

# Quantile regression for longitudinal data using the asymmetric Laplace distribution

MARCO GERACI\*, MATTEO BOTTAI

*Department of Epidemiology and Biostatistics, University of South Carolina,  
800 Sumter Street, Columbia, SC 29208, USA  
geraci@gwm.sc.edu*

## SUMMARY

In longitudinal studies, measurements of the same individuals are taken repeatedly through time. Often, the primary goal is to characterize the change in response over time and the factors that influence change. Factors can affect not only the location but also more generally the shape of the distribution of the response over time. To make inference about the shape of a population distribution, the widely popular mixed-effects regression, for example, would be inadequate, if the distribution is not approximately Gaussian. We propose a novel linear model for quantile regression (QR) that includes random effects in order to account for the dependence between serial observations on the same subject. The notion of QR is synonymous with robust analysis of the conditional distribution of the response variable. We present a likelihood-based approach to the estimation of the regression quantiles that uses the asymmetric Laplace density. In a simulation study, the proposed method had an advantage in terms of mean squared error of the QR estimator, when compared with the approach that considers penalized fixed effects. Following our strategy, a nearly optimal degree of shrinkage of the individual effects is automatically selected by the data and their likelihood. Also, our model appears to be a robust alternative to the mean regression with random effects when the location parameter of the conditional distribution of the response is of interest. We apply our model to a real data set which consists of self-reported amount of labor pain measurements taken on women repeatedly over time, whose distribution is characterized by skewness, and the significance of the parameters is evaluated by the likelihood ratio statistic.

**Keywords:** Asymmetric Laplace distribution; Clinical trials; Markov Chain Monte Carlo; Quantile regression; Random effects.

## 1. INTRODUCTION

In the last few years, quantile regression (QR) has become a more widely used technique to describe the distribution of a response variable given a set of explanatory variables. Median regression is a special case. In their seminal work, Koenker and Bassett (1978) argued on the omnipresent Gaussian assumption for the model error, and formalized asymptotic properties of the least absolute deviations estimator of the

\*To whom correspondence should be addressed.

QR model for independent observations. A comprehensive guide of the subject is provided by Koenker (2005).

The foundations of the methods for independent data are now consolidated, and some computational commands for the analysis of such data are provided by most of the available statistical software (e.g. R/S-PLUS, SAS, and Stata).

We explore the use of QR for the analysis of longitudinal data. These kinds of data occur in a wide variety of studies, such as clinical trials, panel studies, and epidemiological surveys. The within-subject variability due to the measurements on the same subject should be accounted for to avoid bias in the parameter's estimate. As motivating example, in Section 4 we consider a clinical trial study on the effectiveness of a medication for relieving labor pain. The marginal distribution of the response (Figure 2) is characterized by a marked skewness. In addition, the skewness of the conditional distribution changes sign over time. In such a case, the appropriateness of mean regression should be seriously questioned: it would be highly inefficient to estimate the location, and, more important, entirely inadequate to estimate the quantiles of the conditional distribution of the response.

We propose a method to estimate the parameter of conditional quantile functions with subject-specific, location-shift random effects. This approach is analogous to estimating mean regressions with random intercepts. We introduce a likelihood based on the asymmetric Laplace density (ALD), which provides an automatic choice of the optimal level of penalization. In a simulation study, our strategy outperformed some of the best competing models.

Several works have appeared in the literature that propose methods for repeated measurements data. Koenker (2004) considered the penalized interpretation of the classical random-effects estimator in order to estimate quantile functions with subject-specific fixed effects. The variability induced in the estimation process by these effects was controlled by some forms of shrinkage, or regularization, whose degree and type required a suitable choice of a penalization term. The application's areas include reference charts in medicine (Heagerty and Pepe, 1999; Cole and Green, 1992) and clinical trials (Robins *and others*, 1995; Lipsitz *and others*, 1997). Jung (1996) proposed a quasi-likelihood approach for median regression estimation, where the dependency structure is modeled by the covariance matrix of the indicator functions. Lipsitz *and others* (1997) made use of Jung's work and described the use of QR methods to analyze longitudinal data on CD4 cell counts, and following Robins *and others* (1995) accounted for missing at random dropouts by weighting the generalized estimating equations. They estimated the regression parameters' standard errors by using the bootstrap.

The remainder of this paper proceeds as follows: in Section 2, we briefly present the conditional quantile functions for independent data and the ALD. We then propose a model for repeated measurements along with related inferential procedures. In Section 3, we present the results of a simulation. In Section 4, we apply the methodology to clinical trial data, and, in Section 5, we conclude the paper with some final remarks.

## 2. MODEL AND METHOD

### 2.1 The mean absolute error minimization and the ALD for independent data

Consider data in the form  $(\mathbf{x}_i^T, y_i)$ , for  $i = 1, \dots, N$ , where  $y_i$  are independent scalar observations of a continuous random variable with common cumulative distribution function  $F_y$ , whose shape is not exactly known, and  $\mathbf{x}_i^T$  are row  $p$ -vectors of a known design matrix  $X$ .

The linear conditional quantile functions are defined as

$$G_{y_i}(\tau|\mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta}, \quad i = 1, \dots, N, \quad (2.1)$$

where  $0 < \tau < 1$ ,  $G_{y_i}(\cdot) \equiv F_{y_i}^{-1}(\cdot)$ , and  $\boldsymbol{\beta} \in \mathbb{R}^p$  is a column vector of length  $p$  with unknown fixed parameters.

We refer to the  $\tau$ th regression quantile as any solution  $\beta^*$ ,  $\beta^* \in \mathbb{R}^p$ , to the minimization problem

$$\min_{\beta \in \mathbb{R}^p} V_N(\beta), \quad V_N(\beta) \equiv \sum_{i \in (i: y_i \geq \mathbf{x}_i^T \beta)} \tau |y_i - \mathbf{x}_i^T \beta| + \sum_{i \in (i: y_i < \mathbf{x}_i^T \beta)} (1 - \tau) |y_i - \mathbf{x}_i^T \beta|. \quad (2.2)$$

In order to highlight the  $\tau$ -distributional dependency, the parameter  $\beta$  and the solution  $\beta^*$  should be indexed by  $\tau$ , i.e.  $\beta(\tau)$  and  $\beta^*(\tau)$ . For sake of simplicity, however, we will omit this notation in the remainder of the paper.

A possible parametric link between the minimization of the sum of the absolute deviates in (2.2) and the maximum likelihood theory is given by the ALD. This skewed distribution appeared in the papers by Koenker and Machado (1999) and Yu and Moyeed (2001), among others. We say that a random variable  $Y$  is distributed as an ALD with parameters  $\mu$ ,  $\sigma$ , and  $\tau$ , and we write it as  $Y \sim \text{ALD}(\mu, \sigma, \tau)$ , if the corresponding probability density is given by

$$f(y|\mu, \sigma, \tau) = \frac{\tau(1 - \tau)}{\sigma} \exp \left\{ -\rho_\tau \left( \frac{y - \mu}{\sigma} \right) \right\}, \quad (2.3)$$

where  $\rho_\tau(v) = v(\tau - I(v \leq 0))$  is the loss function,  $I(\cdot)$  is the indicator function,  $0 < \tau < 1$  is the skewness parameter,  $\sigma > 0$  is the scale parameter, and  $-\infty < \mu < +\infty$  is the location parameter. The support of the random variable  $Y$  is the real line. It should be noted that the loss function  $\rho$  assigns weight  $\tau$  or  $1 - \tau$  to the observations greater or, respectively, less than  $\mu$  and that  $\Pr(y \leq \mu) = \tau$ . Therefore, the distribution splits along the scale parameter into two parts, one with probability  $\tau$  to the left, and one with probability  $(1 - \tau)$  to the right. See Yu and Zhang (2005) for further properties and generalizations of this distribution.

Set  $\mu_i = \mathbf{x}_i^T \beta$  and  $\mathbf{y} = (y_1, \dots, y_N)$ . Assuming that  $y_i \sim \text{ALD}(\mu_i, \sigma, \tau)$ , then the likelihood for  $N$  independent observations is, bar a proportionality constant,

$$L(\beta, \sigma; \mathbf{y}, \tau) \propto \sigma^{-1} \exp \left\{ -\sum_{i=1}^N \rho_\tau \left( \frac{y_i - \mu_i}{\sigma} \right) \right\}. \quad (2.4)$$

If we consider  $\sigma$  a nuisance parameter, then the maximization of the likelihood in (2.4) with respect to the parameter  $\beta$  is equivalent to the minimization of the objective function  $V_N(\beta)$  in (2.2). Thus, the ALD proves useful as a unifying bridge between the likelihood and the inference for QR estimation. Koenker and Machado (1999) introduced a goodness-of-fit process for QR and related inference processes, and they also considered the likelihood ratio statistic under the parametric assumption of a Laplacean distribution for the error term. For a full Bayesian approach, see Yu and Moyeed (2001), among others.

## 2.2 Use of random effects for longitudinal data

We now consider repeated measurements data in the form  $(\mathbf{x}_{ij}^T, y_{ij})$ , for  $j = 1, \dots, n_i$  and  $i = 1, \dots, N$ , where  $\mathbf{x}_{ij}^T$  are row  $p$ -vectors of a known design matrix  $X_i$  and  $y_{ij}$  is the  $j$ th scalar measurement of a continuous random variable on the  $i$ th subject.

A possible penalized approach is offered by Koenker (2004), who proposed an  $l_1$  penalty term and considered

$$\min_{\alpha, \beta} \sum_{i=1}^N \sum_{j=1}^{n_i} w \rho_\tau(y_{ij} - \mathbf{x}_{ij}^T \beta - \alpha_i) + \lambda \sum_{i=1}^N |\alpha_i|, \quad (2.5)$$

where the weight  $w$  controls the relative influence of the  $\tau$ th quantile on the estimation of the fixed subject-specific effects  $\alpha_i$ , and  $\lambda$  is the penalization parameter.

The parameter  $\lambda$  represents a fundamental issue. Its value needs to be arbitrarily set, yet its choice heavily influences the inference for the parameter  $\beta$ , as confirmed by the simulation results in Section 3. Besides, following a penalized approach, estimating a  $\tau$ -distributional individual effect would be impracticable. Also, as the sample size increases, so does the dimension of the parameter. We circumvented these drawbacks by letting the location-shift effects be random and the parameter of their distribution be dependent on  $\tau$ .

We define the linear mixed quantile functions of the response  $y_{ij}$

$$G_{y_{ij}|u_i}(\tau|\mathbf{x}_{ij}, u_i) = \mathbf{x}_{ij}^T \beta + u_i, \quad j = 1, \dots, n_i, \quad i = 1, \dots, N, \quad (2.6)$$

where  $G_{y_{ij}|u_i}(\cdot) \equiv F_{y_{ij}|u_i}^{-1}(\cdot)$  is the inverse of the cumulative distribution function of the response conditional on a location-shift random effect  $u_i$ .

We assume that  $y_{ij}$ , conditionally on  $u_i$ , for  $j = 1, \dots, n_i$  and  $i = 1, \dots, N$ , are independently distributed according to the ALD

$$f(y_{ij}|\beta, u_i, \sigma) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\rho_\tau\left(\frac{y_{ij} - \mu_{ij}}{\sigma}\right)\right\}, \quad (2.7)$$

where  $\mu_{ij} = \mathbf{x}_{ij}^T \beta + u_i$  is the linear predictor of the  $\tau$ th quantile. In our model, the conditional quantile to be estimated,  $\tau$ , is fixed and known. The random effects induce a correlation structure among observations on the same subject. We assume that the  $u_i$  are identically distributed according to some density  $f_u$  characterized by a  $\tau$ -dependent parameter  $\varphi$  (i.e.  $\varphi(\tau)$ ), and they are mutually independent. Also, we assume that  $\varepsilon_{ij}$  are independent, and  $u_i$  and  $\varepsilon_{ij}$  are independent of one another. In Section 2.4, we elaborate on the distribution of the random effects.

Let  $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})$  and  $f(\mathbf{y}_i|\beta, u_i, \sigma) = \prod_{j=1}^{n_i} f(y_{ij}|\beta, u_i, \sigma)$  be the density for the  $i$ th subject conditional on the random intercept  $u_i$ . The complete-data density of  $(\mathbf{y}_i, u_i)$ , for  $i = 1, \dots, N$ , is then given by

$$f(\mathbf{y}_i, u_i|\eta) = f(\mathbf{y}_i|\beta, u_i, \sigma)f(u_i|\varphi), \quad (2.8)$$

where  $f(u_i|\varphi)$  is the density of  $u_i$  and  $\eta = (\beta, \sigma, \varphi)$  is the parameter of interest.

If we let  $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)$  and  $\mathbf{u} = (u_1, \dots, u_N)$ , the joint density of  $(\mathbf{y}, \mathbf{u})$  based on  $N$  subjects is given by

$$f(\mathbf{y}, \mathbf{u}|\eta) = \prod_{i=1}^N f(\mathbf{y}_i|\beta, u_i, \sigma)f(u_i|\varphi).$$

### 2.3 Estimation

We obtain the marginal density of  $\mathbf{y}$  by integrating out the random effects, leading to  $f(\mathbf{y}|\eta) = \int_{R^N} f(\mathbf{y}, \mathbf{u}|\eta)d\mathbf{u}$ . Thus, the inference about the parameter  $\eta$  should be based on the marginal likelihood,  $L(\eta; \mathbf{y}) = \sum_i^N L(\eta; \mathbf{y}_i)$ . The latter involves an integral that, in general, does not have a closed form solution.

To estimate  $\eta$ , we propose a Monte Carlo EM algorithm, which has been widely applied to generalized linear mixed models (LMMs) (see for example, Meng and Van Dyk, 1998; Booth and Hobert, 1999). The presence of intractable integrals in the E-step can be overcome, at some computational expenses, by using a sampler. Handling missingness due to nonresponse also represents an attractive feature of such an algorithm (see for example, Ibrahim and others, 1999, 2001).

We write the E-step for the  $i$ th subject at the  $(t + 1)$ th EM iteration as

$$\begin{aligned} Q_i(\eta|\eta^{(t)}) &= E\{l(\eta; \mathbf{y}_i, u_i)|\mathbf{y}_i; \eta^{(t)}\} \\ &= \int \{\log f(\mathbf{y}_i|\beta, u_i, \sigma) + \log f(u_i|\varphi)\} \times f(u_i|\mathbf{y}_i, \eta^{(t)}) du_i, \end{aligned} \quad (2.9)$$

where  $\eta^{(t)} = (\beta^{(t)}, \sigma^{(t)}, \varphi^{(t)})$ ,  $l(\eta; \mathbf{y}_i, u_i)$  is the log likelihood for the  $i$ th observation, and the expectation is taken with respect to the “missing” data  $u_i$ , conditioned on the observed data  $\mathbf{y}_i$ .

The approximation of the expected value of the log likelihood can be carried out with a suitable Monte Carlo technique, which depends on the target density to be sampled.

If the random effects were known and the model in (2.6) is correctly specified, we could define

$$\tilde{V}_i(\beta) = \sum_{j \in (j: \tilde{y}_{ij} \geq \mathbf{x}_{ij}^T \beta)}^{n_i} \tau |\tilde{y}_{ij} - \mathbf{x}_{ij}^T \beta| + \sum_{j \in (j: \tilde{y}_{ij} < \mathbf{x}_{ij}^T \beta)}^{n_i} (1 - \tau) |\tilde{y}_{ij} - \mathbf{x}_{ij}^T \beta|, \quad (2.10)$$

where  $\tilde{y}_{ij} = y_{ij} - u_i$ , and apply a linear programming algorithm to solve

$$\min_{\beta \in R^p} \tilde{V}_N(\beta), \quad \tilde{V}_N(\beta) \equiv \sum_{i=1}^N \tilde{V}_i(\beta), \quad (2.11)$$

since the  $\tilde{y}_{ij}$  values would be no longer dependent.

Since the random effects are unobserved, we propose to take a sample  $\mathbf{v}_i = (v_{i1}, \dots, v_{im_i})$  of size  $m_i$  from the density of the random effects conditional to the response

$$f(u_i|\mathbf{y}_i, \eta) \propto f(\mathbf{y}_i|\beta, u_i, \sigma) f(u_i|\varphi), \quad (2.12)$$

and to approximate the expected value in (2.9) with

$$Q_i^*(\eta|\eta^{(t)}) = \frac{1}{m_i} \sum_{k=1}^{m_i} l(\eta; \mathbf{y}_i, v_{ik}).$$

To obtain the maximum likelihood estimate of the parameter  $\eta$  for the  $\tau$ th quantile model, we applied the following iterative procedure:

**Step 1:** Initialize the parameter  $\eta^{(t)} = (\beta^{(t)}, \sigma^{(t)}, \varphi^{(t)})$ , set  $t = 0$ , and plug  $\eta^{(t)}$  in (2.12);

**Step 2:** Draw a sample  $\mathbf{v}_i^{(t)} = (v_{i1}^{(t)}, \dots, v_{im_i}^{(t)})$  from  $f(u_i|\mathbf{y}_i, \eta^{(t)})$ , for  $i = 1, \dots, N$ , independently, by using a sampling technique, e.g. via the Gibbs sampler with the adaptive rejection sampling algorithm (Gilks and others, 1995);

**Step 3:** Solve

$$\operatorname{argmin}_{\beta \in B} \sum_{i=1}^N \frac{1}{m_i} \sum_{k=1}^{m_i} \sum_{j=1}^{n_i} \rho_\tau(\tilde{y}_{ikj}^{(t)} - \mathbf{x}_{ij}^T \beta),$$

where  $\tilde{y}_{ikj}^{(t)} = y_{ij} - v_{ik}^{(t)}$ , set the solution of the minimization problem equal to  $\beta^{(t+1)}$ , and calculate

$$\sigma^{(t+1)} = \frac{1}{\sum_{i=1}^N n_i} \frac{1}{\sum_{i=1}^N m_i} \sum_{i=1}^N \sum_{k=1}^{m_i} \sum_{j=1}^{n_i} \rho_\tau(\tilde{y}_{ikj}^{(t)} - \mathbf{x}_{ij}^T \beta^{(t+1)})$$

and

$$\varphi^{(t+1)} = \hat{\Phi}(\mathbf{v}_i, \dots, \mathbf{v}_N),$$

where  $\hat{\Phi}$  is the maximum likelihood estimate (estimator) of  $\varphi$ , conditional on  $\mathbf{y}$ ;

**Step 4:** Set  $t = t + 1$  and repeat Steps 1–3 until the parameter  $\eta$  reaches the convergence.

The minimization problem in Step 3 is equivalent to a weighted QR with an offset and thus can be easily solved. Then, at each iteration,  $\eta^{(t+1)}$  is the maximum likelihood estimate of  $\eta$ , given  $\eta^{(t)}$ .

The sampling step, obviously, requires evaluating the Markov chain's convergence (see for example, Cowles and Carlin, 1996) and the sensitivity of the parameter's estimates to different burn-in and sample sizes. Alternative strategies for choosing the size of the Gibbs samples are available. One can consider a constant size  $m_i = m$ , for all  $i$  and each EM iteration, or increase  $m_i$  at each EM iteration.

The availability of a parametric likelihood function allowed us to evaluate a likelihood ratio statistic to test the null hypotheses  $H_0: \beta_l = 0, l = 1, \dots, p$ . A bootstrap method can also be used to obtain an estimate of the standard error of the estimates of  $\beta$  and it can be implemented for longitudinal data by regarding the matrix  $(\mathbf{y}_i, \mathbf{X}_i)$  as the basic resampling unit (Lipsitz *and others*, 1997). Buchinsky (1995) studied the performance of various estimators for the asymptotic variance matrix for QR models and found that the bootstrap procedure yielded the best results, performing well for relatively small samples and also when the data exhibit heteroscedasticity.

## 2.4 Random effects distribution

In Section 1, we argued that the usual choice of the normal distribution for the error model is sometimes questionable and, most important, leaves the estimation out to the influence of the outliers. The use of a joint normal distribution within the mixed models simplifies considerably the analytical form of the expected value of the log likelihood when using the EM algorithm and thus reduces the time to convergence. Robustness issues, however, apply to the random effects as well (see for example, Lange *and others*, 1989; Richardson and Welsh, 1995; Richardson, 1997). For our model, we considered the normal distribution, although it is possible to consider a vast range of symmetric as well as asymmetric distributions such as the asymmetric Laplace. Simulation results are reported in Section 3.

Assume that  $u_i \sim N(0, \varphi)$ , identically and independently for  $i = 1, \dots, N$ . The joint density of  $(\mathbf{y}_i, u_i)$  for the  $i$ th subject is given by

$$\begin{aligned} f(\mathbf{y}_i, u_i | \eta) &= f(u_i | \varphi) \prod_{j=1}^{n_i} f(y_{ij} | \beta, u_i, \sigma) \\ &= \left\{ \frac{\tau(1-\tau)}{\sigma} \right\}^{n_i} \frac{1}{\sqrt{2\pi\varphi}} \exp \left[ - \sum_{j=1}^{n_i} \left\{ \rho_\tau \left( \frac{y_{ij} - \mu_{ij}}{\sigma} \right) \right\} - \frac{1}{2\varphi} u_i^2 \right]. \end{aligned} \quad (2.13)$$

Note that the conditional distribution of the random effect  $f(u_i | \mathbf{y}_i, \eta) \propto f(\mathbf{y}_i, u_i | \eta)$  is log-concave in  $u_i$ , where the definition of log-concavity here is derivative-free (Gilks, 1995), i.e. a function  $g$ , which is a density with respect to Lebesgue measure and its domain  $D$  is an interval of the real line, is defined to be log-concave if

$$\ln g(a) - 2 \ln g(b) + \ln g(c) < 0, \quad \forall a, b, c \in D: a < b < c.$$

Thus, the Gibbs sampling in conjunction with the adaptive rejection sampling can be applied and the Monte Carlo E-step carried out by sampling from  $f(u_i | \mathbf{y}_i, \eta)$ , for each subject  $i, i = 1, \dots, N$ .

The maximum likelihood estimator of  $\varphi$ ,  $\hat{\Phi}$ , at Step 3 of the iterative procedure presented in Section 2.3, is then given by

$$\hat{\Phi} = \frac{1}{N} \frac{1}{\sum_{i=1}^N m_i} \sum_{i=1}^N \sum_{k=1}^{m_i} v_{ik}^2.$$

Assuming that  $u_i \sim \text{ALD}(0, \varphi, 0.5)$ , the joint density of  $(\mathbf{y}_i, u_i)$  for the  $i$ th subject would be

$$\begin{aligned} f(\mathbf{y}_i, u_i | \eta) &= f(u_i | \varphi) \prod_{j=1}^{n_i} f(y_{ij} | \beta, u_i, \sigma) \\ &= \left\{ \frac{\tau(1-\tau)}{\sigma} \right\}^{n_i} \frac{1}{4\varphi} \exp \left[ - \sum_{j=1}^{n_i} \left\{ \rho_{\tau} \left( \frac{y_{ij} - \mu_{ij}}{\sigma} \right) \right\} - \frac{1}{2\varphi} |u_i| \right]. \end{aligned} \quad (2.14)$$

The Gibbs sampling can be applied, since the conditional distribution  $f(u_i | \mathbf{y}_i, \eta)$  derived from the joint density in (2.14) is still log-concave in  $u_i$ . Thus, the maximum likelihood estimator of  $\varphi$  is given by

$$\hat{\Phi} = \frac{1}{N} \frac{1}{\sum_{i=1}^N m_i} \sum_{i=1}^N \sum_{k=1}^{m_i} \rho_{0.5}(v_{ik}).$$

Note also that it is possible to consider  $u_i \sim \text{ALD}(0, \varphi, \tau_u)$  and to estimate the degree of skewness  $\tau_u$  of the random effects in the case the assumption of symmetry of  $u$  is untenable.

Consider the median regression for example, i.e.  $\tau = 0.5$ . If we set  $\lambda = \sigma/\varphi$ , the density in (2.14) turns into

$$f(\mathbf{y}_i, u_i | \eta) = \frac{1}{(4\varphi)^{n_i+1} \lambda^{n_i}} \exp \left[ - \frac{1}{2\sigma} \left\{ \sum_{j=1}^{n_i} (|y_{ij} - \mu_{ij}|) + \lambda |u_i| \right\} \right],$$

where the argument of the exponential has a similar form as in the penalized quantile regression (PQR) in (2.5).

### 3. SIMULATION STUDY

We conducted a simulation study to evaluate the performance of the proposed method for two different quantile functions,  $\tau = 0.25$  and  $\tau = 0.5$ . We used  $N = 100$  subjects and  $n_i = n = 23$ , for  $i = 1, \dots, N$ , subject measurements. To generate the data, we used the location-shift model

$$y_{ij} = 1 + 2x_{ij} + u_i + \varepsilon_{ij}, \quad j = 1, \dots, 23, \quad i = 1, \dots, 100. \quad (3.1)$$

The  $x_{ij}$  values were set equal to the values  $x_{i1} = 1, x_{i2} = 2, \dots, x_{i23} = 23$ , constant throughout the simulation study.

Different probability distributions were considered for the error term: the standard normal,  $N(0, 1)$ , the Student's  $t$  distribution with 3 degrees of freedom,  $t_3$ , and the chi-square with 3 degrees of freedom,  $\chi_3^2$ . The random effects were generated by using a standard normal. For the data  $(\mathbf{y}_i, u_i)$ ,  $i = 1, \dots, 100$ , generated using the model in (3.1), we assumed the joint density in (2.13) and we estimated the parameter  $\eta = (\beta, \sigma, \varphi)$  using the method proposed in Section 2.3. We simulated 1000 replications from each distribution of the error term.

For each replication, the Monte Carlo E-step was carried out via the Gibbs sampler (see Section 2.4) with a constant sample size of  $m_i = m = 5000$ ,  $i = 1, \dots, 100$ , and a burn-in size of 2500.

We compared the quantile regression with random effects (QRRE) with the PQR reported in (2.5), using  $\lambda = 0.01$ ,  $\lambda = 5$ , and  $\lambda \rightarrow \infty$ . The latter is equivalent to the ordinary QR, in which the fixed effects are fully shrunk to zero. The choice of these values was made empirically, since there are no theoretically based or practically accepted guidelines as to how to choose the level of shrinkage. The ratio of the estimated standard deviations of  $\varepsilon_{ij}$  and  $u_i$ , when using the normal and the Student's  $t$  distributions, has been sometimes proposed as reference value for  $\lambda$ .

We also compared the proposed model with the ordinary LMM when the error term distribution was symmetric (i.e. Normal and  $t$ -distribution). When the location model was of interest, comparing the performance of the least squares with the least absolute estimators proved to be useful to evaluate the robustness of our model.

In Table 1, for each estimator's component  $\hat{\beta}_k$ ,  $k = 0, 1$ , and for each model, we reported the estimated relative bias averaged over the simulations,

$$\widehat{\text{bias}}(\hat{\beta}_k) = \frac{1}{1000} \sum_{r=1}^{1000} \frac{\hat{\beta}_k^{(r)} - \beta_k}{|\beta_k|},$$

where  $\hat{\beta}_k^{(r)}$  is the estimated quantile for the  $r$ th replication and  $\beta_k$  is the corresponding true quantile of the conditional distribution  $p(y|u)$ . We also reported the estimated relative efficiency, averaged over the

Table 1. *Estimated bias ( $\widehat{\text{bias}}$ ) and efficiency ( $\widehat{\text{eff}}$ ) for different error term distributions and quantile models: PQR with penalization parameter taking values  $\lambda = 0.01, 5$ , and  $\lambda \rightarrow \infty$  (the latter corresponds to ordinary QR), and QRRE. The LMM estimates are also reported for comparisons with the location model. The results are based on 1000 Monte Carlo replications for each distribution of the error term*

Model	$\tau = 0.25$				$\tau = 0.5$			
	$\hat{\beta}_0$		$\hat{\beta}_1$		$\hat{\beta}_0$		$\hat{\beta}_1$	
	$\widehat{\text{bias}}$	$\widehat{\text{eff}}$	$\widehat{\text{bias}}$	$\widehat{\text{eff}}$	$\widehat{\text{bias}}$	$\widehat{\text{eff}}$	$\widehat{\text{bias}}$	$\widehat{\text{eff}}$
$\varepsilon \sim N(0, 1)$								
QRRE	0.0024	1.0000	0.0000	1.0000	0.0082	1.0000	0.0000	1.0000
PQR ( $\lambda = 0.01$ )	2.0910	1.5093	0.0001	1.1710	0.0000	1.3530	0.0000	1.1618
PQR ( $\lambda = 5$ )	1.6162	1.3565	0.0001	1.2448	0.0003	1.2082	0.0001	1.1801
QR	-0.8421	1.1237	0.0000	1.6501	0.0032	1.0370	0.0000	1.4967
LMM	—	—	—	—	0.0032	0.8798	0.0000	0.7373
$\varepsilon \sim t_3$								
QRRE	-0.1392	1.0000	0.0001	1.0000	0.0043	1.0000	0.0001	1.0000
PQR ( $\lambda = 0.01$ )	2.8362	1.5503	0.0001	1.1655	-0.0003	1.4170	0.0001	1.1166
PQR ( $\lambda = 5$ )	1.9915	1.3361	0.0001	1.2398	-0.0009	1.3242	0.0002	1.2938
QR	-1.3092	1.1212	0.0001	1.4947	-0.0016	1.2037	0.0002	1.7608
LMM	—	—	—	—	-0.0015	1.1105	0.0001	1.9311
$\varepsilon \sim \chi_3$								
QRRE	0.0268	1.0000	-0.0001	1.0000	0.0192	1.0000	0.0000	1.0000
PQR ( $\lambda = 0.01$ )	0.3263	1.6483	-0.0001	1.1748	0.0096	1.3437	0.0000	1.1261
PQR ( $\lambda = 5$ )	0.2543	1.3966	-0.0001	1.2430	0.0128	1.2085	0.0001	1.2317
QR	-0.0188	1.1545	-0.0001	1.6921	0.0385	1.1077	0.0000	1.2487



simulations, using the Monte Carlo variance of the QRRE model in the denominator

$$\widehat{\text{eff}}_{\text{model}}(\hat{\beta}_k) = \frac{S_{\text{model}}^2(\hat{\beta}_k)}{S_{\text{QRRE}}^2(\hat{\beta}_k)},$$

where  $S^2(\hat{\beta}_k) = \frac{1}{1000} \sum_{r=1}^{1000} (\hat{\beta}_k^{(r)} - \bar{\beta}_k)^2$  and  $\bar{\beta}_k = \frac{1}{1000} \sum_{r=1}^{1000} \hat{\beta}_k^{(r)}$ .

The results show clearly that our method generally performs better than the other across the simulated scenarios. The  $l_1$  penalization model performs as well as ours for some values of the penalization parameter, but poorly for others, which is an undesirable situation in practical applications.

The relevance of the bias in the mean squared error is particularly evident for the estimates of the first quartile. A poor choice for the penalization parameter may lead to large bias. In fact, in our simulations the relative bias ranged from  $-0.01$  to  $2.83$ . It can be noted that, as  $\lambda$  increases, the bias of the penalized regression decreases from large positive to large negative values. This suggests that there might be a reasonable value of the penalization parameter that ensures no bias. It can be argued that our model automatically provides a nearly optimal degree of shrinking. When estimating the first quartile, the loss of efficiency of the penalized regressions, with respect to our model, ranged from 12 to 65% and from 17 to 69% for the intercept and the slope parameters, respectively.

The gap in terms of mean squared error between our model and the penalized regression decreased when estimating the median. We observed a smaller loss of efficiency and a leveling off of the relative bias for the symmetric distributions. The bias associated with the random effects QR, except for the case in which the error was  $\chi^2$  distributed, turned out to be slightly higher than the others. The reason might be due to the Monte Carlo error that is less conspicuous when the estimated bias is high as in the case of the first quartile.

As expected, the least-squares estimation outperforms the median regression when the error term is normally distributed, but it yields highly inefficient estimates when the distribution has heavy tails.

A small simulation was also performed on the quantile  $\tau = 0.75$ , and the results (not shown here) were equivalent to the ones obtained for the first quartile.

#### 4. LABOR PAIN DATA: COMPARISONS

The labor pain data were reported by Davis (1991) and successively analyzed by Jung (1996). The data set consists of repeated measurements of self-reported amount of pain on  $N = 83$  women in labor, of which 43 were randomly assigned to a pain medication group and 40 to a placebo group. The response was measured every 30 min on a 100-mm line, where 0 means no pain and 100 means extreme pain. A nearly monotone pattern of missing data was found for the response variable and the maximum number of measurements for each woman was six. In Figure 1, a boxplot shows the median pain over time for the placebo and the medication groups. Some degree of distributional dependence on the temporal course of the quartiles of the response is evident, especially for the placebo group.

Jung (1996) considered the median regression model

$$y_{ij} = \beta_0 + \beta_1 R_i + \beta_2 T_{ij} + \beta_3 R_i T_{ij} + \varepsilon_{ij}, \quad (4.1)$$

where  $y_{ij}$  is the amount of pain for the  $i$ th patient at time  $j$ ,  $R_{ij}$  is the treatment indicator taking 1 for placebo and 0 for treatment, and  $T_{ij}$  is the measurement time divided by 30 min, and  $\varepsilon_{ij}$  is a zero-median error term.

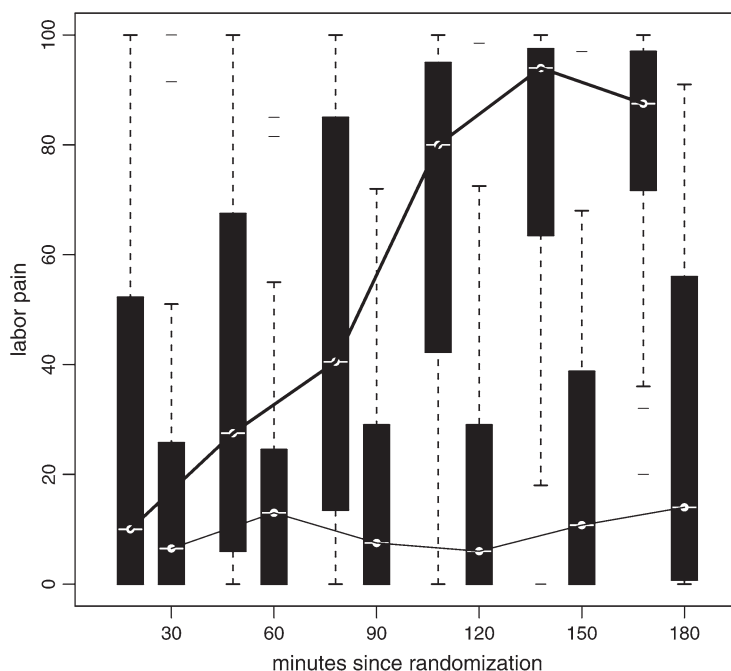


Fig. 1. Boxplot of labor pain score for the placebo (solid, thick line) and the pain medication (solid, thin line) groups.

Let

$$X_i = \begin{bmatrix} 1 & R_i & T_{i1} & R_i T_{i1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & R_i & T_{in_i} & R_i T_{in_i} \end{bmatrix}, \quad \beta = (\beta_0, \beta_1, \beta_2, \beta_3)^T, \quad \text{and} \quad \mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^T,$$

where  $n_i$  ( $\leq 6$ ) is the last measurement time for the  $i$ th woman.

Figure 2 reports the triangular kernel estimates of the marginal density of the response in the entire sample, in the pain medication, and in the placebo group. As optimal level of smoothing, the bandwidth of the kernel function was set equal to twice its standard deviation. Other kernel functions (e.g. Gaussian, rectangular, Epanechnikov, biweight, cosine) and different bandwidths were explored; they all showed similar pictures. The resulting shapes are clearly far from the symmetric, bell-shaped normal distribution. The estimated conditional densities for the pain medication and the placebo group appear to be very different from one another. The former presents a high density peak toward the minimum of the pain score scale, indicating that the treatment is effective. The estimated mean and median pain score are equal to 18.87 and 9.00, respectively. The latter is approximately uniform, with estimated mean and median pain score equal to 49.51 and 45.50, respectively.

Given the longitudinal nature of the observations, we should expect some degree of correlation between the measurements pertaining to the same subject. The application of random-effects regression models to such data has a very widespread success. For the case at hand, however, the estimation of a LMM would drive to misleading, inappropriate inferential results. In fact, the assumption of normality is clearly unreasonable. Further, and most importantly, the median of the pain score might be more informative than the mean for such a skewed distribution, for the mean is highly influenced by unusually large observations, which may occur with substantial probability in skewed populations.

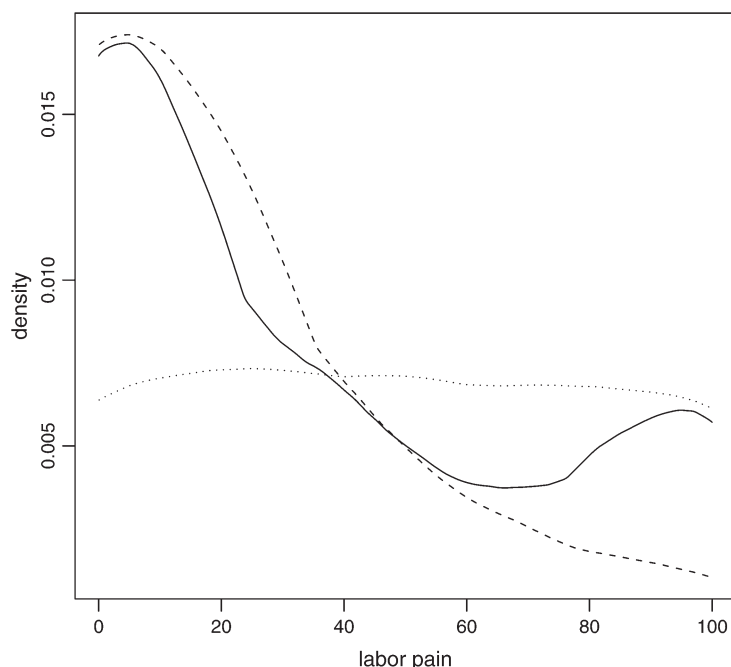


Fig. 2. The density of the labor pain score, estimated with a triangular kernel estimator, is plotted for the entire sample (solid line), for the placebo group only (dotted line), and for the pain medication group only (dashed line).

We then estimated the median regression with random effects

$$G_{y_i|u_i}(0.5|X_i, u_i) = X_i\beta + u_i, \quad i = 1, \dots, 83, \quad (4.2)$$

where  $u_i$  is a random intercept and, for  $(y_i, u_i)$ , we assumed the density given in (2.13). According to the results of the simulation in Section 3, the model in (4.2) and the use of the normal Laplace joint distribution should be superior to the LMM. The advantage is two-fold:

- the quantile model provides an estimate for the median, which is a more informative measure of central tendency than the mean for such a skewed distribution;
- the quantile model makes no assumptions about the distribution of the residual error which allows correct inference about other quantiles.

The Monte Carlo E-step was carried out via the Gibbs sampler with a constant sample size of 10 000, and a burn-in size of 5000.

We also estimated the corresponding penalized median regression with fixed  $u_i$  values. Lacking any practical guidelines as to how to choose the penalizing constant,  $\lambda$ , we set it at two arbitrarily chosen values, 0.01 and 5.

The estimates of the parameter  $\beta$  are reported in Table 2, along with the results of the quasi-likelihood approach showed by Jung (1996) and the least-squares estimates of the random intercept model. The significance of the parameter estimates was obtained by using a likelihood ratio statistic. According to our fitted model, the baseline pain does not differ significantly between the two groups, while time and its interaction with treatment affect the amount of pain. We also evaluated approximated 95% confidence intervals by using a bootstrap technique with a sample of size 499, and no differences were observed.

Table 2. *Parameter estimates for the labor pain data*

Method	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$
Median regression with random effects	14.66*	0.08	1.14*	10.15*
Quasi-likelihood	4.36	-9.85	1.37	15.48*
PQR ( $\lambda = 0.01$ )	6.50	2.17	0.50	10.17*
PQR ( $\lambda = 5$ )	6.75	-0.15	0.25	11.45*
Random-effects regression	12.98*	1.41	2.28*	9.74*

\*significant at 5% level.

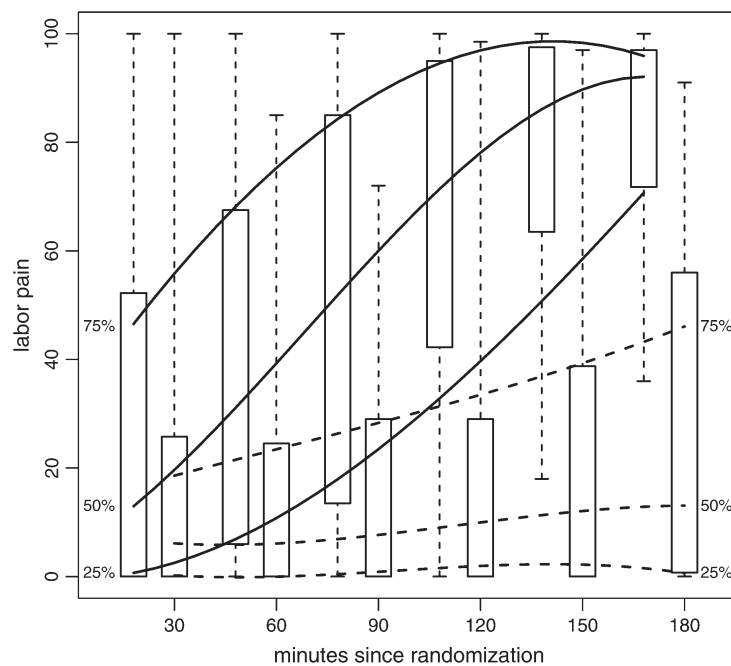


Fig. 3. Boxplot of labor pain score. The lines represent the estimate of the quantile functions for the placebo group (solid) and the pain medication group (dashed).

The median regression with random effects and the random-effects regression tally in terms of the magnitude and the significance of the estimate of  $\beta$ . To some extent, this result supports the use of random effects. Moreover, the rest of the QRs detected the significance of  $\beta_3$  only, which might be due to small power.

By examining the plot in Figure 1, it is evident that a first-degree polynomial model does not adequately describe the median pain score over time. We added square and cubic terms for both the time,  $T$ , and the interaction between time and treatment,  $RT$ , to model (4.1). We also estimated the first and the third quartiles and we used the Akaike information criterion (AIC) to evaluate the fit of each model. The possibility to describe more generally the conditional distribution of the response through the estimation of its quantiles, while accounting for the dependence among the observations, represents a great

advantage of our models with respect to normal mixed regressions. The latter, in fact, cannot realistically approximate the quantiles of a skewed or heavy-tailed distributions.

Figure 3 presents a summary of the QR results. The magnitude and the sign of the skewness of the pain score distribution for the placebo group change over time. In fact, the skewness remains positive in the interval between 30 and 90 min, circa, and it turns negative after 90 min. This result suggests that there is a modest fraction of women that is more sensitive to pain or that is not influenced by the placebo in the first halftime of the treatment, and a modest fraction of women that are less sensitive to pain or influenced by the placebo in the second half. The skewness for the pain medication group is always positive and its magnitude increases over time, suggesting that the medication is effective constantly over time for at least half of the group. A more detailed analysis of the distributional composition of the two groups could be extended to other quantiles. In general, however, the study of the tails of the distribution requires a suitable number of observations to ensure stability of the estimates.

For the labor pain data, the lowest AIC value was obtained when adding the terms  $T^2$ ,  $RT^2$ , and  $RT^3$  for both the first quartile and the median. For the third quartile, the AIC was minimized when adding the terms  $T^2$ ,  $RT^2$ , and  $T^3$ . On one hand, the level of pain for the most sensitive women of the placebo group seems to increase rapidly over time and then to stabilize at the median value, suggesting that they did not experience a placebo effect or they did it to a lesser extent. On the other hand, the third quartile of the medication group seems still on a rise at the end of the observational time. The estimate and the significance of the coefficients (not shown here) of a polynomial of degree higher than one must be interpreted with caution in this case. The value of the self-measured pain is, in fact, constrained to be less than or equal to 100 and the curvature of the upper quartile for the placebo group might be unnatural. These results, therefore, are purely an indication of the inadequacy of the model in (4.1).

We monitored convergence of the Gibbs sampler through the diagnostic techniques suggested by Cowles and Carlin (1996) and we did not observe any lack of convergence. Lag-1 autocorrelation was always within  $\pm 0.05$  for the first and the third quartiles and within  $\pm 0.07$  for the median model.

## 5. FINAL REMARKS

Extending the QR methods to include random effects is currently an active and promising area of research. We proposed a novel linear QR model with a subject-specific random intercept that accounts for within-group correlation. In order to estimate the model parameter, we applied an iterative procedure. At each step, a sample from the conditional density of the random effects was drawn via a Gibbs sampler with an adaptive rejection metropolis sampling algorithm. Then, the predicted random effects were subtracted from the response variable and a simple linear programming algorithm was applied to estimate the regression quantiles. Finally, we used the likelihood ratio statistic to test the null hypotheses that the fixed effects are equal to zero. Alternatively, a bootstrap technique can be used to estimate the variance of the  $\beta$ 's estimator.

In our experience, the ALD seems to be very promising within the regression quantile models, since it clearly represents a suitable error law for the least absolute estimator. Also, the regression median with random effects is seen to be more efficient than the least-squares estimator in the LMM for any distribution for which the median is more efficient than the mean in the location model.

Further studies should be aimed at alleviating the computational burden related to the presence of the multiple integral with respect to the random effects. The combined use of generalized directional derivatives and numerical integration techniques (e.g. the Clark derivative and the adaptive quadrature) represents a possible solution to the maximization of the marginal likelihood, which could be easily implemented with standard statistical languages such as R. The extensions to conditional quantile functions with more than a single random effect and to the use of different link functions also represent interesting areas of development.

## ACKNOWLEDGMENTS

*Conflict of Interest:* None declared.

## REFERENCES

- BOOTH, J. G. AND HOBERT, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society, Series B* **61**, 265–85.
- BUCHINSKY, M. (1995). Estimating the asymptotic covariance matrix for quantile regression models: a Monte Carlo study. *Journal of Econometrics* **68**, 303–38.
- COLE, T. J. AND GREEN, P. J. (1992). Smoothing reference centile curves: the LMS method and penalized likelihood. *Statistics in Medicine* **11**, 1305–19.
- COWLES, M. K. AND CARLIN, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association* **91**, 883–904.
- DAVIS, C. S. (1991). Semi-parametric and non-parametric methods for the analysis of repeated measurements with applications to clinical trials. *Statistics in medicine* **10**, 1959–80.
- GILKS, W. R. (1995). Derivative-free adaptive rejection sampling for Gibbs sampling. In: Bernardo, J. M., Berger, J. O., David, A. P. and Smith, A. F. M. (editors), *Bayesian Statistics 4*. Oxford: Clarendon. pp. 641–9.
- GILKS, W. R., BEST, N. G. AND TAN, K. K. C. (1995). Adaptive rejection metropolis sampling within Gibbs sampling. *Applied Statistics* **44**, 455–72.
- HEAGERTY, P. J. AND PEPE, M. S. (1999). Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in US children. *Journal of the Royal Statistical Society, Series C* **48**, 533–51.
- IBRAHIM, J. G., CHEN, M. H. AND LIPSITZ, S. R. (1999). Monte Carlo EM for missing covariates in parametric regression models. *Biometrics* **99**, 104–11.
- IBRAHIM, J. G., CHEN, M. H. AND LIPSITZ, S. R. (2001). Missing responses in generalized linear mixed models when the missing data mechanism is nonignorable. *Biometrika* **88**, 551–64.
- JUNG, S. (1996). Quasi-likelihood for median regression models. *Journal of the American Statistical Association* **91**, 251–7.
- KOENKER, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis* **91**, 74–89.
- KOENKER, R. (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. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association* **94**, 1296–309.
- LANGE, K. L., LITTLE, R. J. A. AND TAYLOR, J. M. G. (1989). Robust statistical modeling using the *t*-distribution. *Journal of the American Statistical Association* **84**, 881–96.
- 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. *Journal of the Royal Statistical Society, Series C* **46**, 463–76.
- MENG, X. L. AND VAN DYK, D. (1998). Fast EM-type implementations for mixed effects models. *Journal of the Royal Statistical Society, Series B* **60**, 559–68.
- RICHARDSON, A. M. (1997). Bounded influence estimation in the mixed linear model. *Journal of the American Statistical Association* **92**, 154–61.
- RICHARDSON, A. M. AND WELSH, A. H. (1995). Robust restricted maximum likelihood in mixed linear models. *Biometrics* **51**, 1429–39.

- ROBINS, J. M., ROTNITZKY, A. AND ZHAO, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **90**, 106–21.
- YU, K. AND MOYEED, R. A. (2001). Bayesian quantile regression. *Statistics and Probability Letters* **54**, 437–47.
- YU, K. AND ZHANG, J. (2005). A three-parameter asymmetric Laplace distribution and its extension. *Communications in Statistics—Theory and Methods* **34**, 1867–79.

[Received November 11, 2005; revised April 18, 2006; accepted for publication April 21, 2006]