# Supplementary Materials for *Quantile Regression in Linear Mixed Models: A Stochastic Approximation EM approach*

Christian E. Galarza, Victor H. Lachos and Dipankar Bandyopadhyay

Note: The numbers in parentheses inside the text in the material below refer to the equation numbers in the main paper.

## APPENDIX A Some results on SAEM implementation

### A.1 A Gibbs Sampler Algorithm

In order to draw a sample from $f(\mathbf{b}_i,\mathbf{u}_i|\mathbf{y}_i)$ we can use 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

$$\begin{aligned} f(\mathbf{b}_i|\mathbf{y}_i,\mathbf{u}_i) &\propto f(\mathbf{y}_i|\mathbf{b}_i,\mathbf{u}_i)\, f(\mathbf{b}_i), \\ &\propto \phi_{n_i}\left(\mathbf{y}_i|\,\mathbf{X}_i^\top\boldsymbol{\beta}_p+\mathbf{Z}_i\mathbf{b}_i+\vartheta_p\mathbf{u}_i,\sigma\tau_p^2 D(\mathbf{u}_i)\right)\times\phi_q(\mathbf{b}_i|\mathbf{0},\boldsymbol{\Psi}) \end{aligned} \tag{A.1}$$

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*

$$\phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{b},\boldsymbol{\Omega})\phi_q(\mathbf{b}|\mathbf{0},\boldsymbol{\Psi})=\phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta},\boldsymbol{\Sigma})\phi_q(\mathbf{b}|\boldsymbol{\mu}_1(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}),\boldsymbol{\Lambda}) \tag{A.2}$$

*where*

$$\boldsymbol{\mu}_1=\boldsymbol{\Lambda}\mathbf{Z}^T\boldsymbol{\Omega}^{-1},\quad \boldsymbol{\Sigma}=\boldsymbol{\Omega}+\mathbf{Z}\boldsymbol{\Psi}\mathbf{Z}^T,\quad \boldsymbol{\Lambda}=(\boldsymbol{\Psi}^{-1}+\mathbf{Z}^T\boldsymbol{\Omega}^{-1}\mathbf{Z})^{-1}. \tag{A.3}$$

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

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

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

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 $\mathbf{u}_i|\mathbf{y}_i,\mathbf{b}_i = \begin{bmatrix} u_{i1}|y_{i1},\mathbf{b}_i & u_{i2}|y_{i2},\mathbf{b}_i & \cdots & u_{in_i}|y_{in_i},\mathbf{b}_i \end{bmatrix}^\top$ given that $u_{ij}\big|y_{ij},\mathbf{b}_i \perp u_{ik}|y_{ik},\mathbf{b}_i$ 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

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

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

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

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

**(4)** Construct $\mathbf{u}_i^{(1,k)}|\mathbf{y}_i,\mathbf{b}_i^{(1,k)}$ as $\begin{bmatrix} u_{i1}^{(1,k)}|y_{i1},\mathbf{b}_i^{(1,k)} & u_{i2}^{(1,k)}|y_{i2},\mathbf{b}_i^{(1,k)} & \cdots & u_{in_i}^{(1,k)}|y_{in_i},\mathbf{b}_i^{(1,k)} \end{bmatrix}^\top$

**(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 $\mathbf{b}_i,\mathbf{u}_i|\theta^{(k)},\mathbf{y}_i$.

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.

## A.2 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 $\sigma$ as defined in Yu and 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

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

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.3 Computing the conditional expectations

Due the independence between $u_{ij}\big|y_{ij},\mathbf{b}_i$ 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 $\mathscr{E}(\mathbf{u}_i)$ and $\mathscr{E}(\mathbf{D}_i^{-1})$ in the following way. Using matrix expectation properties, we define these expectations as

$$\mathscr{E}(\mathbf{u}_i) = [\mathscr{E}(u_{i1})\ \mathscr{E}(u_{i1})\ \cdots\ \mathscr{E}(u_{in_i})]^\top \tag{A.4}$$

and

$$\mathscr{E}(\mathbf{D}_i^{-1}) = \mathrm{diag}(\mathscr{E}(\mathbf{u}_i^{-1})) = \begin{bmatrix} \mathscr{E}(u_{i1}^{-1}) & 0 & \ldots & 0 \\ 0 & \mathscr{E}(u_{i2}^{-1}) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \mathscr{E}(u_{in_i}^{-1}) \end{bmatrix}. \tag{A.5}$$

We already have $u_{ij}|y_{ij},\mathbf{b}_i \sim GIG(\frac{1}{2},\chi_{ij},\psi)$, where $\chi_{ij}$ and $\psi$ are defined in (14). Then, using (5), we compute the moments involved in the equations above as $\mathscr{E}(u_{ij}) = \frac{\chi_{ij}}{\psi}(1+\frac{1}{\chi_{ij}\psi})$ and $\mathscr{E}(u_{ij}^{-1}) = \frac{\psi}{\chi_{ij}}$. Thus, for iteration $k$ of the algorithm and for the $\ell$th Monte Carlo realization, we can compute $\mathscr{E}(\mathbf{u}_i)^{(\ell,k)}$ and $\mathscr{E}[\mathbf{D}_i^{-1}]^{(\ell,k)}$ using equations (A.4)-(A.5) where

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

## A.4 The empirical information matrix

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

$$\ell_{ci}(\boldsymbol{\theta}) \;=\; -\frac{3}{2}n_i \log\sigma - \frac{1}{2\sigma\tau_p^2}\zeta_i^\top \mathbf{D}_i^{-1}\zeta_i - \frac{1}{2}\log\big|\boldsymbol{\Psi}\big| - \frac{1}{2}\mathbf{b}_i^\top \boldsymbol{\Psi}^{-1}\mathbf{b}_i - \frac{1}{\sigma}\mathbf{u}_i^\top \mathbf{1}_{n_i} \tag{A.6}$$

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

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

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

$$\begin{aligned}\frac{\partial \ell_{ci}(\boldsymbol{\theta})}{\partial \alpha_j} &= \frac{\partial \boldsymbol{\Psi}}{\partial \alpha_j}\frac{\partial \ell_{ci}(\boldsymbol{\theta})}{\partial \boldsymbol{\Psi}} \\ &= \frac{\partial \boldsymbol{\Psi}}{\partial \alpha_j}\frac{1}{2}\mathrm{tr}\{\boldsymbol{\Psi}^{-1}(\mathbf{b}_i\mathbf{b}_i^\top - \boldsymbol{\Psi})\boldsymbol{\Psi}^{-1}\}. \qquad \text{(A.7)}\end{aligned}$$

where, using the fact that $\mathrm{tr}\{\mathbf{ABCD}\} = (\mathrm{vec}(\mathbf{A}^\top))^\top(\mathbf{D}^\top \otimes \mathbf{B})(\mathrm{vec}(\mathbf{C}))$, (A.7) can be rewritten as

$$\frac{\partial \ell_{ci}(\boldsymbol{\theta})}{\partial \alpha_j} = (\mathrm{vec}(\frac{\partial \boldsymbol{\Psi}}{\partial \alpha_j}^\top))^\top \frac{1}{2}(\boldsymbol{\Psi}^{-1} \otimes \boldsymbol{\Psi}^{-1})(\mathrm{vec}(\mathbf{b}_i\mathbf{b}_i^\top - \boldsymbol{\Psi})). \qquad \text{(A.8)}$$

Let $\mathscr{D}_q$ be the elimination matrix (Lavielle, 2014) that transforms the vectorized $\boldsymbol{\Psi}$ (written as $\mathrm{vec}(\boldsymbol{\Psi})$) into its half-vectorized form vech($\boldsymbol{\Psi}$), such that $\mathscr{D}_q\mathrm{vec}(\boldsymbol{\Psi}) = \mathrm{vech}(\boldsymbol{\Psi})$. Using the fact that for all $j = 1, \ldots, \frac{1}{2}q(q+1)$, the vector $(\mathrm{vec}(\frac{\partial \boldsymbol{\Psi}}{\partial \alpha_j})^\top)^\top$ corresponds to the $j$th row of the elimination matrix $\mathscr{D}_q$, we can generalize the derivative in (A.8) for the vector of parameters $\boldsymbol{\alpha}$ as

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

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).

# APPENDIX B Figures

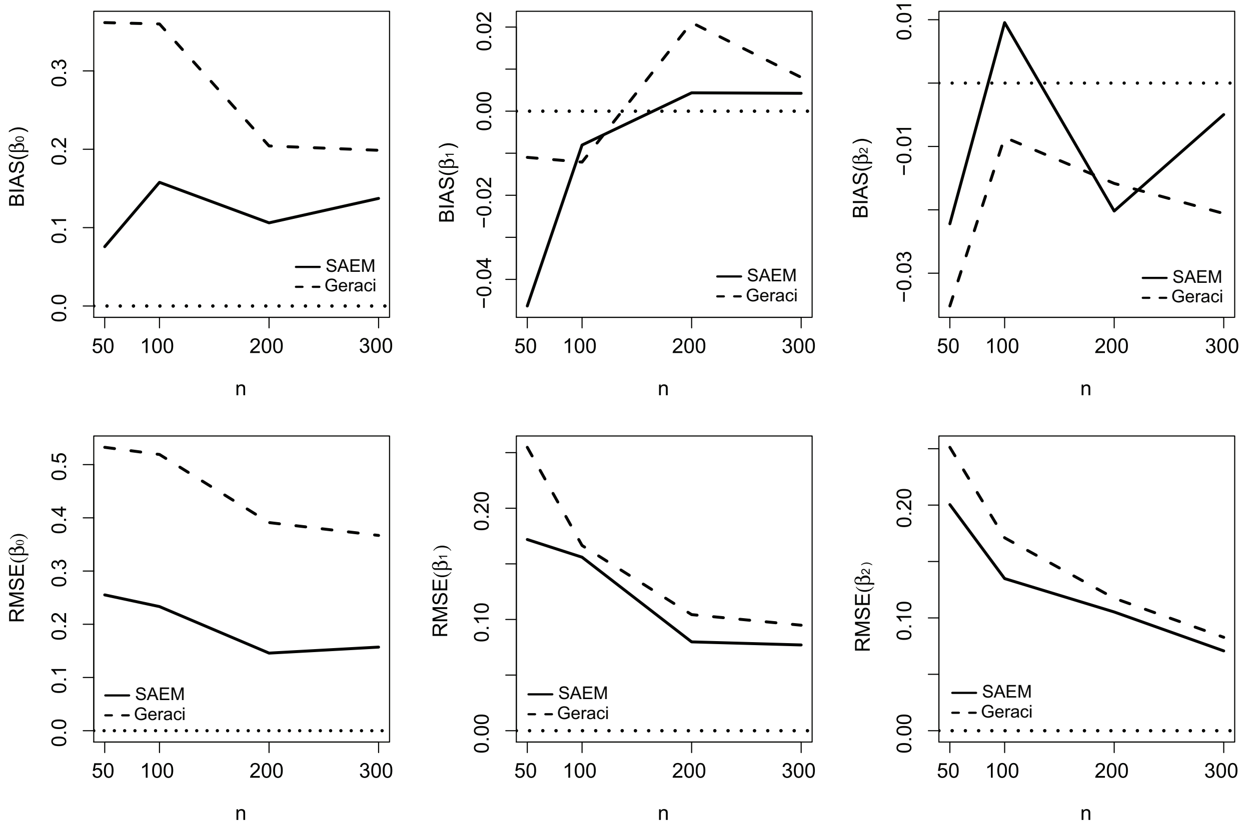


Figure B.1: Comparison of the Bias (upper row) and RMSE (lower row) at the 95-th quantile from fitting the QR-LMM and the Geraci (2014) model for the fixed effects $\beta_0$, $\beta_1$ and $\beta_2$.

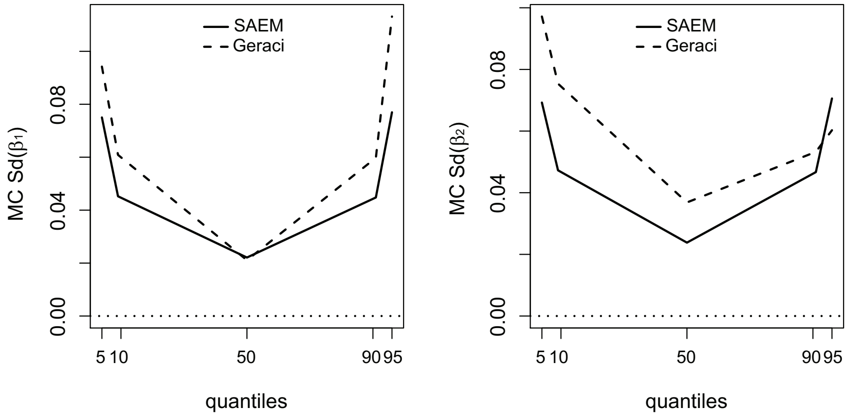


Figure B.2: Comparison of the Monte Carlo standard deviation for the estimatives of $\beta_1$ and $\beta_2$ obtained by the SAEM procedure and the Geraci (2014) algorithm for the set of quantiles 5, 10, 50, 90 and 95.

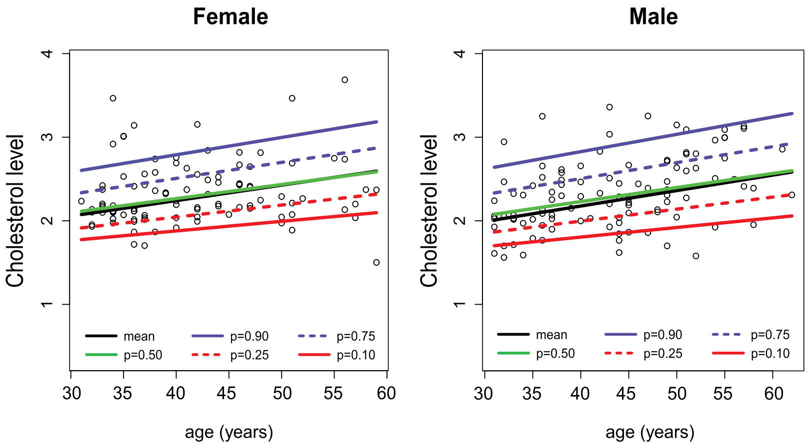


Figure B.3: Fitted mean regression overlayed with five different quantile regression lines for the Cholesterol data, by gender.

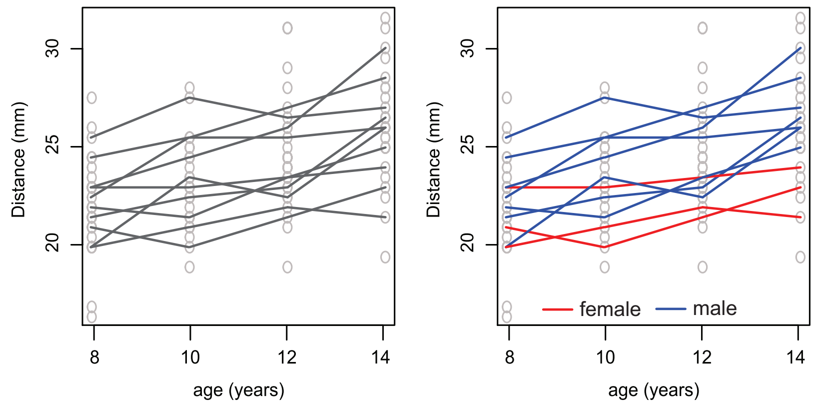


Figure B.4: Orthodontic distance growth data: Individual profiles for 10 random children (Panel a); Individual profiles for the same children, by gender (Panel b).

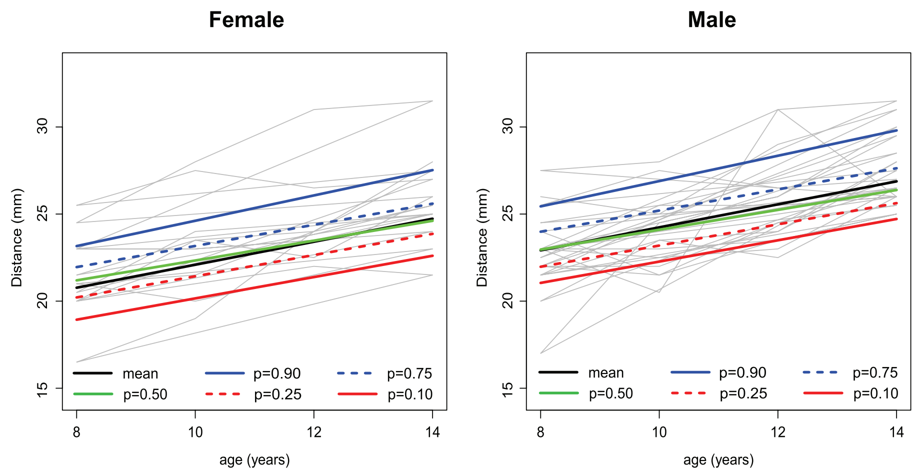


Figure B.5: Fitted mean regression overlayed with five different quantile regression lines for the Orthodontic distance growth data, by gender.

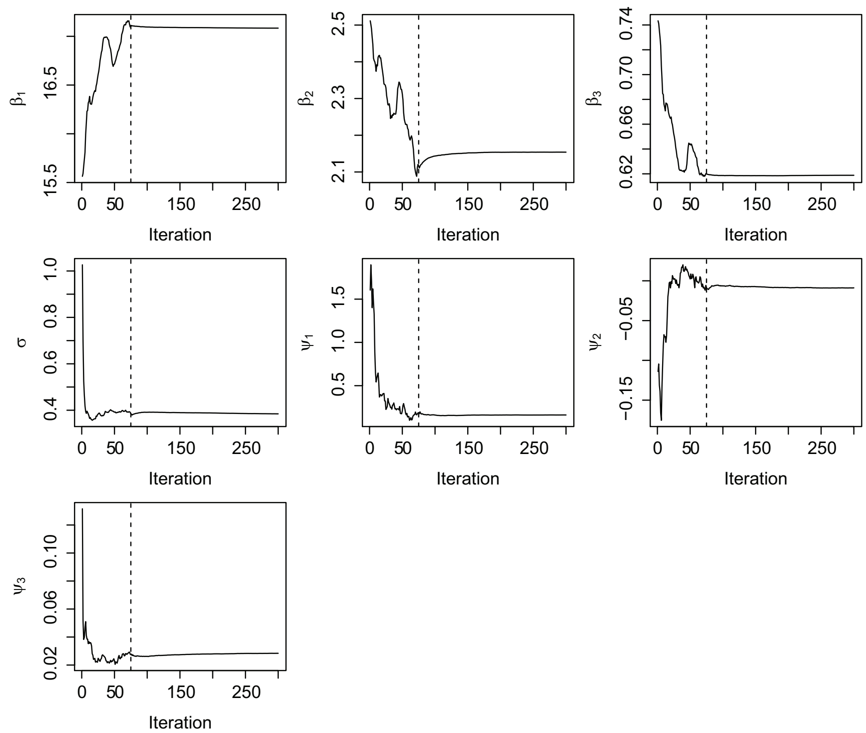


Figure B.6: Graphical summary of convergence for the fixed effect parameters, variance components of the random effects, and nuisance parameters, generated from the `qrLMM` package for the orthodontic distance growth data. The vertical dashed line delimits the beginning of the almost sure convergence,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

# APPENDIX C 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
         z1       z2
z1  0.16106 -0.00887
z2 -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
```

# References


Geraci, M. (2014). Linear quantile mixed models: The lqmm package for laplace quantile regression. *Journal of Statistical Software 57*(13), 1–29.

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

Yu, K. and J. Zhang (2005). A three-parameter asymmetric Laplace distribution and its extension. *Communications in Statistics - Theory and Methods 34*(9-10), 1867–1879.