Bayesian Tobit quantile regression using g-prior distribution with ridge parameter

Rahim Alhamzawi*, Keming Yu

Department of Mathematics, Brunel University, Uxbridge UB8 3PH, UK

Abstract

A Bayesian approach is proposed for coefficient estimation in Tobit quantile regression model. The proposed approach is based on placing a g-prior distribution depends on the quantile level on the regression coefficients. The prior is generalized by introducing a ridge parameter to address important challenges that may arise with censored data, such as multicollinearity and overfitting problems. Then, a stochastic search variable selection approach is proposed for Tobit quantile regression model based on g-prior. An expression for the hyperparameter g is proposed to calibrate the modified g-prior with a ridge parameter to the corresponding g-prior. Some possible extensions of the proposed approach are discussed, including the continuous and binary responses in quantile regression. The methods are illustrated using several simulation studies and a microarray study. The simulation studies and the microarray study indicate that the proposed approach performs well.

Keywords: g-prior, Gibbs sampler, Ridge parameter, Tobit quantile regression, Variable selection.

1. Introduction

Quantile regression (QReg) models have received much attention in the literature since the seminal work of Koenker and Bassett (1978), and a variety of parametric, semiparametric

---

*Corresponding author.

Email address: ralhamzawi@yahoo.com Tel.: +44(0)1895 265618 (Rahim Alhamzawi)

and nonparametric estimation approaches have been proposed over the years. The Tobit QReg model provides an efficient way of coping with left-censored data, and can be viewed as a linear QReg model where only the data on the dependent variable is incompletely observed. A great body of work exists on Tobit QReg methods and we refer to Powell (1986), Hahn (1995), Buchinsky and Hahn (1998), Bilias et al. (2000), Yu and Stander (2007) and Wang and Fygenson (2009) for an overview.

Consider the model,

$$y_i = \max\{y^0, y_i^*\}, \quad i = 1, \cdots, n, \tag{1}$$

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i,$$

where y_i is the observed outcome of interest, y_i^* is the corresponding latent unobserved outcome of subject i, y^0 is a known censoring point, $\boldsymbol{x}_i'$ is a $1 \times k$ vector denoting the ith row of the $n \times k$ matrix of predictors $\boldsymbol{X}$, $\boldsymbol{\beta}$ is a vector of unknown parameters of interest evaluate at pth quantile, and ε_i is the error term whose distribution is restricted to have the pth quantile equal to zero, that is, $\int_{-\infty}^{0} f_p(\varepsilon_i)d\varepsilon_i = p$. Following Powell (1986), it can be shown that the regression coefficients vector $\boldsymbol{\beta}$ can be estimated by the solution to the following problem

$$\sum_{i=1}^{n} \rho_p(y_i - \max\{y^0, y_i^*\}), \tag{2}$$

where $\rho_p(\cdot)$ is the usual check function of Koenker and Bassett (1978), which is defined by

$$\rho_p(\varepsilon) = \frac{|\varepsilon| + (2p - 1)\varepsilon}{2}. \tag{3}$$

Yu and Stander (2007) observed that the posterior estimator of $\boldsymbol{\beta}$ obtained by assigning a likelihood function that is based on the asymmetric Laplace distribution at specific value

2

of p, serves as the pth Tobit QRreg estimate. Here, the density function of the ALD is given by (Yu and Moyeed, 2001)

$$f(\varepsilon) = \tau^{-1}p(1-p)\exp\{-\tau^{-1}\rho_p(\varepsilon)\}, \tag{4}$$

The authors assigned flat priors, independent of the value of p, for the Tobit QRreg coefficients vector and sampling $\boldsymbol{\beta}$ using the Metropolis-Hastings (MH) method. It is well known that flat priors could be useful for coefficient estimation in Tobit QRreg and other models but they cannot be used in subset selection techniques, owing to the fact that proper priors are needed to evaluate Bayes factors (Ibrahim and Chen, 2000). Yu and Stander (2007) also suggested families of symmetric prior distributions on the Tobit QRreg coefficients vector, such as normal and Laplace priors. Although these priors may lead to proper posterior, they are independent of the values of quantiles. That is, the prior is the same for modelling different order of quantiles. This approach may result in inflexibility in quantile modelling. Because the posterior distribution of $\boldsymbol{\beta}$, the variance-covariance matrix of $\boldsymbol{\beta}$ and the marginal distribution of the data depend on the quantile level, extreme quantiles should have different parameter values from the median. In addition, a correct estimate of the prior distribution perhaps leads to support insufficiency of data (Agliari and Parisetti, 1988). Thus, it is crucial to elicit a prior distribution for QRreg coefficients that is as informative as possible, and more crucially, that depends on the quantile level. For such situations and others, Alhamzawi and Yu (2013) proposed a modification of Zellner's g-prior (Zellner, 1986) to be a quantile dependent prior for quantile regression (QReg) models. This modification can be written as

$$\boldsymbol{\beta}|\tau, \boldsymbol{V}, \boldsymbol{X} \sim N(0, 2\tau g(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1}), \quad v_i \sim \text{Exp}(\tau^{-1}p(1-p)), \quad i = 1, ..., n, \tag{5}$$

where $\boldsymbol{V} = \text{diag}(v_1^{-1}, ..., v_n^{-1})$ and $g > 0$ is a known scaling factor. Here, $\text{Exp}(\theta)$ denotes the

density of an exponential distribution with rate parameter θ. In this paper, we use this prior to develop the Bayesian analysis of the Tobit QReg model. Then, we generalized the g-prior by introducing a ridge parameter to address some issues that may arise with censored data such as, multicollinearity and overfitting problems. We also developed an expression for the hyperparameter g to calibrate the modified g-prior with a ridge parameter to a corresponding g-prior. Furthermore, we discuss some possible extensions of our approach, including the continuous and binary responses in QReg.

The remainder of the paper is structured as follows. In Section 2, the structure of our hierarchical Bayesian Tobit QReg model is described in detail and an extension of the modified g-prior is suggested. A choice for the hyper-parameter g is presented and the Bayesian MCMC estimation procedure is outlined. An approach for model selection is presented in Section 3. Some possible extensions of the proposed approach to QReg with continuous and binary responses are discussed. Section 4 evaluates the methods using simulation examples, and Section 5 gives the microarray study. A brief discussion is introduced in Section 6.

2. Methods

2.1. Hierarchical Bayesian Modelling

At the pth quantile, we model conditional Tobit quantiles of the response by using the Asymmetric Laplace Distribution (ALD) for the errors. More specifically, the location-scale mixture representation of the ALD (Kozumi and Kobayashi, 2011; Kotz and Podgrski, 2001) is used to build an efficient Gibbs sampler. This representation can be written as

$$\tau^{-n} \exp\{-\sum_{i=1}^{n} \frac{|\varepsilon_i| + (2p-1)\varepsilon_i}{2\tau}\} \quad = \prod_{i=1}^{n} \int_{0}^{\infty} \frac{1}{\tau\sqrt{4\tau\pi v_i}} \exp\{-\frac{(\varepsilon_i - \xi v_i)^2}{4\tau v_i} - \zeta v_i\} dv_i, \quad (6)$$

where $\xi = (1 - 2p)$ and $\zeta = \tau^{-1}p(1 - p)$; see Alhamzawi and Yu (2013) for some details. Formula (6) has the advantage that there is no need to worry about the prior distribution of

v_i as it is extracted in the same formula. It is easy to show that the exact prior of v_i in formula

(6) is exponential with mean $\tau^{-1}p(1-p)$ and variance $(\tau^{-2}p^2(1-p)^2)$ and strongly depends

on the value of p. However, the mode of v_i is 0 for any p. To complete the prior specification,

we take a quantile dependent prior for $\boldsymbol{\beta}$ so that $\pi(\boldsymbol{\beta}|\tau, \boldsymbol{V}, \boldsymbol{X}) = N(0, 2\tau g(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1})$. This

choice for $\boldsymbol{\beta}$ has several attractive features. First, under this setting the prior distribution

of $\boldsymbol{\beta}$ depends on the quantile level. Thus, we have different priors for different quantiles.

For example, a 95% QReg model should have different parameter values from the median

quantile, and thus the priors used for modelling the quantiles should be different. Second,

this prior is conditionally conjugate, a property that allows for a simple and efficient Gibbs

sampling algorithm for fitting the model. Third, in the case of $\boldsymbol{V} = \boldsymbol{I}_n$ and $2\tau = \sigma^2$, the

proposed prior is reduced to the original g-prior, i.e. $\pi(\boldsymbol{\beta}|\sigma^2, \boldsymbol{X}) = N(0, g\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$.

Finally, as $g \longrightarrow \infty$, the proposed prior distribution of $\boldsymbol{\beta}$ converges to Jeffrey's prior of

the form $\pi(\boldsymbol{\beta}|\boldsymbol{V}_0, \boldsymbol{X}) \propto |\boldsymbol{X}'\boldsymbol{V}_0\boldsymbol{X}|^{1/2}$ with $\boldsymbol{V}_0 = \text{diag}((2\tau v_1)^{-1}, \cdots, (2\tau v_n)^{-1})$, which is a

popular noninformative prior for Bayesian analysis. To summarize, our hierarchical Bayesian

modelling is given by

$$
\begin{aligned}
&y_i = \max\{y^0, y_i^*\}, \quad i = 1, \cdots, n, \\
&y_i^*|\boldsymbol{\beta}, \tau, v_i \sim N(\boldsymbol{x}_i'\boldsymbol{\beta} + \xi v_i, 2\tau v_i), \\
&\boldsymbol{\beta}|\tau, \boldsymbol{V}, \boldsymbol{X} \sim N(0, 2\tau g(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1}), \\
&v_i|\tau \sim \text{Exp}(\tau^{-1}p(1-p)), \\
&p(\tau) \propto \tau^{-1}.
\end{aligned}
\tag{7}
$$

Since our target in the SSVS approach required computation of the marginal distribution

of the data $p(\boldsymbol{y}^*|\tau, \boldsymbol{v})$ with $\boldsymbol{v} = (v_1, \cdots, v_n)'$, the following lemma gives the closed-form of

$p(\boldsymbol{y}^*|\tau, \boldsymbol{v})$ under the proposed prior.

Lemma 1. Under the quantile dependent prior (5), the conditional distribution of the data

$p(\boldsymbol{y}^*|\tau, \boldsymbol{v})$ is given by

$$p(\boldsymbol{y}^*|\tau, \boldsymbol{v}) = \frac{(1+g)^{-k/2}}{(4\pi)^{n/2}}(\prod_{i=1}^{n}(\tau v_i)^{-1/2})$$

$$\times \exp\{-(\boldsymbol{y}^* - \xi\boldsymbol{v})'(\frac{\boldsymbol{V}}{4\tau} - \frac{g\boldsymbol{V}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}}{4(1+g)\tau})(\boldsymbol{y}^* - \xi\boldsymbol{v})\}. \qquad (8)$$

The proof of Lemma 1 is straightforward and can be accomplished by integrating out the quantile coefficients vector $\boldsymbol{\beta}$ as in Smith and Kohn (1996).

2.2. Introducing a ridge parameter

In the original g-prior, the matrix $\boldsymbol{X}'\boldsymbol{X}$ suffers from singularity in case of multicollinearity or overfitting problems ($k >> n$). For this reason, Gupta and Ibrahim (2007) proposed a modification of the original Zellner's g-prior, motivated by the ridge parameter λ which comes from ideas of *ridge regression* to deal with multicollinearity and overfitting problems. The authors showed that their technique allows consistent subset selection and coefficient estimation for overfitting problems. Baragatti and Pommeret (2012) considered the influence of λ on the subset selection and suggested a technique to select the scaling factor. Similar to Gupta and Ibrahim (2007), in the case of singularity of the matrix $\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}$, we modified our prior with the ridge parameter ($\lambda > 0$). More specifically, we propose the following prior for $\boldsymbol{\beta}$:

$$\boldsymbol{\beta}|\tau, \boldsymbol{V}, \boldsymbol{X} \sim N(0, 2\tau g_\lambda(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X} + 2\lambda\boldsymbol{I}_k)^{-1}), \qquad (9)$$

where $g_\lambda > 0$ is a known scaling factor characterised by the parameter λ and $\boldsymbol{I}_k$ is the $k \times k$ identity matrix. In this paper, we assume $g_\lambda \neq g$.

Clearly in order for the conditional distribution of the quantile coefficients vector $\boldsymbol{\beta}$ under the prior (5) and the conditional distribution of $\boldsymbol{\beta}$ under the prior (9) to have identical conditional distributions, we need $g(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1} = g_\lambda(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X} + 2\lambda\boldsymbol{I}_k)^{-1}$. The following

lemma characterises the relationship among the three parameters g, g_λ and λ.

Lemma 2. The posterior distribution of the quantile coefficients vector $\boldsymbol{\beta}$ under the prior (5) and the posterior distribution of $\boldsymbol{\beta}$ under the prior (9) are identical distributions if and only if

$$g_\lambda \boldsymbol{I}_k = g[\boldsymbol{I}_k + 2\lambda(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1}]. \tag{10}$$

This lemma provides a technique to elicit g_λ and the proof of Lemma 2 is straightforward. By taking the trace of left and right hand sides of (10), we are led to

$$\hat{g}_\lambda = \frac{1}{k}\mathbb{E}[gk + \frac{2g\lambda}{\text{tr}(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})}], \tag{11}$$

where the expectation in Equation (11) is taken with respect to the posterior distribution of $\boldsymbol{V}$.

2.3. Choice of g and λ

Various values of g have been used in the context of variable selection and estimation. For example, Kass and Wasserman (1995) proposed the general idea of the unit information prior, i.e. $g = n$. Smith and Kohn (1996) performed variable selection using splines and suggested that the value of g is in the range $10 \leq g \leq 1000$. Following this suggestion, a number of authors set $g = 100$ (see for example, Lee et al., 2003; Gupta et al., 2007; Chen et al., 2011, among others). Although for linear regression models, placing an Inverse Gamma prior on g, $g \sim \text{InvGa}(1/2, n/2)$, leads to a multivariate Cauchy prior on $\boldsymbol{\beta}$ which is recommended as a robust prior for Bayesian variable selection in regression models (Clyde and George, 2004; Zellner and Siow, 1980), the corresponding marginal likelihood $f(\boldsymbol{y}|\boldsymbol{\gamma})$ has no closed form, where $\boldsymbol{\gamma}$ is a latent k-vector with binary entries: $\gamma_j = 1$ if the jth

covariate is active in the regression and $\gamma_j = 0$ otherwise. For this reason, Liang et al. (2008) suggested the idea of the hyper-g prior. Cui and George (2008) proposed an inverse-gamma prior distribution on $(1 + g)$ (not g) and Celeux et al. (2012) suggested a Jeffrey prior on g. In this paper, we choose $g = 100$ and we choose $\lambda = 1/k$ which has been suggested by Baragatti and Pommeret (2012) which lies between 0 and 1 as recommended for Bayesian robustness (Gupta and Ibrahim, 2007).

2.4. Posterior Computation

The hierarchical modelling (7) produces an efficient MCMC algorithm by updating the latent variables y_i^* and v_i as well as the other parameters $\boldsymbol{\beta}$ and τ from their full conditional distributions.

- **Updating y_i^***

  Let $\Upsilon(.)$ denotes to a degenerate distribution, then the latent variable $y_i^*, i = 1, \cdots, n$, has a conditional distribution given by

$$y_i^*|y_i, \boldsymbol{\beta}, \tau, \boldsymbol{V} \sim \begin{cases} \Upsilon(y_i), & \text{if } y_i > y^0; \\ N(\boldsymbol{x}_i'\boldsymbol{\beta} + \xi v_i, 2\tau v_i)I(y_i^* \le y^0), & \text{otherwise,} \end{cases} \tag{12}$$

- **Updating $\boldsymbol{\beta}$**

  The full conditional distribution of $\boldsymbol{\beta}$ is $\mathrm{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\Sigma} = 2\tau[\frac{g_\lambda + 1}{g_\lambda}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X} + \frac{2\lambda}{g_\lambda}\boldsymbol{I}_k]^{-1} \quad \text{and} \quad \boldsymbol{\mu} = (2\tau)^{-1}\boldsymbol{\Sigma}\boldsymbol{X}'\boldsymbol{V}(\boldsymbol{y}^* - \xi\boldsymbol{v}). \tag{13}$$

  Here, $\boldsymbol{y}^* = (y_1^*, \cdots, y_n^*)'$.

- **Updating τ**

$$\tau | \boldsymbol{y}^*, \boldsymbol{\beta}, \boldsymbol{v} \sim \text{InvGa}((3n+k)/2, \frac{1}{4}(\boldsymbol{y}^* - \boldsymbol{X}\boldsymbol{\beta} - \xi\boldsymbol{v})'\boldsymbol{V}(\boldsymbol{y}^* - \boldsymbol{X}\boldsymbol{\beta} - \xi\boldsymbol{v})$$
$$+ \frac{1}{4g}\boldsymbol{\beta}'(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})\boldsymbol{\beta} + p(1-p)\sum_{i=1}^{n} v_i).$$

- **Updating v_i**

For $i = 1, ..., n$, each $v_i \sim \text{GIG}(\nu, \varrho_1, \varrho_2)$, where $\nu = 0$, $\varrho_1^2 = ((y_i^* - \boldsymbol{x}_i'\boldsymbol{\beta})^2 + \boldsymbol{\beta}'\boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{\beta}/g)/(2\tau)$ and $\varrho_2^2 = 1/(2\tau)$.

Recall that if $x \sim \text{GIG}(\nu, \varrho_1, \varrho_2)$ then the probability density function of x is given by (Barndorff-Nielsen and Shephard, 2001)

$$f(x|\nu, \varrho_1, \varrho_2) = \frac{(\varrho_2/\varrho_1)^\nu}{2K_\nu(\varrho_1\varrho_2)} x^{\nu-1} \exp\left\{-\frac{1}{2}(x^{-1}\varrho_1^2 + x\varrho_2^2)\right\},$$

where $x > 0$, $-\infty < \nu < \infty$, $\varrho_1, \varrho_2 \geq 0$ and $K_\nu(.)$ is so called *"modified Bessel function of the third kind"*. During MCMC iteration we update g_λ using $g_\lambda = [gk + 2g\lambda/\text{tr}(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})]$ where $g = n$ and $\lambda = 1/k$. In the case of nonsingularity of the matrix $\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}$, we set $\lambda = 0$ and $g_\lambda = g$.

3. Stochastic Search Variable Selection (SSVS)

3.1. Priors specification and Bayesian sampler

The approach of Ji et al. (2012) allows for Bayesian variable selection in Tobit QReg; however, their approach had the disadvantages of relying on priors that are independent of the value of quantiles. This approach may result in inflexibility in quantile modelling. Our initial goal is to address this problem using the proposed prior in Section 2.

Given $p \in (0, 1)$ and $\tau = 1$, we consider the following prior distribution assumptions:

- The prior distribution of $\boldsymbol{\beta_\gamma}$ is taken as $\boldsymbol{\beta_\gamma}|\boldsymbol{V}, \boldsymbol{X_\gamma} \sim N(0, 2g_\lambda(\boldsymbol{X_\gamma'}\boldsymbol{V}\boldsymbol{X_\gamma}+2\lambda\boldsymbol{I}_{k_\gamma})^{-1})$, where $v_i \sim \mathrm{Exp}(p(1-p))$ for $i = 1, ..., n$.

- The prior of $\boldsymbol{\gamma}$ is taken as $p(\boldsymbol{\gamma}|\pi) \propto \pi^{k_\gamma}(1-\pi)^{k-k_\gamma}$ (George and McCulloch, 1993, 1997), where $\pi \sim \mathrm{Beta}(b_{01}, b_{02})$.

Under prior assumptions, we are able to use a MCMC based computation technique to update $\boldsymbol{y^*}, \boldsymbol{\beta_\gamma}, \boldsymbol{V}$ and π from the posterior:

- **Updating y_i^***

  Under $\boldsymbol{\gamma}$, the full conditional distribution of $y_i^*, i = 1, \cdots, n$, is reduced to

  $$y_i^*|y_i, \boldsymbol{\beta_\gamma}, \boldsymbol{V} \sim \begin{cases} \Upsilon(y_i), & \text{if } y_i > y^0; \\ N(\boldsymbol{x}_{i,\boldsymbol{\gamma}}'\boldsymbol{\beta_\gamma} + \xi v_i, 2v_i)I(y_i^* \le y^0), & \text{otherwise,} \end{cases} \tag{14}$$

- **Updating $\boldsymbol{\beta_\gamma}$**

  The full conditional distribution of $\boldsymbol{\beta_\gamma}$ is $\mathrm{N}_{k_\gamma}(\boldsymbol{\mu_\gamma}, \boldsymbol{\Sigma_\gamma})$, where

  $$\boldsymbol{\Sigma_\gamma} = 2[\frac{g_\lambda + 1}{g_\lambda}\boldsymbol{X_\gamma'}\boldsymbol{V}\boldsymbol{X_\gamma} + \frac{2\lambda}{g_\lambda}\boldsymbol{I}_{k_\gamma}]^{-1},$$
  $$\text{and } \boldsymbol{\mu_\gamma} = 2^{-1}\boldsymbol{\Sigma_\gamma}\boldsymbol{X_\gamma'}\boldsymbol{V}(\boldsymbol{y^*} - \xi\boldsymbol{v}).$$

- **Updating $\boldsymbol{v}$**

  The full conditional distribution of each v_i can be obtained from the full conditional distribution of v_i in the subsection 2.4 by setting $\tau = 1$ and replacing $\boldsymbol{x}_i'$ and $\boldsymbol{\beta}$ everywhere with $\boldsymbol{x}_{i,\boldsymbol{\gamma}}'$ and $\boldsymbol{\beta_\gamma}$, respectively.

- **Updating γ_j**

Each $\gamma_j, j = 1, \cdots, k$, has a full conditional distribution given by

$$p(\gamma_j = 1 | \boldsymbol{y}, \boldsymbol{y}^*, \boldsymbol{\beta_\gamma}, \boldsymbol{v}, \boldsymbol{\gamma}_{-j}) = \frac{1}{1 + h_j},$$

$$h_j = \frac{p(\boldsymbol{y}^* | \boldsymbol{y}, \boldsymbol{\beta_\gamma}, \boldsymbol{v}, \gamma_j = 0, \boldsymbol{\gamma}_{-j}) p(\boldsymbol{\beta_\gamma} | \gamma_j = 0, \boldsymbol{\gamma}_{-j}) p(\gamma_j = 0, \boldsymbol{\gamma}_{-j})}{p(\boldsymbol{y}^* | \boldsymbol{y}, \boldsymbol{\beta_\gamma}, \boldsymbol{v}, \gamma_j = 1, \boldsymbol{\gamma}_{-j}) p(\boldsymbol{\beta_\gamma} | \gamma_j = 1, \boldsymbol{\gamma}_{-j}) p(\gamma_j = 1, \boldsymbol{\gamma}_{-j})}.$$

- **Updating π**

  Since $\pi \sim \text{Beta}(b_{01}, b_{02})$, then, under model $\boldsymbol{\gamma}$ the full conditional distribution of π is $\text{Beta}(k_\gamma + b_{01}, k - k_\gamma + b_{02})$.

3.2. Model extensions

3.2.1. Subset selection in QReg

The proposed method in subsection 3.1 can be used, with some modifications, to find subset selection in QReg with continuous outcome variable. By ignoring the link function $y_i = \max\{y^0, y_i^*\}$ and replacing y^* everywhere with y, our approach offers an alternative way for subset selection in QReg model with continuous outcome to deal with multicollinearity and overfitting problems.

3.2.2. Subset selection in Binary QReg

In this subsection, we show that our technique reported in subsection 3.1 can be extended to subset selection for binary QReg model. Binary QReg models have received considerable interest in the literature and we refer to Manski (1975, 1985), Kordas (2006) and Benoit and Poel (2011) for an overview. Suppose y_i is a binary outcome variable (e.g. normal and cancer), then the binary QReg takes the form of (Manski, 1985)

$$y_i^* = \boldsymbol{x}_i' \boldsymbol{\beta} + \varepsilon_i, \tag{15}$$

$$y_i = 1 \text{ if } y_i^* \geq 0, \quad y_i = 0 \text{ otherwise.}$$

Under the above model, the proposed method in subsection 3.1 can be used to find promising subset in binary QReg by using the link function $y_i \sim 1(y_i^* \geq 0)$ and sampling $y_i^*, i = 1, \cdots, n$, as follows

$$y_i^*|y_i = 1, v_i, \boldsymbol{\beta_\gamma} \sim N(\boldsymbol{x}_{i,\gamma}'\boldsymbol{\beta_\gamma} + \xi v_i, 2v_i) \text{ truncated at the left by } 0,$$

$$y_i^*|y_i = 0, v_i, \boldsymbol{\beta_\gamma} \sim N(\boldsymbol{x}_{i,\gamma}'\boldsymbol{\beta_\gamma} + \xi v_i, 2v_i) \text{ truncated at the right by } 0.$$

4. Numerical Illustrations

4.1. Inference

4.1.1. Example 1

In this example, we consider our Bayesian Tobit QReg approach using g prior (BTQRg) and Bayesian Tobit QReg approach (BTQ) using a symmetric prior distribution, $\boldsymbol{\beta} \sim N_k(\boldsymbol{0}, 100\boldsymbol{I})$, as reported in Kozumi and Kobayashi (2011). These approaches were compared with the standard Tobit QReg approach (crq) employing Powell's method (Koenker, 2011). The simulation design follows the setting of Bilias et al. (2000) and Yu and Stander (2007). Data are simulated from the model

$$y_i = \max\{0, y_i^*\},$$

$$y_i^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i,$$

where $x_{1i} \sim \text{Bernoulli}(0.5)$ centered at zero, $x_{2i} \sim N(0, 1)$ and $(\beta_0, \beta_1, \beta_2) = (1, 1, 1)$. Three error distributions are considered: the standard normal N(0,1), a heteroscedastic normal distribution $(1 + x_2)N(0, 1)$ and a normal mixture distribution $0.75N(0, 1) + 0.25N(0, 4)$. Under these distributions the censoring level is approximately 30%. For each error distribution, we simulate 250 data sets assuming the sample size is n = 100. We fit the models at three different quantiles $p = 0.50, 0.75$ and $p = 0.95$. The MCMC algorithms are run for 17000

iterations, discarding the first 2,000 as burn in. Methods are evaluated based on the relative average bias

$$\widehat{\text{bias}}(\hat{\beta}_m) = \frac{1}{250} \sum_{r=1}^{250} \frac{\hat{\beta}_m^r - \beta_m}{\mid \beta_m \mid},$$

and the estimated relative efficiency

$$\widehat{\text{eff}}_{\text{model}}(\hat{\beta}_m) = \frac{S^2_{\text{model}}(\hat{\beta}_m)}{S^2_{\text{BTQRg}}(\hat{\beta}_m)},$$

where $\hat{\beta}_m^r, m = 0, 1$ is the estimated quantile for the rth replication, β_m is the true value, $S^2(\hat{\beta}_m) = \frac{1}{250} \sum_{r=1}^{250} (\hat{\beta}_m^r - \bar{\beta}_m)^2$ and $\bar{\beta}_m = \frac{1}{250} \sum_{r=1}^{250} \hat{\beta}_m^r$.

The simulation results for β_0, β_1 and β_2 are summarized in Table (1). Clearly, the biases due to three approaches are more or less the same. However, the proposed method (BTQRg) generally behaves much better than the other approaches (crq and BTQ) in terms of the absolute bias. Across the three error distributions, it can be noted that the absolute bias obtained from our proposed method is much smaller at extreme quantiles than the competing approaches. Most noticeably, when $p = 0.95$ the absolute bias generated by the proposed method for all parameters is much smaller than the absolute bias generated by the other approaches. In addition, the proposed method appears more efficient than the other methods. For example, when the error is standard normal and estimating the median, the loss of efficiency of the standard Tobit QReg (crq), with respect to our model, was 107% for β_2 and larger for the other parameters. We can also see that the the loss of efficiency of the BTQ, with respect to our model, was 1.23% for β_2 and larger for the other parameters. It may suggest that the Bayesian approach using a quantile dependent prior distribution has an advantages for modelling the extreme quantiles, which is desirable situation when attention is focused on the extreme quantiles.

Table 1: Estimated bias and relative efficiency for different error distributions. The proposed approach (BTQRg) is compared with two other approaches: the frequentist Tobit QReg approach (crq) and the Bayesian approach using symmetric prior distribution for the regression coefficients (BTQ).

Model	p	bias (eff.) β_0	bias (eff.) β_1	bias (eff.) β_2
$\varepsilon \sim N(0,1)$				
crq	0.50	-0.01466 (2.15047)	0.03376 (2.18275)	0.01217 (2.06735)
BTQ	0.50	0.00597 (1.14267)	0.00996 (1.06784)	-0.00789 (1.01235)
BTQRg	0.50	0.00640 (1.00000)	-0.01036 (1.00000)	-0.05614 (1.00000)
crq	0.75	-0.01052 (1.72184)	-0.00405 (1.32100)	-0.01455 (1.59715)
BTQ	0.75	-0.02690 (1.35108)	0.06365 (1.57798)	0.09263 (1.85198)
BTQRg	0.75	0.00284 (1.00000)	-0.05996 (1.00000)	-0.01000 (1.00000)
crq	0.95	0.01208 (1.13496)	-0.02656 (1.21129)	-0.12986 (0.97514)
BTQ	0.95	0.09551 (0.91855)	0.05976 (1.03759)	0.07646 (1.09315)
BTQRg	0.95	0.00201 (1.00000)	0.00079 (1.00000)	-0.04465 (1.00000)
$\varepsilon \sim (1+x_2)N(0,1)$				
crq	0.50	-0.15563 (9.78132)	0.16054 (21.48298)	0.01826 (2.33432)
BTQ	0.50	-0.07266 (1.32119)	0.09796 (2.85943)	0.05578 (1.22709)
BTQRg	0.50	0.07632 (1.00000)	-0.01073 (1.00000)	-0.01023 (1.00000)
crq	0.75	-0.01138 (1.21019)	0.05568 (2.65549)	-0.07541 (1.62722)
BTQ	0.75	-0.05331 (1.30429)	0.28926 (2.96351)	0.03059 (1.67248)
BTQRg	0.75	0.04686 (1.00000)	-0.04863 (1.00000)	-0.01050 (1.00000)
crq	0.95	0.08300 (1.32362)	-0.01790 (1.18841)	-0.31592 (1.98608)
BTQ	0.95	0.18887 (0.89628)	0.27505 (1.01940)	-0.21792 (1.40077)
BTQRg	0.95	0.13293 (1.00000)	-0.00770 (1.00000)	-0.14665 (1.00000)
$\varepsilon \sim 0.75N(0,1)$ $+0.25N(0,4)$				
crq	0.50	-0.02559 (2.81012)	0.00989 (2.73070)	-0.00652 (1.82914)
BTQ	0.50	-0.01951 (1.49375)	0.00640 (1.36406)	-0.00642 (0.90888)
BTQRg	0.50	0.00207 (1.00000)	-0.00570 (1.00000)	-0.01228 (1.00000)
crq	0.75	-0.13335 (1.42499)	0.01562 (1.54906)	-0.01233 (1.78678)
BTQ	0.75	-0.14603 (1.86671)	0.08703 (1.85623)	0.10127 (1.89332)
BTQRg	0.75	-0.09073 (1.00000)	-0.04322 (1.00000)	-0.00878 (1.00000)
crq	0.95	-0.18130 (1.58685)	-0.05828 (1.15029)	-0.11391 (0.99865)
BTQ	0.95	-0.10930 (1.08060)	0.05055 (0.95292)	0.10552 (1.05016)
BTQRg	0.95	-0.10797 (1.00000)	-0.00999 (1.00000)	-0.02471 (1.00000)

We may also look at the estimation of the regression coefficients and 95% credible intervals (95% CrI). Table 2 compares the posterior means of the Bayesian methods with the true values of $\boldsymbol{\beta}$ when $p = 0.50, 0.75$ and $p = 0.95$. It can be observed that our estimates are very close to the true values and our credible intervals are generally shorter than the intervals given by the BTQ. Although our intervals are shorter, but its more informative than the intervals given by the BTQ especially for the most extreme quantile ($p = 0.95$). For example, when the error is standard normal and estimating the 0.95 quantiles for β_0, our intervals capture the true parameter value 100% of the time, while BTQ intervals capture the true parameter value 98% of the time, suggesting a good performance of our proposed method. It can be seen clearly from Table 1 and Table 2, the proposed Bayesian method outperformed the BTQ method. It yields considerably lower biases, high relative efficiency and much more precise credible interval.

For convergence diagnosis, across the three error distributions, we run our MCMC algorithm with 3 different sets of initial values each for 17,000 iterations discarding the first 2,000 gave values of the Gelman-Rubin (Gelman and Rubin, 1992) diagnostic of between 1.00 and 1.01. The posterior mean for each parameter was virtually identical for each chain, suggestting that the constructed chains are stationary and the mixing is quite good. The posterior histograms of parameters β_0, β_1 and β_2 in Figure 1 also support this conclusion.

4.2. Model selection

4.2.1. Example 2 (Subset selection for left-censored response with $k < n$)

In this example, we consider two simulation studies. We simulate 8 covariates from a multivariate normal with pairwise covariance between any two predictors x_h and x_l being $\rho^{|h-l|}$, for $\rho = 0.5$. For both simulations, we generated 250 datasets each with n = 200

Table 2: True parameter values, posterior means and 95% credible intervals (95% CrI) for Example 1. The results are averaged over 250 independent simulations. %in is the number of times each 95% credible interval capture the true value out of 250 replications.

Error	Model	p	β_0 Mean (95% CrI)	%in	β_1 Mean (95% CrI)	%in	β_2 Mean (95% CrI)	%in
$\varepsilon \sim N(0,1)$								
	$\boldsymbol{\beta}^{true}$	0.50	1.000		1.000		1.000	
	BTQ	0.50	1.013 (0.713, 1.342)	100	0.983 (0.662, 1.349)	100	1.000 (0.528, 1.399)	100
	BTQRg	0.50	1.001 (0.775, 1.215)	100	0.990(0.775, 1.214)	100	1.006 (0.757, 1.236)	100
	$\boldsymbol{\beta}^{true}$	0.75	1.675		1.000		1.000	
	BTQ	0.75	1.592 (1.511, 2.035)	100	0.982 (0.719, 1.314)	100	0.995 (0.549, 1.386)	100
	BTQRg	0.75	1.642 (1.450, 1.952)	100	0.991 (0.748, 1.253)	100	1.031 (0.779, 1.211)	100
	$\boldsymbol{\beta}^{true}$	0.95	2.645		1.000		1.000	
	BTQ	0.95	3.050 (2.145, 3.552)	98	1.081 (0.624, 1.335)	100	1.122 (0.499, 1.791)	100
	BTQRg	0.95	2.650 (2.327, 3.312)	100	1.024 (0.516, 1.492)	100	1.090 (0.544, 1.486)	100
$\varepsilon \sim (1+x_2)N(0,1)$								
	$\boldsymbol{\beta}^{true}$	0.50	1.000		1.000		1.000	
	BTQ	0.50	0.939 (0.689, 1.335)	100	1.069 (0.974, 1.250)	100	1.040 (0.689, 1.448)	100
	BTQRg	0.50	0.989 (0.803, 1.248)	100	1.019 (0.812, 1.181)	100	1.003 (0.792, 1.263)	100
	$\boldsymbol{\beta}^{true}$	0.75	1.675		1.000		1.675	
	BTQ	0.75	1.648 (1.442, 1.891)		1.012 (0.635, 1.432)		1.495 (0.995, 1.873)	
	BTQRg	0.75	1.681 (1.479, 1.984)		1.010 (0.743, 1.226)		1.561 (1.175, 1.711)	
	$\boldsymbol{\beta}^{true}$	0.95	2.645		1.000		2.645	
	BTQ	0.95	3.333 (2.457, 3.678)	96	1.362 (0.519, 1.771)	100	2.122 (1.153, 2.351)	98
	BTQRg	0.95	2.779 (2.659, 3.875)	100	1.019 (0.481, 1.695)	100	2.155 (1.264, 2.161)	100
$\varepsilon \sim 0.75N(0,1)+0.25N(0,4)$								
	$\boldsymbol{\beta}^{true}$	0.50	1.000		1.000		1.000	
	BTQ	0.50	1.003 (0.661,1.457)	98	1.006 (0.721, 1.425)	100	1.005 (0.671, 1.372)	100
	BTQRg	0.50	0.999 (0.786 1.207)	100	1.011 (0.783, 1.206)	100	1.001 (0.767, 1.224)	100
	$\boldsymbol{\beta}^{true}$	0.75	1.843		1.000		1.000	
	BTQ	0.75	(1.153, 2.001)	94	(0.636,1.155)	99	(0.711, 1.356)	99
	BTQRg	0.75	(1.396, 1.874)	99	(0.782, 1.259)	100	(0.748, 1.256)	100
	$\boldsymbol{\beta}^{true}$	0.95	3.056		1.000		1.000	
	BTQ	0.95	2.918 (1.992, 3.451)	91	1.086 (0.571, 1.749)	98	1.150 (0.421, 1.581)	98
	BTQRg	0.95	2.974 (2.191, 3.129)	98	1.011 (0.522, 1.443)	100	1.020 (0.579, 1.416)	100
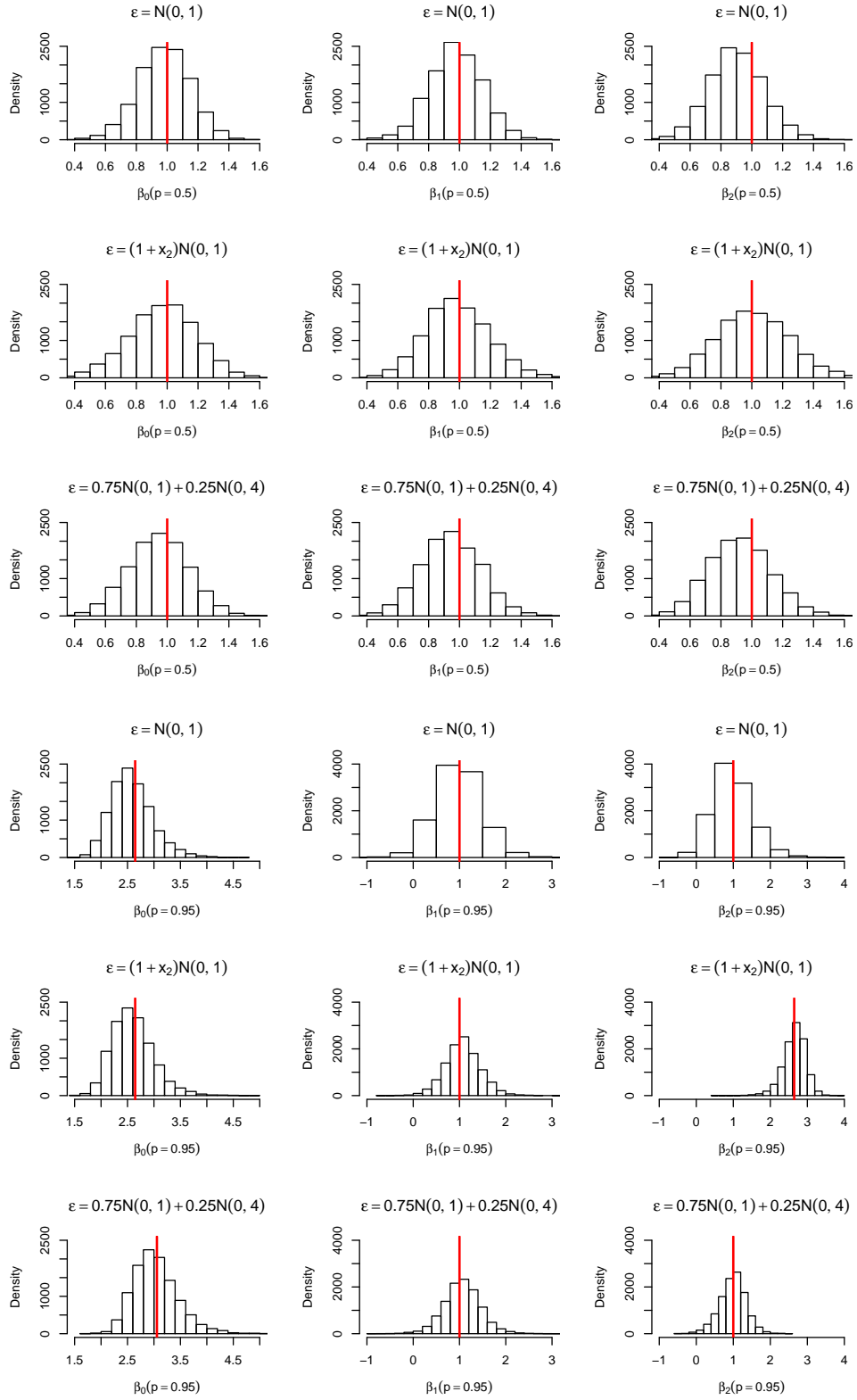
Figure 1: Posterior histograms of β_0, β_1 and β_2 at quantiles 0.50 and 0.95 for Example 1 using the proposed method.

observations from the true censoring model

$$y_i = \max\{0, y_i^*\}, \quad i = 1, \cdots, n,$$

$$y_i^* = \boldsymbol{x}_i' \boldsymbol{\beta} + \varepsilon_i. \tag{16}$$

For the first simulation, we set $\boldsymbol{\beta} = (1, 5, 0, 0, 0, 0, 0, 0, 0)'$, while for the second simulation we set $\boldsymbol{\beta} = (1, 3, 1.5, 0, 0, 2, 0, 0, 0)'$, so that the first true regression coefficient refers to the intercept value. Three error distributions are considered: N(0,1) distribution, t_3 distribution and χ_3^2 distribution. In this example, we compare our Bayesian variable selection approach in Tobit QReg using g prior (BTQRg) and Bayesian variable selection in Tobit QReg approach (BVST) using a symmetric prior distribution as reported in Ji et al. (2012). The results of the standard Tobit QReg approach (crq) are also reported. Methods are evaluated based on the median of mean absolute deviations (MMAD), that is, median($1/200 \sum_{i=1}^{200} |\boldsymbol{x}_i' \hat{\boldsymbol{\beta}} - \boldsymbol{x}_i' \boldsymbol{\beta}^{true}|$), where the median is taken over the 250 simulations.

The results of the MMADs and the standard deviations of the MMADs are listed in Table 3. For the MMADs and the standard deviations (sd) criteria, the proposed method (BTQRg) generally performs better than BVST and crq for all the distributions under consideration. In addition, the average number of correct zero coefficients selected by our approach is greater than the average number of correct zero coefficients selected by BVST for all the distributions under consideration. It can be seen that as the quantiles become more extreme, the BVST approach yields a low average of correct zero coefficients compared with the proposed approach, suggesting a good performance of the proposed algorithm.

Instead of looking at the MMADs, the standard deviations and the average number of correct zero coefficients, we may also look at the top model picked out by the Bayesian methods. From Table 4, we see that both methods pick the correct model. However, we can observe that our method tends to perform better in terms of posterior model probability for

the correct model, than the BVST especially for the extreme quantile ($p = 0.95$). Hence our prior plays a crucial role to correctly identify the correct model with higher posterior model probability, even for the extreme quantiles.

4.2.2. Example 3 ($k > n$)

The setup in this example is the same as the Example 2, except we increase the number of variables to $k = 250$ and choose $\boldsymbol{\beta}$ such as $\boldsymbol{\beta} = (1, 3, 0, 0, 0, 0, 0, 3, 0, 0, 0, 0, 0, 3, \underbrace{0, 0, ..., 0}_{236})$. This allows us to investigate the performance in the case $k > n$.

From Table 5, the performance of our method appears quite good compared to the BVST. We observe that the median of mean absolute deviations (MMAD) produced using the BVST method is much higher than our method. We also see that our method tends to produce lower standard deviations than BVST.

4.2.3. Example 4 ($k > n$)

The setup in this example is the same as Example 3 but we ignore the link function (i.e., continuous response) and we set $n = 50$, $k = 100$ and $\boldsymbol{\beta} = (1, 3, 0, 0, 0, 0, 0, 3, 0, 0, 0, 0, 0, 3, \underbrace{0, 0, ..., 0}_{86})$. This setup allows us to investigate the performance of the proposed method in the quantile regression model with the case $k > n$. In this example, one simulation dataset is used to compare our approach reported in subsection 3.2.1 with the stochastic search variable selection reported in Reed et al. (2009) using the SSVSquantreg function in the R package MCMCpack (Martin et al., 2011). The methods were compared using marginal inclusion probabilities (MIP). In this example, we considered two quantiles, these were 0.50 and 0.95. We ran both algorithms for 17,000 iterations, discarding the first 2,000 as burn in. The results of the marginal inclusion probabilities are summarized in Table 6. Clearly, we can see that our approach tends to perform better in terms of the estimated marginal inclusion probabilities than SSVSquantreg, especially for $p = 0.95$.

Table 3: MMADs, standard deviation of MADs (SD) and the average number of correct zero coefficients (correct) for the simulated data in Example 2, where $p = 0.50$, 0.75 and 0.95.

Model	p	$\varepsilon \sim N(0,1)$ MMAD (SD)	correct	$\varepsilon \sim t_3$ MMAD (SD)	correct	$\varepsilon \sim \chi_3^2$ MMAD (SD)	correct
Sim. 1							
crq	0.50	0.327 (0.112)	-	0.336 (0.191)	-	0.415 (0.289)	-
BVST	0.50	0.158 (0.083)	6.88	0.177 (0.105)	6.84	0.552 (0.386)	6.96
BTQRg	0.50	0.094 (0.082)	6.96	0.142 (0.094)	6.88	0.347 (0.264)	7.00
crq	0.75	0.311 (0.119)	-	0.363 (0.157)	-	0.444 (0.287)	-
BVST	0.75	0.267 (0.254)	6.84	0.322 (0.318)	6.80	0.357 (0.360)	6.68
BTQRg	0.75	0.146 (0.112)	7.00	0.184 (0.162)	6.96	0.297 (0.271)	6.92
crq	0.95	0.366 (0.189)	-	0.404 (0.280)	-	0.450 (0.520)	-
BVST	0.95	0.383 (0.318)	6.88	0.359 (0.393)	6.60	0.386 (0.388)	6.20
BTQRg	0.95	0.223 (0.168)	7.00	0.223 (0.217)	6.92	0.343 (0.290)	6.89
Sim. 2							
crq	0.50	0.296 (0.095)	-	0.343 (0.120)	-	0.408 (0.271)	-
BVST	0.50	0.159 (0.084)	4.84	0.315 (0.301)	4.64	0.371 (0.357)	4.46
BTQRg	0.50	0.149 (0.080)	4.96	0.221 (0.121)	4.84	0.296 (0.253)	4.99
crq	0.75	0.294 (0.092)	-	0.350 (0.121)	-	0.286 (0.263)	-
BVST	0.75	0.283 (0.211)	4.68	0.320 (0.299)	4.72	0.381 (0.338)	4.66
BTQRg	0.75	0.172 (0.081)	4.96	0.218 (0.111)	4.92	0.338 (0.257)	4.96
crq	0.95	0.345 (0.132)	-	0.386 (0.293)	-	0.477 (0.435)	-
BVST	0.95	0.384 (0.311)	4.76	0.390 (0.374)	4.66	0.437 (0.425)	4.36
BTQRg	0.95	0.234 (0.175)	5.00	0.258 (0.291)	4.96	0.381 (0.384)	4.96

Table 4: Top two models in Example 2 for Tobit QReg.

Sim.	Model	p	$\varepsilon \sim N(0,1)$ Variables	Pb	$\varepsilon \sim t_3$ Variables	Pb	$\varepsilon \sim \chi_3^2$ Variables	Pb
1	BVST	0.50	Inter., x_1	0.90	Inter., x_1	0.87	Inter., x_1	0.85
	BTQRg	0.50	Inter., x_1	0.97	Inter., x_1	0.96	Inter., x_1	0.91
	BVST	0.75	Inter., x_1	0.85	Inter., x_1	0.84	Inter., x_1	0.74
	BTQRg	0.75	Inter., x_1	0.97	Inter., x_1	0.96	Inter., x_1	0.90
	BVST	0.95	Inter., x_1	0.82	Inter., x_1	0.76	Inter., x_1	0.60
	BTQRg	0.95	Inter., x_1	0.97	Inter., x_1	0.96	Inter., x_1	0.87
2	BVST	0.50	Inter., $x_1, x2, x5$	0.89	Inter., $x_1, x2, x5$	0.89	Inter., $x_1, x2, x5$	0.86
	BTQRg	0.50	Inter., $x_1, x2, x5$	0.93	Inter., $x_1, x2, x5$	0.93	Inter., $x_1, x2, x5$	0.91
	BVST	0.75	Inter., $x_1, x2, x5$	0.85	Inter., $x_1, x2, x5$	0.88	Inter., $x_1, x2, x5$	0.85
	BTQRg	0.75	Inter., $x_1, x2, x5$	0.92	Inter., $x_1, x2, x5$	0.93	Inter., $x_1, x2, x5$	0.91
	BVST	0.95	Inter., $x_1, x2, x5$	0.84	Inter., $x_1, x2, x5$	0.81	Inter., $x_1, x2, x5$	0.52
	BTQRg	0.95	Inter., $x_1, x2, x5$	0.93	Inter., $x_1, x2, x5$	0.93	Inter., $x_1, x2, x5$	0.91

Table 5: MMADs and SD for the simulated data in Example 3, where $p = 0.50, 0.75$ and 0.95.

Model	p	$\varepsilon \sim N(0,1)$ MMAD (SD)	$\varepsilon \sim t_3$ MMAD (SD)	$\varepsilon \sim \chi_3^2$ MMAD (SD)
BVST	0.50	0.192 (0.089)	0.323 (0.235)	0.473 (0.358)
BTQRg	0.50	0.183 (0.093)	0.211 (0.155)	0.337 (0.294)
BVST	0.75	0.325 (0.339)	0.361 (0.289)	0.552 (0.403)
BTQRg	0.75	0.258 (0.278)	0.328 (0.293)	0.369 (0.413)
BVST	0.95	0.529 (0.363)	0.631 (0.346)	0.670 (0.619)
BTQRg	0.95	0.497 (0.321)	0.589 (0.334)	0.611(0.680)

Table 6: Marginal inclusion probabilities to the true model for simulated data in Example 1 using three different error distributions. Our approach for Bayesian quantile regression (BQRg) is compared with SSVSquantreg.

			Variables			
Error distribution	Method	p	Intercept	x_1	x_7	x_{13}
$\varepsilon \sim N(0,1)$						
	SSVSquantreg	0.50	0.72	1.00	1.00	1.00
	BQRg	0.50	1.00	1.00	1.00	1.00
$\varepsilon \sim t_3$						
	SSVSquantreg	0.50	0.65	1.00	1.00	1.00
	BQRg	0.50	1.00	1.00	1.00	1.00
$\varepsilon \sim \chi_3^2$						
	SSVSquantreg	0.50	0.80	1.00	1.00	1.00
	BQRg	0.50	1.00	1.00	1.00	1.00
$\varepsilon \sim N(0,1)$						
	SSVSquantreg	0.95	1.00	0.11	0.71	0.18
	BQRg	0.95	1.00	1.00	1.00	1.00
$\varepsilon \sim t_3$						
	SSVSquantreg	0.95	1.00	0.09	0.71	0.22
	BQRg	0.95	1.00	1.00	1.00	1.00
$\varepsilon \sim \chi_3^2$						
	SSVSquantreg	0.95	1.00	0.11	0.28	0.12
	BQRg	0.95	1.00	1.00	1.00	1.00

5. Leukemia data set

We illustrate the performance of the proposed model for simultaneous gene selection and estimation on the leukemia dataset reported in Golub et al. (1999). The dataset describe the expression levels of 7129 human genes in 72 acute leukemia mRNA samples with either acute myeloid leukemia (AML, 25 cases) or acute lymphoblastic leukemia (ALL, 47 cases). The data further splits into a training dataset and a testing dataset. The training dataset contains 38 samples of which 27 are ALL and the remaining 11 are AML, while testing dataset consists of 34 samples 20 are ALL and the remaining 14 are AML (Golub et al., 1999).

Table (7) list the top 10 most significant genes selected by the proposed model for $p \in$

$\{0.25, 0.50, 0.75\}$, together with the relevant genes selected by Golub et al. (1999), Lee et al. (2003) and Yang and Song (2010) when $p = 0.50$ which corresponds to the center of the distribution. The leading gene is Zyxin. Zyxin is a protein that is in humans and is encoded by the ZYX gene (Maccalma, 1996). This gene also one of the strong genes in Golub et al. (1999), Lee et al. (2003), Bae and Mallick (2004), and Yang and Song (2010), among others. Bae and Mallick (2004) used only Zyxin for prediction and had only 3 misclassifications, while Golub et al. (1999) used 50 genes for prediction and had 5 misclassifications. This ensure that Zyxin plays an important role in classification. A more complete picture of gene effects can be provided by the first and third quartiles. We note that the leading gene is Macmarcks when $p = 0.25$, while the leading gene is CST3 Cystatin C (amyloid angiopathy and cerebral hemorrhage) when $p = 0.75$. We used the top five selected genes for each $p \in \{0.25, 0.50, 0.75\}$ to perform predictions on the test data. The proposed model misclassified only one observation when $p = 0.50$ and 0.75 (the observation number 4 and the observation number 30, respectively) and two observations when $p = 0.25$ (the observations number 4 and 29). The classification results indicate that the proposed model perform well. In addition, since we use only five genes, our prediction results appear to improve predictions done by Golub et al. (1999) who used 50 genes and had 5 misclassifications and also our results give a more complete picture of the underlying distribution.

6. Discussion

This paper has introduced a new method for Bayesian model selection of Tobit QReg. Our approach relies on a modification of Zellners informative g-prior in QReg reported in Alhamzawi and Yu (2013) to take into account different priors for different quantile levels. The proposed prior is first developed for settings in which $p < n$, and then extended to deal with multicollinearity and over-fitting problems by defining a ridge parameter in the prior construction. We have also discussed some extensions of our method, including the continuous and binary responses in QReg. The analysis of simulation studies and microarray study shows strong support for the use of the proposed method to inference for quantile regression models. The proposed method generally behaves much better than the other approaches in terms of the absolute bias and relative efficiency. Clear advantages over an earlier approach proposed by Reed et al. (2009) and Ji et al. (2012) include quantile dependent priors and efficiency of posterior computation. The advantage of the method is that the prior distribution changes automatically when we change the quantile. Thus, we have different priors for different quantiles. In particular, the prior distribution and efficient Bayesian computation represent a useful alternative to methods that rely on symmetric priors for regression coefficients, which are reported by (Kozumi and Kobayashi, 2011; Reed et al., 2009; Ji et al., 2012, among others). The work considered in this paper opens the door to new research directions for subset selection and coefficient estimation in QReg models by using the modified g-prior. For example, the approach can be extended to the Bayesian QReg models with right-censored or interval censored responses. There are, also, many other possible extensions such as using the modified g-prior in Bayesian single index QReg or Bayesian nonparametric QReg.

Table 7: The top 10 significant genes selected by the proposed model.

p	Rank	Index	Gene description
0.25	1	804	Macmarcks
	2	1685	Termianl transferase mRNA
	3	3847	HoxA9 mRNA
	4	2354	CCND3 Cyclin D3
	5	1779	MPO Myeloperoxidase
	6	4847	Zyxin
	7	2402	Azurocidin gene
	8	760	CYSTATIN A
	9	1882	CST3 Cystatin C (amyloid angiopathy and cerebral hemorrhage)
	10	6041	APLP2 Amyloid beta (A4) precursor-like protein 2
0.50	1	4847	Zyxin [a,b,c]
	2	760	CYSTATIN A[b,c]
	3	804	Macmarcks [a,b,c]
	4	4052	Catalase (EC 1.11.1.6) 5′flank and exon 1 mapping to chromosome 11, band p13 [a,c]
	7	1882	CST3 Cystatin C (amyloid angiopathy and cerebral hemorrhage)[a,b,c]
	6	1144	SPTAN1 Spectrin, alpha, non-erythrocytic 1 (alpha-fodrin)[b]
	5	1745	LYN V-yes-1 Yamaguchi sarcoma viral related oncogene homolog
	8	1630	Inducible protein mRNA[b]
	9	2288	DF D component of complement (adipsin)[b]
	10	1953	Fc-epsilon-receptor gamma-chain mRNA
0.75	1	1882	CST3 Cystatin C (amyloid angiopathy and cerebral hemorrhage)
	2	4377	ME491 gene extracted from H.sapiens gene for Me491/CD63 antigen
	3	1834	CD33 CD33 antigen (differentiation antigen)
	4	760	CYSTATIN A
	5	4336	ARHG Ras homolog gene family, member G (rho G)
	6	4847	Zyxin
	7	6041	APLP2 Amyloid beta (A4) precursor-like protein 2
	8	3847	HoxA9 mRNA
	9	1953	Fc-epsilon-receptor gamma-chain mRNA
	10	4328	PROTEASOME IOTA CHAIN

[a] Golub et al. (1999), [b] Lee (2009), [c] Yang and Song (2010)

Bibliography

Agliari, A. and C. C. Parisetti (1988). A-g reference informative prior: a note on Zeliner's g-prior. *The Statistician 37*, 271–275.

Alhamzawi, R. and K. Yu (2013). Conjugate priors and variable selection for Bayesian quantile regression. *Computational Statistics & Data Analysis 64*, 209–219.

Bae, K. and B. Mallick (2004). Gene selection using a two-level hierarchical Bayesian model. *Bioinformatics 20*, 3423–3430.

Baragatti, M. C. and D. Pommeret (2012). A study of variable selection using g-prior distribution with ridge parameter. *Computational Statistics & Data Analysis 56*, 1920–1934.

Barndorff-Nielsen, O. E. and N. Shephard (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, Series B 63*, 167–241.

Benoit, D. F. and D. V. D. Poel (2011). Binary quantile regression: a Bayesian approach based on the asymmetric Laplace distribution. *Journal of Applied Econometrics 27*, 1174–1188.

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

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

Celeux, G., M. El-Anbari, J. Marin, and C. Robert (2012). Regularization in regression: comparing Bayesian and frequentist methods in a poorly informative situation. *Bayesian Analysis 7*, 477–502.

Chen, R.-B., C.-H. Chu, T.-Y. Lai, and Y. N. Wu (2011). Stochastic matching pursuit for Bayesian variable selection. *Statistics and Computing 21*, 247–259.

Clyde, M. and E. I. George (2004). Model uncertainty. *Statistical Science 19*, 81–94.

Cui, W. and E. I. George (2008). Empirical Bayes vs. fully Bayes variable selection. *Journal of Statistical Planning and Inference 138*, 888–900.

Gelman, A. and D. B. Rubin (1992). Inference from iterative simulations using multiple sequences. *Statistical Science 7*, 457–511.

George, E. I. and R. E. McCulloch (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association 88*, 881–889.

George, E. I. and R. E. McCulloch (1997). Approaches for Bayesian variable selection. *Statistica Sinica 7*, 339–373.

Golub, T. R., D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caliguri, C. D. Bloomfield, and E. S. Lander (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science 286*, 531–537.

Gupta, M. and J. G. Ibrahim (2007). Variable selection in regression mixture modeling for the discovery of gene regulatory networks. *Journal of the American Statistical Association 102*, 867–880.

Gupta, M., P. Qu, and J. Ibrahim (2007). A temporal hidden markov regression model for the analysis of gene regulatory networks. *Biostatistics 8*, 805–820.

Hahn, J. (1995). Bootstrapping quantile regression estimators. *Econometric Theory 11*, 105–121.

Ibrahim, J. G. and M.-H. Chen (2000). Power prior distributions for regression models. *Statistical Science 15*, 46–60.

Ji, Y., N. Lin, and B. Zhang (2012). Model selection in binary and Tobit quantile regression using the Gibbs sampler. *Computational Statistics & Data Analysis 56*, 827–839.

Kass, R. E. and L. Wasserman (1995). A reference Bayesian test for nested hypothesis and its relationship to the Schwarz criterion. *Journal of the American Statistical Association 90*, 928–934.

Koenker, R. (2011). `quantreg`: Quantile regression. R package version 4.71.

Koenker, R. and G. J. Bassett (1978). Regression quantiles. *Econometrica 46*, 33–50.

Kordas, G. (2006). Smoothed binary regression quantiles. *Journal of Applied Econometrics 21*, 387–407.

Kotz, S., K. T. J. and Podgrski (2001). *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and.* Financ, Birkhauser, Boston, Mass, USA.

Kozumi, H. and G. Kobayashi (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation 81*, 1565–1578.

Lee, D. J. (2009). Testing parameter stability in quantile models: An application to the U.S. inflation process. Working papers 2009-26, University of Connecticut, Department of Economics.

Lee, K., N. Sha, E. R. Dougherty, M. Vannucci, and B. K. Mallik (2003). Gene selection: a Bayesian variable selection approach. *Bioinformatics. 19*, 90–97.

Liang, F., R. Paulo, G. Molina, M. A. Clyde, and J. O. Berger (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association 103*, 410–423.

Maccalma, T. e. a. (1996). Molecular characterization of human zyxin. *Biol. Chem. 271*, 31470–31478.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics 3*, 205–228.

Manski, C. F. (1985). Semiparametric analysis of discrete response: asymptotic properties ofthe maximum score estimator. *Journal of Econometrics 27*, 313–333.

Martin, A., K. Quinn, and J. Park (2011). `MCMCpack`: Markov chain Monte Carlo. R package version 1.0-10.

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

Reed, C., D. Dunson, and K. Yu (2009). Bayesian variable selection in quantile regression. Technical report, Brunel University, Department of Mathematical Sciences.

Smith, M. and R. Kohn (1996). Nonparametric regression using Bayesian variable selection. *Journal of Econometrics 75*, 317–343.

Wang, H. J. and M. Fygenson (2009). Inference for censored quantile regression models in longitudinal studies. *The Annals of Statistics 37*, 756–781.

Yang, A. J. and X. Y. Song (2010). Bayesian variable selection for disease classification using gene expression data. *Bioinformatics 26*, 215–222.

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

Yu, K. and J. Stander (2007). Bayesian analysis of a Tobit quantile regression model. *Journal of Econometrics 137*, 260–276.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In Bayesian inference and decision techniques. *Stud. Bayesian Econometrics Statist. 6*, 233–243.

Zellner, A. and A. Siow (1980). Posterior odds ratios for selected regression hypotheses. *In Bayesian Statistics. Proceedings of the First Valencia International Meeting Held in Valencia (Spain) (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 585-603. Valencia Univ. Press*.