$\mathcal{AP}$
ijpam.eu

BAYESIAN QUANTILE REGRESSION WITH SCALE MIXTURE OF UNIFORM PRIOR DISTRIBUTIONS

Fadel Hamid Hadi Alhusseini

Department of Statistics and Economic Informatics
University of Craiova
ROMANIA

Abstract: In this paper, we propose a new Bayesian Lasso quantile regression method for variable selection assigning independent scale-mixture of uniform distributions for the regression coefficients. Then, a simple and efficient MCMC algorithm was presented for Bayesian sampler. Simulation studies and a real data set are used to investigate the performance of the proposed method compared to some other existing methods. Both simulated and real data examples show that the proposed method performs quite good compared to the other methods under a variety of scenarios.

1. Introduction

Quantile regression (QReg) models have become very common since the pioneering research of [16]. It have been employed in many different fields: Microarray study [29], agricultural economics [19], ecological studies [10], body mass index [32, 33], growth chart [30], and so on. Compared to the standard mean regression models, QReg models belong to a robust family [17]. Unlike standard mean regression, QReg does not need any assumption about the residual distribution providing greater statistical efficiency than standard mean regression when the error is non-normal.

The linear QReg model assumes that the outcome of interest y_i can be written as

$$y_i = \boldsymbol{x}_i' \boldsymbol{\beta}_\tau + \varepsilon_i, \qquad \tau \in (0, 1), \tag{1}$$

where $\boldsymbol{x}_i'$ is a $1 \times k$ vector of explanatory variables, $\boldsymbol{\beta}_\tau$ is a $k \times 1$ vector of unknown quantities and τ is the quantile level. Here, ε_i is the residual term whose density is restricted to have the τth quantile equal to 0. Similar to the standard mean regression, QReg aims at evaluating the conditional quantiles of the outcome of interest y_i given a explanatory vector $\boldsymbol{x}_i$. It can be proved [16] that the QReg coefficients of $\boldsymbol{\beta}_\tau$ can be estimated by

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{x}_i' \boldsymbol{\beta}_\tau), \tag{2}$$

where $\rho_\tau(s)$ is the check function defined by $\rho_\tau(s) = s\{\tau - I(\tau < 0)\}$. For simplicity of notation, we will omit τ from the notation $\boldsymbol{\beta}_\tau$ in the remainder of this paper.

One significant issue in building a regression model is the selection of the active regressors. The selection process aims to increase the prediction accuracy and to get high interpretation [5]. Nowadays, there has been considerable attention on sparse methods that includes all regressors in the model and uses informative priors to shrink inactive regression coefficients toward zero. See, for example, Lasso [27], the adaptive Lasso [36], dantzig selector [11], matrix completion [12], compressive sensing [7], Lasso QReg [22] and adaptive Lasso QReg [31]. A comprehensive account of these and other recent methods can be found in [28]. Similarly, from a Bayesian framework, [25] proposed Bayesian Lasso for linear regression models by specifying scale mixture of normal (SMN) prior distributions on the regression coefficients and independent exponential prior distributions on their variances. [26] developed Bayesian adaptive Lasso by allowing different shrinkage parameters for different coefficients. Based on the later approaches [21] suggested Bayesian Lasso QReg and [5] proposed Bayesian adaptive Lasso QReg. Some further extensions of the Lasso QReg models have also suggested by [8], [4] and [2], among others. Compared to the frequentist counterparts, the Bayesian models usually offer a valid measure of standard error based on a MCMC. It also offers a convenient method of incorporating regression coefficients uncertainty into predictive inferences. Moreover, the Bayesian formulation offers a flexible way of estimating the penalty parameter along with regression coeffecients.

Our objective is to develop a Bayesian formulation for regularization in linear QReg. Recently, for linear regression, [24] provided a different approach

of Lasso-based model by employing the scale mixture of uniform formulation of the Laplace density. The performance of this method was illustrated via simulation studies and a real dataset. [24] show that the new method performs quite well compared to some other existing methods. In this paper, we propose a new formulation for Bayesian lasso QReg by employing the scale mixture of uniform formulation. Then we develop a fully Bayesian treatment that leads to a simple and efficient Gibbs sampling algorithm with tractable full conditional posterior distributions.

The rest of this paper is organized as follows. In Section 2, we briefly review the Bayeian QReg method and propose a hierarchical model based on the scale mixture of uniform distribution. The Gibbs sampler is introduced in Section 3. We illustrate the performance of the proposed method by simulation studies in Section 4 followed by a real data example in Section 5. In Section 6, we conclude the paper.

2. Methods

2.1. Bayesian QReg

Within the Bayesian QReg formulation, a popular choice of the error distribution has been skewed Laplace distribution (SLD); see [34]. The probability density function (pdf) of a SLD is

$$f(\varepsilon_i|\sigma) = \tau(1-\tau)\sigma \exp\{-\sigma\rho_\tau(\varepsilon_i)\}, \quad \varepsilon_i \in \mathbb{R}, \tag{3}$$

where σ is the scale parameter and τ determines the quantile level. The joint distribution of $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)'$ is

$$f(\boldsymbol{\varepsilon}|\sigma) = \tau^n(1-\tau)^n\sigma^n \exp\{-\sum_{i=1}^n \sigma\rho_\tau(\varepsilon_i)\}. \tag{4}$$

Following [18] the check loss function (2) is closely equivalent to (4). In particular, maximizing (4) is equivalent to minimizing (2). The relationship between (2) and (4) can be employed to represent the QReg method in the likelihood framework. Recently [20] provided a Bayesian approach for QReg by reformulated the SLD as SMNs family. More specifically, let $\varepsilon_i \sim \mathrm{N}((1-2\tau)m_i, 2\sigma^{-1}m_i)$ then the SLD for ε_i arises when $m_i \sim Exp(\tau(1-\tau)\sigma)$. Under this motivation, [20] proposed a geometrically ergodic MCMC relying on conditional conjugacy to draw the parameters from the posteriors. Rewriting the SLD for the errors as

a SMN and introducing the exponential mixing density result in the following hierarchy:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + (1 - 2\tau)m_i + \sqrt{2\sigma^{-1}m_i}\epsilon_i,$$
$$m_i|\sigma \sim \tau(1 - \tau)\sigma \exp\{-\tau(1 - \tau)\sigma m_i\}, \tag{5}$$
$$\epsilon_i \sim \frac{1}{\sqrt{2\pi}}\exp\{-\frac{1}{2}\epsilon_i^2\}.$$

Under the above hierarchical formulation, the posterior distribution of interest $p(\boldsymbol{\beta}_\tau|\sigma, m_1, \cdots, m_n)$ is a multiverate normal.

2.2. Bayesian QReg with the Lasso Penalty

The Lasso QReg [22] is given by

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}) + \lambda||\boldsymbol{\beta}||, \tag{6}$$

where λ ($\lambda \geq 0$) is the shrankage parameter. [21] proposed the Bayesian Lasso for linear QReg model by putting a Laplace prior takes the form $p(\beta_j|\lambda, \sigma) = (\sigma\lambda/2)\exp\{-\sigma\lambda|\beta_j|\}$ on the jth QReg coefficient. More specifically, they putting a scale mixture of normal prior distributions on $\boldsymbol{\beta}_\tau$ and exponential prior distributions for the variance parameters assuming they are independent. In this paper, we put a Laplace prior on β_j takes the form $p(\beta_j|\lambda) = \lambda/2\exp\{-\lambda|\beta_j|\}$ and develope an alternative hierarchical Bayesian model of the Lasso model.

Following [24] and [3], the Laplace prior on β_j can be written as:

$$\frac{\lambda}{2}\exp\{-\lambda|\beta_j|\} \quad = \frac{\lambda}{2}\int_{u_j > |\beta_j|} \lambda \exp\{-\lambda u_j\}du_j,$$
$$= \int_{-u_j < \beta_j < u_j} \frac{1}{2u_j}\frac{\lambda^2}{\Gamma(2)}u_j^{2-1}\exp\{-\lambda u_j\}du_j, \tag{7}$$

where u_j is a mixing variable. We further put Gamma priors on σ and λ. Using (5) and (7), our Bayesian hierarchical model can be formulated as follows:

$$y_i|\boldsymbol{\beta}_\tau, \sigma, m_i \sim \mathrm{N}(\boldsymbol{x}_i'\boldsymbol{\beta}_\tau + (1 - 2\tau)m_i, 2\sigma^{-1}m_i),$$
$$m_i|\sigma \sim \mathrm{Exp}(\tau(1 - \tau)\sigma),$$
$$\beta_j|u_j \sim \mathrm{Uniform}(-u_j, u_j), \tag{8}$$
$$u_j|\lambda \sim \mathrm{Gamma}(2, \lambda),$$

$$\sigma \sim \sigma^{a-1} \exp(-b\sigma),$$
$$\lambda \sim \lambda^{c-1} \exp(-d\lambda).$$

Here, $\text{Exp}(\tau(1-\tau)\sigma)$ refers to the exponential distribution with rate parameter $\tau(1-\tau)\sigma$.

3. Posterior Inference

The posterior distribution of $\boldsymbol{\beta}, \sigma, \boldsymbol{m} = (m_1, ..., m_n)'$, $\boldsymbol{u} = (u_1, ..., u_p)'$ and λ can be updated using an efficient MCMC-based computation technique.

- Updating m_i^{-1}: The full conditional distribution (FCD) of m_i^{-1} is inverse Gaussian distribution " refered to as $\text{InvGa}(\mu_i', \varphi')$", where $\varphi' = \sigma/2$ and $\mu_i' = 1/\sqrt{(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2}$. Here, the pdf of the InvGa is given by [13]:

$$f(x|\mu_i', \varphi') = \sqrt{\frac{\varphi'}{2\pi}} x^{-3/2} \exp\{\frac{-\varphi'(x - \mu_i')^2}{2(\mu_i')^2 x}\}, \quad x, \varphi', \mu_i' > 0. \qquad (9)$$

- Updating β_j: The full conditional distribution of β_j is truncated normal with mean $\bar{\beta}_j$ and variance $S_{\beta_j}^2$ where

$$\bar{\beta}_j = \left(S_{\beta_j}^2 \sum_{i=1}^n \frac{\sigma x_{ij}(y_i - (1-2\tau)m_i - \sum_{j \neq l} x_{il}\beta_l)}{2m_i}\right) I\{|\beta_j| < u_j\} \qquad \text{and}$$

$$S_{\beta_j}^2 = \left(\sum_{i=1}^n \frac{\sigma x_{ij}^2}{2m_i}\right)^{-1}.$$

- Updating u_j: The full conditional distribution of u_j is a left-truncated exponential distribution given by

$$u_j|\boldsymbol{\beta}, \lambda \sim \text{Exp}(\lambda)I\{u_j > |\beta_j|\}$$

Updating u_j can be done by using inversion method [24] as follows:

1. Update u_j^* from $\text{Exp}(\lambda)$,
2. Set $u_j = u_j^* + |\beta_j|$

- Updating σ: The full conditional distribution of σ is gamma with shape parameter $a+3n/2$ and rate parameter $\sum_{i=1}^n \left((y_i-(1-2\tau)m_i-\boldsymbol{x}_i'\boldsymbol{\beta})^2/(2m_i)+ \tau(1-\tau)m_i\right) + b$.

- Updating λ: The full conditional distribution of λ is Gamma($c + 2p, d + \sum_{j=1}^{p} |\beta_j|$).

4. Simulation Studies

In this section, the performance of the proposed model is investigated by simulations. The proposed model is compared with Bayesian Lasso quantile regression reported in [21]. The results of the classical frequentist approache using the R function `rq()` in the R package `quantreg` and the results of Bayesian quantile regression using the R function `MCMCquantreg` in the R package `MCMCpack` are also reported. We consider three choices of τ, 0.5, 0.75 and 0.95. For each simulation study, we generated the error term ε_i from three distributions: a normal distribution N(μ, 9) with μ adjusted so that the τth quantile is zero, a $t_{(3)}$ distribution with three degrees of freedom, and a $\chi^2_{(3)}$ distribution with three degrees of freedom. For each simulation study and each choice of the error distribution, we run 100 simulations. In each simulation study, our algorithm is run for 11000 iterations and the first 1000 were removed as burn in. Methods are evaluated based on median of mean absolute deviations (MMAD), where MMAD=median(mean($|\boldsymbol{x}_i'\hat{\boldsymbol{\beta}} - \boldsymbol{x}_i'\boldsymbol{\beta}^{true}|$)).

4.1. Simulation 1

In this simulation study, we illustrate the performance of the proposed approach for sparse models. In particular, we consider the true model

$$y_i = 3x_{1i} + 1.5x_{2i} + 2x_{5i} + \varepsilon_i,$$

where $i = 1, 2, \cdots, 100$.

We simulate 8 independent variables from a multivariate normal with mean zero and Cov(x_h, xg) = $0.5^{|h-g|}$. Thus the correct coefficients of predictors, including the intercept term, are $\boldsymbol{\beta} = (0, 3, 1, 0, 0, 1, 0, 0, 0)'$.

The MMADs and the standard deviations (SD) of the MMADs are reported in Table 1. It is readily observed that in all the quantile levels and error distributions under considerations, the proposed method (LassoU) uniformly achieves smaller MMADs and SD, which shows that LassoU estimates more accurately. In general, the two Bayesian regularized QReg methods (LassoU and LassoN) perform better than the other two methods. Compared to the `rq` results, the Bayesian methods still perform well even when the error distribution

Table 1: MMADs for Simulation 1. Standard deviations of the MMADs are displayed in parentheses. The results are averaged over 100 independent simulations.

Methods	τ	$\varepsilon_i \sim N(\mu, 9)$	$\varepsilon_i \sim t_{(3)}$	$\varepsilon_i \sim \chi^2_{(3)}$
LassoU	0.50	0.559 (0.120)	0.249 (0.067)	2.819 (0.193)
LassoN	0.50	0.564 (0.150)	0.260 (0.067)	2.907 (0.194)
rq	0.50	0.747 (0.160)	0.346 (0.100)	3.010 (0.244)
MCMCquantreg	0.50	0.650 (0.120)	0.285 (0.071)	2.921 (0.205)
LassoU	0.75	0.961 (0.178)	0.313 (0.063)	3.001 (0.242)
LassoN	0.75	0.981 (0.190)	0.320 (0.064)	3.003 (0.245)
rq	0.75	1.016 (0.247)	0.362 (0.105)	3.006 (0.285)
MCMCquantreg	0.75	1.063 (0.180)	0.325 (0.077)	2.843 (0.234)
LassoU	0.95	1.458 (0.401)	0.764 (0.311)	3.591 (0.560)
LassoN	0.95	1.477 (0.405)	0.765 (0.299)	3.629 (0.567)
rq	0.95	1.555 (0.423)	0.932 (0.410)	3.712 (0.594)
MCMCquantreg	0.95	2.014 (0.360)	1.139 (0.574)	2.839 (0.548)

assumption is violated. For convergence diagnosis, we present the multivariate potential scale reduction factor (MPSRF) reported by [9] for one particular run.

From Figure (1), we can observe that the MPSRF becomes stable and close to 1 after about 2000 iterations for each $p \in \{0.50, 0.75, 0.95\}$, which shows that the convergence of the posterior distribution for the proposed method was quick and the mixing was good.

4.2. Simulation 2

For simulation 2, the performance of the proposed approach is illustrate for very sparse models. Here, we consider the true model

$$y_i = 3x_{1i} + \varepsilon_i,$$

where $i = 1, 2, \cdots, 100$. Similar to Simulation 1, 8 independent variables are simulated from a multivariate normal with mean zero and $\text{Cov}(x_h, xg) = 0.5^{|h-g|}$. The correct coefficients, including the intercept term, are $\boldsymbol{\beta} = (0, 3, 0, 0, 0, 0, 0, 0, 0)'$.

From Table 2, it can be observed that our proposed method has smallest MMADs in 8 out of 9 comparative experiments, which shows that LassoU es-
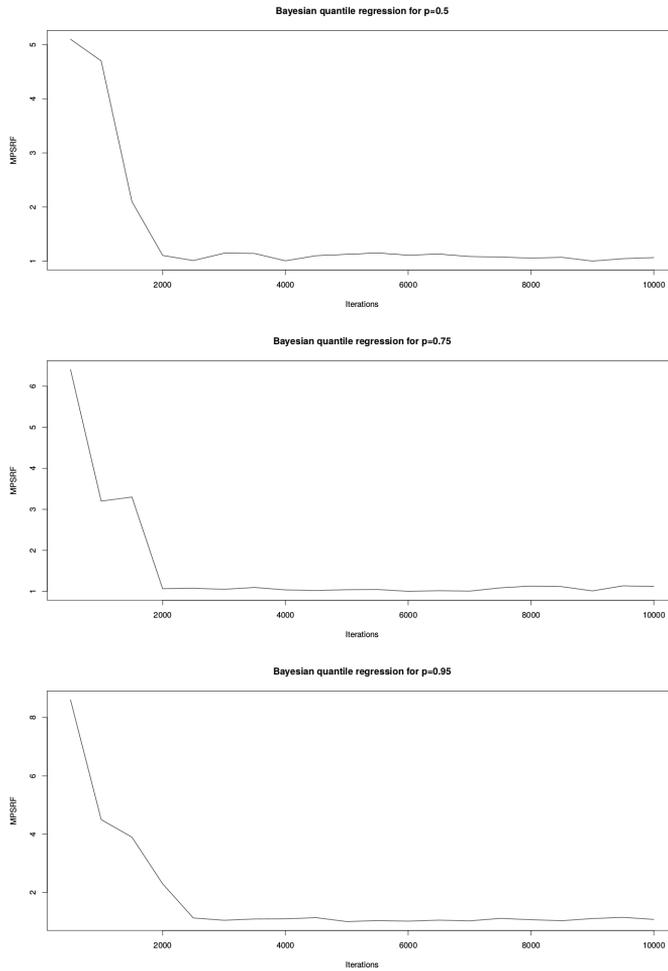
Figure 1: MPSRF for the proposed method in Simulation 1.

timates are more accurately than the other methods. Although none of these three error distributions are assumed in the Bayesian QReg methods, the simulation results show that the performance of the Bayesian methods are quite robust to the error distribution assumption.

Table 2: MMADs for Simulation 2. Standard deviations of the MMADs are displayed in parentheses. The results are averaged over 100 independent simulations.

Methods	τ	$\varepsilon_i \sim N(\mu, 9)$	$\varepsilon_i \sim t_{(3)}$	$\varepsilon_i \sim \chi^2_{(3)}$
LassoU	0.50	0.507 (0.122)	0.274 (0.072)	2.689 (0.132)
LassoN	0.50	0.609 (0.126)	0.293 (0.075)	2.721 (0.138)
rq	0.50	0.806 (0.176)	0.352 (0.090)	2.832 (0.204)
MCMCquantreg	0.50	0.748 (0.140)	0.321 (0.078)	2.741 (0.138)
LassoU	0.75	0.786 (0.151)	0.322 (0.078)	3.015 (0.177)
LassoN	0.75	0.809 (0.147)	0.355 (0.084)	3.017 (0.177)
rq	0.75	1.018 (0.178)	0.419 (0.098)	2.948 (0.277)
MCMCquantreg	0.75	0.981 (0.140)	0.376 (0.073)	2.847 (0.182)
LassoU	0.95	1.730 (0.355)	0.736 (0.236)	3.386 (0.625)
LassoN	0.95	1.755 (0.346)	0.736 (0.238)	3.375 (0.641)
rq	0.95	1.732 (0.427)	0.867 (0.272)	3.603 (0.567)
MCMCquantreg	0.95	2.205 (0.350)	1.213 (0.372)	2.615 (0.556)

4.3. Simulation 3

For simulation 3, the performance of the proposed approach is illustrate for dense sparse models. Here, we consider the true model

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

where $\boldsymbol{\beta} = (0.00, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85)'$, including the intercept term.

From Table 3, it can be observed that the proposed method has smallest MMADs in 8 out of 9 comparative experiments, which shows again that our estimates are more accurately than the other methods. Most noticeably, when $\tau = 0.95$ the median of mean absolute deviation (MMAD) generated by the proposed method for the three error distributions is much smaller than the median of mean absolute deviation generated by other methods.

Instead of looking at the estimation of MMAD and SD for the Simulation studies 1, 2 and 3, we could also look at the estimation of $\boldsymbol{\beta}$ in direct way. Since the simulation results would be too many to list in a table, we only choose the case where $\tau = 0.95$ and the error distribution follows $N(\mu, 9)$ for illustrations. From Table 4, we see that our method performs well in estimating the regression

Table 3: MMADs for Simulation 3. Standard deviations of the MMADs are displayed in parentheses. The results are averaged over 100 independent simulations.

Methods	τ	$\varepsilon_i \sim N(\mu, 9)$	$\varepsilon_i \sim t_{(3)}$	$\varepsilon_i \sim \chi^2_{(3)}$
LassoU	0.50	0.704 (0.157)	0.301 (0.131)	2.816 (0.130)
LassoN	0.50	0.723 (0.167)	0.316 (0.129)	2.812 (0.135)
rq	0.50	0.767 (0.227)	0.325 (0.151)	2.954 (0.149)
MCMCquantreg	0.50	0.706 (0.194)	0.302 (0.135)	2.854 (0.139)
LassoU	0.75	0.871 (0.155)	0.343 (0.152)	2.870 (0.163)
LassoN	0.75	0.875 (0.159)	0.344 (0.152)	2.901 (0.165)
rq	0.75	1.017 (0.216)	0.380 (0.153)	2.878 (0.262)
MCMCquantreg	0.75	0.930 (0.192)	0.357 (0.145)	2.809 (0.189)
LassoU	0.95	1.569 (0.260)	0.722 (0.523)	3.240 (0.504)
LassoN	0.95	1.576 (0.262)	0.730 (0.528)	3.241 (0.550)
rq	0.95	1.745 (0.266)	0.888 (0.692)	3.344 (0.454)
MCMCquantreg	0.95	2.049 (0.320)	1.008 (0.764)	3.501 (0.519)

coefficients compared to the other approaches. Compared to the three other approaches, our estimates are very close to the true parameter values.

5. Air Pollution Data

In this section, we use QReg methods to analyze the air pollution data [23] previously analyzed in [1]. This dataset was measured by the *Public Roads Administration in Norway* and consists of 500 observations, 7 covariates and one response variable. The response variable is the log (concentration of NO2 per hour), and the seven covariates are: the log (number of cars per hour) referred to as x_1, temperature (x_2), wind speed in meters per second (x_3), the temperature difference (x_4), wind direction (x_5), time of day in hours (x_6), and day number (x_7).

Similar to Section 4, in this section we compare four methods: rq, MCMC-quantreg, LassoN, and our approach. The methods are evaluted based on the mean squared error (MSE) and 95% intervals for $\tau \in \{0.50, 0.75, 0.95\}$. The results of the MSE are listed in Table 5. From Table 5, in general, we can see that our proposed method tends to produce lower MSE and standard devia-

Table 4: Posterior means for the Simulation studies 1, 2 and 3 when the error is $N(\mu, 9)$ and $\tau=0.95$. The results are averaged over 100 simulations.

Methods	τ	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
True Values	0.95	4.935	3.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LassoU	0.95	4.693	3.146	0.254	-0.099	0.318	0.135	-0.169	-0.083	-0.083
LassoN	0.95	4.685	3.159	0.240	-0.102	0.339	0.129	-0.186	-0.078	-0.097
rq	0.95	4.435	3.100	0.185	-0.251	0.459	-0.056	-0.262	0.190	0.179
MCMCquantreg	0.95	5.318	3.252	0.068	-0.097	0.102	0.242	-0.120	-0.441	-0.123
True Value	0.95	4.935	3.000	1.500	0.000	0.000	2.000	0.000	0.000	0.000
LassoU	0.95	5.005	2.730	1.507	-0.072	0.107	1.871	0.053	-0.211	-0.012
LassoN	0.95	4.979	2.746	1.524	-0.080	0.108	1.847	0.082	-0.181	-0.015
rq	0.95	4.752	2.975	1.747	-0.272	0.219	1.825	-0.292	-0.666	-0.063
MCMCquantreg	0.95	5.624	2.886	1.589	-0.299	0.060	1.966	-0.107	-0.276	-0.026
True Value	0.95	4.935	0.850	0.850	0.850	0.850	0.850	0.850	0.850	0.850
LassoU	0.95	4.680	0.912	0.868	0.825	0.845	0.829	0.838	0.815	0.854
LassoN	0.95	4.683	0.912	0.779	0.871	0.771	0.632	0.797	0.800	1.055
rq	0.95	4.660	1.007	0.892	1.012	0.692	0.675	0.749	0.870	1.121
MCMCquantreg	0.95	5.340	0.768	0.787	0.829	0.919	0.799	1.008	0.817	0.936

tion than the other methods, suggesting a good performance of the proposed method. On the other hand, we can observe that the Bayesian Lasso (LassoN) does not perform well for $\tau = 0.75$ and 0.95. In fact, the MSEs and standard deviations produce by LassoN are very high compared to the other approaches. From Table 6, it can be seen that our estimates are very close to the rq estimated and our credible intervals are much narrower than the intervals given by the other three methods. Although, our intervals are narrower than the others, the estimates of rq and MCMCquantreg lie inside our intervals. Hence, the analysis shows strong support for the apply of the proposed method to inference for QReg.

Table 5: MSEs and standard deviations (SD) for the air pollution data.

Methods	$\tau = 0.50$ MSE (SD)	$\tau = 0.75$ MSE (SD)	$\tau = 0.95$ MSE (SD)
LassoU	**0.649** (**0.989**)	0.943 (**1.416**)	**2.443** (**2.568**)
LassoN	0.650 (1.001)	0.964 (1.453)	3.247 (3.184)
rq	0.655 (1.012)	**0.939** (1.428)	2.448 (2.592)
MCMCquantreg	0.651 (1.007)	0.959 (1.440)	2.729 (2.724)

6. Conclusion and Discussion

In this paper, we propose a new hierarchical model of Bayesian Lasso quantile regression by utilizing the scale mixture of uniform representation of the Laplace density. This representation leads to a simple and efficient Gibbs sampling algorithm for Bayesian Lasso quantile regression with tractable full conditional posterior distributions. The proposed method does not put any distributional assumption on the response of interest, and thus is able to accommodate non-normal errors, which are popular in many areas.

Simulation studies show that the proposed Gibbs sampler is effective in shrinkage and estimation the regression coefficients under a variety of scenarios. Our simulation scenarios also indicate that our method works well even when the true distribution for the error term is not ALD. This phenomenon was also observed by [35], [21], [5], [14], [6] and [15], among others. In addition, the analysis of the air pollution data shows strong support for the use of our proposed method.

Table 6: Estimates and 95% intervals for the 0.50, 0.75 and 0.95 QReg parameters of the air pollution data.

Parameters	LassoU Mean (95% CrI)	LassoN Mean (95% CrI)	rq Mean (lower bd, upper bd)	MCMCquantreg Mean (95% CrI)
$\tau = 0.50$				
β_0	1.360 (0.711, 1.974)	1.525 (0.659, 3.837)	1.617 (0.546, 2.638)	1.798 (0.882, 2.104)
β_1	0.306 (0.222, 0.390)	0.287 (-0.010, 0.405)	0.275 (0.131, 0.422)	0.259 (0.197, 0.343)
β_2	0.001 (-0.011, 0.012)	0.002 (-0.010, 0.016)	0.003 (-0.018, 0.026)	0.004 (0.008, 0.015)
β_3	-0.106 (-0.144, -0.067)	-0.110 (-0.157, -0.071)	-0.112 (-0.184, -0.043)	-0.118 (-0.151, -0.075)
β_4	0.039 (-0.042, 0.120)	0.031 (-0.121, 0.123)	0.035 (-0.120, 0.191)	0.011 (-0.059, 0.100)
β_5	0.000 (-0.001, 0.001)	-0.000 (-0.001, 0.001)	-0.000 (-0.002, 0.001)	0.000 (-0.001, 0.001)
β_6	-0.000 (-0.012, 0.012)	0.000 (-0.012, 0.016)	0.002 (-0.020, 0.024)	-0.002 (0.013, 0.011)
β_7	0.000 (-0.000, 0.001)	0.000 (-0.000, 0.001)	0.000 (-0.000, 0.001)	0.000 (-0.000, 0.001)
$\tau = 0.75$				
β_0	2.011 (1.337, 2.652)	3.730 (1.571, 23.367)	2.189 (1.083, 3.305)	2.281 (1.413, 2.843)
β_1	0.294 (0.197, 0.394)	0.084 (-2.288, 0.375)	0.269 (0.095, 0.446)	0.263 (0.156, 0.372)
β_2	0.011 (-0.002, 0.023)	0.016 (-0.001, 0.080)	0.010 (-0.016, 0.034)	0.016 (0.000, 0.023)
β_3	-0.106 (-0.148, -0.061)	-0.114 (-0.190, -0.068)	-0.107 (-0.189, -0.017)	-0.112 (-0.157, -0.068)
β_4	0.047 (-0.034, 0.125)	-0.003 (-0.410, 0.115)	0.033 (-0.123, 0.183)	0.049 (-0.059, 0.120)
β_5	-0.000 (-0.001, 0.001)	-0.001 (-0.006, 0.001)	-0.000 (-0.002, 0.002)	-0.001 (-0.002, 0.001)
β_6	0.004 (-0.010, 0.018)	0.006 (-0.013, 0.030)	0.006 (-0.020, 0.034)	0.005 (-0.010, 0.027)
β_7	0.000 (-0.000, 0.001)	0.000 (-0.002, 0.001)	0.000 (-0.001, 0.001)	0.000 (-0.000, 0.001)
$\tau = 0.95$				
β_0	2.020 (1.389, 2.582)	3.327 (1.446, 20.129)	2.178 (1.046, 3.306)	2.281 (1.413, 2.843)
β_1	0.292 (0.201, 0.378)	0.127 (-1.915, 0.394)	0.271 (0.102, 0.447)	0.263 (0.156, 0.372)
β_2	0.011 (-0.002, 0.022)	0.015 (-0.001, 0.066)	0.010 (-0.016, 0.034)	0.016 (0.000, 0.023)
β_3	-0.106 (-0.148, -0.062)	-0.110 (-0.168, -0.063)	-0.108 (-0.192, -0.021)	-0.112 (-0.157, -0.068)
β_4	0.046 (-0.036, 0.122)	0.003 (-0.399, 0.120)	0.034 (-0.120, 0.186)	0.049 (-0.059, 0.120)
β_5	-0.000 (-0.001, 0.000)	-0.001 (-0.005, 0.001)	-0.000 (-0.002, 0.001)	-0.001 (-0.002, 0.001)
β_6	0.004 (-0.010, 0.018)	0.008 (-0.010, 0.045)	0.006 (-0.021, 0.033)	0.005 (-0.010, 0.027)
β_7	0.000 (-0.000, 0.001)	0.000 (-0.002, 0.001)	0.000 (-0.001, 0.001)	0.000 (-0.000, 0.001)

References

[1] Al-Hamzawi, R. J. T. (2013). Prior elicitation and variable selection for bayesian quantile regression. Ph.D. thesis, Brunel University, School of Information Systems, Computing and Mathematics.

[2] Alhamzawi, R. (2014). Bayesian elastic net Tobit quantile regression. *Communications in Statistics-Simulation and Computation*, (just-accepted): 00–00.

[3] Alhamzawi, R. and Ali, H. T. M. (2017). The Bayesian Elastic Net Regression. *Communications in Statistics-Simulation and Computation*, (just-accepted).

[4] Alhamzawi, R. and Yu, K. (2014). Bayesian Lasso-mixed quantile regression. *Journal of Statistical Computation and Simulation*, 84(4): 868–880.

[5] Alhamzawi, R., Yu, K., and Benoit, D. F. (2012). Bayesian adaptive Lasso quantile regression. *Statistical Modelling*, 12(3): 279–297.

[6] Alshaybawee, T., Midi, H., and Alhamzawi, R. (2016). Bayesian elastic net single index quantile regression. *Journal of Applied Statistics*, 1–19.

[7] Baraniuk, R. G. (2007). Compressive sensing. *IEEE signal processing magazine*, 24(4).

[8] Benoit, D. F., Alhamzawi, R., and Yu, K. (2013). Bayesian lasso binary quantile regression. *Computational Statistics*, 28(6): 2861–2873.

[9] Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4): 434–455.

[10] Cade, B. S. and Noon, B. R. (2003). A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, 1(8): 412–420.

[11] Candes, E. and Tao, T. (2007). The Dantzig selector: statistical estimation when p is much larger than n. *The Annals of Statistics*, 2313–2351.

[12] Candès, E. J. and Recht, B. (2009). Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6): 717–772.

[13] Chhikara, R. (1988). *The Inverse Gaussian Distribution: Theory: Methodology, and Applications*, volume 95. CRC Press.

[14] Hashem, H., Vinciotti, V., Alhamzawi, R., and Yu, K. (2016). Quantile regression with group lasso for classification. *Advances in Data Analysis and Classification*, 10(3): 375–390.

[15] Ji, Y., Lin, N., and Zhang, B. (2012). Model selection in binary and tobit quantile regression using the Gibbs sampler. *Computational Statistics & Data Analysis*, 56(4): 827–839.

[16] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 33–50.

[17] Koenker, R. and Geling, O. (2001). Reappraising medfly longevity: a quantile regression survival analysis. *Journal of the American Statistical Association*, 96(454): 458–468.

[18] Koenker, R. and Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the american statistical association*, 94(448): 1296–1310.

[19] Kostov, P. and Davidova, S. (2013). A quantile regression analysis of the effect of farmers attitudes and perceptions on market participation. *Journal of Agricultural Economics*, 64(1): 112–132.

[20] Kozumi, H. and Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of statistical computation and simulation*, 81(11): 1565–1578.

[21] Li, Q., Xi, R., Lin, N., et al. (2010). Bayesian regularized quantile regression. *Bayesian Analysis*, 5(3): 533–556.

[22] Li, Y. and Zhu, J. (2008). L1-norm quantile regression. *Journal of Computational and Graphical Statistics*, 17(1).

[23] Lindmark, A. and Karlsson, M. (2011). truncSP: Semi-parametric estimators of truncated regression models. R package version 1.1.

[24] Mallick, H. and Yi, N. (2014). A new Bayesian lasso. *STATISTICS AND ITS INTER- FACE*, 7(4): 571–582.

[25] Park, T. and Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, 103(482): 681–686.

[26] Sun, W., Ibrahim, J. G., and Zou, F. (2010). Genomewide multiple-loci mapping in experimental crosses by iterative adaptive penalized regression. *Genetics*, 185(1): 349–359.

[27] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

[28] — (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3): 273–282.

[29] Wang, H. and He, X. (2007). Detecting differential expressions in GeneChip microarray studies: a quantile approach. *Journal of the American Statistical Association*, 102(477): 104–112.

[30] Wei, Y., Pere, A., Koenker, R., and He, X. (2006). Quantile regression methods for reference growth charts. *Statistics in medicine*, 25(8): 1369–1382.

[31] Wu, Y. and Liu, Y. (2009). Variable selection in quantile regression. *Statistica Sinica*, 19(2): 801.

[32] Yu, K., Dang, W., Zhu, H., and Al Hamzawi, R. (2013). Comment on article by spokoiny, wang and härdle. *Journal of Statistical Planning and Inference*, 7(143): 1140–1144.

[33] Yu, K., Liu, X., Alhamzawi, R., Becker, F., and Lord, J. (2016). Statistical meth- ods for body mass index: a selective review. *Statistical methods in medical research*, 0962280216643117.

[34] Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, 54(4): 437–447.

[35] Yuan, Y. and Yin, G. (2010). Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics*, 66(1): 105–114.

[36] Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476): 1418–1429.