Journal of Statistical Computation and Simulation
Vol. 81, No. 11, November 2011, 1565–1578

Gibbs sampling methods for Bayesian quantile regression

Hideo Kozumi* and Genya Kobayashi

Graduate School of Business Administration, Kobe University, Rokko, Kobe 657-8501, Japan

(*Received 27 July 2009; final version received 23 May 2010*)

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.

Keywords: 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 [1], 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.* [2] and Koenker [3] for an overview.

Let y_i be a response variable and $\mathbf{x}_i$ a $k \times 1$ vector of covariates for the ith observation. Let $q_p(\mathbf{x}_i)$ denote the pth $(0 < p < 1)$ 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 modelled 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 pth quantile equal to zero, that is, $\int_{-\infty}^{0} f_p(\epsilon_i)\, \mathrm{d}\epsilon_i = p$.

*Correspondence author. Email: kozumi@kobe-u.ac.jp

ISSN 0094-9655 print/ISSN 1563-5163 online

http://dx.doi.org/10.1080/00949655.2010.496117
http://www.tandfonline.com

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

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

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

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

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 programming methods are commonly applied to obtain quantile regression estimates for $\boldsymbol{\beta}_p$ [4,5].

From a Bayesian point of view, Walker and Mallick [6], Kottas and Gelfand [7], and Hanson and Johnson [8] considered median regression, which is a special case of quantile regression with $p = 0.5$, and discussed non-parametric modelling for the error distribution based on either Pólya tree or Dirichlet process priors. Regarding general quantile regression, Yu and Moyeed [9] proposed a Bayesian modelling approach by noting that minimizing Equation (1) is equivalent to maximizing a likelihood function under the asymmetric Laplace error distribution [10]. Recently, Kottas and Krnjajić [11] developed Bayesian semi-parametric models for quantile regression using Dirichlet process mixtures for the error distribution.

As discussed in Yu and Moyeed [9], 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 [9] considered Markov chain Monte Carlo (MCMC) methods for posterior inference. Specifically, they used a random walk Metropolis algorithm with a Gaussian proposal density centred 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 a Gibbs sampling algorithm 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 analyse 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. In Section 4, we conduct numerical studies to assess the performance of the proposed methods. Finally, brief conclusions are given in Section 5.

2. Posterior inference

2.1. *Mixture representation*

Following Yu and Moyeed [9], 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

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

where $\rho_p(\cdot)$ is defined in Equation (2). The parameter p determines the skewness of distribution, and the pth 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 [12].

As shown in Kotz *et al.* [13], 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 a Gibbs sampling algorithm for the quantile regression model, we utilize a mixture representation based on exponential and normal distributions, which is found in Kotz *et al.* [13] and summarized as follows.

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

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

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

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

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 [9], 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 Equation (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

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

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

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

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 Equation (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

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

and $K_\nu(\cdot)$ is a modified Bessel function of the third kind [14]. There exist efficient algorithms to simulate from a generalized inverse Gaussian distribution [15,16], so that our Gibbs sampler defined in Equations (7) and (9) can be easily applied to quantile regression estimation.

We note that Tsionas [10] 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 [10] 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 [17]. In addition, Tsionas [10] mentioned that his Gibbs sampler is complicated and can be 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 in introducing a scale parameter $\sigma > 0$ into the model, our approach can incorporate it by rewriting Equation (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 for developing a Gibbs sampling algorithm 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 $\tilde{\mathbf{B}}_p^{-1} = \sum_{i=1}^n (\mathbf{x}_i\mathbf{x}_i'/\tau^2\sigma v_i) + \mathbf{B}_{p0}^{-1}$ and $\tilde{\boldsymbol{\beta}}_p = \hat{\mathbf{B}}_p\{\sum_{i=1}^n (\mathbf{x}_i(y_i - \theta v_i)/\tau^2\sigma v_i) + \mathbf{B}_{p0}^{-1}\boldsymbol{\beta}_{p0}\}$. Similar 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

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

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

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

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 [18] showed that all posterior moments of $\boldsymbol{\beta}_p$ exist for Tobit quantile regression models. Recently, Park and Casella [19] considered this prior in the context of Lasso estimation [20] and discussed the choice of hyperparameters.

As shown in Park and Casella [19], the double exponential density can be expressed as

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

where ω_j has an exponential distribution with mean $2/\lambda_0^2$. 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

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

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 [21–24]. Yu and Stander [18] 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

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

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 Equation (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 [25] can be applied to the sampling of y_i^*, that is,

$$y_i^*|\mathbf{y}, \boldsymbol{\beta}_p, \mathbf{v}, \sigma \sim y_i I(y_i > 0) + \mathcal{TN}_{(-\infty,0]}(\mathbf{x}_i'\boldsymbol{\beta}_p + \theta v_i, \tau^2\sigma v_i) I(y_i = 0), \tag{16}$$

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 Equations (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 [18], which relies on Metropolis–Hastings algorithms and requires to control tuning parameters for each value of p.

4. Numerical examples

We present four examples to illustrate the Gibbs sampling methods developed in Sections 2 and 3. The first two examples estimate the quantile regression model to assess the performance of our approach. The other examples consider the Tobit quantile regression model using both simulated and real data sets. In each example, we implement the proposed Gibbs sampler with starting values of ones for all the parameters. Preliminary runs of the algorithm showed that it converges within several hundred iterations and the results are robust against different choices of starting values.

4.1. *Simulation 1*

In this simulation study, the response y_i is generated from the model

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i, \quad i = 1, \ldots, 100, \tag{17}$$

where each covariate x_{ij} is simulated from a standard normal distribution and all the β_j are set to one. We consider three different distributions for ϵ_i: (i) the standard normal, $\mathcal{N}(0, 1)$, (ii) the Student's t distribution with three degrees of freedom, t_3, and (iii) a heteroscedastic normal, $(1 + x_{i2})\mathcal{N}(0, 1)$. The normal prior given by Equation (6) is used for $\boldsymbol{\beta}_p$ and the hyperparameters are chosen as $\boldsymbol{\beta}_{p0} = \mathbf{0}$ and $\mathbf{B}_{p0} = 100\mathbf{I}$.

We considered the quantile regression model without a scale parameter and fitted it to the simulated data set for $p = 0.1, 0.5$ and 0.9. To assess the sampling efficiency of the proposed algorithm, we calculated Monte Carlo standard errors [26,27] and inefficiency factors for $\boldsymbol{\beta}_p$ by running the Gibbs sampler for 12,000 iterations with an initial burn-in of 2000 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 [28]. The results are summarized in Table 1 together with those obtained from the random walk sampler of Yu and Moyeed [9]. When

Table 1. Simulated data: Monte Carlo standard errors (MCSEs) and inefficiency factors (IFs) are shown.

		β_1		β_2		β_3	
Method	p	MCSE	IF	MCSE	IF	MCSE	IF
$\epsilon \sim \mathcal{N}(0, 1)$							
Gibbs	0.1	0.0019	4.068	0.0017	5.821	0.0016	3.112
	0.5	0.0027	5.331	0.0019	2.128	0.0024	3.437
	0.9	0.0025	8.340	0.0015	6.858	0.0030	7.582
Random walk	0.1	0.0022	11.119	0.0020	8.353	0.0021	8.781
	0.5	0.0031	8.104	0.0024	5.824	0.0033	13.542
	0.9	0.0028	8.663	0.0018	8.485	0.0033	10.539
$\epsilon \sim t_3$							
Gibbs	0.1	0.0045	5.393	0.0034	4.608	0.0047	5.280
	0.5	0.0022	1.674	0.0019	2.295	0.0021	4.086
	0.9	0.0035	7.839	0.0033	9.961	0.0038	5.370
Random walk	0.1	0.0049	8.740	0.0037	4.979	0.0057	7.904
	0.5	0.0030	9.891	0.0027	12.954	0.0031	7.606
	0.9	0.0038	0.061	0.0037	11.753	0.0047	10.359
$\epsilon \sim (1 + x_2)\mathcal{N}(0, 1)$							
Gibbs	0.1	0.0039	7.489	0.0025	3.584	0.0030	6.282
	0.5	0.0027	2.661	0.0026	2.124	0.0021	2.767
	0.9	0.0034	6.225	0.0021	4.547	0.0024	3.091
Random walk	0.1	0.0046	10.891	0.0031	7.556	0.0036	10.557
	0.5	0.0034	10.451	0.0037	13.511	0.0031	12.766
	0.9	0.0037	12.579	0.0025	10.925	0.0028	11.283

applying the random walk sampler, we adjusted proposal step sizes in order that the inefficiency factors achieve the lowest values. Table 1 shows that both the Monte Carlo standard errors and the inefficiency factors of the Gibbs sampler are smaller than those of the random walk sampler in all cases. These results indicate the superiority of our Gibbs sampler to the random walk sampler.

Next, we conducted a simulation experiment intended to compare the empirical performance of our approach with that of a classical method. In each replication, simulated data were generated from Equation (17) using the three error distributions. For Bayesian estimation, we applied the Gibbs sampling method to the quantile regression models with and without a scale parameter σ and calculated the posterior means from a sample of 10,000 draws after 2000 burn-in iterations. The prior distributions for $\boldsymbol{\beta}_p$ and σ were specified as $\boldsymbol{\beta}_p \sim \mathcal{N}(\mathbf{0}, 100\mathbf{I})$ and $\sigma \sim \mathcal{IG}(3/2, 0.1/2)$, respectively. As for a classical method, the minimization of Equation (1) was implemented by the `quantreg` package for the R language [29]. For each of these methods, we evaluated the bias and root mean-squared error (RMSE) for $\boldsymbol{\beta}_p$ based on 1000 replications. Table 2 reports the simulation results for $p = 0.1, 0.5$ and 0.9. We observe that all the methods are severely biased in the case of heteroscedastic errors and that the classical method yields lower biases than the Bayesian methods. Comparing the RMSE of the methods, we see that the Bayesian methods

Table 2. Simulated data: biases and RMSEs are shown.

		β_1		β_2		β_3	
Method	p	Bias	RMSE	Bias	RMSE	Bias	RMSE
$\epsilon \sim \mathcal{N}(0, 1)$							
Classical	0.1	0.012	0.172	0.000	0.172	−0.008	0.181
	0.5	0.003	0.122	0.000	0.127	0.002	0.142
	0.9	−0.017	0.168	0.001	0.176	−0.003	0.191
Gibbs without σ	0.1	−0.097	0.181	−0.010	0.138	0.012	0.148
	0.5	0.003	0.113	0.000	0.113	0.001	0.124
	0.9	0.093	0.172	0.011	0.142	−0.021	0.158
Gibbs with σ	0.1	−0.001	0.163	−0.001	0.156	−0.004	0.166
	0.5	0.004	0.117	0.001	0.117	0.001	0.129
	0.9	−0.003	0.156	0.002	0.159	−0.006	0.174
$\epsilon \sim t_3$							
Classical	0.1	−0.046	0.309	−0.009	0.282	0.005	0.341
	0.5	−0.009	0.139	−0.006	0.141	−0.003	0.158
	0.9	0.029	0.320	0.003	0.290	−0.008	0.338
Gibbs without σ	0.1	−0.192	0.339	−0.020	0.235	0.026	0.290
	0.5	−0.007	0.126	−0.005	0.127	−0.003	0.147
	0.9	0.177	0.337	0.014	0.236	−0.032	0.292
Gibbs with σ	0.1	−0.084	0.304	−0.012	0.254	0.013	0.313
	0.5	−0.007	0.128	−0.005	0.129	−0.003	0.149
	0.9	0.067	0.308	0.007	0.258	−0.014	0.310
$\epsilon \sim (1 + x_2)\mathcal{N}(0, 1)$							
Classical	0.1	−0.142	0.301	0.583	0.635	0.048	0.243
	0.5	0.004	0.155	0.006	0.164	0.001	0.073
	0.9	0.147	0.290	−0.584	0.631	−0.042	0.238
Gibbs without σ	0.1	−0.283	0.371	0.610	0.652	0.057	0.199
	0.5	0.004	0.145	0.006	0.159	0.001	0.088
	0.9	0.279	0.356	−0.608	0.643	−0.060	0.194
Gibbs with σ	0.1	−0.171	0.306	0.589	0.636	0.048	0.219
	0.5	0.004	0.147	0.006	0.159	0.001	0.081
	0.9	0.171	0.290	−0.586	0.626	−0.047	0.213

outperform the classical method in almost all cases. In addition, the inclusion of the additional parameter σ reduces the bias without the loss of accuracy. Although not reported here, we tried other error distributions and hyperparameters to assess the sensitivity of the proposed method and obtained similar results.

4.2. *Patent data*

As an real data example, we consider the data examined by Wang *et al*. [30]. 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*. [31]. Following Tsionas [10], we analyse the relationship between patents and research and development (R&D) by estimating the model

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

where N is the number of patent applications, RD is the R&D spending and RD/SALE is the ratio of R&D to sales. In this study, we included a scale parameter σ and used the same priors as in Tsionas [10]. To fit the model, we ran the Gibbs sampler for 15,000 iterations and discarded the first 5000 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 decline to zero by lag 10. Furthermore, a comparison with Figure 3 of Tsionas [10] shows that the autocorrelations from our Gibbs sampler decay more quickly than those from the one of Tsionas [10]. This is because the Gibbs sampler of Tsionas [10] updates each element of $\boldsymbol{\beta}_p$ separately, whereas our algorithm draws all the elements jointly.

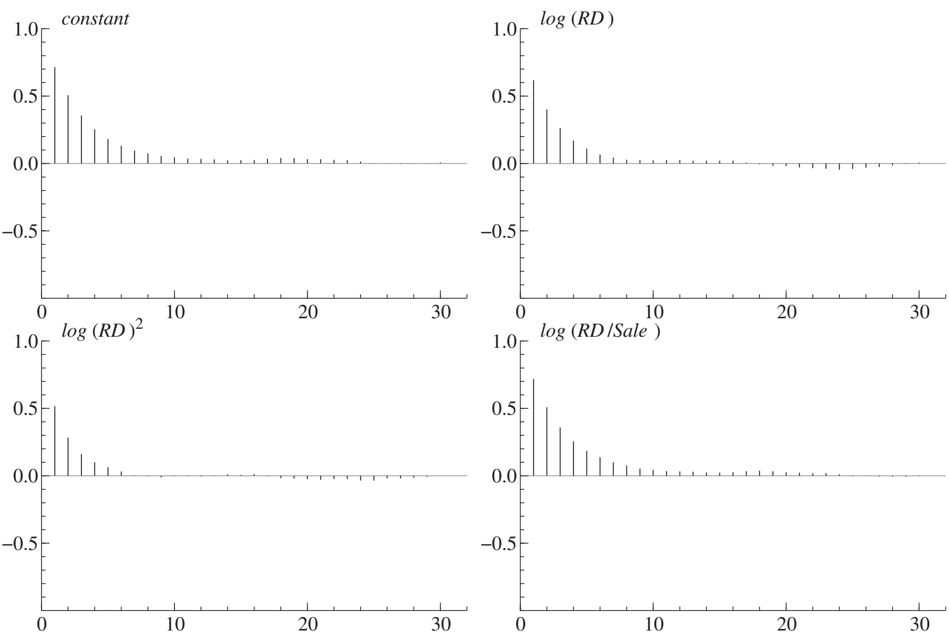

Figure 1. Patent data: sample autocorrelation functions are shown for $p = 0.5$.

4.3. *Simulation 2*

To investigate the performance of our method in Tobit regression models, we again conducted a simulation experiment. We used the same settings as before, except that the response variable was generated from Equation (17) left censored at zero. The proportion of censored observations was roughly 25% in all cases. In each replication, we ran the Gibbs samplers for 12,000 iterations with a burn-in of 2000 iterations to estimate the Tobit quantile regression models with and without a scale σ. For comparison, we also considered the Powell [21] estimator, which is commonly used in practice and can be implemented by the function `crq` in the `quantreg` package.

Table 3 summarizes the simulation results obtained from 1000 replications. From the table, we observe that the overall performance of the Powell estimator is satisfactory though its RMSEs are large when $p = 0.1$. In contrast, the Gibbs sampler without σ yields negatively biased estimates of β_1 for $p = 0.1$. Furthermore, the magnitude of the bias and the corresponding RMSE become the largest among the three methods when the error is the heteroscedastic normal. The reason for the negative bias may be explained as follows: as seen from Equation (16), the Gibbs sampler without σ simulates the latent variables y_i^* from $\mathcal{TN}_{(-\infty,0]}(\mathbf{x}_i'\boldsymbol{\beta}_p + \theta v_i, \tau^2 v_i)$ if observations are censored. Since the value of τ^2 becomes 22.2 when $p = 0.1$, simulated values of y_i^* can be small unless v_i is sufficiently small. Consequently, the posterior mean of the intercept becomes negatively

Table 3. Simulated censored data: biases and RMSEs are shown.

Method	p	β_1 Bias	β_1 RMSE	β_2 Bias	β_2 RMSE	β_3 Bias	β_3 RMSE
$\epsilon \sim \mathcal{N}(0, 1)$							
Powell	0.1	−0.011	1.018	−0.049	0.536	−0.061	0.754
	0.5	−0.014	0.303	−0.018	0.439	−0.015	0.620
	0.9	−0.031	0.215	−0.039	0.212	−0.035	0.260
Gibbs without σ	0.1	−2.288	2.364	1.010	1.058	1.297	1.356
	0.5	−0.113	0.185	0.110	0.177	0.121	0.202
	0.9	0.067	0.166	0.062	0.177	0.028	0.181
Gibbs with σ	0.1	0.147	0.232	−0.075	0.184	−0.074	0.203
	0.5	0.000	0.127	0.004	0.127	0.007	0.145
	0.9	−0.001	0.156	0.004	0.166	−0.003	0.182
$\epsilon \sim t_3$							
Powell	0.1	−0.095	2.934	−0.169	1.306	−0.276	1.281
	0.5	−0.009	0.262	−0.011	0.251	−0.003	0.309
	0.9	0.042	0.348	−0.062	0.299	−0.083	0.397
Gibbs without σ	0.1	−2.336	2.427	1.009	1.086	1.281	1.387
	0.5	−0.105	0.195	0.089	0.184	0.098	0.211
	0.9	0.158	0.331	0.065	0.279	0.009	0.322
Gibbs with σ	0.1	0.242	0.326	−0.124	0.265	−0.114	0.294
	0.5	−0.023	0.147	0.012	0.149	0.013	0.170
	0.9	0.075	0.310	0.019	0.274	−0.010	0.324
$\epsilon \sim (1 + x_2)\mathcal{N}(0, 1)$							
Powell	0.1	−0.533	5.923	0.715	2.945	−0.238	1.100
	0.5	−0.021	0.240	−0.006	0.303	0.002	0.391
	0.9	0.020	0.263	−0.286	0.462	−0.015	0.314
Gibbs without σ	0.1	−2.972	3.034	1.453	1.531	1.972	2.032
	0.5	−0.157	0.230	0.180	0.274	0.269	0.325
	0.9	0.159	0.256	−0.311	0.423	0.129	0.263
Gibbs with σ	0.1	−0.134	0.253	0.622	0.659	0.096	0.261
	0.5	−0.066	0.168	0.090	0.212	0.128	0.206
	0.9	0.095	0.228	−0.343	0.450	0.044	0.236

biased. This consideration suggests that the Gibbs sampler would produce better results if the scale parameter was incorporated into the model. In fact, we see from Table 3 that including the scale parameter substantially reduces the bias of the Bayes estimate and improves the estimation accuracy as expected. These findings differ from those obtained from the simulation study of non-censored data, where the inclusion of σ has little effect on the estimation accuracy. Finally, it is observed that the Gibbs sampler with the scale parameter attains the smallest RMSE in almost all cases.

4.4. *Labour supply data*

In this subsection, we examine the data set from Mroz [32] who carried out a systematic analysis of theoretical and statistical assumptions used in empirical models of female labour supply. This data set comprised 753 married women between the ages of 30 and 60, and has been analysed by many researchers. As the response variable, we consider the total number of hours the wife worked for a wage outside the home during the year 1975, which is measured in 100 h. 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. We include as covariates 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 six years old ($kidslt6$) and the number of children over six years old ($kidsge6$). Since the response variable is censored, we investigate the relationship between the number of hours worked and the covariates from estimating the Tobit quantile regression model.

For the regression coefficients, we considered both the normal and double exponential priors 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. Taking into account the results of the previous subsection, we incorporated a scale parameter σ to obtain efficient estimates for all range of p and assumed that $\sigma \sim \mathcal{IG}(3/2, 0.1/2)$. All results are based on an MCMC sample of 10,000 draws obtained after a burn-in of 5000 iterations.

We compared the proposed Bayesian method with the Powell estimator, and Table 4 shows the estimation results of median regression. As a benchmark, we also present the maximum likelihood (ML) estimates of the standard Tobit model. From the table, we observe that both the priors produce similar posterior estimates. Moreover, a comparison of the Bayes estimates with the ML estimates shows that both estimates have the same sign and are very similar. However, as for the Powell estimates, the coefficient on $nwifeinc$ has a different sign and the estimates on $educ$ and $exper$ are smaller than the corresponding Bayes and ML estimates. It is well known that the objective function for the censored quantile regression problem is no longer convex [3], so that our results suggest that the Powell estimates might be local minima.

Table 4. Labour supply data: estimation results are shown.

	Normal		Double exponential		Powell		ML	
Covariate	Mean	SD	Mean	SD	Estimate	SE	Estimate	SE
$constant$	11.951	4.031	11.298	4.572	18.898	3.786	9.653	4.464
$nwifeinc$	−0.098	0.044	−0.099	0.044	0.030	0.034	−0.088	0.045
$educ$	0.863	0.205	0.872	0.219	0.162	0.207	0.806	0.216
$exper$	1.413	0.180	1.414	0.180	0.854	0.161	1.316	0.173
$exper^2$	−0.018	0.006	−0.018	0.006	−0.002	0.006	−0.019	0.005
age	−0.610	0.069	−0.598	0.073	−0.532	0.065	−0.544	0.074
$kidslt6$	−9.724	1.135	−9.613	1.197	−7.125	0.749	−8.940	1.119
$kidsge6$	−0.426	0.395	−0.400	0.397	−0.090	0.286	−0.162	0.386

As mentioned in Section 1, quantile regression can offer more detailed information on the distribution of response variable than mean regression. To study closely the relationship between the wife's hours of work and the covariates, Figure 2 plots the posterior estimates of $\boldsymbol{\beta}_p$ against various values of p under the normal prior. We observe some interesting evidences from the figure, which cannot be detected by mean regression. For example, $educ$ is positively associated with the wife's hours of work across all the quantiles, but its posterior mean becomes smaller as the value of p increases. This indicates that the education background becomes less important for the wife who works longer. We can see a similar pattern in the result for $exper$ though its influence appears to be of an inverted-U shape. On the other hand, $nwifeinc$, $kidslt6$ and age have negative effects on the wife's working time. While they are highly related in the middle quantiles, the effects of $kidslt6$ and age become smaller in the tails of response distribution. Table 4 shows that the ML estimate on $nwifeinc$ is significantly different from zero at the 5% level, but the 95% credible interval for $nwifeinc$ contains zero when the value of p exceeds around 0.6. Finally, the results

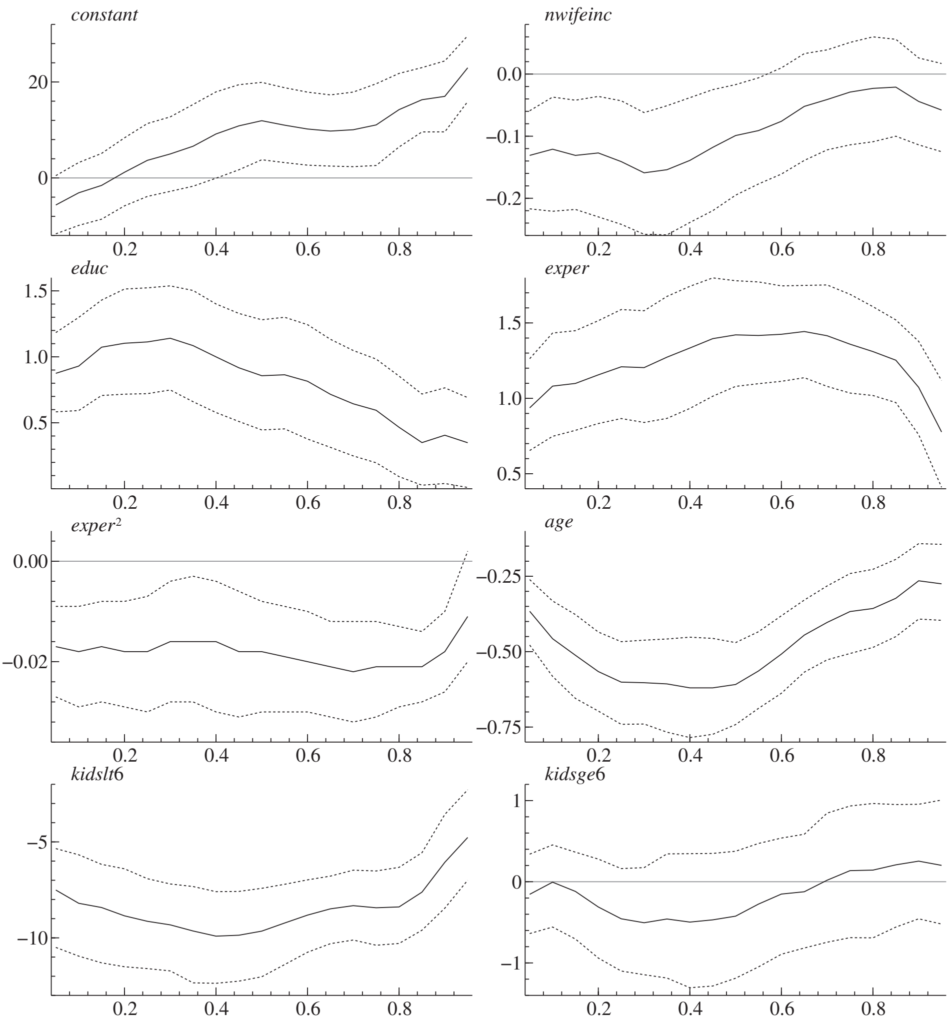

Figure 2. Labour supply data: posterior means (solid line) and 95% credible intervals (dotted lines) are plotted.

for $exper^2$ 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. The simulation studies and empirical examples have demonstrated the superiority of the proposed methods to the existing methods.

Although quantile regression originally does not require any distributional assumption, our approach assumes a 'pseudo'-asymmetric Laplace distribution for the error term to achieve the right quantiles. Our simulation studies have shown that the finite sample performance of the proposed methods is reasonably well even when the error distribution is away from the asymmetric Laplace distribution. In addition, recent work by Chernozhukov and Hong [33] provides a theoretical justification for our approach in a large sample.

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.

Acknowledgements

The authors are grateful to an associated editor and two anonymous referees for their useful comments, which improved an earlier version of the paper. 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 [34].

References

[1] R. Koenker and G. Bassett, *Regression quantiles*, Econometrica 46 (1978), pp. 33–50.
[2] K. Yu, Z. Lu, and J. Stander, *Quantile regression: Applications and current research areas*, Statistician 52 (2003), pp. 331–350.
[3] R. Koenker, *Quantile Regression*. Cambridge University Press, New York, 2005.
[4] R. Koenker and B.J. Park, *An interior point algorithm for nonlinear quantile regression*, J. Econom. 71 (1996), pp. 265–283.
[5] S. Portnoy and R. Koenker, *The Gaussian hare and the Laplacian tortoise: Computability of squared-error versus absolute-error estimators (with discussion)*, Statist. Sci. 12 (1997), pp. 279–300.
[6] S.G. Walker and B.K. Mallick, *A Bayesian semiparametric accelerated failure time model*, Biometrics 55 (1999), pp. 477–483.
[7] A. Kottas and A.E. Gelfand, *Bayesian semiparametric median regression modeling*, J. Am. Statist. Assoc. 96 (2001), pp. 1458–1468.
[8] T. Hanson and W.O. Johnson, *Modeling regression error with a mixture of Pólya trees*, J. Am. Statist. Assoc. 97 (2002), pp. 1020–1033.
[9] K. Yu and R.A. Moyeed, *Bayesian quantile regression*, Stat. Probab. Lett. 54 (2001), pp. 437–447.
[10] E.G. Tsionas, *Bayesian quantile inference*, J. Statist. Comput. Simul. 73 (2003), pp. 659–674.
[11] A. Kottas and M. Krnjajić, *Bayesian semiparametic modelling in quantile regression*, Scand. J. Stat. 36 (2008), pp. 297–319.
[12] K. Yu and J. Zhang, *A three-parameter asymmetric Laplace distribution and its extension*, Commun. Stat. Theory Methods 34 (2005), pp. 1867–1879.
[13] S. Kotz, T.J. Kozubowski, and K. Podgórski, *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Boston: Engineering, and Finance*, Birkhäuser, Boston, 2001.

[14] O.E. Barndorff-Nielsen and N. Shephard, *Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics*, J. R. Statist. Soc. B63 (2001), pp. 167–241.

[15] J. Dagnapur, *An easily implemented generalized inverse Gaussian generator*, Commun. Stat. Simul. Comput. 18 (1989), pp. 703–710.

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

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

[18] K. Yu and J. Stander, *Bayesian analysis of a Tobit quantile regression model*, J. Econom. 137 (2005), pp. 260–276.

[19] T. Park and G. Casella, *The Bayesian Lasso*, J. Am. Statist. Assoc. 103 (2008), pp. 681–686.

[20] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. R. Statist. Soc. B58 (1996), pp. 267–288.

[21] J. Powell, *Censored regression quantiles*, J. Econom. 32 (1986), pp. 143–155.

[22] J. Hahn, *Bootstrapping quantile regression estimators*, Econom. Theory 11 (1995), pp. 105–121.

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

[24] Y. Bilias, S. Chen, and Z. Ying, *Simple resampling methods for censored regression quantiles*, J. Econom. 68 (2000), pp. 303–338.

[25] S. Chib, *Bayes inference in the Tobit censored regression model*, J. Econom. 51 (1992), pp. 79–99.

[26] C.J. Geyer, *Practical Markov chain Monte Carlo*, Statist. Sci. 7 (1992), pp. 473–511.

[27] J.M. Flegal, M. Haran, and G.L. Jone, *Markov chain Monte Carlo: Can we trust the third significant figure?* Statist. Sci. 23 (2008), pp. 250–260.

[28] S. Chib, *Markov chain Monte Carlo methods: Computation and inference*, in: *Handbook of Econometrics*, Vol. 5, J.J. Heckman and E. Leamer, eds., North Holland, Amsterdam, 2001, pp. 3569–3649.

[29] R. Koenker, *quantreg: Quantile Regression*. R package version 4.30; available at http://CRAN.R-project.org/package=quantreg, 2009.

[30] P. Wang, I. Cockburn, and M.K. Puterman, *Analysis of patent data – a mixed-Poisson-regression-model approach*, J. Bus. Econ. Stat. 16 (1998), pp. 27–41.

[31] B.H. Hall, C. Cummins, E. Laderman, and J. Mundy, *The R&D master file documentation*, Tech. Working Paper 72, National Bureau of Economic Research, Cambridge, MA, 1998.

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

[33] V. Chernozhukov and H. Hong, *An MCMC approach to classical estimation*, J. Econom. 115 (2003), pp. 293–346.

[34] J.A. Doornik, *Object-Oriented Matrix Programming Using Ox*, 3rd edn. Timberlake Consultants Press, London, 2007.