Graduate School of
Business Administration

KOBE UNIVERSITY

ROKKO KOBE JAPAN

2009-2

Gibbs Sampling Methods for Bayesian Quantile Regression

Hideo Kozumi Genya Kobayashi

Discussion Paper Series

Gibbs Sampling Methods for Bayesian Quantile Regression

Hideo Kozumi[*,a], Genya Kobayashi[a]

[a]*Graduate School of Business Administration, Kobe University, 2-1 Rokko, Kobe 657-8501, Japan*

Abstract

This paper considers quantile regression models using an asymmetric Laplace distribution from a Bayesian point of view. We develop a simple and efficient Gibbs sampling algorithm for fitting the quantile regression model based on a location-scale mixture representation of the asymmetric Laplace distribution. It is shown that the resulting Gibbs sampler can be accomplished by sampling from either normal or generalized inverse Gaussian distribution. We also discuss some possible extensions of our approach, including the incorporation of a scale parameter, the use of double exponential prior, and a Bayesian analysis of Tobit quantile regression. The proposed methods are illustrated by both simulated and real data.

Key words: Asymmetric Laplace distribution, Bayesian quantile regression, double exponential prior, generalized inverse Gaussian distribution, Gibbs sampler, Tobit quantile regression

1. Introduction

Since the seminal work of Koenker and Bassett (1978), quantile regression has received increasing attention both from a theoretical and from an empirical viewpoint. It is a statistical procedure based on minimizing sums of asymmetrically weighted absolute residuals and can be used to explore the relationship between quantiles of the response distribution and available covariates. Since a set of quantiles often provides more complete description of the response distribution than the mean, quantile regression offers a practically important alternative to classical mean regression. There exists a large literature on quantile regression methods and we refer to Yu *et al*. (2003) and Koenker (2005) for an overview.

Let y_i be a response variable and $\mathbf{x}_i$ a $k \times 1$ vector of covariates for the i-th observation. Let $q_p(\mathbf{x}_i)$ denote the p-th $(0 < p < 1)$ quantile regression function of y_i given $\mathbf{x}_i$. Suppose that the relationship between $q_p(\mathbf{x}_i)$ and $\mathbf{x}_i$ can be modeled as $q_p(\mathbf{x}_i) = \mathbf{x}_i'\boldsymbol{\beta}_p$, where $\boldsymbol{\beta}_p$ is a vector of unknown parameters of interest. Then we consider the quantile regression model given by

$$y_i = \mathbf{x}_i'\boldsymbol{\beta}_p + \epsilon_i, \quad (i = 1, \ldots, n),$$

where ϵ_i is the error term whose distribution (with density, say, $f_p(\cdot)$) is restricted to have the p-th quantile equal to zero, that is, $\int_{-\infty}^{0} f_p(\epsilon_i) d\epsilon_i = p$.

[*]Corresponding author

Email addresses: kozumi@kobe-u.ac.jp (Hideo Kozumi), genyako@hotmail.co.jp (Genya Kobayashi)

January 30, 2009

The error density $f_p(\cdot)$ is often left unspecified in the classical literature. Thus, quantile regression estimation for $\boldsymbol{\beta}_p$ proceeds by minimizing

$$\sum_{i=1}^{n} \rho_p \left(y_i - \mathbf{x}_i' \boldsymbol{\beta}_p \right), \tag{1}$$

where $\rho_p(\cdot)$ is the check (or loss) function defined by

$$\rho_p(u) = u \left\{ p - I(u < 0) \right\}, \tag{2}$$

and $I(\cdot)$ denotes the usual indicator function. Since, however, the check function is not differentiable at zero, we cannot derive explicit solutions to the minimization problem. Therefore, linear programing methods are commonly applied to obtain quantile regression estimates for $\boldsymbol{\beta}_p$ (see, *e.g.*, Koenker and Park (1996) and Portnoy and Koenker (1997)).

From a Bayesian point of view, Walker and Mallick (1999), Kottas and Gelfand (2001), and Hanson and Johnson (2002) considered median regression, which is a special case of quantile regression with $p = 0.5$, and discussed nonparametric modeling for the error distribution based on either Pólya tree or Dirichlet process priors. Regarding general quantile regression, Yu and Moyeed (2001) proposed a Bayesian modeling approach by noting that minimizing (1) is equivalent to maximizing a likelihood function under the asymmetric Laplace error distribution (see also Tsionas (2003)).

As discussed in Yu and Moyeed (2001), the use of the asymmetric Laplace distribution for the error terms provides a natural way to deal with the Bayesian quantile regression problem. However, the resulting posterior density for $\boldsymbol{\beta}_p$ is not analytically tractable due to the complexity of the likelihood function. Therefore, Yu and Moyeed (2001) considered Markov chain Monte Carlo (MCMC) methods for posterior inference. Specifically, they used a random walk Metropolis 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 p through the likelihood function. As a result, tuning parameters of proposals such as a proposal step size need to be adjusted so as to attain some appropriate acceptance rates for each value of p, and this limits the applicability of the random walk sampler in practice.

This paper considers Bayesian quantile regression models using the asymmetric Laplace distribution and proposes MCMC methods that are not only computationally efficient but also easy to implement. In particular, we develop the Gibbs sampling algorithms based on a location-scale mixture representation of the asymmetric Laplace distribution. It is shown that the mixture representation provides fully tractable conditional posterior densities and considerably simplifies the existing estimation procedures for quantile regression models. To our knowledge, no previous work has employed a location-scale mixture representation to analyze quantile regression models. Furthermore, we show that our approach can readily incorporate a scale parameter and can be directly extended to Tobit quantile regression.

The rest of the paper is organized as follows. In section 2 we present a mixture representation of an asymmetric Laplace distribution and derive full conditional densities for parameters. Section 3 discusses some possible extensions of our approach, including the incorporation of a scale parameter, the use of double exponential prior for regression coefficients, and a Bayesian analysis of Tobit quantile regression models. Some numerical results for both simulated and real data are presented in Section 4, and some conclusions are given in the final section.

2. Posterior inference

2.1. Mixture representation

Following Yu and Moyeed (2001), we consider the linear model given by

$$y_i = \mathbf{x}_i'\boldsymbol{\beta}_p + \epsilon_i, \quad (i = 1, \ldots, n),$$

and assume that ϵ_i has the asymmetric Laplace distribution with density

$$f_p(\epsilon_i) = p(1-p)\exp\left\{-\rho_p(\epsilon_i)\right\}, \tag{3}$$

where $\rho_p(\cdot)$ is defined in (2). The parameter p determines the skewness of distribution and the p-th quantile of this distribution is zero. It is also known that the mean and variance of the asymmetric Laplace distribution with density (3) are given, respectively, by

$$E(\epsilon_i) = \frac{1-2p}{p(1-p)} \quad \text{and} \quad \mathrm{Var}(\epsilon_i) = \frac{1-2p+2p^2}{p^2(1-p)^2}.$$

Some other properties of the asymmetric Laplace distribution can be found in Yu and Zhang (2005).

As shown in Kotz *et al*. (1998), the asymmetric Laplace distribution has various mixture representations. For example, if ξ and η are independent and identical standard exponential distributions, $\xi/p - \eta/(1-p)$ has the asymmetric Laplace distribution. To develop Gibbs sampling algorithms for the quantile regression model, we utilize a mixture representation based on exponential and normal distributions, which is given by the following proposition.

Proposition. *Let z be an standard exponential variable and u a standard normal variable. If a random variable ϵ follows the asymmetric Laplace distribution with density (3), then we can represent ϵ as a location-scale mixture of normals given by*

$$\epsilon = \theta z + \tau\sqrt{z}u,$$

where

$$\theta = \frac{1-2p}{p(1-p)} \quad \textit{and} \quad \tau^2 = \frac{2}{p(1-p)}.$$

Proof. See Appendix. □

From this result, the response y_i can be equivalently rewritten as

$$y_i = \mathbf{x}_i'\boldsymbol{\beta}_p + \theta z_i + \tau\sqrt{z_i}u_i, \tag{4}$$

where $z_i \sim \mathcal{E}(1)$ and $u_i \sim \mathcal{N}(0,1)$ are mutually independent, and $\mathcal{E}(\psi)$ denotes an exponential distribution with mean ψ. As the conditional distribution of y_i given z_i is normal with mean $\mathbf{x}_i'\boldsymbol{\beta}_p + \theta z_i$ and variance $\tau^2 z_i$, the joint density of $\mathbf{y} = (y_1, \ldots, y_n)'$ is given by

$$f(\mathbf{y}|\boldsymbol{\beta}_p, \mathbf{z}) \propto \left(\prod_{i=1}^{n} z_i^{-\frac{1}{2}}\right) \exp\left\{-\sum_{i=1}^{n} \frac{(y_i - \mathbf{x}_i'\boldsymbol{\beta}_p - \theta z_i)^2}{2\tau^2 z_i}\right\}, \tag{5}$$

where $\mathbf{z} = (z_1, \ldots, z_n)'$.

2.2. Gibbs sampler

To proceed a Bayesian analysis, we assume the prior

$$\boldsymbol{\beta}_p \sim \mathcal{N}(\boldsymbol{\beta}_{p0}, \mathbf{B}_{p0}), \tag{6}$$

where $\boldsymbol{\beta}_{p0}$ and $\mathbf{B}_{p0}$ are the prior mean and covariance of $\boldsymbol{\beta}_p$, respectively. As proved in Yu and Moyeed (2001), all posterior moments of $\boldsymbol{\beta}_p$ exist under the normal prior (6). Using data augmentation, a Gibbs sampling algorithm for the quantile regression model is constructed by sampling $\boldsymbol{\beta}_p$ and $\mathbf{z}$ from their full conditional distributions. Since (4) is a normal linear regression model conditionally on z_i, it is not difficult to derive the full conditional density of $\boldsymbol{\beta}_p$ given by

$$\boldsymbol{\beta}_p|\mathbf{y}, \mathbf{z} \sim \mathcal{N}(\hat{\boldsymbol{\beta}}_p, \hat{\mathbf{B}}_p), \tag{7}$$

where

$$\hat{\mathbf{B}}_p^{-1} = \sum_{i=1}^{n} \frac{\mathbf{x}_i\mathbf{x}_i'}{\tau^2 z_i} + \mathbf{B}_{p0}^{-1}, \quad \hat{\boldsymbol{\beta}}_p = \hat{\mathbf{B}}_p \left\{\sum_{i=1}^{n} \frac{\mathbf{x}_i(y_i - \theta z_i)}{\tau^2 z_i} + \mathbf{B}_{p0}^{-1}\boldsymbol{\beta}_{p0}\right\}.$$

From (5) together with a standard exponential density, the full conditional distribution of z_i is proportional to

$$z_i^{-1/2} \exp\left\{-\frac{1}{2}\left(\hat{\delta}_i^2 z_i^{-1} + \hat{\gamma}_i^2 z_i\right)\right\}, \tag{8}$$

where $\hat{\delta}_i^2 = (y_i - \mathbf{x}_i'\boldsymbol{\beta})^2/\tau^2$ and $\hat{\gamma}_i^2 = 2 + \theta^2/\tau^2$. Since (8) is the kernel of a generalized inverse Gaussian distribution, we have

$$z_i|\mathbf{y}, \boldsymbol{\beta}_p \sim \mathcal{GIG}\left(\frac{1}{2}, \hat{\delta}_i, \hat{\gamma}_i\right), \tag{9}$$

where the probability density function of $\mathcal{GIG}(\nu, a, b)$ is given by

$$f(x|\nu, a, b) = \frac{(b/a)^\nu}{2K_\nu(ab)} x^{\nu-1} \exp\left\{-\frac{1}{2}(a^2x^{-1} + b^2x)\right\}, \quad x > 0, \ -\infty < \nu < \infty, \ a, b \geq 0,$$

and $K_\nu(\cdot)$ is a modified Bessel function of the third kind (see Barndorff-Nielsen and Shephard (2001)). There exist efficient algorithms to simulate from a generalized inverse Gaussian distribution (see, *e.g.*, Dagpunar (1989) and Hörmann *et al.* (2004)), so that our Gibbs sampler defined in (7) and (9) can be easily applied to quantile regression estimation.

We note that Tsionas (2003) also developed a Gibbs sampling algorithm for the quantile regression model by employing a different representation of the asymmetric Laplace distribution. Although it does not require Metropolis–Hastings steps, the Gibbs sampler proposed by Tsionas (2003) needs to update each element of $\boldsymbol{\beta}_p$ separately. Thus his sampling algorithm may produce highly correlated draws and become less efficient than our algorithm (see Liu *et al.* (1994)). Furthermore, Tsionas (2003) mentioned that his Gibbs sampler is complicated and can be somewhat slow when the number of observations is large.

3. Some extensions

3.1. Inference with scale parameter

In the previous section, we have considered the quantile regression model without taking into account a scale parameter. If one may be interested to introduce a scale parameter $\sigma > 0$ into the model, our approach can incorporate it by rewriting (4) as

$$y_i = \mathbf{x}_i'\boldsymbol{\beta}_p + \sigma\theta z_i + \sigma\tau\sqrt{z_i}u_i. \tag{10}$$

However this expression is not convenient to develop Gibbs samplers as the scale parameter appears in the conditional mean of y_i. Therefore, we reparameterize (10) as

$$y_i = \mathbf{x}_i'\boldsymbol{\beta}_p + \theta v_i + \tau\sqrt{\sigma v_i}u_i,$$

where $v_i = \sigma z_i$. To complete the model specification, we assume that $\boldsymbol{\beta}_p \sim \mathcal{N}(\boldsymbol{\beta}_{p0}, \mathbf{B}_{p0})$ and $\sigma \sim \mathcal{IG}(n_0/2, s_0/2)$, where $\mathcal{IG}(a, b)$ denotes an inverse Gamma distribution with parameters a and b.

We now need to sample $\boldsymbol{\beta}_p$, $\mathbf{v} = (v_1, \ldots, v_n)'$ and σ from their conditional distributions. The usual Bayesian calculations show that the full conditional density of $\boldsymbol{\beta}_p$ is given by

$$\boldsymbol{\beta}_p|\mathbf{y}, \mathbf{v}, \sigma \sim \mathcal{N}(\tilde{\boldsymbol{\beta}}_p, \tilde{\mathbf{B}}_p), \tag{11}$$

where $\tilde{\mathbf{B}}_p^{-1} = \sum_{i=1}^n \frac{\mathbf{x}_i\mathbf{x}_i'}{\tau^2\sigma v_i} + \mathbf{B}_{p0}^{-1}$ and $\tilde{\boldsymbol{\beta}}_p = \hat{\mathbf{B}}_p\left\{\sum_{i=1}^n \frac{\mathbf{x}_i(y_i-\theta v_i)}{\tau^2\sigma v_i} + \mathbf{B}_{p0}^{-1}\boldsymbol{\beta}_{p0}\right\}$. Similarly to the previous section, we can easily obtain that

$$v_i|\mathbf{y}, \boldsymbol{\beta}_p, \sigma \sim \mathcal{GIG}\left(\frac{1}{2}, \tilde{\delta}_i, \tilde{\gamma}_i\right), \tag{12}$$

where $\tilde{\delta}_i^2 = (y_i - \mathbf{x}_i'\boldsymbol{\beta})^2/\tau^2\sigma$, and $\tilde{\gamma}_i^2 = 2/\sigma + \theta^2/\tau^2\sigma$. By noting that $v_i \sim \mathcal{E}(\sigma)$, the full conditional density of σ is proportional to

$$\left(\frac{1}{\sigma}\right)^{\frac{n_0}{2}+\frac{3}{2}n+1} \exp\left[-\frac{1}{\sigma}\left\{\frac{s_0}{2} + \sum_{i=1}^n v_i + \sum_{i=1}^n \frac{(y_i - \mathbf{x}_i'\boldsymbol{\beta}_p - \theta v_i)^2}{2\tau^2 v_i}\right\}\right],$$

so that we have

$$\sigma|\mathbf{y}, \boldsymbol{\beta}_p, \mathbf{v} \sim \mathcal{IG}\left(\frac{\tilde{n}}{2}, \frac{\tilde{s}}{2}\right), \tag{13}$$

where $\tilde{n} = n_0 + 3n$ and $\tilde{s} = s_0 + 2\sum_{i=1}^n v_i + \sum_{i=1}^n (y_i - \mathbf{x}_i'\boldsymbol{\beta}_p - \theta v_i)^2/\tau^2 v_i$. Consequently, the introduction of scale parameter does not cause any difficulties in our Gibbs sampling algorithm.

3.2. Double exponential prior

Instead of a normal prior, we consider an important alternative, that is, a double exponential prior for $\boldsymbol{\beta}_p$. The density of double exponential prior is given by

$$\pi(\boldsymbol{\beta}_p) = \prod_{j=1}^{k} \pi(\beta_{pj}) \propto \prod_{j=1}^{k} \exp\left(-\lambda_0|\beta_{pj} - \beta_{pj0}|\right),$$

where β_{pj} is the jth element of $\boldsymbol{\beta}_p$, and β_{pj0} and λ_0 are hyperparameters. Using this prior distribution, Yu and Stander (2005) showed that all posterior moments of $\boldsymbol{\beta}_p$ exist for Tobit quantile regression models. Recently, Park and Casella (2008) considered this prior in the context of Lasso estimation (Tibshirani (1996)) and discussed the choice of hyperparameters.

As shown in Park and Casella (2008), the double exponential density can be expressed as

$$\pi(\beta_{pj}) = \int_0^\infty \frac{1}{\sqrt{2\pi\omega_j}} \exp\left\{-\frac{(\beta_{pj} - \beta_{pj0})^2}{2\omega_j}\right\} \exp\left(-\frac{\lambda_0^2\omega_j}{2}\right) d\omega_j$$

where ω_j has an exponential distribution with mean $2/\lambda^2$. This suggests the following hierarchical representation of the prior:

$$\boldsymbol{\beta}_p|\boldsymbol{\omega} \sim \mathcal{N}(\boldsymbol{\beta}_{p0}, \boldsymbol{\Omega}),$$
$$\omega_i \sim \mathcal{E}(2/\lambda_0^2),$$

where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_k)'$ and $\boldsymbol{\Omega}$ is a diagonal matrix with the jth element ω_j. It follows from this specification that the full conditional distributions of $\boldsymbol{\beta}_p$, σ and $\mathbf{v}$ are the same as those under the normal prior with $\mathbf{B}_{p0}$ replaced with $\boldsymbol{\Omega}$. Also, the full conditional density of ω_j is proportional to

$$\omega_j^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left\{(\beta_{pj} - \beta_{pj0})^2\omega_j^{-1} + \lambda_0^2\omega_j\right\}\right],$$

which implies that

$$\omega_j|\mathbf{y}, \boldsymbol{\beta}_p, \sigma, \mathbf{v} \sim \mathcal{GIG}\left(\frac{1}{2}, |\beta_{pj} - \beta_{pj0}|, \lambda_0\right).$$

3.3. Tobit quantile regression

Tobit quantile regression models have received much attention in the classical literature (see, *e.g.*, Powell (1986), Hahn (1995), Buchinsky and Hahn (1998) and Bilias *et al.* (2000)). Yu and Stander (2005) proposed a Bayesian framework for Tobit quantile regression based on the asymmetric Laplace distribution. Here we show that our methodology is directly extended to the analysis of Tobit quantile regression models.

As in the standard Tobit model, we assume that the response y_i is generated according to

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > 0, \\ 0 & \text{if } y_i^* \leq 0, \end{cases}$$
$$y_i^* = \mathbf{x}_i'\boldsymbol{\beta}_p + \epsilon_i, \qquad (14)$$

where y_i^* is a latent variable. To develop a Tobit quantile regression model, we again assume that ϵ_i has the asymmetric Laplace distribution and rewrite (14) as

$$y_i^* = \mathbf{x}_i'\boldsymbol{\beta}_p + \theta v_i + \tau\sqrt{\sigma v_i}u_i, \tag{15}$$

where $u_i \sim \mathcal{N}(0, 1)$ and $v_i \sim \mathcal{E}(\sigma)$.

Since the model (15) has a regression form conditionally on v_i, the method developed by Chib (1992) can be applied to the sampling of y_i^*, that is,

$$y_i^*|\mathbf{y}, \boldsymbol{\beta}_p, \mathbf{v}, \sigma \sim y_i I(y_i > 0) + \mathcal{TN}_{(-\infty,0]}(\mathbf{x}_i'\boldsymbol{\beta}_p + \theta v_i, \tau^2\sigma v_i)I(y_i = 0),$$

where $\mathcal{TN}_{(a,b]}(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2 truncated on the interval $(a, b]$. Moreover, by assuming that $\boldsymbol{\beta}_p \sim \mathcal{N}(\boldsymbol{\beta}_{p0}, \mathbf{B}_{p0})$ and $\sigma \sim \mathcal{IG}(n_0/2, s_0/2)$, the full conditional distributions of $\boldsymbol{\beta}_p$, $\mathbf{v}$ and σ can be easily obtained from (11)–(13) by replacing y_i everywhere with y_i^*. Thus our approach offers a simple sampling method for Tobit quantile regression models compared with that of Yu and Stander (2005), which relies on Metropolis-Hastings algorithms and requires to control tuning parameters for each value of p.

4. Numerical examples

We present three examples to illustrate the Gibbs sampling methods developed in Sections 2 and 3. The first example uses a simulated data set and the second is based on the real data analyzed by Wang *et al*. (1998). While the first and second examples estimate the quantile regression models, the third example considers the Tobit quantile regression model using the data from Mroz (1987).

4.1. Simulated data

A data set of $n = 100$ observations was generated from the model

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \epsilon, \quad \epsilon \sim \mathcal{N}(0, 1),$$

where each covariate was simulated from a uniform distribution on $(-2,\ 2)$ and all the β_j were set to one. The normal prior (6) was used for $\boldsymbol{\beta}_p$ and the hyperparameters were chosen as $\boldsymbol{\beta}_{p0} = \mathbf{0}$ and $\mathbf{B}_{p0} = 100\mathbf{I}$. We fitted the quantile regression model without a scale parameter for $p = 0.05$, 0.5 and 0.95.

To assess the sampling efficiency of our algorithm, we estimated inefficiency factors for $\boldsymbol{\beta}_p$ by running the Gibbs sampler for 15000 iterations with an initial burn-in of 5000 iterations. The inefficiency factor is defined as a ratio of the numerical variance of the sample mean from the Markov chain to the variance from independent draws (see Chib (2001)). The results are summarized in Table 1 together with those obtained from the random walk sampler of Yu and Moyeed (2001). When applying the random walk sampler, we adjusted proposal step sizes in order that the inefficiency factors achieve the lowest values.

It is observed that the inefficiency factors for the Gibbs sampler are smaller than those for the random walk sampler in all the cases. This indicates that our Gibbs sampler is more efficient than the random walk sampler. It is also found that both the Gibbs sampler and the random walk algorithm attain the smallest inefficiency factors in the case of $p = 0.5$.

4.2. Patent data

As an real data example, we consider the data examined by Wang *et al.* (1998). The data set contains information about the number of patent applications from 70 pharmaceutical and biomedical companies in 1976. A more detailed explanation of the data can be found in Hall *et al.* (1988). Following Tsionas (2003), we analyze the relationship between patents and research and development (R&D) by estimating the model

$$\log(1+N) = \beta_1 + \beta_2 \log(RD) + \beta_3 \log(RD)^2 + \beta_4 \log\left(\frac{RD}{SALE}\right) + \epsilon,$$

where N is the number of patent applications, RD is R&D spending, and $RD/SALE$ is the ratio of R&D to sales. In this study, we included a scale parameter σ and assumed that $\boldsymbol{\beta}_p \sim \mathcal{N}(\mathbf{0}, 100\mathbf{I})$ and $\sigma \sim \mathcal{IG}(5/2, 0.1/2)$. Using the prior specification, we ran the Gibbs sampler for 20000 iterations and discarded the first 50000 iterations as a burn-in period.

Figure 1 shows the sample autocorrelation functions of $\boldsymbol{\beta}_p$ for $p = 0.5$. It is observed that the autocorrelations from the Gibbs sampler generally decline to zero by lag 20. Although different priors are employed in Tsionas (2003), a comparison with the results shown in Figure 3 of Tsionas (2003) shows that the autocorrelations from our Gibbs sampler decay more quickly than those from the Gibbs sampler of Tsionas (2003). This is because the Gibbs sampler of Tsionas (2003) updates each element of $\boldsymbol{\beta}_p$ separately while our algorithm draws all the elements jointly.

4.3. Labor supply data

To illustrate our method for Tobit quantile regression models, we consider the data set from Mroz (1987), which is comprised of 753 married women between the ages of 30 and 60. The response variable is given by the total number of hours the wife worked for a wage outside the home during year 1975 and measured in 100 hours. Of the 753 women in the sample, 325 of the women worked zero hours and the corresponding responses are treated as left censored at zero. As the explanatory variables, we include a constant term, income which is not due to the wife ($nwifeinc$), years of education ($educ$), years of work experience ($exper$), wife's age (age), the number of children under 6 years old ($kidslt6$), and the number of children over 6 years old ($kidsge6$).

For the regression coefficients $\boldsymbol{\beta}_p$, we considered both the normal and the double exponential prior, and specified the following hyperparameters: $\boldsymbol{\beta}_{p0} = \mathbf{0}$, $\mathbf{B}_{p0} = 100\mathbf{I}$ and $\lambda_0 = 0.14$. Note that the prior variances are the same under both the priors. Furthermore, we incorporated a scale parameter σ and assumed that $\sigma \sim \mathcal{IG}(10/2, 0.02/2)$. All results are based on an MCMC sample of 10000 draws obtained after a burn-in of 5000 iterations.

Table 2 shows the posterior estimates of the parameters. From the table, we see that both the priors yield very similar posterior estimates and notice that posterior means of some parameters (*e.g.*, the coefficients on $educ$ and $kidslt6$) are different across the different values of p. To study more closely the relationships between the quantiles and the posterior estimates, Figure 2 plots the posterior estimates of $\boldsymbol{\beta}_p$ against various values of p under the normal prior. We observe from the figure that $nwifeinc$, $kidslt6$ and age are negatively associated with the wife's work in the lower and middle quantiles, but $nwifeinc$ and $kidslt6$ become less related in the higher quantiles. On the contrary, $educ$ and $exper$ have positive effects on the wife's working time in the lower and middle quantiles, but they are also less related in the higher quantiles. Finally, the results for $expersq$ and $kidsge6$ show almost flat paths compared with the other covariates, indicating that they have constant effects on the wife's propensity to work.

5. Conclusions

We have developed a Gibbs sampling method for quantile regression models based on the location-scale mixture representation of the asymmetric Laplace distribution. The proposed Gibbs sampling algorithm is easy to implement in practice since one cycle of the algorithm can be accomplished by the simulation from a normal and a generalized inverse Gaussian distribution. We have also discussed some extensions of our approach, which are inference with scale parameters, the use of double exponential priors, and Tobit quantile regression analysis. Finally, we have provided illustrations with simulated and real data sets and shown the superiority of our Gibbs sampler to the existing sampling methods.

The mixture representation utilized in this paper allows us to express a quantile regression model as a normal regression model. Therefore, our approach can be further extended to more complicated models such as nonlinear models, and this is left for future research.

Acknowledgments

This research is partly supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan. The computational results are obtained using Ox version 5.10 (Doornik (2007)).

[2009.1.30. 906]

Appendix

Suppose that a random variable ϵ has the probability density function given by

$$f_p(\epsilon) = p(1-p)\exp\left\{-\rho_p(\epsilon)\right\},$$

where $\rho_p(\epsilon) = \epsilon\{p - I(\epsilon < 0)\}$. Then the characteristic function of ϵ is derived as

$$\begin{aligned}
\varphi(t) &= E[e^{it\epsilon}] \\
&= \int_{-\infty}^{\infty} e^{it\epsilon} f_p(\epsilon)\, d\epsilon \\
&= \int_{-\infty}^{0} p(1-p)e^{it\epsilon+(1-p)\epsilon} d\epsilon + \int_{0}^{\infty} p(1-p)e^{it\epsilon-p\epsilon} d\epsilon \\
&= p(1-p)\left\{\frac{1}{it+(1-p)} + \frac{1}{p-it}\right\} \\
&= \left\{\frac{1}{p(1-p)}t^2 - i\frac{1-2p}{p(1-p)}t + 1\right\}^{-1},
\end{aligned}$$

where $i^2 = -1$.

Let z be an standard exponential variable and u a standard normal variable. The characteristic function of $\epsilon' = \theta z + \tau\sqrt{z}u$ can be expressed as

$$\begin{aligned}
\phi(t) &= E\left[e^{it(\theta z+\tau\sqrt{z}u)}\right] \\
&= \int_0^{\infty} e^{it\theta z} E\left[e^{it\tau\sqrt{z}u}\right] e^{-z} dz.
\end{aligned}$$

Since u follows a standard normal distribution, we have

$$E\left[e^{it\tau\sqrt{z}u}\right] = e^{-\frac{1}{2}t^2\tau^2 z}.$$

Thus, the characteristic function of ϵ' is obtained as

$$\begin{aligned}
\phi(t) &= \int_0^{\infty} e^{-z\left(1+\frac{1}{2}t^2\tau^2-i\theta t\right)} dz \\
&= \left(\frac{1}{2}\tau^2 t^2 - i\theta t + 1\right)^{-1}.
\end{aligned}$$

These two characteristic functions are equivalent when $\theta = \frac{1-2p}{p(1-p)}$ and $\tau^2 = \frac{2}{p(1-p)}$.

References

[1] Barndorff-Nielsen, O.E. and Shephard, N. (2001). “Non–Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics,” *Journal of the Royal Statistical Society* **B63**, 167–241.

[2] Bilias, Y., Chen, S., and Ying, Z. (2000). “Simple resampling methods for censored regression quantiles,” *Journal of Econometrics* **68**, 303–338.

[3] Buchinsky, M. and Hahn, J. (1998). “An alternative estimator for censored quantile regression,” *Econometrica* **66**, 653–671.

[4] Chib, S. (1992). “Bayes inference in the Tobit censored regression model,” *Journal of Econometrics* **51**, 79–99.

[5] Chib, S. (2001). “Markov chain Monte Carlo methods: computation and inference,” in J.J. Heckman and E. Leamer (eds.), *Handbook of Econometrics* **5**, 3569–3649. Amsterdam: North Holland.

[6] Dagnapur, J. (1989). “An easily implemented generalized inverse Gaussian generator,” *Communications in Statistics–Simulation* **18**, 703–710.

[7] Doornik, J.A. (2007). *Object-Oriented Matrix Programming Using Ox* (3rd ed.). London: Timberlake Consultants Press and Oxford.

[8] Hahn, J. (1995). “Bootstrapping quantile regression estimators,” *Econometric Theory* **11**, 105–121.

[9] Hall, B. H., Cummins, C., Laderman, E., and Mundy, J. (1988). ”The R&D master file documentation,” Technical Working Paper 72. National Bureau of Economic Research, Cambridge, MA.

[10] Hanson, T. and Johnson, W.O. (2002). “Modeling regression error with a mixture of Pólya trees,” *Journal of the American Statistical Association* **97**, 1020–1033.

[11] Hörmann, W., Leydold, J., and Derflinger, G. (2004). *Automatic Nonuniform Random Variate Generation*. Berlin: Springer.

[12] Koenker, R. (2005). *Quantile Regression*. New York: Cambridge University Press.

[13] Koenker, R. and Bassett, G. (1978). “Regression quantiles,” *Econometrica* **46**, 33–50.

[14] Koenker, R. and Park, B.J. (1996). “An interior point algorithm for nonlinear quantile regression,” *Journal of Econometrics* **71**, 265–283.

[15] Kottas, A. and Gelfand, A.E. (2001). “Bayesian semiparametric median regression modeling,” *Journal of American Statistical Association* **96**, 1458–1468.

[16] Kotz, S., Kozubowski, T.J., and Podgórski, K. (2001). *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Boston: Engineering, and Finance*. Birkhäuser.

[17] Liu, J.S., Wong, W.H., and Kong, A. (1994). "Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes," *Biometrika* **81**, 27–40.

[18] Mroz, T. (1987). "The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions," *Econometrica* **55**, 765–799.

[19] Park, T. and Casella, G. (2008). "The Bayesian Lasso," *Journal of the American Statistical Association* **103**, 681–686.

[20] Portnoy, S. and Koenker, R. (1997). "The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators (with discussion)," *Statistical Science* **12**, 279–300.

[21] Powell, J. (1986). "Censored regression quantiles," *Journal of Econometrics* **32**, 143–155.

[22] Tibshirani, R. (1996). "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society* **B58**, 267–288.

[23] Tsioans, E. (2003). "Bayesian quantile inference," *Journal of Statistical Computation and Simulation* **73**, 659–674.

[24] Walker, S.G. and Mallick, B.K. (1999). "A Bayesian semiparametric accelerated failure time model," *Biometrics* **55**, 477–483.

[25] Wang, P., Cockburn, I., and Puterman, M. K. (1998). "Analysis of patent data – A mixed-Poisson-regression-model approach,“ *Journal of Business and Economic Statistics* **16**, 2741.

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

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

[28] Yu, K. and Stander, J. (2005). "Bayesian analysis of a Tobit quantile regression model," *Journal of Econometrics* **137**, 260–276.

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

Table 1: Simulated data: Inefficiency factors are shown.

	Gibbs sampler			Random walk sampler		
p	0.05	0.5	0.95	0.05	0.5	0.95
β_1	11.902	3.244	14.141	59.486	16.348	45.394
β_2	12.150	3.904	16.663	56.312	14.832	51.253
β_3	20.085	3.540	11.262	71.708	24.904	20.252
β_4	13.743	5.564	17.461	81.104	24.578	59.490
β_5	10.572	3.615	11.486	39.880	21.221	69.212

Table 2: Labor supply data: Posterior means and standard deviations are shown.

Normal prior						
Explanatory	$p = 0.05$		$p = 0.5$		$p = 0.95$	
Variable	Mean	SD	Mean	SD	Mean	SD
constant	-5.833	3.083	11.883	4.120	22.803	3.558
nwifeinc	-0.132	0.043	-0.099	0.045	-0.059	0.036
educ	0.858	0.139	0.859	0.209	0.331	0.178
exper	0.948	0.155	1.413	0.181	0.798	0.181
$exper^2$	-0.018	0.004	-0.018	0.006	-0.012	0.006
age	-0.354	0.057	-0.607	0.072	-0.269	0.065
*kidslt*6	-7.237	1.336	-9.630	1.142	-4.855	1.145
*kidsge*6	-0.131	0.278	-0.437	0.394	0.223	0.400
Double exponential prior						
Explanatory	$p = 0.05$		$p = 0.5$		$p = 0.95$	
Variable	Mean	SD	Mean	SD	Mean	SD
constant	-5.290	3.201	11.274	4.551	23.898	3.653
nwifeinc	-0.129	0.041	-0.100	0.045	-0.057	0.037
educ	0.839	0.150	0.883	0.216	0.307	0.177
exper	0.929	0.142	1.410	0.182	0.769	0.188
$exper^2$	-0.017	0.004	-0.018	0.006	-0.011	0.006
age	-0.361	0.058	-0.600	0.074	-0.282	0.067
*kidslt*6	-7.179	1.243	-9.611	1.160	-4.820	1.236
*kidsge*6	-0.159	0.256	-0.402	0.404	0.160	0.380

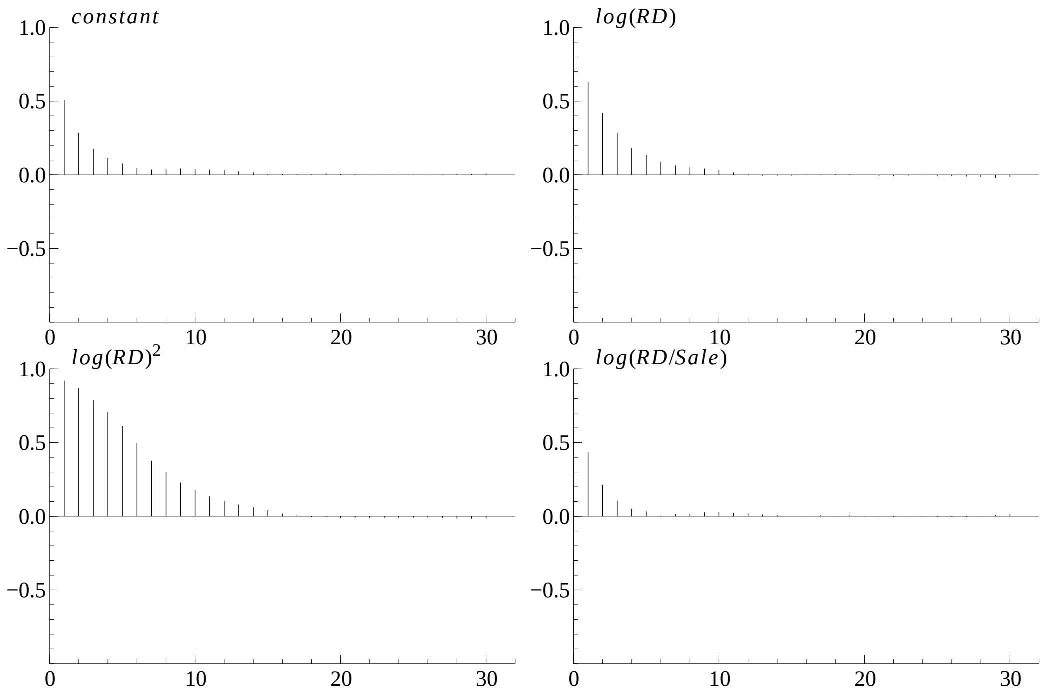

Figure 1: Patent data: Sample autocorrelation functions are shown for $p = 0.5$.

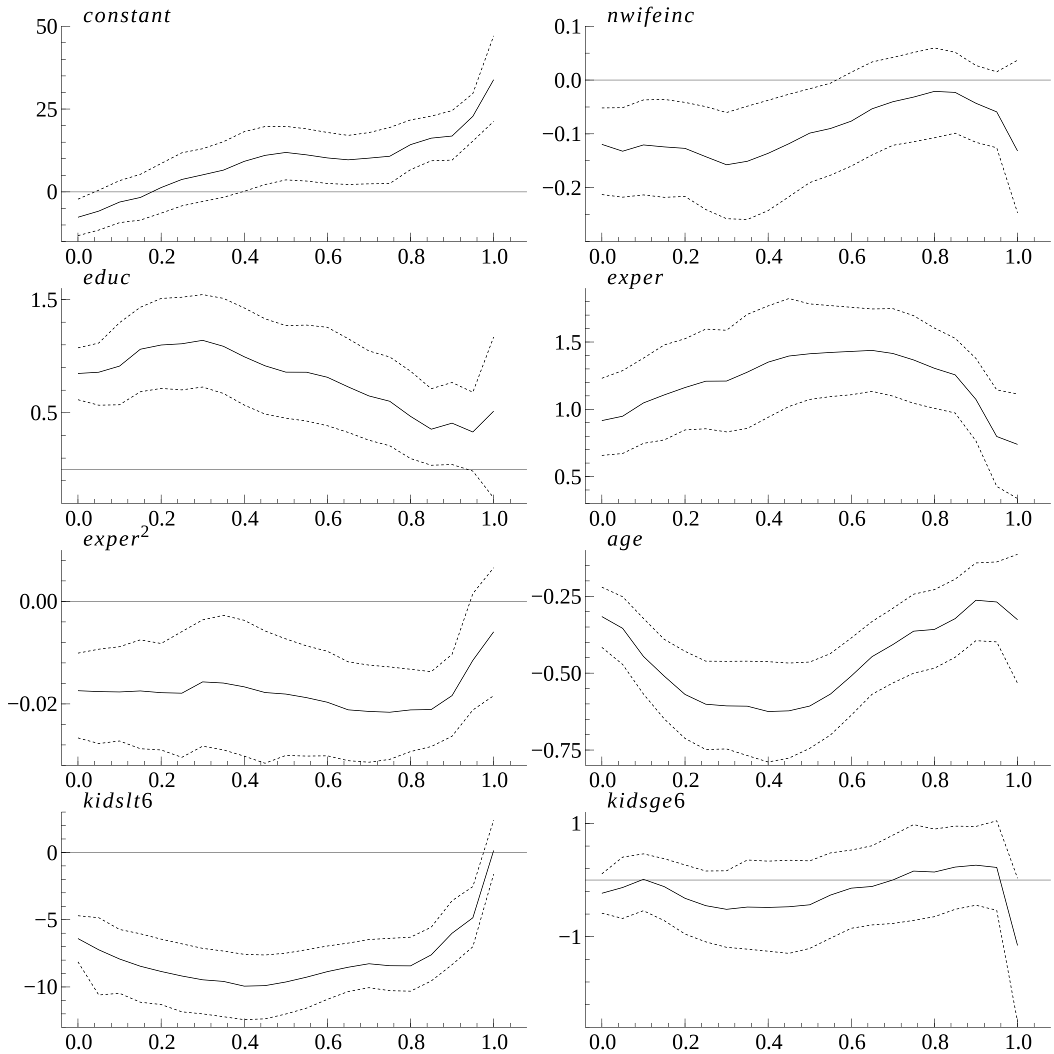

Figure 2: Labor supply data: Posterior means (solid line) and 95% credible intervals (dotted lines) are plotted for the normal prior.