect to the conditional quantile being more significant for lower quantiles. Using the information provided by the 95th percentile, we conclude that the soybean plants have a mean leaf weight around 19.35 grams for the Forrest genotype and 23.25 grams for the Introduction genotype. Therefore, the asymptotic difference for the two genotypes is about 4 grams. Finally, it is important to stress that the convergence of the fixed effect estimates and variance components was analyzed using graphical criteria as shown in Figure 10 (Appendix A4).

Fig. 5 Soybean data: Fitted quantile regression for several quantiles



6.2 HIV viral load study

The dataset analyzed in this section comes from a clinical trial (ACTG 315), studied in previous works by Wu (2002) and Lachos et al. (2013). In this study, the HIV viral load of 46 HIV-1 infected patients under antiretroviral treatment (protease inhibitor and reverse transcriptase inhibitor drugs) is analyzed. The viral load and some other covariates were measured several times after the start of treatment. Wu (2002) found that the only significant covariate for modeling the virus load was CD4. Figure 6 shows the profile of viral load in log10 scale and CD4 cell count/100 per mm^3 versus time (in days/100) for six randomly selected patients. As can be seen, an inverse relationship exists between the viral load and the CD4 cell count, *i.e.*, high CD4 cell count leads to lower levels of viral load. This is because the CD4 cells (also called T-cells) alert the immune system in case of invasion by viruses and/or bacteria. Consequently, a lower CD4 count means a weaker immune system.

In order to fit the ACTG 315 data, we propose a bi-phasic nonlinear model considered by Wu (2002) and also used by Lachos et al. (2013). The proposed NLME model

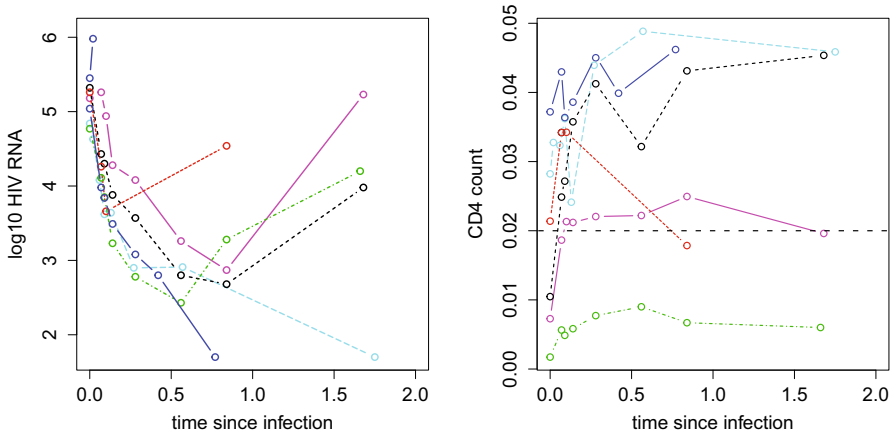


Fig. 6 ACTG 315 data. Profiles of viral load (response) in log₁₀ scale and CD4 cell count (in cells/100mm³) for six randomly selected patients

is given by,

$$y_{ij} = \log_{10} \left(e^{(\varphi_{1i} - \varphi_{2i}t_{ij})} + e^{(\varphi_{3i} - \varphi_{4i}t_{ij})} \right) + \epsilon_{ij}, \tag{11}$$

for $i = 1, \dots, 46, j = 1, \dots, n_i$ with $\varphi_{1i} = \beta_1 + b_{1i}, \varphi_{2i} = \beta_2 + b_{2i}, \varphi_{3i} = \beta_3 + b_{3i}, \varphi_{4i} = \beta_4 + \beta_5 CD4_{ij} + b_{4i}$, where the observed value y_{ij} represents the log-10 transformation of the viral load for the i -th patient at time j , $CD4_{ij}$ is the CD4 cell count (in cells/100mm³) for the i -th patient at time j and ϵ_{ij} is the measurement error term. As in the previous case, $\beta_p = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^T$ and $\mathbf{b}_i = (b_{1i}, b_{2i}, b_{3i}, b_{4i})^T$ denote the fixed and random effects vectors respectively, and $\mathbf{CD4}_i = (CD4_{i1}, \dots, CD4_{in_i})^T$. The matrices \mathbf{A}_i and \mathbf{B}_i are defined as

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

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 latent or long-lived infected cells if therapy was successful, respectively. For more details about the model in (11) see Grossman et al. (1999) and Perelson et al. (1997).

Figure 7 shows the fitted regression lines for quantiles 0.10, 0.25, 0.50, 0.75 and 0.90 for the ACTG 315 data. In order to plot this, first, we have fixed the CD4 covariate using the predicted sequence from a linear regression (including a quadratic term) for explaining the CD4 cell count over time. It can be seen how the quantile estimated functions follow the data behavior and make it easy to estimate a specific viral load quantile at any time of the experiment. Extreme quantile functions bound most of the observed profiles and evidence possible influential observations.

The results after fitting the QR-NLME model over the grid of quantiles $p = \{0.05, 0.10, \dots, 0.95\}$ are shown in Fig. 9 in Appendix A4. The first phase viral decay

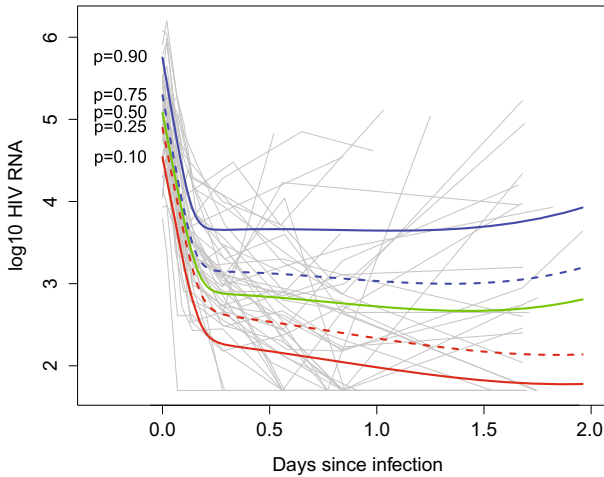


Fig. 7 ACTG 315 data: Fitted quantile regression functions

rate is positive and its effect tends to increase proportionally along quantiles. Moreover, the second phase viral decay rate is positively correlated with the CD4 count and therefore with the duration of therapy. Consequently, more days of treatment imply a higher CD4 cell count and therefore a higher second phase viral decay. The CD4 cell process for this model has 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 its effect becomes constant for greater quantiles. As in the previous case, the convergence of estimates for all parameters was also assessed using the graphical criteria.

7 Conclusions

In this paper, we investigate quantile regression under nonlinear mixed effects models from a likelihood-based perspective. The AL distribution and SAEM algorithm are combined efficiently to propose an exact ML estimation method. 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, this paper is the first attempt at exact ML estimation in the context of QR-NLME models. The methods developed can be readily implemented in R through the package `qrNLMM()`, making our approach quite powerful and accessible to practitioners. We apply our method to two datasets from longitudinal studies, obtaining interesting results from the point of view of quantile estimation. Moreover, in the two considered applications, similar conclusions to the previous analyses of these datasets were obtained.

As was observed by a referee, Mu & He (2007) (see also Liu and Bottai 2009) noted that nonlinear QR models can be modeled using a monotone transformation over the QR linear models. Although this approach is very interesting and easy to understand,

our problem considers not only a nonlinear behavior of the response but also on the random effects. In addition, we consider the assumption of the AL distribution to estimate the model parameters. In this context, an in-depth study of the properties of parameter estimation of fixed effects within the context of mixed effect models seems to be an interesting topic for future research.

Finally, there are a large number of possible extensions of the current work. For modeling both skewness and heavy tails in the random effects, the use of scale mixtures of skew-normal (SMSN) distributions (Lachos et al. 2010) is a feasible choice. Also, HIV viral loads studies include covariates (CD4 cell counts) that often come with substantial measurement errors (Wu 2002). How to incorporate measurement errors 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 paper, but certainly is an interesting topic for future research.

Acknowledgements We thank the Editor and two anonymous referees whose constructive comments and suggestions led to an improved presentation of the paper. The research of C. Galarza was supported by Grant 2015/17110-9 from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP-Brazil). L. M. Castro acknowledges Grant Fondecyt 1170258 from the Chilean government. The research of F. Louzada was supported by Grant 305351/2013-3 from from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil) and by Grant 2013/07375-0 from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP-Brazil).

Appendix

A1: 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 the algorithm to the global maxima solution. Without considering the random effect term, *i.e.*, $\mathbf{b}_i = \mathbf{0}$, let $\mathbf{y}_i \sim \text{AL}(\boldsymbol{\eta}(\boldsymbol{\beta}_p, \mathbf{0}), \sigma, p)$. Next, considering the ML estimates for $\boldsymbol{\beta}_p$ and σ as defined in Yu and Zhang (2005) for this model, we follow the steps below for the QR-LME model implementation:

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

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

2. Using the initial value of $\widehat{\boldsymbol{\beta}}_p^{(0)}$ 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 $\boldsymbol{\Psi}^{(0)}$.

A2: Computing the conditional expectations

Due to the independence between $u_{ij} | y_{ij}, \mathbf{b}_i$ and $u_{ik} | y_{ik}, \mathbf{b}_i$, for all $j, k = 1, 2, \dots, n_i$ and $j \neq k$, we can write $\mathbf{u}_i | \mathbf{y}_i, \mathbf{b}_i = [u_{i1} | y_{i1}, \mathbf{b}_i \ u_{i2} | y_{i2}, \mathbf{b}_i \ \dots \ u_{in_i} | y_{in_i}, \mathbf{b}_i]^\top$. 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:

$$\mathcal{E}(\mathbf{u}_i) = [\mathcal{E}(u_{i1}) \ \mathcal{E}(u_{i1}) \ \dots \ \mathcal{E}(u_{in_i})]^\top, \tag{12}$$

$$\mathcal{E}(\mathbf{D}_i^{-1}) = \begin{bmatrix} \mathcal{E}(u_{i1}^{-1}) & 0 & \dots & 0 \\ 0 & \mathcal{E}(u_{i2}^{-1}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathcal{E}(u_{in_i}^{-1}) \end{bmatrix}. \tag{13}$$

We have $u_{ij} | y_{ij}, \mathbf{b}_i \sim \text{GIG}(\frac{1}{2}, \chi_{ij}, \psi)$, where χ_{ij} and ψ are defined in Sect. 3.2. Then, using (2), 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 the k -th iteration of the algorithm and for the ℓ -th Monte Carlo realization, we can compute $\mathcal{E}(\mathbf{u}_i)^{(\ell,k)}$ and $\mathcal{E}[\mathbf{D}_i^{-1}]^{(\ell,k)}$ using equations (12)-(13) where

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

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

A3: The empirical information matrix

Using (5), the complete log-likelihood function can be rewritten as

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

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

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

with \mathbf{J}_i as defined in Sect. 3.2. and

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

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

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

Next, taking derivatives with respect to a specific α_j from α based on the chain rule, we have

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

where, using the fact that $\text{tr}\{\mathbf{ABCD}\} = (\text{vec}(\mathbf{A}^\top))^\top (\mathbf{D}^\top \otimes \mathbf{B})(\text{vec}(\mathbf{C}))$, (14) can be rewritten as

$$\frac{\partial \ell_{ci}(\theta)}{\partial \alpha_j} = (\text{vec}(\frac{\partial \Psi}{\partial \alpha_j}^\top))^\top \frac{1}{2} (\Psi^{-1} \otimes \Psi^{-1})(\text{vec}(\mathbf{b}_i \mathbf{b}_i^\top - \Psi)). \quad (14)$$

Let \mathcal{D}_q be the elimination matrix (Lavielle 2014), which transforms the vectorized Ψ (written as $\text{vec}(\Psi)$) into its half-vectorized form $\text{vech}(\Psi)$, so that $\mathcal{D}_q \text{vec}(\Psi) = \text{vech}(\Psi)$. Using the fact that for all $j = 1, \dots, \frac{1}{2}q(q+1)$, the vector $(\text{vec}(\frac{\partial \Psi}{\partial \alpha_j}^\top))^\top$ corresponds to the j -th row of the elimination matrix \mathcal{D}_q , we can generalize the derivative in (14) for the vector of parameters α as

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

Finally, at each iteration, we can compute the empirical information matrix by approximating the score for the observed log-likelihood by the stochastic approximation given in (9).

A4: Figures

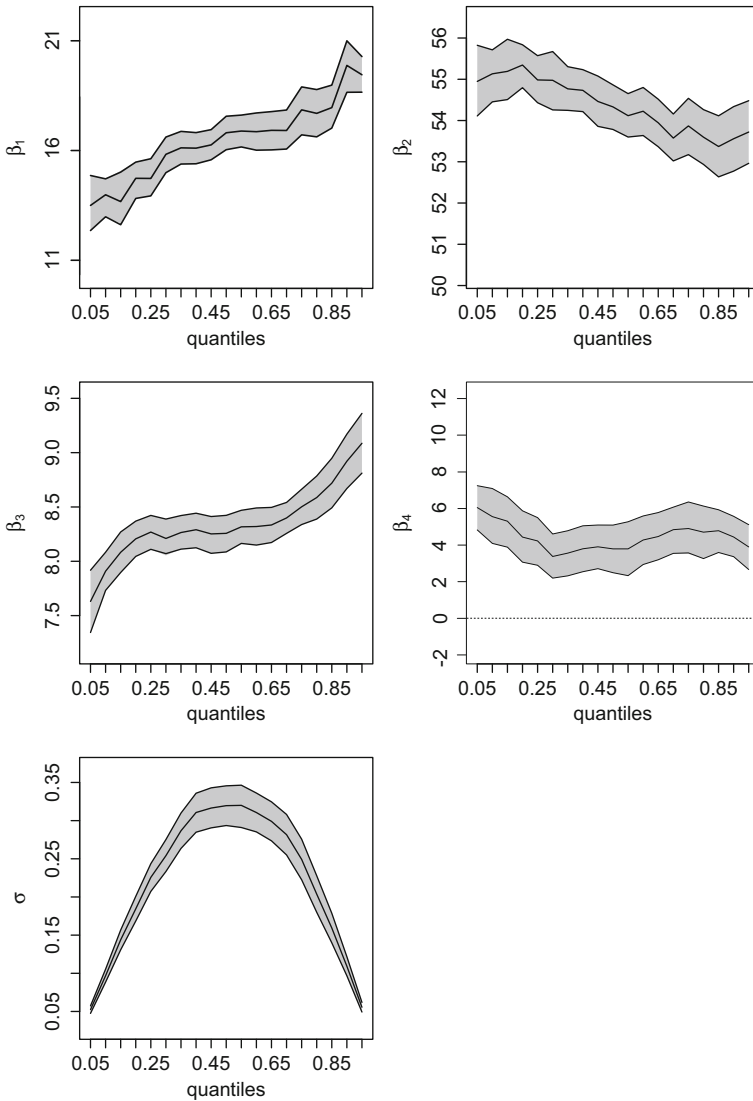


Fig. 8 Soybean data: Point estimates (center solid line) and 95% confidence intervals for model parameters after fitting the QR-NLME model. The interpolated curves are spline-smoothed

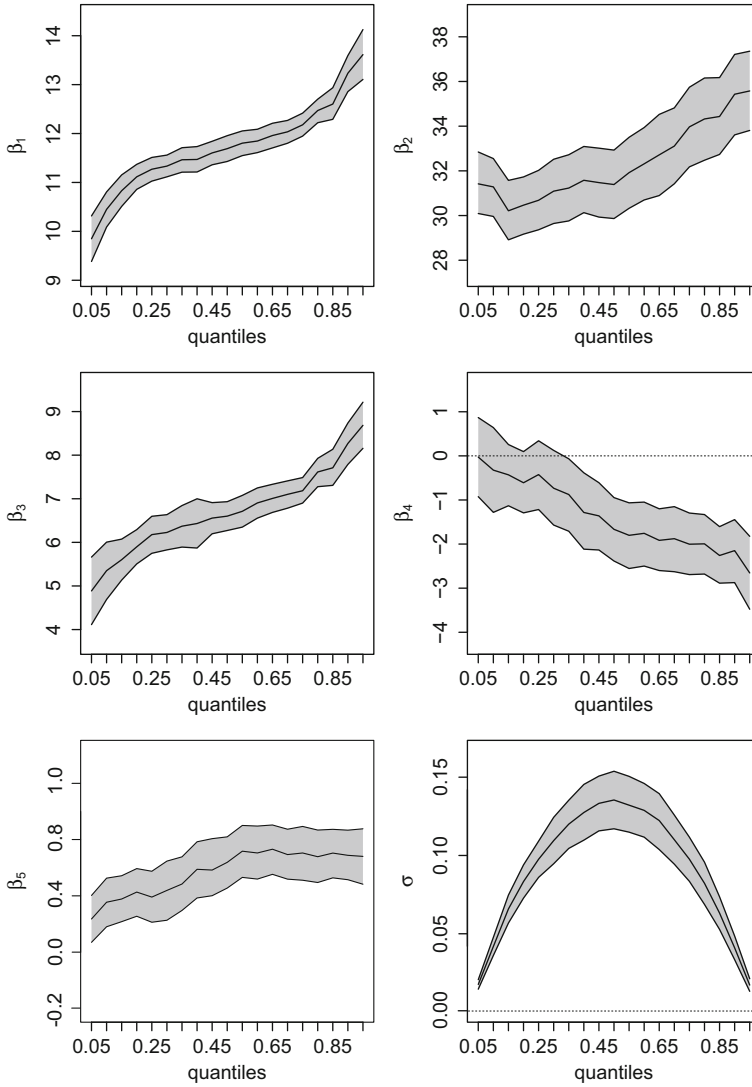


Fig. 9 ACTG 315 data: Point estimates (center solid line) and 95% confidence intervals for model parameters after fitting the QR-NLME model. The interpolated curves are spline-smoothed

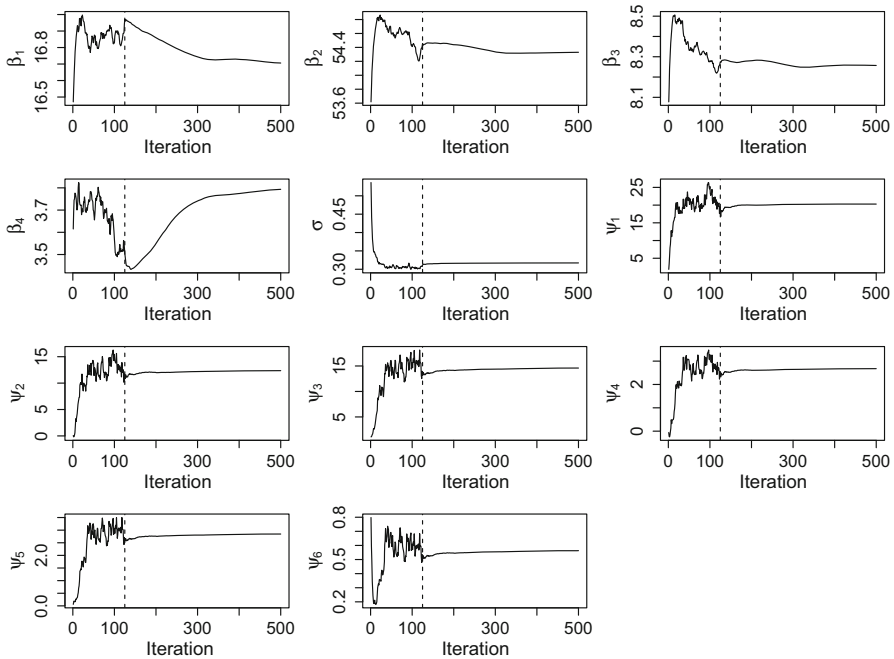


Fig. 10 Graphical summary for the convergence of the fixed effect estimates, variance components of the random effects, and nuisance parameters performing a median regression ($p = 0.50$) for the soybean data. The vertical dashed line delimits the beginning of the almost sure convergence as defined by the cut-point parameter $c = 0.25$

A5: Sample output from R package qrnLMM()

 Quantile Regression for Nonlinear Mixed Model

Quantile = 0.5
 Subjects = 48 ; Observations = 412

- Nonlinear function

```
nlmodel = 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	18.58381	0.55089	33.73433	0
beta 2	54.26468	0.28691	189.13608	0

```
beta 3  8.23245    0.08848  93.04351    0
```

```
sigma = 0.31615
```

```
Random effects Variance-Covariance Matrix matrix
```

```
      b1      b2      b3
b1 23.75098 11.43940 2.58884
b2 11.43940 14.91302 2.86235
b3  2.58884  2.86235 0.55971
```

```
-----
Model selection criteria
-----
```

```
      Loglik      AIC      BIC      HQ
Value -623.899 1267.798 1308.009 1283.704
```

```
-----
Details
-----
```

```
Convergence reached? = TRUE
Iterations = 694 / 1000
Criteria = 8e-05
MC sample = 20
Cut point = 0.25
Processing time = 54.04346 mins
```

References

- Aghamohammadi A, Mohammadi S (2017) Bayesian analysis of penalized quantile regression for longitudinal data. *Stat Pap* 58(4):1035–1053
- Allasonnière S, Kuhn E, Trouvé A (2010) Construction of Bayesian deformable models via a stochastic approximation algorithm: a convergence study. *Bernoulli* 16(3):641–678
- Andriyana Y, Gijbels I, Verhasselt A (2016) Quantile regression in varying-coefficients models: non-crossing quantile curves and heteroscedasticity. *Stat Pap*. <https://doi.org/10.1007/s00362-016-0847-7>
- Barndorff-Nielsen OE, Shephard N (2001) Non-gaussian ornstein-uhlenbeck-based models and some of their uses in financial economics. *J R Stat Soc Ser B* 63(2):167–241
- Booth JG, Hobert JP (1999) Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *J R Stat Soc Ser B* 61(1):265–285
- Davidian M, Giltinan D (1995) *Nonlinear models for repeated measurement data*. CRC Press, Boca Raton
- Davidian M, Giltinan D (2003) *Nonlinear models for repeated measurement data: an overview and update*. *J Agric Biol Environ Stat* 8(4):387–419
- Delyon B, Lavielle M, Moulines E (1999) Convergence of a stochastic approximation version of the EM algorithm. *Ann Stat* 8:94–128
- Dempster A, Laird N, Rubin D (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B* 39:1–38
- Fu L, Wang Y (2012) Quantile regression for longitudinal data with a working correlation model. *Comput Stat Data Anal* 56(8):2526–2538
- Galarza C, Lachos VH, Cabral C, Castro L (2017) Robust quantile regression using a generalized class of skewed distributions. *Statistics* 6:113–130
- Galvao A (2011) Quantile regression for dynamic panel data with fixed effects. *J Econ* 164(1):142–157
- Galvao A, Montes-Rojas GV (2010) Penalized quantile regression for dynamic panel data. *J Stat Plan Inference* 140(11):3476–3497

- Geraci M, Bottai M (2007) Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics* 8(1):140–154
- Geraci M, Bottai M (2014) Linear quantile mixed models. *Stat Comput* 24(3):461–479
- Grossman Z, Polis M, Feinberg M, Grossman Z, Levi I, Jankelevich S, Yarchoan R, Boon J, de Wolf F, Lange J, Goudsmit J, Dimitrov D, Paul W (1999) Ongoing HIV dissemination during HAART. *Nat Med* 5(10):1099–1104
- Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57(1):97–109
- Huang Y, Dagne G (2011) A Bayesian approach to joint mixed-effects models with a skew-normal distribution and measurement errors in covariates. *Biometrics* 67(1):260–269
- Koenker R (2004) Quantile regression for longitudinal data. *J Multivar Anal* 91(1):74–89
- Koenker R (2005) Quantile regression. Cambridge University Press, New York
- Kozubowski T, Nadarajah S (2010) Multitude of Laplace distributions. *Stat Pap* 51:127–148
- Kuhn E, Lavielle M (2004) Coupling a stochastic approximation version of EM with an MCMC procedure. *ESAIM Probab Stat* 8:115–131
- Kuhn E, Lavielle M (2005) Maximum likelihood estimation in nonlinear mixed effects models. *Comput Stat Data Anal* 49(4):1020–1038
- Lachos VH, Ghosh P, Arellano-Valle RB (2010) Likelihood based inference for skew-normal independent linear mixed models. *Stat Sin* 20(1):303–322
- Lachos VH, Castro LM, Dey DK (2013) Bayesian inference in nonlinear mixed-effects models using normal independent distributions. *Comput Stat Data Anal* 64:237–252
- Lavielle M (2014) Mixed effects models for the population approach. Chapman and Hall/CRC, Boca Raton
- Lipsitz SR, Fitzmaurice GM, Molenberghs G, Zhao LP (1997) Quantile regression methods for longitudinal data with drop-outs: application to CD4 cell counts of patients infected with the human immunodeficiency virus. *J R Stat Soc Ser C* 46(4):463–476
- Liu Y, Bottai M (2009) Mixed-effects models for conditional quantiles with longitudinal data. *Int J Biostat*. <https://doi.org/10.2202/1557-4679.1186>
- Louis TA (1982) Finding the observed information matrix when using the EM algorithm. *J R Stat Soc Ser B* 44(2):226–233
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. *J Chem Phys* 21:1087–1092
- Meza C, Osorio F, De la Cruz R (2012) Estimation in nonlinear mixed-effects models using heavy-tailed distributions. *Stat Comput* 22:121–139
- Mu Y, He X (2007) Power transformation toward a linear regression quantile. *J Am Stat Assoc* 102:269–279
- Perelson AS, Essunger P, Cao Y, Vesanan M, Hurley A, Saksela K, Markowitz M, Ho DD (1997) Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* 387(6629):188–191
- Pinheiro J, Bates D (1995) Approximations to the log-likelihood function in the nonlinear mixed effects model. *J Comput Gr Stat* 4:12–35
- Pinheiro JC, Bates DM (2000) Mixed-effects models in S and S-PLUS. Springer, New York
- R Core Team (2017) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Searle SR, Casella G, McCulloch C (1992) Variance components. Wiley, New York
- Sriram K, Ramamoorthi R, Ghosh P (2013) Posterior consistency of Bayesian quantile regression based on the misspecified asymmetric Laplace distribution. *Bayesian Anal* 8(2):479–504
- Wang J (2012) Bayesian quantile regression for parametric nonlinear mixed effects models. *Stat Methods Appl* 21:279–295
- Wu L (2002) A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to AIDS studies. *J Am Stat Assoc* 97(460):955–964
- Wu L (2010) Mixed effects models for complex data. Chapman & Hall/CRC, Boca Raton
- Yu K, Moyeed R (2001) Bayesian quantile regression. *Stat Probab Lett* 54(4):437–447
- Yu K, Zhang J (2005) A three-parameter asymmetric Laplace distribution and its extension. *Commun Stat Theory Methods* 34(9–10):1867–1879
- Yuan Y, Yin G (2010) Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics* 66(1):105–114