Statistics and Its Interface Volume 14 (2021) 459–474

# Prior conditioned on scale parameter for Bayesian quantile LASSO and its generalizations

Zhongheng Cai* and Dongchu Sun

Several undesirable issues exist in the Bayesian quantile LASSO and its two generalizations, quantile group LASSO and bridge quantile regression (Alhamzawi et al. [1]; Alhamzawi and Algamal [2]; Li et al. [21]). In this paper, we numerically show that, the joint posterior may be multimodal using unconditional prior for the regression coefficients and the posterior estimates may be sensitive to the hyperparameters in Gamma prior frequently used for the scale parameter. We also theoretically illustrate that the joint posterior may be improper when an invariant prior is used for the scale parameter, especially when predictors outnumber observations. To resolve the issues in a unified framework, we propose applying the priors conditioned on the scale parameter for the coefficients along with invariant prior to the scale parameter. We justify the prior choice under one general likelihood including asymmetric Laplace density and the common class of conditioned priors by establishing the corresponding sufficient and necessary condition of the posterior propriety. In addition, we develop ready-to-use partially collapsed Gibbs sampling algorithms for all methods to aid computations. Simulation studies and a real data example demonstrate that our methods usually outperform the original Bayesian approaches.

Keywords and phrases: Asymmetric Laplace density, Bayesian quantile LASSO, Bridge regression, Group LASSO, Posterior propriety.

## 1. INTRODUCTION

While normal linear regression has been used for several centuries, quantile regression, first proposed by Koenker and Bassett [14], has emerged as a useful supplement to ordinary mean regression and has widespread applications. Given a fixed quantile level $p$, the linear quantile regression for the response $y_i$ is

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \mu_i,\ i = 1, \cdots, n,$$

where $\boldsymbol{x}_i \in \mathbb{R}^m$ is the explanatory variables with unknown regression coefficients $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_m)'$ and $\mu_i$ is the independent error term whose distribution is restricted to have the $p$th quantile zero. From a Bayesian perspective, one could either model the error distribution as mixture distributions based on the Dirichlet process (Kottas and Krnjajić [16]) or assume $\mu_i$ follows asymmetric Laplace distribution (ALD) proposed by Yu and Moyeed [36]. We focus on the latter for guaranteeing posterior consistency of Bayesian estimators (Sriram et al. [30]) and robustness (Yu and Moyeed [36]) even if the random errors do not follow ALD. The density of ALD$(\mu, \sigma, p)$ with location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma > 0$ and $p$th quantile is

$$p(y|\mu, \sigma, p) = p(1-p)\sigma \exp(-\sigma\rho_p(y-\mu)), \tag{1}$$

where $\rho_p(x)$ is called a check function,

$$\rho_p(x) = \begin{cases} px, & \text{if } x > 0, \\ (p-1)x, & \text{if } x \le 0. \end{cases} \tag{2}$$

Koenker and Bassett [14] firstly proposed to solve

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_p(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}), \tag{3}$$

to obtain the estimator of linear quantile coefficients. In constrast, with ALD, the likelihood function based on $\boldsymbol{y} = (y_1, \cdots, y_n)'$ is

$$p(\boldsymbol{y}|p, \boldsymbol{\beta}, \sigma) = p^n(1-p)^n\sigma^n \exp(-\sigma \sum_{i=1}^{n} \rho_p(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})). \tag{4}$$

Clearly, with a constant prior on $\boldsymbol{\beta}$, the posterior mode of $\boldsymbol{\beta}$ given $\sigma$ is the minimizer of (3).

Nowadays, regularization approaches for quantile regression have received considerable attention for improving prediction accuracy and selecting active covariates among a large set of candidate covariates (Jiang et al. [18]; Jiang et al. [19]; Jiang et al. [20]; Li et al. [22]; Li and Zhu [23]; Wang et al. [33]; Wu and Liu [34]). It can be formulated as,

$$\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{n} \rho_p(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}) + \lambda f(\boldsymbol{\beta}) \right\}, \tag{5}$$

where $\lambda > 0$ is the regularization parameter and $f(\boldsymbol{\beta})$ is a given nonnegative penalty for $\boldsymbol{\beta}$. Three penalty functions are considered in this paper.

*Corresponding author.

(a). Least absolute shrinkage and selection operator (LASSO) (Tibshirani [31]) corresponds to

$$f(\boldsymbol{\beta}) = \sum_{i=1}^{m} |\beta_i|. \tag{6}$$

(b). Suppose that the predictors are classified into $G$ groups and $\boldsymbol{\beta}_g$ is the coefficient vector of the $g$th group. Define $\boldsymbol{x}_i = (\boldsymbol{x}'_{i1}, \cdots, \boldsymbol{x}'_{iG})'$, $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \cdots, \boldsymbol{\beta}'_G)'$, $||\boldsymbol{\beta}_g||_{\boldsymbol{K}_g} = \sqrt{\boldsymbol{\beta}'_g \boldsymbol{K}_g \boldsymbol{\beta}_g}$, $\boldsymbol{K}_g$ is a known $d_g \times d_g$ positive definite matrix and $\sum_{g=1}^{G} d_g = m$. Then the group LASSO (Yuan and Lin [37]) corresponds to

$$f(\boldsymbol{\beta}) = \sum_{g=1}^{G} ||\boldsymbol{\beta}_g||_{\boldsymbol{K}_g}. \tag{7}$$

(c). The bridge quantile regression corresponds to

$$f(\boldsymbol{\beta}) = \sum_{i=1}^{m} |\beta_i|^q, \tag{8}$$

for some $q > 0$.

To solve (5) under the Bayesian framework, the prior for $\boldsymbol{\beta}$ given $(\sigma, \lambda)$

$$p(\boldsymbol{\beta}|\sigma, \lambda) = C_{\sigma,\lambda} \exp(-\sigma\lambda f(\boldsymbol{\beta})), \ \boldsymbol{\beta} \in \mathbb{R}^m, \tag{9}$$

is utilized, where $C_{\sigma,\lambda}$ is the normalizing constant depending on $(\sigma, \lambda)$. The posterior mode of $\boldsymbol{\beta}$ is the solution to (5) given $(\sigma, \lambda)$. In the fully Bayesian analysis, it is often to set $\eta = \lambda\sigma$ and add the independent Gamma prior either on $(\sigma, \eta^2)$ in Bayesian quantile LASSO (BQL) and Bayesian quantile group LASSO (BQGL) (Alhamzawi et al. [2]; Li et al., [21]) as,

$$\sigma \sim \text{Gamma}(a, b), \ \eta^2 \sim \text{Gamma}(c, d), \tag{10}$$

or $(\sigma, \eta)$ in Bayesian bridge quantile regression (BBQR) (Alhamzawi and Algamal [1]) as,

$$\sigma \sim \text{Gamma}(a, b), \ \eta \sim \text{Gamma}(c, d), \tag{11}$$

where a random variable $X \sim \text{Gamma}(\alpha, \beta)$ has the density $\beta^\alpha \Gamma^{-1}(\alpha) x^{\alpha-1} \exp(-\beta x), x > 0$. In fact, the prior (9) turns out to be irrelevant to $\sigma$ when using the parametrization $\eta = \sigma\lambda$. We refer it to the unconditional prior for $\boldsymbol{\beta}$. With the unconditional prior for $\boldsymbol{\beta}$ and Gamma prior for $\sigma$, several issues arise.

- The joint posterior may be multimodal. It slows down the convergence of the Gibbs sampler and the point estimate may be computed through multiple modes, which leads to the inaccurate estimators (Kyung et al. [15]; Park and Casella [27]).
- The choices of the hyperparameters $a$ and $b$ may have strong influences on the posterior estimates.

We demonstrate the two issues above via simulation studies in Section 4. For the first, it was firstly observed by Park and Casella [27] in the Bayesian LASSO. We observe the similar phenomenon in BQL. For the second, it is common to utilize invariant prior (Berger [5]) on $\sigma$ for an objective purpose. However, as we prove in Section 2, the posterior $(\boldsymbol{\beta}, \sigma, \eta^2|\boldsymbol{y})$ or $(\boldsymbol{\beta}, \sigma, \eta|\boldsymbol{y})$ does not exist with the invariant prior for $\sigma$ when the predictors outnumber the observations (large $p$ and small $n$). We argue that the posterior impropriety deserves more attention since it exists not only in the BQL and its generalizations but also in the counterpart of Bayesian mean regularized regression such as Bayesian bridge regression (BBR) (Polson et al. [28]). To demonstrate this point, we propose one general density function including normal distribution and ALD as special cases with one general prior for $\boldsymbol{\beta}$ accommodating LASSO, group LASSO and bridge penalty. We further establish conditions for posterior propriety when the unconditional prior is imposed on the regression coefficients. To solve all aforementioned problems, we propose the prior for $\boldsymbol{\beta}$ conditioned on scale parameter and obtain the sufficient and necessary conditions for the existence of the joint posterior.

The remainder of this paper is as follows. In Section 2, we formally define a general class of likelihood with a generalized prior for the regression coefficients and derive the conditions of posterior propriety with respect to Bayesian regularized approaches. In Section 3, we develop the partially collapsed Gibbs sampling algorithm for BQL with its two generalizations. In Sections 4 and 5, a wide range of simulation studies and one real data example are conducted. In Section 6, we draw the conclusions and discuss some feasible future directions of our work.

## 2. PRIOR CONDITIONED ON THE SCALE PARAMETER

Suppose that $e_i$ in $\boldsymbol{e} = (e_1, \cdots, e_n)'$ are independently and identically distributed (i.i.d). We consider a general class of density $\zeta(\boldsymbol{e}) = \prod_{i=1}^{n} \zeta(e_i)$ and assume that there exists $q_1 < 0, q_2 \in \mathbb{R}, q_3 < 0$ and $q_4 \in \mathbb{R}$ such that

$$q_1 ||\boldsymbol{e}||_l^{k_1} + q_2 \le \log(\zeta(\boldsymbol{e})) \le q_3 ||\boldsymbol{e}||_r^{k_1} + q_4 \tag{12}$$

for any $\boldsymbol{e} \in \mathbb{R}^n$, where $||.||_l$ and $||.||_r$ are the pre-norms (Horn and Johnson [11]) on $\mathbb{R}^n$ and $k_1 > 0$ is fixed. One pre-norm is a real-valued function on a finite-dimensional real or complex vector space satisfying the three hypotheses of positivity, homogeneity, and continuity. All norms are the pre-norms, however, the converse is not true. For example, $||\boldsymbol{a}||_q = (\sum_{i=1}^{n} |a_i|^q)^{1/q}$ for $0 < q < 1$ is a pre-norm not a norm.

(12) includes standard normal distribution, $e_i \sim N(0, 1)$, by taking $||.||_l = ||.||_r = ||.||_2, k_1 = 2, q_1 = q_3 = -0.5, q_2 = q_4 = -n\ln(\sqrt{2\pi})$ and $e_i$ ∼ALD$(0, 1, p)$ by taking $||.||_l = ||.||_r = ||.||_1, k_1 = 1, q_1 = -\max(p, 1-p), q_3 = -\min(p, 1-$

$p), q_2 = q_4 = n\ln(p(1-p))$, respectively. With the scale parameter $\sigma$ and linear trend $\boldsymbol{X\beta}$, the likelihood of $(\sigma, \boldsymbol{\beta})$ is

$$\sigma^n \prod_{i=1}^{n} \zeta(\sigma(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})). \tag{13}$$

For $\boldsymbol{\beta}$, we consider the generalized unconditional prior which is proportional to

$$\eta^{\frac{m}{k_2}} \exp(-\eta F(\boldsymbol{\beta})), \tag{14}$$

where $F(\boldsymbol{\beta})$ satisfies that $F^{1/k_2}(\boldsymbol{\beta})$ is a pre-norm on $\mathbb{R}^m$ for some $k_2 > 0$. (14) includes LASSO, group LASSO and bridge penalty by $k_2$ taking $1, 1$, and $q$ with $F(\boldsymbol{\beta})$ taking (6), (7) and (8), respectively. In addition, to include the prior (10) and (11), we impose the independent Gamma prior on $(\sigma^{k_1}, \eta^h)$

$$(\sigma^{k_1}, \eta^h) \sim \text{Ga}(a, b)\text{Ga}(c, d), \tag{15}$$

with two choices of $h$, $h \geq 2$ or $h = 1$. Here, we allow $a, c$ to be the real and $b, d$ to be nonnegative to include some limiting cases of Gamma priors such as invariant prior. With the likelihood (13) and prior (14), (15), we have the following theorem.

**Theorem 1.** *Assume the design matrix* $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)$ *has the rank* $r$, $a \in \mathbb{R}, b \geq 0, c \in \mathbb{R}$ *and* $d \geq 0$ *in* (15). *Define SSE* $= \boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{y}$.

*(a). When* $h \geq 2$, *if the posterior of* $(\boldsymbol{\beta}, \sigma^{k_1}, \eta^h|\boldsymbol{y})$ *is proper, Conditions A, B and C hold.*
*Condition A. One of the following holds:*
*(A1.)* $d > 0, n + ak_1 > 0$;
*(A2.)* $d = 0, c < 0$.
*Condition B. One of the following holds:*
*(B1.) SSE*$+b > 0, r + hk_2c > 0$;
*(B2.) SSE*$+b = 0, n - r + ak_1 < 0$.
*Condition C.* $n + ak_1 + hk_2c > 0$.

*(b). When* $h = 1$, *Condition A, B and C are also sufficient for the posterior propriety of* $(\boldsymbol{\beta}, \sigma^{k_1}, \eta|\boldsymbol{y})$.

In Theorem 1, $(a)$ provides necessary conditions of posterior propriety for $h \geq 2$ and $(b)$ indicates that Conditions in $(a)$ are also sufficient for $h = 1$. One case of interest is that an invariant prior is imposed for $\sigma^{k_1}$ $(a = b = 0)$. Table 1 presents the corresponding justifications of posterior propriety using a normal distribution or ALD as the likelihood. Notice that implications of Theorem 1 include but not limited to Table 1.

For BQL and BQGL, Li et al. [21] employed the prior in (15) with $k_1 = 1, h = 2$. For BBQR, Alhamzawi and Algamal [1] utilized $k_1 = 1, h = 1$ for BBQR. They all chose $a = b = c = d = 0.1$. Li et al. [21] stated that the vague Gamma prior was employed for an almost noninformative purpose. However, as demonstrated in Table 1, the joint posterior is improper with invariant priors ($a = b = c = d = 0$ in (15)) for $(\sigma, \eta)$. Hence, this specification for the hyperparameters is unfavorable. Furthermore, it is inconvenient to utilize invariant prior for scale component $\sigma$ with the unconditional prior (14) for $\boldsymbol{\beta}$. The reason is that the joint posterior does not exist when SSE=0 even with a proper Gamma prior for $\eta^h$. Therefore, we advise extra caution when one intends to use prior (15) with an invariant prior for $\sigma$.

To enable a safe use for invariant prior for $\sigma$, we propose a general prior for $\boldsymbol{\beta}$,

$$\lambda^{\frac{m}{k_2}} \sigma^m \exp(-\lambda F(\sigma\boldsymbol{\beta})). \tag{16}$$

The main difference between prior (14) and (16) is that our prior is associated with scale parameter $\sigma$ and hence named as conditional prior. For the prior on $\sigma$, we adopt the invariant prior, $p(\sigma) \propto \sigma^{-1}$, which is preferable when little information is available. For the prior on $\lambda$, we consider two cases,

1. When $k_1 = k_2 = k$, we assume a general type prior, $g(\lambda)$, for $\lambda$, where $g(.)$ is a pre-specified function.
2. When $k_1 \neq k_2$, we impose a prior Gamma$(a_0, b_0), a_0 > 0, b_0 > 0$.

We have the following theorem.

**Theorem 2.** *When* $\zeta(e)$ *is log-concave,* $F^{1/k_2}(\boldsymbol{\beta})$ *is a norm on* $\mathbb{R}^m$, $k_2 \geq 1$ *and* $m + n \geq k_1$, *the posterior of* $(\boldsymbol{\beta}, \sigma^{k_1}|\boldsymbol{y})$ *given* $\lambda$ *is unimodal. In addition,*

*(a). When* $k_1 = k_2 = k$, *the sufficient and necessary condition for the posterior propriety of* $(\boldsymbol{\beta}, \sigma^k, \lambda|\boldsymbol{y})$ *is*

- *When SSE=0, there exists* $\epsilon > 0$ *such that*

$$\int_0^{\epsilon} \lambda^{\frac{r-n}{k}} g(\lambda) d\lambda + \int_{\epsilon}^{\infty} g(\lambda) d\lambda < \infty. \tag{17}$$

- *When SSE* $> 0$, *there exists* $\epsilon > 0$ *such that*

$$\int_0^{\epsilon} \lambda^{\frac{r}{k}} g(\lambda) d\lambda + \int_{\epsilon}^{\infty} g(\lambda) d\lambda < \infty. \tag{18}$$

*(b). When* $k_1 \neq k_2$, *the sufficient and necessary condition for the posterior propriety of* $(\boldsymbol{\beta}, \sigma^{k_1}, \lambda|\boldsymbol{y})$ *is when SSE=0,* $a_0k_2 + r > n$.

With Theorem 2, Table 2 displays conditions for posterior propriety when $L(y)$ is normal or ALD for a quick reference. One case of interest is SSE=0. It occurs when predictors outnumber observations and $\boldsymbol{X}$ has full row rank $(r = n)$. We can see from Table 2, when $k_1 = k_2$, we only need a proper prior for $\lambda$ to ensure a proper joint posterior. When $k_1 \neq k_2$, there is no constraint on Gamma prior. Theorem 2 also implies that, under our framework, the joint posterior given $\lambda$ for BQL, BQGL and BBQR $(q \geq 1)$ is unimodal. Therefore, compared with existing literature of BQL, BQGL and BBQR, our proposed priors guarantee a proper posterior with an invariant prior for $\sigma$ when SSE=0. In addition, with proposed priors, there is no concern for the multimodality regarding BQL, BQGL and BBQR with $q \geq 1$.

Table 1. Summary of the posterior propriety under ALD and normal distribution by different penalties in Theorem 1. The symbol "+" indicates positive values and "∗" means that it holds when $h = 1$

| Density | Penalty | c | d | SSE=0 | SSE>0 |
|---|---|---|---|---|---|
| ALD | LASSO | 0 | 0 | improper | improper |
| | | + | + | improper | proper$^*$ |
| | Group LASSO | 0 | 0 | improper | improper |
| | | + | + | improper | proper$^*$ |
| | Bridge | 0 | 0 | improper | improper |
| | | + | + | improper | proper$^*$ |
| Normal | LASSO | 0 | 0 | improper | improper |
| | | + | + | improper | proper$^*$ |
| | Group LASSO | 0 | 0 | improper | improper |
| | | + | + | improper | proper$^*$ |
| | Bridge | 0 | 0 | improper | improper |
| | | + | + | improper | proper$^*$ |

Table 2. Summary of the posterior propriety under ALD and normal distribution by different penalties in Theorem 2

| Likelihood | Penalty | Is $k_1 = k_2$ | Prior for $\lambda$ | Is posterior unimodal | Is posterior proper when SSE=0 and $r = n$ |
|---|---|---|---|---|---|
| ALD | LASSO | Yes | $g(\lambda)$ | Yes | Yes, if $g(\lambda)$ is proper |
| | Group LASSO | Yes | $g(\lambda)$ | Yes | Yes, if $g(\lambda)$ is proper |
| | Bridge | No | Gamma$(a_0, b_0)$ | Yes, if $q \geq 1$ | Yes |
| Normal | LASSO | Yes | $g(\lambda)$ | Yes | Yes, if $g(\lambda)$ is proper |
| | Group LASSO | Yes | $g(\lambda)$ | Yes | Yes, if $g(\lambda)$ is proper |
| | Bridge | No | Gamma$(a_0, b_0)$ | Yes, if $q \geq 1$ | Yes |

**Remark 1.** Polson et al. [28] proposed BBR based on two distinct scale mixture representations of the generalized Gaussian density. They employed the prior (15) with $a = b = 0$. However, with the Gibbs sampling developed by Polson et al. [28], Mallick and Yi [26] found that posterior samples did not converge via simulation studies for large $p$ small $n$ case. Table 1 shows that the joint posterior is improper in this case, which provides a plausible explanation for this failure in convergence.

**Remark 2.** When $\zeta(e)$ is normal, Theorem 1 and 2 corresponds to the context of Bayesian regularized regression. In fact, the advantage of conditional priors for regression coefficients with invariant prior for the scale parameter has been identified as unimodality. For example, Park and Casella [27] proposed these priors should be used for the Bayesian LASSO to obtain a unimodal posterior. However, to the best of our knowledge, there is no theoretical evidence to justify the posterior propriety of such choice. Therefore, Theorem 1 and 2 serve as a complement rationale for using conditional prior instead of unconditional prior along with the invariant prior for $\sigma$.

**Remark 3.** Notice that Theorem 2 states the conditions of posterior propriety depends on whether $k_1$ is equal to $k_2$. When $k_1 = k_2$, we apply a general prior $g(\lambda)$ for $\lambda$ and establish the corresponding posterior propriety. However, when $k_1 \neq k_2$, we employ the Gamma$(a_0, b_0)$ for $\lambda$ instead of a general prior. The major challenge of developing a general prior for $k_1 \neq k_2$ lies in transforming the intractable terms $||\boldsymbol{y} - \boldsymbol{X\beta}||^{k_1}$ and $F(\boldsymbol{\beta})$ in the joint posterior to tractable forms. This can be easily achieved for $k_1 = k_2$ but not for $k_1 \neq k_2$. Alternatively, when $k_1 \neq k_2$, we specify a Gamma prior to relate $||\boldsymbol{y} - \boldsymbol{X\beta}||^{k_1}$ and $F(\boldsymbol{\beta})$ to tractable expressions. As for specifying a general prior for $\lambda$, it is of great interest and deserves further investigation.

## 3. PARTIALLY COLLAPSED GIBBS SAMPLING

The Gibbs sampling for our proposed methods could be performed by introducing latent variables. However, the sampling for non-standard full conditional distributions, such as $\lambda$, is computationally expensive. The partially collapsed Gibbs sampling improve the efficiency for those distributions by marginalizing out part of latent variables. In this section, for a fast computation, we derive the partially collapsed Gibbs sampling algorithms for BQL and its two generalizations, respectively.

### 3.1 Bayesian quantile LASSO

We first consider quantile regression with the LASSO penalty. To facilitate the Gibbs sampling involving ALD, Kozumi and Kobayashi [17] proved that ALD can be decomposed as a mixture of an exponential and a scaled normal distribution. To be specific, if $Y$ is distributed as ALD$(\mu, \sigma, p)$,

$$Y \overset{d}{=} \xi_1 s^{-1} + \xi_2 \sigma^{-\frac{1}{2}} \sqrt{s^{-1}} z + \mu, \tag{19}$$

where

$$\xi_1 = \frac{1-2p}{p(1-p)},\ \xi_2 = \sqrt{\frac{2}{p(1-p)}},$$

$s^{-1} \sim \text{Exp}(\sigma), z \sim N(0,1)$, $s$ and $z$ are independent. This representation allows us to express the quantile regression model as the normal distribution with the latent variable $s^{-1}$ distributed as exponential.

For the prior of $\boldsymbol{\beta}$, which corresponds to $F(\boldsymbol{\beta}) = \sum_{i=1}^m |\beta_i|$ and $k_2 = 1$ in (16), Li et al. [21] utilized the scale mixture of normal (SMN) distribution for the Laplace distribution (Andrew and Mallows [4]) to obtain the Gibbs sampler for BQL. Mallick and Yi [25] represented the Laplace distribution as the scale mixture of uniform (SMU) distribution. They illustrated that SMU provides slightly better posterior predictions than SMN and has good mixing property via extensive simulation studies. Hence, we employed the SMU for the prior of $\beta_i$ by introducing the latent variable $\nu_i$ as follows,

$$\beta_i|\nu_i \sim U(-\nu_i, \nu_i),\ \nu_i \sim \text{Gamma}(2, \sigma\lambda), \tag{20}$$

where $U(-\nu_i, \nu_i)$ is the uniform distribution on $(-\nu_i, \nu_i)$. For the prior of $\lambda$, we take $g(.)$ to be $\text{Gamma}(a, b), a > 0, b > 0$. Based on Theorem 2$(a)$, the sufficient and necessary condition for the posterior propriety of $(\boldsymbol{\beta}, \sigma, \lambda|\boldsymbol{y})$ is $a + r > n$ if SSE=0. Hence, the shape parameter needs to be chosen carefully when SSE=0. As suggested by Park and Casella [27], the prior density for $\lambda$ should approach 0 sufficiently fast as $\lambda \to \infty$ and be relatively flat. We choose $a = \max(n - r, 1)$ when SSE=0 and $a = 1$ when SSE>0. For the selection of $b$, notice that the $\lambda$ in our model can be exactly interpreted as the regularization parameter in (5). Therefore, we could borrow information from the tuned $\lambda$. We set the mean of the prior to be the best option of $\lambda$ in (5) chosen by some criteria including cross validation, BIC or EBIC (Chen and Chen [7]). Let $\lambda_{opt}$ be the best $\lambda$ chosen in (5), we set $b = \lambda_{opt}/a$.

By introducing the latent variables $\boldsymbol{s} = (s_1, \cdots, s_n)$ from the likelihood using (19) and the latent variables $\boldsymbol{\nu} = (\nu_1, \cdots, \nu_m)$ from the prior using (20), the Gibbs samplers for all the parameters but $\lambda$ can be drawn efficiently. We marginalize the latent variables $(\boldsymbol{s}, \boldsymbol{\nu})$ in $(\lambda, \boldsymbol{s}, \boldsymbol{\nu}|\sigma, \boldsymbol{\beta}, \boldsymbol{y})$ using partially collapsed Gibbs sampling (van Dyk and Park [32]). Define $\rho_p(\boldsymbol{x}) = \sum_{i=1}^n \rho_p(x_i)$, $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{\beta}_{[-k]}$ as the parameter vector $\boldsymbol{\beta}$ excluding the component $\beta_k$. We call that a random variable is distributed as generalized inverse Gaussian distribution (Jørgensen [12]), $\text{GIG}(u, v, w)$, if the density is proportional to $y^{w-1}\exp(-(2vy)^{-1} - (2vu^2)^{-1}y)$, $u > 0, v > 0, w \in \mathbb{R}$. The partially collapsed Gibbs sampling procedure in our case is described as follows:

(a). Sample $(\lambda|\sigma, \boldsymbol{\beta}; \boldsymbol{y})$ from $\text{Gamma}(m+a, \sum_{i=1}^m |\sigma\beta_i| + b)$.

(b). Sample $(\sigma|\boldsymbol{\beta}, \boldsymbol{s}, \boldsymbol{\nu}; \boldsymbol{y})$ from truncated Gamma distribution, $\text{Gamma}(k_1, k_2)I(\sigma < \min_i(\nu_i|\beta_i|^{-1}))$, where

$$k_1 = \frac{3}{2}n + m,$$
$$k_2 = \sum_{i=1}^n s_i^{-1} + \frac{p(1-p)}{4}\sum_{i=1}^n (y_i - \xi_1 s_i^{-1} - \boldsymbol{x}_i'\boldsymbol{\beta})^2 s_i).$$

(c). Sample $(\beta_k|\lambda, \sigma, \boldsymbol{\beta}_{[-k]}, \boldsymbol{s}, \boldsymbol{\nu}; \boldsymbol{y})$ from the truncated normal $N(\mu_k, \sigma_k^2)I(|\beta_k| < \nu_k\sigma^{-1})$, where

$$\mu_k = \sigma_k^2\sigma\xi_2^{-2}\sum_{i=1}^n \tilde{y}_{ik}x_{ij}s_i,$$
$$\sigma_k^2 = \frac{1}{\sigma\xi_2^{-2}\sum_{i=1}^n x_{ik}^2 s_i},$$
$$\tilde{y}_{ik} = y_i - \xi_1 s_i^{-1} - \sum_{j=1, j\neq k}^m x_{ij}\beta_j.$$

(d). Given $(\lambda, \sigma, \boldsymbol{\beta}, \boldsymbol{\nu}; \boldsymbol{y})$, sample

$$s_i \overset{indep}{\sim} \text{GIG}(\sqrt{(\xi_1^2 + 2\xi_2^2)(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2}, (\sigma\xi_1^2\xi_2^{-2} + 2\sigma)^{-1}, -1/2), i = 1, \cdots, n.$$

(e). Given $(\lambda, \sigma, \boldsymbol{\beta}, \boldsymbol{s}; \boldsymbol{y})$, $\nu_1, \cdots, \nu_m$ are independent and $\nu_k$ is distributed as $\text{Exp}(\lambda)I(\nu_k > |\sigma\beta_k|)$. The sampling can be done by

- Sample $\nu_k^*$ from $\text{Exp}(\lambda)$,
- Set $\nu_k = \nu_k^* + |\sigma\beta_k|$.

Since the joint distribution of $\boldsymbol{\beta}$ given other parameters is

$$N((\boldsymbol{X}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}\tilde{\boldsymbol{y}}, (\boldsymbol{X}'\boldsymbol{R}\boldsymbol{X})^{-1})\prod_{i=1}^m I(|\beta_k| < \frac{\nu_k}{\sigma}), \tag{21}$$

where $\boldsymbol{R} = \text{diag}(\xi_2^2 s_1^{-1}, \cdots, \xi_2^2 s_N^{-1})$ and $\tilde{\boldsymbol{y}} = \boldsymbol{y} - \xi_1\boldsymbol{s}^{-1}$, the inverse of $\boldsymbol{X}'\boldsymbol{R}\boldsymbol{X}$ does not exist if $\boldsymbol{X}$ is not full column rank. Hence, we update the elements of $\boldsymbol{\beta}$ one by one in $(c)$ rather than all together.

## 3.2 Bayesian quantile group LASSO

The first generalization of LASSO is the group LASSO, which considers the group structure among the predictors. Suppose that the predictors are grouped into $G$ groups. With the ALD in (4) acting as the likelihood, the conditional prior for $\boldsymbol{\beta}$ is proportional to (16) with $F(\boldsymbol{\beta}) = \sum_{g=1}^G ||\boldsymbol{\beta}_g||_{\boldsymbol{K}_g}$ and $k_2 = 1$. Using SMN for the generalized Laplace distribution (Li et al. [21]), the prior can be represented as,

$$\boldsymbol{\beta}_g|\nu_g \sim N(\boldsymbol{0}, \nu_g^{-1}\boldsymbol{K}_g),\ \nu_g^{-1} \sim \text{Gamma}(\frac{d_g+1}{2}, \frac{\lambda^2\sigma^2}{2}). \tag{22}$$

Similar to BQL, we adopt the Gamma prior, Gamma$(a, b)$, for $\lambda$. We propose setting $a = \max(n-r, 1)$ when SSE=0 and $a = 1$ when SSE>0. For $b$, we set $b = \lambda_{opt}/a$ where $\lambda_{opt}$ is chosen in (5) by some criteria.

With incorporating the latent variables $\boldsymbol{s} = (s_1, \cdots, s_n)$, $\boldsymbol{\nu} = (\nu_1, \cdots, \nu_G)$ through (19) and (22), we marginalize out the latent variables $(\boldsymbol{s}, \boldsymbol{\nu})$ in $(\sigma, \lambda, \boldsymbol{s}, \boldsymbol{\nu}|\boldsymbol{\beta}, \boldsymbol{y})$ and the partially collapsed Gibbs sampling for $(\boldsymbol{\beta}, \sigma, \lambda, \boldsymbol{\nu}, \boldsymbol{s}|\boldsymbol{y})$ is as follows:

(a). Sample $(\sigma|\boldsymbol{\beta}, \lambda; \boldsymbol{y})$ from Gamma$(n+m, \rho_p(\boldsymbol{y} - \boldsymbol{X\beta}) + \lambda \sum_{g=1}^{G} ||\boldsymbol{\beta}_g||_{\boldsymbol{K}_g})$.

(b). Sample $(\lambda|\boldsymbol{\beta}, \sigma; \boldsymbol{y})$ from

$$\text{Gamma}(a + m, \sigma \sum_{i=g}^{G} ||\boldsymbol{\beta}_g||_{\boldsymbol{K}_g} + b).$$

(c). Sample $(\boldsymbol{\beta}_g|\sigma, \lambda, \boldsymbol{\beta}_{[-k]}, \boldsymbol{s}, \boldsymbol{\nu}; \boldsymbol{y})$ from $N(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$, where

$$\boldsymbol{\mu}_g = \boldsymbol{\Sigma}_g \sigma \xi_2^{-2} \sum_{i=1}^{n} \tilde{y}_{ig} x_{ig} s_i,$$

$$\boldsymbol{\Sigma}_g^{-1} = \nu_g \boldsymbol{K}_g + \sigma \xi_2^{-2} \sum_{i=1}^{n} s_i \boldsymbol{x}_{ig} \boldsymbol{x}'_{ig},$$

$$\tilde{y}_{ig} = y_i - \xi_1 s_i^{-1} - \sum_{j=1, j\neq g}^{G} \boldsymbol{x}'_{ij} \boldsymbol{\beta}_j.$$

(d). Given $(\sigma, \lambda, \boldsymbol{\beta}, \boldsymbol{\nu}; \boldsymbol{y})$, sample

$$s_i \overset{indep}{\sim} \text{GIG}(\sqrt{(\xi_1^2 + 2\xi_2^2)/(y_i - \boldsymbol{x}'_i\boldsymbol{\beta})^2}, (\sigma \xi_1^2 \xi_2^{-2} + 2\sigma)^{-1}, -1/2), i = 1, \cdots, n.$$

(e). Given $(\sigma, \lambda, \boldsymbol{\beta}, \boldsymbol{s}; \boldsymbol{y})$, sample

$$\nu_g \overset{indep}{\sim} \text{GIG}(\sigma\lambda(\boldsymbol{\beta}'_g \boldsymbol{K}_g \boldsymbol{\beta}_g)^{-1/2}, \sigma^{-2}\lambda^{-2}, -1/2), \quad g = 1 \cdots, G$$

### 3.3 Bayesian bridge quantile regression

The second generalization of the LASSO is penalized quantile regression by solving (5) with penalty (8), which is an analogy of the bridge regression (Frank and Friedman [9]). Since $k_1 \neq k_2$ for bridge quantile regression in general, the prior we adopt for $(\boldsymbol{\beta}, \sigma, \lambda)$ is,

$$p(\boldsymbol{\beta}|\lambda, \sigma) = \prod_{i=1}^{m} \frac{\lambda^{\frac{1}{q}} \sigma}{2\Gamma(\frac{1}{q} + 1)} \exp(-\lambda |\sigma \beta_i|^q), \; p(\sigma) \propto \frac{1}{\sigma}, \; \lambda \sim \text{Gamma}(a_0, b_0). \tag{23}$$

By Theorem 2(b), the sufficient and necessary condition for the posterior propriety of $(\boldsymbol{\beta}, \sigma, \lambda|\boldsymbol{y})$ is $a_0 q + r > n$ if SSE=0. In a similar spirit to BQL, we choose $a_0 = \max((n-r)/q, 1)$ for SSE=0 and $a_0 = 1$ for SSE>0. For $b_0$, one could employ the empirical Bayes approach (Casella [6]) to estimate $b_0$, where the updating rule for $b_0$ is

$$b_0^{(k)} = \frac{a_0}{E_{b_0^{(k-1)}}(\lambda|\boldsymbol{y})}, \tag{24}$$

and $b_0^{(k)}$ is the estimate of $b_0$ at the $k$th iteration. Empirical Bayes approach may be computationally expensive because many Gibbs sampler runs are needed. As an alternative choice, we set $b_0$ to some small values in the sense of a flat prior. Our sensitivity analysis in Section 4.2 showed that the posterior estimates are relatively robust to the choice of $a_0$ and small $b_0$.

We develop the partially collapsed Gibbs sampling to draw the posterior samples from $(\boldsymbol{\beta}, \sigma, \lambda|\boldsymbol{y})$. Mallick and Yi [26] extended (20) as below,

$$\frac{\tau^{\frac{1}{q}}}{2\Gamma(\frac{1}{q}+1)} \exp(-\tau|t|^q) = \int_{\nu > |t|^q} \frac{\tau^{\frac{1}{q}+1}}{2\nu^{\frac{1}{q}} \Gamma(1 + \frac{1}{q})} \nu^{\frac{1}{q}} \exp(-\tau\nu) d\nu. \tag{25}$$

Using (25), generalized Gaussian (GG) prior can be represented as the SMU. Specifically, if the density of $X$ is proportional to $\exp(-\tau|x|^q)$ given $q > 0$ and $\tau > 0$,

$$X \overset{d}{=} u\nu^{\frac{1}{q}}, \tag{26}$$

where $u \sim U[-1, 1]$ and $\nu \sim$ Gamma$(1 + 1/q, \tau)$ are independent. In fact, this representation is universal for the unimodal density by Khinchine's theorem (Devroye [8], page 172). It implies that a random variable $T$ is unimodal if and only if it is distributed as $SY$, $S \sim U[0, 1]$ and the distribution of $Y$ depends on $T$. $S$ and $Y$ are independent random variables.

With the latent variables $\boldsymbol{s} = (s_1, \cdots, s_n)$ from the likelihood and latent variables $\boldsymbol{\nu} = (\nu_1, \cdots, \nu_m)$ from the prior using (25), all the parameters can be sampled efficiently except for $\lambda$. Similar to BQL, the procedure of partially collapsed Gibbs sampling is as follows,

(a) Sample $(\lambda|\sigma, \boldsymbol{\beta}; \boldsymbol{y})$ from

$$\text{Gamma}(m/q + a_0, \sum_{i=1}^{m} |\sigma\beta_i|^q + b_0).$$

(b) Sample $(\sigma|\boldsymbol{\beta}, \boldsymbol{s}, \boldsymbol{\nu}; \boldsymbol{y})$ from truncated Gamma distribution, Gamma$(k_3, k_4) I(\sigma < \min_i \frac{\nu_i^{1/q}}{|\beta_i|})$, where

$$k_3 = \frac{3}{2}n + m,$$

$$k_4 = \sum_{i=1}^{n} s_i^{-1} + \frac{p(1-p)}{4} \sum_{i=1}^{n} (y_i - \xi_1 s_i^{-1} - \boldsymbol{x}'_i\boldsymbol{\beta})^2 s_i.$$

(c). Sample $(\beta_k|\lambda, \sigma, \boldsymbol{\beta}_{[-k]}, \boldsymbol{s}, \boldsymbol{\nu}; \boldsymbol{y})$ from the truncated normal $N(\mu_k, \sigma_k^2)I(|\beta_k| < \nu_k^{1/q}\sigma^{-1})$, where

$$\mu_k = \sigma_k^2 \sigma \xi_2^{-2} \sum_{i=1}^{n} \tilde{y}_{ik} x_{ij} s_i,$$
$$\sigma_k^2 = \frac{1}{\sigma \xi_2^{-2} \sum_{i=1}^{n} x_{ik}^2 s_i},$$
$$\tilde{y}_{ik} = y_i - \xi_1 s_i^{-1} - \sum_{j=1, j\neq k}^{m} x_{ij}\beta_j.$$

(d). Given $(\lambda, \sigma, \boldsymbol{\beta}, \boldsymbol{\nu}; \boldsymbol{y})$, sample

$$s_i \overset{indep}{\sim} \mathrm{GIG}(\sqrt{(\xi_1^2 + 2\xi_2^2)(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2}, (\sigma\xi_1^2\xi_2^{-2} + 2\sigma)^{-1}, -1/2), i = 1, \cdots, n.$$

(e). Given $(\lambda, \sigma, \boldsymbol{\beta}, \boldsymbol{s}; \boldsymbol{y})$, $\nu_1, \cdots, \nu_m$ are independent and $\nu_k$ is distributed as $\mathrm{Exp}(\lambda)I(\nu_k > |\sigma\beta_k|^q)$. The sampling can be done by

- Sample $\nu_k^*$ from $\mathrm{Exp}(\lambda)$,
- Set $\nu_k = \nu_k^* + |\sigma\beta_k|^q$.

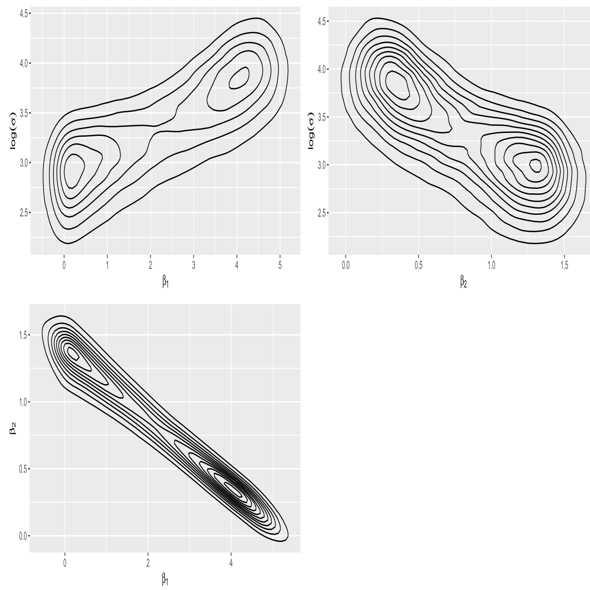

*Figure 1. Pairwise contour plot of a posterior density of $(\boldsymbol{\beta}, \log(\sigma))$ with $p = 0.5$. The logarithm of $\sigma$ is used only because it provides better visual scaling.*

## 4. SIMULATION STUDIES

### 4.1 Multimodality of the joint posterior

In this subsection, we conduct the simulation studies to demonstrate that the unconditional prior for $\boldsymbol{\beta}$ can result in multimodality of the joint posterior. We generate the data with the linear model as follows

$$(27) \qquad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim \mathrm{ALD}(0, \sigma^{-1} = 0.03, p = 0.5).$$

We take $p = 0.5, \boldsymbol{\beta} = (5, 0), \eta = 3,\ \mathrm{tr}(\boldsymbol{X}'\boldsymbol{X}) = 1$ and $a = b = 0$ in (10) (Bimodality can occur even if the prior for $\sigma$ is proper). In Figure 1, the joint posterior of $(\beta_1, \sigma)$, $(\beta_2, \sigma)$ and $(\beta_1, \beta_2)$ exhibits severe bimodality. For $\beta_1$, the center of left corner is near 0, which is due to the shrinkage of the LASSO penalty. In contrast, for $\beta_2$, the posterior is still bimodal even if the true value is zero. We also conduct the simulations when $p = 0.1$ and $p = 0.9$ (not shown here), both display the bimodality.

### 4.2 Sensitivity tests of the hyperparemeters

In this subsection, we test the sensitivity of hyperparameters of Gamma prior in (10) and (23) on the posterior estimates. We equally divide $x \in [-2, 2]$ into 50 pieces and the data are generated from

$$(28) \qquad y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i,\ \epsilon_i \overset{iid}{\sim} \mathrm{ALD}(0, \sigma^{-1} = 0.03, p = 0.5),\quad i = 1, \cdots, 50,$$

with $\boldsymbol{x}_i = ((1 + e^{-4(x_i - 0.3)})^{-1}, (1 + e^{3(x_i-0.2)})^{-1}, (1 + e^{-4(x_i-0.7)})^{-1}, (1 + e^{5(x_i-0.8)})^{-1})'$ and $\boldsymbol{\beta} = (1, 1, 1, 1)'$. It indicates that the true curve is

$$f(x) = (1 + e^{-4(x-0.3)})^{-1} + (1 + e^{3(x-0.2)})^{-1} + (1 + e^{-4(x-0.7)})^{-1} + (1 + e^{5(x-0.8)})^{-1}.$$

In fact, this function was utilized in Jullion and Lambert [13] to test the sensitivity of hyperparameters of the Gamma prior on the scale component in Bayesian P-spline. Figure 2 showed one representative sensitivity analysis. In the figure, Graphs (a) and (b) used BQL, which is proposed by Li et al. [21] and employed the unconditional prior (14) for $\boldsymbol{\beta}$. Here, note that the sensitivity refers to the hyperparameters in Gamma prior in (10). Without of the loss of generality, in Graphs (a) and (b), we keep $a = 0.1$ fixed with $b$ varied and $b = 0.1$ fixed with a varied $a$, respectively. For both cases, we set $c = d = 0.1$. These values are chosen the same as those recommended by Li et al. [21] for an illustrative and comparative purpose in terms of potential limitations of using a Gamma prior. As we can see from the top panel in Figure 2, the fitted curves fluctuate a lot according to the choice of hyperparameters for $a$ and $b$. In contrast, Graphs (c) and (d) are generated from BBQR with the prior (23). Here, the hyperparameter refers to $\lambda$ in (23). Follow the recommendation from Park and Casella [27], we choose $a_0 = \max(n - r, 1)$ when SSE=0 and $a_0 = 1$ when SSE>0 in Graph (c). We keep $a_0 = 1$ fixed with $b_0$ varied. Also, to explore the sensitivity of hyperparemeter $a_0$, without the loss of generality, in Graph (d), we keep $b_0 = 0.1$ fixed and $a_0$ varied, respectively. In this case, the resulting fitted curves are relatively robust to the choice of hyperparameter $a_0$ and $b_0$.

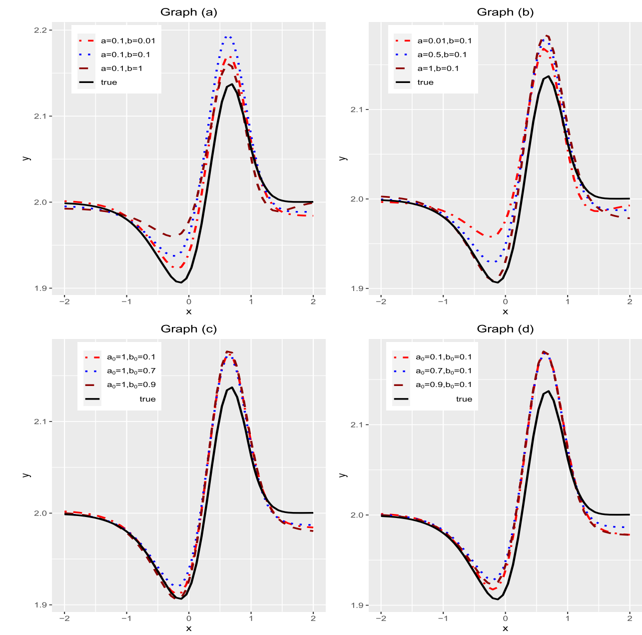


*Figure 2. All the curves are fitted with $p = 0.5$. The true curve is the black solid line for four subgraphs. Graph (a): Gamma prior (10) for $\sigma$ with $a = 0.1, b = 0.01, c = d = 0.1$ (dash-dotted), $a = 0.1, b = 0.1, c = d = 0.1$ (dotted) and $a = 0.1, b = 1, c = d = 0.1$ (dash). Graph (b): Gamma prior (10) for $\sigma$ with $a = 0.01, b = 0.1, c = d = 0.1$ (dash-dotted), $a = 0.5, b = 0.1, c = d = 0.1$ (dotted) and $a = 1, b = 0.1, c = d = 0.1$ (dash). Graph(c): Gamma prior (23) for $\lambda$ with $a_0 = 1, b_0 = 0.1$ (dash-dotted), $a_0 = 1, b_0 = 0.7$ (dotted) and $a_0 = 1, b_0 = 0.9$ (dashed). The curves are fitted when $q = 0.5$. Graph(d): Gamma prior (23) for $\lambda$ with $a_0 = 0.1, b_0 = 0.1$ (dash-dotted), $a_0 = 0.7, b_0 = 0.1$ (dotted) and $a_0 = 0.9, b_0 = 0.1$ (dashed). The curves in Graphs (c)–(d) are fitted when $q = 0.5$.*

## 4.3 Comparisons with the original Bayesian approaches

In this subsection, we compare our new methods for Bayesian regularized quantile regression with the existing Bayesian regularized quantile regression methods using Monte Carlo simulation studies. The six candidate methods we employ are listed as below,

(a). New Bayesian quantile LASSO (NBQL).
(b). Original Bayesian quantile LASSO (OBQL) in Li et al. [21].
(c). New Bayesian quantile group LASSO (NBQGL).
(d). Original Bayesian quantile group LASSO (OBQGL) in Li et al. [21].
(e). New Bayesian bridge quantile regression (NBBQR).
(f). Original Bayesian bridge quantile regression (OBBQR) in Alhamzawi and Algamal [1].

The data in the simulation studies are generated by

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + a_p + \mu_i, i = 1, \cdots, n, \tag{29}$$

where $\mu_i$ is the random error and $a_p$ is utilized to make the $p$th quantile of $\mu_i$ zero.

#### 4.3.1 Independent and identically distributed random errors

We consider four simulation studies in the i.i.d random error settings.

- Simulation 1: $\boldsymbol{\beta} = (3, 1.5, 0, 0, 2, 0, 0, 0)$ for the sparse case.
- Simulation 2: $\boldsymbol{\beta} = (0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85)$ for the dense case.
- Simulation 3: $\boldsymbol{\beta} = (\underbrace{5, \cdots, 5}_{5}, \underbrace{0, \cdots, 0}_{20}, \underbrace{5, \cdots, 5}_{5})$ for more predictors than sample size case.
- Simulation 4: $\boldsymbol{\beta} = ((-1.2, 1.8, 0), (0, 0, 0), (0.5, 1, 0), (0, 0, 0), (1, 1, 0))$ for group structure case.

In Simulations 1–3, each row in $\boldsymbol{X}$ is generated independently from $N(\mathbf{0}, \boldsymbol{\Sigma}_1)$, where the $(i, j)$ element of $\boldsymbol{\Sigma}_1$ is $0.5^{|i-j|}$. In Simulation 4, we first draw the latent variable $\boldsymbol{Z} = (Z_1, \cdots, Z_5)$ from $N(\mathbf{0}, \boldsymbol{\Sigma})$, where the $(i, j)$ element of $\boldsymbol{\Sigma}$ is $0.5^{|i-j|}$. Then, each $Z_j$ is trichotomized as 0, 1 and 2, which depends on whether it is smaller than $\Phi^{-1}(1/3)$, between $\Phi^{-1}(1/3)$ and $\Phi^{-1}(2/3)$, or larger than $\Phi^{-1}(2/3)$. Here, $\Phi(x)$ is the cumulative function of standard normal distribution. The rows of $\boldsymbol{X}$ are given by $(I(Z_1 = 0), I(Z_1 = 1), I(Z_1 = 2), \cdots, I(Z_5 = 0), I(Z_5 = 1), I(Z_5 = 2))$. Within each simulation study, we consider four different choices for the distribution of $\mu_i$.

- normal distribution with $N(0, 2^2)$.
- normal mixture distribution, $0.1N(0, 1) + 0.9N(0, 2^2)$.
- ALD$(0, \sigma^{-1} = 1, p = 0.5)$, so that the variance of random error is 4.
- ALD mixture with 0.1ALD$(0, \sigma^{-1} = 1, p = 0.5)$+0.9 ALD$(0, \sigma^{-1} = \sqrt{2}, p = 0.5)$.

We simulate a training dataset with 20 observations, a validation dataset with 20 observations to obtain the optimal $\lambda_{opt}$ by 10-fold cross validation in the BQL and BQGL and a testing dataset with 200 observations. We consider three choices of $p$, $p = 0.1$, $p = 0.3$ and $p = 0.5$. In Simulations 1–3, there is no group structure in the predictors, so BQGL reduces to BQL. In Simulation 4, as suggested in Yuan and Lin [37], we choose $\boldsymbol{K}_g = d_g \boldsymbol{I}_{d_g}$ where $d_g$ is the dimension of $\boldsymbol{\beta}_g$. For BBQR, $q \in (0, 1)$ enjoys many desirable statistical properties such as oracle, sparsity, and unbiasedness. We choose $q = 0.5$ to preserve these properties and for a representative purpose (Xu et al. [35]). Although the unimodality of BBQR under is only proved for $q \geq 1$, we ensure the unimodality of posterior in the simulation study by directly plotting the density functions of each $\beta_i$ in $\boldsymbol{\beta}$. For an illustration, we offer marginal density plots for the first element

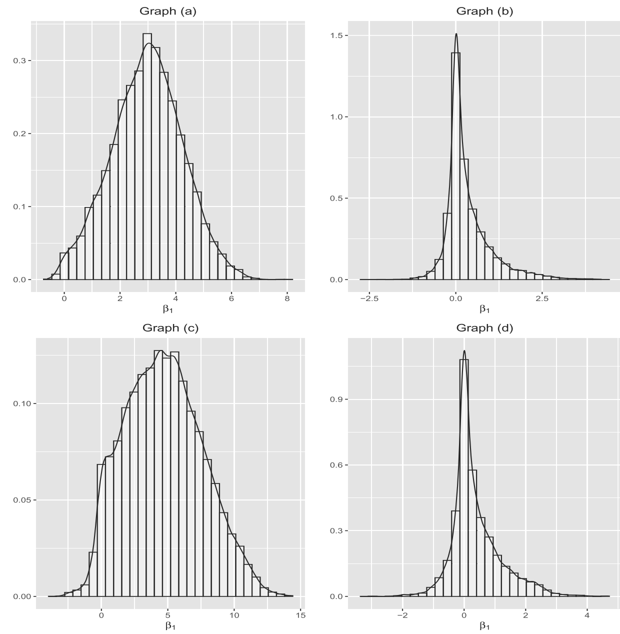


*Figure 3. Marginal density plot of $\beta_1 \in \boldsymbol{\beta}$ with ALD random error and $p = 0.1$. Graphs (a)–(d) corresponding to Simulations 1–4, respectively.*

$\beta_1$ in $\boldsymbol{\beta}$ as a visual aid across Simulations 1–4. In detail, we used the first dataset among 50 datasets considered in Section 4.3.1 with ALD random errors and quantile level 0.1. From Figure 3, we may find that all the marginal densities of $\beta_1$ are unimodal. It is worth investigating whether unimodality holds for $q \in (0, 1)$ from a theoretical viewpoint. We evaluate the performance by median of mean absolute deviation (MMAD) and median of mean square error (MMSE) based on the 50 replications,

$$\text{MMAD} = \text{median}\left(\frac{1}{200}\sum_{i=1}^{200} |\boldsymbol{x}_i'\hat{\boldsymbol{\beta}} - \boldsymbol{x}_i'\boldsymbol{\beta}|\right),$$

$$\text{MMSE} = \text{median}\left(\frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{m}\right).$$

In Simulation 1, we may find NBQL have better performance compared with OBQL in terms of MMAD (12 out of 12 comparisons) and MMSE (9 out of 12 comparisons). NBBQR have smaller MMAD in 12 out of 12 comparisons and MMSE in 9 out of 12 comparisons. Moreover, our methods tend to outperform original Bayesian approaches in more and more cases as $p$ is away from 0.5. Similar findings can be found for Simulation 2. In Simulation 3, our two methods defeat the original approaches for all the selected quantiles except two cases, which show great advantages of our method in the large $p$ and small $n$ case. Moreover, for the overall performances in Simulations 1–3, there seems to be a tendency that all the methods behave better when the selected quantile is close to 0.5.

Simulation 4 corresponds to the case with group structures among the predictors. BQGL gives the better estimations than the other methods and NBQGL tends to behave the best in the most of times. For BQL and BBQR, NBQL and NBBQR tend to have the smaller MMAD while OBQL and OBBQR performs better in terms of MMSE in each comparison.

#### 4.3.2 Heterogeneous random errors

We consider the case with non-i.i.d random errors. The data set were simulated according to the model proposed in Wu and Liu [34],

$$y_i = 1 + x_{1i} + x_{2i} + x_{3i} + (1 + x_{3i})\epsilon_i, \tag{30}$$

where $x_{1i} \sim N(0, 1)$, $x_{3i} \sim$U(0,1) and $x_{2i} = x_{1i} + x_{3i} + z_i$, $z_i \sim N(0, 1)$, $\epsilon_i \sim N(\mu, 1)$, with $\mu$ chosen so that $p$th quantile zero. There are also five more independent noise random variables, $x_4, \cdots, x_8$, distributed as $N(0, 1)$. Each simulated sample is partitioned into a training set with 20 observations, a validation set with 20 observations, and a testing set with 200 observations. Methods are evaluated based on MMAD and the test error, which refers to the median of average check loss (MACL) on the testing dataset,

$$\text{MACL} = \text{median}(\text{mean}(\rho_p(\boldsymbol{y} - \boldsymbol{X\beta}))).$$

From Table 7, our methods have the better predictions in each comparison in terms of MMAD for all the selected quantiles and have the smaller MACLs in 4 out of 6 comparisons. Moreover, we find that NBQL has the smallest error in terms of MMAD for each quantile, which reveals the efficiency of our new methods.

## 5. BOSTON HOUSING DATASET

We evaluate the performances of the six candidate methods, NBQL, OBQL, NBBQR, OBBQR, quantile LASSO (QL) and LASSO on the Boston Housing data. Boston Housing dataset was first analyzed by Harrison and Rubinfeld [10] in a study on the influence of clean air on house prices and became a standard dataset in investigating the normality assumption of residuals for robust estimation methods. The Boston housing dataset contains the census tracts in the greater Boston area with 506 cases. The response variable is defined as the logarithm of the median housing price (LMHP) in the tract associated with 13 explanatory variables representing various measurements: per capita crime (CRIM), proportions of residential land zoned for lots over 25,000 square feet per town (ZN), proportions of nonretail business acres per town (INDUS), a factor indicating whether tract borders Charles River (CHAS), nitric oxides concentration (parts per 10 million) per town (NOX), average numbers of rooms per dwelling (RM), proportions of

Table 3. MMADs and MMSEs from Simulation 1 with $p = 0.1, 0.3$ and 0.5. In the parentheses are standard deviations of the MMADs and MMSEs obtained by 500 bootstrap resampling. The bold numbers correspond to the smallest MMAD and MMSE in each comparison

| | Method | Error distribution | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Normal | | Normal mixture | | ALD | | ALD mixture | |
| | | MMAD | MMSE | MMAD | MMSE | MMAD | MMSE | MMAD | MMSE |
| $p = 0.1$ | NBQL | **1.143**(0.0587) | **.3593**(.0840) | **1.3077**(.0387) | **.4811**(.0334) | **1.6193**(.1398) | **.8210**(.1355) | **2.0468**(.1515) | **.9672**(.1714) |
| | OBQL | 1.165(.0493) | .3653(.0701) | 1.3331(.0661) | .4900(.0628) | 1.6461(.1382) | .8390(.1385) | 2.1180(.1244) | 1.0073(.1492) |
| | NBBQR($q = 0.5$) | **1.239**(.0566) | .4339(.0606) | **1.2530**(.0636) | **.4577**(.0546) | **1.7492**(.0845) | **.8291**(.0891) | **2.1165**(.1424) | **1.0258**(.1483) |
| | OBBQR($q = 0.5$) | 1.241(.0539) | **.4330**(.0577) | 1.2578(.0603) | .4652(.0558) | 1.7555(.1094) | .8364(.1094) | 2.2035(.1268) | 1.0753(.1311) |
| $p = 0.3$ | NBQL | **1.0315**(.0566) | **.2986**(.0589) | **1.1913**(.0653) | .3872(.0571) | **1.2348**(.1120) | **.3908**(.1171) | **1.7735**(.0750) | **.6927**(.0742) |
| | OBQL | 1.0452(.0553) | .2995(.0618) | 1.1951(.0747) | **.3772**(.0713) | 1.2692(.1107) | .3979(.1209) | 2.0663(.1797) | .8331(.1910) |
| | NBBQR($q = 0.5$) | **1.0444**(.0689) | .3169(.0680) | **1.1745**(.0519) | **.3936**(.0529) | **1.3091**(.1452) | **.4013**(.1439) | **1.8983**(.1174) | .8438(.1649) |
| | OBBQR($q = 0.5$) | 1.0959(.0664) | **.2929**(.0707) | 1.1822(.0526) | .4089(.0498) | 1.3152(.1569) | .4606(0.1554) | 2.1255(.1831) | **.7710**(.1382) |
| $p = 0.5$ | NBQL | **1.1453**(.0522) | **.3318**(.0455) | **1.0809**(.1102) | **.3069**(.0653) | **1.0407**(.0935) | .2996(.0968) | **1.4219**(.0882) | .5257(.0745) |
| | OBQL | 1.1620(.0421) | .3360(0.0506) | 1.0934(.1052) | .3187(0.1041) | 1.0611(.0974) | **.2928**(.1072) | 1.4573(.0725) | **0.4625**(.0920) |
| | NBBQR($q = 0.5$) | **1.1219**(.0551) | **.3499**(.0609) | **1.0725**(.0692) | **.2872**(.0664) | **1.0784**(.1024) | **.3124**(.1052) | **1.4723**(.0957) | **.5293**(.0886) |
| | OBBQR($q = 0.5$) | 1.1349(.0426) | .3601(.0500) | 1.1092(.1238) | .2879(.0652) | 1.0889(.1112) | .3159(.0984) | 1.4823(.0892) | .5592(.0863) |

Table 4. MMADs and MMSEs from Simulation 2 with $p = 0.1, 0.3$ and 0.5. In the parentheses are standard deviations of the MMADs and MMSEs obtained by 500 bootstrap resampling. The bold numbers correspond to the smallest MMAD and MMSE in each comparison

| | Method | Error distribution | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Normal | | Normal mixture | | ALD | | ALD mixture | |
| | | MMAD | MMSE | MMAD | MMSE | MMAD | MMSE | MMAD | MMSE |
| $p = 0.1$ | NBQL | **1.5302**(.0770) | **.3427**(.0801) | **1.1499**(.0474) | **.2498**(.0564) | **1.4271**(.0727) | **.3469**(.0741) | **1.9286**(.1233) | .5161(.1269) |
| | OBQL | 1.5568(.1051) | .3766(.1051) | 1.1569(.0562) | .2599(.0605) | 1.4501(.0501) | .3485(.1069) | 2.1043(.1297) | **.5011**(.1234) |
| | NBQBR($q = 0.5$) | **1.7043**(.0985) | **.4443**(.1031) | **1.3548**(.0562) | **.3644**(.0606) | **1.6736**(.0572) | .4952(.0801) | **2.1849**(.1167) | **.6159**(.1126) |
| | OBQBR($q = 0.5$) | 1.7164(.1122) | .4479(.1254) | 1.3812(.0537) | .3901(.0561) | 1.7117(.0601) | **.4667**(.0733) | 2.3183(.1203) | .6187(.1222) |
| $p = 0.3$ | NBQL | **1.0109**(.0719) | **.2509**(.0588) | **1.0446**(.0477) | **.2347**(.0625) | 1.2421(.0693) | **.3166**(.0578) | **1.7522**(.0647) | .4699(.1011) |
| | OBQL | 1.0398(.06221) | .2605(.0687) | 1.0503(.0547) | .2464(.0505) | **1.2338**(.0631) | .3344(.0629) | 1.9584(.1035) | **.4223**(.07877) |
| | NBBQR($q = 0.5$) | **1.2012**(.0745) | **.3485**(.06698) | **1.1695**(.0599) | **.3248**(.0622) | **1.5116**(.0774) | .4266(.0778) | **1.9734**(.0901) | **.5281**(.0929) |
| | OBBQR($q = 0.5$) | 1.2166(.0676) | .3593(.0722) | 1.2157(.0667) | .3294(.0668) | 1.5173(.0836) | **.4091**(.0956) | 2.2444(.1189) | .6135(.1166) |
| $p = 0.5$ | NBQL | **1.0277**(.0572) | **.2399**(.0582) | 1.0129(.0385) | **.2368**(.0447) | **1.1301**(.0335) | **.3246**(.0549) | 1.3620(.1039) | **.3055**(.3055) |
| | OBQL | 1.0556(.0470) | .2703(.0595) | **1.0095**(.0511) | .2523(.0675) | 1.1319(.0437) | .3464(.0691) | **1.3474**(.0884) | .3276(.3276) |
| | NBBQR($q = 0.5$) | **1.2007**(.0494) | **.3656**(.0551) | **1.1631**(.0581) | **.3381**(.0504) | 1.3578(.0425) | **.4413**(.0636) | 1.6641(.1031) | **.4615**(.4615) |
| | OBBQR($q = 0.5$) | 1.2189(.0527) | .3903(.0634) | 1.1977(.0461) | .3397(.0675) | **1.3318**(.0445) | .4471(.0619) | **1.6431**(.0843) | .4895(.1042) |

Table 5. MMADs and MMSEs from Simulation 3 with $p = 0.1, 0.3$ and 0.5. In the parentheses are standard deviations of the MMADs and MMSEs obtained by 500 bootstrap resampling. The bold numbers correspond to the smallest MMAD and MMSE in each comparison

| | Method | Error distribution | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Normal | | Normal mixture | | ALD | | ALD mixture | |
| | | MMAD | MMSE | MMAD | MMSE | MMAD | MMSE | MMAD | MMSE |
| $p = 0.1$ | NBQL | **7.4779**(.3780) | **3.1625**(.3863) | **5.8921**(.2509) | **2.1615**(.2029) | **6.2122**(.1946) | **2.2825**(.1987) | **9.3929**(.1872) | **3.6611**(.2053) |
| | OBQL | 7.7125(.4873) | 3.2823(.4799) | 6.0044(.2148) | 2.3335(.2373) | 6.2224(.2961) | 2.3592(.2912) | 9.6099(.2511) | 3.7234(.2356) |
| | NBBQR($q = 0.5$) | **7.6591**(.3609) | **4.4418**(.3489) | **5.2168**(.2694) | **2.4616**(.2100) | **6.5778**(.2192) | **3.4921**(.2200) | **9.8055**(.2991) | **4.7411**(.3026) |
| | OBBQR($q = 0.5$) | 8.8848(.6677) | 5.5295(.6615) | 5.3179(.1972) | 2.5577(.2636) | 7.0862(.2457) | 3.8539(.2536) | 9.9535(.2155) | 4.8084(.2137) |
| $p = 0.3$ | NBQL | **6.4815**(.2007) | **2.1457**(.2253) | **5.7440**(.1465) | **2.0637**(.1475) | **6.4395**(.2652) | **2.4138**(.2552) | **5.0492**(.2816) | 1.7124(.1672) |
| | OBQL | 6.7701(.1747) | 2.7660(.1715) | 5.9356(.2802) | 2.3216(.2503) | 6.5778(.2070) | 2.6229(.1945) | 5.3888(.1639) | **1.6195**(.2791) |
| | NBBQR($q = 0.5$) | **6.7237**(.3166) | **3.2115**(.1782) | **5.0394**(.2699) | **2.1982**(.2548) | **6.8224**(.2612) | **3.1276**(.2725) | **5.4150**(.4051) | **2.3685**(.4106) |
| | OBBQR($q = 0.5$) | 6.9624(.2093) | 3.2506(.3254) | 5.4999(.4102) | 2.7484(.3764) | 7.0062(.2852) | 3.6604(.2986) | 5.8272(.3733) | 2.5466(.3603) |
| $p = 0.5$ | NBQL | **6.4848**(.1327) | 2.1096(.1219) | **5.4239**(.2592) | **1.8175**(.1966) | **6.6349**(.2329) | **2.3996**(.2639) | **6.9855**(.2221) | **2.7138**(.2285) |
| | OBQL | 6.6855(.1439) | **1.9313**(.1721) | 5.5085(.1814) | 1.9139(.2745) | 6.8397(.2434) | 2.8514(.2310) | 7.1836(.2401) | 2.7290(.2381) |
| | NBBQR($q = 0.5$) | **6.7548**(.1921) | **2.8133**(.1934) | **4.8667**(.1790) | **1.8850**(.1840) | **7.2224**(.2380) | **3.5822**(.2221) | **7.2151**(.3925) | **3.0601**(.3996) |
| | OBBQR($q = 0.5$) | 7.0114(.1310) | 2.9513(.1100) | 4.9570(.1821) | 1.9407(.1915) | 8.0200(.4426) | 4.3862(.4527) | 7.4434(.4326) | 3.2155(.4189) |

*Table 6. MMADs and MMSEs from Simulation 4 with $p = 0.1, 0.3$ and 0.5. In the parentheses are standard deviations of the MMADs and MMSEs obtained by 500 bootstrap resampling. The bold numbers correspond to the smallest MMAD and MMSE in each comparison*

| | Method | Error distribution | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Normal | | Normal mixture | | ALD | | ALD mixture | |
| | | MMAD | MMSE | MMAD | MMSE | MMAD | MMSE | MMAD | MMSE |
| $p = 0.1$ | NBQL | **1.1473**(.0478) | .5087(.0202) | **1.2381**(.0301) | .5106(.0134) | **1.2594**(.0403) | .5003(.0141) | **1.1545**(.0817) | **.5257**(.0152) |
| | OBQL | 1.2155(.0129) | **.4415**(.0176) | 1.2396(.0447) | **.4391**(.0186) | 1.4690(.1364) | **.4991**(.0304) | 1.3803(.2261) | .6099(.0769) |
| | NBBQR($q = 0.5$) | **1.1623**(.0425) | .5186(.0176) | **1.2286**(.0202) | .5229(.0084) | 1.3451(.0968) | .5180(.0105) | **2.7796**(.1235) | **.9075**(.1235) |
| | OBBQR($q = 0.5$) | 1.2175(.0314) | **.4976**(.0224) | 1.2620(.1408) | **.4698**(.0335) | **1.2660**(.0598) | **.5106**(.0298) | 2.8942(.1213) | .9649(.1213) |
| | NBQGL | **1.0792**(.0424) | **.4239**(.0169) | **1.2169**(.0878) | **.4004**(.0288) | **1.2257**(.0182) | **.4633**(.0968) | **1.1483**(.0533) | **.5282**(.0036) |
| | OBQGL | 1.1130(.0307) | .4253(.0307) | 1.2179(.0586) | .4060(.1408) | 1.2744(.0894) | .5080(.1364) | 1.1855(.1394) | .5408(.0347) |
| $p = 0.3$ | NBQL | **1.2329**(.0102) | .5179(.0041) | **1.1372**(.0451) | .4652(.0192) | **1.2473**(.0182) | .4932(.0199) | **1.1394**(.0473) | **.4986**(.0081) |
| | OBQL | 1.2343(.0303) | **.4916**(.0106) | 1.1940(.0282) | **.4169**(.0184) | 1.2666(.0519) | **.3887**(.0315) | 1.1544(.0146) | .5122(.0041) |
| | NBBQR($q = 0.5$) | **1.2509**(.0226) | .5271(.0016) | **1.1941**(.0393) | .5107(.0085) | **1.2471**(.0473) | **.4478**(.0335) | **1.1437**(.0023) | .5256(.0022) |
| | OBBQR($q = 0.5$) | 1.2518(.0706) | **.5206**(.0058) | 1.2046(.0179) | **.4477**(.0164) | 1.3015(.0802) | .5197(.0109) | 1.1484(.0082) | **.5203**(.0043) |
| | NBQGL | 1.2302(.0013) | **.4351**(.0171) | **1.047**(.0751) | .3964(.0122) | **1.2224**(.0316) | .3889(.0473) | **.8983**(.0718) | .4704(.0718) |
| | OBQGL | **1.2289**(.0043) | .4478(.0706) | 1.0858(.0664) | **.3816**(.0751) | 1.2297(.0042) | **.3776**(.0802) | 1.0081(.0829) | **.4435**(.0829) |
| $p = 0.5$ | NBQL | **1.0285**(.0576) | .3421(0.0405) | 1.0254(.0651) | .4049(.0241) | **1.0675**(.0344) | .4651(.0146) | **1.1290**(.0209) | .4834(.0097) |
| | OBQL | 1.0331(.0442) | **.3323**(.0307) | **.8664**(.0805) | **.3545**(.0266) | 1.1317(.0336) | **.4225**(.0208) | 1.1378(.0207) | **.4637**(.0151) |
| | NBBQR($q = 0.5$) | **.9761**(.0871) | .3663(.0411) | **.8728**(.0992) | .4049(.0386) | **1.0591**(.0624) | **.4497**(.0209) | 1.1339(.0091) | .5222(.0047) |
| | OBBQR($q = 0.5$) | 1.0286(.0397) | **.3522**(.0296) | .9466(.0805) | **.3898**(.0433) | 1.1729(.0283) | .5020(.0101) | **1.1191**(.0208) | **.5118**(.0086) |
| | NBQGL | **.8897**(.0657) | **.3345**(.0442) | **.8067**(.0605) | **.3407**(.0678) | **1.0133**(.0485) | **.3855**(.0189) | **.5379**(.0258) | .4565(.0258) |
| | OBQGL | .9466(.0621) | .3367(.0397) | .8158(.0733) | .3415(.0605) | 1.0758(.0429) | .4046(.0429) | .7223(.0350) | **.4395**(.0350) |

*Table 7. MMADs and MACLs from heterogeneous random errors with $p = 0.3, 0.5$ and 0.7. In the parentheses are standard deviations of the MMADs and MMSEs obtained by 500 bootstrap resampling. The bold numbers correspond to the smallest MMAD and MACL in each comparison*

| | Method | MMAD | MACL |
|---|---|---|---|
| $p = 0.3$ | NBQL | **.9656**(.0979) | **.5774**(.0522) |
| | OBQL | 1.0525(.0854) | .5897(.0682) |
| | NBBQR($q = 0.5$) | **1.1056**(.1476) | **.5982**(.0748) |
| | OBBQR($q = 0.5$) | 1.1108(.1288) | .6039(.0921) |
| $p = 0.5$ | NBQL | **.7618**(.0455) | .3855(.0243) |
| | OBQL | .7711(.0521) | **.3809**(.0187) |
| | NBBQR($q = 0.5$) | **.7881**(.0631) | **.3941**(.0287) |
| | OBBQR($q = 0.5$) | .8431(.0561) | .4215(.0258) |
| $p = 0.7$ | NBQL | **.7865**(.0839) | **.3234**(.0243) |
| | OBQL | .8447(.0847) | .3301(.0299) |
| | NBBQR($q = 0.5$) | **.7905**(.1071) | .3149(.0361) |
| | OBBQR($q = 0.5$) | .8037(.0884) | **.2947**(.0355) |

owner-occupied units built prior to 1940 (AGE), weighted distances to five Boston employment centers (DIS), index of accessibility to radial highways per town (RAD), full-value property tax rate per USD 10,000 per town (TAX), pupil-teacher ratios per town (PTRATIO), $1000(bk-0.63)^2$ where bk is the proportion of blacks by town (BLACK) and percentage values of lower status population (LSTAT). This dataset is available in the R package MASS.

We consider three choices of $p$, $p = 0.4$, 0.5 and 0.7 to be the representatives of quantile level smaller than median, median level and greater than median levels, respectively. We take $q = 0.5$ in NBBQR and OBBQR. We choose the optimal regularization parameter in quantile LASSO and LASSO using 10-fold cross validation. For each method, the median absolute deviation (MAD) is recorded, where the MAD is defined as

$$\text{MAD} = \text{median}(|y_i - \hat{y}_i|),\ i = 1, \cdots, 506. \tag{31}$$

The MADs for each method with different quantiles are summarized in Table 8. We may find that all the Bayesian methods outperform frequentist approaches uniformly in all the quantiles. Moreover, compared with the original Bayesian approaches, our methods have the smaller MADs in 5 out of 6 comparisons, which indicate that our methods could provide more accurate prediction.

Additionally, we compute posterior medians of the coefficients with their 95% credible intervals for the Bayesian methods and coefficient estimators for the frequentist methods. Note that the Bayesian methods can provide the interval estimations simultaneously while the frequentist methods usually do not have simply implemented interval estimators. Figure 4 illustrates these estimations with $p = 0.7$. For brevity, we drop the names of above predictors and keep the corresponding number to indicate each predictor. In this figure, we add a slight horizontal shift to the estimators given by NBQL, NBBQR($q = 0.5$), OBQL and OBBQR($q = 0.5$) to make it more readable.

From Figure 4, we can see that all the Bayesian methods tend to behave similarly and the estimation are pretty close. For predictors 2, 8 and 9, quantile LASSO tends to behave differently compared with Bayesian methods, since the estimators lie outside the 95% credible intervals. For LASSO, it gives different estimation in terms of predictors 5, 6, 12

*Table 8. MADs from the six methods. The bold number is the smallest number in the category*

| | Method | MAD |
|---|---|---|
| $p = 0.4$ | NBQL | .0835 |
| | OBQL | .0853 |
| | NBBQR($q = 0.5$) | **.0831** |
| | OBBQR($q = 0.5$) | .0836 |
| | QL | .0889 |
| | LASSO | .0933 |
| $p = 0.5$ | NBQL | .0833 |
| | OBQL | .0848 |
| | NBBQR($q = 0.5$) | **.0822** |
| | OBBQR($q = 0.5$) | .0837 |
| | QL | .0879 |
| | LASSO | .0933 |
| $p = 0.7$ | NBQL | .0838 |
| | OBQL | .0849 |
| | NBBQR($q = 0.5$) | .0826 |
| | OBBQR($q = 0.5$) | **.0821** |
| | QL | .0878 |
| | LASSO | .0933 |

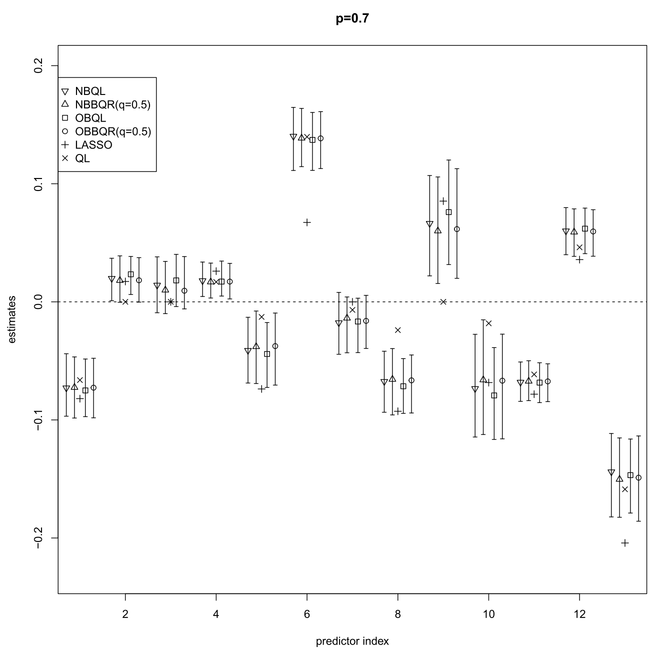


*Figure 4. The estimates of the predictor effects for the Boston Housing data using different methods with $p = 0.7$. The 95% credible intervals given by NBQL, NBBQR($q = 0.5$), OBQL and OBBQR($q = 0.5$) are plotted.*

and 13 compared with Bayesian methods. The similar performances can also be found in $p = 0.4$ and $p = 0.5$. For the variable selection, all the Bayesian regularized quantile methods and LASSO select all the predictors except predictors 4 and 6. In contrast, QL selects all the predictors except predictors 2, 3, 9.

## 6. CONCLUSION AND DISCUSSION

In this paper, we have mainly investigated the Bayesian methods for quantile regression with LASSO penalty and its two generalizations. We have presented the multimodality of the joint posterior and the sensitivity of hyperparameter specifications existing in the current popular methods. We also have demonstrated that using the invariant prior is not favorable for the scale component by establishing the conditions of the posterior propriety under the generalized likelihood and prior. To solve all problems with one framework, we propose the conditional prior for the regression coefficients and derive the conditions for posterior propriety for our proposed methods. Researchers who are interested in the topic could easily use Theorem 1 and 2 for an immediate guidance in terms of choosing appropriate priors. Meanwhile, the guidance could be applied to the context of Bayesian mean regression without efforts. For example, we have provided theoretical evidence about the improper joint posterior in Bayesian bridge regression (Polson et al. [28]) when predictors outnumbered observations. Additionally, the partially collapsed Gibbs sampling procedures have been developed for a fast and convenient use. Simulation studies and real data example have shown that our new Bayesian methods generally perform better than the existing Bayesian regularized quantile regression methods.

Our work highlights the importance of using the conditional priors for Bayesian regularized quantile regression for its sound characteristics of the posterior distribution. It also opens a gate for researchers to generalize the conditional priors to other related contexts. For instance, adaptive regularization has been proven to enjoy the oracle property (Zou [38]), one possible direction is to extend the theoretical results to the regularization with adaptive penalties such as the Bayesian quantile adaptive LASSO (Alhamzawi et al. [3]), adaptive group LASSO and adaptive bridge regression. Furthermore, we do not consider the crossing problem for different quantiles. Without special restriction, quantile regression functions estimated at different orders can cross each other. It disobeys the rule of the probability. Therefore, another concern is to develop the theoretical results for the noncrossing Bayesian regularized quantile regression (Liu and Wu [24]; Reich et al. [29]).

## APPENDIX A

*Proof of Theorem 1.* We write $\gtrsim$ if right hand side quantity is bounded by the left hand side up to a universal constant.

For (a), without loss of generality, we take $||.||_l = ||.||_r = ||.||$. By (12), we only need to consider the integrability of $(\boldsymbol{\beta}, \sigma^{k_1}, \eta^h)$ in

$$\sigma^n \eta^{\frac{m}{k_2}} \exp(\sigma^{k_1} q_i ||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}||^{k_1} - \eta F(\boldsymbol{\beta}))(\sigma^{k_1})^{a-1} \exp(-\sigma^{k_1} b)(\eta^h)^{c-1} \exp(-\eta^h d), i = 1, 3. \tag{32}$$

With loss of generality, we set $q_1 = q_3 = -1$. Since

$$\sqrt{ab} \le \left(\frac{a^h + b^h}{2}\right)^{1/h}, a > 0, b > 0, h > 0,$$

we have

(33)

$$(32) \ge \sigma^n \eta^{\frac{m}{k_2}} \exp(-\sigma_1^k ||\boldsymbol{y} - \boldsymbol{X\beta}||^{k_1} - 4^{-\frac{1}{h}}(1 + \eta^h F^h(\boldsymbol{\beta}))^{\frac{2}{h}}) (\sigma^{k_1})^{a-1} \exp(-\sigma^{k_1} b)(\eta^h)^{c-1} \exp(-\eta^h d) \gtrsim \sigma^n \eta^{\frac{m}{k_2}} \exp(-\sigma_1^k ||\boldsymbol{y} - \boldsymbol{X\beta}||^{k_1} - 4^{-\frac{1}{h}} \eta^h F^h(\boldsymbol{\beta}))(\sigma^{k_1})^{a-1} \exp(-\sigma^{k_1} b)(\eta^h)^{c-1} \exp(-\eta^h d).$$

The second inequality is due to the monotonicity of $a^x, a > 1$ and $h \ge 2$. Integrating out $\eta^h$ in (33) yields the result proportional to

(34)

$$\frac{1}{(4^{-\frac{1}{h}} F^h(\boldsymbol{\beta}) + d)^{\frac{m}{k_2 h} + c}} \sigma^n \exp(-\sigma^{k_1} ||\boldsymbol{y} - \boldsymbol{X\beta}||^{k_1})(\sigma^{k_1})^{a-1} \exp(-\sigma^{k_1} b).$$

For $a > 0, b > 0$, consider $a^l + b^l$ and $(a+b)^l$ for any given $l > 0$. When $l \ge 1$, we have $2^{1-l}(a+b)^l \le a^l + b^l \le (a+b)^l$. When $0 < l < 1$, $(a+b)^l \le a^l + b^l \le 2^{1-l}(a+b)^l$. Hence,

(35) $$\min(1, 2^{1-l}) \le (a^l + b^l)/(a+b)^l \le \max(1, 2^{l-1}),\ l > 0.$$

Therefore, we can replace

$$4^{-\frac{1}{h}} F^h(\boldsymbol{\beta}) + d,$$

with

$$(4^{-\frac{1}{k_2 h^2}} F^{\frac{1}{k_2}}(\boldsymbol{\beta}) + d^{\frac{1}{k_2 h}})^{k_2 h}.$$

In addition, $F^{1/k_2}(\boldsymbol{\beta})$ is a pre-norm on $\mathbb{R}^m$, by Theorem 5.4.4 in Horn and Johnson [11], there exist $A_1 > 0$ and $A_2 > 0$ such that

$$A_1 ||\boldsymbol{\beta}||_2 \le F^{\frac{1}{k_2}}(\boldsymbol{\beta}) \le A_2 ||\boldsymbol{\beta}||_2.$$

Hence, we can replace

$$4^{-\frac{1}{h}} F^h(\boldsymbol{\beta}) + d,$$

with

$$(4^{-\frac{2}{k_2 h^2}} A_i^2 ||\boldsymbol{\beta}||_2^2 + d^{\frac{2}{k_2 h}})^{\frac{k_2 h}{2}}, i = 1, 2.$$

That's to say, the integrability of (34) is equivalent to the integrability of

(36)

$$\frac{1}{(4^{-\frac{2}{k_2 h^2}} A_i^2 ||\boldsymbol{\beta}||_2^2 + d^{\frac{2}{k_2 h}})^{\frac{m + h k_2 c}{2}}} \sigma^n \exp(-\sigma^{k_1} ||\boldsymbol{y} - \boldsymbol{X\beta}||^{k_1}) (\sigma^{k_1})^{a-1} \exp(-\sigma^{k_1} b).$$

with respect to $(\boldsymbol{\beta}, \sigma^{k_1})$. For sake of simplicity, we set $A_i = 1, i = 1, 2$, since the values have no influence on the conclusion. When integrating out $\sigma^{k_1}$ in (36), $n + ak_1 > 0$ is necessary. We integrate out $\sigma^{k_1}$ using the similar argument for integrating out $\eta^h$ and it boils down to considering

(37)

$$\frac{1}{(4^{-\frac{2}{k_2 h^2}} ||\boldsymbol{\beta}||_2^2 + d^{\frac{2}{k_2 h}})^{\frac{m + h k_2 c}{2}}} \frac{1}{(||\boldsymbol{y} - \boldsymbol{X\beta}||_2^2 + b^{\frac{k_1}{2}})^{\frac{a k_1 + n}{2}}}.$$

In contrast, if integrating out $(s, t)$ in

(38) $$s^{\frac{m + h k_2 c}{2} - 1} \exp(-(4^{-\frac{2}{k_2 h^2}} ||\boldsymbol{\beta}||_2^2 + d^{\frac{2}{k_2 h}}) s) t^{\frac{a k_1 + n}{2} - 1} \exp(-(||\boldsymbol{y} - \boldsymbol{X\beta}||_2^2 + b^{\frac{k_1}{2}}) t),$$

the result is proportional to (37). Therefore, we only need to consider (38), which is tractable.

Set $(\delta, t) = (s/t, t)$ and integrate out $\boldsymbol{\beta}$ in (38), we obtain

(39) $$\frac{t^{\frac{a k_1 + n + h k_2 c}{2} - 1}}{|\boldsymbol{X'X} + 4^{-\frac{2}{k_2 h^2}} \delta \boldsymbol{I}_m|^{\frac{1}{2}}} \exp(-(\boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X'X} + 4^{-\frac{2}{k_2 h^2}} \delta \boldsymbol{I}_m)^{-1} \boldsymbol{X'}) \boldsymbol{y} + b^{\frac{k_1}{2}} + d^{\frac{2}{k_2 h}} \delta) t) \delta^{\frac{m + h k_2 c}{2} - 1}.$$

(39) is integrable with respect to $t$ iff $ak_1 + n + hk_2 c > 0$ and it yields

(40) $$\frac{\delta^{\frac{m + h k_2 c}{2} - 1}}{|\boldsymbol{X'X} + 4^{-\frac{2}{k_2 h^2}} \delta \boldsymbol{I}_m|^{\frac{1}{2}} Q_1},$$

where

$$Q_1 = (\boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X'X} + 4^{-\frac{2}{k_2 h^2}} \delta \boldsymbol{I}_m)^{-1} \boldsymbol{X'}) \boldsymbol{y}) + b^{\frac{k_1}{2}} + d^{\frac{2}{k_2 h}} \delta)^{\frac{a k_1 + n + h k_2 c}{2}}.$$

By the fact $(\boldsymbol{I} + \boldsymbol{AB})^{-1} \boldsymbol{A} = \boldsymbol{A}(\boldsymbol{I} + \boldsymbol{BA})^{-1}$,

(41) $$\boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X'X} + 4^{-\frac{2}{k_2 h^2}} \delta \boldsymbol{I}_m)^{-1} \boldsymbol{X'}) \boldsymbol{y} = \boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{XX'}(\boldsymbol{XX'} + 4^{-\frac{2}{k_2 h^2}} \delta \boldsymbol{I}_n)^{-1}) \boldsymbol{y}.$$

By spectral decomposition of $\boldsymbol{XX'} = \boldsymbol{Q\Lambda Q'}$, $\boldsymbol{\Lambda} = diag(\tau_1, \cdots, \tau_r, 0, \cdots, 0)$. Assume $\boldsymbol{Q}'\boldsymbol{y} = \boldsymbol{d}$, using the fact that (41) is SSE when $\delta = 0$, we obtain

(42) $$(41) = SSE + \sum_{i=1}^{r} \frac{4^{-\frac{2}{k_2 h^2}} \delta}{4^{-\frac{2}{k_2 h^2}} \delta + \tau_i} d_i^2.$$

By spectral decomposition, $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{S\Lambda}_1 \boldsymbol{S}'$, where $\boldsymbol{\Lambda}$ and $\boldsymbol{\Lambda}_1$ share the same nonzero eigenvalues.

(43) $$(40) \propto \frac{\delta^{\frac{r + h k_2 c}{2} - 1}}{\sqrt{\prod_{i=1}^{r} (\tau_i + 4^{-\frac{2}{k_2 h^2}} \delta)} Q_2},$$

where

$$Q_2 = (SSE + \sum_{i=1}^{r} \frac{4^{-\frac{2}{k_2 h^2}}\delta}{4^{-\frac{2}{k_2 h^2}}\delta + \tau_i} d_i^2 + b^{\frac{k_1}{2}} + d^{\frac{2}{k_2 h}}\delta)^{\frac{ak_1+n+hk_2 c}{2}}.$$

We need discuss whether SSE+$b^{k_1/2}$ or SSE+$b$ and $d$ are positive or not. We have the following cases:
(1) SSE+$b = 0$, $d = 0$;
(2) SSE+$b = 0$, $d > 0$;
(3) SSE+$b > 0$, $d = 0$;
(4) SSE+$b > 0$, $d > 0$.
We take the first as an example. When SSE+$b = 0$ and $d = 0$, (43) $= O(\delta^{(r-ak_1-n)/2-1})$ as $\delta \to 0$ and (43) $= O(\delta^{hk_2c/2-1})$ as $\delta \to \infty$. Hence, $n - r + ak_1 < 0$ and $c < 0$. Other cases can be obtained in the same way. Notice that $n + ak_1 > 0$ is implied by Condition A, B and C, the result hold.

For (b), we could use the same argument in (a) with $h = 1$ except the the integration of (32) with respect to $\eta$ is proportional to (34). □

## APPENDIX B

*Proof of Theorem 2.* For (a), we only need to consider the integrability of $(\boldsymbol{\beta}, \sigma^k, \lambda|\boldsymbol{y})$ in

$$(44)\quad (\sigma^k)^{\frac{m+n}{k}-1}\lambda^{\frac{m}{k}}\exp(-\sigma^k||\boldsymbol{y}-\boldsymbol{X\beta}||^k - \lambda\sigma^k F(\boldsymbol{\beta}))g(\lambda).$$

Integrating out $\sigma^k$ in (44), the result is proportional to

$$(45)\quad \frac{\lambda^{\frac{m}{k}}}{(||\boldsymbol{y}-\boldsymbol{X\beta}||^k + \lambda F(\boldsymbol{\beta}))^{\frac{m+n}{k}}} g(\lambda).$$

By (35), we just need to consider

$$(46)\quad \frac{\lambda^{\frac{m}{k}}}{(||\boldsymbol{y}-\boldsymbol{X\beta}|| + \lambda^{\frac{1}{k}} F^{\frac{1}{k}}(\boldsymbol{\beta}))^{m+n}} g(\lambda).$$

Since $F^{1/k}(\boldsymbol{\beta})$ is pre-norm on $\mathbb{R}^m$, we only need to consider

$$(47)\quad \frac{\lambda^{\frac{m}{k}}}{(||\boldsymbol{y}-\boldsymbol{X\beta}||_2 + \lambda^{\frac{1}{k}}||\boldsymbol{\beta}||_2)^{m+n}} g(\lambda).$$

Using (35) again, we just need focus on

$$(48)\quad \frac{\lambda^{\frac{m}{k}}}{(||\boldsymbol{y}-\boldsymbol{X\beta}||_2^2 + \lambda^{\frac{2}{k}}||\boldsymbol{\beta}||_2^2)^{\frac{m+n}{2}}} g(\lambda).$$

In contrast, if integrating out $s$ in

$$(49)\quad s^{\frac{m+n}{2}-1}\exp(-(||\boldsymbol{y}-\boldsymbol{X\beta}||_2^2 + \lambda^{\frac{2}{k}}||\boldsymbol{\beta}||_2^2)s)\lambda^{\frac{m}{k}} g(\lambda),$$

the result is proportional to (48). Integrating out $(\boldsymbol{\beta}, s)$ in (49) yields the result proportional to

$$(50)\quad \frac{\lambda^{\frac{m}{k}}}{|\boldsymbol{X}'\boldsymbol{X} + \lambda^{\frac{2}{k}}\boldsymbol{I}_m|^{\frac{1}{2}}(\boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X} + \lambda^{\frac{2}{k}}\boldsymbol{I}_m)^{-1}\boldsymbol{X}')\boldsymbol{y})^{\frac{n}{2}}} g(\lambda).$$

By (42), we obtain

$$(51)\quad (50) \propto \frac{\lambda^{\frac{r}{k}}}{\sqrt{\prod_{i=1}^{r}(\tau_i + \lambda^{\frac{2}{k}})}(SSE + \sum_{i=1}^{r}\frac{\lambda^{\frac{2}{k}}}{\tau_i + \lambda^{\frac{2}{k}}} d_i^2)^{\frac{n}{2}}} g(\lambda).$$

When SSE=0, (51) $= O(\lambda^{(r-n)/k})$ as $\lambda \to 0$ and (51) $= O(1)$ as $\lambda \to \infty$. When SSE>0, (51) $= O(\lambda^{r/k})$ as $\lambda \to 0$ and (51) $= O(1)$ as $\lambda \to \infty$. Hence, the results hold.

For (b), we only need to consider the integrability of

$$(52)\quad (\sigma^{k_1})^{\frac{m+n}{k_1}-1}\exp(-\sigma^{k_1}||\boldsymbol{y}-\boldsymbol{X\beta}||^{k_1} - \lambda\sigma^{k_2}F(\boldsymbol{\beta}))\lambda^{\frac{m}{k_2}+a_0-1}\exp(-b_0\lambda).$$

Integrating out $\lambda$ yields the result proportional to

$$(53)\quad \frac{1}{(\sigma^{k_2}F(\boldsymbol{\beta}) + b_0)^{\frac{m}{k_2}+a_0}}(\sigma^{k_1})^{\frac{m+n}{k_1}-1}\exp(-\sigma^{k_1}||\boldsymbol{y}-\boldsymbol{X\beta}||^{k_1}).$$

Using the same argument in (a), we only need to consider

$$(54)\quad \frac{1}{(\sigma^{k_1}||\boldsymbol{\beta}||^{k_1} + b_0^{\frac{k_1}{k_2}})^{\frac{m+a_0k_2}{k_1}}}(\sigma^{k_1})^{\frac{m+n}{k_1}-1}\exp(-\sigma^{k_1}||\boldsymbol{y}-\boldsymbol{X\beta}||^{k_1}).$$

Hence, we only consider

$$(55)\quad (\sigma^{k_1})^{\frac{m+n}{k_1}-1}\exp(-\sigma^{k_1}||\boldsymbol{y}-\boldsymbol{X\beta}||^{k_1} - \lambda\sigma^{k_1}||\boldsymbol{\beta}||^{k_1})\lambda^{\frac{m+a_0k_2}{k_1}-1}\exp(-b_0^{\frac{k_1}{k_2}}\lambda).$$

Integrating $\sigma^{k_1}$ in (55) yields

$$(56)\quad \frac{1}{(||\boldsymbol{y}-\boldsymbol{X\beta}||^{k_1} + \lambda||\boldsymbol{\beta}||^{k_1})^{\frac{m+n}{k_1}}}\lambda^{\frac{m+a_0k_2}{k_1}-1}\exp(-b_0^{\frac{k_1}{k_2}}\lambda).$$

The remaining is exactly the same as (a).

To prove the unimodality of the joint posterior of $(\boldsymbol{\beta}, \sigma^{k_1})$ give $\lambda$, we follow the similar argument in Park and Casella [27]. We only need to prove the upper level set of joint posterior $p(\boldsymbol{\beta}, \sigma^{k_1}|\boldsymbol{y})$

$$\{(\boldsymbol{\beta}, \sigma^{k_1})|\ p(\boldsymbol{\beta}, \sigma^{k_1}|\boldsymbol{y}) > c\},$$

is connected for any $c > 0$. The logarithm of posterior is

$$(57)\quad (\frac{n+m}{k_1} - 1)\log(\sigma^{k_1}) + \ln(\zeta(\sigma(\boldsymbol{y}-\boldsymbol{X\beta}))) - \lambda F(\sigma\boldsymbol{\beta}),$$

after dropping the terms free of $(\boldsymbol{\beta}, \sigma^{k_1})$. The coordinate transformation defined by

$$(58)\quad \boldsymbol{\phi} = \sigma\boldsymbol{\beta}, \tau = \sigma,$$

is one-to-one and thus the unimodality in the original coordinate is equivalent to unimodality in the transformed coordinates. For the transformed coordinates, (57) becomes

$$(n+m-k_1)\log(\tau)+\ln(\zeta(\tau \boldsymbol{y}-\boldsymbol{X}\boldsymbol{\phi}))-\lambda F(\boldsymbol{\phi}). \tag{59}$$

The first term and the third term is concave in $(\boldsymbol{\phi},\tau)$, since $F^{1/k_2}(\boldsymbol{\phi})$ is a norm and $x^{k_2}$ is convex in $x$. The second term is also concave in $(\boldsymbol{\phi},\tau)$. $\tau\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\phi}$ is the linear transformation of $(\boldsymbol{\phi},\tau)$ and $\ln(\zeta(y))$ is concave. Hence, (59) is concave in $(\boldsymbol{\phi},\tau)$. □


## ACKNOWLEDGEMENT

The authors sincerely thank the associate editor and two anonymous reviewers for their thoughtful and constructive suggestions they provided that led to considerable improvements in the presentation of our work. Dongchu Sun's work was partially supported by Chinese 111 Project B14019 and Chinese NSF grant 11671147.

*Received 1 November 2020*


## REFERENCES


[1] Alhamzawi, R. and Algamal, Z. Y. (2019). Bayesian bridge quantile regression. *Communications in Statistics – Simulation and Computation.* **48** 994–956. MR3938064

[2] Alhamzawi, R., Alhamzawi, A. and Ali, H. T. M. (2019). New Gibbs sampling methods for bayesian regularized quantile regression. *Computers in Biology and Medicine.* **110** 52–65.

[3] Alhamzawi, R., Yu, K. and Benoit, D. F. (2012). Bayesian adaptive Lasso quantile regression. *Statistical Modelling.* **12** 279–297. MR3179503

[4] Andrews, D. F. and Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological).* **36** 99–102. MR0359122

[5] Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer-Verlag Inc. MR0804611

[6] Casella, G. (2001). Empirical Bayes Gibbs sampling. *Biostatistics.* **2** 485–500.

[7] Chen, J. and Chen, Z. (2014). Extended BIC for small-n-large-p sparse GLM. *Statistica Sinica.* **22** 555–574. MR2954352

[8] Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer-Verlag New York Inc. MR0836973

[9] Frank, I. and Friedman, J. (1993). A statistical view of some chemometrics regression tools. *Technometrics.* **35** 109–135.

[10] Harrison, D and Rubinfeld, D (1978). Hedonic prices and the demand for clean air. *Journal of Environmental Economics and Management.* **5** 81–102.

[11] Horn, R. and Johnson, C. (1994). *Matrix Analysis*. Cambridge University Press MR2978290

[12] Jørgensen, B. (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Springer-Verlag New York Inc. MR0648107

[13] Jullion, A. and Lambert, P. (2007). Robust specification of the roughness penalty prior distribution in spatially adaptive Bayesian P-splines models. *Computational Statistics and Data Analysis.* **51** 2542–2558. MR2338987

[14] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Biometrika.* **46** 33–50. MR0474644

[15] Kyung, M. and Gill, J. and Ghosh, M and Casella, G(2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis.* **5** 369–412. MR2719657

[16] Kottas, A. and Krnjajić, M. (2009). Bayesian semiparametric modelling in quantile regression. *Scandinavian Journal of Statistic.* **36** 297–319. MR2528986

[17] Kozumi, H. and Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation.* **81** 1565–1578. MR2851270

[18] Jiang, X., Jiang, J. and Song, X. (2012). Oracle model selection for nonlinear models based on weighted composite quantile regression. *Statistica Sinica.* **22** 1479–1506. MR3027096

[19] Jiang, L., Bondell, H. and Wang, H. (2014). Interquantile shrinkage and variable selection in quantile regression. *Computation Statistics and Data Analysis.* **69** 208–219. MR3146889

[20] Jiang, L., Wang, H. and Bondell, H. (2013). Interquantile shrinkage in regression models. *Journal of Computational and Graphical Statistics.* **22** 970–986. MR3173752

[21] Li, Q, Xi, R and Lin, N (2010). Bayesian regularized quantile regression. *Bayesian Analysis.* **5** 533–556. MR2719666

[22] Li, Y, Liu, Y and Zhu, J (2007). Quantile regression in reproducing kernel Hilbert spaces. *Journal of the American Statistical Association.* **102** 255–268. MR2293307

[23] Li, Y and Zhu, J (2008). $l_1$-norm quantile regressions. *Journal of Computational and Graphical Statistics.* **17** 163–185. MR2424800

[24] Liu, Y and Wu, Y (2011). Simultaneous multiple non-crossing quantile regression estimation using kernel constraints. *Journal of Nonparametric Statistics.* **23** 415–437. MR2801302

[25] Mallick, H. and Yi, N. (2014). A new Bayesian lasso. *Statistics and Its Interface.* **7** 571–582. MR3302383

[26] Mallick, H. and Yi, N. (2018). Bayesian bridge regression. *Journal of Applied Statistics.* **45** 988–1008. MR3774527

[27] Park, T. and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association.* **103** 681–686. MR2524001

[28] Polson, N., Scott, J. and Windle, J. (2014). The Bayesian bridge. *Journal of the Royal Statistical Society: Series B.* **76** 713–733. MR3248673

[29] Reich, B., Fuentes, M. and Dunson, D. (2011). Bayesian spatial quantile regression. *Journal of the American Statistical Association.* **106** 6–20. MR2816698

[30] Sriram, K., Ramamoorthi, R.V. and Ghosh, P. (2013). Posterior consistency of Bayesian quantile regression based on the misspecified asymmetric Laplace density. *Bayesian Analysis.* **8** 479–504. MR3066950

[31] Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B.* **58** 267–288. MR1379242

[32] van Dyk, D. A. and Park, T. (2008). Partially collapsed Gibbs samplers. *Journal of the American Statistical Association.* **103** 790–796. MR2524010

[33] Wang, H., Zhou, H. and Li, Y. (2013). Variable selection for censored quantile regression. *Statistica Sinica.* **23** 145–167. MR3076162

[34] Wu, Y and Liu, Y. (2009). Variable selection in quantile regression. *Statistica Sinica.* **19** 801–817. MR2514189

[35] Xu, Z., Zhang, H., Wang, Y., Chang, X. and Liang, Y. (2010). $L_{1/2}$ regularization. *Science China Information Sciences.* **53** 1159–1169. MR2671451

[36] Yu, K. and Moyeed, R. (2001). Bayesian quantile regression. *Statistics and Probability Letters.* **54** 437–447. MR1861390

[37] Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B.* **68** 49–67. MR2212574

[38] Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association.* **101** 1418–1429. MR2279469

Zhongheng Cai
School of Statistics
Faculty of Economics and Management
East China Normal University
Shanghai
China
E-mail address: zhonghengcai123@gmail.com

Dongchu Sun
School of Statistics
Faculty of Economics and Management
East China Normal University
Shanghai
and
Department of Statistics
University of Nebraska-Lincoln
Nebraska
USA
E-mail address: dsun9@unl.edu