Bayesian analysis of a Tobit quantile regression model

Keming Yu[a] and Julian Stander[b*]
[a]Department of Mathematical Sciences, Brunel University, UK
[b]School of Mathematics and Statistics, University of Plymouth, UK

Abstract

This paper develops a Bayesian framework for Tobit quantile regression. Our approach is organized around a likelihood function that is based on the asymmetric Laplace distribution, a choice that turns out to be natural in this context. We discuss families of prior distribution on the quantile regression vector that lead to proper posterior distributions with finite moments. We show how the posterior distribution can be sampled and summarized by Markov chain Monte Carlo methods. A method for comparing alternative quantile regression models is also developed and illustrated. The techniques are illustrated with both simulated and real data. In particular, in an empirical comparison, our approach out-performed two other common classical estimators.

Keywords: Asymmetric Laplace distribution; Bayes factor; Bayesian inference; Bayesian model comparison; Markov chain Monte Carlo methods; Quantile regression; Tobit model.

JEL classification: C14, C24.

1 Introduction

This paper is concerned with the following problem. Suppose that y^* and y are random variables connected by the censoring relationship

$$y = \max\left\{y^0, \, y^*\right\},$$

where y^0 is a known censoring point, and that we are given a sample of independent observations $\mathbf{y} = (y_1, ..., y_n)$ and associated covariates $\mathbf{x} = (x_1, ..., x_n)$, where $y_i = \max\{y_i^0, y_i^*\}$ and x_i is a k vector. The objective is to model and estimate the θth conditional quantile function of y given the sample $(\mathbf{y}, \mathbf{x})$, for $0 < \theta < 1$. Following Powell (1986) and Buchinsky and Hahn (1998), we assume that for the θth quantile, the partially latent y_i^* is generated according to the model

$$y_i^* = x_i' \beta_\theta + \epsilon_{\theta i}$$

* *corresponding author:* Dr Keming Yu, Department of Mathematical Sciences, Brunel University, Uxbridge, UB8 3PH, UK, tel: 44-1895-266128, fax: 44-1895-269732, email: keming.yu@brunel.ac.uk

where the θth conditional quantile of $\epsilon_{\theta i}$, denoted $\text{quant}_\theta(\epsilon_{\theta i}|x_i)$, is zero. The study of this important and interesting "Tobit quantile regression" problem has been taken up by Powell (1986), Hahn (1995), Buchinsky and Hahn (1998) and Bilias *et al.* (2000), amongst others, and has led to a body of frequentist parametric and semiparametric methods for estimating the conditional quantile function. The purpose of this paper is to describe the first Bayesian approach for estimating a Tobit quantile regression model, extending and complementing the method developed by Chib (1992) for the standard Tobit model.

Our approach to this problem relies on the use of the asymmetric Laplace distribution as the distribution of the error $\epsilon_{\theta i}$. As we show in Section 2, this choice is quite natural in the context of the quantile regression problem. For given families of prior distributions on the quantile regression parameter β_θ, we provide conditions under which the posterior distribution is proper. We show how appropriate Markov chain Monte Carlo (MCMC) methods can be used to simulate and summarize the posterior distribution (Tierney, 1994). Chib and Greenberg (1995) provide an excellent tutorial on Metropolis-Hastings algorithm. The approach of Chib (1995) for Gibbs output, as extended by Chib and Jeliazkov (2001) for Metropolis-Hastings chains, is also used to estimate the marginal likelihood of our model. This leads to a Bayesian framework for comparing alternative Tobit quantile regression models.

The rest of the paper is organized as follows. Assuming a Bayesian structure, in Section 2 we present details of the likelihood and prior distributions that we consider. We show that this choice leads to proper posteriors with finite moments. In Section 3 we outline the MCMC scheme that we adopt to perform the necessary Bayesian computations. We then present a series of simulation studies that illustrate the implementation of the proposed approach. The marginal likelihood which allows Bayesian Tobit model selection is derived in Section 4. The methods are applied to a real data set in Section 5. Section 6 provides additional discussion.

2 Inferential framework

Powell's (1986) estimator for the population parameter β_θ, as well as the alternative estimators proposed by Buchinsky and Hahn (1998) and Bilias *et al.* (2000), are based on the check (or loss) function $\rho_\theta(\lambda) = \{\theta - I(\lambda < 0)\}\lambda$, where I is the usual indicator function. An intuitive estimator for the Tobit quantile regression model is given by

$$\hat{\beta}_\theta = \arg\min_\beta \sum_{i=1}^{n} \rho_\theta(y_i - \max\{y_i^0,\, x_i'\beta\}). \tag{1}$$

Buchinsky and Hahn (1998) pointed out that the objective function (1) is not convex in β with the result that obtaining a global minimizer can be difficult. Instead, Buchinsky and Hahn (1998) noted that $x_i'\beta_\theta$ is the θ^*th conditional quantile of y_i if $\max\{y_i^0,\, x_i'\beta_\theta\}$ is the θth conditional quantile of y_i^* given x_i, where $\theta^* \equiv \{h_0(x_i) - (1-\theta)\}/h_0(x_i)$ in which $h_0(x) \equiv Pr[y_i^* > y_i^0|x_i = x]$, so that the parameter β_θ can be estimated via $\hat{\beta}_{\theta^*} = \arg\min_\beta \sum_{i=1}^n \rho_{\theta^*}(y_i - x_i'\beta)$. However, since $h_0(x)$ and therefore θ^* are unknown, a first stage estimation for $h_0(.)$ and therefore θ^* has to be performed before estimating β_θ. Further, the asymptotic covariance matrices of both the Powell and the Buchinsky and Hahn estimators, which depend on the error densities of ϵ_θ, are therefore difficult to estimate reliably. In fact, bootstrap procedure was first proposed to improve the reliability of Powell estimation by Hahn (1995) and Buchinsky (1995) who provided theoretical justification and simulation evidence respectively. Then Bilias *et al.* (2000) proposed a modified bootstrap procedure to increase the reliability of Buchinsky and Hahn estimators. Our proposed Bayesian inference no longer has this lack of convexity problem and can provide credible limits using a posterior sample straightway.

To introduce the Bayesian framework for the Tobit censoring model, that is, to derive a parametric distribution based likelihood for any posterior inference, let us recall the asymmetric Laplace distributions discussed in Yu and Zhang (2005). A random variable U is said to follow the simplest form of an asymmetric Laplace distribution if its probability density is given by

$$f_\theta(u) = \theta(1-\theta)\, \exp\{-\rho_\theta(u)\}. \tag{2}$$

When $\theta = 1/2$, (2) reduces to $\exp(-|u|/2)/4$, which is the density function of a standard symmetric Laplace distribution. For all other values of θ, the density in (2) is asymmetric. The θth quantile of this distribution is zero. That is, for any random variable Y, if $q_\theta(Y)$ is the θth quantile of Y, then $q_\theta(Y)$ is the θth quantile of $q_\theta(Y) + U$. Hence introducing a location parameter μ into the density (2) so that it becomes $f_\theta(u) = \theta(1-\theta)\, \exp\{-\rho_\theta(u-\mu)\}$, means that estimating the θth quantile of Y reduces to estimating the location parameter of U. Empirical studies based on different distributions such as normal, Gamma and Cauchy for Y, and simulating (2) using $\frac{\xi}{\theta} - \frac{\eta}{1-\theta}$, where ξ and η are independent exponential distribution with mean 1, have confirm this link.

For the Tobit quantile regression model we observe y_i instead of y_i^*, where $y_i = \max\{y_i^0,\, y_i^*\}$. If $x_i'\beta_\theta$ is the θth quantile of y_i^*, then $q(y_i) \equiv \max\{y_i^0,\, x_i'\beta_\theta\}$ is the θth quantile of y_i, so the θth quantile of $q_\theta(y_i) + U$. Hence a parametric distribution based likelihood function in terms of parameter β and known censoring points y_i^0 is given by

$$L(\mathbf{y}|\beta) = \theta^n(1-\theta)^n \exp\left\{-\sum_{i=1}^{n} \rho_\theta(y_i - \max\{y_i^0, x_i'\beta\})\right\}. \tag{3}$$

The posterior distribution $\pi(\beta|\mathbf{y})$ of $\beta = \beta_\theta$ given $\mathbf{y}$ can be obtained using Bayes theorem as

$$\pi(\beta|\mathbf{y}) \propto L(\mathbf{y}|\beta)\, \pi(\beta), \tag{4}$$

where $\pi(\beta)$ is the prior distribution of β_θ.

This approach does not require us to use a Dirichlet process for the prior, as Kottas and Gelfand (2001) did. Actually, this posterior distribution only involves the Tobit quantile regression model parameter β_θ which is generally assumed to be a $p+1$-dimensional vector.

2.1 Some key results

Although a standard conjugate prior distribution is not available for the Tobit quantile regression formulation, MCMC methods may be used to draw samples from the posterior distributions. This, principal, allows us to use virtually any prior distribution. However, we should select priors that yields proper posteriors.

In this section we show that we can choose the prior $\pi(\beta)$ from a class of known distributions, in order to get proper posteriors.

The likelihood $L(\boldsymbol{y}|\beta)$ in (3) is not continuous on the whole real line, but has a finite or a countably infinite set of discontinuities, thus is Riemann integrable.

First, the posterior is proper if and only if

$$0 < \int_{R^{p+1}} \pi(\beta|\mathbf{y}) d\beta < \infty, \tag{5}$$

or, equivalently, if and only if,

$$0 < \int_{R^{p+1}} L(\mathbf{y}|\beta)\, \pi(\beta)\, d\beta < \infty.$$

Moreover, we require that all posterior moment exist. That is,

$$E\Big[(\prod_{j=0}^{p} |\beta_j|^{r_j}) |\mathbf{y} \Big] < \infty, \tag{6}$$

where $(r_0, \ldots, r_p)$ denotes the order of the moments of $\beta = (\beta_0, \ldots, \beta_p)$.

We now establish a bound for the integral $\int_{R^{p+1}} \prod_{j=0}^{p} |\beta_j|^{r_j} L(\mathbf{y}|\beta)\, \pi(\beta)\, d\beta$ that allows us to obtain proper posterior moments.

Lemma 1: Let the function $g(t) = \exp(-|t|)$, and $f(t) = \theta(1-\theta) \exp(-t[\theta - I(t < 0)])$, then $f(t)$ has upper bound $g(h_1(\theta)t)$ and lower bound $g(h_2(\theta)t)$.

Proof: Write $f(t)$ as a mixture form of g:

$$\begin{aligned} f(t) &= \theta(1-\theta) \left\{ \exp(-\theta t)\, I(t \geq 0) + \exp((1-\theta)|t|)\, I(t < 0) \right\} \\ &= \theta(1-\theta) \left\{ g(\theta t)\, I(t \geq 0) + g((1-\theta)t)\, I(t < 0) \right\}. \end{aligned}$$

Note that $g(t) = g(|t|)$ is a decreasing function of t or $|t|$, and so for $\theta \leq \frac{1}{2}$, $g((1-\theta)t) = g((1-\theta)t)\, I(t \geq 0) + g((1-\theta)t)\, I(t < 0) < g(\theta t)\, I(t \geq 0) + g((1-\theta)t)\, I(t < 0) \leq g(\theta t)\, I(t \geq 0) + g(\theta t)\, I(t < 0) = g(\theta t)$. Hence $f(t)$ has upper bound $g(h_1(\theta)t)$ and lower bound $g(h_2(\theta)t)$. A similar argument establishes the same result for $\theta > \frac{1}{2}$.

Lemma 2: For any constant $a > 0$ and $m > p$,

$$\int \prod_{k=0}^{p} |\beta_k|^{r_k} \prod_{i=1}^{m} \exp(-a|y_i - x_i'\beta|)\, d\beta < \infty.$$

Proof: Without loss of generality, we can take $p = 1$, $x_i'\beta = \beta_0 + \beta_1 x_{1i}$,

$$\begin{aligned} \textstyle\int_{R^2} \quad & |\beta_0|^{r_0}|\beta_1|^{r_1} \exp(-a \sum_{i=1}^{m} |y_i - x_i'\beta|) d\beta_0 d\beta_1 \\ \leq \quad & \int_{R^2} |\beta_0|^{r_0} \exp(-a|\beta_0 + x_{11}\beta_1 - y_1|)\, |\beta_1|^{r_1} \exp(-a|\beta_0 + x_{12}\beta_1 - y_2|) d\beta_0 d\beta_1. \end{aligned} \quad (7)$$

Note that since the double-integration $\int_{R^2} |U|^{r_0} \exp(-|U+V-c_1|)|V|^{r_1} \exp(-|U+V-c_2|)\, dUdV$ is finite for any constants c_1, c_2, $r_0 \geq 0$ and $r_1 \geq 0$, the integral (7) is finite. Hence Lemma 2 is proved.

Theorem 1 below establishes that in the absence of any realistic prior information we could legitimately use an improper uniform prior distribution for all the components of β. This choice may be appealing as the resulting posterior distribution is proportional to the likelihood surface.

Theorem 1: Assume that the prior for β is improper and uniform, that is, $\pi(\beta) \propto 1$, then all posterior moments of β exist., i.e., equation (6) holds.

Proof: We need to prove $\int_{R^{p+1}} \prod_{j=0}^{p} |\beta_j|^{r_j} \exp\left\{-\sum_{i=1}^{n} \rho_\theta(y_i - \max\{y_i^0, x_i'\beta\})\right\} d\beta$ is finite. Note that $\sum_{i=1}^{n} \rho_\theta(y_i - \max\{y_i^0, x_i'\beta\}) = \sum_{i\in\mathcal{C}} \rho_\theta(y_i - x_i'\beta) + \sum_{i\notin\mathcal{C}} \rho_\theta(y_i - y_i^0)$, where the set $\mathcal{C} = \{i : x_i'\beta > y_i^0\}$. Hence, it suffices to prove that $\int \prod_{j=0}^{p} |\beta_j|^{r_j} \exp\left\{-\sum_{i\in\mathcal{C}} \rho_\theta(y_i - x_i'\beta)\right\} d\beta$ is finite. According to Lemma 1, this is true if and only if $\int_{R^{p+1}} \prod_{j=0}^{p} |\beta_j|^{r_j}\, g(h(\theta) \sum_{i\in\mathcal{C}}(y_i - x_i'\beta)) d\beta$ is finite, and this is true according to Lemma 2. where we omit the subscript on h for notational simplicity.

Noting that any terms like $\exp(-b|\beta - c|)$ adding to the inside of the integral product of lemma 2, Lemma 2 still holds for any constant $b > 0$ and c. Theorems 2 and 3 below provide that the existence of posterior and posterior moments based on double-exponential and normal priors.

Theorem 2: When the elements of β are assumed prior independent, and each $\pi(\beta_i) \propto \exp(-\frac{|\beta_i - \mu_i|}{\lambda_i})$, a double-exponential with fixed μ_i and $\lambda_i > 0$, all posterior moments of β exist.

Theorem 3: Assume that the prior for β is multivariate normal $N(\mu, \Sigma)$ with fixed μ and Σ, then all posterior moments of β exist.

In particular, when the elements of β are assumed a prior independent and univariate normal, all posterior moments of β exist.

3 Simulation studies

We have found it simplifies the algorithm to assume zero to be the fixed censoring points by a simple transformation of any non-zero censoring points. To see this, let $\tilde{Y}_i = Y_i - y_i^0$, $\tilde{X}_i = (X_i, y_i^0)$, and $\tilde{\beta} = (\beta', -1)'$ Then the Tobit θth quantile regression of $\tilde{Y}_i$ given $\tilde{X}_i$ reduces to $\max\{X_i'\beta,\, y_i^0\} - y_i^0$=$\max\{\tilde{X}_i{}'\tilde{\beta}, 0\}$.

We proceed by adopting an MCMC scheme that constructs a Markov chain with equilibrium distribution the posterior $\pi(\beta|\boldsymbol{y})$; in fact, our approach is similar to the MCMC algorithm for frequentist problems described by Chernozhukov and Hong (2003). After running the Markov chain for a certain *burn-in* period so that it can reach equilibrium, one obtains samples from $\pi(\beta|\boldsymbol{y})$.

We describe our approach to Bayesian Tobit quantile regression through a simulation study that was considered by Bilias *et al.* (2000). They took $p = 2$, so that $\beta_\theta = (\beta_0, \beta_1, \beta_2)$, and generate data according to

$$y = \max\{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon, 0\},$$

where x_1 is a Bernoulli random variable taking -1 and 1 each with probability 1/2 and x_2 is a standard normal random variable. They set $(\beta_0, \beta_1, \beta_2) = (1, 1, 1)$ and used three different distributions for ϵ. These are the standard normal distribution $N(0, 1)$, a heteroscedastic normal $(1+x_2)N(0, 1)$ and a normal mixture $0.75N(0, 1)+0.25N(0, 4)$. The censoring level is approximately 30% for all designs.

We take the Tobit θ^{th} quantile regression of y given $x = (x_1, x_2)$ to be given by $\max\{\beta_0 + \beta_1 x_1 + \beta_2 x_2,\ 0\}$.

We first chose independent improper uniform priors for the components of β_θ in all the experiments in accordance with Theorem 2. We simulated realizations from the posterior distribution using the MH algorithm. In particular, the parameters were updated using a random-walk Metropolis algorithm with a Gaussian proposal density centred at the current parameter value. In all the examples, time series plots indicated that the Markov chain converged very rapidly, usually within the first few iterations. However, in every case we adopted a burn-in of 3000 iterations and then collected a sample of 2000 values of each of the elements of $\beta_\theta = (\beta_0, \beta_1, \beta_2)$.

For θ=0.05, 0.5 and 0.95, we estimated θ^{th} quantiles using posterior means. Figures 1, 2 and 3 show the posterior histograms of β_θ for the above three error models together with the true parameter values. The true parameter values of β_θ were obtained according to the true θth quantile regression of y in the model which is denoted as $q_\theta(y) = \max\{\beta_0+\beta_1 x_1+\beta_2 x_2+q_\theta(\epsilon),\quad 0\}$. For example, when ϵ follows the heteroscedastic normal $(1+x_2)N(0, 1)$, $q_\theta(y) = \max\{(\beta_0+\Phi^{-1}(\theta))+\beta_1 x_1+(\beta_2+\Phi^{-1}(\theta))x_2,\ 0\}$, so the true parameter values for this model are $\beta_0 = 1+\Phi^{-1}(\theta)$, $\beta_1 = 1$, and $\beta_2 = 1+\Phi^{-1}(\theta)$, in which $\Phi^{-1}(\theta)$ is the θ^{th} quantile of $N(0, 1)$. Table 1 compares the posterior means with the true values of β_θ and also gives standard deviations and 95% credible intervals under each of three error models. As expected, the proposed Bayesian inference works well. For example, all the absolute biases for estimating β_θ turn out to be in the range [0.02, 0.17]. The empirical coverage probabilities based on 1000 0.95-level confidence intervals for the regression coefficient β lies in the range of (0.936,   0.959). The posterior histograms of parameters β_θ in Figures 1, 2 and 3 also support this conclusions.

We repeated the above analysis using a multivariate normal prior in accordance with Theorem 4. For simplicity, we used independent normal priors here. In general, the mean vector and the covariance matrix of the multivariate normal prior distribution can been estimated using a simple "plug-in" or "initial estimation" rule. That is, we use the uncensored observations in the data set to perform an ordinary quantile regression, and then we "plugged-in" the parameter and covariance matrix estimates from this fitting to the prior distribution. Just as with the independent uniform prior, our experience was that this method is very successful in terms of speed of convergence of the MCMC algorithm. Table 2 presents results analogous to those in Table 1 for this normal prior. Except for β_1 for the heterosedastic model with $\theta = 0.95$ and 0.05, all absolute biases for estimating β_θ are in the range [0.01, 0.27].

We also checked our methodology with the double-exponential prior discussed in Theorem 3. We found only small differences in the inferences obtained compared with those when the independent normal prior was used.

3.1 Empirical comparison

Buchinsky and Hahn (1998) performed Monte Carlo experiments to compare their estimator with the one proposed by Powell (1986). One of models they used was given by

$$y = \max\{-0.75,\, y^*\}$$

and

$$y^* = 1 + x_1 + 0.5x_2 + \epsilon,$$

where the regressors in x_i were each drawn from a standard normal distribution and the error term has multiplicative heteroskedasticity obtained by taking $\epsilon = \xi\, v(x)$ with $\xi \sim N(0, 25)$ and $v(x) = 1+0.5(x_1+x_1^2+x_2+x_2^2)$. For estimating the median regression for this model, Table 3 summaries the biases, root mean square errors (RMSE) and 95% credible intervals for β_0 and β_1 obtained from the following three approaches: BH method (Buchinsky and Hahn, 1998), Powell's estimator (Power, 1986) and the proposed Bayesian method with uniform prior. Power's estimator is carried via an iterative linear programming algorithm suggested by Buchinsky (1994). BH's estimator involes in kernel estimation of $h_0(x) \equiv= Pr[y_i^* > y_i^0 | x_i = x] = EI(y_i^* > y_i^0)|x_i = x]$, which is a specific case of standard regression mean smooth. The values relating to BH and Powell were also reported in Table 1 of Buchinsky and Hahn (1998). In particular, the BH method used log-likelihood cross-validated bandwidth selection for kernel estimation of θ^*, and the 95% confidence intervals for both BH and Powell estimators are based on their asymptotic normality theory. The results from Bayesian inference are based on a burn-in of 1000 iterations and then 3000 sample values.

Clearly, the proposed Bayesian method outperformed the BH and Powell methods. It yields considerably lower biases, lower mean square errors and much more precise credible intervals. S-PLUS code to implement the method with this comparison is available from the authors.

3.2 Inference with scale parameter

One may be interested in introducing a scale parameter into the likelihood function $L(\mathbf{y}|\beta)$ of Section 2 for the proposed Bayesian inference. Suppose $\sigma > 0$ is the scale parameter, it is natural to have the alternative likelihood function as

$$L(\mathbf{y}|\beta, \sigma) = \frac{\theta^n(1-\theta)^n}{\sigma^n} \exp\Big(-\sum_{i=1}^{n} \rho_\theta\big(\frac{y_i - \max\{y_i^0,\, x_i'\beta\}}{\sigma}\big)\Big).$$

The corresponding posterior distribution $\pi(\beta, \sigma|\mathbf{y})$ can be written as

$$\pi(\beta, \sigma|\mathbf{y}) \propto L(\mathbf{y}|\beta, \sigma)\, \pi(\beta, \sigma),$$

where $\pi(\beta, \sigma)$ is the prior distribution of (β_θ, σ) for a particular θ. As what interests us is the regression parameter β, and σ is what is referred to as a nuisance parameter, we may integrate out σ and investigate the marginal posterior $\pi(\beta|\boldsymbol{y})$ only. For example, we have considered a "reference" prior $\pi(\beta, \sigma) \propto \frac{1}{\sigma}$, which gives that

$$\pi(\beta|\mathbf{y}) \propto \left(\sum_{i=1}^{n} \rho_\theta(y_i - \max\{y_i^0,\, x_i'\beta\} \right)^{-n},$$

or

$$\log\, \pi(\beta|\mathbf{y}) \propto -n\, \log \sum_{i=1}^{n} \rho_\theta(y_i - \max\{y_i^0,\, x_i'\beta\}).$$

Implementing MCMC algorithm on this posterior form, we have found that the simulation results in this section are more or less same as those obtained using the posterior density (4).

4 Marginal likelihood and Bayes factors

We now discuss Tobit quantile regression model choice under the proposed Bayesian setting. The issue of model choice can be dealt with by calculating Bayes factors. For this, the marginal likelihood, which is the normalizing constant of the posterior density, is required. The calculation of the marginal likelihood has attracted considerable interest in the recent MCMC literature. In particular, Chib (1995) and Chib and Jeliazkov (2001) have developed a simple approach for estimating the marginal likelihood using the output from the Gibbs sample and the MH algorithm respectively. Under our proposed Bayesian Tobit inference, their approach can be simplified further.

We start our discussion by considering the problem of comparing a collection of models $\{M_1, ..., M_L\}$ that reflect competing hypotheses about the regression form. Under model M_k, suppose that the Tobit θth quantile model is given by

$$\boldsymbol{y}|M_k = \max\{y_{(k)}^0,\, x_{(k)}'\beta_{\theta(k)} + \epsilon_{\theta(k)}\},$$

then the marginal likelihood arising from estimating $\beta_{\theta(k)}$ is defined as

$$m(\mathbf{y}|M_k) = \int L(\mathbf{y}|M_k, \beta_{\theta(k)})\, \pi(\beta_{\theta(k)}|M_k) d\beta_{\theta(k)},$$

which is the normalizing constant of the posterior density.

Our estimation of $m(\mathbf{y}|M_k)$ is based on the work of Chib (1995). In particular, we use the relationship

$$\log m(\mathbf{y}|M_k) = \log L(\mathbf{y}|M_k, \beta^*_{\theta(k)}) + \log \pi(\beta^*_{\theta(k)}|M_k) - \log \pi(\beta^*_{\theta(k)}|\mathbf{y}, M_k),$$

from which the marginal likelihood can be estimated by finding an estimate of the posterior ordinate $\pi(\beta^*_{\theta(k)}|\mathbf{y}, M_k)$. We denote this estimate as $\hat{\pi}(\beta^*_{\theta(k)}|\mathbf{y}, M_k)$. For estimate efficiency, $\beta_{\theta(k)} = \beta^*_{\theta(k)}$ is generally taken to be a point of high density in the support of posterior density. On substituting the latter estimate in $\log m(\mathbf{y}|M_k)$, we get

$$\log \hat{m}(\mathbf{y}|M_k) = n\log(\theta(1-\theta)) - \sum_i \rho_\theta(y_i - \max\{y^0_{i(k)},\, x'_{i(k)}\beta^*_{\theta(k)}\}) \\ + \log\pi(\beta^*_{\theta(k)}|M_k) - \log\hat{\pi}(\beta^*_{\theta(k)}|\mathbf{y}, M_k), \quad (8)$$

in which the first term $n\log(\theta(1-\theta))$ is constant and the sum is over all data points.

Once the posterior ordinate is estimated, we can estimate the Bayes factor of any two models M_k and M_l by

$$\hat{B}_{kl} = \exp\{\log\hat{m}(\mathbf{y}|M_k) - \log\hat{m}(\mathbf{y}|M_l)\}.$$

Now we discuss a simulation-consistent estimate of $\hat{\pi}(\beta^*_{\theta(k)}|\mathbf{y}, M_k)$ for an improper prior and proper prior respectively.

Following equation (7) of Chib and Jeliazkov (2001) and even using an improper prior, we can obtain simulation-consistent estimate of the posterior ordinate as

$$\hat{\pi}(\beta^*|\mathbf{y}) = \frac{G^{-1}\sum_{g=1}^{G}\alpha(\beta^{(g)},\beta^*)\,q(\beta^{(g)},\beta^*)}{J^{-1}\sum_{j=1}^{J}\alpha(\beta^*,\beta^{(j)})},$$

in which

$$\alpha(\beta,\beta^*) = \min\{1,\ \frac{\pi(\beta^*)}{\pi(\beta)}L^*(\beta^*,\beta)\},$$

and

$$L^*(\beta^*,\beta) = \exp\{-\sum_i\Big(\rho_\theta(y_i - \max\{y^0_{i(k)},\, x'_{i(k)}\beta^*_{\theta(k)}\}) - \rho_\theta(y_i - \max\{y^0_{i(k)},\, x'_{i(k)}\beta_{\theta(k)}\})\Big)\}.$$

Where $\{\beta^{(j)}\}$ are samples drawn from $q(\beta^*,\beta)$ and $\{\beta^{(g)}\}$ are samples drawn from the posterior distribution.

However, the estimate can be simplified further if a proper prior is used. In this case, noting that $\frac{q(\beta^*,\beta)}{q(\beta,\beta^*)} = 1$ for our simple random walk proposal density and dropping the dependency on M_k for notational simplicity, we obtain the posterior ordinate as

$$\hat{\pi}(\beta^*|\mathbf{y}) = \frac{\int\alpha(\beta,\beta^*)\pi(\beta|\mathbf{y})d\beta}{\int\alpha^*(\beta^*,\beta)\pi(\beta)d\beta},$$

in which

$$\alpha^*(\beta^*,\beta) = \min\{\frac{1}{\pi(\beta)},\ \frac{1}{\pi(\beta^*)}L^*(\beta,\beta^*)\}.$$

From this it follows that

$$\hat{\pi}(\beta^*|\mathbf{y}) = \frac{E[\alpha(\beta,\beta^*)]}{E[\alpha^*(\beta^*,\beta)]},$$

where the expectation in the numerator is over the posterior distribution $\pi(\beta|\mathbf{y})$ and the expectation in denominator is over the prior $\pi(\beta)$. This implies that a simulation-consistent estimate of the posterior ordinate is given by

$$\hat{\pi}(\beta^*|\mathbf{y}) = \frac{G^{-1}\sum_{g=1}^{G}\alpha(\beta^{(g)},\beta^*)}{J^{-1}\sum_{j=1}^{J}\alpha^*(\beta^*,\beta^{(j)})},$$

where $\{\beta^{(j)}\}$ are samples drawn from a proper prior distribution and $\{\beta^{(g)}\}$ are samples drawn from the posterior distribution.

5 Illustrative example

To illustrate the applications of the proposed methods to model fitting and selection, we made a analysis of the data set on women's labor force participation of Mroz (1987). Part of interests in the analysis of this data is in investigating the relationship between women's working hours and years of education and experience. In the data, Hours is the number of hours the wife worked outside the household in a given year, Y_{Ed} is the years of education, and Y_{Exp} is the years of work experience. If the wife was not working for pay, her hours worked will be left censored at zero. Of the 753 observations, 325 are censored. Therefore the censoring ratio is 43%.

In the SAS/STAT User's Guide (1999), a small part with size 17 of the data is used to illustrate censored mean regression model with the LIFEREG Procedure in SAS/STAT. The procedure was used to fit a linear regression with normal model error assumption. However, the observations on Hours are left skewed and far from normal. We re-visited this analysis by considering the median regression and the 95% quantile regression and used all 753 observations. We still assume that the θth conditional quantile of women's working hours depends on the Y_{Ed} and Y_{Exp} through the following linear equation

$$q_\theta(\boldsymbol{x}) = \max\{0,\ \ \beta_0(\theta) + \beta_1(\theta)Y_{Ed} + \beta_2(\theta)Y_{Exp}\}. \tag{9}$$

We employ the Bayesian Tobit quantile regression with censored point $y_0 = 0$ and uniform prior for the estimation.

The output $\{\beta^{(g)}\}$ is obtained by running the sampler for 5000 cycles after a burn-in of 2000. Our results are summarized in Table 4. The positive values for the estimates of coefficients $\beta_1(\theta)$ and $\beta(\theta)$ of Y_{Ed} and Y_{Exp} clearly show that both education and experience are important for women's working for pay.

Based on equation (8), we got value -225.3684 as the estimate of the logarithm of the marginal likelihood for the model, by taking $\beta^* = \sum_g \beta^{(g)}/3000$ based on the last 3000 samples of the output $\{\beta^{(g)}\}$.

One may introduce a quadratic term for Y_{exp} in the model (9) and fit the underlying quantile regression by

$$q_\theta(\boldsymbol{x}) = \max\{0,\ \ \beta_0(\theta) + \beta_1(\theta)Y_{Ed} + \beta_2(\theta)Y_{Exp} + \beta_3(\theta)Y_{Exp}^2\}. \tag{10}$$

However, we have the estimate -283.1418 of the logarithm of the marginal likelihood for the second model. Hence we see that there is not gain in introducing extra term for modelling underlying linear quantile regression.

6 Discussion

We have described a complete Bayesian approach for Tobit quantile regression modelling. This approach avoids the need to solve a non-convex minimization problem and a density estimation problem. We employ MCMC methodology to produce a sample from the posterior distribution, which we use to estimate such quantities as the posterior mean, marginal probability density functions, posterior correlations, standard deviations and credible intervals.

We also remark that a Bayesian semiparametric Tobit quantile regression approach based on the setting of Kottas and Gelfand (2001) could also be developed for inference in this case. In the semi-parametric approach, one could use a Dirichlet process as a prior. Instead of the likelihood function $L(\boldsymbol{y}|\beta)$, one can use a flexible family of zero median error distributions $p(.;\alpha,\gamma) = \frac{1}{\gamma}f(\frac{.}{\gamma};\alpha)1_{(-\infty,0)}(.) + \gamma f(.\gamma;\alpha)1_{(0,\infty)}(.)$, which is a kernel mixture of the distribution $f(.;\alpha)$ that is unimodal and symmetric around 0. According to Kottas and Gelfand (2001), $\alpha > 0$ is an arbitrary one-dimensional scale parameter and $\gamma > 0$ is a skewed parameter. Any member of this family with $\gamma \neq 1$ is a skewed distribution, with the type and amount skewness depending on the value of γ.

If we choose a split asymmetric Laplace density (2) for the kernel mixture, $p(u;\alpha,\sigma) = f_\theta(\frac{u}{\alpha\sigma})I(u<0) + f_\theta(\frac{u}{\alpha/\sigma})I(u>0)$, a fully Bayesian hierarchical structure is given by

$$y_i|\beta,\alpha_i,\sigma \sim p(y_i - \max\{y_i^0, x_i'\beta\};\alpha_i,\sigma), \quad i = 1,...,n,$$

in which

$$\beta \sim N_{p+1}(\mu,\Sigma),$$
$$\alpha_i \sim DP(\nu G_0),$$

and

$$\sigma \sim Gamma(a,b),$$

where DP stands for Dirichlet process, and the basic distribution G_0 is taken to be an IGamma(s,t), with mean $t/(s-1)$ if $s > 1$.

Richardson (1999) mentioned that popular forms of priors tend to be those which have parameters that can be set straightforwardly and which lead to posteriors with a relatively immediate form. In this sense, the proposed approach is preferable to Bayesian semiparametric methods.

Our sampling scheme was performed on the whole parameter vector in one block. In applications when the dimension of is large, it may be difficult to construct a single block Metropolis-Hastings algorithm that converges rapidly to the target density. In such cases, it is helpful to break up the variate space into smaller blocks and to then construct a Markov chain with these smaller blocks. Suppose, for illustration, that the regression vector parameter β is split into two vector blocks $(\beta^{(1)}, \beta^{(2)})$. The general rule for block choice is that sets of parameters that are highly correlated should be treated as one block when applying the multiple-block Metropolis-Hastings algorithm. Then, for each block, let

$$q_1(\beta^{(1)}, {\beta^{(1)}}^*|\beta^{(2)}), \;\; q_2(\beta^{(2)}, {\beta^{(2)}}^*|\beta^{(1)}),$$

denote the corresponding proposal density. Here each proposal density is allowed to depend on the data and the current value of the remaining block. Also define (by analogy with the single-block case)

$$\alpha(\beta^{(1)}, {\beta^{(1)}}^*|\beta^{(2)}) = \min\left\{1, \;\; \frac{\pi({\beta^{(1)}}^*|\beta^{(2)})\, q_1({\beta^{(1)}}^*, \beta^{(1)}|\beta^{(2)})}{\pi(\beta^{(1)}|\beta^{(2)})\, q_1(\beta^{(1)}, {\beta^{(1)}}^*|\beta^{(2)})}\right\},$$

and

$$\alpha(\beta^{(2)}, {\beta^{(2)}}^*|\beta^{(1)}) = \min\left\{1, \;\; \frac{\pi({\beta^{(2)}}^*|\beta^{(1)})\, q_2({\beta^{(2)}}^*, \beta^{(2)}|\beta^{(1)})}{\pi(\beta^{(2)}|\beta^{(1)})\, q_2(\beta^{(2)}, {\beta^{(2)}}^*|\beta^{(1)})}\right\},$$

as the acceptance probability for block $\beta^{(k)}$ $(k = 1, 2)$ conditioned on the other block. The conditional densities $\pi(\beta^{(1)}|\beta^{(2)})$ and $\pi(\beta^{(2)}|\beta^{(1)})$ that appear in these functions are called the full conditional densities. By Bayes theorem each is proportional to the joint density. For example,

$$\pi(\beta^{(1)}|\beta^{(2)}) \propto \pi(\beta^{(1)}, \beta^{(2)})$$

and, therefore, the acceptance probabilities in $\alpha(\beta^{(1)}, {\beta^{(1)}}^*|\beta^{(2)})$ and $\alpha(\beta^{(2)}, {\beta^{(2)}}^*|\beta^{(1)})$ can be expressed equivalently in terms of the kernel of the joint posterior density because the normalizing constant of the full conditional density (the norming constant in the latter expression) cancels in forming the ratio. With these inputs, one sweep of the multiple-block Metropolis-Hastings algorithm is completed by updating each block, say sequentially in fixed order, using a Metropolis-Hastings step with the above acceptance probabilities, given the current value of the other block. For further details see Chib and Greenberg (1995).

Acknowledgement

Keming Yu's work was partially supported by the Engineering and Physical Sciences Research Council GR-S68293-01. We are extremely grateful to Professor Siddhartha Chib for his extensive comments on the original manuscript. His comments were particularly helpful in comparing other methods for the models.

References

Bilias, Y., Chen, S. and Ying, Z., (2000), Simple Resampling Methods for Censored Regression Quantiles. *Journal of Econometrics* **68**, 303–338.

Buchinsky, M., (1994), Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression. *Econometrica*, **62**, 405–458.

Buchinsky, M., (1995). Estimating the asymptotic covariance matrix for quantile regression models a Monte Carlo study. *Journal of Econometrics* **99**, 373–386.

Buchinsky, M. and Hahn, J., (1998). An Alternative Estimator for Censored Quantile Regression. *Econometrica* **66**, 653–671.

Chernozhukov, V. and Hong, H., (2003). A MCMC Approach to Classical Estimation. *Journal of Econometrics* **115**, 293–346.

Chib, S., (1992). Bayes Inference in the Tobit Censored Regression Model. *Journal of Econometrics* **51**, 79–99.

Chib, S., (1995). Marginal Likelihood from the Gibbs Output. *Journal of the American Statistical Association* **90,** 1313–1321.

Chib, S. and Greenberg, E., (1995). Understanding the Metropolis-Hastings Algorithm. *American Statistician* **49**, 327–335.

Chib, S. and Jeliazkov, I., (2001). Marginal Likelihood from Metropolis-Hastings Output. *Journal of the American Statistical Association* **96,** 270–281.

Hahn, J., (1995). Bootstrapping Quantile Regression Estimators. *Econometric Theory* **11,** 105–121.

Kottas, A. and Gelfand, A. E., (2001). Bayesian Semiparametric Median Regression Model. *Journal of the American Statistical Association* **91**, 689–698.

Powell, J., (1986). Censored Regression Quantiles. *Journal of Econometrics* **32**, 143–155.

Powell, J. L., (2001). *Semiparametric Estimation of Censored Selection Model* in Hsiao, C., K. Morimune, and J.L. Powell, (Eds.), Nonlinear statistical modeling, Cambridge University Press.

Richardson, S., (1999). Contribution to the Discussion of Walker *et al.*, Bayesian Nonparametric Inference for Random Distribution and Related Functions. *Journal of the Royal Statistics Society* B, **61**, 485–527.

SAS/STAT User's Guide, (1999). SAS Institute Inc., Cary, NC, USA, SAS Online Document, Version 8.

Tierney, L., (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics* **22,** 1701-1762.

Yu, K. and Moyeed, R.A., (2001). Bayesian Quantile Regression. *Statistics Probability Letter* **54**, 437–447.

Yu, K., Lu, Z. and Stander, J., (2003). Quantile Regression: Applications and Current Research

Area. *The Statistician* **52**, 331–350.

Yu, K. and Moyeed, R.A., (2001). Bayesian Quantile Regression. *Statistics Probability Letter* **54**, 437–447.

Yu, K. and Zhang, J., (2005). A Three-Parametric Asymmetric Laplace Distribution and Its Extension. *Communication in Statistics–Theory and Methods* **34**, 1867–1879.

Table 1. True parameter values (T.V.) and their posterior means, standard deviations (S.D.) and 95% credible intervals. A uniform prior was adopted.

θ		Normal			Heteroscedastic			Mixture		
		β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
	T.V.	-0.65	1	1	-0.65	1	-0.65	-1.10	1	1
	Mean	-0.61	0.99	1.08	-0.66	1.05	-0.58	-1.00	1.09	0.98
0.05	S.D.	0.49	0.87	0.32	0.60	0.77	0.26	0.54	0.82	0.28
	2.5%	-1.16	-0.43	0.30	-2.53	-0.66	-1.15	-2.26	-0.54	0.50
	97.5%	-0.05	2.32	1.72	0.24	2.45	0.01	-0.11	2.70	1.62
	T.V.	1	1	1	1	1	1	1	1	1
	Mean	1.00	0.98	1.02	0.95	1.04	0.97	1.16	0.94	1.03
0.5	S.D.	0.25	0.39	0.15	0.22	0.29	0.16	0.24	0.29	0.15
	2.5%	0.52	0.21	0.72	0.37	0.44	0.61	0.60	0.29	0.73
	97.5%	1.53	1.67	1.30	1.34	1.67	1.30	1.63	1.50	1.31
	T.V.	2.65	1	1	2.65	1	2.65	3.10	1	1
	Mean	2.78	1.17	1.17	2.65	0.84	2.62	3.15	1.12	1.07
0.95	S.D.	0.79	1.04	0.43	0.54	1.22	0.31	0.80	1.05	0.43
	2.5%	1.64	-1.23	0.46	1.85	-1.70	1.91	1.89	-1.20	0.34
	97.5%	4.17	2.62	2.30	3.98	3.20	3.54	4.89	3.15	2.15

Table 2. True parameter values (T.V.) and their posterior means,
standard deviations (S.D.) and 95% credible intervals. A normal prior was adapted.

θ		Normal			Heteroscedastic			Mixture		
		β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
	T.V.	-0.65	1	1	-0.65	1	-0.65	-1.10	1	1
0.05	Mean	-0.80	1.14	1.02	-0.830	0.74	-0.77	-1.05	1.19	0.81
	S.D.	0.42	0.58	0.27	0.62	0.93	0.32	0.53	0.664	0.32
	2.5%	-1.75	0.06	0.41	-2.20	-0.1.01	-1.44	-2.09	-0.11	0.17
	97.5%	-0.12	2.34	1.49	0.10	2.81	-0.06	-0.04	2.45	1.39
	T.V.	1	1	1	1	1	1	1	1	1
0.5	Mean	0.90	1.27	0.94	0.95	0.97	1.11	1.01	0.86	0.82
	S.D.	0.26	0.29	0.16	0.23	0.28	0.15	0.30	0.37	0.17
	2.5%	0.50	0.67	0.59	0.45	0.43	0.86	0.44	0.19	0.48
	97.5%	1.38	1.60	1.23	1.34	1.57	1.41	1.63	1.06	1.15
	T.V.	2.65	1	1	2.65	1	2.65	3.10	1	1
0.95	Mean	2.78	1.17	1.17	3.00	0.89	2.09	3.09	0.87	0.95
	S.D.	0.79	1.04	0.43	0.55	0.63	0.27	0.45	0.65	0.23
	2.5%	1.64	-1.23	0.46	2.70	-0.53	1.22	2.39	-0.54	0.44
	97.5%	4.17	2.62	2.30	4.79	2.11	2.58	4.06	1.94	1.35

Table 3. Bias, root mean square errors (RMSE) and 95% credible intervals for the parameters β_0 and β_1 of the median regression. Samples were generated from the model considered by Buchinsky and Hahn (1998). Three approaches were used: BH method (Buchinsky and Hahn, 1998), Powell estimator (Powell, 1986) and the proposed bayesian method with uniform prior.

		β_0			β_1		
Sample size		BH	Powell	Bayes	BH	Powell	Bayes
100	Bias	0.14	-0.08	-0.08	0.31	0.33	0.30
	RMSE	2.88	4.11	1.414	2.16	2.85	1.416
	2.5%	-4.49	-6.00	0.07	-3.13	-4.55	0.13
	97.5%	6.76	9.40	2.20	5.65	7.41	2.65
400	Bias	0.20	0.19	-0.18	-0.06	-0.45	-0.06
	RMSE	0.58	0.68	0.39	0.61	0.66	0.22
	2.5%	-0.85	-0.83	0.12	-0.82	-1.12	0.13
	97.5%	4.41	4.31	2.13	2.24	2.36	1.77
600	Bias	0.18	0.20	-0.17	-0.06	-0.47	-0.06
	RMSE	0.48	0.49	0.39	0.50	0.57	0.35
	2.5%	-1.33	-0.14	0.21	-0.37	-0.89	0.13
	97.5%	4.67	3.42	1.99	2.09	1.89	1.48

Table 4. β_θ and their posterior means, standard deviations (S.D.) and 95% confident intervals under a uniform prior for the Mroz data

	$\theta = 0.5$			$\theta = 0.95$		
	β_0	β_1	β_2	β_0	β_1	β_2
Mean	-731.43	39.78	67.04	1331.94	30.46	34.72
S.D.	8.57	0.59	0.34	9.09	0.62	0.41
2.5%	-746.13	38.60	66.58	1310.22	29.41	34.15
97.5%	-715.91	40.94	67.80	1347.68	31.59	35.79

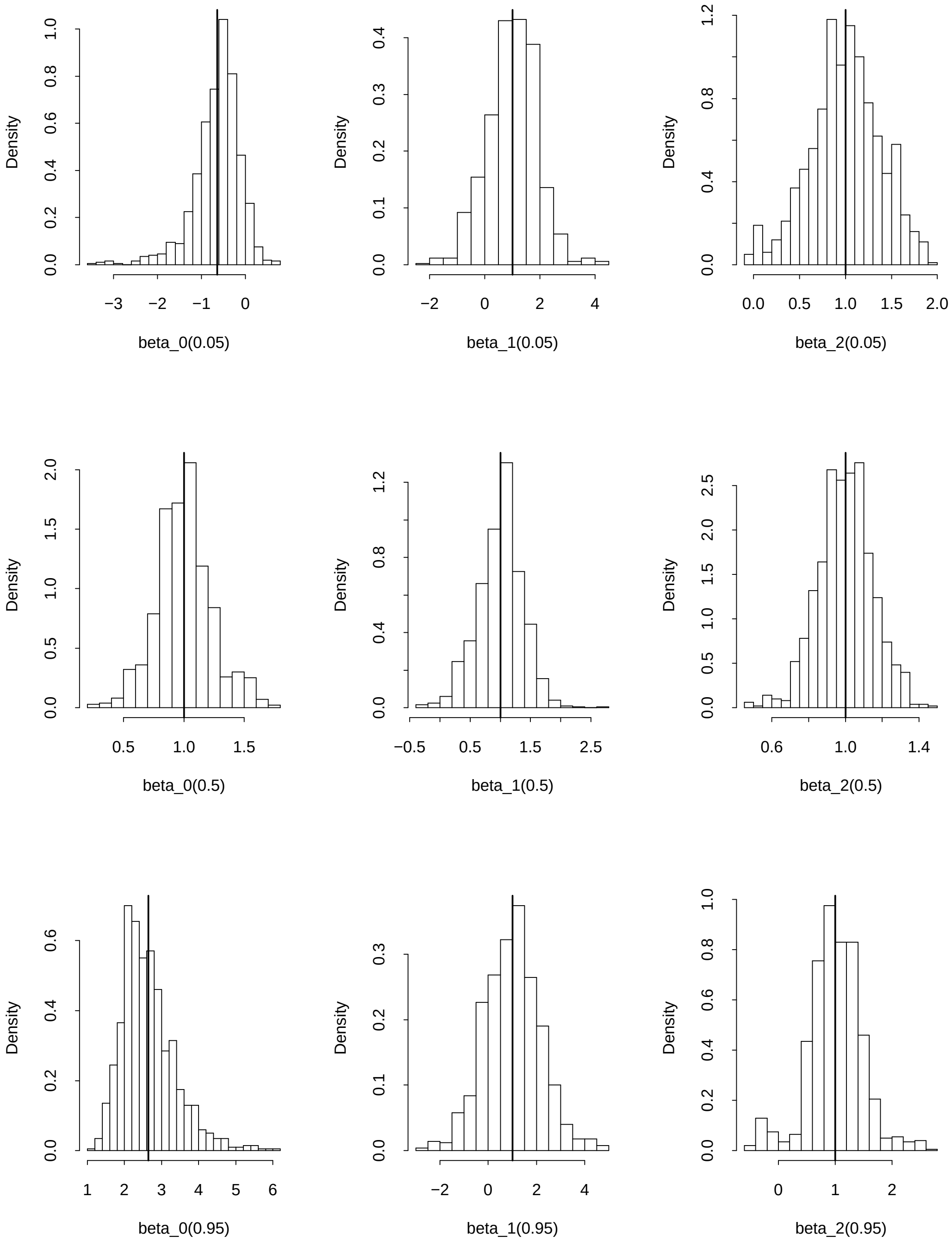

Fig. 1. Normal error model: Posterior histograms of the quantile regression parameters $\beta_\theta = (\beta_0, \beta_1, \beta_2)$ for $\theta = 0.05$, 5 and 0.95 together with the true parameter values indicated by the bold vertical line

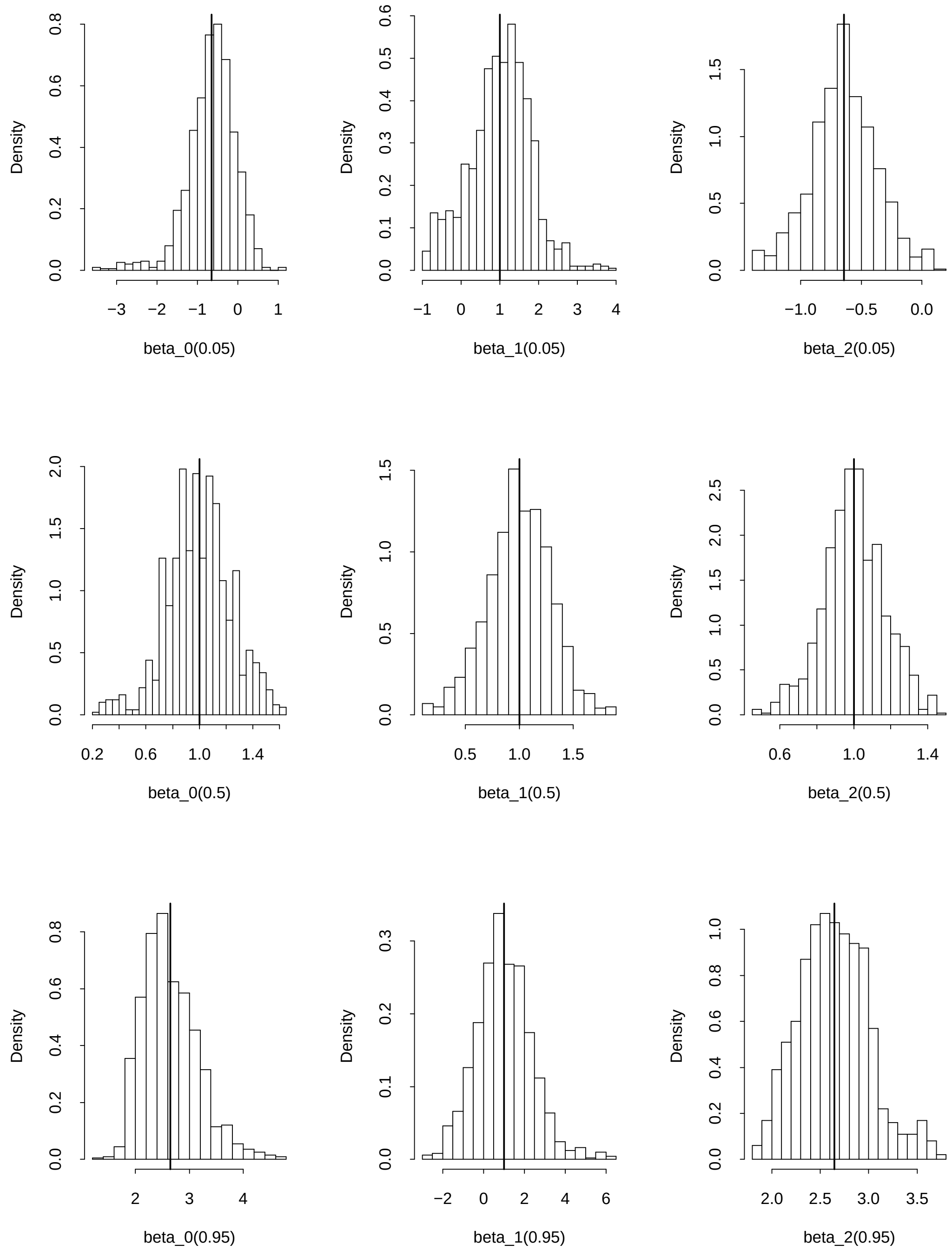

Fig. 2. Heteroscedastic normal error model: Posterior histograms of the quantile regression parameters $\beta_\theta = (\beta_0, \beta_1, \beta_2)$ for $\theta = 0.05$, 0.5 and 0.95 together with the true parameter values indicated by the bold vertical line

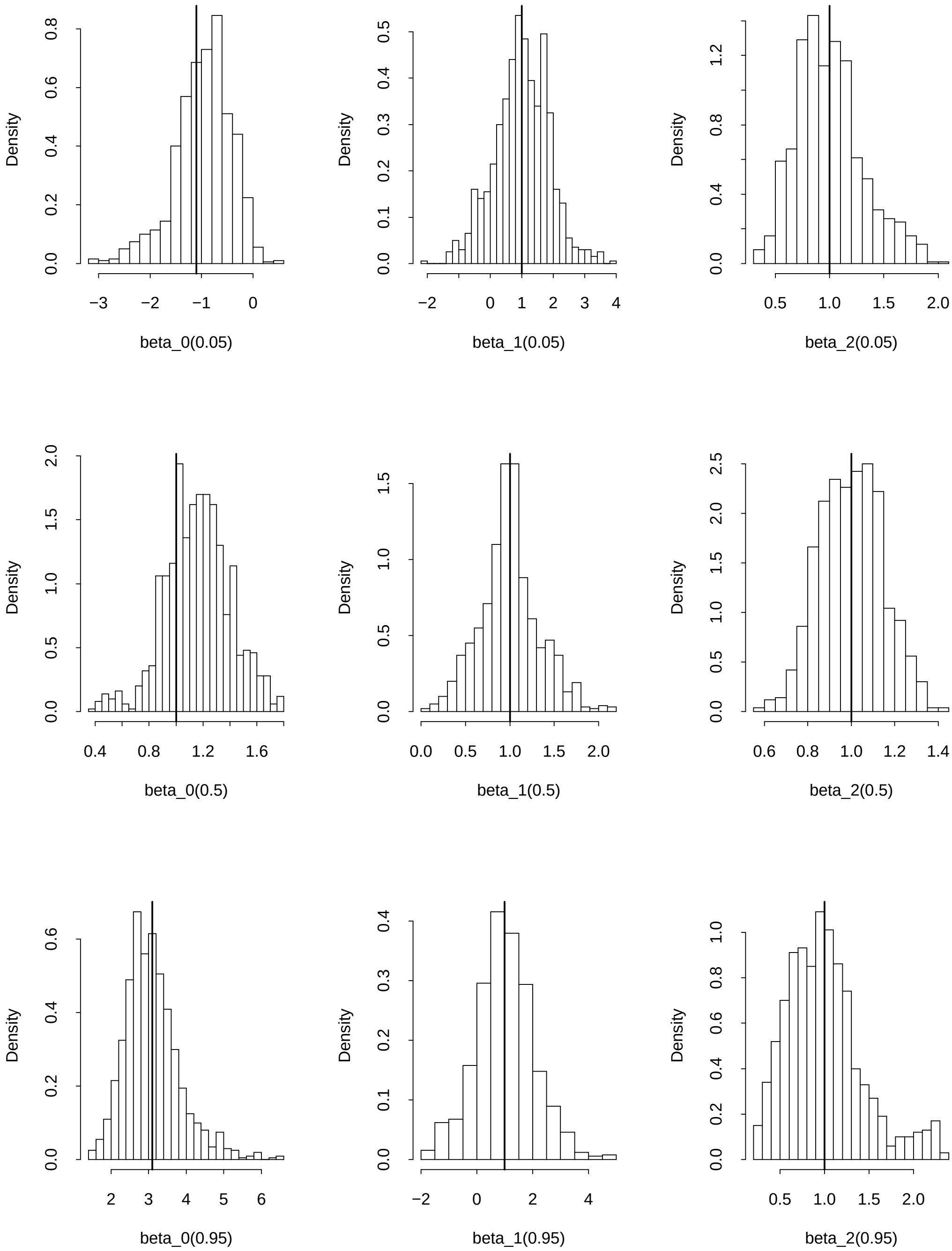

Fig. 3. Mixture normal error model: Posterior histograms of the quantile regression parameters $\beta_\theta = (\beta_0, \beta_1, \beta_2)$ for $\theta = 0.05$, 0.5 and 0.95 together with the true parameter values indicated by the bold vertical line