Journal of Statistical Computation and Simulation
Vol. 82, No. 11, November 2012, 1635–1649

Bayesian quantile regression for longitudinal data models

Youxi Luo[a,b]*, Heng Lian[c] and Maozai Tian[a]

[a]*School of Statistics, Renmin University of China, Beijing 100872, People's Republic of China;* [b]*School of Science, Hubei University of Technology, Wuhan 430068, People's Republic of China;* [c]*Division of Mathematical Sciences, SPMS, Nanyang Technological University, Singapore 637371, Singapore*

(*Received 3 September 2010; final version received 19 May 2011*)

In this paper, we discuss a fully Bayesian quantile inference using Markov Chain Monte Carlo (MCMC) method for longitudinal data models with random effects. Under the assumption of error term subject to asymmetric Laplace distribution, we establish a hierarchical Bayesian model and obtain the posterior distribution of unknown parameters at τ-th level. We overcome the current computational limitations using two approaches. One is the general MCMC technique with Metropolis–Hastings algorithm and another is the Gibbs sampling from the full conditional distribution. These two methods outperform the traditional frequentist methods under a wide array of simulated data models and are flexible enough to easily accommodate changes in the number of random effects and in their assumed distribution. We apply the Gibbs sampling method to analyse a mouse growth data and some different conclusions from those in the literatures are obtained.

Keywords: asymmetric Laplace distribution; quantile regression; Bayesian inference; longitudinal data; Markov Chain Monte Carlo; Metropolis-Hastings algorithm; Gibbs sampler

1. Introduction

Quantile regression is a method that provides a more complete inferential picture than ordinary least-squares regression. It allows the full range of the data to be modelled and so can be beneficial when large or small response values are of particular interest. It also means that quantile regression can be viewed as a data-exploration technique. Koenker [1] presented a comprehensive discussion about the use of quantile regression. Quantile regression can be implemented in a range of different ways, and Yu *et al.* [2] provided an overview of some commonly used quantile regression techniques.

Koenker and Machado [3] introduced a goodness-of-fit process for quantile regression and related inference processes, and they considered the likelihood-ratio statistic under the parametric assumption of an asymmetric Laplace distribution (ALD) for the error term. Thus, it provides a possible approach for Bayesian inference on quantile regression, which has been considered by several authors, including Yu and Moyeed [4], who adopted a parametric approach based on a polynomial quantile regression function. Inference about the posterior distribution of the

*Corresponding author. Email: youxiluo@163.com

ISSN 0094-9655 print/ISSN 1563-5163 online

http://dx.doi.org/10.1080/00949655.2011.590488
http://www.tandfonline.com

parameters of this regression function is made by means of a Markov Chain Monte Carlo (MCMC) algorithm. Dunson and Taylor [5] discussed Bayesian inference for quantiles when the likelihood function is not fully specified. Kottas and Gelfand [6] considered Bayesian semiparametric median regression modelling approach under a Dirichlet process mixture framework. Kottas and Krnjajić [7] extended this approach to general quantiles. Lancaster and Jun [8] took an empirical likelihood approach to conduct a Bayesian inference in quantile regression. They showed that their method works well even in models with an endogenous regressor when the instruments are not too weak. Note also that it is very straightforward to incorporate multiple values of quantiles in Lancaster and Jun's framework.

How to extend the quantile regression methods to longitudinal data models is currently an active and promising area of research. In this paper, we focus on exploring the use of Bayesian quantile regression for the analysis of longitudinal data which include random effects. These data are characterized by repeated measurements on the same subject over time, as may be collected in clinical trials, epidemiological studies, etc. By using the connection between ALD and quantile regression, we develop a fully Bayesian hierarchical model to estimate the parameters of conditional quantile functions with random effects by adopting an ALD for the residual errors and prior distributions for other parameters. This Bayesian quantile mixed-effects model is analogous to the linear mixed model [9]. The within-subject dependence among data is taken into account through incorporation of random effects to avoid bias in the parameter estimate. The proposed approach allows the estimation at different quantiles of conditional distribution, and hence leads to a robust estimation of parameters and conveys a more comprehensive understanding of data. There is a relatively small amount of literature about the extension of quantile regression to longitudinal data or dependent data. Three types of models can be identified among them: the marginal model, the penalized model and the conditional model. Jung [10] proposed a quasi-likelihood approach for median regression estimation, where the dependency structure is modelled by the covariance matrix of the indicator functions. Lipsitz *et al.* [11] made use of Jung's work and described the use of quantile regression methods to analyse longitudinal data on CD4 cell counts, and they followed the work of Robins *et al.* [12] to account for missing at random dropouts by weighting the generalized estimating equations. Karlsson [13] examined a weighted version of quantile regression estimator adjusted to the case of nonlinear longitudinal data. These methods are basically marginal models that capture the overall trend among all subjects for a given quantile. Koenker [14] 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. Geraci and Bottai [15] proposed a conditional model for quantile regression model with random intercept. Liu and Bottai [16] generalized the conditional model proposed by Geraci and Bottai [15] and a quantile regression with multiple random effects is developed. Their approach is to use Monte Carlo EM algorithm. Instead of only drawing samplers from the conditional density of the random effects, we proposed a random walk Metropolis–Hastings sampling algorithm and a full Gibbs sampling algorithm to estimate all of the model unknowns. These approaches provide a more flexible way to analyse longitudinal data models with any random effects distributions, such as Cauchy and other heavy-tailed distributions that are commonly used in speech and image data [17–19].

The remainder of the paper proceeds as follows. Section 2 introduces the connection between quantile regression and ALD. We propose the hierarchical Bayesian quantile regression model for longitudinal data in Section 3 and two MCMC algorithms are also given. Simulation studies are presented in Section 4. A real data example is illustrated in Section 5. Discussions and conclusions are presented in Section 6. Some technical derivations are relegated to the Appendix.

2. Relationship between quantile regression and ALD

The linear conditional quantile function is defined as

$$Q_{y_i}(\tau|\boldsymbol{x}_i) = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\tau, \tag{1}$$

where $Q_{y_i}(\cdot) \equiv F_{y_i}^{-1}(\cdot)$ is the inverse of the cumulative distribution function of the response conditional on $\boldsymbol{x}_i$. In τ-th quantile regression, we have the following minimization problem

$$\min_{\boldsymbol{\beta}\in\mathbf{R}^p} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^{\mathrm{T}}\boldsymbol{\beta}_\tau), \tag{2}$$

where $\rho_\tau(u) = u(\tau - I(u \leq 0))$ is the so-called check function. A connection between the minimization of the sum in Equation (2) and the maximum-likelihood theory is provided by the ALD. This skewed distribution appeared in the paper by Koenker and Machado [3] and Yu and Moyeed [4], among others. We say that a random variable Y is distributed as an ALD with parameters μ, σ, τ, and we denote $Y \sim \mathrm{ALD}(\mu, \sigma, \tau)$, if the corresponding probability density is given by

$$f(y|\mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\rho_\tau\left(\frac{y-\mu}{\sigma}\right)\right\}, \tag{3}$$

where $0 < \tau < 1$ is the skewness parameter, σ is the scale parameter, and $-\infty < \mu < +\infty$ is the location parameter. Set $\mu_i = x_i^{\mathrm{T}}\boldsymbol{\beta}$ and $\boldsymbol{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$. Assuming that $y_i \sim \mathrm{ALD}(\mu_i, \sigma, \tau)$, then the likelihood for n independent observations is, bar a proportionality constant,

$$L(\boldsymbol{\beta}, \sigma; \boldsymbol{y}, \tau) \propto \frac{1}{\sigma^n} \exp\left\{-\sum_{i=1}^{n} \rho_\tau\left(\frac{y_i - x_i^{\mathrm{T}}\boldsymbol{\beta}}{\sigma}\right)\right\}. \tag{4}$$

If we consider σ a nuisance parameter, then the maximization of the likelihood in Equation (4) with respect to the parameter $\boldsymbol{\beta}$ is equivalent to Equation (2).

In quantile regression, it is often of interest to compare slope coefficients for different quantiles. Then how ALD can deal with the case when slope coefficients might be different for different quantile level. In the Bayesian model using ALD, we impose the assumption $y \sim \mathrm{ALD}(\mu, \sigma, \tau)$, which implies that the different quantiles of y conditional on x has the same slope. However, we only compute the τ-quantile of y if $y \sim \mathrm{ALD}(\mu, \sigma, \tau)$ and for different τ, we actually use a different model. Thus as long as Equation (1) is true, the likelihood is consistent in the sense that the maximum likelihood estimator (MLE) will converge to the true β_τ in Equation (1). Thus, when using ALD in Bayesian analysis, we still can get consistent estimation of the quantile function and the slope coefficients might be different for different τ.

3. Longitudinal data models and Bayesian quantile regression methods

3.1. *Hierarchical Bayesian quantile regression model*

Consider the classical linear random effects model,

$$y_{ij} = \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_i + \varepsilon_{ij}, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, n_i, \quad \sum_i n_i = N, \tag{5}$$

where y_{ij} is the j-th scalar measurement of a continuous random variable on the i-th subject, $\boldsymbol{x}_{ij}^{\mathrm{T}}$ are row vectors of a known design matrix of dimension $N \times k$ and $\boldsymbol{\beta}$ is a $k \times 1$ vector of fixed regression coefficients, z_{ij} is a $p \times 1$ covariate associated with random effects, and $\boldsymbol{\alpha}_i$ is a $p \times 1$ vector of random effects.

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

$$Q_{y_{ij}}(\tau|\boldsymbol{x}_{ij}, \boldsymbol{\alpha}_i) = \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + z_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_i. \tag{6}$$

We assume that y_{ij}, conditionally on $\boldsymbol{\alpha}_i$, for $j = 1, \ldots, n_i,\ i = 1, \ldots, n$ are independently distributed according to the ALD

$$f(y_{ij}|\boldsymbol{\beta}, \boldsymbol{\alpha}_i, \sigma) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\rho_\tau\left(\frac{y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - z_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_i}{\sigma}\right)\right\}. \tag{7}$$

Also, we assume that $\boldsymbol{\alpha}_i$ are identically distributed as $f(\boldsymbol{\alpha}_i|\Sigma)$. Assume that we have the priors $\boldsymbol{\beta} \sim \pi(\boldsymbol{\beta}),\ \sigma \sim \pi(\sigma),\ \Sigma \sim \pi(\Sigma)$. Then we have the following hierarchical Bayesian quantile regression model:

$$\begin{aligned}
y_{ij} &\sim \mathrm{ALD}(\mu_{ij}, \sigma, \tau),\\
\mu_{ij} &= \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + z_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_i,\\
\boldsymbol{\beta} &\sim \pi(\boldsymbol{\beta}),\\
\boldsymbol{\alpha}_i|\Sigma &\sim f(\boldsymbol{\alpha}_i|\Sigma),\\
\Sigma &\sim \pi(\Sigma),\\
\sigma &\sim \pi(\sigma).
\end{aligned}$$

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

$$f(\boldsymbol{y}_i, \boldsymbol{\alpha}_i|\boldsymbol{\beta}, \sigma, \Sigma) = f(\boldsymbol{y}_i|\boldsymbol{\beta}, \boldsymbol{\alpha}_i, \sigma) f(\boldsymbol{\alpha}_i|\Sigma). \tag{8}$$

If we let $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n)$, the joint density of $(\boldsymbol{y}, \boldsymbol{\alpha})$ based on n subjects is given by

$$f(\boldsymbol{y}, \boldsymbol{\alpha}|\boldsymbol{\beta}, \sigma, \Sigma) = \prod_{i=1}^{n} f(\boldsymbol{y}_i|\boldsymbol{\beta}, \boldsymbol{\alpha}_i, \sigma) f(\boldsymbol{\alpha}_i|\Sigma). \tag{9}$$

The posterior distribution can then be calculated using Bayes theorem

$$\pi(\boldsymbol{\beta}, \sigma, \Sigma, \boldsymbol{\alpha}|\boldsymbol{y}) \propto f(\boldsymbol{y}, \boldsymbol{\alpha}|\boldsymbol{\beta}, \sigma, \Sigma)\pi(\boldsymbol{\beta})\pi(\sigma)\pi(\Sigma), \tag{10}$$

then all inferences about unknown parameters $\vartheta = (\boldsymbol{\beta}, \sigma, \Sigma, \boldsymbol{\alpha})$ should be based on this posterior distribution. For example, the marginal distribution of fixed parameters $\boldsymbol{\beta}$ can be obtained by integrating out σ, Σ, and $\boldsymbol{\alpha}$ from $\pi(\boldsymbol{\beta}, \sigma, \Sigma, \boldsymbol{\alpha}|\boldsymbol{y})$, using

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}) = \int \cdots \int \pi(\boldsymbol{\beta}, \sigma, \Sigma, \boldsymbol{\alpha}|\boldsymbol{y})\, \mathrm{d}\sigma\, \mathrm{d}\Sigma\, \mathrm{d}\boldsymbol{\alpha}. \tag{11}$$

Next, we need to choose appropriate distribution for the random effects $\boldsymbol{\alpha}$ and priors for $\boldsymbol{\beta}$, σ, and Σ. Specifically, we assume the priors $\boldsymbol{\beta} \sim N_k(\boldsymbol{b}_0, \boldsymbol{B}_0)$, $\sigma \sim \mathrm{IG}(c_0, d_0)$, where $\mathrm{IG}(a, b)$ denotes an inverse Gamma distribution with parameters a and b. The random effects $\boldsymbol{\alpha}_i$ are identically distributed as $N_p(\mathbf{0}, \phi^2\boldsymbol{I})$, and they are mutually independent. More generally, we could set $\boldsymbol{\alpha}_i \sim N_p(\mathbf{0}, \phi^2\boldsymbol{\Sigma})$, however, inferences become more difficult with more parameters to estimate. Finally, we assign the hyper prior $\phi^2 \sim \mathrm{IG}(k_0, w_0)$. Because the ALD has no conjugate prior, the posterior density of both fixed and random effects parameters are very complicated. Besides, apparently, the above integral is too complex to be evaluated analytically. However, Equation (10) lends itself to a full Bayesian analysis by MCMC methods. We deal with this problem using the next two approaches.

3.2. *Metropolis–Hastings algorithm*

According to Equation (10), the posterior joint distribution is

$$\begin{aligned}
\pi(\boldsymbol{\beta}, \sigma, \phi^2, \boldsymbol{\alpha}|\boldsymbol{y}) &\propto f(\boldsymbol{y}, \boldsymbol{\alpha}|\boldsymbol{\beta}, \sigma, \phi^2)\pi(\boldsymbol{\beta})\pi(\sigma)\pi(\phi) \\
&\propto \left(\frac{1}{\sigma}\right)^N \exp\left\{-\frac{1}{\sigma}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\rho_\tau(y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_i)\right\} \\
&\quad \times \left(\frac{1}{\phi^2}\right)^{np/2} \exp\left\{-\frac{1}{2\phi^2}\sum_{i=1}^{n}\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{\alpha}_i\right\} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{b}_0)^{\mathrm{T}}\boldsymbol{B}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{b}_0)\right\} \\
&\quad \times \left(\frac{1}{\sigma}\right)^{c_0-1} \exp\left\{-d_0\frac{1}{\sigma}\right\} \left(\frac{1}{\phi^2}\right)^{k_0-1} \exp\left\{-w_0\frac{1}{\phi^2}\right\},
\end{aligned}$$

The full conditional distribution of all the parameters are given by

$$\begin{aligned}
\boldsymbol{\alpha}_i|\boldsymbol{y}, \boldsymbol{\beta}, \sigma, \phi^2 &\propto \exp\left\{-\frac{1}{2\phi^2}\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{\alpha}_i - \frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau(y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_i)\right\}, \\
\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{\alpha}, \sigma, \phi^2 &\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{b}_0)^{\mathrm{T}}\boldsymbol{B}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{b}_0) - \frac{1}{\sigma}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\rho_\tau(y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_i)\right\}, \\
\sigma|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \phi^2 &\sim \mathrm{IG}\left(N + c_0, d_0 + \sum_{i=1}^{n}\sum_{j=1}^{n_i}\rho_\tau(y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_i)\right), \\
\phi^2|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma &\sim \mathrm{IG}\left(\frac{np}{2} + k_0, w_0 + \frac{1}{2}\sum_{i=1}^{n}\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{\alpha}_i\right).
\end{aligned}$$

The posterior densities of $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}$ do not represent common and well-known distributions but can be updated by the Metropolis–Hastings algorithm (see, for example [20]). We summarize the Metropolis–Hastings algorithm as follows:

1. Sample $\boldsymbol{\beta}$ from $\pi(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{\alpha}, \sigma, \phi^2)$ using the random walk Metropolis–Hastings algorithm with multivariate normal distribution as the proposal density. Given the current value $\boldsymbol{\beta}^{\mathrm{c}}$, we first draw $\boldsymbol{\beta}^{\mathrm{t}}$ from $N_k(\boldsymbol{\beta}^{\mathrm{c}}, \sigma_\beta^2\Omega_\beta)$ and accept $\boldsymbol{\beta}^{\mathrm{t}}$ with probability

$$\boldsymbol{\alpha}(\boldsymbol{\beta}^{\mathrm{c}}, \boldsymbol{\beta}^{\mathrm{t}}) = \min\left\{1, \frac{\pi(\boldsymbol{\beta}^{\mathrm{t}}|\boldsymbol{y}, \boldsymbol{\alpha}, \sigma, \phi^2)}{\pi(\boldsymbol{\beta}^{\mathrm{c}}|\boldsymbol{y}, \boldsymbol{\alpha}, \sigma, \phi^2)}\right\},$$

   where we tune the parameter σ_β to make the acceptance probability fall between 0.15 and 0.5 [21].

2. Sample $\boldsymbol{\alpha}$ from $\pi(\boldsymbol{\alpha}_i|\boldsymbol{y}, \boldsymbol{\alpha}, \sigma, \phi^2)$ using the random walk Metropolis–Hastings algorithm with multivariate normal distribution as the proposal density. Given the current value $\boldsymbol{\alpha}_i^{\mathrm{c}}$, we first draw $\boldsymbol{\alpha}_i^{\mathrm{t}}$ from $N_p(\boldsymbol{\alpha}_i^{\mathrm{c}}, \sigma_\alpha^2\Omega_\alpha)$ and accept $\boldsymbol{\alpha}_i^{\mathrm{t}}$ with probability

$$\boldsymbol{\alpha}(\boldsymbol{\alpha}_i^{\mathrm{c}}, \boldsymbol{\alpha}_i^{\mathrm{t}}) = \min\left\{1, \frac{\pi(\boldsymbol{\alpha}_i^{\mathrm{t}}|\boldsymbol{y}, \boldsymbol{\beta}, \sigma, \phi^2)}{\pi(\boldsymbol{\alpha}_i^{\mathrm{c}}|\boldsymbol{y}, \boldsymbol{\beta}, \sigma, \phi^2)}\right\},$$

   where we tune the parameter σ_α to make the acceptance probability falling between 0.15 and 0.5.
3. Sample σ from $\sigma|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \phi^2$
4. Sample ϕ^2 from $\phi^2|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma$
5. Repeat Steps 1–4 using the most recent values of the conditioning variables.

In Steps 1 and 2, the correlations involved in the proposal covariance matrix Ω_β and Ω_α influence the direction of the proposed movement. We use identity matrix in the following simulation studies and real data example. For a Metropolis algorithm with general proposal variance–covariance matrix, one could refer to Gelman *et al.* [22].

3.3. *Gibbs sampling algorithm*

In Section 3.2, we use a random walk Metropolis–Hastings algorithm with a Gaussian proposal density centered at the current parameter value. Although the random walk sampler is a convenient choice to generate candidate values, the corresponding acceptance probability depends on the value of τ through the likelihood function. As a result, parameters in the proposal distributions need to be carefully adjusted so as to attain appropriate acceptance rates for each value of τ, and this limits the applicability of the random walk sampler in practice. In this section, we develop a Gibbs sampling algorithm based on a location-scale mixture representation of the ALD.

Let $y \sim \mathrm{ALD}(\mu, \sigma, \tau)$, z be a standard normal random variable and e be an exponential random variable with parameter σ. Then we have the following representation [23]:

$$y = \sqrt{\frac{2\sigma e}{\tau(1-\tau)}}z + \frac{1-2\tau}{\tau(1-\tau)}e + \mu. \tag{12}$$

Denoting $k_1 = \frac{1-2\tau}{\tau(1-\tau)}$ and $k_2 = \frac{2}{\tau(1-\tau)}$, the above is rewritten as

$$y = k_1 e + \sqrt{k_2\sigma e}z + \mu. \tag{13}$$

From this result, the response y_{ij} can be equivalently rewritten as

$$y_{ij} = \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_i + k_1 e_{ij} + \sqrt{k_2\sigma e_{ij}}z_{ij}, \tag{14}$$

where $\boldsymbol{\alpha}_i \sim N_p(\boldsymbol{0}, \phi^2\boldsymbol{I})$, $e_{ij} \sim \exp(1/\sigma)$ and $z_{ij} \sim N(0, 1)$ are mutually independent, and thus

$$f(y_{ij}|\boldsymbol{\beta}, \boldsymbol{\alpha}_i, e_{ij}, \sigma) = (2\pi k_2\sigma e_{ij})^{-1/2}\exp\left\{-\frac{1}{2k_2\sigma e_{ij}}(y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_i - k_1 e_{ij})^2\right\}. \tag{15}$$

Denoting $\boldsymbol{e}_i = (e_{i1}, \ldots, e_{in_i})^{\mathrm{T}},\ i = 1, \ldots, n$, $\boldsymbol{e} = (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$. The full hierarchical Bayesian quantile regression model can be written as

$$y_{ij} = \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_i + k_1 e_{ij} + \sqrt{k_2\sigma e_{ij}}z_{ij},$$

$$z_{ij} \sim \prod_{i=1}^{n}\prod_{j=1}^{n_i}\exp\left(-\frac{1}{2}z_{ij}^2\right),$$

$$
\begin{aligned}
\boldsymbol{e}|\sigma &\sim \prod_{i=1}^{n}\prod_{j=1}^{n_i}\frac{1}{\sigma}\exp\left(-\frac{e_{ij}}{\sigma}\right),\\
\boldsymbol{\alpha}_i|\phi &\sim N_p(0,\phi^2\boldsymbol{I}),\\
\boldsymbol{\beta} &\sim N_k(\boldsymbol{b}_0,\boldsymbol{B}_0),\\
\sigma &\sim \mathrm{IG}(c_0,d_0),\\
\phi^2 &\sim \mathrm{IG}(k_0,w_0).
\end{aligned}
$$

A Gibbs sampling algorithm then can be carried out as follows:

1. Generate initial values $\sigma^{(0)}$, $\boldsymbol{\alpha}_i^{(0)}$, $\boldsymbol{\beta}^{(0)}$.
2. Generate e_{ij} from $\pi(e_{ij}|\boldsymbol{y},\boldsymbol{\alpha}_i,\boldsymbol{\beta},\sigma)$.
3. Generate σ from $\pi(\sigma|\boldsymbol{y},\boldsymbol{\alpha},\boldsymbol{e},\boldsymbol{\beta})$.
4. Generate ϕ^2 from $\pi(\phi^2|\boldsymbol{y},\boldsymbol{\alpha},\boldsymbol{\beta},\sigma)$.
5. Generate $\boldsymbol{\alpha}_i$ from $\pi(\boldsymbol{\alpha}_i|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{e}_i,\sigma,\phi^2)$.
6. Generate $\boldsymbol{\beta}$ from $\pi(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{\alpha},\boldsymbol{e},\sigma,\phi^2)$.
7. Repeat Steps 2–6 using the most recent values of the conditioning variables.

The full conditional distribution of $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}$ is normal, those of σ and ϕ^2 are inverse Gamma distributions. The full conditional distribution of e_{ij} is the generalized inverse Gaussian distribution, and there exists efficient algorithms to be simulated from it [24]. All calculation details are included in the Appendix.

4. Simulation study

4.1. *Simple regression model*

We use the following simple linear mixed-effects model to generate the data

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \alpha_{i0} + \alpha_{i1}x_{ij} + \varepsilon_{ij}, \quad i = 1,\dots,20, \;\; j = 1,\dots,5, \tag{16}$$

where $x_{ij} \sim U(0,1)$ and $\boldsymbol{\beta} = (\beta_0,\beta_1)^{\mathrm{T}} = (1,5)^{\mathrm{T}}$. We consider the model with random effects $\boldsymbol{\alpha}_i = (\alpha_{i0},\alpha_{i1})^{\mathrm{T}} \sim N_2(\boldsymbol{0},\boldsymbol{I})$, and error term $\varepsilon_{ij} \sim N(0,1)$. We then carry out quantile regression at three different quantiles, namely $\tau = (0.25, 0.50, 0.75)$. We assume weak prior information and use an $N_2(\boldsymbol{0}, 100\boldsymbol{I})$ prior on $\boldsymbol{\beta}$ and $\mathrm{IG}(0.01, 0.01)$ priors on σ and ϕ^2. The Bayesian quantile regression using Metropolis–Hastings algorithm (BQRMH) or Gibbs sampling algorithm (BQRGS) is run with 20,000 iterations and a burn-in of 10,000. Table 1 compares mean, median, and 95% intervals for the three quantile regression parameters with two Bayesian quantile regression methods. From Table 1, we can see that both methods behave well in the simple linear mixed-effects model. In the simulation processes, we find that the posterior inference based on the two methods is not sensitive to the prior choice. For convergence diagnostics, we present some trace plots and sample autocorrelation functions (Figures 1–4). From the autocorrelation functions plotted, we see that the Gibbs sampling algorithm is more efficient.

4.2. *Multiple regression model*

We use the following model to generate the data

$$y_{ij} = x_{ij}^{(1)}\beta_1 + x_{ij}^{(2)}\beta_2 + x_{ij}^{(3)}\beta_3 + x_{ij}^{(4)}\beta_4 + \alpha_i + \varepsilon_{ij}, \quad i = 1,\dots,5, \;\; j = 1,\dots,30. \tag{17}$$

Table 1. Mean, median,and 95% intervals for the three quantile regression parameters with two Bayesian quantile regression methods.

		BQRMH				BQRGS			
Par	True	Mean	Median	2.5%	97.5%	Mean	Median	2.5%	97.5%
$\beta_0(0.25)$	0.33	0.350	0.351	0.262	1.324	0.355	0.357	0.219	1.455
$\beta_1(0.25)$	5.00	5.290	5.291	4.487	6.109	5.297	5.298	4.308	6.266
$\beta_0(0.50)$	1.00	1.058	1.040	0.481	1.705	1.042	1.046	0.425	1.657
$\beta_1(0.50)$	5.00	5.328	5.330	4.494	6.116	5.288	5.288	4.466	6.093
$\beta_0(0.75)$	1.67	1.628	1.623	1.078	2.402	1.600	1.606	0.879	2.284
$\beta_1(0.75)$	5.00	5.262	5.260	4.327	6.176	5.265	5.269	4.438	6.077

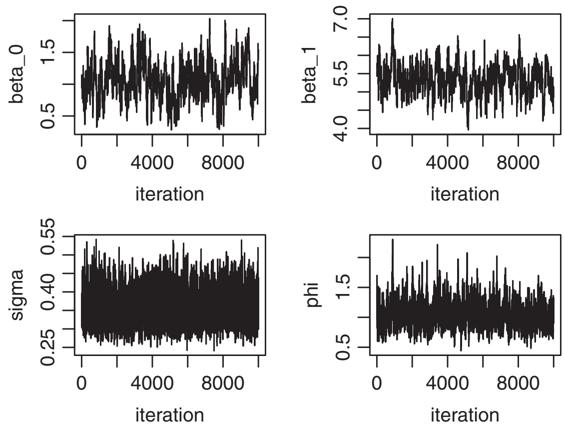

Figure 1. Trace plots of β_0, β_1, σ, and ϕ at 0.5 quantile based on 10,000 iterations from Metropolis–Hastings algorithm in simple regression model simulation study.

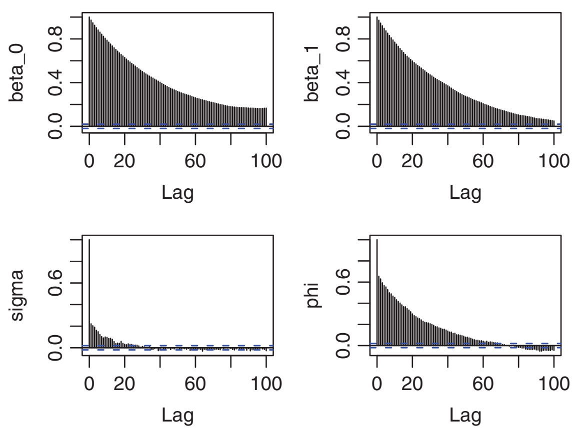

Figure 2. Sample autocorrelation functions of β_0, β_1, σ, and ϕ at 0.5 quantile based on 10,000 iterations from Metropolis–Hastings algorithm in simple regression model simulation study.

We set all covariates $x_{ij}^{(k)} \sim N(0, 1)$, $k = 1, 2, 3, 4$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)^{\mathrm{T}} = (5, 6, 7, 8)^{\mathrm{T}}$. A random intercept effect $\alpha_i \sim N(0, 4)$ is also considered. For the error term ε_{ij}, we consider the following four types of distribution: the standard normal distribution $N(0, 1)$, the Student's t distribution $t(3)$, the Cauchy$(0, 1)$ distribution, and the Laplace$(0, 1)$ distribution. Seven estimators are considered. The ordinary least-squares (LS) estimator simply ignores the random effects

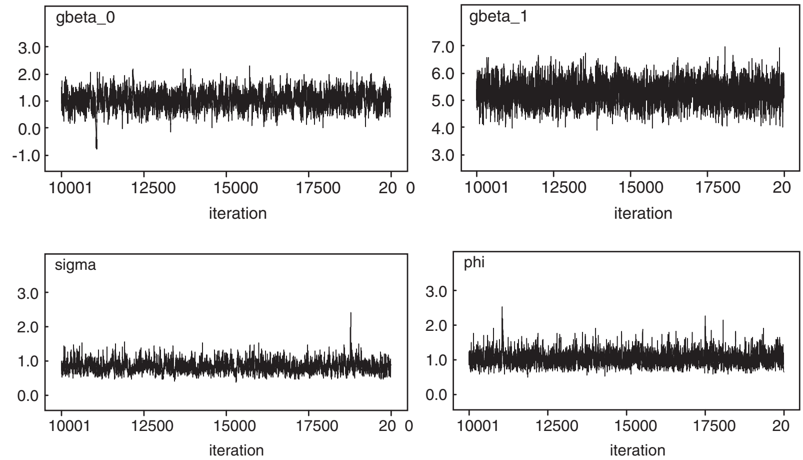

Figure 3. Trace plots of β_0, β_1, σ, and ϕ at 0.5 quantile based on 10,000 iterations from Gibbs sampling algorithm in simple regression model simulation study.

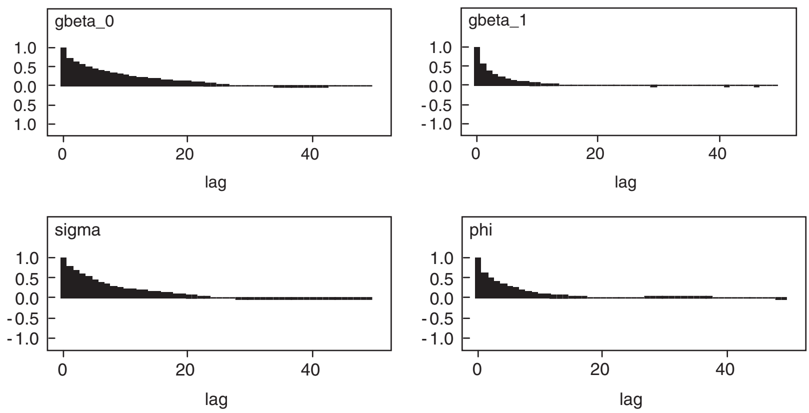

Figure 4. Sample autocorrelation functions of β_0, β_1, σ, and ϕ at 0.5 quantile based on 10,000 iterations from Gibbs sampling algorithm in simple regression model simulation study.

entirely. The least-squares fixed-effects estimator (LSFE) implements LS estimator of the model Equation (17). We also include the classical restricted maximum-likelihood estimator (REML) for random effects model, and the penalized quantile regression estimator (PQR) [14] with the regularization parameter $\lambda = 0.5,\ \infty$ (i.e., no random effects). The latter is equivalent to the ordinary quantile regression estimator (QR), in which the random effects are shrunk to zero. The choice $\lambda = 0.5$ is based on that, owing to difficulty in choosing appropriate regularization parameter, the ratio of the standard deviations of ε_{ij} and α_i has been sometimes proposed as reference value for regularization parameter. Finally, we have the BQRMH and the BQRGS estimators (based on posterior mean from samples drawn). We simulate 100 datasets for each error distribution. When carrying out the Bayesian quantile regression, we assume weak prior knowledge and use a $N(0, 100)$ prior on β and IG(0.01, 0.01) prior on $\sigma,\ \phi^2$. For each replication, both BQRMH and BQRGS are run with 10,000 iterations and a burn-in of 5000.

Table 2 reports the bias and root mean-squared error (RMSE) in each case. It is easy to see that the LS and QR do not perform well in nearly all scenarios because they ignore the random effects entirely. In the Gaussian setting, we see the anticipated efficiency loss owing to estimating the

Table 2. Bias and RMSE of $\boldsymbol{\beta}$ for the seven estimators.

	Bias				RMSE			
Method	β_1	β_2	β_3	β_4	β_1	β_2	β_3	β_4
$N(0,1)$								
LS	−0.0142	−0.0014	−0.0026	0.0045	0.1852	0.1601	0.1572	0.1544
LSFE	0.0024	−0.0001	0.0046	−0.014	0.0768	0.0826	0.0776	0.0819
REML	0.0024	−0.0003	0.0046	−0.0140	0.0769	0.0824	0.0776	0.0819
PQR	0.0100	0.0056	−0.0012	−0.0160	0.0996	0.1132	0.1111	0.1002
QR	−0.0024	−0.0071	−0.0027	−0.0110	0.2413	0.2373	0.2289	0.2204
BQRMH	−0.0299	−0.0361	−0.0457	−0.0680	0.0832	0.0928	0.0990	0.1104
BQRGS	0.0082	0.0041	−0.0008	−0.0160	0.0884	0.1003	0.0982	0.0932
$t(3)$								
LS	0.0246	−0.0262	−0.0200	0.0229	0.3119	0.2870	0.2645	0.2803
LSFE	0.0395	−0.0195	−0.0210	0.0393	0.3152	0.2512	0.2614	0.2385
REML	0.0371	−0.0204	−0.0180	0.0412	0.3063	0.2544	0.2492	0.2453
PQR	0.0029	0.0041	−0.0130	−0.0060	0.1288	0.1297	0.1308	0.1185
QR	−0.0180	−0.0055	−0.0170	−0.0030	0.2526	0.2679	0.2543	0.2486
BQRMH	−0.0850	−0.0973	−0.1190	−0.1170	0.1480	0.1561	0.1870	0.1602
BQRGS	−0.0040	0.0011	−0.0100	0.0031	0.1162	0.1204	0.1261	0.1013
Cauchy(0,1)								
LS	0.2416	0.6105	−0.2136	0.7290	1.2006	2.0477	3.0146	1.6726
LSFE	0.2252	0.6537	−0.1803	0.7536	1.2316	1.9982	2.9003	1.6736
REML	0.2620	0.6334	−0.1987	0.6987	1.1897	2.0498	3.0158	1.6433
PQR	−0.0301	−0.0280	0.0061	0.0733	0.1054	0.1138	0.1538	0.1515
QR	−0.0770	0.0732	−0.0586	0.0334	0.2162	0.2874	0.2830	0.2603
BQRMH	−0.2422	−0.2730	−0.3096	−0.3030	0.2939	0.3237	0.3542	0.3619
BQRGS	−0.0372	−0.0010	−0.0021	0.0416	0.1254	0.1380	0.1543	0.1341
Laplace(0,1)								
LS	0.0051	−0.0310	0.0074	0.0032	0.1943	0.2024	0.1778	0.1977
LSFE	0.0075	−0.0090	−0.0117	0.0170	0.1060	0.1165	0.1059	0.1136
REML	0.0076	−0.0090	−0.0110	0.0166	0.1062	0.1166	0.1057	0.1131
PQR	0.0034	0.0038	−0.0098	0.0048	0.1003	0.1122	0.0951	0.0982
QR	0.0012	−0.011	0.0190	−0.0190	0.2401	0.2538	0.2315	0.2422
BQRMH	−0.0481	−0.062	−0.0864	−0.0740	0.1083	0.1229	0.1237	0.1282
BQRGS	0.0070	0.0043	−0.0080	0.0100	0.0908	0.1037	0.0829	0.0972

median rather than the mean, but this loss in efficiency is generally small. For the $t(3)$, Cauchy(0,1), and Laplace(0,1) errors, the two Bayesian quantile regression methods perform better than LSFE and REML. Especially in the $t(3)$ and Cauchy(0,1) scenarios, the RMSE of BQRMH and BQRGS are far smaller than that of LSFE and REML. That is to say, the Bayesian quantile regression methods are robust to the noise term though we assume an ALD in the longitudinal data models. The biggest competitor of BQRMH and BQRGS is the PQR, who does as well as BQRGS nearly in all scenarios and does slightly better than BQRMH. For our two Bayesian quantile regression methods, BQRGS performs better than BQRMH nearly in all cases. Furthermore, the distributions in the full hierarchical model are all common distributions, which can be implemented easily via WinBUGS software.

5. Analysis of the mouse data

This dataset is reported by Williams and Izenman [25] and analysed by Rao [26] and later by Lee [27]. It consists of weights of 13 male mice measured at intervals of three days over the 21 days from birth to weaning. The purpose of the analysis of this dataset is to find the growth pattern of the mouse. Following Rao [26] and Lee [27], the growth curve for the mouse data is assumed to be a polynomial of second degree and the seven time points set to be $1, \ldots, 7$. Let y_{ij} be the weight of the i-th mouse at time point t_j, $t_j = j, i = 1, \ldots, 13$, $j = 1, \ldots, 7$. We consider the following three models.

Model 1: Quantile growth curve model with no random effects

$$y_{ij} = \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon_{ij}, \quad \text{with} \quad Q_{\varepsilon_{ij}}(\tau|\boldsymbol{x}_{ij}) = 0,$$

where $\boldsymbol{x}_{ij}^{\mathrm{T}} = (1, t_j, t_j^2)$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^{\mathrm{T}}$.

Model 2: Quantile growth curve model with random intercept

$$y_{ij} = \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + z_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_i + \varepsilon_{ij}, \quad \text{with} \quad Q_{\varepsilon_{ij}}(\tau|\boldsymbol{x}_{ij}, \boldsymbol{\alpha}_i) = 0,$$

where $\boldsymbol{x}_{ij}^{\mathrm{T}} = (1, t_j, t_j^2)$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^{\mathrm{T}}$, $z_{ij}^{\mathrm{T}} = 1$, $\boldsymbol{\alpha}_i = \alpha_{i0}$.

Model 3: Quantile growth curve model with random intercept and random slope

$$y_{ij} = \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + z_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_i + \varepsilon_{ij}, \quad \text{with} \quad Q_{\varepsilon_{ij}}(\tau|\boldsymbol{x}_{ij}, \boldsymbol{\alpha}_i) = 0,$$

where $\boldsymbol{x}_{ij}^{\mathrm{T}} = (1, t_j, t_j^2)$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^{\mathrm{T}}$, $z_{ij}^{\mathrm{T}} = (1, t_j)$, $\boldsymbol{\alpha}_i = (\alpha_{i0}, \alpha_{i1})^{\mathrm{T}}$.

In the three models, we assume $\varepsilon_{ij} \sim \mathrm{ALD}(0, \sigma, \tau)$ as defined in Section 2.1 and the random effects follow the multivariate normal distribution $N_p(\boldsymbol{0}, \phi^2\boldsymbol{I})$. We consider the priors $\boldsymbol{\beta} \sim N_3(\boldsymbol{0}, 100\boldsymbol{I})$, $\sigma \sim \mathrm{IG}(0.01, 0.01)$, $\phi^2 \sim \mathrm{IG}(0.01, 0.01)$. All results are based on 20,000 draws obtained after a burn-in of 10,000 iterations. Our main interest is the growth pattern of the mouse.

Table 3 summarizes the estimated fixed regression coefficients at five quantiles, $\tau = (0.15, 0.25, 0.50, 0.75, 0.95)$, for all the three models. Pan and Fang [28] obtained the least squares estimator (LSE) and MLE of $\boldsymbol{\beta}$ for Model 1, where $\hat{\boldsymbol{\beta}}_{\mathrm{LSE}} = (0.0036, 0.2177, -0.0119)^{\mathrm{T}}$, $\hat{\boldsymbol{\beta}}_{\mathrm{MLE}} = (0.0222, 0.2084, -0.0108)^{\mathrm{T}}$. Firstly, we compare our three median ($\tau = 0.5$) BQRGS estimates with LSE and MLE. Figure 5 depicts the five average growth curves, and we see that the curves attain their maximums at time point 10 and decrease after that. We can also see that the five curves are very close to each other before 10 and exhibit slight differences after that. However, in this example, Bayesian quantile regression gives us some other important information. If we compare the growth curves at different quantiles, it is not difficult to find that the growth curves are quite different between high quantiles ($\tau \geq 0.75$) and low quantiles ($\tau \leq 0.25$). Figure 6 presents graphically the growth curves at every quantile for Model 2. Plots for the other two models are very similar. Unlike the high quantile growth curves that first increase and then decrease, the low quantile growth curves show a strong upward trend all along. Finally, we compare the estimates

Table 3. BQRGS of $\boldsymbol{\beta}$ at five quantiles for the three models.

τ	0.15	0.25	0.50	0.75	0.95
Model 1					
β_0	−0.8079	−0.6957	0.0031	0.6247	0.5727
β_1	0.0173	0.0966	0.2189	0.2343	0.2273
β_2	0.0100	0.0009	−0.0122	−0.0138	−0.0131
Model 2					
β_0	−0.75	−0.671	0.0023	0.6276	0.555
β_1	0.0107	0.0845	0.2127	0.2352	0.2285
β_2	0.0104	0.0027	−0.0114	−0.0142	−0.0130
Model 3					
β_0	−0.7428	−0.6759	0.0027	0.6479	0.5634
β_1	−0.0069	0.0823	0.2194	0.2232	0.2355
β_2	0.0125	0.0028	−0.0122	−0.0128	−0.0142

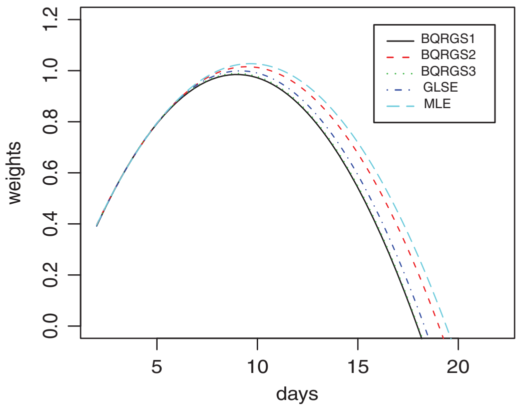

Figure 5. Five average growth curves. BQRGS1–BQRGS3 are the median estimates for Models 1–3, respectively.

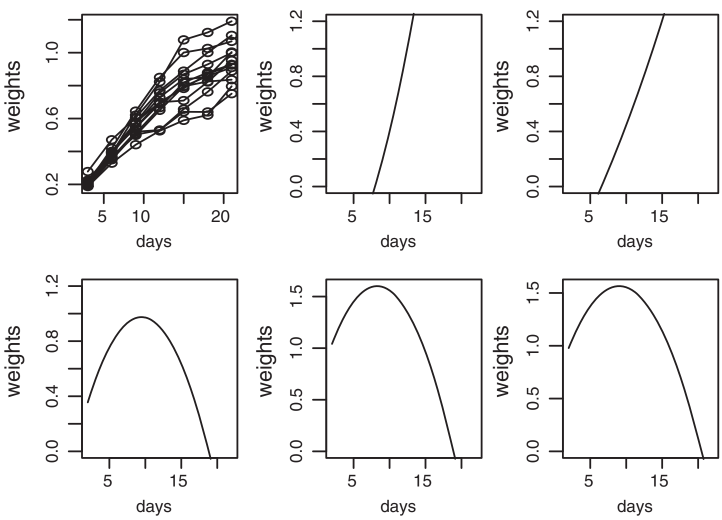

Figure 6. The mouse data and growth curves at five quantiles of Model 2. Top middle: $\tau = 0.15$; top right: $\tau = 0.25$; bottom left: $\tau = 0.50$; bottom middle: $\tau = 0.75$; bottom right:$\tau = 0.95$.

of the three models. Note that in all three models, the estimates of $\boldsymbol{\beta}$ are very similar at the corresponding quantiles, and in fact, the estimate of ϕ in Models 2 and 3 are 0.37 and 0.27, respectively, the random effects are not significant for this dataset.

6. Discussion and conclusion

In this paper, a hierarchical Bayesian quantile regression model is proposed for longitudinal data by assuming the ALD for the error term, in which random effects are incorporated into the model to account for the dependence among data. This approach is analogous to the traditional hierarchical Bayesian linear mixed model for the mean but allows the estimation at different quantiles of the conditional distribution and hence provides a more robust estimator and offers a more comprehensive understanding of the data. Though the ALD has no conjugate prior, our two MCMC algorithms deal with this problem successfully. The methods are relatively straightforward to implement and, unlike frequentist approaches to quantile regression, do not rely on estimation or approximation of the asymptotic variances of the estimated parameters. They also provide inferences and predictions, which fully take into account parameter uncertainty. Our methods are more general than that previously described by Geraci and Bottai [15] and permit greater flexibility in the analysis of random effects. In addition to longitudinal data, the proposed methods can be applied to other types of dependent data, such as clustered, hierarchical, and spatial data. In the simulation study, the proposed methods effectively describe the underlying conditional distribution and provide more efficient estimates than many previous methods when the errors are overdispersed. Although fully parametric, which may cause some concerns on the use of ALD, simulation results show that the Bayesian methods are quite insensitive to this assumption and behave well for data generated from other distributions.

With the Metropolis–Hastings algorithm, one practical issue is that, the tuning parameters in the proposal distribution need to be adjusted for each quantile. An adaptive and automatic method should be developed. Another future direction is to develop Bayesian quantile regression for non-parametric or semi-parametric models with random effects, which we are currently investigating.

Acknowledgements

We would like to thank the reviewer, Associate Editor and Editor Richard G. Krutchkoff for their thoughtful and useful comments. The first author especially thanks Professor Haibin Wang for his valuable comments on MCMC method during his visit to Nanyang Technological University. The work was partially supported by National Natural Science Foundation of China (No.10871201), Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (No. 10XNL018), the Major Project of Humanities Social Science Foundation of Ministry of Education (No. 08JJD910247), Key Project of Chinese Ministry of Education (No.108120), and Beijing Natural Science Foundation (No. 1102021).

References

[1] R. Koenker, *Quantile Regression*, Cambridge University Press, New York, 2005.
[2] K. Yu, Z. Lu, and J. Stander, *Quantile regression: Applications and current research area*, The Statist. 52 (2003), pp. 331–350.
[3] R. Koenker and J. Machado, *Goodness of fit and related inference processes for quantile regression*, J. Am. Statist. Assoc. 94 (1999), pp. 1296–1309.
[4] K. Yu and R.A. Moyeed, *Bayesian quantile regression*, Stat. Probab. Lett. 54 (2001), pp. 437–447.
[5] D.B. Dunson and J. Taylor, *Approximate Bayesian inference for quantiles*, J. Nonparametric Stat. 17 (2005), pp. 385–400.
[6] A. Kottas, and A.E. Gelfand, *Bayesian semiparametric median regression modeling*, J. Am. Stat. Assoc. 96 (2001), pp. 1458–1468.
[7] A. Kottas and M. Krnjajić, *Bayesian semiparametric modeling in quantile regression*, Scand. J. Stat. 36 (2009), pp. 297–319.
[8] T. Lancaster and S.J. Jun, *Bayesian quantile regression methods*, J. Appl. Econometrics 25 (2010), pp. 287–307.
[9] G. Verbeke and G. Molenberghs, *Linear Mixed Models for Longitudinal Data*, Springer, New York, 2000.
[10] S. Jung, *Quasi-likelihood for median regression models*, J. Am. Stat. Assoc. 91 (1996), pp. 251–257.
[11] S.R. Lipsitz, G.M. Fitzmaurice, G. Molenberghs, and L.P. Zhao, *Quantile regression methods for longitudinal data with drop-outs: application to CD4 cell counts of patients infected with the human immunodeficiency virus*, J. R. Stat. Soc. Ser. C 46 (1997), pp. 463–476.
[12] J.M. Robins, A. Rotnitzky, and L.P. Zhao, *Analysis of semiparametric regression models for repeated outcomes in the presence of missing data*, J. Am. Stat. Assoc. 90 (1995), pp. 106–121.
[13] A. Karlsson, *Nonlinear quantile regression estimation of longitudinal data*, Commun. Stat.-Simul. Comput. 37 (1996), pp. 114–131.
[14] R. Koenker, *Quantile regression for longitudinal data*, J. Multivariate Anal. 91 (2004), pp. 74–89.
[15] M. Geraci and M. Bottai, *Quantile regression for longitudinal data using the asymmetric Laplace distribution*, Biostatistics 8 (2007), pp. 140–154.
[16] Y. Liu and M. Bottai, *Mixed-effects models for conditional quantiles with longitudinal data*, Int. J. Biostatistics 5 (2009), Article 28.
[17] K.L. Lange, R.J.A. Little, and J.M.G. Taylor, *Robust statistical modeling using t-distribution*, J. Am. Stat. Assoc. 84 (1989), pp. 881–896.
[18] J.C. Pinheiro, C. Liu, and Y.N. Wu, *Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution*, J. Comput. Graph. Stat. 10 (2001), pp. 249–276.

[19] P.X.-K. Song, P. Zhang, and A. Qu, *Maximum likelihood inference in robust linear mixed-effects models using multivariate t distributions*, Stat. Sin. 17 (2007), pp. 929–943.
[20] W.K. Hestings, *Monte Carlo sampling methods using Markov chains and their application*, Biometrika 57 (1970), pp. 97–109.
[21] W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman and Hall, New York, 1996.
[22] A. Gelman, J. Carlin, H. Stern, and D. Rubin, *Bayesian Data Analysis*, Chapman and Hall, London, 1995.
[23] H. Kozumi and G. Kobayashi, *Gibbs sampling methods for the Bayesian quantile regression*, Technical report, Kobe University, Japan, 2009.
[24] J. Dagpunar, *An easily implemented generalized inverse Gaussian generator*, Commun. Stat.-Simul. Comput. 18 (1989), pp. 703–710.
[25] J.S. Williams and A.J. Izenman, *A class of linear spectral models and analysis for the study of longitudinal data*, Biometrics 45 (1981), pp. 831–849.
[26] C.R. Rao, *Prediction of future observations in polynomial growth curve models*, Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on Statistics: Applications and New Directions, Indian Statistical Institute, Calcutta, 1984, pp. 512–520.
[27] J.C. Lee, *Prediction and estimation of growth curve with special covariance structure*, J. Am. Stat. Assoc. 83 (1988), pp. 432–440.
[28] J.X. Pan and K.T. Fang, *Growth Curve Model and Statistical Diagnostics*, Springer, New York, 2002.
[29] D.V. Lindley and A.F.M. Smith, *Bayes estimates for the linear model (with discussion)*, J. R. Stat. Soc. Ser. B 34 (1972), pp. 1–41.
[30] O.E. Barndorff-Nielsen and N. Shephard, *Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics*, J. R. Stat. Soc. Ser. B 63 (2001), pp. 167–241.

Appendix: The full conditional distribution in Gibbs sampling algorithm

From the full hierarchical model, we can get

$$
\begin{aligned}
f(y_{ij}|\boldsymbol{\beta},\boldsymbol{\alpha}_i,e_{ij},\sigma) &= (2\pi k_2\sigma e_{ij})^{-1/2}\exp\left\{-\frac{1}{2k_2\sigma e_{ij}}(y_{ij}-\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}-z_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_i-k_1e_{ij})^2\right\}.\\
f(\boldsymbol{y}_i|\boldsymbol{\beta},\boldsymbol{\alpha}_i,\boldsymbol{e}_i,\sigma) &= \prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{\beta},\boldsymbol{\alpha}_i,e_{ij},\sigma)\\
&= (2\pi k_2\sigma)^{-\frac{n_i}{2}}\left(\prod_{j=1}^{n_i}e_{ij}\right)^{-1/2}\exp\left\{-\frac{1}{2k_2\sigma}\sum_{j=1}^{n_i}\left(\frac{y_{ij}-\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}-z_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_i-k_1e_{ij}}{\sqrt{e_{ij}}}\right)^2\right\}.\\
f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{e},\sigma) &= \prod_{i=1}^{n}\prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{\beta},\boldsymbol{\alpha}_i,e_{ij},\sigma)\\
&= (2\pi k_2\sigma)^{-N/2}\left(\prod_{i=1}^{n}\prod_{j=1}^{n_i}e_{ij}\right)^{-1/2}\exp\left\{-\frac{1}{2k_2\sigma}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left(\frac{y_{ij}-\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}-z_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_i-k_1e_{ij}}{\sqrt{e_{ij}}}\right)^2\right\}.
\end{aligned}
\tag{A1}
$$

Since the full conditional distribution of y_{ij} is a normal distribution, it is not difficult to derive the full conditional density of $\boldsymbol{\beta}$ given by

$$\pi(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{\alpha},\boldsymbol{e},\sigma) \propto f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{e},\sigma)\pi(\boldsymbol{\beta}).$$

According to Lindley and Smith [29],

$$\pi(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{\alpha},\boldsymbol{e},\sigma) \sim N_k(\boldsymbol{Bb},\boldsymbol{B}) \tag{A2}$$

where

$$
\begin{aligned}
\boldsymbol{B}^{-1} &= \frac{1}{k_2\sigma}\sum_i\sum_j\frac{\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathrm{T}}}{e_{ij}}+\boldsymbol{B}_0^{-1},\\
\boldsymbol{b} &= \frac{1}{k_2\sigma}\sum_i\sum_j\frac{\boldsymbol{x}_{ij}(y_{ij}-z_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_i-k_1e_{ij})}{e_{ij}}+\boldsymbol{B}_0^{-1}\boldsymbol{b}_0
\end{aligned}
$$

The full conditional density of $\boldsymbol{\alpha}_i$ is given by

$$\pi(\boldsymbol{\alpha}_i|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{e}_i, \sigma, \phi^2) \propto f(\boldsymbol{y}_i|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{e}, \sigma)\pi(\boldsymbol{\alpha}_i|\phi^2).$$

As the full conditional distribution of $\boldsymbol{y}_i$ and $\pi(\boldsymbol{\alpha}_i|\phi^2)$ are multivariate normal distributions, again use the result of Lindley and Smith [29],

$$\pi(\boldsymbol{\alpha}_i|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{e}_i, \sigma, \phi^2) \sim N_p(\boldsymbol{A}\boldsymbol{a}, \boldsymbol{A}) \tag{A3}$$

where

$$\boldsymbol{A}^{-1} = \frac{1}{k_2\sigma}\sum_j \frac{z_{ij}z_{ij}^{\mathrm{T}}}{e_{ij}} + \frac{1}{\phi^2}\boldsymbol{I},$$

$$\boldsymbol{a} = \frac{1}{k_2\sigma}\sum_j \frac{z_{ij}(y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_i - k_1 e_{ij})}{e_{ij}}$$

From Equation (A1) together with an exponential density, the full conditional distribution of e_{ij} is given by

$$\begin{aligned}\pi(e_{ij}|y_{ij}, \boldsymbol{\beta}, \boldsymbol{\alpha}_i, \sigma) &\propto f(y_{ij}|\boldsymbol{\beta}, \boldsymbol{\alpha}_i, e_{ij}, \sigma) f(e_{ij}|\sigma)\\ &\propto (e_{ij})^{-1/2}\exp\left\{-\frac{1}{2k_2\sigma e_{ij}}(y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_i - k_1 e_{ij})^2\right\}\exp\left\{-\frac{e_{ij}}{\sigma}\right\}\\ &\propto e_{ij}^{-1/2}\exp\left\{-\frac{1}{2}(\hat{\delta}_{ij}^2 e_{ij}^{-1} + \hat{\gamma}_{ij}^2 e_{ij})\right\},\end{aligned}$$

where

$$\hat{\delta}_{ij}^2 = \frac{(y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_i)^2}{k_2\sigma}, \quad \hat{\gamma}_{ij}^2 = \frac{k_1^2}{k_2\sigma} + \frac{2}{\sigma}.$$

Since it is the kernel of a generalized inverse Gaussian (GIG) distribution, we have

$$\pi(e_{ij}|\boldsymbol{y}, \boldsymbol{\alpha}_i, \boldsymbol{\beta}) \sim \mathrm{GIG}\left(\frac{1}{2}, \hat{\delta}_{ij}, \hat{\gamma}_{ij}\right), \tag{A4}$$

where the probability density function of $\mathrm{GIG}(\varrho, m, n)$ is given by

$$f(x|\varrho, m, n) = \frac{(n/m)^{\varrho}}{2K_{\varrho}(mn)} x_{\varrho-1}\exp\left\{-\frac{1}{2}(m^2x^{-1} + n^2x)\right\}, \quad x > 0,\ -\infty < \varrho < \infty,\ m, n \geq 0,$$

and $K_{\varrho}(\cdot)$ is a modified Bessel function of the third kind [30]. For the other parameters σ and ϕ^2,

$$\begin{aligned}\pi(\sigma|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{e}, \boldsymbol{\beta}, \phi^2) &\propto f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{e}, \sigma)\pi(\sigma)\\ &\sim \mathrm{IG}(\nu, \omega),\end{aligned}$$

where

$$\nu = \frac{N}{2} + c_0, \quad \omega = d_0 + \frac{1}{2k_2}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{(y_{ij} - \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_i - k_1 e_{ij})^2}{e_{ij}},$$

and

$$\begin{aligned}\pi(\phi^2|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma) &\propto \left(\prod_{i=1}^{n} f(\boldsymbol{\alpha}_i|\phi^2)\right)\pi(\phi^2)\\ &\sim \mathrm{IG}(\upsilon, \varpi),\end{aligned}$$

where

$$\upsilon = \frac{np}{2} + k_0, \quad \varpi = w_0 + \frac{1}{2}\sum_{i=1}^{n}\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{\alpha}_i.$$