Linear quantile mixed models*

Marco Geraci
University College London
and
Matteo Bottai
University of South Carolina and Karolinska Institutet

June 1, 2011

Abstract

Dependent data arise in many studies. For example, children with the same parents or living in neighboring geographic areas tend to be more alike in many characteristics than individuals chosen at random from the population at large; observations taken repeatedly on the same individual are likely to be more similar than observations from different individuals. Frequently adopted sampling designs, such as cluster, multilevel, spatial, and repeated measures (or longitudinal or panel), may induce this dependence, which , the analysis of the data needs to take into due account. In a previous publication (Geraci and Bottai, *Biostatistics* 2007), we proposed a conditional quantile regression model for continuous responses where a random intercept was included along with fixed-coefficient predictors to account for between-subjects dependence in the context of longitudinal data analysis. Conditional on the random intercept, the response was assumed to follow an asymmetric Laplace distribution. The approach hinged upon the link existing between the minimization of weighted least absolute deviations, typically used in quantile regression, and the maximization of a Laplace likelihood. As a follow up to that study, here we consider an extension of those models to more complex dependence structures in the data, which are modeled by including multiple random effects in the linear conditional quantile functions. Differently from the Gibbs

sampling expectation-maximization approach proposed previously, the estimation of the fixed regression coefficients and of the random effects covariance matrix is based on a combination of Gaussian quadrature approximations and optimization algorithms. The former include Gauss-Hermite and Gauss-Laguerre quadratures for, respectively, normal and doubleexponential (i.e., symmetric Laplace) random effects; the latter include a modified compass search algorithm and general purpose optimizers. As a result, some of the computational burden associated with large Gibbs sample sizes is avoided. We also discuss briefly an estimation approach based on generalized Clarkes derivatives. Finally, a simulation study is presented and some preliminary results are shown.

1 Introduction

Conditional quantile regression (QR) pertains the estimation of unknown quantiles of an outcome as a function of a set of covariates and a vector of fixed regression coefficients. QR estimation, generally, makes no assumption on the shape of the distribution of the outcome (Koenker and Bassett, 1978). Their capability to provide a rich description of the distributional effects at play without investing too much effort in looking for and fitting the 'right' probability model for the data, has contributed to make QR models attractive in several fields. See for example Koenker (2005) and Yu et al. (2003) for an overview of recent applications.

In the last few years, the need for extending the capabilities of QR for independent data to deal with clustered sampling designs has led to several and quite distinct approaches. These can be roughly classified into two groups: distribution–free and likelihood–based. The former include fixed effects (Koenker, 2005; Lamarche, 2010a,b) and weighted (Lipsitz et al., 1997; Karlsson, 2008) approaches. The latter are mainly based on the asymmetric Laplace (AL) density (Geraci and Bottai, 2007; Liu and Bottai, 2009; Yuan and Yin, 2010; Lee and Neocleous, 2010) or other parametric distributions (Reich et al., 2010a; Tzavidis et al., 2010). These categories are by no means mutually exclusive nor the subdivision is exhaustive. For example, penalty methods as those proposed by Koenker (2004) might have a strict relationship with the asymmetric Laplace regression with double–exponential random effects as suggested by Geraci and Bottai (2007). Yet, this has not been fully explored. Also, other approaches might involve modeling the moments of a distribution function using a parametric family (e.g., Rigby and Stasinopoulos, 2005) and, by inversion, deriving the quantiles of the response. Here, we just stress that the spirit of a likelihood–based

approach to QR is different from a model fitting exercise. In the first case, a probability distribution "wrapped" around a kernel function that either is the loss function of a classical QR problem (Yu and Moyeed, 2001; Geraci and Bottai, 2007) or some approximation of it (Tzavidis et al., 2010) allows for maximum likelihood estimation of the regression parameters; in the other case, a distribution function is fitted to the data by modeling, generally, location and variability or, via higher moments, other features such as skewness and kurtosis (Rigby and Stasinopoulos, 2005).

In Section 2, we briefly review the AL–based approach of Geraci and Bottai (2007) and we introduce a generalization of the model proposed therein. In Section 3, we describe an estimation process based on numerical integration and nonsmooth optimization. A simulation study is offered in Section 4. All computations were performed using the package `lqmm` (Geraci, 2011) for the statistical programming environment R (R Development Core Team, 2010).

2 Linear quantile mixed models

Random-effects models with asymmetric Laplace error

A continuous random variable $y \in \mathbb{R}$ is said to follow an asymmetric Laplace density with parameters (μ, σ, τ), $y \sim \text{AL}(\mu, \sigma, \tau)$, if its density can be expressed as

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

where $\rho_\tau(v) = v(\tau - I(v < 0)) = 0.5|v| + 0.5(2\tau - 1)v$ is the check or loss function, $I(\cdot)$ is the indicator function, $0 < \tau < 1$ is the skew parameter, $\sigma > 0$ is the scale parameter and $-\infty < \mu < +\infty$ is the location parameter. See Yu and Zhang (2005) for more details on this distribution. For a random variable y composed of n independent variates y_i with common skew and scale parameters, but generic location $\mu = (\mu_1, \ldots, \mu_n)$, $y_i \sim \text{AL}(\mu_i, \sigma, \tau)$, $i = 1, \ldots, n$, we use the simplified notation

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

where $\sigma_n(\tau) = \tau^n(1-\tau)^n/\sigma^n$ and $\rho_\tau\left(y-\mu\right) = \sum_{i=1}^n \rho_\tau\left(y_i-\mu_i\right)$ denotes the sum of the values taken by the function ρ_τ at each element of its argument.

In a previous paper (Geraci and Bottai, 2007) (GB07 hereafter) we proposed a random intercept QR model for longitudinal data using the AL for

the conditional response. We assumed the conditional quantile regression function

$$\mathrm{F}^{-1}_{y|u}(\tau|x,u) = X\beta + u,$$

where (y, X) represents longitudinal data, u a vector of subject-specific random effects and F^{-1} denotes the inverse of the unknown distribution function of $y|u$. The τ-th regression quantile of $y|u$ was then estimated under the convenient assumption $y|u \sim \mathrm{AL}(X\beta + u, \sigma, \tau)$, where the parameters β and σ have a frequentist interpretation. A Bayesian nature to this model, however, has been erroneously attributed by others (Reich et al., 2010b).

The link between the weighted absolute deviations minimization problem and the maximum likelihood estimation of the regression coefficients β has been described elsewhere (Koenker and Machado, 1999; Yu and Moyeed, 2001).

Model generalization

In this section we extend our model to include multiple nested random effects. As notational standard we will use 1_n to denote the $n \times 1$ vector of ones and I_n to denote the $n \times n$ identity matrix. Unless the specification of the dimension is essential to the intelligibility of the formulas, the subscript will be omitted.

Consider clustered data in the form $(x'_{ij}, z'_{ij}, y_{ij})$, for $j = 1, \ldots, n_i$ and $i = 1, \ldots, M$, $N = \sum_i n_i$, where x'_{ij} is the ith row of a known $n_i \times p$ matrix X_i, z'_{ij} is the ith row of a known $n_i \times q$ matrix Z_i and y_{ij} is the jth observation of the ith cluster. Mixed models (MM) represent a common and well-known class of regression models used to analyze data coming from similar designs. A typical linear formulation of a MM for clustered data is given by

$$y_i = x'_{ij}\beta + z'_{ij}u_i + \epsilon_{ij}, j = 1, \ldots, n_i, i = 1, \ldots, M$$

where β and u_i, $i = 1, \ldots, M$, are, respectively, fixed and random effects associated to p and q model covariates and y is assumed to follow a multivariate normal distribution characterized by some parameter θ. Within this framework, the target of the analysis is the mean of the response, whether conditional, i.e. $E(y_i|u)$ or marginal, i.e. $E(y_i)$, will depend on the purpose of the analysis. Note that the use of the terms marginal and conditional refers to the manipulation of the likelihood object for the MM specified above that is, from a modeling standpoint, conditional. Throughout the paper, we will use such distinction in the same context. For a discussion

about conditional and marginal modeling see for example Lee and Nelder (2004).

There are several reasons why an analyst might consider a random-effects approach for their data and, yet, be wanting to go beyond the usual assumptions of a mixed model for the mean:

- The error distribution is not well approximated by a Gaussian bell.
- There are data points that have a substantial leverage on the least square estimate of the model.
- The mean is not a sufficient summary of the complex distributional effects exerted by the covariates or
- it is not a meaningful location for the distribution of the response (e.g., the distribution is skewed).
- If heteroscedasticity is present, this requires additional modeling effort and reduces the degrees of freedom in the data.

Similar grounds were offered by GB07 when considering a quantile regression approach to longitudinal designs. Here we follow the same approach but we provide a more general and computationally efficient framework within which to model and to estimate regression quantiles of a clustered continuous outcome.

In our model, τ's value is fixed depending on the τth conditional quantile to be estimated, thus it will be omitted when writing out conditional distributions. We assume that the y_i's, $y_i = (y_{11}, \ldots, y_{1n_i})'$, $i = 1, \ldots, M$, conditionally on a $q \times 1$ vector of random effects u_i, are independently distributed according to a joint AL with location and scale parameters given by $\mu_i = X_i\theta_x + Z_i u_i$ and σ, where $\theta_x \in \mathbb{R}^p$ is a vector of unknown fixed effects. Also, we assume that $u_i = (u_{i1}, \ldots, u_{iq})'$, for $i = 1, \ldots, M$, is a random vector independently distributed according to $p(u_i|\Psi)$, where Ψ is a $q \times q$ covariance matrix. We assume that the random effects are zero-median vectors. Other scenarios may include the case in which u are not zero–centered (see for example Lamarche, 2010a; Koenker, 2005, p.281) and/or are not symmetric. We will briefly resume this issue further on.

If we let $u = (u_1', \ldots, u_M')'$, $y = (y_1', \ldots, y_M')'$, $X = [X_1'| \ldots |X_M']'$ and $Z = \bigoplus_{i=1}^{M} Z_i$, $\mu = X\theta_x + Zu$, the joint density of (y, u) based on M clusters for the linear quantile mixed model (LQMM) is given by

$$p(y, u|\theta_x, \sigma, \Psi) = p(y|\theta_x, \sigma, u)p(u|\Psi) = \prod_{i=1}^{M} p(y_i|\theta_x, \sigma, u_i)p(u_i|\Psi). \qquad (2.1)$$

Throughout the paper we will assume that $\Psi \in S^q_{++}$, where S^q_{++} is the set of real symmetric positive-definite $q \times q$ matrices. Note that all the parameters in equation (2.1) are τ-dependent.

The formulation of the quantile regression with random effects in (2.1) is general. For $q = 1$, $Z_i = 1_{n_i}$, $i = 1, \ldots, M$, the LQMM corresponds to the model in GB07. We obtain the i-th contribution to the marginal likelihood by integrating out the random effects, leading to

$$L_i(\theta_x, \sigma, \Psi | y_i) = \int_{R^q} p(y_i, u_i | \theta_x, \sigma, \Psi) du_i, \tag{2.2}$$

where R^q denotes the q-dimensional Euclidean space. We denote the marginal log-likelihood with $\ell_i(\theta_x, \sigma, \Psi | y) = \log L_i(\theta_x, \sigma, \Psi | y)$, $i = 1, \ldots, M$.

3 Estimation

In GB07 we proposed to estimate the parameter of the random intercept QR model by using a MCEM algorithm to avoid the evaluation of a multidimensional integral. This approach on the one hand simplifies the analytical impasse, but on the other hand, like any MCMC method it can be computationally burdensome.

The generalization of the QR model with random intercepts to account for more complex structures of the random effect vector seem to bring additional difficulties to the maximization of the marginal likelihood in (2.2). In this study we will explore alternative computational techniques based on different optimization approaches.

The integral we want to estimate for the marginal distribution of y_i in model (2.1) can be written as

$$p_y(y_i | \theta_x, \sigma, \Psi) = \sigma_{n_i}(\tau) \int_{R^q} \exp\left\{ -\frac{1}{\sigma} \rho_\tau \left(y_i - \mu_i \right) \right\} p(u_i | \Psi) du_i, \tag{3.1}$$

where the subscript y is used to avoid confusion with the conditional distribution of $y|u$.

It is useful to introduce Theorem 6 of Prékopa (1973) in order to characterize the density in (3.1).

Theorem (Prékopa, 1973)**.** *Let $f(x, y)$ be a function of $p + q$ variables where x is an p-component and y is an q-component vector. Suppose that f is logarithmic concave in $\mathbb{R}^{p+q}$ and let A be a convex subset of $\mathbb{R}^q$. Then the function of the variable x:*

$$\int_A f(x, y) dy$$

is logarithmic concave in the entire space $\mathbb{R}^p$. $\square$

It follows that, if $p(u_i|\Psi)$ is log–concave in u, the integrand function in (3.1) will be log–concave in y and Prékopa's Theorem will apply to $p_y(y_i|\theta_x, \sigma, \Psi)$. As we shall see, the random-effects distributions considered in this study are log–concave. Extension of the the log-concavity property to the likelihood function must be applied with caution for the joint log–likelihood is not a concave function of the scale and variance parameters, unless it undergoes a (simple) parametric transformation.

3.1 Numerical integration

The integral in equation (3.1) has the form $\int_{\mathbb{R}^q} f(u)w(u) \prod_q du_q$. By choosing a suitable weighting function or kernel $w(u)$ and upon a change of the integration variable where necessary, the q-dimensional Gaussian quadrature formula, based on q successive applications of simple one–dimensional rules, provides the approximation

$$\int_{\mathbb{R}^q} f(u)w(u)du \approx \sum_{k_1=1}^{K} \cdots \sum_{k_q=1}^{K} f(v_{k_1,\ldots,k_q}) \prod_{l=1}^{q} w_{k_l},$$

where K is a given integer. The abscissas $v_{k_1,\ldots,k_q} = (v_{k_1}, \ldots, v_{k_q})'$ and the weights w_{k_l}, $k_l = 1, \ldots, K$, $l = 1, \ldots, q$, are chosen so that the approximation is exact if $f(u)$ is a polynomial of a given total order. More precisely, the product rule defined above would be exact for a tensor product of univariate polynomials.

For this reason the product rule entails a 'curse of dimensionality', an exponential increase of the number of evaluations of the integrand function. For example, we would need 3,200,000 function evaluations for a 5-dimensional quadrature rule based on 20 nodes but 64,000,000 for adding only one random effect, let alone the total number of evaluations necessary to convergence if the parameter's estimation algorithm is iterative (as it will be seen to be the case). A possible relief from such cumbersome computational burden is offered by integration on sparse grids (Heiss and Winschel, 2008). In the next sections we will provide formulas for the standard Gaussian quadrature and refer the reader to Appendix for a skecth of Heiss and Winschel's (2008) idea.

The choice of an appropriate distribution for the random effects u is not straightforward. In our previos publication, we recognized that robustness issues might apply not only to the error model but also to the random effects.

As a robust alternative to the Gaussian choice, we suggested the use of the symmetric Laplace. This choice led to a regression model which, after a simple transformation (Geraci and Bottai, 2007, p.146), was similar to the penalized model proposed by Koenker (2004).

In the following, we will focus explicitly on two types of distributions, namely Gaussian and Laplacian. It is immediate to verify that these choices correspond to applying, respectively, a Gauss-Hermite and a Gauss-Laguerre quadrature to the integral in (3.1). More general considerations can be done with regard to the use of symmetric as well as asymmetric kernels belonging to the exponential family. Also, we will briefly introduce and discuss a possible adaptive approach to the Gaussian quadrature.

Normal random effects

Under the assumption of normal random effects, the approximation of the integral in equation (3.1) by Gauss-Hermite quadrature is, bar a proportionality constant, given by

$$\begin{aligned}
&\int_{R^q} (2\pi)^{-q/2} |\Psi|^{-1/2} \exp\left\{-\frac{1}{\sigma}\rho_\tau \left(y_i - X_i\theta_x - Z_i u_i\right)\right\} \exp\left(-u_i'\Psi^{-1}u_i/2\right) du_i \\
&= \int_{R^q} (2\pi)^{-q/2} \exp\left\{-\frac{1}{\sigma}\rho_\tau \left[y_i - X_i\theta_x - Z_i\left(\Psi^{T/2}v\right)\right]\right\} \exp\left(-\|v\|^2/2\right) dv \\
&\simeq \sum_{k_1=1}^{K} \cdots \sum_{k_q=1}^{K} \exp\left\{-\frac{1}{\sigma}\rho_\tau \left[y_i - X_i\theta_x - Z_i\left(\Psi^{T/2}v_{k_1,\ldots,k_q}\right)\right]\right\} \times \prod_{l=1}^{q} w_{k_l},
\end{aligned}$$

with nodes $v_{k_1,\ldots,k_q} = (v_{k_1}, \ldots, v_{k_q})'$ and weights w_{k_l}, $l = 1, \ldots, q$.

The (marginal) log-likelihood for all clusters is approximated by

$$\begin{aligned}
\ell_{\text{app}}(\theta_x, \sigma, \Psi|y) = \sum_{i}^{M} \log \Bigg\{ &\sum_{k_1=1}^{K} \cdots \sum_{k_q=1}^{K} p\left(y_i|\theta_x, \sigma, \Psi^{T/2}v_{k_1,\ldots,k_q}\right) \\
&\times \prod_{l=1}^{q} w_{k_l} \Bigg\}.
\end{aligned} \tag{3.2}$$

The integral above can be recognized as a normal-Laplace convolution (Reed, 2006). This type of distribution is known in a special form in meta-analysis (Demidenko, 2004). The temptation to follow an approach based on the (closed form) likelihood of this distribution is strong. However, we desist from pursuing such attempt here as it will require further investigation.

Robust random effects

We consider independent random effects, i.e. $\Psi = \text{diag}(\psi_1, \ldots, \psi_q)$ under the assumption of symmetric Laplace distribution. Since each one-dimensional integral in (3.1) can be split around zero, the Gauss-Laguerre quadrature can be applied to the interval $[0, \infty)$ and, by symmetry, to the interval $(-\infty, 0]$. This results in the following approximation

$$\begin{aligned}
&\int_{R^q} \left(\prod_{l=1}^{q} \frac{1}{4\psi_l} \right) \exp\left\{ -\frac{1}{\sigma} \rho_\tau \left(y_i - X_i\theta_x - Z_i u_i \right) \right\} \exp\left(-\sum_{l=1}^{q} |u_{il}|/2\psi_l \right) du_i \\
&= \int_{R^q} 4^{-q} \exp\left\{ -\frac{1}{\sigma} \rho_\tau \left[y_i - X_i\theta_x - Z_i \left(\Psi^T v \right) \right] \right\} \exp\left(-\sum_{l=1}^{q} |v_l|/2 \right) dv \\
&\simeq \sum_{k_1=1}^{K} \cdots \sum_{k_q=1}^{K} \exp\left\{ -\frac{1}{\sigma} \rho_\tau \left[y_i - X_i\theta_x - Z_i \left(\Psi^T v_{k_1,\ldots,k_q} \right) \right] \right\} \prod_{l=1}^{q} w_{k_l},
\end{aligned}$$

with nodes $v_{k_1,\ldots,k_q} = (v_{k_1}, \ldots, v_{k_q})'$ and weights w_{k_l}, $l = 1, \ldots, q$, opportunely chosen.

The log-likelihood for all clusters is approximated by

$$\begin{aligned}
\ell_{\text{app}}(\theta_x, \sigma, \Psi | y) = \sum_{i}^{M} \log \Bigg\{ &\sum_{k_1=1}^{K} \cdots \sum_{k_q=1}^{K} p\left(y_i | \theta_x, \sigma, \Psi^T v_{k_1,\ldots,k_q} \right) \\
&\times \prod_{l=1}^{q} w_{k_l} \Bigg\}.
\end{aligned} \tag{3.3}$$

There is an increasing number of papers on variations of the univariate Laplace distribution (Kozubowski and Nadarajah, 2008). A valid multivariate extension of the asymmetric Laplace distribution that imposes a general covariance structure is not straightforward and would require some elaboration. For example, Kotz et al. (2000) proposed a multivariate asymmetric Laplace distribution whose density in the n-variate symmetric case is given by

$$p(y|\Sigma) = 2(2\pi)^{-n/2} |\Sigma|^{-1/2} \left(y'\Sigma^{-1} y/2 \right)^{\lambda/2} K_\lambda \left(\sqrt{2y'\Sigma^{-1} y} \right), \tag{3.4}$$

with $\text{Cov}(y) = \Sigma$, $\lambda = 1 - n/2$ and $K_\lambda(t)$ is the modified Bessel function of the third kind. For $n = 1$ and $\Sigma = \sigma^2$, (3.4) reduces to

$$\frac{1}{2} \frac{\sqrt{2}}{\sigma} e^{-\frac{\sqrt{2}}{\sigma}|y|},$$

that is an AL$(0, \frac{\sigma}{2\sqrt{2}}, 0.5)$. It can be shown that for a $n \times n$ real matrix G, $\mathrm{Cov}(Gy) = G\Sigma G'$ (Kotz et al., 2000).

The use of equation (3.4) for correlated random effects is uncertain. Even though it is easy to re-scale Ψ to a diagonal matrix, the joint density does not factorize into q AL variates (Eltoft et al., 2006) and, therefore, the q–dimensional quadrature can not be based on q successive applications of one–dimensional rules. Therefore, at the moment, we do not consider any of the available proposals for a generalized Laplace distribution as suitable for our purposes. See for example Liu and Bottai (2009) for some results on the use of Kotz et al.'s (2000) multivariate Laplace distribution.

Adaptive quadrature

Generalized linear models and nonlinear mixed models are a typical example of where quadrature-based approximations of high-dimensional integrals provide an efficient and reliable computational strategy. At some extent, the quadrature gaussian rule can be considered as the deterministic equivalent of Monte Carlo integration algorithm; similarly the adaptive quadrature is seen as the efficient counterpart of importance sampling (Pinheiro and Bates, 1995; Pinheiro and Chao, 2006). In normal (linear or nonlinear) mixed models the adaptive gaussian rule is easily implemented by centering and scaling the grid of abscissas using an "importance" distribution that approximates the integrand obtained by applying a second-order Taylor expansion around the conditional modes of the random effects.

In this section, we consider an adaptation of (3.2) and (3.3) when centering the grid of abscissas v around the conditional (on Ψ) modes of u_i rather than 0. Although this is a topic of current research, we provide a brief sketch of the idea.

The adaptive quadrature rule for the Gaussian and the Laplace models is simply obtained by centering the abscissas v, scaled by the relevant variance matrix, with the conditional modes

$$\hat{u}_i = \arg\min_{u_i} \frac{1}{\sigma} \rho_\tau \left(y_i - X_i \theta_x - Z_i u_i \right) + u_i' \Psi^{-1} u_i / 2$$

or

$$\hat{u}_i = \arg\min_{u_i} \frac{1}{\sigma} \rho_\tau \left(y_i - X_i \theta_x - Z_i u_i \right) + \sum_{l=1}^{q} |u_{il}| / 2\psi_l$$

respectively, that is $\hat{u}_i + \Psi^{1/2} v_{k_1,\ldots,k_q}$ or $\hat{u}_i + \Psi v_{k_1,\ldots,k_q}$. This approach is particularly useful if the u_i's are not assumed to be zero–median random vectors.

The scaling of the random effects could be ameliorated when using a variance matrix consistent with the scale of the integrand or a linear approximation of it. This is particularly true if the variance of the random effects is large (see for example Heiss and Winschel, 2008). The lack of (strict) differentiability of the loss function in a LQMM prevents us from pursuing an approach similar, say, to that described by Pinheiro and Bates (1995). Nonetheless, it would be reasonable to introduce a smooth approximation of ρ_τ and go from there.

For example, one could consider the smooth function used in Chen and Wei (2005)

$$\kappa_{\omega,\tau}(v) = \begin{cases} v(\tau-1) - \frac{1}{2}(\tau-1)^2\omega & \text{if } v \leq (\tau-1)\omega, \\ \frac{v^2}{2}\omega & \text{if } (\tau-1)\omega \leq v \leq \tau\omega, \\ v\tau - \frac{1}{2}\tau^2\omega & \text{if } v \geq \tau\omega, \end{cases}$$

where $v \in \mathbb{R}$, $\omega > 0$ is a scalar "tuning" parameter and with the understanding that $\kappa_{\omega,\tau}(v) = \sum_{i=1}^n \kappa_{\omega,\tau}(v_i)$ if $v \in \mathbb{R}^n$.

Let us define the function

$$g\left(\theta_x, \Psi, \sigma, y_i, u_i\right) = \frac{1}{\sigma}\rho_\tau\left(y_i - X_i\theta_x - Z_i u_i\right) + u_i'\Psi^{-1}u_i/2$$

and its smooth approximation

$$h\left(\theta_x, \Psi, \sigma, y_i, u_i\right) = \frac{1}{\sigma}\kappa_{\omega,\tau}\left(y_i - X_i\theta_x - Z_i u_i\right) + u_i'\Psi^{-1}u_i/2.$$

In place of Ψ in the scaling of the v, we use the Hessian of h

$$H_i \equiv H\left(\theta_x, \Psi, \sigma, y_i\right) = \left.\frac{\partial h\left(\theta_x, \Psi, \sigma, y_i, u_i\right)}{\partial u_i \partial u_i'}\right|_{u_i = \hat{u}_i}$$

evaluated at the modes $\hat{u}_i = \arg\min_{u_i} h\left(\theta_x, \Psi, \sigma, y_i, u_i\right)$.

Let $v_{k_1,\ldots,k_q} = (v_{k_1}, \ldots, v_{k_q})' =$ and w_{k_l}, $l = 1, \ldots, q$ be the nodes and the weights of the Gauss-Hermite quadrature rule. Under the assumption of normal random effects, a possible adaptive Gaussian quadrature rule is given by

$$\begin{aligned} &\int_{R^q} (2\pi)^{-q/2}\left|\Psi\right|^{-1/2} \exp\left\{-\frac{1}{\sigma}\rho_\tau\left(y_i - X_i\theta_x - Z_i u_i\right)\right\} \exp\left(-u_i'\Psi^{-1}u_i/2\right) du_i \\ &= \int_{R^q} (2\pi)^{-q/2}\left|H_i\Psi\right|^{-1/2} \exp\left\{-g\left(\theta_x, \Psi, \sigma, y_i, \hat{u}_i + H_i^{-1/2}v\right) + \|v\|^2/2\right\} \\ &\quad \times \exp\left(-\|v\|^2/2\right) dv \\ &\simeq \left|H_i\Psi\right|^{-1/2} \sum_{\mathbf{k}} \exp\left\{-g\left(\theta_x, \Psi, \sigma, y_i, \tilde{v}_{i\mathbf{k}}\right)\right\} \tilde{w}_{\mathbf{k}} \end{aligned}$$

where $\tilde{v}_{i\mathbf{k}} = \hat{u}_i + H_i^{-1/2} v_{k_1,\ldots,k_q}$ and $\tilde{w}_{\mathbf{k}} = \exp\left(\|v_{\mathbf{k}}\|^2/2\right) \prod_{l=1}^{q} w_{k_l}$.

The (marginal) log-likelihood for all clusters is approximated by

$$\ell_{\text{app}}(\theta_x, \sigma, \Psi|y) = \log \sigma_N(\tau) - \frac{M}{2} \log|\Psi| - \frac{1}{2} \sum_{i=1}^{M} \log|H_i| + \sum_{i}^{M} \log \left\{ \sum_{\mathbf{k}} \exp\left[-g\left(\theta_x, \Psi, \sigma, y_i, \tilde{v}_{i\mathbf{k}}\right)\right] \tilde{w}_{\mathbf{k}} \right\}.$$

Under Laplacian random effects, one could smooth the kernel of u using κ again

$$h\left(\theta_x, \Psi, \sigma, y_i, u_i\right) = \frac{1}{\sigma} \kappa_{\omega_y,\tau}\left(y_i - X_i\theta_x - Z_i u_i\right) + \kappa_{\omega_u,0.5}\left(u_i\right),$$

however the feasibility and benefit of such approach in this case are unclear at this stage of the research.

Note that if the smooth function h was too used for approximating ρ in the exponential argument, it would be possible to expand the objective function with a Laplace (second-order) approximation, which, at least in the case of a generalized linear MM, corresponds the one point adaptive Gaussian quadrature (Pinheiro and Bates, 1995).

3.2 Nonsmooth or subgradient optimization

The nondifferentiability of the loss function ρ_τ at points where $y_{ij} - x'_{ij}\theta_x - z'_{ij}\left(\Psi^{1/2} v\right) = 0$ interferes with the standard theory of smooth optimization. Subgradient optimization and derivative–free optimization techniques (e.g., coordinate and pattern–search methods, modified Nelder–Mead methods, implicit filtering) have been developed to tackle nonstandard optimization problems. In Geraci and Bottai (2007), we considered the Clarke's derivatives as a viable approach to nonsmooth analysis.

In his book, first published in 1983, Clarke (1990) developed a general theory of nonsmooth analysis that leads to a powerful and elegant approach to mathematical programming. Here, we focus our attention on theorems for Lipschitz functions which play an important role in Clarke's treatise.

We begin characterizing the (approximated) marginal likelihood as a Lipschitz function. Let θ_z denotes the vector of m non-redundant elements of the matrix $\Psi^{1/2}$. Consider the i-th contribution to the likelihood in (3.2), rewritten as

$$\ell_{\text{app},i}\left(\theta, \sigma\right) = \log \sigma_{n_i}(\tau) + \log \sum_{\mathbf{k}} \exp\left\{ -\frac{1}{\sigma} \rho_\tau \left(y_i - \tilde{X}_{i,\mathbf{k}} \theta \right) \right\} \prod_{l=1}^{q} w_{k_l} \quad (3.5)$$

where $\tilde{X}_{i,\mathbf{k}} = [X_i | (v'_{\mathbf{k}} \otimes Z_i) T_q]$ has row vectors $\tilde{x}'_{ij,\mathbf{k}}$, $v_{\mathbf{k}}$ is a $q \times 1$ vector of nodes, $\theta = (\theta'_x, \theta'_z)'$. T_q is a matrix of order $q^2 \times m$ so that vec $\left(\Psi^{T/2}\right) = T_q \theta_z$ (Gauss-Hermite) or vec $\left(\Psi^{T}\right) = T_q \theta_z$ (Gauss-Laguerre). Since the likelihood function is strictly differentiable with respect to σ, we focus on θ alone for brevity.

It can be noted that $\ell_{\text{app},i}$, as a function of θ, is a real-valued function given by the composition $g \circ h$, where $h : \mathbb{R}^{p+m} \to \mathbb{R}^{n_i K^q}$ and $g : \mathbb{R}^{n_i K^q} \to \mathbb{R}$. Each component function $h_{j,\mathbf{k}}$ of h,

$$h_{j,\mathbf{k}}(\theta) = \rho_\tau \left\{ y_{ij} - \tilde{x}'_{ij,\mathbf{k}} \theta \right\} / \sigma$$

is Lipschitz near θ, so is $g\left(h(\theta)\right) = \log \sigma_{n_i}(\tau) + \log \sum_{\mathbf{k}} \exp\left\{ - \sum_j h_{j,\mathbf{k}}(\theta) \right\} \cdot \prod_{l=1}^{q} w_{k_l}$ near $h(\theta)$.

Then, we calculate the generalized gradient for Lipschitz functions (Clarke, 1990). We have that (i) $-g$ is convex (i.e., g is concave), thus regular at $h(\theta)$, (ii) each $h_{j,\mathbf{k}}$ is regular at θ, and (iii) every element λ of $\partial(-g)\left(h(\theta)\right)$ has nonnegative components

$$\lambda_{r,\mathbf{h}} = \partial(-g)\left(h(\theta)\right)_{r,\mathbf{h}} = \frac{\prod_{l=1}^{q} w_{h_l} \exp\left\{ -h_{r,\mathbf{h}}(\theta) - \sum_{j \neq r} h_{j,\mathbf{h}}(\theta) \right\}}{\sum_{\mathbf{k}} \exp\left\{ -\frac{1}{\sigma} \rho_\tau \left(y_i - \tilde{X}_{i,\mathbf{k}} \theta \right) \right\} \prod_{l=1}^{q} w_{k_l}},$$

for $r = 1, \ldots, n_i$ and $\mathbf{h} = (h_1, \ldots, h_q)'$, $h_l = 1, \ldots, K$, $l = 1, \ldots, q$.

For the chain rule I (Clarke, 1990, Theorem 2.3.9, p.42), it follows that

$$\partial(-g)\left(h(\theta)\right) = \mathcal{C}\left\{ \sum \lambda_{j,\mathbf{k}} \xi_{j,\mathbf{k}} : \xi_{j,\mathbf{k}} \in \partial h_{j,\mathbf{k}}(\theta), \lambda \in \partial(-g)\left(h(\theta)\right) \right\},$$

where the summation is extended to all j's and $\mathbf{k}$'s and $\mathcal{C}$ denotes the weak*-closed convex hull. For the chain rule II (Clarke, 1990, Theorem 2.3.10, p.45)

$$\partial h_{j,\mathbf{h}}(\theta) = \begin{cases} -\dfrac{\tau}{\sigma} \tilde{x}_{ij,\mathbf{h}} & \text{if } y_{ij} - \tilde{x}'_{ij,\mathbf{h}} \theta > 0 \\ -\dfrac{(\tau - 1)}{\sigma} \tilde{x}_{ij,\mathbf{h}} & \text{if } y_{ij} - \tilde{x}'_{ij,\mathbf{h}} \theta < 0 \\ -\left(\omega + \tau - \dfrac{1}{2} \right) \tilde{x}_{ij,\mathbf{h}} : |\omega| \leq 1 & \text{if } y_{ij} - \tilde{x}'_{ij,\mathbf{h}} \theta = 0 \end{cases}$$

If a local minimum or maximum is attained at $\hat{\theta}$, solution to the minimization problem

$$\min_{\theta} \left\{ - \sum_{i}^{M} \ell_{\text{app},i}(\theta, \sigma) \,|\theta_x \in \mathbb{R}^p, \Psi(\theta_z) \in S^q_{++} \right\},$$

then $0 \in \partial - \sum_i^M \ell_{\text{app},i}$ (Clarke, 1990). On a practical level, the constraint $\Psi(\theta_z) \in S^q_{++}$ can be imposed *a posteriori* by calculating the nearest symmetric positive definite matrix (Higham, 2002) after θ_z is estimated or, *a priori*, by a reparameterization of Ψ (Pinheiro and Bates, 1996; Pourahmadi, 1999). Equality and inequality constraints can be accommodated by the Lagrangian rule provided by Theorem 6.1.1 in Clarke (1990, p.228).

Following the subgradient approach of Rockafellar (1970), Koenker (2005) gives optimality conditions of the quantile regression problem for independent data.

3.3 Interpretation of parameters

Consider the classical random intercept model

$$\begin{aligned} y|u &= \mu + Zu + \varepsilon \\ u &\sim N\left(0, \psi_u^2 I\right) \\ \varepsilon &\sim N\left(0, \psi^2 I\right), \end{aligned}$$

where $\varepsilon \perp u$, with marginal distribution $y \sim N\left(\mu, \psi^2 I + \psi_u^2 ZZ'\right)$.

The parameters of such model have a straightforward interpretation: μ is the mean effect at the population level, ψ_u^2 is a measure of the dispersion of the cluster–specific random effects and related to the intra–cluster correlation γ, $\gamma = \psi_u^2/(\psi^2 + \psi_u^2)$, and ψ^2 is the 'white' noise.

Each cluster have a conditional distribution $N\left(\mu + u_i, \psi^2 I\right)$, therefore the 'atomic' τ-th quantile is $q_{ij}(\tau) \equiv q_{i\cdot}(\tau) = \mu + \psi\Phi^{-1}(\tau) + u_i$, where Φ^{-1} denotes the inverse of the cumulative distribution function of a standard normal. Conditionally on u_i, the y_{ij}'s are independent. Thus, the τ-th sample quantile $\hat{q}_{i\cdot}(\tau)$ is an estimator of $q_{i\cdot}(\tau)$, which, asymptotically,

$$\hat{q}_{i\cdot}(\tau) \dot{\sim} N\left(q_{i\cdot}(\tau), \frac{\tau(1-\tau)}{n_i \left[p_{y_{i\cdot}|u_i}(q_{i\cdot}(\tau))\right]^2}\right).$$

Note that this result is valid for any $0 < \tau < 1$ and continuous p. If we took the average of M such estimators we would obtain, for a large M,

$$\frac{1}{M}\sum_i^M \hat{q}_{i\cdot}(\tau) \dot{\sim} N\left(\bar{q}(\tau), \nu\right)$$

where $\bar{q}(\tau) = \frac{1}{M}\sum_i q_{i\cdot}(\tau) \approx \mu + \psi\Phi^{-1}(\tau)$ and $\nu = \frac{1}{M^2}\sum_i \frac{\tau(1-\tau)}{n_i\left[p_{y_{i\cdot}|u_i}(q_{i\cdot}(\tau))\right]^2} = \frac{1}{M^2}\sum_i \frac{\tau(1-\tau)2\pi\psi^2 e^{\Phi^{-1}(\tau)}}{n_i}$. It is intended that, under the above random in-

tercept model, the approximation of the mean is valid on average (that is, $E_u\{\bar{q}(\tau)\}$).

Let us now turn to the linear quantile mixed models. Our starting model is conditional. The marginal likelihood (3.1), estimated under models (3.2) or (3.3), implicitly assumes that the τ-th regression quantiles of the clusters 'gravitate' around a common regression quantile which, in the case of the normal intercept model, would be $\mu + \psi\Phi^{-1}(\tau)$, clearly different from the τ-th quantile of the marginal model $\mu + \sqrt{\psi^2 + \psi_u^2}\Phi^{-1}(\tau)$. Oberhofer and Haupt's (2005) showed that the unconditional quantile estimator for dependent random variables is asymptotically unbiased.

The scale parameter σ does not have, in general, a straightforward interpretation since the use of the Laplace distribution for the conditional response responds to the need for a likelihood approach to quantile regression rather than to the observation that the data is effectively Laplacian. But what if it is? Consider the linear median mixed model ($\tau = 0.5$). The (asymptotic) variance of the quantile estimator of $q_{i\cdot}(0.5)$ based on n_i observations is $\frac{\tau(1-\tau)}{n_i\left[\frac{\tau(1-\tau)}{\sigma}\right]^2} = \frac{4\sigma^2}{n_i}$ and the variance of a variable $y \sim \mathrm{AL}(\mu, \sigma, \tau)$ is given by $\mathrm{var}(y) = \frac{\sigma^2(1-2\tau+2\tau^2)}{(1-\tau)^2\tau^2}$ (Yu and Zhang, 2005). It follows that $\mathrm{var}(y) = 8\sigma^2$ for $\tau = 0.5$. The relative efficiency of this estimator under Laplacian and normal hypotheses is then $\frac{4\sigma^2}{n_i}/\frac{\pi\psi^2}{2n_i} = 8\sigma^2/\pi\psi^2$. The median estimator would achieve the same asymptotic efficiency under the two distributions if $\sigma \approx 0.62\psi$ or, equivalently in terms of variances, if $8\sigma^2 = \pi\psi^2$.

4 Simulation study

The data were generated according to

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_i + \epsilon_{ij}$$

where $\beta = (100, 2)$, $x_{ij} = \delta_i + \zeta_{ij}$, $\delta_i \sim N(0,1)$, $\zeta_{ij} \sim N(0,1)$. Random intercepts, u, and error terms, ϵ, were independently drawn from a $N(0,5)$ (N), a Student's t_3 (t), and a χ^2_2 (X). The number of replicated datasets is denoted with R. The Gauss-Hermite quadrature ("normal") was used. The quantiles $\tau \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$ were estimated. The estimation method was either derivative free ("df") or gradient search ("gs").

The relative absolute bias $|\hat{\beta}_i - \tilde{\beta}_i|/\tilde{\beta}_i$ and the standard deviation $\sqrt{var(\hat{\beta}_i)}$, $i = 0, 1$, are reported in Tables 1–12 (the first letter of the distribution combination refers to u, the second to ϵ). The 'true' β's, $\tilde{\beta}$, were calculatated

as follows: $\tilde{\beta}_0(\tau) = 100 + F_\epsilon^{-1}(\tau)$, where $F_\epsilon^{-1}()$ is the quantile function of ϵ, and $\tilde{\beta}_1(\tau) = 2$, for all τ's (i.e., homoscedastic model) .

quantile	N-N	t-N	N-t	t-t	N-X	t-X
0.10	0.0008	0.0005	0.0053	0.0024	0.0036	0.0011
0.25	0.0012	0.0003	0.0028	0.0017	0.0002	0.0012
0.50	0.0004	0.0004	0.0007	0.0004	0.0039	0.0030
0.75	0.0012	0.0006	0.0034	0.0022	0.0043	0.0028
0.90	0.0020	0.0001	0.0064	0.0045	0.0043	0.0025

Table 1: Relative bias for β_0. Method = df, n = 50, m = 5, R = 300.

quantile	N-N	t-N	N-t	t-t	N-X	t-X
0.10	0.6037	0.4989	0.6163	0.4875	0.5753	0.4170
0.25	0.5254	0.3608	0.5595	0.3645	0.5022	0.3415
0.50	0.4826	0.3382	0.5381	0.3602	0.5107	0.3610
0.75	0.4905	0.3523	0.5259	0.3353	0.5316	0.3645
0.90	0.5760	0.4501	0.5820	0.5447	0.6671	0.6350

Table 2: Std deviation for β_0. Method = df, n = 50, m = 5, R = 300.

quantile	N-N	t-N	N-t	t-t	N-X	t-X
0.10	0.0038	0.0020	0.0007	0.0029	0.0028	0.0009
0.25	0.0002	0.0026	0.0016	0.0008	0.0009	0.0025
0.50	0.0062	0.0006	0.0020	0.0024	0.0005	0.0021
0.75	0.0006	0.0030	0.0057	0.0037	0.0003	0.0038
0.90	0.0148	0.0088	0.0024	0.0026	0.0033	0.0093

Table 3: Relative bias for β_1. Method = df, n = 50, m = 5, R = 300.

quantile	N-N	t-N	N-t	t-t	N-X	t-X
0.10	0.2344	0.2044	0.1753	0.1623	0.1179	0.1074
0.25	0.1925	0.1751	0.1377	0.1199	0.1198	0.1046
0.50	0.1775	0.1577	0.1227	0.1052	0.1277	0.1208
0.75	0.1877	0.1752	0.1338	0.1153	0.1774	0.1746
0.90	0.2120	0.2063	0.1740	0.1544	0.2697	0.2704

Table 4: Std deviation for β_1. Method = df, n = 50, m = 5, R = 300.

quantile	N-N	t-N	N-t	t-t	N-X	t-X
0.10	0.0006	0.0006	0.0029	0.0030	0.0013	0.0021
0.25	0.0003	0.0005	0.0010	0.0012	0.0009	0.0007
0.50	0.0002	0.0001	0.0005	0.0002	0.0020	0.0016
0.75	0.0004	0.0004	0.0019	0.0015	0.0021	0.0018
0.90	0.0006	0.0006	0.0032	0.0037	0.0019	0.0018

Table 5: Relative bias for β_0. Method = df, n = 300, m = 10, R = 500.

quantile	N-N	t-N	N-t	t-t	N-X	t-X
0.10	0.3273	0.2949	0.3536	0.3063	0.3221	0.2867
0.25	0.2806	0.1738	0.3241	0.2317	0.3307	0.2348
0.50	0.2714	0.1712	0.3136	0.2291	0.3166	0.1907
0.75	0.2638	0.1936	0.3001	0.2453	0.2897	0.1975
0.90	0.3122	0.3030	0.3274	0.3940	0.3383	0.3360

Table 6: Std deviation for β_0. Method = df, n = 300, m = 10, R = 500.

quantile	N-N	t-N	N-t	t-t	N-X	t-X
0.10	0.0003	0.0016	0.0032	0.0002	0.0005	0.0002
0.25	0.0003	0.0013	0.0007	0.0001	0.0006	0.0010
0.50	0.0011	0.0014	0.0004	0.0004	0.0008	0.0008
0.75	0.0000	0.0006	0.0006	0.0004	0.0020	0.0008
0.90	0.0020	0.0005	0.0016	0.0005	0.0021	0.0014

Table 7: Relative bias for β_1. Method = df, n = 300, m = 10, R = 500.

quantile	N-N	t-N	N-t	t-t	N-X	t-X
0.10	0.0717	0.0764	0.0607	0.0562	0.0353	0.0347
0.25	0.0598	0.0565	0.0429	0.0392	0.0377	0.0338
0.50	0.0557	0.0488	0.0381	0.0341	0.0405	0.0376
0.75	0.0586	0.0571	0.0441	0.0387	0.0624	0.0545
0.90	0.0773	0.0765	0.0583	0.0552	0.0947	0.0970

Table 8: Std deviation for β_1. Method = df, n = 300, m = 10, R = 500.

quantile	N-N	t-N	N-t	t-t	N-X	t-X
0.10	0.0016	0.0009	0.0044	0.0034	0.0035	0.0026
0.25	0.0004	0.0004	0.0014	0.0012	0.0005	0.0006
0.50	0.0001	0.0001	0.0002	0.0001	0.0019	0.0016
0.75	0.0003	0.0005	0.0021	0.0014	0.0023	0.0018
0.90	0.0015	0.0006	0.0044	0.0035	0.0026	0.0017

Table 9: Relative bias for β_0. Method = gs, n = 300, m = 10, R = 500.

quantile	N-N	t-N	N-t	t-t	N-X	t-X
0.10	0.3032	0.2707	0.3215	0.2770	0.3441	0.3586
0.25	0.2733	0.1658	0.3268	0.2236	0.3275	0.2356
0.50	0.2551	0.1638	0.3164	0.2120	0.3054	0.1790
0.75	0.2586	0.1844	0.3119	0.2355	0.2745	0.1912
0.90	0.2973	0.5716	0.3193	0.2884	0.3129	0.4666

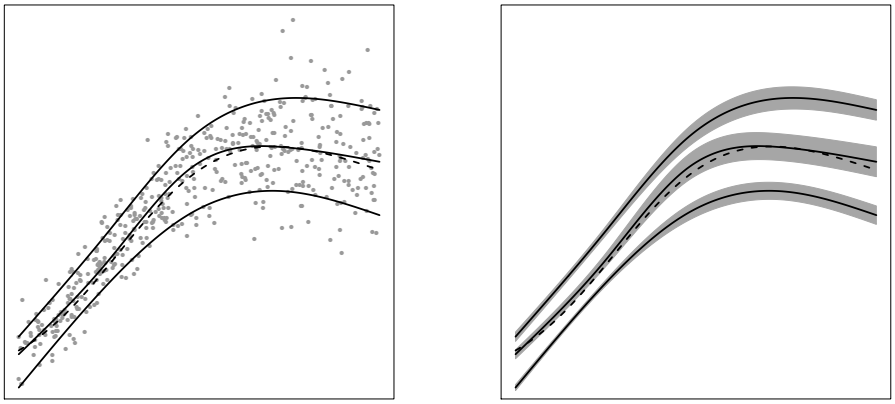
Table 10: Std deviation for β_0. Method = gs, n = 300, m = 10, R = 500.

quantile	N-N	t-N	N-t	t-t	N-X	t-X
0.10	0.0009	0.0005	0.0023	0.0005	0.0018	0.0007
0.25	0.0001	0.0011	0.0002	0.0003	0.0010	0.0014
0.50	0.0008	0.0011	0.0010	0.0004	0.0005	0.0009
0.75	0.0001	0.0007	0.0000	0.0002	0.0016	0.0010
0.90	0.0004	0.0004	0.0029	0.0011	0.0030	0.0004

Table 11: Relative bias for β_1. Method = gs, n = 300, m = 10, R = 500.

quantile	N-N	t-N	N-t	t-t	N-X	t-X
0.10	0.0736	0.0764	0.0587	0.0553	0.0386	0.0363
0.25	0.0601	0.0567	0.0436	0.0392	0.0379	0.0337
0.50	0.0561	0.0494	0.0383	0.0346	0.0401	0.0375
0.75	0.0592	0.0576	0.0448	0.0393	0.0622	0.0550
0.90	0.0777	0.0778	0.0598	0.0559	0.0948	0.0974

Table 12: Std deviation for β_1. Method = gs, n = 300, m = 10, R = 500.



(a) (b)

Figure 1: Smoothed regression quantiles ($\tau \in \{0.1, 0.5, .9\}$) with data points (a) and 95% confidence bands (b). The true mean function is represented by a dashed line.

5 Special cases

5.1 Smoothing splines

We show a simple application of the LQMMs presented in Section 2. We generated $n = 10$ observations clustered within $M = 50$ groups using the nonlinear heteroscedastic model

$$y_{ij} = \exp\left\{\frac{5x_{ij}}{1 + 2x^2}\right\} + u_i + x_{ij}\epsilon_{ij}, \tag{5.1}$$

where $x_{ij} \sim U(0, 1)$, $u_i \sim N(0, .25)$ and $\epsilon_{ij} \sim N(0, 1)$, independently. With such model, the within-group correlation $cor\left(y_{ij}, y_{ij'}\right)$, $j \neq j'$, will vary in the range 0.2–1, depending on x_{ij}.

We approximated the nonlinear function in (5.1) with a natural cubic spline model with four degrees of freedom, break-points placed at the quartiles of x and boundary knots at the extremes of x's range. The B-spline basis matrix S was then included in the quantile mixed model $y = S\beta + u + \epsilon$. Figure 1 shows a summary of the regression models fitted with a 9-knot Gauss-Hermite quadrature for three quantiles ($\tau \in \{0.1, 0.5, .9\}$) and 95% confidence bands estimated by using 50 bootstrap replications. Due to the

simmetry of the error term, the estimated median and the (true) mean are very close. However, the spread of the distribution differs across the values of x.

5.2 Spatial modelling

Spatial modelling refers, loosely, to the case in which geographical information is introduced in the model. We might consider the case in which x is a vector of geographical coordinates or, more simply, when the grouping factor is some geographical unit (e.g., postal codes, wards or counties). In the latter case, each random term u of the LQMM would be associated to a small-area effect and the resulting variance–covariance matrix would be interpreted as a spatial correlation matrix at the quantile of interest. The modelling of Ψ, therefore, should take into account the spatial association between areas (e.g., contiguity). See for example Lee and Neocleous (2010) for an application of Bayesian quantile regression (Yu and Moyeed, 2001) to environmental epidemiology using the results of Machado and Silva (2005).

Quantile smoothing of surfaces and related inference is a recent topic (He et al., 1998; He and Portnoy, 2000; Koenker and Mizera, 2004). An application of triogram smoothing splines to poverty mapping is described by Geraci and Salvati (2007). A Bayesian quantile modeling of ozone concentration surfaces is given by Reich et al. (2010b).

Suppose we want to estimate the quantiles of an outcome under the following nonparametric model

$$y_i = f(x_i) + \epsilon_i$$

where x_i' is a vector of geographical coordinates. We could introduce a bivariate smoother for f as, for example, a generalized covariance function with quadratic penalty and adopt an approach as described in Ruppert et al. (2003) or a triogram smoothing spline model (Koenker and Mizera, 2004) with L_1 penalized coefficients. Differently from the classical L_2 penalty used for estimating nonparametric mean functions, L_1 penalization (Koenker et al., 1994) has a more direct and natural application to quantiles from a modeling and estimation standpoint. However, in both circumstances, the random effects would play the role of spline coefficients and the type of quadrature to be used would naturally follow the metric of the associated penalty term.

APPENDIX

Consider the integration problem as in section 3.1 and let $G_k = \sum_{v \in \mathbb{V}_k} f(v) \cdot w(v)$ be the univariate Gaussian quadrature rule with accuracy level k where $\mathbb{V}_k$ is a set of nodes and $w(\cdot)$ is a weighting function (e.g., Hermite or Laguerre). The product rule for q-variate Gaussian quadrature with accuracy level k is given by the tensor product of q univariate rules

$$T_{q,k} = (G_k \otimes \ldots \otimes G_k).$$

Following the seminal work of Smolyak (1963), Heiss and Winschel (2008) proposed an integration rule based on the difference of successive univariate quadrature rules $D_k = G_k - G_{k-1}$, with $G_0 = 0$ and $k \in \mathbb{N}$. Given a sequence of Gaussian quadrature rules G_k such that each $\mathbb{V}_k$ used by G_k has k nodes, the Smolyak rule with accuracy level $K \in \mathbb{N}$ for q-dimensional Gaussian quadrature is defined as

$$S_{q,K} = \sum_{m=0}^{K-1} \sum_{\mathbf{k} \in \mathbb{N}_m^q} \left(D_{k_1} \otimes \ldots \otimes D_{k_q}\right) \tag{A.1}$$

where $\mathbb{N}_m^q = \left\{\mathbf{k} \in \mathbb{N}^q : \sum_{l=1}^q k_l = q + m\right\}$ and $\mathbf{k} = (k_1, \ldots, k_q)'$. For accuracy level $K = 1$, $S_{q,1} = T_{q,1}$. As K grows to infinity, the number of nodes in the set $\mathbb{V}_{q,K} = \bigcup_{m=K-q}^{K-1} \bigcup_{\mathbf{k} \in \mathbb{N}_m^q} \left(\mathbb{V}_{k_1} \otimes \cdots \otimes \mathbb{V}_{k_q}\right)$ used by the sparse grid rule (A.1) does not increase exponentially as does with the product rule, but only polynomially. Heiss and Winschel (2008) showed that for a given accuracy K and rising q, the logarithm of nodes in $V_{q,K}$ is of order $O(\log(q))$ (Theorem 2).

References

Chen, C. and Wei, Y. (2005), "Computational issues for quantile regression," *Sankhya: The Indian Journal of Statistics*, 67, 399–417.

Clarke, F. H. (1990), *Optimization and Nonsmooth analysis*, Philadelphia: SIAM.

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

Eltoft, T., Kim, T., and Lee, T.-W. (2006), "On the Multivariate Laplace Distribution," *IEE Signal Processing Letters*, 13, 300–303.

Geraci, M. (2011), *lqmm: Linear Quantile Mixed Models*, R package version 1.0.

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

Geraci, M. and Salvati, N. (2007), "The geographical distribution of the consumption expenditure in Ecuador: Estimation and mapping of the regression quantiles," *Statistica Applicata/Italian Journal of Applied Statistics*, 19, 167–183.

He, X. M., Ng, P., and Portnoy, S. (1998), "Bivariate quantile smoothing splines," *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 60, 537–550.

He, X. M. and Portnoy, S. (2000), "Some asymptotic results on bivariate quantile splines," *Journal of Statistical Planning and Inference*, 91, 341–349.

Heiss, F. and Winschel, V. (2008), "Likelihood approximation by numerical integration on sparse grids," *Journal of Econometrics*, 144, 62–80.

Higham, N. (2002), "Computing the nearest correlation matrix - a problem from finance," *IMA Journal of Numerical Analysis*, 22, 329–343.

Karlsson, A. (2008), "Nonlinear quantile regression estimation of longitudinal data," *Communications in Statistics-Simulation and Computation*, 37, 114–131.

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

— (2005), *Quantile Regression*, New York: Cambridge University Press.

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

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

Koenker, R. and Mizera, I. (2004), "Penalized triograms: total variation regularization for bivariate smoothing," *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 66, 145–163.

Koenker, R., Ng, P., and Portnoy, S. (1994), "Quantile smoothing splines," *Biometrika*, 81, 673–680.

Kotz, S., Kozubowski, T. J., and Podgórski, K. (2000), "An Asymmetric Multivariate Laplace Distribution," Tech. Rep. 367, Department of Statistics and Applied Probability, University of California at Santa Barbara.

Kozubowski, T. and Nadarajah, S. (2008), "Multitude of Laplace distributions," , 1–22Published Online DOI:10.1007/s00362-008-0127-2.

Lamarche, C. (2010a), "Penalized Quantile Regression Estimation for a Model with Endogenous Individual Effects," Unpublished.

— (2010b), "Robust penalized quantile regression estimation for panel data," *Journal of Econometrics*, 157, 396–498.

Lee, D. and Neocleous, T. (2010), "Bayesian quantile regression for count data with application to environmental epidemiology," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59, 905–920.

Lee, Y. and Nelder, J. A. (2004), "Conditional and marginal models: Another view," *Statistical Science*, 19, 219–228.

Lipsitz, S. R., Fitzmaurice, G. M., Molenberghs, G., and Zhao, L. P. (1997), "Quantile regression methods for longitudinal data with drop-outs: Application to CD4 cell counts of patients infected with the human immunodeficiency virus," *Applied Statistics-Journal of the Royal Statistical Society Series C*, 46, 463–476.

Liu, Y. and Bottai, M. (2009), "Mixed-effects models for conditional quantiles with longitudinal data," *The International Journal of Biostatistics*, 5, 1–22.

Machado, J. A. F. and Silva, J. (2005), "Quantiles for counts," *Journal of the American Statistical Association*, 100, 1226–1237.

Oberhofer, W. and Haupt, H. (2005), "The asymptotic distribution of the unconditional quantile estimator under dependence," *Statistics & Probability Letters*, 73, 243–250.

Pinheiro, J. and Bates, D. (1995), "Approximations to the log-likelihood function in the nonlinear mixed-effects model," *Journal of Computational and Graphical Statistics*, 4, 12–35.

Pinheiro, J. C. and Bates, D. M. (1996), "Unconstrained parametrizations for variance-covariance matrices," *Statistics and Computing*, 6, 289–296.

Pinheiro, J. C. and Chao, E. C. (2006), "Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models," *Journal of Computational and Graphical Statistics*, 15, 58–81.

Pourahmadi, M. (1999), "Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation," *Biometrika*, 86, 677–690.

Prékopa, A. (1973), "Logarithmic concave measures and functions," *Acta Scientiarum Mathematicarum*, 34, 334–343.

R Development Core Team (2010), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Reed, W. (2006), "The normal-Laplace distribution and its relatives," in *Advances in Distribution Theory, Order Statistics, and Inference*, ed. N. Balakrishnan, E. Castillo, J. S. e., New York, NY: Birkhauser Boston, pp. 61–74.

Reich, B. J., Bondell, H. D., and Wang, H. J. (2010a), "Flexible Bayesian quantile regression for independent and clustered data," *Biostatistics*, 11, 337–352.

Reich, B. J., Fuentes, M., and Dunson, D. B. (2010b), "Bayesian Spatial Quantile Regression," *Journal of the American Statistical Association*.

Rigby, R. and Stasinopoulos, D. (2005), "Generalized additive models for location, scale and shape," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54, 507–554.

Rockafellar, R. (1970), *Convex Analysis*, Princeton, NJ: Princeton University Press.

Ruppert, D., Wand, M., and Carroll, R. (2003), *Semiparametric regression*, New York: Cambridge University Press.

Smolyak, S. (1963), "Quadrature and interpolation formulas for tensor products of certain classes of functions," *Soviet Mathematics Doklady*, 4, 240–243.

Tzavidis, N., Salvati, N., Geraci, M., and Bottai, M. (2010), "M-Quantile and expectile random effects regression for multilevel data," Tech. Rep. M10/07, Southampton Statistical Sciences Research Institute.

Yu, K., Lu, Z., and Stander, J. (2003), "Quantile regression: applications and current research areas," *The Statistician*, 52, 331–350.

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

Yu, K. M. and Moyeed, R. A. (2001), "Bayesian quantile regression," *Statistics & Probability Letters*, 54, 437–447.

Yuan, Y. and Yin, G. (2010), "Bayesian quantile regression for longitudinal studies with nonignorable missing data," *Biometrics*, 66, 105–114.