Linear Quantile Mixed Models: The lqmm Package for Laplace Quantile Regression

Marco Geraci
University College London

Abstract

Inference in quantile analysis has received considerable attention in the recent years. Linear quantile mixed models (Geraci and Bottai 2013) represent a flexible statistical tool to analyse data from sampling designs such as multilevel, spatial, panel or longitudinal which induce some form of clustering. In this paper, I will show how to estimate conditional quantile functions with random effects using the R package **lqmm**. Modelling, estimation and inference are discussed in detail using a real data example. A thorough description of the optimization algorithms is also provided.

Keywords: asymmetric Laplace distribution, non-smooth optimization, random effects.

1. Introduction

In classical statistics, a common assumption is that sample observations are drawn independently from the same population. However, dependent data arise in many studies. For example, clinical observations, such as blood pressure or insulin measurements, taken repeatedly on the same individuals are likely to be more similar than observations from different individuals; environmental measurements (e.g., air pollution or rainfall measurements) that are taken in the same geographic area will show substantial degree of spatial correlation. Groups of dependent observations are commonly called clusters.

Mixed-effects models (or mixed models) (Pinheiro and Bates 2000; Demidenko 2004) are highly popular and flexible regression models used to analyse the conditional mean of clustered outcome variables. The extent of applications of mixed models is vast, ranging from medical studies, in which responses to an exposure or a treatment show a strong degree of heterogeneity between subjects due to unobserved genetic factors, to agricultural field trials, environmental and wildlife ecology studies, to mention a few.

Mixed models are based on the assumption that predictors affect the conditional distribution of the outcome only through its location parameter (i.e., the mean). Empirical evidence shows that this assumption is inappropriate in a number of real-world applications. For example, the negative effects of maternal smoking during pregnancy or lack of prenatal care have been amply documented. In particular, these factors have been shown to decrease the average weight of infants at birth. However, infants who rank lower in the distribution (i.e., low birthweight infants) are affected by smoking and lack of prenatal care at a greater extent than average-weighting infants, and the latter at a greater extent than those who rank higher in the distribution (Abrevaya 2001; Koenker and Hallock 2001; Geraci 2013a). In general,

individuals who rank differently according to some outcome variable (e.g., blood pressure, body mass index, size of a tumour) might be affected by risk factors (e.g., age, gender, socio-economic status) to a different extent or even in opposite ways. Quantile regression (QR) is a statistical tool that extends regression for the mean to the analysis of the entire conditional distribution of the outcome variable. Therefore, location, scale and shape of the distribution can be examined through the analysis of conditional quantile models to provide a complete picture of the distributional effects.

The application of QR methods to clustered data is an emerging area of research in statistics. There have been several proposals of QR for dependent data, including Lipsitz, Fitzmaurice, Molenberghs, and Zhao (1997), Koenker (2004), Geraci and Bottai (2007), Reich, Bondell, and Wang (2010), and Canay (2011). Recently, Geraci and Bottai (2013) developed a class of models, called linear quantile mixed models (LQMMs), which extends quantile regression models with random intercepts (Geraci 2005; Geraci and Bottai 2007) to include random slopes, and introduced new computational approaches. These are based on the asymmetric Laplace (AL) likelihood (Hinkley and Revankar 1977) which has a well-known relationship with the L_1-norm objective function described by Koenker and Bassett (1978).

Briefly, consider a sample of observations $\left(\mathbf{x}_i^\top, y_i\right)$, $i = 1, \ldots, M$, drawn independently from a population with continuous distribution function $F_{Y_i|\mathbf{x}_i}$. The latter is assumed to be unknown, with quantile function given by its inverse $Q_{Y_i|\mathbf{x}_i} \equiv F^{-1}_{Y_i|\mathbf{x}_i}$. In linear QR problems, the goal is to estimate models of the type $Q_{Y_i|\mathbf{x}_i}(\tau) = \mathbf{x}_i^\top \boldsymbol{\beta}^{(\tau)}$, where τ, $0 < \tau < 1$, denotes the quantile level of interest. The τth regression quantile is defined as any solution of (Koenker and Bassett 1978)

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^{M} \rho_\tau \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta} \right), \tag{1}$$

where $\rho_\tau(v) = \tau \max(v, 0) + (1-\tau)\max(-v, 0)$ is the asymmetrically weighted L_1 loss function. Nonlinear QR is discussed by Koenker (2005).

Over the past few years, a number of QR methods based on the AL distribution appeared in the literature (Koenker and Machado 1999; Yu and Moyeed 2001; Geraci and Bottai 2007; Lee and Neocleous 2010; Farcomeni 2012; Wang 2012). A continuous random variable $w \in \mathbb{R}$ is said to follow an AL density with parameters (μ, σ, τ), $w \sim \text{AL}(\mu, \sigma, \tau)$, if its density can be expressed as

$$p(w|\mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{ -\frac{1}{\sigma} \rho_\tau (w - \mu) \right\},$$

where $\mu \in \mathbb{R}$ is the location parameter, $\sigma \in \mathbb{R}_+$ is the scale parameter, and $0 < \tau < 1$ is the skewness parameter. Mean and variance of this distribution are given by $\mathsf{E}(w) = \mu + \sigma \frac{1-2\tau}{\tau(1-\tau)}$ and $\mathsf{VAR}(w) = \frac{\sigma^2(1-2\tau+2\tau^2)}{(1-\tau)^2\tau^2}$ (Yu and Zhang 2005). Note that $\rho_\tau(\cdot)$ is the loss function introduced in (1). It is easy to verify that the location parameter μ is the τth quantile of w, i.e., $\Pr(w \leq \mu) = \tau$. Therefore, for a fixed τ, it is convenient to estimate the τth regression quantile from the model

$$y_i = \mu_i^{(\tau)} + \epsilon_i^{(\tau)}, \qquad i = 1, \ldots, M, \tag{2}$$

where $\mu_i^{(\tau)} = \mathbf{x}_i^\top \boldsymbol{\beta}^{(\tau)}$ and $\epsilon_i^{(\tau)} \sim \text{AL}(0, \sigma, \tau)$. The AL assumption is ancillary as it is not assumed that $F_{Y_i|\mathbf{x}_i}$ is truly AL. Computationally, however, the maximum likelihood estimate

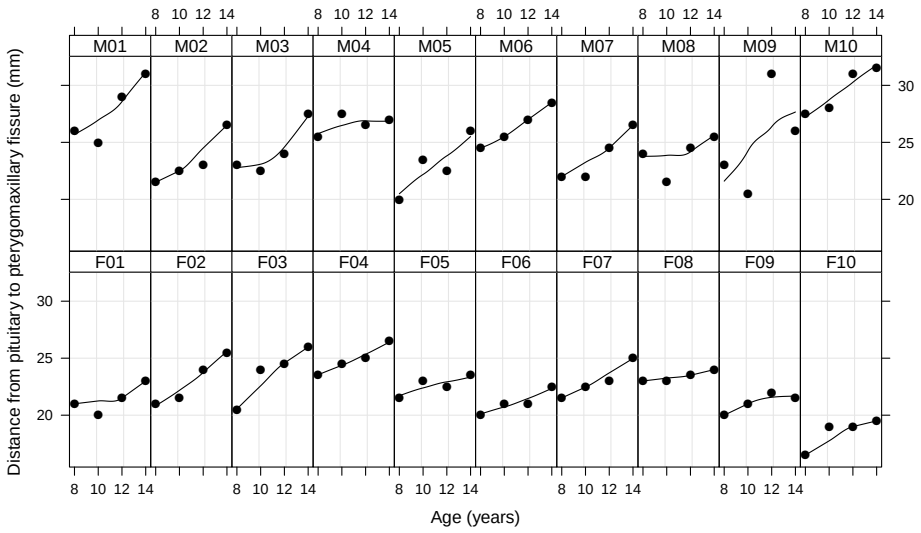

Figure 1: Trellis plot of pituitary-pterygomaxillary fissure distance conditional on age in 10 girls (subjects F1-F10) and 10 boys (subjects M01-M10). A loess smooth is shown in each panel.

(MLE) of $\boldsymbol{\beta}^{(\tau)}$ from Equation 2 is equivalent to the solution of (1). Such computational equivalence provided Geraci (2005) with a (quasi-)likelihood framework within which estimating the conditional quantiles of longitudinal outcomes.

This paper's focus is on the package **lqmm** (Geraci 2013b) for the R statistical programming environment (R Core Team 2013). In the next section, the dataset that will be used to illustrate LQMM methods is briefly described. The models and the estimation algorithms are introduced in Sections 3 and 4, respectively. Section 5 is dedicated to **lqmm** methods. Short examples and related R code are given throughout to illustrate individual commands. A more extensive data analysis using LQMMs is offered in Section 6. The notation used in this paper, as well as the labeling adopted in the **lqmm** package, follows closely that of Geraci and Bottai (2013). Vectors and matrices are denoted in bold, $\mathbf{I}_k$ denotes the $k \times k$ identity matrix, and $\bigoplus_{i=1}^{n} \mathbf{A}_i$ is the direct sum of the n matrices $\mathbf{A}_i$.

2. Orthodontic growth data

These data collect repeated measurements of the distance between the centre of the pituitary to the pterygomaxillary fissure, two points that are identified on x-ray exposures of the side of the head, in 27 children (16 boys, 11 girls). Measurements were taken by researchers of the University of North Carolina Dental School at four different ages (8, 10, 12, 14 years), giving 108 observations in total, to study growth patterns by sex. The dataset was reported

in Potthoff and Roy (1964) and used for illustration of mixed modelling methods by Pinheiro and Bates (2000). The dataset is available in the package **nlme** (Pinheiro, Bates, DebRoy, Sarkar, and R Core Team 2013) as well as in **lqmm**. I load the former as it provides useful functions for objects of class `groupedData` and then I run the summary of the dataset as follows:

```
R> library("nlme")
R> data("Orthodont")
R> Orthodont$Subject <- as.character(Orthodont$Subject)
R> Orthodont <- update(Orthodont, units = list(x = "(years)", y = "(mm)"),
+ order.groups = F)
R> summary(Orthodont)
```

```
    distance          age          Subject       Sex
 Min.   :16.50   Min.   : 8.0   F01    : 4   Male  :64
 1st Qu.:22.00   1st Qu.: 9.5   F02    : 4   Female:44
 Median :23.75   Median :11.0   F03    : 4
 Mean   :24.02   Mean   :11.0   F04    : 4
 3rd Qu.:26.00   3rd Qu.:12.5   F05    : 4
 Max.   :31.50   Max.   :14.0   F06    : 4
                                (Other):84
```

A Trellis plot (Sarkar 2008) of selected individual temporal trajectories in the pituitary-pterygomaxillary fissure distance is shown in Figure 1. This was obtained with the function `nlme::plot.nfnGroupedData`. To simplify some of the examples, only a subset (i.e., girls) is used (see also Pinheiro and Bates 2000, p. 35). The full dataset is analysed in Section 6.

3. Models for clustered data

3.1. Linear mixed models

Consider clustered data in the form $\left(\mathbf{x}_{ij}^\top, \mathbf{z}_{ij}^\top, y_{ij}\right)$, for $j = 1, \ldots, n_i$ and $i = 1, \ldots, M$, $N = \sum_i n_i$, where $\mathbf{x}_{ij}^\top$ is the jth row of a known $n_i \times p$ matrix $\mathbf{X}_i$, $\mathbf{z}_{ij}^\top$ is the jth row of a known $n_i \times q$ matrix $\mathbf{Z}_i$ and y_{ij} is the jth observation of the ith response vector $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$. Mixed effects models represent a common and well-known class of regression models used to analyse data coming from similar designs. A typical formulation of a linear mixed model (LMM) for clustered data is given by

$$y_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + \mathbf{z}_{ij}^\top \mathbf{u}_i + \epsilon_{ij}, \qquad j = 1, \ldots, n_i,\ i = 1, \ldots, M,$$

where $\boldsymbol{\beta}$ and $\mathbf{u}_i$, $i = 1, \ldots, M$, are, respectively, fixed and random effects associated with p and q model covariates and the response vector $\mathbf{y}_i$ is assumed to follow a multivariate normal distribution characterised by some parameter $\boldsymbol{\theta}$. The dependence among the observations within the i-th cluster is induced by the random effect vector $\mathbf{u}_i$ which is shared by all observations within the same cluster. However, the random effects and the within-cluster

errors are assumed to be independent for different clusters and to be mutually independent for the same cluster (Pinheiro and Bates 2000).

In matrix form, the model above can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}, \tag{3}$$

where $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_M)^\top$, $\mathbf{X} = \left[\mathbf{X}_1^\top | \ldots | \mathbf{X}_M^\top\right]^\top$, $\mathbf{Z} = \bigoplus_{i=1}^M \mathbf{Z}_i$, and $\mathbf{u} = (\mathbf{u}_1^\top, \ldots, \mathbf{u}_M^\top)^\top$. From a conditional point of view, the LMM is a location-shift model. That is, the predictors $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ and $\mathcal{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_q\}$, where often $\mathcal{X} \supseteq \mathcal{Z}$, are assumed to shift the conditional expectation $\mathsf{E}(\mathbf{y}|\mathbf{u}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$ only, without affecting the response distribution in any other respect. However, a particular marginal model can be derived from (3) (Lee and Nelder 2004). Suppose $\mathbf{X} = \mathbf{Z}$ and $\mathsf{COV}(\boldsymbol{\epsilon}) = \psi_\epsilon^2 \mathbf{I}_N$. Equation 3 can be rewritten as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}^*$, where $\boldsymbol{\epsilon}^* = \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$, with $\mathsf{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathsf{COV}(\boldsymbol{\epsilon}^*) = \mathbf{X}\mathsf{COV}(\mathbf{u})\mathbf{X}^\top + \psi_\epsilon^2 \mathbf{I}_N$. The within-cluster random term then confers a specific heteroscedastic structure to the model's error term $\boldsymbol{\epsilon}^*$, despite the *iid* assumptions on $\boldsymbol{\epsilon}$.

For example, consider the orthodontic growth data described earlier. Using the package **lme4** (Bates, Maechler, and Bolker 2013), a random-intercept model is fitted by restricted maximum likelihood (REML) to estimate the growth rate in girls. The random effects u_i, $i = 1, \ldots, 11$, are therefore assumed $\mathcal{N}(0, \psi_u^2)$. Note also that the variable `age` is centred at 11 years to remove the correlation between intercept and slope.

```
R> library("lme4")
R> Orthodont$age.c <- Orthodont$age - 11
R> Orthodont.sub <- subset(Orthodont, Orthodont$Sex == "Female")
R> fit.lmer <- lmer(distance ~ age.c + (1|Subject), data = Orthodont.sub)
R> fit.lmer

Linear mixed model fit by REML
Formula: distance ~ age.c + (1 | Subject)
   Data: Orthodont.sub
   AIC   BIC logLik deviance REMLdev
 149.2 156.4 -70.61      138   141.2
Random effects:
 Groups   Name        Variance Std.Dev.
 Subject  (Intercept) 4.27857  2.06847
 Residual             0.60845  0.78003
Number of obs: 44, groups: Subject, 11

Fixed effects:
            Estimate Std. Error t value
(Intercept) 22.64773    0.63463   35.69
age.c        0.47955    0.05259    9.12


Correlation of Fixed Effects:
      (Intr)
age.c 0.000
```

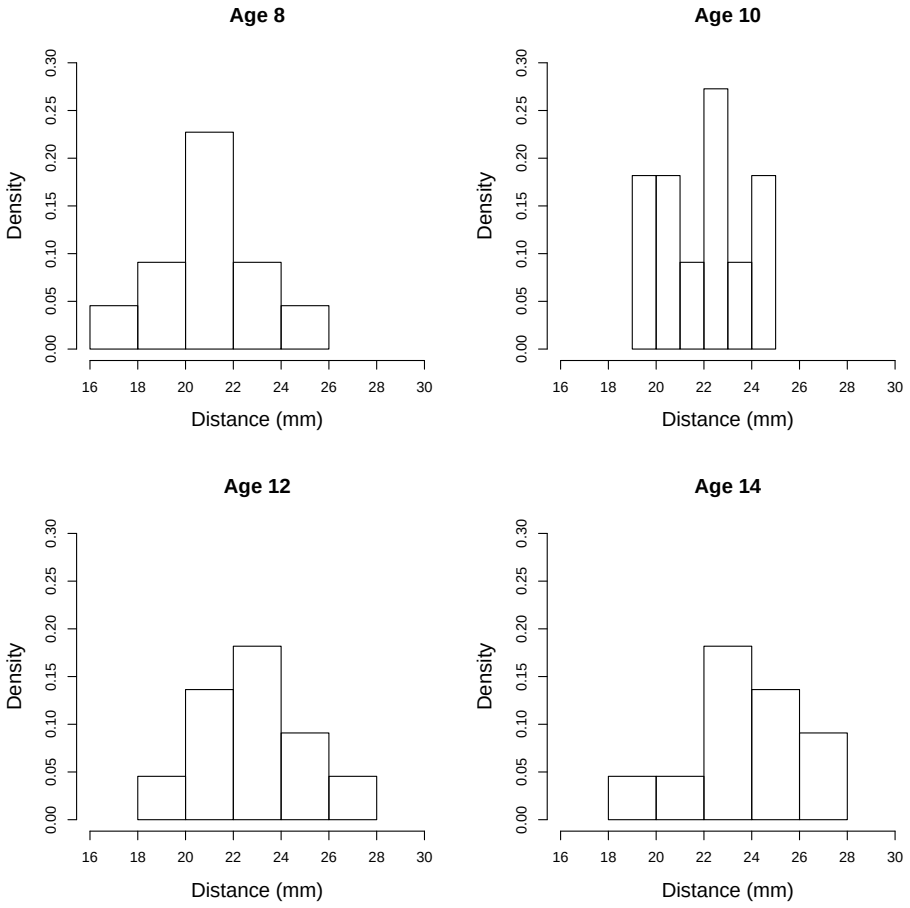

Figure 2: Histogram of pituitary-pterygomaxillary fissure distance in girls by age.

The interpretation of the results reported above is as follows. The *mean* distance at age 11 years in this sample is $\hat{\beta}_0 = 22.65$ mm while the *mean* slope or growth rate over the observed time period is $\hat{\beta}_1 = 0.48$ mm per year, which are average characteristics of the population that these subjects represent. At the individual level, girl-specific trajectories are randomly shifted around the population mean curve with an estimated variance $\hat{\psi}^2_u = 4.28$ mm^2. The intraclass correlation coefficient (ICC) is given by $\hat{\psi}^2_u/(\hat{\psi}^2_u + \hat{\psi}^2_\varepsilon) = 4.28/(4.28 + 0.61) = 0.87$, which suggests that measurements within the same subject are strongly correlated at the *mean* of the marginal distribution of the response. In other words, 87% of the total variability in the intercepts is due to unobserved heterogeneity between subjects with respect to the population average.

These results, however informative, do not tell us what happens in the rest of the conditional distribution. Most importantly, the data show signs of non-normality (Figure 2) for which the assumptions of a location-shift model may prove inappropriate.

3.2. Linear quantile mixed models

As seen in Section 1, the convenience of using an AL distribution allows estimating the

τth conditional quantile using maximum likelihood methods. Let us assume that the $\mathbf{y}_i$'s, $i = 1, \ldots, M$, conditionally on a $q \times 1$ vector of random effects $\mathbf{u}_i$, are independently distributed according to an unknown continuous distribution $F_{\mathbf{y}_i|\mathbf{u}_i}$ (therefore the Gaussian assumption for $\mathbf{y}_i|\mathbf{u}_i$ of a LMM is abandoned). Independence is also assumed for the within-cluster errors, though in principle extensions to allow for within-cluster correlation can be considered. A joint AL model for $\mathbf{y}_i|\mathbf{u}_i$ is introduced, with location and scale parameters given by $\boldsymbol{\mu}_i^{(\tau)} = \mathbf{X}_i\boldsymbol{\theta}_x^{(\tau)} + \mathbf{Z}_i\mathbf{u}_i$ and $\sigma^{(\tau)}$, respectively, where $\boldsymbol{\theta}_x^{(\tau)} \in \mathbb{R}^p$ is a vector of unknown fixed effects. The τth LQMM is given by

$$\mathbf{y} = \boldsymbol{\mu}^{(\tau)} + \boldsymbol{\epsilon}^{(\tau)}, \tag{4}$$

where $\boldsymbol{\mu}^{(\tau)} = \left(\boldsymbol{\mu}_1^{(\tau)}, \ldots, \boldsymbol{\mu}_M^{(\tau)}\right)^\top$, which can be compactly written in matrix form as $\boldsymbol{\mu}^{(\tau)} = \mathbf{X}\boldsymbol{\theta}_x^{(\tau)} + \mathbf{Z}\mathbf{u}$. The skewness parameter τ is set *a priori* and defines the quantile level to be estimated. Also, $\mathbf{u}_i = (u_{i1}, \ldots, u_{iq})^\top$, for $i = 1, \ldots, M$, is assumed to be a zero-median random vector independent from the model's error term and distributed according to $p\left(\mathbf{u}_i|\boldsymbol{\Psi}^{(\tau)}\right)$, where $\boldsymbol{\Psi}^{(\tau)}$ is a $q \times q$ covariance matrix. Note that all the parameters are τ-dependent. The random effects vector $\mathbf{u}$ depends on τ through $\boldsymbol{\Psi}^{(\tau)}$. The superscript τ will be omitted when this is not source of confusion.

From the LQMM in Equation 4, the joint density of $(\mathbf{y}, \mathbf{u})$ based on M clusters in the τth is given by

$$\begin{aligned} p(\mathbf{y}, \mathbf{u}|\boldsymbol{\theta}_x^{(\tau)}, \sigma^{(\tau)}, \boldsymbol{\Psi}^{(\tau)}) &= p(\mathbf{y}|\boldsymbol{\theta}_x^{(\tau)}, \sigma^{(\tau)}, \mathbf{u}) p(\mathbf{u}|\boldsymbol{\Psi}^{(\tau)}) \\ &= \prod_{i=1}^{M} p(\mathbf{y}_i|\boldsymbol{\theta}_x^{(\tau)}, \sigma^{(\tau)}, \mathbf{u}_i) p(\mathbf{u}_i|\boldsymbol{\Psi}^{(\tau)}). \end{aligned} \tag{5}$$

It is worth stressing that, although the conditional distribution $F_{y_{ij}|\mathbf{u}_i}$ is assumed to be unknown, its τ-th quantile is conveniently estimated as the location parameter $\mu_{ij}^{(\tau)} = \mathbf{x}_{ij}^\top\boldsymbol{\theta}_x^{(\tau)} + \mathbf{z}_{ij}^\top\mathbf{u}_i$ of an AL distribution with scale $\sigma^{(\tau)}$ and (given) skewness τ. Since the model's interpretation is conditional, one could also define the parameter $\theta_x^{(\tau)}$ as the τ-th quantile of $y_{ij}|\mathbf{u}_i = \mathbf{0}$.

Following the definition of the joint density in Equation 5, the next step consists in adopting an estimation strategy for the parameters of interest, which prompts considerations on how to deal with the unobserved random effects $\mathbf{u}$. There is a rich literature on mixed models estimation. The typical approach is to integrate the random effects out from the joint distribution and then optimize the integrated (log-) likelihood. The marginal likelihood of a LMM, for example, has a closed-form expression (Pinheiro and Bates 2000) and (restricted) maximum likelihood estimation is carried out using iterative algorithms such as EM (Dempster, Laird, and Rubin 1977) or Newton-Raphson (Laird and Ware 1982) algorithms. In models where integrals do not have a closed-form solution, one needs to resort to approximate methods such as marginal and penalized quasi-likelihood methods, Markov Chain Monte Carlo methods, and numerical integration. This is the case of generalized linear mixed models (Booth and Hobert 1999; Rabe-Hesketh, Skrondal, and Pickles 2002; Pinheiro and Chao 2006), nonlinear mixed models (Pinheiro and Bates 1995) and models for non-Gaussian continuous responses (Staudenmayer, Lake, and Wand 2009).

The first attempt to fit quantile regression models with random intercepts led to a Monte Carlo EM procedure (Geraci and Bottai 2007), which, however, can be computationally intensive

and inefficient. A different approach based on Gaussian quadrature has been proposed by Geraci and Bottai (2013) and implemented in the R package **lqmm**.

Briefly, the log-likelihood for M clusters integrated over $\mathbb{R}^q$ is given by

$$\ell\left(\boldsymbol{\theta}_x^{(\tau)}, \sigma^{(\tau)}, \boldsymbol{\Psi}^{(\tau)} | \mathbf{y}\right) = \sum_i^M \left[\log \sigma_{n_i}^{(\tau)} + \log \int_{\mathbb{R}^q} \exp\left\{-\frac{1}{\sigma}\rho_\tau\left(\mathbf{y}_i - \boldsymbol{\mu}_i^{(\tau)}\right)\right\} p\left(\mathbf{u}_i | \boldsymbol{\Psi}^{(\tau)}\right) du_i\right],$$

where I used the simplified notation $\sigma_{n_i}^{(\tau)} = \tau^{n_i}(1-\tau)^{n_i}/\sigma^{n_i}$ and $\rho_\tau\left(\mathbf{y}_i - \boldsymbol{\mu}_i^{(\tau)}\right) = \sum_{j=1}^{n_i} \rho_\tau\left(y_{ij} - \mu_{ij}^{(\tau)}\right)$. For fitting purposes, the covariance matrix of the random effects is parameterized in terms of an m-dimensional vector of non-redundant parameters $\boldsymbol{\theta}_z$, i.e., $\boldsymbol{\Psi}(\boldsymbol{\theta}_z)$. The parameter of interest is defined as $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_x^\top, \boldsymbol{\theta}_z^\top\right)^\top$.

It is immediate to verify that there is a strict relationship between the choice of a distribution for $\mathbf{u}$ and the type of quadrature. The assumption of normal random effects, $\mathbf{u} \sim N(\mathbf{0}, \boldsymbol{\Psi})$, which is often introduced in mixed models, leads directly to a Gauss-Hermite quadrature (Geraci and Bottai 2013) for the approximate log-likelihood

$$\ell_{\text{app}}(\boldsymbol{\theta}, \sigma | \mathbf{y}) = \sum_i^M \log\left\{\sum_{k_1=1}^K \cdots \sum_{k_q=1}^K p\left(\mathbf{y}_i | \boldsymbol{\theta}_x, \sigma, \left(\boldsymbol{\Psi}^\top\right)^{1/2} v_{k_1,\ldots,k_q}\right) \times \prod_{l=1}^q w_{k_l}\right\}, \quad (6)$$

with nodes $v_{k_1,\ldots,k_q} = (v_{k_1}, \ldots, v_{k_q})^\top$ and weights w_{k_l}, $l = 1, \ldots, q$, respectively. The constant K is an integer giving the number of points for each of the q one-dimensional integrals over the real line. As a robust alternative to the Gaussian distribution, it is also possible to consider the symmetric Laplace (double-exponential) distribution which leads to an approximation similar to Equation 6, but where the nodes and weights are those from a Gauss-Laguerre quadrature rule. In **lqmm**, the quadrature nodes and weights are obtained using the command `gauss.quad` from **statmod** (Smyth and with contributions from Yifang Hu and Peter Dunn and Belinda Phipson 2013).

3.3. The main call `lqmm`

I now illustrate the basic arguments in the main command `lqmm`. After loading the package **lqmm**

```
R> library("lqmm")

Package lqmm (1.04) loaded. Type citation("lqmm") on how to cite this package
```

the documentation for `lqmm` is accessed through `help("lqmm")`. The arguments of this function can also be displayed on the R console:

```
R> args(lqmm)

function (fixed, random, group, covariance = "pdDiag", tau = 0.5,
    nK = 7, type = "normal", rule = 1, data = sys.frame(sys.parent()),
    subset, weights, na.action = na.fail, control = list(), contrasts = NULL,
    fit = TRUE)
```

Let us start with a simple model for the conditional median of `distance` using the orthodontic growth data. As in Section 3.1, a random intercept is included in the linear predictor to account for the correlation of repeated observations within `Subject`. The first two arguments, `fixed` and `random`, are formula objects which define, respectively, the fixed and the random parts of the linear predictor $\boldsymbol{\mu}_i^{(\tau)} = \mathbf{X}_i\boldsymbol{\theta}_x^{(\tau)} + \mathbf{Z}_i\mathbf{u}_i$, $i = 1, \ldots, M$, while the clustering or grouping variable is defined in the argument `group`. All variables are taken from the optional data frame `data`. Finally, the quantile level τ is specified using the argument `tau` (by default, the median):

```
R> fit.lqmm <- lqmm(fixed = distance ~ age.c, random = ~ 1, group =
+ Subject, tau = 0.5, nK = 7, type = "normal", data = Orthodont.sub)
R> fit.lqmm

Call: lqmm(fixed = distance ~ age.c, random = ~1, group = Subject,
    tau = 0.5, nK = 7, type = "normal", data = Orthodont.sub)

Quantile 0.5

Fixed effects:
(Intercept)         age.c
    22.9410        0.4417

Covariance matrix of the random effects:
(Intercept)
      2.341

Residual scale parameter: 0.2969 (standard deviation 0.8397)
Log-likelihood: -68.19

Number of observations: 44
Number of groups: 11
```

Before entering in the details of the estimation algorithms, let us read the output obtained from the above call to `lqmm` which returns an R list of class `lqmm`. The estimated fixed effects $\hat{\boldsymbol{\theta}}_x = (22.94, 0.44)^\top$ show that the *median* distance at age 11 years in girls is 22.94 mm while the *median* slope or growth rate is 0.44 mm per year. The random effects have an estimated variance of $\hat{\psi}_u^2 = 2.34$ mm^2. The ICC is given by $\hat{\psi}_u^2/(\hat{\psi}_u^2 + \hat{\psi}^2) = 2.34/(2.34 + 0.84^2) = 0.77$. One of course might ask where the number 0.84 comes from. Recall that a measure of the scale of the residuals obtained after solving (1) can be calculated as (Koenker 2005)

$$\frac{1}{M}\sum_{i=1}^{M} \rho_\tau\left(y_i - \mathbf{x}_i^\top\hat{\boldsymbol{\beta}}\right),$$

which is also the MLE of the scale parameter of an AL distribution (Yu and Zhang 2005). As seen previously, σ is related to the variance of an AL, which provides a model-based 'residual variance' to derive the ICC above. Given $\hat{\sigma} = 0.2969$ and $\tau = 0.5$, one can also calculate

```
R> sqrt(varAL(sigma = 0.2969, tau = 0.5))

[1] 0.83976
```

The estimated parameters are stored in the fitted object `fit.lqmm` as `fit.lqmm$theta` (θ), `fit.lqmm$theta_x` (θ_x), `fit.lqmm$theta_z` (θ_z), and `fit.lqmm$scale` (σ). The generic function `coefficients` (or `coef`) can be used to obtain the fixed effects while the function `cov.lqmm` extracts the matrix $\boldsymbol{\Psi}$, as shown below. The use of `cov.lqmm` is recommended since, as mentioned before, $\boldsymbol{\Psi}$ is parametrised in terms of $\boldsymbol{\theta}_z$.

```
R> coef(fit.lqmm)

(Intercept)       age.c
 22.9410472   0.4417377

R> cov.lqmm(fit.lqmm)

(Intercept)
   2.340926
```

In general, the interpretation of the LQMM parameters is specific to the quantile being estimated (Geraci and Bottai 2013). For example, the output below shows that rate of growth in girls is 0.50 mm per year at the third quartile ($\tau = 0.75$). The random effects have an estimated variance of $\hat{\psi}_u^2 = 2.21$ mm^2, yielding an ICC equal to 0.71. The basic idea of LQMM is that the covariates might exert different effects at different quantiles of the outcome distribution, as assessed in standard quantile regression (Koenker and Bassett 1978), and that the degree of unobserved heterogeneity might also be characterised with τ-specific variance parameters.

```
R> lqmm(fixed = distance ~ age.c, random = ~ 1, group = Subject,
+ tau = 0.75, nK = 7, type = "normal", data = Orthodont.sub)
Call: lqmm(fixed = distance ~ age.c, random = ~1, group = Subject,
    tau = 0.75, nK = 7, type = "normal", data = Orthodont.sub)

Quantile 0.75

Fixed effects:
(Intercept)       age.c
      23.22        0.50

Covariance matrix of the random effects:
(Intercept)
      2.207

Residual scale parameter: 0.2233 (standard deviation 0.9416)
Log-likelihood: -68.06

Number of observations: 44
Number of groups: 11
```

Covariance matrix	Argument `covariance`	$\mathsf{VAR}(u_l)$ $l = 1, \ldots, q$	$\mathsf{COV}(u_l, u_{l'})$ $l \neq l'$	m	Argument `type`
Multiple of an identity	`pdIdent`	ψ_u^2	0	1	`normal` or `robust`
Compound symmetry	`pdCompSymm`	ψ_u^2	ϕ	1 ($q = 1$) or 2 ($q > 1$)	`normal` only
Diagonal	`pdDiag`	ψ_l^2	0	q	`normal` or `robust`
General positive-definite	`pdSymm`	ψ_l^2	$\phi_{ll'}$	$q(q+1)/2$	`normal` only

Table 1: Summary table of the covariance structures available in **lqmm** and type of quadrature available.

Let us now focus on the arguments `type` and `covariance` which are relevant to the choice of the distribution of the random effects and, ultimately, to numerical integration. The number of quadrature nodes K is specified with `nK` and this is set to 7 by default. Since guidance on how to choose K in LQMMs is given by Geraci and Bottai (2013), here I will not discuss this issue further. However, a general recommendation is to start with low values of K (say, $K < 10$), as the total size of the quadrature grid, K^q, grows exponentially with the number of random effects. The argument `type` takes the character string `"normal"` for Gaussian random effects (i.e., Gauss-Hermite quadrature) or `"robust"` for Laplace random effects (i.e., Gauss-Laguerre quadrature). However, while the Gauss-Hermite quadrature allows for all types of covariance matrix $\boldsymbol{\Psi}$ implemented in **lqmm**, the Gauss-Laguerre quadrature can be used for uncorrelated random effects only (see Geraci and Bottai 2013, for further details). Table 1 gives a summary of the options currently available in the package. The types of covariance matrices specified in `covariance` are named as in **nlme**. The table also includes the number of unique parameters for each type of matrix, that is, the dimension m of $\boldsymbol{\theta}_z$.

Another argument related to the choice of the quadrature rule is `rule`, which introduces integration on sparse grids (Heiss and Winschel 2008) and nested integration rules for Gaussian weights (Genz and Keister 1996) as implemented in the package **SparseGrid** (Ypma 2012). This part of the code, which has not yet been extensively tested, aims at introducing computational relief when the size of the quadrature grid is large. Further studies are needed to assess the performance of these approaches.

4. Estimation algorithms

4.1. Optimization control

In **lqmm**, there are currently two algorithms to minimize the negative integrated log-likelihood in Equation 6. The default is the gradient-search method described by Geraci and Bottai (2013), which makes use of the Clarke's derivative of the objective function to find the path of steepest descent. The alternative is Nelder-Mead optimization, as implemented in `optim`, which belongs to the class of direct search methods.

The argument `control` in `lqmm` takes a named list of optimization control parameters. Such a list is also produced by the function `lqmmControl`, whose arguments and corresponding default values are displayed below:

```
R> args(lqmmControl)
```

```
function (method = "gs", LP_tol_ll = 1e-05, LP_tol_theta = 1e-05,
    check_theta = FALSE, LP_step = NULL, beta = 0.5, gamma = 1,
    reset_step = FALSE, LP_max_iter = 500, UP_tol = 1e-04, UP_max_iter = 20,
    startQR = FALSE, verbose = FALSE)
```

The argument `method` specifies the optimization algorithm: gradient-based (`method = "gs"`) or derivative-free (`method = "df"`) minimization. The basic computing engines `lqmm.fit.gs` and `lqmm.fit.df` are called within `lqmm` but they can be used outside of the main call. These functions and the relevant `lqmmControl`'s parameters are described in the next two sections.

4.2. Gradient-search optimization

The function `lqmm.fit.gs` executes the gradient-based estimation of $\boldsymbol{\theta}$. It is a wrapper for the C function `gradientSd_h` and it performs pre- and post-estimation checks. The gradient-search algorithm (Geraci and Bottai 2013) works as follows. From a current parameter value, the algorithm searches the positive semi-line in the direction of the gradient for a new parameter value at which the likelihood is larger. The algorithm stops when the change in the likelihood is less than a specified tolerance. At iteration t, let $s\left(\boldsymbol{\theta}^t\right)$ denote the Clarke gradient of the negative approximate log-likelihood, rewritten compactly as $\ell_{\text{app}}\left(\boldsymbol{\theta}^t, \sigma^0\right)$, given σ^0. The minimization steps for $\boldsymbol{\theta}$ are:

1. Set $\boldsymbol{\theta} = \boldsymbol{\theta}^0$; $\delta = \delta^0$; $\sigma = \sigma^0$; $t = 0$.
2. If $\ell_{\text{app}}\left\{\boldsymbol{\theta}^t - \delta^t s\left(\boldsymbol{\theta}^t\right), \sigma^0\right\} \geq \ell_{\text{app}}\left(\boldsymbol{\theta}^t, \sigma^0\right)$
   (a) then set $\delta^{t+1} = a\delta^t$;
   (b) else if $|\ell_{\text{app}}\left(\boldsymbol{\theta}^t, \sigma^0\right) - \ell_{\text{app}}\left\{\boldsymbol{\theta}^t - \delta^t s\left(\boldsymbol{\theta}^t\right), \sigma^0\right\}| < \omega_1$
      (i) then set $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t - \delta s\left(\boldsymbol{\theta}^t\right)$; return $\boldsymbol{\theta}^{t+1}$; stop;
      (ii) else set $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t - \delta s\left(\boldsymbol{\theta}^t\right)$; $\delta^{t+1} = b\delta^t$.
3. Set $t = t + 1$; go to step 2.

A check on the convergence of the parameter $\boldsymbol{\theta}$ can be introduced in step (b) by verifying $\max |\ \delta^t s\left(\boldsymbol{\theta}^t\right)\ | < \omega_2$, where $\omega_2 > 0$ controls the tolerance. By default, this check is not carried out in `lqmm` but it can be changed by setting `check_theta = TRUE` in `lqmmControl`.

The iterative loop for $\boldsymbol{\theta}$ is the 'lower' level of the optimization. The 'upper' level of the algorithm consists in updating the scale parameter: once the algorithm finds a solution for $\boldsymbol{\theta}$, given σ^0, the scale parameter is estimated residually to obtain σ^1. If the change in the parameter, is sufficiently small, say $|\sigma^0 - \sigma^1| < \omega_3$, with $\omega_3 > 0$, the algorithm stops. Otherwise another iterative loop for θ is initialized by setting $\sigma = \sigma^1$ and so forth until convergence is achieved for σ as well.

The starting values $\boldsymbol{\theta}^0$ and σ^0 are specified in `lqmm.fit.gs`'s arguments `theta_0` and `sigma_0`, respectively. The starting values for $\boldsymbol{\theta}_x$ and σ are provided automatically by `lqmm`, either using ordinary least-squares estimates or, if `startQR = TRUE`, L_1-norm estimates using `lqm` (see Section 7). The elements of `theta_z` are all set equal to one. The initial step $\delta^0 > 0$ is specified in the argument `LP_step` of `lqmmControl`. If not provided by the user, this is set equal to the standard deviation of the response variable. Additionally, it is possible to instruct the algorithm to reset δ to its initial value (`reset_step = TRUE`) at step (2.ii) of the algorithm above, i.e., $\delta^{t+1} = \delta^0$.

The values for the tolerance parameters ω_1, ω_2, and ω_3 can be passed to the arguments `LP_tol_ll`, `LP_tol_theta` and `UP_tol`, respectively. The contraction step factor $a \in (0, 1)$ and the expansion step factor $b \geq 1$ are specified in the arguments `beta` and `gamma`, respectively. Finally, the maximum numbers of iterations for the lower and upper loops are given in `LP_max_iter` and `UP_max_iter`, respectively. It is possible to monitor the value of the objective function as the algorithm proceeds by setting `verbose = TRUE`.

The function `lqmm.fit.gs` can be called directly by the user. This allows, for example, specifying arbitrary starting values for $\boldsymbol{\theta}$ and σ. The list of arguments can be created by first calling `lqmm` with `fit = FALSE`. This object, opportunely modified, can be then passed to `lqmm.fit.gs`:

```
R> fit.args <- lqmm(fixed = distance ~ age.c, random = ~ 1, group =
+ Subject, tau = 0.5, nK = 7, type = "normal", data = Orthodont.sub,
+ fit = FALSE)
R> fit.args$theta_0

(Intercept)       age.c
 22.6477273   0.4795455   1.0000000

R> fit.args$theta_0[3] <- 0.001
R> do.call("lqmm.fit.gs", args = fit.args)

$theta
[1] 2.260618e+01 4.646043e-01 7.221280e-05

$scale
[1] 0.8368506

$gradient
[1] 0.8706574 0.2220348 0.4181307

$logLik
[1] -97.15939

$opt
$opt$low_loop
[1] 22
```

```
$opt$upp_loop
[1] 2
```

The output above shows that setting the starting value for `theta_z` (which in this case is the third element of `theta`) to 0.001 causes the algorithm to converge to a different optimum for $\boldsymbol{\theta}_z$, in the vicinity of the starting point itself. Care, therefore, must be taken in defining the initial values. The output also reports the number of iterations at convergence for the upper loop (`opt$upp_loop`) and that for the last cycle of the lower loop (`opt$low_loop`). If the algorithm fails to converge, a warning will be produced. By way of example, let us set the maximum number of iterations for $\boldsymbol{\theta}$ to a small value, say `LP_max_iter = 10`:

```
R> fit.args$control$LP_max_iter <- 10
R> do.call("lqmm.fit.gs", args = fit.args)

$theta
[1] 22.622664418  0.465670539  0.002287113

$scale
[1] 0.837245

$gradient
[1]  1.19432138  1.19432138 -0.08808333

$logLik
[1] -97.18022

$opt
$opt$low_loop
[1] -1

$opt$upp_loop
[1] 2


Warning message:
In errorHandling(OPTIMIZATION$low_loop, "low", control$LP_max_iter,  :
Lower loop did not converge in: lqmm. Try increasing max number of iterations
 (10) or tolerance (1e-05)
```

The warning message will suggest using less restrictive convergence criteria, while reporting in parentheses those last used. Note that `opt$low_loop` is equal to -1, which is the value that the function `errorHandling` interprets as 'algorithm did not converge', as opposed to -2 which is the code for 'algorithm did not start'.

4.3. Derivative-free optimization

As mentioned before, this algorithm is based on Nelder-Mead optimization routines. The function `lqmm.fit.df` is a wrapper for the command `optim`, which in turn minimizes the negative log-likelihood as returned by the C function `ll_h_R`. It proceeds similarly to gradient-search by alternating a loop for $\boldsymbol{\theta}$ and a step to update σ. The parameters `LP_tol_ll`, `LP_max_iter` and `verbose` in `lqmmControl` are passed to `optim` via the arguments `abstol`, `maxit` and `trace`, respectively.

Two successive calls to `fit.lqmm.df` using the list `fit.args` are shown below: the first with `theta_z` left equal to 0.001, the second with `theta_z` changed back to 1. The maximum number of iterations is restored to 500.

```
R> fit.args$control$LP_max_iter <- 500
R> fit.args$control$verbose <- TRUE
R> do.call("lqmm.fit.df", args = fit.args)

Upper loop = 1
  Nelder-Mead direct search function minimizer
function value for initial parameters = 97.525938
  Scaled convergence tolerance is 1.45325e-06
Stepsize computed as 2.264773
[...]
Exiting from Nelder Mead minimizer
    178 function evaluations used
(1) logLik = -84.378
[snip]
Upper loop = 4
[...]
Exiting from Nelder Mead minimizer
    219 function evaluations used
(4) logLik = -68.16
$theta
(Intercept) I(age - 11)
  22.937500    0.437500    1.515932

$scale
[1] 0.2963305

$logLik
[1] -68.15952

$opt
$opt$low_loop
[1] 219

$opt$upp_loop
[1] 4
```

```
R> fit.args$theta_0[3] <- 1
R> do.call("lqmm.fit.df", args = fit.args)

[...]
$theta
(Intercept) I(age - 11)
 22.9374987   0.4375005   1.5159317

$scale
[1] 0.2963305

$logLik
[1] -68.15952

$opt
$opt$low_loop
[1] 147

$opt$upp_loop
[1] 3
```

These two calls produce similar estimates of the parameters. However, convergence in the first call is attained with a larger overall number of function evaluations (889 against 675, respectively).

In comparison, gradient search needed in total 93 evaluations of the likelihood and score functions when starting from

```
R> fit.args$theta_0

(Intercept)       age.c
 22.6477273   0.4795455   1.0000000
```

converging to

```
R> do.call("lqmm.fit.gs", args = fit.args)

[...]
$theta
[1] 22.9410472  0.4417377  1.5300087

$scale
[1] 0.2968949

$gradient
[1] -1.1507525 -0.8904712  2.8303527
```

```
$logLik
[1] -68.19345

$opt
$opt$low_loop
[1] 21

$opt$upp_loop
[1] 3
```

It is possible to transform `theta_z` back to $\hat{\psi}_u^2$ using the function `covHandling`, for example:

```
R> covHandling(theta = c(1.5300087), n = 1, cov_name = fit.args$cov_name,
+ quad_type = "normal")

[1] 2.340927
```

When the estimated covariance matrix is not positive definite, this function will produce a warning and will apply an approximation to the nearest symmetric positive definite matrix by using `corpcor::make.positive.definite` (Schäfer, Opgen-Rhein, Zuber, Ahdesmäki, Silva, and Strimmer. 2013).

5. Methods for lqmm objects

5.1. The summary and bootstrap functions

Consider first a call to `lmer` using the orthodontic growth data, in which both random intercepts and slopes are specified in the linear predictor:

```
R> fit.lmer <- lmer(distance ~ age.c + (age.c|Subject), data = Orthodont.sub)
R> summary(fit.lmer)

Linear mixed model fit by REML
Formula: distance ~ age.c + (age.c | Subject)
   Data: Orthodont.sub
   AIC   BIC logLik deviance REMLdev
 149.4 160.1 -68.71    134.6   137.4
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 Subject  (Intercept) 4.319035 2.07823
          age.c       0.025898 0.16093  0.530
 Residual             0.446591 0.66827
Number of obs: 44, groups: Subject, 11

Fixed effects:
```

```
            Estimate Std. Error t value
(Intercept) 22.64773    0.63466   35.68
age.c        0.47955    0.06621    7.24

Correlation of Fixed Effects:
      (Intr)
age.c 0.384
```

To estimate a similar model for the quartiles of `distance` conditional on `age.c` (age centred on 11 years), the vector `c(0.25, 0.5, 0.75)` is passed to the argument `tau`. The resulting `lqmm` object will contain the estimates of the three quantile models, which are fitted sequentially and independently using the same formulas for fixed and random effects. The covariance model of the random effects `"pdSymm"` will have in this case $m = 3$ unique parameters (two variances, one for the intercept and one for the slope, and one intercept-slope covariance).

```
R> fit.lqmm <- lqmm(distance ~ age.c, random = ~ age.c, group = Subject,
+ covariance = "pdSymm", tau = c(0.25, 0.5, 0.75), nK = 7, type = "normal",
+ data = Orthodont.sub, control = lqmmControl(method = "df"))
```

The function `summary.lqmm` produces a summary object which provides standard errors, $(1-\alpha)\%$ confidence intervals and p values for the coefficients and the scale parameter of each quantile model. Inference on parameters is based on block-bootstrap (Geraci and Bottai 2013) which is currently the only method implemented in the package. The number of bootstrap replications (default 50) can be specified in `summary` using the additional argument `R`. The results of the likelihood ratio test and AIC values are also produced.

```
R> system.time(print(fit.lqmm.s <- summary(fit.lqmm, R = 100, seed = 52)))

Call: lqmm(fixed = distance ~ age.c, random = ~age.c, group = Subject,
  covariance = "pdSymm", tau = c(0.25, 0.5, 0.75), nK = 7, type = "normal",
  data = Orthodont.sub, control = lqmmControl(method = "df"))

tau = 0.25

Fixed effects:
              Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept) 22.80948    0.82165    21.17915     24.4398 < 2.2e-16 ***
age.c        0.46518    0.13831     0.19075      0.7396  0.001096 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

tau = 0.5

Fixed effects:
              Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept) 23.11215    0.83347    21.45836     24.7659 < 2.2e-16 ***
age.c        0.53738    0.10559     0.32787      0.7469  1.71e-06 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

tau = 0.75

Fixed effects:
                Value Std. Error lower bound upper bound  Pr(>|t|)
(Intercept) 24.273541   0.830865   22.624924     25.9222 < 2.2e-16 ***
age.c        0.575486   0.093357    0.390246      0.7607 1.539e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Null model (likelihood ratio):
[1] 25.869 (p = 3.653e-07) 28.732 (p = 8.314e-08)  8.175 (p = 4.247e-03)
AIC:
[1] 146.4 (df = 6) 141.6 (df = 6) 154.0 (df = 6)
   user  system elapsed
  60.70    0.00   60.71
There were 31 warnings (use warnings() to see them)
```

It is interesting to note in the output above that the magnitude of the slope for age increases with increasing quantile level. The elapsed CPU time was about one minute to run 100 replications for three quantiles (approximately 0.04 seconds per sample). The 'non-convergence' warnings

```
R> warnings()

Warning messages:
1: In errorHandling(OPTIMIZATION$low_loop, "low", control$LP_max_iter,  ... :
  Lower loop did not converge in: lqmm. Try increasing max number of
iterations (500) or tolerance (1e-05)
[...]
```

may be of less concern if they occur during bootstrap. This may happen when the algorithm requires a certain number of data points to estimate the specified regression model but one or more bootstrap samples do not provide adequate information because of a particular configuration of their units. Clearly, this situation is more likely to happen when estimating tail quantiles with a number of parameters that is relatively high given the size and design of the dataset. In the first instance, one can assess the summary output by using less stringent optimization parameters. This is shown below by increasing the number of maximum iterations and the tolerance for the previous example. As a result, the same estimates are obtained but with no warnings:

```
R> fit.lqmm$control$LP_tol_ll <- 1e-3
R> fit.lqmm$control$LP_max_iter <- 1000
R> summary(fit.lqmm, R = 100, seed = 52)
```

The function `boot.lqmm` executes the bootstrapping, producing an object of class `boot.lqmm` that is stored within the `summary` output (e.g., `fit.lqmm.s$B`). The function `boot.lqmm` can also be applied directly to a fitted `lqmm` object:

```
R> fit.boot <- boot.lqmm(fit.lqmm, R = 100, seed = 52, startQR = FALSE)
```

Among the `boot.lqmm`'s arguments, it is worth calling the attention on `startQR`. If set to `TRUE`, the fitted parameters in the `lqmm` object are used as starting values for each bootstrap sample. On the one hand this may speed up the fitting process. However, it may also cause the algorithm to converge too often to a similar optimum, which would ultimately result in underestimated standard errors. Finally, all bootstrap estimates are stored in an array of dimension `c(R, p + m, nt)`, where `R` represents the number of bootstrap replications, `p + m` the length of `theta` and `nt` the length of `tau`. Estimated parameters for fixed (`theta_x`) and random (`theta_z`) effects can be extracted separately from a `boot.lqmm` object using the function `extractBoot.lqmm`, while a summary can be produced using the function `summary.lqmm`:

```
R> extractBoot.lqmm(fit.lqmm.s$B, "random")
R> extractBoot.lqmm(fit.boot, "random")
R> summary(fit.lqmm.s$B)
R> summary(fit.boot)
```

Bootstrap estimates can also be used to perform inference on the difference between regression quantiles. A 95% bootstrap confidence interval for the interquartile regression coefficients can be computed as follows:

```
R> B <- extractBoot.lqmm(fit.boot, "fixed")[,'age.c',c(1,3)]
R> quantile(apply(B, 1, diff), probs = c(.025,0.975))

      2.5%      97.5%
-0.4141825  0.1731508
```

5.2. Prediction functions

Other important functions include `raneff.lqmm`, `predict.lqmm` and `residuals.lqmm`. Here I introduce the former in detail and briefly describe the other two.

Prediction of random effects in LQMMs is still an ongoing research issue. Geraci and Bottai (2013) provided provisional guidance on the development of an approach based on best prediction. The best linear predictor (BLP) of $\mathbf{u}$ for the τ-th LQMM is given by

$$\mathbf{u}_{\text{BLP}}^{(\tau)} = \boldsymbol{\Psi}^{(\tau)}\mathbf{Z}^{\top}\boldsymbol{\Sigma}^{-1}\left\{\mathbf{y} - \mathbf{X}\boldsymbol{\theta}_x^{(\tau)} - \mathsf{E}\left(\boldsymbol{\epsilon}^{(\tau)}\right)\right\}, \tag{7}$$

where $\boldsymbol{\Sigma} \equiv \mathsf{COV}(\mathbf{y}) = \mathbf{Z}\boldsymbol{\Psi}^{(\tau)}\mathbf{Z}^{\top} + \psi_{\epsilon}^{(\tau)}\mathbf{I}_N$ and $\psi_{\epsilon}^{(\tau)} = \frac{\sigma^2(1-2\tau+2\tau^2)}{(1-\tau)^2\tau^2}$.

For a median random-intercepts model, the BLP in Equation 7 simplifies to

$$\mathbf{u}_{\text{BLP}}^{(0.5)} = \mathbf{Z}^{\top}\left\{\mathbf{Z}\mathbf{Z}^{\top} + \lambda\mathbf{I}_N\right\}^{-1}\left\{\mathbf{y} - \mathbf{X}\boldsymbol{\theta}_x^{(0.5)}\right\}, \tag{8}$$

where $\lambda = \psi_\epsilon^{(0.5)}/\psi_u^2$, provided that $\psi_u^2 > 0$. Note that $\psi_\epsilon^{(0.5)} = 8\sigma^2$.

The BLP in Equation 8 resembles that obtained from a LMM. It is therefore reasonable to expect that predicted random effects from a mean and median mixed models should be comparable when the parameters' estimates are similar. The code to obtain predicted subject-specific intercepts and slopes of the mean and median models for the orthodontic growth data is shown below:

```
R> uhat.lqmm <- raneff.lqmm(fit.lqmm)[['0.50']]
R> uhat.lmer <- ranef(fit.lmer)$Subject
R> cbind(uhat.lqmm, uhat.lmer)
```

```
    (Intercept)        age.c (Intercept)       age.c
F01 -1.67336060 -0.151487226 -1.24648085 -0.07545064
F02 -0.06175306  0.145982503  0.37668930  0.15411858
F03  0.65729898  0.199918276  1.10996123  0.19327565
F04  1.65962130  0.024167200  2.16096461  0.04634387
F05 -0.50600768 -0.166157311 -0.04428503 -0.09397927
F06 -1.91022033 -0.159960086 -1.48909355 -0.08088568
F07 -0.10411736  0.003396539  0.34952012  0.03985892
F08  0.18762578 -0.197773117  0.67268540 -0.12337800
F09 -1.92716605 -0.216994472 -1.49996122 -0.12658955
F10 -4.38453818 -0.206149328 -4.02837614 -0.10367577
F11  3.11467111  0.189073132  3.63837614  0.17036188
```

The BLP of the random effects is implemented in the function `raneff.lqmm`, which takes a `lqmm` object as the only argument. If more than one quantile model has been fitted, the output of `raneff.lqmm` will be a named list of predictions, with names given by `tau`.

Prediction of the response can be carried out using the function `predict.lqmm`. The argument `level` specifies whether predictions should be returned at the 'population' level (`level = 0`), that is $\hat{\mathbf{y}}^{(\tau)} = \mathbf{X}\hat{\boldsymbol{\theta}}_x^{(\tau)}$, or at the 'cluster' level (`level = 1`), that is $\hat{\mathbf{y}}^{(\tau)} = \mathbf{X}\hat{\boldsymbol{\theta}}_x^{(\tau)} + \mathbf{Z}\hat{\mathbf{u}}_{\text{BLP}}^{(\tau)}$. This is similar to the argument `level` in `nlme::predict.lme` (the reader is referred to Geraci and Bottai (2013) for a discussion on the interpretation of the regression coefficients at the population level in LQMM). Analogously, residuals can be calculated at one or the other level by using the function `residuals.lqmm` as shown below:

```
R> predict(fit.lqmm, level = 0)
R> residuals(fit.lqmm, level = 0)
```

By way of example, Figure 3 shows the mean and quartile regression lines predicted for each of the 11 girls (that is, `level = 1`) in the orthodontic growth data (see Figure 1.14 in Pinheiro and Bates 2000, p. 38). While medians and means tally in magnitude and sign, for some subjects the distance between the first and third quartiles varies by age, which suggests the presence of a slight heteroscedasticity in the conditional distribution $y|u$.

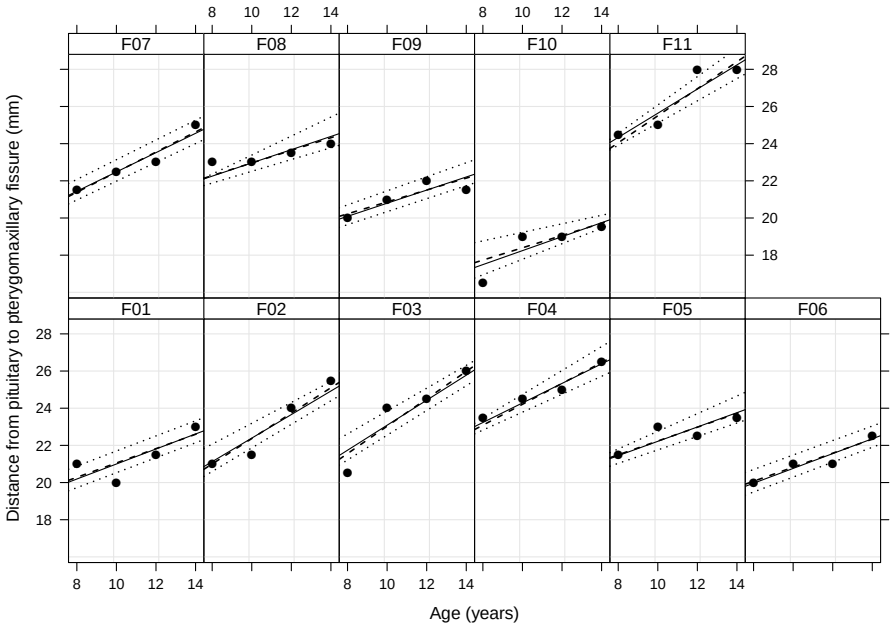

Figure 3: Trellis plot of subject-specific predicted curves (`level = 1`) of pituitary-pterygomaxillary fissure distance in girls (mean, solid line; median, dashed line; first and third quartiles, dotted lines).

6. Modelling conditional quantiles

So far individual **lqmm** commands have been described separately. In this section the focus is on performing a quantile analysis and discussing related modelling choices.

Consider three linear quantile models with fixed and random effects for age, sex and age-sex interaction. The random effects are assumed normally distributed with either identity (Model 1), or compound symmetry (Model 2) or diagonal (Model 3) covariance matrices. A fourth model (Model 4) includes the same fixed effects, but no interaction in the vector $\mathbf{u}$, which is assumed normally distributed with diagonal covariance matrix. The number of quadrature nodes `nK` was set to 9, while the estimation control parameters were defined as

```
R> ctrl <- lqmmControl(method = "df", LP_tol_ll = 1e-4, LP_tol_theta = 1e-4,
+ check_theta = TRUE, LP_max_iter = 1000)
```

Three quartiles were estimated as follows

```
R> fit.lqmm.id <- lqmm(distance ~ age.c * Sex, random = ~ age.c * Sex,
+ group = Subject, covariance = "pdIdent", tau = c(0.25, 0.5, 0.75),
+ nK = 9, type = "normal", data = Orthodont, control = ctrl)
R> fit.lqmm.csymm <- lqmm(distance ~ age.c * Sex, random = ~ age.c * Sex,
+ group = Subject, covariance = "pdCompSymm", tau = c(0.25, 0.5, 0.75),
+ nK = 9, type = "normal", data = Orthodont, control = ctrl)
R> fit.lqmm.diag <- lqmm(distance ~ age.c * Sex, random = ~ age.c * Sex,
+ group = Subject, covariance = "pdDiag", tau = c(0.25, 0.5, 0.75),
+ nK = 9, type = "normal", data = Orthodont, control = ctrl)
R> fit.lqmm.diag2 <- lqmm(distance ~ age.c * Sex, random = ~ age.c,
+ group = Subject, covariance = "pdDiag", tau = c(0.25, 0.5, 0.75),
+ nK = 9, type = "normal", data = Orthodont, control = ctrl)
```

Table 2 shows estimated parameters for the LQMMs described above. Standard errors were based on `R = 50` bootstrap replications. The values reported in the last column were obtained from analogous LMMs with equivalent fixed and random effects structures. However, Model 4 was fitted using the function `nlme::varIdent` (e.g., Pinheiro and Bates 2000, p. 177) to allow for different variances by sex for the within-group error.

The estimates from Model 1 suggest that at $\tau = 0.75$ the population growth rate in boys is faster than the rate at lower quantiles. In other words, the effort required to stay on the same quantile level between any two time points is greater for those who rank higher in the outcome distribution than for those who rank lower in the distribution (clearly, this does not imply that if a given subject ranks, say, 75th at age 8 years will necessarily rank similarly at a later time). However, the difference in slopes between the first and third quartiles contributes to a mere 0.66 mm over a 6-year time period which may not be clinically relevant. Growth trajectories in girls start from lower levels and proceed at slower pace than in boys at all quantiles. The estimated variance of the random effects, assumed to be the same for age, sex and age-sex interaction, is very small or null at $\tau = 0.25$, but larger at higher quartiles. As compared to Models 2-4, Model 1 is the most parsimonious but it provides the worst AIC values for $\tau = 0.25$ and $\tau = 0.75$.

	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	LMM
Model 1 – `covariance = "pdIdent"`				
intercept	23.43 (0.81)	24.97 (0.57)	26.25 (0.74)	24.97 (0.30)
age	0.64 (0.13)	0.66 (0.12)	0.75 (0.16)	0.78 (0.25)
sex	−2.05 (0.93)	−1.97 (0.75)	−2.25 (0.98)	−2.32 (0.54)
age:sex	−0.27 (0.58)	−0.16 (0.20)	−0.25 (0.35)	−0.30 (0.49)
ψ_u^2 (age, sex, age:sex)	0.00	2.15	0.96	0.91
log-likelihood	−242.73	−224.33	−239.72	−240.59
AIC	497.45	460.65	491.44	493.19
Model 2 – `covariance = "pdCompSymm"`				
intercept	23.73(0.71)	24.92 (0.62)	26.33 (0.67)	24.97 (0.30)
age	0.74 (0.35)	0.64 (0.10)	0.76 (0.17)	0.78 (0.25)
sex	−1.66 (0.82)	−1.93 (0.80)	−2.50 (1.02)	−2.32 (0.54)
age:sex	−0.32 (0.54)	−0.14 (0.30)	−0.31 (0.25)	−0.30 (0.49)
ψ_u^2 (diagonal)	1.47	2.20	1.23	0.90
ϕ (off-diagonal)	0.01	0.06	0.06	0.01
log-likelihood	−230.76	−223.97	−237.60	−240.59
AIC	475.51	461.94	489.20	495.18
Model 3 – `covariance = "pdDiag"`				
intercept	24.49 (0.91)	25.24 (0.67)	26.25 (0.62)	24.97 (0.46)
age	0.75 (0.10)	0.75 (0.09)	0.75 (0.09)	0.78 (0.08)
sex	−1.99 (0.94)	−2.24 (0.77)	−2.94 (0.88)	−2.32 (0.78)
age:sex	−0.25 (0.14)	−0.25 (0.12)	−0.31 (0.11)	−0.30 (0.13)
ψ_1^2 (intercept)	2.91	2.12	3.81	2.92
ψ_2^2 (age)	0.06	0.00	0.03	0.03
ψ_3^2 (sex)	2.90	2.15	3.36	1.09
ψ_4^2 (age:sex)	0.00	0.00	0.00	0.00
log-likelihood	−209.62	−201.43	−205.70	−216.30
AIC	437.24	420.86	429.41	450.60
Model 4 – `covariance = "pdDiag"`				
intercept	24.75 (0.89)	25.23 (0.75)	26.24 (0.71)	24.97 (0.51)
age	0.75 (0.12)	0.73 (0.09)	0.75 (0.11)	0.78 (0.10)
sex	−2.13 (0.86)	−2.30 (1.07)	−2.86 (0.99)	−2.32 (0.76)
age:sex	−0.37 (0.15)	−0.28(0.12)	−0.37 (0.15)	−0.30 (0.12)
ψ_1^2 (intercept)	3.82	2.82	3.73	3.44
ψ_2^2 (age)	0.00	0.00	0.00	0.02
log-likelihood	−210.71	−203.97	−207.20	−205.76
AIC	435.42	421.95	428.39	429.52

Table 2: Fitted models of pituitary-pterygomaxillary fissure distance conditional on age (baseline: 11 years) and sex (baseline: boys) for three quartiles and the mean (LMM). Standard errors are in parentheses.

Model 2, which allows random effects to be correlated, yet imposes the same variance parameters, does not offer an improvement with respect to Model 1, except for $\tau = 0.25$.

Models 3 and 4, in contrast, show substantially lower AIC values, and these are similar between the two models at all quartiles. Note also that, as compared to the first two models, Models 3 and 4 provide a better AIC value for the mean. Furthermore, the AIC for Model 4 (429.52) is lower than that for Model 3 (450.60). This improvement is explained by the heteroscedastic component of the model. Without it, Model 4 provides an AIC value of 448.58. The estimates of the fixed slopes for age, in either boys or girls, are approximately constant across quartiles. The estimated variance parameters also indicate that the random intercepts vary considerably between subjects, slightly less for $\tau = 0.5$ as compared to the other two quartiles and the mean. In contrast, there is little or null variation between subject-specific slopes. The conclusion that can be drawn based on the orthodontic growth data is that there is no evidence of unequal growth rates in terms of pituitary-pterygomaxillary fissure distance for either boys or girls ranking below and above the median. There seems to be, however, a reduced level of heterogeneity near the median.

The analysis can be further extended to assess individual growth trajectories. Each panel in Figure 4 plots individual measurements for boys and girls at different ages, together with quartile regression lines predicted at the population level (that is, `level = 0`) using Model 4 (Table 2). Note that the slopes differ by sex but not by subject. This type of plot provides centile curves analogous to those typically used for screening purposes. For example, the growth paths of subjects 'F01' and 'F10' lie below the first quartile at all ages. A plot like the one in Figure 3, on the other hand, can be used for 'conditional' screening (Wei and He 2006). Note that in the conditional plot the measurements for subjects 'F01' and 'F10' rank within the first and third conditional quartiles or very close to them.

7. Quantile regression for independent data

The package **lqmm** deals mainly with clustered data. However, some functions are also provided to estimate quantile regression models for independent data via maximization of the AL likelihood with a gradient-search algorithm. See also the package **quantreg** (Koenker 2013) for estimation based on linear programming algorithms.

The main function `lqm` has the typical syntax of R regression fitting functions, with some arguments analogous to those in `lqmm`:

```
R> args(lqm)
```

```
function (formula, data, subset, na.action, weights = NULL, tau = 0.5,
    contrasts = NULL, control = list(), fit = TRUE)
```

Methods and functions for objects of type `lqm` include `coefficients`, `predict`, `summary`, `logLik`, `AIC`, and `boot.lqm`.

Finally the function `lqm.counts` implements the quantile methods for independent count data proposed by Machado and Santos Silva (2005), using `lqm.fit.gs` as computing engine.

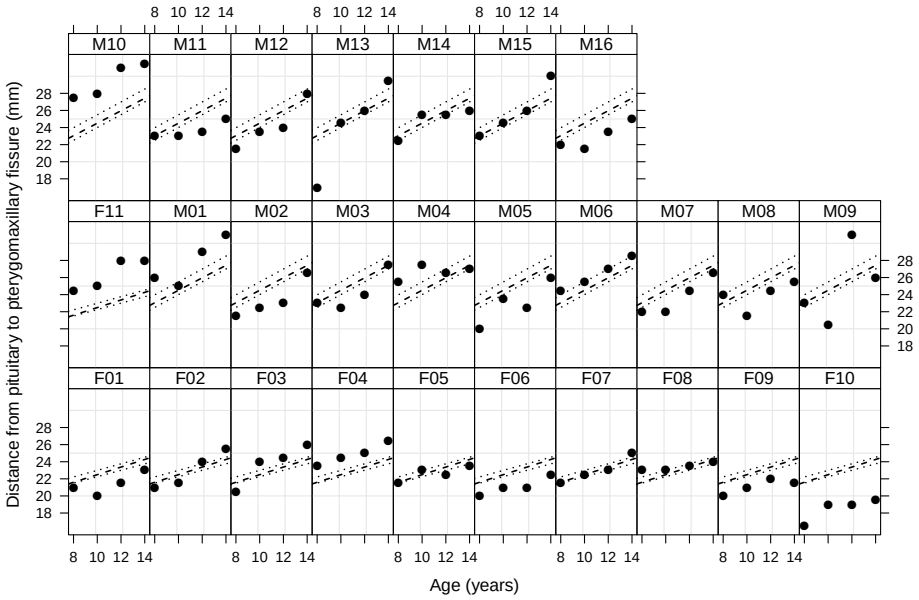

Figure 4: Trellis plot of predicted marginal curves (`level = 0`) of pituitary-pterygomaxillary fissure distance in boys and girls (median, dashed line; first and third quartiles, dotted lines).

8. Conclusion

The **lqmm** package implements methods developed by Geraci and Bottai (2013) for conditional quantile estimation with clustered data, originally proposed by Geraci (2005). The R code is written in S3-style, while main fitting procedures are coded in C. Ongoing methodological work in LQMMs includes developing approaches with a reduced computational burden for both model's estimation and standard error calculation. Future extensions of the **lqmm** package will also provide functions for complex survey estimation and multiple imputation as proposed by Geraci (2013a).

Acknowledgments

The Centre for Paediatric Epidemiology and Biostatistics benefits from funding support from the Medical Research Council in its capacity as the MRC Centre of Epidemiology for Child Health (G0400546). The UCL Institute of Child Health receives a proportion of funding from the Department of Health's NIHR Biomedical Research Centres funding scheme. I would like to thank an anonymous referee for their helpful comments and suggestions on an earlier version of the article.

References

Abrevaya J (2001). "The Effects of Demographics and Maternal Behavior on the Distribution of Birth Outcomes." *Empirical Economics*, **26**(1), 247–257.

Bates D, Maechler M, Bolker B (2013). ***lme4****: Linear Mixed-Effects Models Using S4 Classes*. R package version 0.999999-2, URL http://CRAN.R-project.org/package=lme4.

Booth JG, Hobert JP (1999). "Maximizing Generalized Linear Mixed Model Likelihoods with an Automated Monte Carlo EM Algorithm." *Journal of the Royal Statistical Society B*, **61**, 265–285.

Canay IA (2011). "A Simple Approach to Quantile Regression for Panel Data." *Econometrics Journal*, **14**(3), 368–386.

Demidenko E (2004). *Mixed Models. Theory and Applications*. John Wiley & Sons, Hoboken, NJ.

Dempster AP, Laird NM, Rubin DB (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical B*, **39**(1), 1–38.

Farcomeni A (2012). "Quantile Regression for Longitudinal Data Based on Latent Markov Subject-Specific Parameters." *Statistics and Computing*, **22**(1), 141–152.

Genz A, Keister BD (1996). "Fully Symmetric Interpolatory Rules for Multiple Integrals over Infinite Regions with Gaussian Weight." *Journal of Computational and Applied Mathematics*, **71**(2), 299–309.

Geraci M (2005). *Prediction in Semiparametric and Nonparametric Modelling with Random Effects*. Doctoral dissertation, University of Florence, Italy.

Geraci M (2013a). "Estimation of Regression Quantiles in Complex Surveys with Data Missing at Random: An Application to Birthweight Determinants." *Statistical Methods in Medical Research*. doi:10.1177/0962280213484401.

Geraci M (2013b). ***lqmm****: Linear Quantile Mixed Models*. R package version 1.04, URL http://CRAN.R-project.org/package=lqmm.

Geraci M, Bottai M (2007). "Quantile Regression for Longitudinal Data Using the Asymmetric Laplace Distribution." *Biostatistics*, **8**(1), 140–154.

Geraci M, Bottai M (2013). "Linear Quantile Mixed Models." *Statistics and Computing*. doi:10.1007/s11222-013-9381-9.

Heiss F, Winschel V (2008). "Likelihood Approximation by Numerical Integration on Sparse Grids." *Journal of Econometrics*, **144**(1), 62–80.

Hinkley DV, Revankar NS (1977). "Estimation of the Pareto Law from Underreported Data: A Further Analysis." *Journal of Econometrics*, **5**(1), 1–11.

Koenker R (2004). "Quantile Regression for Longitudinal Data." *Journal of Multivariate Analysis*, **91**(1), 74–89.

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

Koenker R (2013). ***quantreg**: Quantile Regression*. R package version 4.98, URL http://CRAN.R-project.org/package=quantreg.

Koenker R, Bassett G (1978). "Regression Quantiles." *Econometrica*, **46**(1), 33–50.

Koenker R, Hallock KF (2001). "Quantile Regression." *Journal of Economic Perspectives*, **15**(4), 143–156.

Koenker R, Machado JAF (1999). "Goodness of Fit and Related Inference Processes for Quantile Regression." *Journal of the American Statistical Association*, **94**(448), 1296–1310.

Laird NM, Ware JH (1982). "Random-Effects Models for Longitudinal Data." *Biometrics*, **38**(4), 963–974.

Lee D, Neocleous T (2010). "Bayesian Quantile Regression for Count Data with Application to Environmental Epidemiology." *Journal of the Royal Statistical Society C*, **59**(5), 905–920.

Lee Y, Nelder JA (2004). "Conditional and Marginal Models: Another View." *Statistical Science*, **19**(2), 219–228.

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." *Journal of the Royal Statistical Society C*, **46**(4), 463–476.

Machado JAF, Santos Silva JMC (2005). "Quantiles for Counts." *Journal of the American Statistical Association*, **100**(472), 1226–1237.

Pinheiro J, Bates D (1995). "Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model." *Journal of Computational and Graphical Statistics*, **4**(1), 12–35.

Pinheiro J, Bates D (2000). *Mixed-Effects Models in S and S-PLUS*. Springer-Verlag, New York, NY.

Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2013). ***nlme**: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-109, URL http://CRAN.R-project.org/package=nlme.

Pinheiro JC, Chao EC (2006). "Efficient Laplacian and Adaptive Gaussian Quadrature Algorithms for Multilevel Generalized Linear Mixed Models." *Journal of Computational and Graphical Statistics*, **15**(1), 58–81.

Potthoff RF, Roy SN (1964). "A Generalized Multivariate Analysis of Variance Model Useful Especially for Growth Curve Problems." *Biometrika*, **51**(3/4), 313–326.

Rabe-Hesketh S, Skrondal A, Pickles A (2002). "Reliable Estimation of Generalized Linear Mixed Models Using Adaptive Quadrature." *The Stata Journal*, **2**(1), 1–21.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Reich BJ, Bondell HD, Wang HJ (2010). "Flexible Bayesian Quantile Regression for Independent and Clustered Data." *Biostatistics*, **11**(2), 337–352.

Sarkar D (2008). *Lattice: Multivariate Data Visualization with R*. Springer-Verlag, New York, NY.

Schäfer J, Opgen-Rhein R, Zuber V, Ahdesmäki M, Silva APD, Strimmer K (2013). ***corpcor**: Efficient Estimation of Covariance and (Partial) Correlation*. R package version 1.6.6, URL http://CRAN.R-project.org/package=corpcor.

Smyth G, with contributions from Yifang Hu and Peter Dunn and Belinda Phipson (2013). ***statmod**: Statistical Modeling*. R package version 1.4.17, URL http://CRAN.R-project.org/package=statmod.

Staudenmayer J, Lake EE, Wand MP (2009). "Robustness for General Design Mixed Models Using the t-distribution." *Statistical Modelling*, **9**(3), 235–255.

Wang J (2012). "Bayesian Quantile Regression for Parametric Nonlinear Mixed Effects Models." *Statistical Methods & Applications*, **21**(3), 279–295.

Wei Y, He XM (2006). "Conditional Growth Charts (with Discussion)." *The Annals of Statistics*, **34**(5), 2069–2097.

Ypma J (2012). ***SparseGrid**: Sparse Grid Integration in R*. R package version 0.8.1, URL http://CRAN.R-project.org/package=SparseGrid.

Yu KM, Moyeed RA (2001). "Bayesian Quantile Regression." *Statistics & Probability Letters*, **54**(4), 437–447.

Yu KM, Zhang J (2005). "A Three-Parameter Asymmetric Laplace Distribution and Its Extension." *Communications in Statistics-Theory and Methods*, **34**(9-10), 1867–1879.

Affiliation:

Marco Geraci
Centre for Paediatric Epidemiology and Biostatistics
Institute of Child Health
University College London
London WC1N 1EH, UK
E-mail: m.geraci@ucl.ac.uk
URL: http://marcogeraci.wordpress.com