Computers in Biology and Medicine 110 (2019) 52–65

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

New Gibbs sampling methods for bayesian regularized quantile regression

Rahim Alhamzawi[a,*], Ahmed Alhamzawi[b], Haithem Taha Mohammad Ali[c]

[a] Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Iraq
[b] Department of Mathematics, College of Science, University of Al-Qadisiyah, Iraq
[c] College of Computers and Information Technology, Nawroz University, Duhok, Iraq

ARTICLE INFO

Keywords:
Regularization
Adaptive lasso
Lasso
Elastic net
Gibbs sampler

ABSTRACT

In this paper, we propose new Bayesian hierarchical representations of lasso, adaptive lasso and elastic net quantile regression models. We explore these representations by observing that the lasso penalty function corresponds to a scale mixture of truncated normal distribution (with exponential mixing densities). We consider fully Bayesian treatments that lead to new Gibbs sampler methods with tractable full conditional posteriors. The new methods are then illustrated with both simulated and real data. Results show that the new methods perform very well under a variety of simulations, such as the presence of a moderately large number of predictors, collinearity and heterogeneity.

1. Introduction

After its inception in Ref. [1]; quantile regression (QReg) models have received considerable attention and widely applied method to investigate the whole conditional distribution of an outcome variable. A vast number of publications has been widely applied to investigate QReg in many different areas, such as climate change, geology, medicine, ecology, econometrics, economics, finance, growth chart, microarray study and survival analysis. QReg offers a more comprehensive picture of the underlying relationships of interest that can be especially useful when the upper or lower quantiles of the response variable depend on the predictors very differently from the center. Compared to the classical mean regression, QReg models belong to a robust model family [2]. It does not impose any distributional assumption on the error term, except that the conditional quantile of the error term is zero. In addition, classical mean regression limits comparison to mean levels, while QReg identifies the heterogeneous effects of the predictors at different quantile levels of the outcome of interest.

Suppose that we have a sample of independent observations , where y_i denote the response variable of interest and $\boldsymbol{x}_i = (x_{i1}, \cdots, x_{ik})'$ denote the corresponding vector of explanatory variables. Given a fixed quantile level $(0 < p < 1)$, the simple linear quantile regression model is

$$y_i = \boldsymbol{x}'_i\boldsymbol{\beta} + \varepsilon_i, \tag{1}$$

where $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_k)'$ and ε_i's are independent with their pth quantiles equal to 0. Under model (1), the linear conditional quantile functions are defined as

$$G_{y_i}(p|\boldsymbol{x}_i) = \boldsymbol{x}'_i\boldsymbol{\beta}, \tag{2}$$

where $G_{y_i}(p|\boldsymbol{x}_i)$ is the quantile function of y_i given $\boldsymbol{x}_i$ evaluated at p $(0 < p < 1)$. The quantile coefficient vector $\boldsymbol{\beta}$ can be estimated by solving

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

where $\rho_p(\cdot)$ is the quantile check loss function

$$\rho_p(w) = \begin{cases} pw, & \text{if } w \geq 0, \\ (p-1)w, & \text{if } w < 0. \end{cases} \tag{4}$$

Nowadays, high-dimensional data becomes very common and frequently collected in many different areas, such as economics, tumor classification, biomedical imaging and signal processing [3]. Variable selection plays a crucial role in building a QReg model. Specifically, it is assumed that only an unknown subset of candidate covariates are active in the regression, so that the variable selection problem is to select this unknown subset. Various methods have been proposed over the years for dealing with variable selection in linear regression models. Among these, regularization approaches have received considerable attention in the recent years. The popular regularization approach is lasso [4], which stands for least absolute shrinkage and selection operator. The lasso estimate is the solution to $\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} (y_i - \boldsymbol{x}'_i\boldsymbol{\beta})^2 + \lambda||\boldsymbol{\beta}||_1$, where $\lambda \geq 0$ is the tuning parameter and $||\boldsymbol{\beta}||_1 = \sum_{j=1}^{k} |\beta_j|$. Several improvement approaches of the lasso are as follows [5]. proposed the so-called elastic net to solve the collinearity problems [6]. proposed a group lasso

* Corresponding author.
E-mail address: rahim.alhamzawi@qu.edu.iq (R. Alhamzawi).

https://doi.org/10.1016/j.compbiomed.2019.05.011
Received 9 November 2018; Received in revised form 11 May 2019; Accepted 12 May 2019

method for group selection [7]. introduced adaptive lasso regression and showed that its oracle properties under some regularity conditions. Other related approaches include the fused lasso [8], graphical lasso [6], dantzig selector [9], and matrix completion [10,11]. Noting the penalty form $\lambda\sum_{j=1}^{k}|\beta_j|$ [4], suggested that the lasso regression estimates for linear models can be interpreted as a Bayesian posterior mode estimates when specifying a Laplace prior distribution to each β_j [12]. provided a Gibbs sampler for Bayesian lasso by putting scale mixture of normal prior distributions on the parameters and independent exponential prior distributions on their variances. Recently [13], developed a Bayesian method to solve the elastic net model using Gibbs sampler and [14] proposed a new Bayesian method to solve the elastic net model by utilizing a scale mixture of uniform representation of the Laplace density [15] in the elastic net prior.

With regard to the QReg, the lasso estimate is the solution to $\min_{\beta}\sum_{i=1}^{n}\rho_p(y_i - \boldsymbol{x}'_i\boldsymbol{\beta}) + \lambda\|\boldsymbol{\beta}\|_1$ [16]. Several improvement approaches of this technique are as follows [17]. illustrated the oracle properties of the SCAD and adaptive lasso QReg. [18] studied the statistical properties of the group Lasso for high dimensional sparse quantile regression models [19]. considered a weighted penalized composite QReg (CQReg) estimator and its oracle properties [20]. proposed a weighted CQReg estimation method and study model selection for nonlinear models with a diverging number of parameters and [21] proposed a general adaptive lasso QReg estimator and its oracle properties. From a Bayesian perspective [22], proposed Bayesian approaches using the asymmetric Laplace distribution for the errors to solve the lasso, elastic net and group lasso QReg models using Gibbs samples, and [23] proposed the Bayesian adaptive lasso QReg by assigning scale mixture of normal priors on the regression coefficents and independent exponential priors on their variances [24]. More recently [25], proposed Bayesian lasso mixed QReg and [26] studied Bayesian lasso and adaptive lasso composite QReg. In this paper, we propose new Bayesian QReg models for lasso, adaptive lasso and elastic net penalties. We explore these models by observing that the lasso penalty function corresponds to a scale mixture of truncated normal distribution (with exponential mixing densities). New Gibbs sampling methods are proposed for these three types of regularized QReg. Empirical results and real data analyses show that the new algorithms perform well compared to other existing regularized Bayesian and non-Bayesian QReg methods in terms of prediction accuracy.

In Section 2, we describe in detail the new Bayesian QReg models for lasso, adaptive lasso and elastic net penalties, and derive the corresponding Gibbs samplers. In Section 3, we illustrate the performance of the new proposed methods through a simulation study. Empirical results show that the new methods perform very well under a variety of scenarios, including moderate or severe sparsity, full predictor spaces, collinearity, more predictors than observations and heterogeneous random errors. Section 4 gives two real data examples. Conclusions are put in Section 5. An appendix contains a technical proof and full Conditional densities for lasso, adaptive lasso and elastic net QReg.

2. Methods

2.1. Bayesian formulation of the QReg

Since (3) is not differentiable at 0, a closed-form solution is not available for the unknown quantity $\boldsymbol{\beta}$ [1]. However, the minimization in (3) can be solved by a linear programming algorithm [27]. A parametric method to inference on the QReg coefficients arises if the distribution of the error term ε_i is specified. A popular choice for the distribution of the error term ε_i is the asymmetric Laplace distribution (ALD); see Refs. [28–31]. The density function of an ALD is

$$f(\varepsilon_i|\sigma) = p(1-p)\sigma\exp\{-\sigma\rho_p(\varepsilon_i)\}, \tag{5}$$

where σ is the scale parameter. Let $\boldsymbol{y} = (y_1,\cdots,y_n)'$, then the likelihood corresponding to modeling the pth quantile is

$$f(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{\beta},\sigma) = p^n(1-p)^n\sigma^n\exp\left\{-\sigma\sum_{i=1}^{n}\rho_p(y_i - \boldsymbol{x}'_i\boldsymbol{\beta})\right\}, \tag{6}$$

where $\boldsymbol{X}$ is the $n\times k$ matrix of covariates with ith row $\boldsymbol{x}'_i$. Maximizing the likelihood function (6) is equivalent to minimizing (3). However, direct use of the likelihood function (6) is rather inconvenient for Bayesian QReg. Specifically, the conditional distribution for the regression coefficients is not analytically tractable due to the complexity of the above likelihood function. By Ref. [32]; ε_i can be rewritten by a scale mixture of normals,

$$\varepsilon_i = \delta_1 v_i + \delta_2\sigma^{-1/2}\sqrt{v_i}\,\epsilon_i, \tag{7}$$

where $\delta_1 = (1-2p)/(p-p^2)$ and $\delta_2 = \sqrt{2/(p-p^2)}$. Here, v_i and ϵ_i follow an exponential distribution with rate parameter (σ) and a standard normal distribution, respectively. Specifically, Model (1) can be reformulated as,

$$y_i = \boldsymbol{x}'_i\boldsymbol{\beta} + \delta_1 v_i + \delta_2\sigma^{-1/2}\sqrt{v_i}\,\epsilon_i, \qquad i = 1,\ \cdots, n. \tag{8}$$

Since the conditional distribution of y_i given v_i is normal with mean $\boldsymbol{x}'_i\boldsymbol{\beta} + \delta_1 v_i$ and variance $\delta_2^2 v_i/\sigma$, the likelihood corresponding to modeling the pth quantile is

$$f(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{v},\boldsymbol{\beta},\sigma) \propto \prod_{i=1}^{n}\left(\frac{\sigma}{v_i}\right)^{\frac{1}{2}}\exp\left(-\frac{\sigma(y_i - \boldsymbol{x}'_i\boldsymbol{\beta} - \delta_1 v_i)^2}{2\delta_2^2 v_i}\right), \tag{9}$$

where $\boldsymbol{v} = (v_1,\cdots,v_n)'$.

2.2. New bayesian regularized QReg methods

2.2.1. A new bayesian lasso QReg

The lasso regularized QReg [16] is given by

$$\min_{\beta}\sum_{i=1}^{n}\rho_p(y_i - \boldsymbol{x}'_i\boldsymbol{\beta}) + \lambda\sum_{j=1}^{k}|\beta_j|, \qquad \lambda > 0. \tag{10}$$

Following [4,22] pointed out that solving the lasso regularized QReg problem is equivalent to finding the marginal posterior mode of the regression coefficients when the regression coefficients have independent Laplace priors,

$$\pi(\boldsymbol{\beta}|\eta) = (\eta/2)^k\exp\left\{-\eta\sum_{j=1}^{k}|\beta_j|\right\}, \tag{11}$$

where $\eta = \sigma\lambda$. Similar to Refs. [12,22] provided a Gibbs sampler for Bayesian lasso QReg by reformulating the Laplace priors as scale mixture of normal priors on the regression coefficients and independent exponential priors on their variances. In this paper, along the same line as [22]; we propose a new hierarchical representation of Bayesian lasso QReg.

Proposition. A Laplace density can be written as a scale mixture of truncated normal distribution (with exponential mixing densities), i.e.

$$\frac{\eta}{2}\exp(-\eta|\beta_j|) = \int_0^{\infty}\int_{u_j>\sqrt{\eta^2}|\beta_j|}\frac{1}{\sqrt{2\pi s_j}}\exp\left(-\frac{\beta_j^2}{2s_j}\right) \times\exp\left(-\frac{u_j}{2}\right)\frac{\eta^2}{8}\exp\left(-\frac{\eta^2 s_j}{8}\right)du_j\,ds_j, \tag{12}$$

A straightforward proof of this result is provided in Appendix A. We put a Gamma prior on σ with shape parameter a and rate parameter b and a Gamma prior on η^2 with shape parameter c and rate parameter d. Thus, the reformulation of the Laplace density suggests the following hierarchical representation of the full model.

$$
\begin{aligned}
&y_i = \boldsymbol{x}'_i\boldsymbol{\beta} + \delta_1 v_i + \delta_2\sigma^{-1/2}\sqrt{v_i}\varepsilon_i, \qquad \varepsilon_i \sim \mathrm{N}(0,1)\\
&\boldsymbol{v} \sim \prod_{i=1}^{n} \mathrm{Exponential}(\sigma),\\
&\boldsymbol{\beta}|\eta^2, \boldsymbol{u}, \boldsymbol{s} \sim \prod_{j=1}^{k} N(0, s_j) I\left\{|\beta_j| < \frac{u_j}{\sqrt{\eta^2}}\right\},\\
&\boldsymbol{s}|\eta^2 \sim \prod_{j=1}^{k} \mathrm{Exponential}\left(\frac{\eta^2}{8}\right),\\
&\boldsymbol{u} \sim \prod_{j=1}^{k} \mathrm{Exponential}\left(\frac{1}{2}\right),\\
&\sigma \sim \mathrm{Gamma}(a, b),\\
&\eta^2 \sim \mathrm{Gamma}(c, d).
\end{aligned} \tag{13}
$$

We sample from the above Bayesian hierarchical formulation using the following Gibbs sampler. All details are included in Appendix B.

Algorithm 1. (MCMC sampling for the Bayesian lasso QReg).

- Sample $\beta_j|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{v}, u_j, s_j, \eta^2, \sigma \sim N\left(\mu_j, \tau^2_{\beta_j}\right) I\{|\beta_j| < u_j/\sqrt{\eta^2}\}$, where

$$\tau^{-2}_{\beta_j} = \sum_{i=1}^{n} \frac{\sigma x_{ij}^2}{\delta_2^2 v_i} + \frac{1}{s_j}, \quad \text{and} \quad \mu_j = \tau^2_{\beta_j} \sum_{i=1}^{n} \frac{\sigma \tilde{y}_i x_{ij}}{\delta_2^2 v_i}.$$

where $\tilde{y}_i = y_i - \sum_{h=1,h\neq j}^{k} x_{ih}\beta_h - \delta_1 v_i$.

•Sample $\sigma|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{v} \sim \mathrm{Gamma}(g_1, g_2)$, where

$$g_1 = \frac{3n}{2} + a \quad \text{and} \quad g_2 = \sum_{i=1}^{n}\left(\frac{(y_i - \boldsymbol{x}'_i\boldsymbol{\beta} - \delta_1 v_i)^2}{2\delta_2^2 v_i} + v_i\right) + b$$

•Sample $v_i|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}, \sigma \sim \mathrm{GIG}(0.5, \gamma_{1i}, \gamma_2)$, where $\gamma_{1i} = \sigma(y_i - \boldsymbol{x}'_i\boldsymbol{\beta})^2/\delta_2^2$, and $\gamma_2 = \sigma(\delta_1^2/\delta_2^2 + 2)$.
•Sample $s_j|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}, \eta^2 \sim \mathrm{GIG}(0.5, \gamma_{1j}, \gamma_2)$, where $\gamma_{1j} = \beta_j^2$, and $\gamma_2 = \eta^2/4$.
•Sample $u_j|\boldsymbol{y}, \boldsymbol{X}, \beta_j, \eta^2$ from the left-truncated exponential distribution $\mathrm{Exponential}(0.5) I\{u_j > \sqrt{\eta^2}|\beta_j|\}$ using inversion method which can be done by sampling u_j^* from an exponential distribution with rate parameter 0.5 and setting $u_j = u_j^* + \sqrt{\eta^2}|\beta_j|$.
•Sample from the right-truncated gamma distribution $\mathrm{Gamma}\left(c + k, d + \frac{1}{8}\sum_{j=1}^{k} s_j\right) I\left\{\eta^2 < \mathrm{Min}_j\left(u_j^2/\beta_j^2\right)\right\}$.

Although the proposed Gibbs sampler is very different from the old Gibbs sampler reported in Ref. [22]; the posterior distribution $p(\boldsymbol{\beta}, \sigma, \boldsymbol{v}|\boldsymbol{y}, \boldsymbol{X})$ is exactly the same for both. Thus, any observed differences in estimation of the regression coefficients can be contributed to the features of the different algorithms utilized to get draws from the corresponding posteriors.

2.3. A new bayesian adaptive lasso QReg

Although lasso has gained a high degree of success in many situations, it usually either includes a number of inactive predictors to reduce the estimation bias or over-shrinks the parameters of the correct predictors to produce a model with correct size [5]. These drawbacks are partially addressed by adaptive lasso [5], which extends lasso by allowing different penalization parameters for different regression coefficients [23]. considered the following adaptive lasso regularized QReg from a Bayesian perspective.

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_p(y_i - \boldsymbol{x}'_i\boldsymbol{\beta}) + \sum_{j=1}^{k} \lambda_j|\beta_j|, \qquad \lambda_1, \cdots, \lambda_k > 0. \tag{14}$$

Specifically, they proposed Bayesian adaptive lasso QReg using a conditional Laplace prior distribution of the form $\pi(\boldsymbol{\beta}|\eta) = \prod_{j=1}^{k} \eta_j/2 \exp\{-\eta_j|\beta_j|\}$, where $\eta_j = \sigma\lambda_j$. Under this prior, the posterior of $\boldsymbol{\beta}$ becomes

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{X}, \sigma, \lambda_1, \cdots, \lambda_k) \propto \exp\left\{-\sigma\sum_{i=1}^{n} \rho_p(y_i - \boldsymbol{x}'_i\boldsymbol{\beta}) - \sigma\sum_{j=1}^{k} \lambda_j|\beta_j|\right\} \tag{15}$$

Maximizing the posterior (15) is then equivalent to minimizing (14). From the equality (12), we have

$$
\begin{aligned}
&\frac{\eta_j}{2}\exp(-\eta_j|\beta_j|)\\
&= \int_0^\infty \int_{u_j > \sqrt{\eta_j^2}|\beta_j|} \frac{1}{\sqrt{2\pi s_j}} \exp\left(-\frac{\beta_j^2}{2s_j}\right)\exp\left(-\frac{u_j}{2}\right)\frac{\eta_j^2}{8}\exp\left(-\frac{\eta_j^2 s_j}{8}\right) du_j ds_j,\\
&\propto \int_0^\infty \int_{u_j > \sqrt{\eta_j^2}|\beta_j|} \mathrm{N}(\beta_j; 0, s_j)\mathrm{Exponential}\left(u_j; \frac{1}{2}\right)\mathrm{Exponential}\left(s_j; \frac{\eta_j^2}{8}\right) du_j ds_j.
\end{aligned} \tag{16}
$$

We put a Gamma prior on σ with shape parameter a and rate parameter b. We further put a Gamma prior on η_j with shape parameter c and rate parameter d. Thus, the reformulation of the Laplace density suggests the following hierarchical representation of the full model:

$$
\begin{aligned}
&y_i = \boldsymbol{x}'_i\boldsymbol{\beta} + \delta_1 v_i + \delta_2\sigma^{-1/2}\sqrt{v_i}\varepsilon_i,\\
&\boldsymbol{v} \sim \prod_{i=1}^{n} \mathrm{Exponential}(\sigma),\\
&\varepsilon \sim \prod_{i=1}^{n} \mathrm{N}(0,1),\\
&\boldsymbol{\beta}|\boldsymbol{\eta}^2, \boldsymbol{s}, \boldsymbol{u} \sim \prod_{j=1}^{k} N(0, s_j) I\left\{|\beta_j| < \frac{u_j}{\sqrt{\eta_j^2}}\right\},\\
&\boldsymbol{s}|\boldsymbol{\eta}^2 \sim \prod_{j=1}^{k} \mathrm{Exponential}\left(\frac{\eta_j^2}{8}\right),\\
&\boldsymbol{u} \sim \prod_{j=1}^{k} \mathrm{Exponential}\left(\frac{1}{2}\right),\\
&\sigma \sim \mathrm{Gamma}(a, b),\\
&\boldsymbol{\eta}^2 \sim \prod_{j=1}^{k} \mathrm{Gamma}(c, d),
\end{aligned} \tag{17}
$$

where $\boldsymbol{\eta}^2 = (\eta_1^2, \cdots, \eta_k^2)$. All details of the full conditional distributions are included in Appendix C. We sample from the above Bayesian hierarchical formulation using the following Gibbs sampler.

Algorithm 2. (MCMC sampling for the Bayesian adaptive lasso QReg).

•The sampling steps of β_j, σ, v_i, s_j, and u_j are the same as those in Algorithm 1 with η^2 replaced with η_j^2.
•Sample $\eta_j^2|\boldsymbol{y}, \boldsymbol{X}, \beta_j, u_j, s_j$ from the right-truncated gamma distribution $\mathrm{Gamma}\left(c + 1, d + \frac{1}{8}s_j\right) I\left\{\eta_j^2 < \left(u_j^2/\beta_j^2\right)\right\}$.

2.4. A new bayesian elastic net QReg

[5] proposed the elastic net as an improved version of the lasso [4] for analysing data with much more predictors than observations. Following [5]; the elastic net QReg can be written as

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_p(y_i - \boldsymbol{x}'_i\boldsymbol{\beta}) + \lambda_1\sum_{j=1}^{k} |\beta_j| + \lambda_2\sum_{j=1}^{k} \beta_j^2, \qquad \lambda_1, \lambda_2 > 0. \tag{18}$$

[22] introduced the Bayesian elastic net QReg by assigning a conditional prior distribution on β_j takes the form $\pi(\beta_j|\eta_1, \eta_2) = C(\eta_1, \eta_2) 2^{-1}\eta_1 \exp\left(-\eta_1|\beta_j| - \eta_2\beta_j^2\right)$, where $\eta_1 = \sigma\lambda_1$, $\eta_2 = \sigma\lambda_2$ and $C(\eta_1, \eta_2)$ is a normalizing constant depending on η_1 and η_2, which is given by

$$C(\eta_1, \eta_2) = \Gamma^{-1}\left(1/2, \frac{\eta_1^2}{4\eta_2}\right)\left(\frac{\eta_1^2}{4\eta_2}\right)^{-1/2}\exp\left(-\frac{\eta_1^2}{4\eta_2}\right). \tag{19}$$

Here, $\Gamma(1/2, \eta_1^2/(4\eta_2))$ is the upper incomplete gamma function. In this paper, along the same line as [22]; we propose a new hierarchical representation of Bayesian elastic net QReg. Following the mixture representation of the Laplace prior in (12), the prior distribution $\pi(\beta_j|\eta_1,\eta_2) = C(\eta_1,\eta_2)2^{-1}\eta_1 \exp\left(-\eta_1|\beta_j| - \eta_2\beta_j^2\right)$ can be written as

$$\pi(\beta_j,\eta_1,\eta_2) = C(\eta_1,\eta_2)\frac{\eta_1}{2}\exp\left\{-\eta_1|\beta_j| - \eta_2\beta_j^2\right\}$$

$$= C(\eta_1,\eta_2)\int_0^\infty \int_{u_j>\sqrt{\eta_1^2}|\beta_j|}\frac{1}{\sqrt{2\pi s_j}}\exp\left(-\frac{1+2\eta_2 s_j}{2s_j}\beta_j^2\right)\exp\left(-\frac{u_j}{2}\right)$$

$$\times \frac{\eta_1^2}{8}\exp\left(\frac{-\eta_1^2 s_j}{8}\right)du_j ds_j.$$

Letting $t_j = 1 + 2\eta_2 s_j$, the prior becomes

$$\pi(\beta_j,\eta_1,\eta_2)$$

$$= C(\eta_1,\eta_2)\int_1^\infty \int_{u_j>\sqrt{\eta_1^2}|\beta_j|}\frac{t_j^{-1/2}}{\sqrt{2\pi(t_j-1)/(2\eta_2 t_j)}}\exp\left(-\frac{1}{2}\left[\frac{t_j-1}{2\eta_2 t_j}\right]^{-1}\beta_j^2\right)$$

$$\times \exp\left(-\frac{u_j}{2}\right)\frac{\eta_1^2}{16\eta_2}\exp\left(\frac{-\eta_1^2}{16\eta_2}(t_j-1)\right)du_j dt_j.$$

We set $\widetilde{\eta_1} = \eta_1^2/(4\eta_2)$ and put Gamma priors on σ, $\widetilde{\eta_1}$, and η_2 with shape parameters a, c_1, c_2 and rate parameters b, d_1, d_2, respectively.

To summarize, our Bayesian hierarchical formulation is provided below.

$$y_i = \boldsymbol{x'}_i\boldsymbol{\beta} + \delta_1 v_i + \delta_2\sigma^{-1/2}\sqrt{v_i}\varepsilon_i,$$

$$\boldsymbol{v} \sim \prod_{i=1}^n \text{Exponential}(\sigma),$$

$$\varepsilon \sim \prod_{i=1}^n N(0,1),$$

$$\boldsymbol{\beta}|\boldsymbol{t},\boldsymbol{u},\eta_2$$

$$\sim \prod_{j=1}^k \frac{1}{\sqrt{2\pi(t_j-1)/(2\eta_2 t_j)}}\exp\left(-\frac{1}{2}\left[\frac{t_j-1}{2\eta_2 t_j}\right]^{-1}\beta_j^2\right)$$

$$I\left\{|\beta_j| < \frac{u_j}{\sqrt{4\widetilde{\eta}_1\eta_2}}\right\},$$

$$\boldsymbol{t}|\tilde{\eta}_1 \sim \prod_{j=1}^k \Gamma^{-1}(1/2,\tilde{\eta}_1)t_j^{-1/2}\tilde{\eta}_1^{1/2}\exp\left\{-\frac{\tilde{\eta}_1 t_j}{4}\right\}I(t_j>1),$$

$$\boldsymbol{u} \sim \prod_{j=1}^k \text{Exponential}\left(\frac{1}{2}\right),$$

$$\sigma \sim \text{Gamma}(a,b)$$

$$\tilde{\eta}_1 \sim \text{Gamma}(c_1,d_1),$$

$$\eta_2 \sim \text{Gamma}(c_2,d_2),$$

where $\boldsymbol{t} = (t_1,\cdots,t_k)$. All details of the full conditional distributions are included in Appendix D. We sample from the above Bayesian hierarchical formulation using the following Gibbs sampler.

Algorithm 3. (MCMC sampling for the Bayesian elastic net QReg).

•Sample $\beta_j|\boldsymbol{y},\boldsymbol{X},\boldsymbol{v},u_j,s_j,\widetilde{\eta_1},\eta_2,\sigma \sim N\left(\mu_j,\tau_{\beta_j}^2\right)I\{|\beta_j| < u_j/\sqrt{4\widetilde{\eta}_1\eta_2}\}$, where

$$\tau_{\beta_j}^{-2} = \sum_{i=1}^n \frac{\sigma x_{ij}^2}{\delta_2^2 v_i} + \left[\frac{t_j-1}{2\eta_2 t_j}\right]^{-1}, \quad \text{and} \quad \mu_j = \tau_{\beta_j}^2\sum_{i=1}^n\frac{\sigma\tilde{y}_i x_{ij}}{\delta_2^2 v_i}.$$

where $\tilde{y}_i = y_i - \sum_{h=1,h\neq j}^k x_{ih}\beta_h - \delta_1 v_i$.

•Sample $\sigma|\boldsymbol{y},\boldsymbol{X},\boldsymbol{\beta},\boldsymbol{v} \sim \text{Gamma}(g_1,g_2)$, where

$$g_1 = \frac{3n}{2} + a \quad \text{and} \quad g_2 = \sum_{i=1}^n\left(\frac{(y_i - \boldsymbol{x'}_i\boldsymbol{\beta} - \delta_1 v_i)^2}{2\delta_2^2 v_i} + v_i\right) + b$$

•Sample $v_i|\boldsymbol{y},\boldsymbol{X},\boldsymbol{\beta},\sigma\sim$ GIG$(0.5,\gamma_{1i},\gamma_2)$, where $\gamma_{1i} = \sigma(y_i - \boldsymbol{x'}_i\boldsymbol{\beta})^2/\delta_2^2$, and $\gamma_2 = \sigma(\delta_1^2/\delta_2^2 + 2)$.
•Sample $t_j - 1|\boldsymbol{y},\boldsymbol{X},\beta_j,\tilde{\eta}_1,\eta_2\sim$ GIG$(0.5,\gamma_{1j},\gamma_2)$, where $\gamma_{1j} = 2\eta_2\beta_j^2$ and $\gamma_2 = \tilde{\eta}_1/2$.
•Sample $u_j|\boldsymbol{y},\boldsymbol{X},\beta_j,\tilde{\eta}_1,\eta_2$ from the left-truncated exponential distribution Exponential$(0.5)I\{u_j > \sqrt{4\tilde{\eta}_1\eta_2}|\beta_j|\}$ using inversion method which can be done by sampling u_j^* from an exponential distribution with rate parameter 0.5 and setting $u_j = u_j^* + \sqrt{4\tilde{\eta}_1\eta_2}|\beta_j|$.
•Sample $\tilde{\eta}_1$, which has conditional posterior distribution

$$\pi(\tilde{\eta}_1|\boldsymbol{y},\boldsymbol{X},\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{t},\eta_2) \propto \Gamma^{-k}(1/2,\tilde{\eta}_1)\tilde{\eta}_1^{k/2+c_1-1}\exp\left(-\tilde{\eta}_1\left[d_1 + \sum_{j=1}^k t_j\right]\right)I$$

$$\left\{\tilde{\eta}_1 < \text{Min}_j\left(u_j^2/\left(4\eta_2\beta_j^2\right)\right)\right\}$$

Although $\pi(\tilde{\eta}_1|\boldsymbol{y},\boldsymbol{X},\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{t},\eta_2)$ has no closed form, it is log-concave. Thus we sample $\tilde{\eta}_1$ by Adaptive Rejection Sampling (ARS) [40].

•Sample $\eta_2|\boldsymbol{y},\boldsymbol{X},\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{t},\tilde{\eta}_1$ from the right-truncated gamma distribution $\text{Gamma}\left(c_2 + k/2, d_2 + \sum_{j=1}^k t_j\left(t_j-1\right)^{-1}\beta_j^2\right)I\left\{\eta_2 < \text{Min}_j\left(u_j^2/\left(4\tilde{\eta}_1\beta_j^2\right)\right)\right\}$.

3. Simulation studies

In this section, we investigate the prediction accuracy of the proposed methods with comparison to some Bayesian and non-Bayesian methods. The methods in the comparison include:

1. The new Bayesian lasso QReg (NBLQ).
2. The new Bayesian adaptive lasso QReg (NBALQ).
3. The new Bayesian elastic net QReg (NBENQ).
4. The original Bayesian lasso QReg (OBLQ) reported in Ref. [22].
5. The original Bayesian elastic net QReg (OBENQ) reported in Ref. [22].
6. The lasso QReg (lassoQ).
7. The standard QReg (RQ).
8. The lasso regression (lasso).
9. The elastic net regression (EN).

The data are simulated from the true model

$$y_i = \boldsymbol{x'}_i\boldsymbol{\beta} + \varepsilon_i, \quad i,\cdots,n, \tag{20}$$

where ε_i's have the pth quantile equal to 0. 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 median of mean absolute deviations (MMAD), where the median is taken over 100 simulations. In other words, MMAD = median(mean($|\boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta}^{true}|$)). The Bayesian estimates $\hat{\boldsymbol{\beta}}$ are posterior means using 10,000 samples of the Gibbs sampler after burn-in of 1000 iterations. Methods are fitted on the training data set and MMADs are calculated on the testing data set.

3.1. Simulation 1 (sparse case)

In this simulation study, we simulated data from the true model (20), where $\boldsymbol{\beta}$ = (3,1.5, 0,0,2,0,0,0,0,0) and the covariates are simulated independently from $N(0,\boldsymbol{\Sigma})$. The (i,j)th element of $\boldsymbol{\Sigma}$ is 0. $5^{|i-j|}$ and the error term ε_i is restricted to have the pth quantile equal to zero. We assume that the response variable is centered and the covariates are standardized, that is, $\sum_{i=1}^{n} y_i = 0$, $\sum_{i=1}^{n} x_{ij} = 0$ and $\sum_{i=1}^{n} x_{ij}^2/n = 1$, $j = 1, \cdots, k$. Following [22]; we consider four different choices for the error distribution ε_i.

1. $\varepsilon_i \sim N(\mu, 9)$, with μ chosen so that the pth quantile is 0.
2. $\varepsilon_i \sim 0.1N(\mu, 1) + 0.9N(\mu, 5)$, with μ chosen so that the pth quantile is 0.
3. $\varepsilon_i \sim$ Laplace(μ, 3), with μ chosen so that the pth quantile is 0.
4. $\varepsilon_i \sim$ 0.1Laplace(μ, 1) + 0.9Laplace(μ, $\sqrt{5}$), with μ chosen so that the pth quantile is 0.

Table 2 lists the number of times each method has smallest mean absolute deviation (SMAD) and the frequency of correctly-fitted models (FC) over 100 replications. For the Bayesian methods and frequentist QReg (RQ), variable selection is conducted based on the credible intervals and confidence intervals, respectively. It can be seen that the results suggest that the new Gibbs sampler for Bayesian lasso perform better than the other methods in terms of SMAD and FC. Specifically, we see that the new Gibbs sampler for Bayesian lasso has the smallest mean absolute deviation in 8 out of 12 simulation setups and has the highest frequency in identifying the correct model in 7 out of 12 simulation setups.

The inefficiency factors (IFs) were calculated for the regression coefficients to evaluate the sampling efficiency of the Bayesian methods. The inefficiency factor (IF) is defined as a ratio of the numerical variance of the sample mean from the MCMC to the variance from independent draws. The results of the inefficiency factors are summarized in Table 3. It can be observed that the IFs of the proposed algorithms are smaller than those of the old algorithms in all three quantiles under consideration. These results indicate the superiority of our proposed algorithms. Convergence of the proposed algorithms in

Table 2
Frequency of smallest mean absolute deviation (SMAD) and frequency of correctly-fitted models (FC) over 100 replications.

p	Method	Error Distribution			
		normal	normal mixture	Laplace	Laplace mixture
		SMAD (FC)	SMAD (FC)	SMAD (FC)	SMAD (FC)
0.50	NBLQ	14 (26)	20 (23)	21 (19)	22 (31)
	NBALQ	12 (14)	17 (19)	12 (22)	13 (28)
	NBENQ	13 (17)	09 (12)	11 (17)	10 (23)
	OBLQ	09 (12)	10 (16)	14 (10)	13 (20)
	OBENQ	08 (10)	08 (16)	12 (10)	09 (20)
	lassoQ	19 (04)	13 (20)	17 (02)	17 (12)
	RQ	02 (04)	01 (05)	03 (02)	05 (01)
	lasso	12 (32)	13 (28)	06 (14)	04 (18)
	EN	11 (28)	09 (20)	04 (08)	07 (12)
0.75	NBLQ	11 (22)	19 (18)	16 (31)	16 (22)
	NBALQ	13 (16)	10 (13)	11 (19)	10 (20)
	NBENQ	11 (13)	12 (15)	10 (10)	11 (11)
	OBLQ	09 (14)	13 (12)	11 (16)	11 (14)
	OBENQ	07 (12)	11 (10)	12 (10)	14 (12)
	lassoQ	23 (04)	10 (08)	12 (04)	11 (06)
	RQ	04 (04)	02 (03)	06 (01)	08 (05)
	lasso	12 (18)	12 (14)	15 (26)	07 (18)
	EN	10 (10)	11 (04)	07 (10)	12 (08)
0.95	NBLQ	18 (29)	15 (13)	15 (21)	23 (18)
	NBALQ	10 (18)	19 (14)	13 (27)	14 (16)
	NBENQ	12 (20)	16 (11)	13 (23)	14 (16)
	OBLQ	09 (16)	15 (06)	16 (18)	17 (14)
	OBENQ	12 (16)	13 (00)	09 (16)	14 (12)
	lassoQ	11 (01)	01 (00)	02 (04)	02 (00)
	RQ	07 (01)	05 (00)	07 (01)	09 (00)
	lasso	10 (26)	09 (11)	13 (36)	01 (06)
	EN	11 (20)	07 (08)	09 (26)	06 (06)

Table 1
MMADs and the standard deviations of the MADs (in the parentheses) obtained by bootstrap resampling 500 times for Simulation 1.

pz	Method	Error Distribution			
		normal	normal mixture	Laplace	Laplace mixture
		MMAD (SD)	MMAD (SD)	MMAD (SD)	MMAD (SD)
0.50	NBLQ	**0.3896** (0.0125)	0.9027 (0.0477)	0.4247 (0.0211)	1.1451 (0.0459)
	NBALQ	0.3911 (0.0131)	**0.8660** (0.0628)	**0.4210** (0.0157)	1.1545 (0.0566)
	NBENQ	0.4023 (0.0133)	0.9172 (0.0546)	0.4544 (0.0118)	1.1502 (0.0412)
	OBLQ	0.3963 (0.0109)	0.9578 (0.0460)	0.4255 (0.0221)	**1.1440** (0.0484)
	OBENQ	0.4127 (0.0152)	0.9994 (0.0362)	0.4665 (0.0181)	1.1962 (0.0528)
	lassoQ	0.3897 (0.0229)	0.9741 (0.0331)	0.5260 (0.0274)	1.5765 (0.0672)
	RQ	0.4856 (0.0151)	1.2782 (0.0359)	0.5620 (0.0136)	1.5601 (0.0496)
	lasso	0.3932 (0.0228)	0.9739 (0.0335)	0.5283 (0.0268)	1.5690 (0.0661)
	EN	0.4345 (0.0283)	1.0795 (0.0502)	0.6253 (0.0457)	1.5384 (0.0943)
0.75	NBLQ	0.3936 (0.0164)	1.1022 (0.0226)	0.4456 (0.0273)	1.3104 (0.0574)
	NBALQ	**0.3802** (0.0103)	**1.0866** (0.0280)	**0.4404** (0.0246)	1.3019 (0.0665)
	NBENQ	0.4014 (0.0167)	1.1123 (0.0251)	0.4631 (0.0309)	**1.2481** (0.0348)
	OBLQ	0.3939 (0.0142)	1.0880 (0.0196)	0.4380 (0.0253)	1.3040 (0.0412)
	OBENQ	0.4030 (0.0130)	1.1860 (0.0435)	0.4874 (0.0238)	1.2960 (0.0442)
	lassoQ	0.3984 (0.0182)	1.1748 (0.0508)	0.5442 (0.0257)	1.5911 (0.0880)
	RQ	0.4994 (0.0104)	1.5137 (0.0356)	0.5644 (0.0240)	1.6425 (0.0718)
	lasso	0.3973 (0.0177)	1.1691 (0.0502)	0.5521 (0.0260)	1.5893 (0.0842)
0.95	EN	0.4178 (0.0217)	1.2481 (0.0300)	0.5631 (0.0391)	1.5319 (0.0802)
	NBLQ	0.3803 (0.0117)	1.3896 (0.0450)	0.4287 (0.0180)	1.5638 (0.0505)
	NBALQ	**0.3597** (0.0148)	**1.3061** (0.0609)	**0.3994** (0.0234)	1.5581 (0.0798)
	NBENQ	0.3738 (0.0119)	1.3664 (0.0473)	0.4463 (0.0237)	**1.5548** (0.0557)
	OBLQ	0.3701 (0.0126)	1.3538 (0.0722)	0.4203 (0.0222)	1.5774 (0.0388)
	OBENQ	0.3970 (0.0120)	1.4118 (0.0438)	0.4497 (0.0207)	1.5989 (0.0462)
	lassoQ	0.3936 (0.0153)	1.5366 (0.0480)	0.5320 (0.0322)	2.0508 (0.0835)
	RQ	0.4823 (0.0207)	1.7885 (0.0547)	0.5477 (0.0179)	2.2735 (0.0684)
	lasso	0.3832 (0.0157)	1.5427 (0.0476)	0.5332 (0.0347)	2.0513 (0.0846)
	ENS	0.4408 (0.0272)	1.5082 (0.0580)	0.5741 (0.0259)	1.9656 (0.0900)

The results are listed in Table 1 which clearly suggest that the proposed methods outperform the other methods in general across all scenarios of this simulation study. Most noticeably, the proposed new Bayesian adaptive lasso QReg (NBALQ) performs better than the other methods. It has the smallest MMAD in 8 out of 12 simulation setups. In two situations, the proposed new Bayesian elastic net QReg (NBENQ) performs slightly better than the NBALQ. Overall, the Bayesian approaches significantly outperform the frequentist approaches in terms of prediction accuracy.

Table 3
Inefficiency factors are shown.

p	Method	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
0.50	NBLQ	3.193	3.617	3.185	3.227	3.519	3.662	3.289	3.771
	NBALQ	3.662	4.005	3.971	3.667	3.409	3.718	3.919	3.948
	NBENQ	4.011	3.892	3.615	3.157	3.472	3.163	3.227	3.227
	OBLQ	5.387	5.782	4.892	5.371	5.449	4.817	4.779	5.138
	OBENQ	5.104	5.392	5.861	4.992	5.753	4.973	4.819	5.003
0.75	NBLQ	8.916	8.265	8.713	9.360	8.715	8.332	8.419	8.573
	NBALQ	8.337	8.618	8.338	8.197	8.119	9.004	8.817	9.113
	NBENQ	8.518	9.107	8.819	9.219	8.663	8.916	8.991	9.461
	OBLQ	10.471	10.229	11.931	11.945	10.337	9.789	11.119	10.997
	OBENQ	11.015	10.873	10.116	11.562	10.452	10.837	10.673	9.661
0.95	NBLQ	12.169	12.883	12.720	12.927	13.201	12.825	10.673	11.660
	NBALQ	11.937	12.027	11.893	12.883	11.946	13.617	11.993	12.338
	NBENQ	12.993	11.883	13.782	11.927	12.937	12.773	11.827	12.563
	OBLQ	14.930	15.662	14.913	15.903	15.663	14.783	14.937	15.787
	OBENQ	15.383	16.407	15.689	15.668	14.689	14.839	15.937	15.788

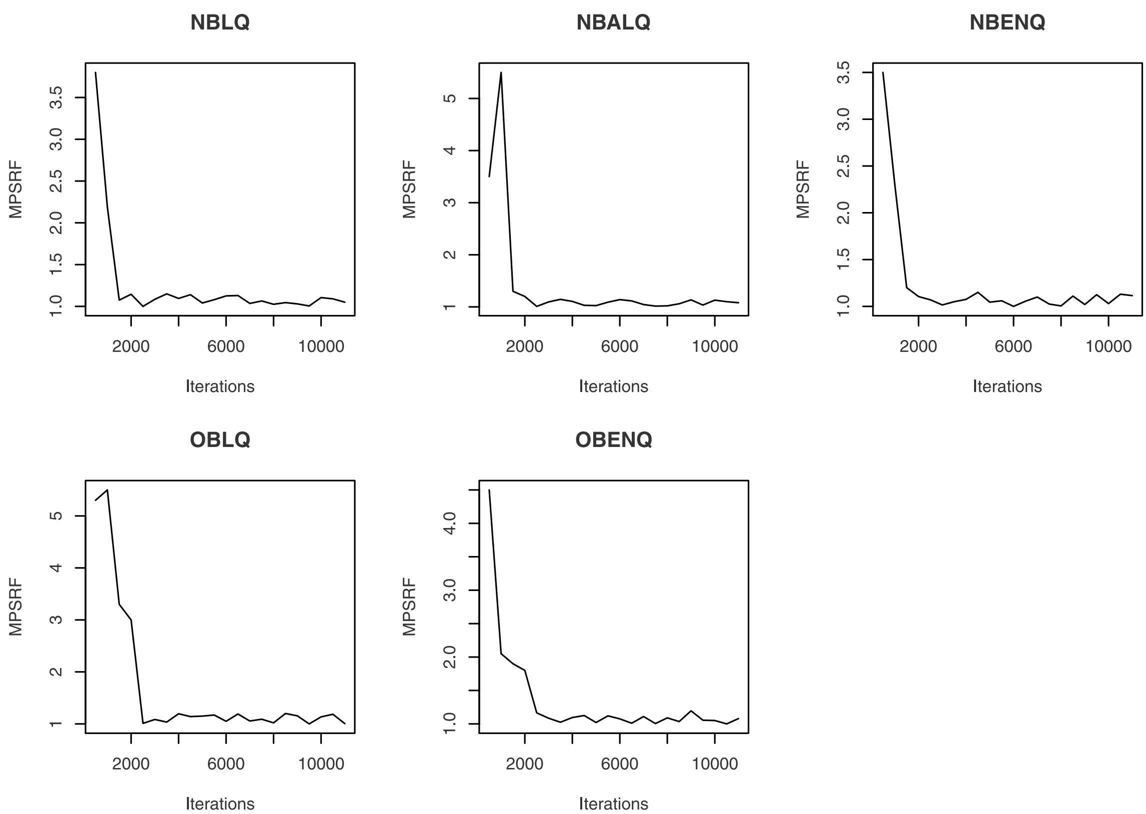

Fig. 1. MPSRF for the Bayesian methods in Simulation 1. The given quantile is $p = 0.50$.

the simulations was observed using the multivariate potential scale reduction factor (MPSRF) [33]. Fig. 1 shows the MPSRF for Simulation 1 when $p = 0.50$ and $\varepsilon_i \sim N(\mu, 9)$. From Fig. 1, it can be seen that the convergence of the proposed algorithms to the posterior was very fast and the chain mixing was well compared to the existing algorithms (OBLQ and OBENQ). Similar patterns are also seen for the situations not shown here.

3.2. Simulation 2 (very sparse case)

The setup for the second scenario is the same as the first, except we set $\boldsymbol{\beta} = (5,0,0,0,0,\ 0,0,0,0,0)$. The results are listed in Table 4 which clearly suggest that the proposed methods outperform the others. Again, in general, the proposed NBALQ performs better than the other methods. It has the smallest MMAD in 10 out of 12 simulation setups.

3.3. Simulation 3 (dense case)

The setup for this simulation study is the same as the first, except we set $\boldsymbol{\beta} = (1.15, 1.00, 0.85, 0.60, 0.45, 1.15, 1.00, 0.85, 0.60, 0.45)$.

Table 5 summarizes the experimental results for this simulation study. We can observe that the proposed methods perform better than the others and all the Bayesian methods outperform their frequentist counterpart. Again, NBALQ performs better than the other methods. It has the smallest MMAD in 9 out of 12 simulation setups.

3.4. Simulation 4 (high correlation case)

The setup for this simulation study is the same as the first, except we simulated the covariates from the multivariate normal distribution with mean 0, variance 1 and pairwise correlations between x_i and x_j equal to 0.95.

Table 4
MMADs and the standard deviations of the MADs (in the parentheses) obtained by bootstrap resampling 500 times for Simulation 2.

p	Method	Error Distribution			
		normal	normal mixture	Laplace	Laplace mixture
		MMAD (SD)	MMAD (SD)	MMAD (SD)	MMAD (SD)
0.50	NBLQ	0.3332 (0.0107)	0.7577 (0.0292)	0.3601 (0.0167)	0.8643 (0.0403)
	NBALQ	0.3278 (0.0155)	**0.6972** (0.0286)	0.3332 (0.0221)	**0.7334** (0.0408)
	NBENQ	0.3594 (0.0114)	0.8126 (0.0350)	0.3574 (0.0141)	0.9136 (0.0417)
	OBLQ	0.3449 (0.0131)	0.7848 (0.0246)	0.3394 (0.0096)	0.8178 (0.0490)
	OBENQ	0.3678 (0.0149)	0.8918 (0.0222)	0.3740 (0.0135)	1.0696 (0.0305)
	lassoQ	0.4183 (0.0118)	1.1172 (0.0387)	0.4698 (0.0157)	1.3367 (0.0609)
	RQ	**0.2759** (0.0139)	0.7919 (0.0492)	**0.2998** (0.0250)	0.7589 (0.0421)
	lasso	0.3755 (0.0297)	0.9323 (0.0569)	0.5422 (0.0402)	1.1302 (0.0710)
	EN	0.3783 (0.0224)	0.9492 (0.0486)	0.5378 (0.0349)	1.2080 (0.0667)
0.75	NBLQ	0.3109 (0.0108)	0.8688 (0.0344)	0.3759 (0.0162)	0.9360 (0.0614)
	NBALQ	**0.2935** (0.0098)	**0.7187** (0.0401)	**0.3481** (0.0159)	**0.8401** (0.0589)
	NBENQ	0.3187 (0.0112)	0.8715 (0.0341)	0.3794 (0.0164)	0.9355 (0.0586)
	OBLQ	0.3221 (0.0091)	0.8364 (0.0287)	0.3724 (0.0149)	0.9657 (0.0715)
	OBENQ	0.3431 (0.0136)	1.0098 (0.0403)	0.3986 (0.0203)	1.1453 (0.0530)
	lassoQ	0.4193 (0.0088)	1.2982 (0.0378)	0.4972 (0.0161)	1.4104 (0.0725)
	RQ	0.3138 (0.0179)	0.9513 (0.0565)	0.3967 (0.0263)	1.1363 (0.0682)
	lasso	0.3641 (0.0243)	0.8612 (0.0710)	0.5380 (0.0362)	1.2197 (0.0716)
	EN	0.3585 (0.0234)	1.0356 (0.0904)	0.5473 (0.0458)	1.2520 (0.0585)
0.95	NBLQ	0.3216 (0.0133)	0.9305 (0.0519)	0.3499 (0.0093)	1.1154 (0.0562)
	NBALQ	**0.3197** (0.0152)	**0.7150** (0.0512)	**0.3254** (0.0098)	**0.8538** (0.0522)
	NBENQ	0.3426 (0.0123)	1.1023 (0.0634)	0.3653 (0.0140)	1.1922 (0.0705)
	OBLQ	0.3387 (0.0175)	0.9499 (0.0549)	0.3565 (0.0117)	1.1150 (0.0590)
	OBENQ	0.3480 (0.0155)	1.2358 (0.0518)	0.3820 (0.0133)	1.3911 (0.0693)
	lassoQ	0.4088 (0.0198)	1.5608 (0.0738)	0.4693 (0.0240)	1.9858 (0.0902)
	RQ	0.5110 (0.0140)	2.5940 (0.0750)	0.7885 (0.0307)	3.7911 (0.1193)
	lasso	0.3617 (0.0340)	1.1120 (0.0810)	0.4344 (0.0341)	1.5986 (0.0777)
	EN	0.3950 (0.0269)	1.2068 (0.0659)	0.4845 (0.0530)	1.7321 (0.0912)

Table 5
MMADs and the standard deviations of the MADs (in the parentheses) obtained by bootstrap resampling 500 times for Simulation 3.

p	Method	Error Distribution			
		normal	normal mixture	Laplace	Laplace mixture
		MMAD (SD)	MMAD (SD)	MMAD (SD)	MMAD (SD)
0.50	NBLQ	0.7245 (0.0637)	0.5063 (0.0371)	0.6789 (0.0389)	0.5617 (0.0501)
	NBALQ	0.7183 (0.0605)	**0.4639** (0.0356)	**0.6719** (0.0371)	0.5217 (0.0532)
	NBENQ	0.7225 (0.0593)	0.4891 (0.0401)	0.7013 (0.0355)	0.5312 (0.0587)
	OBLQ	0.7495 (0.0616)	0.5172 (0.0339)	0.7554 (0.0432)	0.5783 (0.0577)
	OBENQ	0.7799 (0.0748)	0.4891 (0.0410)	0.6873 (0.0393)	**0.5134** (0.0461)
	lassoQ	**0.7038** (0.0750)	0.5465 (0.0547)	0.7495 (0.1043)	0.7327 (0.0642)
	RQ	0.7048 (0.0835)	0.5939 (0.0302)	0.7202 (0.0964)	0.7663 (0.0599)
	lasso	0.8658 (0.1019)	0.5092 (0.0335)	0.7901 (0.0513)	0.5852 (0.0771)
	EN	0.7078 (0.0741)	0.5447 (0.0588)	0.7791 (0.1178)	0.7693 (0.0701)
0.75	NBLQ	**0.6011** (0.0399)	0.6393 (0.0524)	0.5381 (0.0934)	0.5771 (0.0701)
	NBALQ	0.6395 (0.0409)	**0.5944** (0.0509)	**0.5209** (0.0842)	**0.5739** (0.0611)
	NBENQ	0.6274 (0.0402)	0.6281 (0.0604)	0.5433 (0.0847)	0.7382 (0.0599)
	OBLQ	0.6316 (0.0480)	0.7033 (0.0695)	0.6603 (0.1123)	0.6381 (0.0629)
	OBENQ	0.6077 (0.0413)	0.6546 (0.0649)	0.5354 (0.0894)	0.5882 (0.0642)
	lassoQ	0.6857 (0.0463)	0.5950 (0.1042)	0.7296 (0.1294)	0.7166 (0.0937)
	RQ	0.6259 (0.0620)	0.6387 (0.0734)	0.8856 (0.0855)	0.8165 (0.0888)
	lasso	0.7049 (0.0585)	0.6972 (0.0417)	0.6337 (0.0787)	0.6308 (0.0553)
	EN	0.6857 (0.0463)	0.5951 (0.1042)	0.7296 (0.1294)	0.7166 (0.0937)
0.95	NBLQ	0.5667 (0.0402)	0.7119 (0.0548)	0.6135 (0.0605)	0.8035 (0.0.603)
	NBALQ	**0.5634** (0.0553)	**0.6825** (0.0619)	**0.5987** (0.0703)	**0.7748** (0.0857)
	NBENQ	0.6001 (0.0132)	0.7481 (0.0679)	0.6073 (0.0647)	0.7911 (0.0522)
	OBLQ	0.6321 (0.0559)	0.7736 (0.0654)	0.6515 (0.0668)	0.8514 (0.0739)
	OBENQ	0.5792 (0.0361)	0.7145 (0.0831)	0.6074 (0.0721)	0.7802 (0.0591)
	lassoQ	0.6019 (0.0947)	0.6918 (0.0827)	0.9453 (0.1547)	1.2307 (0.1243)
	RQ	0.6428 (0.0919)	0.8092 (0.1911)	0.9441 (0.0762)	1.5213 (0.1427)
	lasso	0.7035 (0.0374)	0.8417 (0.1242)	0.6206 (0.0582)	0.8842 (0.0350)
	EN	0.6019 (0.0947)	0.6918 (0.0827)	0.9453 (0.1547)	1.2307 (0.1243)

Table 6
MMADs for Simulation 4. In the parentheses are standard deviations of the MADs obtained by bootstrap resampling 500 times.

p	Method	Error Distribution			
		normal	normal mixture	Laplace	Laplace mixture
		MMAD (SD)	MMAD (SD)	MMAD (SD)	MMAD (SD)
0.50	NBLQ	0.4467 (0.0343)	1.1464 (0.2304)	0.4171 (0.0386)	1.4857 (0.4152)
	NBALQ	0.4417 (0.0365)	1.1294 (0.2324)	**0.4014** (0.0329)	1.4785 (0.3739)
	NBENQ	**0.4040** (0.0320)	1.0007 (0.2051)	0.4191 (0.0380)	**1.0354** (0.3507)
	OBLQ	0.4940 (0.0530)	1.2393 (0.3201)	0.4423 (0.1056)	1.6583 (0.4799)
	OBENQ	0.4046 (0.0250)	1.0843 (0.1008)	0.4963 (0.0432)	1.1216 (0.4035)
	lassoQ	0.5367 (0.1663)	1.2535 (0.2047)	0.8233 (0.2668)	1.9290 (0.2590)
	RQ	0.5190 (0.0259)	1.3615 (0.1048)	0.5021 (0.0401)	1.5493 (0.1198)
	lasso	0.5367 (0.1663)	1.2535 (0.2047)	0.8233 (0.2668)	1.9290 (0.2590)
	EN	0.6109 (0.0751)	**0.9269** (0.1652)	0.6927 (0.0885)	1.1385 (0.2469)
0.75	NBLQ	0.4590 (0.0493)	1.3381 (0.1557)	0.4504 (0.0717)	1.3035 (0.1007)
	NBALQ	0.4519 (0.0481)	1.4380 (0.1074)	0.4467 (0.0651)	1.3112 (0.1339)
	NBENQ	0.5119 (0.0455)	1.0090 (0.1152)	**0.4297** (0.0483)	**1.0777** (0.1290)
	OBLQ	0.4555 (0.0533)	1.2842 (0.2831)	0.4950 (0.0952)	1.3604 (0.1828)
	OBENQ	0.5028 (0.0579)	1.1141 (0.0964)	0.5425 (0.0497)	1.4538 (0.1979)
	lassoQ	0.4594 (0.1595)	1.2507 (0.2618)	0.7751 (0.1982)	1.6195 (0.1947)
	RQ	0.4410 (0.0225)	1.5006 (0.1219)	0.5623 (0.0678)	1.5808 (0.1573)
	lasso	0.4594 (0.1595)	1.2507 (0.2618)	0.7751 (0.1982)	1.6195 (0.1947)
	EN	**0.4042** (0.0472)	**0.7659** (0.0259)	0.5630 (0.0307)	1.2455 (0.2305)
0.95	NBLQ	0.5017 (0.0475)	1.1292 (0.2546)	0.5420 (0.1021)	1.1417 (0.1268)
	NBALQ	0.5146 (0.0410)	1.0149 (0.2029)	0.5586 (0.0468)	**1.0109** (0.1318)
	NBENQ	**0.4937** (0.0444)	**1.0064** (0.2046)	**0.5152** (0.0586)	1.2291 (0.1713)
	OBLQ	0.5082 (0.0510)	1.1334 (0.4551)	0.5410 (0.1229)	1.3715 (0.2703)
	OBENQ	0.5415 (0.0520)	1.7180 (0.2032)	0.5746 (0.0402)	1.6670 (0.3787)
	lassoQ	0.9421 (0.1830)	1.4699 (0.2924)	0.9335 (0.1333)	2.3301 (0.2748)
	RQ	0.5375 (0.0278)	1.6958 (0.2202)	0.6015 (0.0449)	2.0663 (0.1671)
	lasso	0.9421 (0.1830)	1.4699 (0.2924)	0.9335 (0.1333)	2.3301 (0.2748)
	EN	0.5352 (0.0533)	1.1450 (0.1792)	0.6128 (0.0721)	1.5326 (0.2304)

The results of this simulation study are listed in Table 6. This table shows that, in terms of the MMAD, the new Bayesian elastic net QReg (NBENQ) method performs better than the other methods in general, especially for $p = 0.95$. In this simulation study, NBENQ has the smallest MMAD in 7 out of 12 simulation setups.

3.5. Simulation 5 (small n large k case)

Here we set $\boldsymbol{\beta} = \left(\underbrace{5, \cdots, 5}_{5}, \underbrace{0, \cdots, 0}_{20}, \underbrace{5, \cdots, 5}_{5}\right)'$, leaving other setups exactly same as Simulation 1.

Table 7
MMADs and the standard deviations of the MADs (in the parentheses) obtained by bootstrap resampling 500 times for Simulation 5.

p	Method	Error Distribution			
		normal	normal mixture	Laplace	Laplace mixture
		MMAD (SD)	MMAD (SD)	MMAD (SD)	MMAD (SD)
0.50	NBLQ	1.1951 (0.0464)	2.8569 (0.1421)	1.8200 (0.2092)	4.1713 (0.1565)
	NBALQ	**0.9933** (0.0368)	**2.0720** (0.1504)	**1.3995** (0.1475)	**2.9149** (0.1665)
	NBENQ	1.2011 (0.0550)	2.7728 (0.1769)	1.7282 (0.2475)	3.9405 (0.2035)
	OBLQ	1.2118 (0.0460)	2.8495 (0.1561)	1.7526 (0.1896)	3.9165 (0.2310)
	OBENQ	1.2358 (0.0799)	3.1102 (0.2403)	1.9204 (0.2556)	4.5815 (0.2226)
	lassoQ	2.2533 (0.1835)	2.7465 (0.4698)	2.3619 (0.2329)	4.1852 (0.4205)
	lasso	2.2533 (0.1835)	2.7465 (0.4698)	2.3619 (0.2329)	4.1852 (0.4205)
	EN	3.6174 (0.0557)	4.0387 (0.2501)	3.6752 (0.1378)	4.9782 (0.2205)
0.75	NBLQ	1.2272 (0.1013)	3.3010 (0.2042)	1.5669 (0.0672)	4.2133 (0.3797)
	NBALQ	**0.9569** (0.0615)	**2.4630** (0.1680)	**1.2472** (0.0744)	**2.8940** (0.1681)
	NBENQ	1.2546 (0.1022)	3.5215 (0.2654)	1.6100 (0.0942)	4.4178 (0.4241)
	OBLQ	1.2807 (0.0924)	3.3745 (0.2352)	1.6508 (0.1116)	4.0921 (0.6180)
	OBENQ	1.3041 (0.0857)	3.7569 (0.2385)	1.6768 (0.1172)	4.6787 (0.5010)
	lassoQ	2.0040 (0.1555)	3.5075 (0.1141)	1.9094 (0.2493)	4.7625 (0.4554)
	lasso	2.0040 (0.1555)	3.5075 (0.1141)	1.9094 (0.2493)	4.7625 (0.4554)
	EN	3.6642 (0.0391)	4.6993 (0.1639)	3.5050 (0.0904)	5.5977 (0.6178)
0.95	NBLQ	1.2809 (0.1273)	4.3376 (0.2565)	1.6888 (0.1241)	5.6210 (0.2540)
	NBALQ	**1.0338** (0.0915)	**2.8228** (0.2304)	**1.2378** (0.0796)	**5.1112** (0.6638)
	NBENQ	1.1544 (0.1290)	4.3812 (0.2352)	1.6130 (0.0997)	5.6035 (0.2977)
	OBLQ	1.2653 (0.1575)	4.1757 (0.2515)	1.7394 (0.1539)	5.6021 (0.3020)
	OBENQ	1.2133 (0.1568)	4.7279 (0.2294)	1.7221 (0.1249)	5.9260 (0.2192)
	lassoQ	1.8536 (0.2030)	4.2865 (1.6664)	1.7546 (0.1564)	6.1333 (0.8765)
	lasso	1.8536 (0.2030)	4.2865 (1.6664)	1.7546 (0.1564)	6.1333 (0.8765)
	EN	3.5497 (0.0812)	5.2208 (0.4541)	3.5498 (0.1393)	6.9135 (0.3882)

Table 7 summarizes the experimental results for this simulation study. It can be sseen that the NBALQ method significantly outperform the other methods in terms of prediction accuracy. It has the smallest MMAD in all simulation setups of this example. We can also see that the Bayesian methods significantly outperform the frequentist counterparts. Another comment on Table 7 is that, due to the case $p > n$, the design matrix is singular. Thus, the frequentist quantile regression approach (RQ) fails in this situation, while the other eight methods still work. This result showing the advantage of the regularization methods over the non-regularization methods.

3.6. Simulation 6 (heterogeneous random errors)

In this simulation study, we consider the case with non-i.i.d random errors. The data set were simulated according to the model [17,22,34],

$$y_i = 1 + x_{1i} + x_{2i} + x_{3i} + (1 + x_{3i})\varepsilon_i,$$

where $x_{1i} \sim \mathrm{N}(0,1)$, $x_{3i} \sim \mathrm{uniform}[0,1]$ and $x_{2i} = x_{1i} + x_{3i} + w_i$, where $w_i \sim \mathrm{N}(0,1)$. $\varepsilon_i \sim N(\mu, 1)$, with μ chosen so that the pth quantile is 0. Here, x_{1i}, x_{3i}, w_i, and ε_i were mutually independent. Five more independent noise covariates ($x_4, \cdots, x_8 \sim \mathrm{N}(0,1)$) were included to the model. 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 [17,22]. Here, the test error refers to the median of the average check loss (MACL) on the independent testing data set, i.e. $\mathrm{MACL} = \mathrm{median}(\mathrm{mean}(\rho_p(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})))$, where the median is taken over 100 simulations.

The results of this simulation study are listed in Table 8. The simulation results show that, in terms of MMAD and the test error, the new Bayesian regularized QReg methods still perform well even in the case of heterogeneity.

4. Real data analyses

Here, we apply the proposed methods to two benchmark datasets viz. Boston housing data [35] and prostate cancer data [36]. Methods are evaluated based on median of mean squared errors (MMSE) and the average check loss (test error) on the independent testing data set [17,22].

4.1. Boston Housing data

The corrected version of the this data set has 506 rows and 15 columns (one response variable and 15 covariates), and is available in the mlbench package [37] in R. The outcome of interest is the log-transformed corrected median value of owner-occupied housing in USD 1000 (LCMEDV). The 15 covariates are the point longitudes in decimal degrees (x_1), point latitudes in decimal degrees (x_2), per capita crime rate (x_3), proportion of residential land zoned for lots over 25,000 sq.ft (x_4), proportions of non-retail business acres per town which is constant for all Boston tracts (x_5), Charles River ($x_6 = 1$ if tract bounds river; 0 otherwise), nitric oxides concentration per town (x_7), average numbers of rooms per dwelling (x_8), proportion of owner-occupied units built prior to 1940 (x_9), the weighted distances to 5 Boston employment centres (x_{10}), index of accessibility to radial highways (x_{11}), full-value property-tax rate per 10,000$ (x_{12}), pupil-teacher ratio by town (x_{13}), 1000 (proportion of blacks − 0.63) (x_{14}) and the lower status of the population (x_{15}). We centered the response variable to have mean zero, while the covariates have been standardized.

We run 10-fold cross-validation to evaluate the performance of the

Table 8
MMADs and test errors for the Simulation 6.

Method	$p = 0.50$		$p = 0.75$		$p = 0.95$	
	MMAD	Test Error	MMAD	Test Error	MMAD	Test Error
NBLQ	0.5823	0.6660	**0.6266**	0.5447	0.9273	0.2160
NBALQ	0.5812	0.6652	0.6289	**0.5409**	**0.9004**	**0.2133**
NBENQ	**0.5806**	**0.6610**	0.6298	0.5418	0.9010	0.2150
OBLQ	0.6060	0.6728	0.6313	0.5454	0.9373	0.2165
OBENQ	0.5851	0.6677	0.6402	0.5445	0.9027	0.2198
lassoQ	0.6142	0.6782	0.6762	0.5450	1.0556	0.2377
RQ	0.7702	0.7139	0.8526	0.5980	1.2906	0.2987
lasso	1.6413	0.9582	1.7220	0.8935	1.8501	0.4267
EN	1.6686	0.9660	1.6946	0.8605	1.8418	0.4326

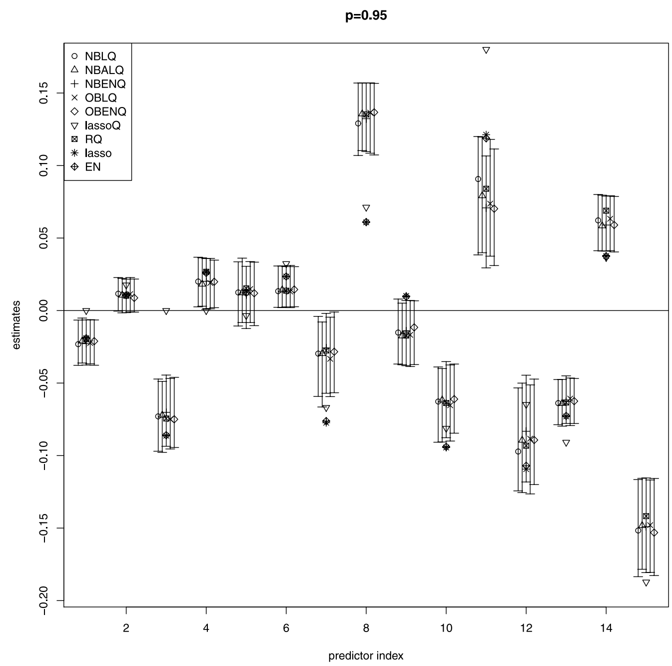

Fig. 2. For the Boston Housing data, posterior mean estimates and corresponding 95% equal-tailed credible intervals for the Bayesian methods. The given quantile is $p = 0.95$.

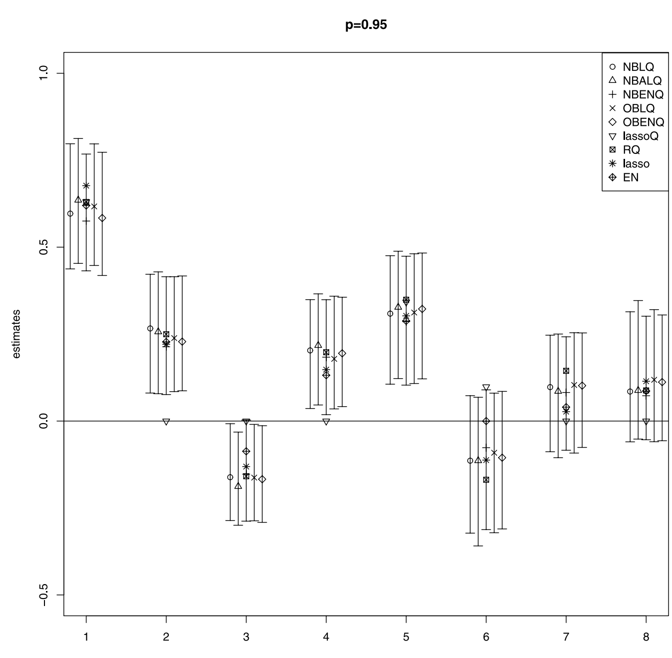

Fig. 3. For the prostate cancer data, posterior mean estimates and corresponding 95% equal-tailed credible intervals for the Bayesian methods. The given quantile is $p = 0.95$.

proposed methods. The results are summarized in Table E9 (Appendix E), which show that the proposed methods peroform better than existing Bayesian and non-Baysian methods in terms of prediction accuracy. Fig. 2 describe the point and 95% equal-tailed credible interval estimations for the parameters when $p = 0.95$. We can observe that the proposed methods give very similar posterior mean estimates compared to the original Bayesian lasso and elastic net methods. These interval estimates can be used as a guide for variable selection. From Fig. 2, we also see that the Bayesian approaches identify the important and unimportant covariates based on the credible intervals and any seen differences must be attributed to the properties of the different MCMC samplers used to obtain draws from the posterior distributions.

4.2. Prostate cancer data

We consider the prostate cancer data [36] to demonstrate the performance of the proposed methods. This data set has 97 rows and 9 columns (one response variable and 8 clinical measures), and is available in the bayesQR package [38] in R. The response variable is the level of prostate specic antigen and the eight clinical measures are log cancer volume (x_1), log prostate weight (x_2), age (x_3), log of the amount of benign prostatic hyperplasia (x_4), seminal vesicle invasion (x_5), log of capsular penetration (x_6), Gleason score (x_7), and percentage of Gleason scores 4 or 5 (x_8). We centered the response variable to have mean zero, while the covariates have been standardized.

We run 5-fold cross-validation to evaluate the performance of the proposed methods. The results are summarized in Table E.10 (Appendix E), which show that the performance of the proposed methods appear quite good compared to the other methods. Fig. 3 describe the point and 95% equal-tailed credible interval estimations for the parameters when $p = 0.95$. From this figure we can observe that the Bayesian quantile regression methods tend to behave similarly while the non Bayesian methods also behave similarly. We can observe that the proposed methods give very similar posterior mean estimates compared to the original Bayesian lasso and elastic net methods. The observed differences in parameter estimates and 95% credible intervals can be contributed to the properties of the different Gibbs samplers used to obtain samples from the corresponding posteriors.

5. Conclusion

In this paper, we have presented new Bayesian hierarchical representations of lasso, adaptive lasso and elastic net QReg models. We explored these representations by observing that the lasso penalty function corresponds to a scale mixture of truncated normal distribution (with exponential mixing densities). Simulation studies real data examples show that the proposed methods are effective in prediction accuracy under a variety of scenarios. In many situations, the proposed methods are outperform the other Bayesian and non-Bayesian methods. This attractive performance of the proposed methods can be contributed (up to Monte Carlo error) to the properties of the different Gibbs samplers used to obtain samples from the corresponding posteriors.

Appendix A. Proof of Proposition

It is well known that [15].

$$2\exp\left\{-\frac{\eta|x|}{2}\right\} = \int_{z>\sqrt{\eta^2}|x|}\exp\left\{-\frac{z}{2}\right\}dz,$$

and [24].

$$\frac{\eta}{4}\exp\left\{-\frac{\eta|x|}{2}\right\} = \int_0^\infty \frac{1}{\sqrt{2\pi s}}\exp\left(-\frac{x^2}{2s}\right)\frac{\eta^2}{8}\exp\left(-\frac{\eta^2}{8}s\right)ds.$$

The pdf of a Laplace distribution with mean 0 and variance $1/\eta$ can be written as

$$\frac{\eta}{2}\exp\{-\eta|x|\} = \frac{\eta}{2}\exp\left\{-\frac{\eta\mid x\mid}{2}\right\}\exp\left\{-\frac{\eta\mid x\mid}{2}\right\},$$
$$= \int_0^\infty\int_{z>\sqrt{\eta^2}|x|}\frac{1}{\sqrt{2\pi s}}\exp\left(-\frac{x^2}{2s}\right)\exp\left\{-\frac{z}{2}\right\}\frac{\eta^2}{8}\exp\left(-\frac{\eta^2}{8}s\right)dzds.$$

Appendix B. Conditional densities for the lasso QReg

The derivations below follow the ordering as listed in Algorithm 1.

The full conditional distribution of each β_j, denoted by $\pi(\beta_j|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}_{-j}, \sigma, \boldsymbol{v}, s_j, u_j, \eta^2)$, is proportional to $\pi(\boldsymbol{y}|\boldsymbol{\beta}, \sigma, \boldsymbol{v})\pi(\beta_j|s_j, u_j, \eta^2)$, where $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)$ and $\boldsymbol{\beta}_{-j}$ is the vector $\boldsymbol{\beta}$ excluding the element β_j. Thus, we have

$$\pi(\beta_j|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}_{-j}, \sigma, \boldsymbol{v}, s_j, u_j, \eta^2) \propto \pi(\boldsymbol{y}|\boldsymbol{\beta}, \sigma, \boldsymbol{v})\pi(\beta_j|s_j, u_j, \eta^2)$$
$$\propto\exp\left\{-\frac{\sigma}{2}\sum_{i=1}^n\frac{(y_i-\boldsymbol{x}'_i\boldsymbol{\beta}-\delta_1 v_i)^2}{\delta_2^2 v_i}\right\}\exp\left\{-\frac{\beta_j^2}{2s_j}\right\}I\left\{|\beta_j|<\frac{u_j}{\sqrt{\eta^2}}\right\},$$
$$\propto\exp\left\{-\frac{1}{2}\left[\left(\sum_{i=1}^n\frac{\sigma x_{ij}^2}{\delta_2^2 v_i}+\frac{1}{s_j}\right)\beta_j^2-2\sum_{i=1}^n\frac{\sigma\tilde{y}_i x_{ij}}{\delta_2^2 v_i}\beta_j\right]\right\}I\left\{|\beta_j|<\frac{u_j}{\sqrt{\eta^2}}\right\},$$

where $\boldsymbol{x}_i = (x_{i1}, \cdots, x_{ik})'$ and $\tilde{y}_i = y_i - \sum_{h=1,h\neq j}^k x_{ih}\beta_h - \delta_1 v_i$. Then the full conditional distribution for β_j is a truncated normal with mean μ_j and variance $\tau_{\beta_j}^2$, where

$$\tau_{\beta_j}^{-2} = \sum_{i=1}^n\frac{\sigma x_{ij}^2}{\delta_2^2 v_i}+\frac{1}{s_j}, \quad \text{and} \quad \mu_j = \tau_{\beta_j}^2\sum_{i=1}^n\frac{\sigma\tilde{y}_i x_{ij}}{\delta_2^2 v_i}.$$

The fully conditional posterior of σ is as follows

$$\pi(\sigma|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{v}) \propto \pi(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}, \sigma, \boldsymbol{v})\pi(\sigma)$$
$$\propto (\sigma)^{\frac{3n}{2}+a-1} \exp\left\{-\left[\sum_{i=1}^{n} \left(\frac{\sigma(y_i - \boldsymbol{x}'_i\boldsymbol{\beta} - \delta_1 v_i)^2}{2\delta_2^2 v_i} + \sigma v_i\right) + b\right]\right\}$$

Hence,

$$\sigma|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{v} \sim \text{Gamma}\left(\frac{3n}{2} + a, \sum_{i=1}^{n}\left(\frac{(y_i - \boldsymbol{x}'_i\boldsymbol{\beta} - \delta_1 v_i)^2}{2\delta_2^2 v_i} + v_i\right) + b\right)$$

The full conditional distribution of each v_i, denoted by $\pi(v_i|y_i, \boldsymbol{\beta}, \sigma)$, is proportional to $\pi(y_i|v_i, \boldsymbol{\beta}, \sigma)\pi(v_i|\sigma)$. Thus, we have

$$\begin{aligned}\pi(v_i|y_i, \boldsymbol{\beta}, \sigma) &\propto v_i^{-1/2} \exp\left\{-\frac{\sigma}{2}\left(\frac{y_i - \boldsymbol{x}'_i\boldsymbol{\beta} - \delta_1 v_i}{\delta_2\sqrt{v_i}}\right)^2 - \sigma v_i\right\}\\ &\propto v_i^{-1/2} \exp\left\{-\frac{\sigma}{2}\left(\frac{(y_i - \boldsymbol{x}'_i\boldsymbol{\beta})^2 + \delta_1^2 v_i^2 - 2\delta_1 v_i(y_i - \boldsymbol{x}'_i\boldsymbol{\beta})}{\delta_2^2 v_i} + 2v_i\right)\right\}\\ &\propto v_i^{-1/2} \exp\left\{-\frac{\sigma}{2}\left[\frac{(y_i - \boldsymbol{x}_i{}'\boldsymbol{\beta})^2}{\delta^2} v_i^{-1} + \left(\frac{\delta_1^2}{\delta_2^2} + 2\right)v_i\right]\right\}\\ &\propto v_i^{-1/2} \exp\left\{-\frac{1}{2}[\varrho_1^2 v_i^{-1} + \varrho_2^2 v_i]\right\},\end{aligned}$$

Thus, the full conditional distribution of each v_i is a generalized inverse Gaussian distribution GIG $(\nu, \varrho_1, \varrho_2)$, where $\nu = 0.5$, $\varrho_1^2 = \sigma(y_i - \boldsymbol{x}'_i\boldsymbol{\beta})^2/\delta_2^2$ and $\varrho_2^2 = \sigma\left(\frac{\delta_1^2}{\delta_2^2} + 2\right)$. Recall that if $x\sim$ GIG $(\nu, \varrho_1, \varrho_2)$ then the pdf of x is given by Ref. [39],

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

where $x > 0$, $-\infty < \nu < \infty$, $\varrho_1, \varrho_2 \geq 0$ and $K_\nu(.)$ is so called "*modified Bessel function of the third kind*".

The fully conditional posterior of each s_j is as follows:

$$\begin{aligned}\pi(s_j|\beta_j, \eta^2) &\propto \pi(\beta_j|s_j)\pi(s_j|\eta^2)\\ &\propto \frac{1}{\sqrt{2\pi s_j}}\exp\left(-\frac{\beta_j^2}{2s_j}\right)\exp\left(-\frac{\eta^2}{8}s_j\right)\\ &\propto \frac{1}{\sqrt{s_j}}\exp\left\{-\frac{1}{2}\left[\beta_j^2 s_j^{-1} + \frac{\eta^2}{4}s_j\right]\right\}\end{aligned}$$

Hence, the full conditional $\pi(s_j|\beta_j, \eta^2)$ is a generalized inverse Gaussian distribution. GIG $(\nu, \varrho_1, \varrho_2)$, where $\nu = 0.5$, $\varrho_1^2 = \beta_j^2$ and $\varrho_2^2 = \eta^2/4$.

The fully conditional posterior of u_j is as follows,

$$\begin{aligned}\pi(u_j|\beta_j, \eta^2) &\propto \pi(\beta_j|\eta^2, u_j)\pi(u_j)\\ &\propto \text{Exponential}\left(\frac{1}{2}\right)I\{u_j > \sqrt{\eta^2}|\beta_j|\}.\end{aligned}$$

The fully conditional posterior of η^2 is as follows:

$$\begin{aligned}\pi(\eta^2|\boldsymbol{u}, \boldsymbol{s}, \boldsymbol{\beta}) &\propto \pi(\boldsymbol{\beta}|\boldsymbol{u}, \boldsymbol{s}, \eta^2)\pi(\boldsymbol{s}|\eta^2)\pi(\eta^2)\\ &\propto \text{Gamma}\left(k + c, \frac{1}{8}\sum_{j=1}^{k} s_j + d\right)I\left\{\eta^2 < \text{Min}_j\left(\frac{u_j^2}{\beta_j^2}\right)\right\}.\end{aligned}$$

Appendix C. Conditional densities for the adaptive lasso QReg

The full conditional distribution of each β_j, denoted by $\pi\left(\beta_j|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}_{-j}, \sigma, \boldsymbol{v}, s_j, u_j, \eta_j^2\right)$, is proportional to $\pi(\boldsymbol{y}|\boldsymbol{\beta}, \sigma, \boldsymbol{v})\pi\left(\beta_j|s_j, u_j, \eta_j^2\right)$, where $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)$ and $\boldsymbol{\beta}_{-j}$ is the vector $\boldsymbol{\beta}$ excluding the element β_j. Thus, we have

$$\begin{aligned}\pi\left(\beta_j|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}_{-j}, \sigma, \boldsymbol{v}, s_j, u_j, \eta_j^2\right) &\propto \pi(\boldsymbol{y}|\boldsymbol{\beta}, \sigma, \boldsymbol{v})\pi\left(\beta_j|s_j, u_j, \eta_j^2\right)\\ &\propto \exp\left\{-\frac{\sigma}{2}\sum_{i=1}^{n}\frac{(y_i - \boldsymbol{x}'_i\boldsymbol{\beta} - \delta_1 v_i)^2}{\delta_2^2 v_i}\right\}\exp\left\{-\frac{\beta_j^2}{2s_j}\right\}I\left\{|\beta_j| < \frac{u_j}{\sqrt{\eta_j^2}}\right\},\\ &\propto \exp\left\{-\frac{1}{2}\left[\left(\sum_{i=1}^{n}\frac{\sigma x_{ij}^2}{\delta_2^2 v_i} + \frac{1}{s_j}\right)\beta_j^2 - 2\sum_{i=1}^{n}\frac{\sigma\tilde{y}_i x_{ij}}{\delta_2^2 v_i}\beta_j\right]\right\}I\left\{|\beta_j| < \frac{u_j}{\sqrt{\eta_j^2}}\right\},\end{aligned}$$

where $\boldsymbol{x}_i = (x_{i1}, \cdots, x_{ik})'$ and $\tilde{y}_i = y_i - \sum_{h=1,h\neq j}^{k} x_{ih}\beta_h - \delta_1 v_i$. Then the full conditional distribution for β_j is a truncated normal with mean μ_j and variance $\tau_{\beta_j}^2$, where

$$\tau_{\beta_j}^{-2} = \sum_{i=1}^{n}\frac{\sigma x_{ij}^2}{\delta_2^2 v_i} + \frac{1}{s_j}, \quad \text{and} \quad \mu_j = \tau_{\beta_j}^2\sum_{i=1}^{n}\frac{\sigma\tilde{y}_i x_{ij}}{\delta_2^2 v_i}.$$

The fully conditional posterior of each s_j is as follows:

$$\pi\left(s_j|\beta_j, \eta_j^2\right) \propto \pi(\beta_j|s_j)\pi\left(s_j|\eta_j^2\right)$$
$$\propto \frac{1}{\sqrt{2\pi s_j}}\exp\left(-\frac{\beta_j^2}{2s_j}\right)\exp\left(-\frac{\eta_j^2}{8}s_j\right)$$
$$\propto \frac{1}{\sqrt{s_j}}\exp\left\{-\frac{1}{2}\left[\beta_j^2 s_j^{-1} + \frac{\eta_j^2}{4}s_j\right]\right\}$$

Hence, the full conditional $\pi\left(s_j|\beta_j, \eta_j^2\right)$ is a generalized inverse Gaussian distribution. GIG $(\nu, \varrho_1, \varrho_2)$, where $\nu = 0.5$, $\varrho_1^2 = \beta_j^2$ and $\varrho_2^2 = \eta_j^2/4$.

The fully conditional posterior of u_j is as follows,

$$\pi\left(u_j|\beta_j, \eta_j^2\right) \propto \pi\left(\beta_j|\eta_j^2, u_j\right)\pi(u_j)$$
$$\propto \text{Exponential}\left(\frac{1}{2}\right)I\left\{u_j > \sqrt{\eta_j^2}|\beta_j|\right\}.$$

The fully conditional posterior of η_j^2 is as follows:

$$\pi\left(\eta_j^2|u_j, s_j, \beta_j\right) \propto \pi\left(\beta_j|u_j, s_j, \eta_j^2\right)\pi\left(s_j|\eta_j^2\right)\pi\left(\eta_j^2\right)$$
$$\propto \text{Gamma}\left(1 + c, \frac{1}{8}s_j + d\right)I\left\{\eta_j^2 < \left(\frac{u_j^2}{\beta_j^2}\right)\right\}.$$

The full conditional posterior distributions of the latent variable v_i and the scale parameter σ are again the same as in the lasso QReg case.

Appendix D. Conditional densities for the elastic net QReg

The full conditional distribution of each β_j, denoted by $\pi(\beta_j|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}_{-j}, \sigma, \boldsymbol{v}, s_j, u_j, t_j, \widetilde{\eta_1}, \eta_2)$, is proportional to $\pi(\boldsymbol{y}|\boldsymbol{\beta}, \sigma, \boldsymbol{v})\pi(\beta_j|u_j, t_j, \widetilde{\eta_1}, \eta_2)$. Thus, we have

$$\pi(\beta_j|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}_{-j}, \sigma, \boldsymbol{v}, u_j, t_j, \widetilde{\eta_1}, \eta_2) \propto (\boldsymbol{y}|\boldsymbol{\beta}, \sigma, \boldsymbol{v})\pi(\beta_j|u_j, t_j, \widetilde{\eta_1}, \eta_2)$$
$$\propto \exp\left(-\frac{\sigma}{2}\sum_{i=1}^{n}\frac{(y_i - \boldsymbol{x}_i'\boldsymbol{\beta} - \delta_1 v_i)^2}{\delta_2^2 v_i}\right)\exp\left(-\frac{1}{2}\left[\frac{t_j - 1}{2\eta_2 t_j}\right]^{-1}\beta_j^2\right)I\left\{|\beta_j| < \frac{u_j}{\sqrt{4\widetilde{\eta_1}\eta^2}}\right\},$$
$$\propto \exp\left(-\frac{1}{2}\left[\left(\sum_{i=1}^{n}\frac{\sigma x_{ij}^2}{\delta_2^2 v_i} + \left[\frac{t_j - 1}{2\eta_2 t_j}\right]^{-1}\right)\beta_j^2 - 2\sum_{i=1}^{n}\frac{\sigma\tilde{y}_i x_{ij}}{\delta_2^2 v_i}\beta_j\right]\right)I\left\{|\beta_j| < \frac{u_j}{\sqrt{4\widetilde{\eta_1}\eta^2}}\right\},$$

Thus, the full conditional distribution for β_j is a truncated normal with mean μ_j and variance $\tau_{\beta_j}^2$, where

$$\tau_{\beta_j}^{-2} = \sum_{i=1}^{n}\frac{\sigma x_{ij}^2}{\delta_2^2 v_i} + \left[\frac{t_j - 1}{2\eta_2 t_j}\right]^{-1}, \quad \text{and} \quad \mu_j = \tau_{\beta_j}^2\sum_{i=1}^{n}\frac{\sigma\tilde{y}_i x_{ij}}{\delta_2^2 v_i}.$$

The full conditional distribution of each $t_j - 1$, denoted by $\pi(t_j - 1|\beta_j, u_j, \tilde{\eta}_1, \eta_2)$, is given by

$$\pi(t_j - 1|\beta_j, u_j, \tilde{\eta}_1, \eta_2) \propto \frac{1}{\sqrt{t_j - 1}}\exp\left(-\frac{1}{2}\left[\frac{t_j - 1}{2\eta_2 t_j}\right]^{-1}\beta_j^2\right)\exp\left(-\frac{\tilde{\eta}_1 t_j}{4}\right)I\{t_j > 1\}$$

$$\propto \frac{1}{\sqrt{t_j - 1}}\exp\left(-\frac{1}{2}\left[\frac{\tilde{\eta}_1(t_j - 1)}{2} + \frac{2\eta_2\beta_j^2}{t_j - 1}\right]\right)I\{t_j > 1\}$$

Thus, the full conditional distribution of $t_j - 1$ is a generalized inverse Gaussian distribution GIG $(\nu, \varrho_1, \varrho_2)$, where $\nu = 0.5$, $\varrho_1^2 = 2\eta_2\beta_j^2$ and $\varrho_2^2 = \tilde{\eta}_1/2$.

The full conditional distribution of $\tilde{\eta}_1$, denoted by $\pi(\tilde{\eta}_1|t_j)$, is given by

$$\pi(\tilde{\eta}_1|t_j) \propto \tilde{\eta}_1^{a_1 - 1}\exp(-b_1\tilde{\eta}_1)\prod_{j=1}^{k}\Gamma^{-1}(1/2, \tilde{\eta}_1)\tilde{\eta}_1^{1/2}\exp\left\{-\frac{\tilde{\eta}_1 t_j}{4}\right\}$$

$$\propto \Gamma^{-1}(1/2, \tilde{\eta}_1)\tilde{\eta}_1^{k/2 + a_1 - 1}\exp\left\{-\tilde{\eta}_1\left[b_1 + \sum_{j=1}^{k}t_j\right]\right\}$$

The full conditional distribution of η_2, denoted by $\pi(\eta_2|t_j, \beta_j, u_j, \eta_1)$, is given by

$$\pi(\eta_2|t_j, \beta_j, u_j, \eta_1) \propto \eta_2^{a_2 - 1}\exp\left(-b_2\eta_2\right)\prod_{j=1}^{k}\eta_2^{1/2}\exp\left(-\frac{1}{2}\left[\frac{t_j - 1}{2\eta_2 t_j}\right]^{-1}\beta_j^2\right)I\left\{|\beta_j| < \frac{u_j}{\sqrt{\eta_1^2}}\right\},$$

$$\propto \eta_2^{k/2+a_2-1} \exp\left(-\eta_2\left[b_2 + \sum_{j=1}^{k} t_j(t_j-1)^{-1} \beta_j^2\right]\right) I\left\{|\beta_j| < \frac{u_j}{\sqrt{\eta_1^2}}\right\}.$$

Thus, the full conditional distribution of η_2 is a truncated Gamma distribution.

Appendix E

Table E.9
MMSEs and test errors based on 10-fold cross-validation results for the Boston Housing data.

p	Methods	MMSE (SD)	Test Error (SD)
0.50	NBLQ	0.0347 (0.0158)	**0.0659** (0.0125)
	NBALQ	0.0353 (0.0153)	0.0661 (0.0126)
	NBENQ	**0.0345** (0.0152)	0.0662 (0.0125)
	OBLQ	0.0354 (0.0153)	0.0662 (0.0126)
	OBENQ	0.0355 (0.0162)	0.0663 (0.0134)
	lassoQ	0.0359 (0.0152)	0.0676 (0.0134)
	RQ	0.0357 (0.0153)	0.0662 (0.0132)
	lasso	0.0355 (0.0162)	0.0682 (0.0132)
	EN	0.0389 (0.0168)	0.0679 (0.0133)
0.75	NBLQ	0.0352 (0.0158)	**0.0658** (0.0130)
	NBALQ	0.0345 (0.0156)	0.0662 (0.0129)
	NBENQ	**0.0344** (0.0163)	0.0662 (0.0136)
	OBLQ	0.0355 (0.0162)	0.0663 (0.0131)
	OBENQ	0.0353 (0.0164)	0.0661 (0.0135)
	lassoQ	0.0369 (0.0152)	0.0662 (0.0132)
	RQ	0.0357 (0.0154)	0.0702 (0.0101)
	lasso	0.0355 (0.0162)	0.0682 (0.0135)
	EN	0.0415 (0.0130)	0.0677 (0.0133)
0.95	NBLQ	0.0352 (0.0162)	0.0660 (0.0131)
	NBALQ	0.0351 (0.0158)	0.0659 (0.0131)
	NBENQ	**0.0345** (0.0163)	**0.0657** (0.0130)
	OBLQ	0.0352 (0.0166)	0.0664 (0.0136)
	OBENQ	0.0356 (0.0158)	0.0662 (0.0132)
	lassoQ	0.0362 (0.0155)	0.0909 (0.0104)
	RQ	0.0357 (0.0151)	0.0662 (0.0132)
	lasso	0.0355 (0.0162)	0.0681 (0.0136)
	EN	0.0582 (0.0162)	0.0678 (0.0133)

Table E.10
MMSEs and test errors based on 5-fold cross-validation results for the prostate cancer data.

p	Methods	MMSE (SD)	Test Error (SD)
0.50	NBLQ	0.4746 (0.1558)	0.2913 (0.0538)
	NBALQ	0.4711 (0.1560)	0.2765 (0.0534)
	NBENQ	0.4755 (0.1545)	**0.2725** (0.0533)
	OBLQ	0.5327 (0.1575)	0.2772 (0.0540)
	OBENQ	0.4931 (0.1666)	0.2918 (0.0604)
	lassoQ	0.6006 (0.2040)	0.3159 (0.0647)
	RQ	**0.4593** (0.1479)	0.2778 (0.0522)
	lasso	0.5793 (0.1819)	0.2928 (0.0600)
	EN	0.5732 (0.1963)	0.3065 (0.0661)
0.75	NBLQ	0.4755 (0.1433)	0.2706 (0.0501)
	NBALQ	0.4787 (0.1539)	**0.2701** (0.0508)
	NBENQ	**0.4428** (0.1593)	0.2704 (0.0499)
	OBLQ	0.4831 (0.1716)	0.2749 (0.0501)
	OBENQ	0.4779 (0.1525)	0.2835 (0.0504)
	lassoQ	0.6110 (0.1999)	0.3246 (0.0721)
	RQ	0.4593 (0.1479)	0.2778 (0.0522)
	lasso	0.5756 (0.1984)	0.2928 (0.0714)
	EN	0.5756 (0.1965)	0.3065 (0.0655)
0.95	NBLQ	0.4779 (0.1528)	0.2703 (0.0455)
	NBALQ	**0.4450** (0.1396)	**0.2701** (0.0476)
	NBENQ	0.4498 (0.1502)	0.2732 (0.0498)
	OBLQ	0.5284 (0.1555)	0.2872 (0.0505)
	OBENQ	0.5081 (0.1651)	0.2856 (0.0539)
	lassoQ	0.6596 (0.2202)	0.3210 (0.0722)
	RQ	0.4593 (0.1479)	0.2778 (0.0522)
	lasso	0.5756 (0.1880)	0.3010 (0.0604)
	EN	0.5829 (0.1951)	0.3003 (0.0664)

References

[1] R. Koenker, G. Bassett, Regression quantiles. *Econometrica*, J. Econom. Soc. (1978) 33–50.
[2] R. Koenker, Quantile Regression, Cambridge University press, New York, 2005.
[3] D.L. Donoho, High-dimensional data analysis: the curses and blessings of dimensionality, AMS Math. Chall. Lect. 1 (32) (2000) 375.
[4] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B (1996) 267–288.
[5] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. R. Stat. Soc. Ser. B 67 (2) (2005) 301–320.
[6] M. Yuan, Y. Lin, Model selection and estimation in regression with grouped variables, J. R. Stat. Soc. Ser. B 68 (1) (2006) 49–67.
[7] H. Zou, The adaptive lasso and its oracle properties, J. Am. Stat. Assoc. 101 (476) (2006) 1418–1429.
[8] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, K. Knight, Sparsity and smoothness via the fused lasso, J. R. Stat. Soc. Ser. B 67 (1) (2005) 91–108.
[9] E. Candes, T. Tao, The dantzig selector: statistical estimation when p is much larger than n, Ann. Stat. (2007) 2313–2351.
[10] E.J. Candès, T. Tao, The power of convex relaxation: near-optimal matrix completion, IEEE Trans. Inf. Theory 56 (5) (2010) 2053–2080.
[11] R. Mazumder, J.H. Friedman, T. Hastie, Sparsenet: coordinate descent with nonconvex penalties, J. Am. Stat. Assoc. 106 (495) (2011) 1125–1138.
[12] T. Park, G. Casella, The bayesian lasso, J. Am. Stat. Assoc. 103 (482) (2008) 681–686.
[13] Q. Li, N. Lin, The Bayesian elastic net, Bayesian Anal. 5 (1) (2010) 151–170.
[14] R. Alhamzawi, H.T.M. Ali, The Bayesian elastic net regression, Commun. Stat. Simulat. Comput. 47 (4) (2017) 1168–1178.
[15] H. Mallick, N. Yi, A new bayesian lasso, Stat. Interface 7 (4) (2014) 571.
[16] Y. Li, J. Zhu, L_1-norm quantile regression, J. Comput. Graph. Stat. 17 (1) (2008) 163–185.
[17] Y. Wu, Y. Liu, Variable selection in quantile regression, Stat. Sin. (2009) 801–817.
[18] K. Kato, Group Lasso for High Dimensional Sparse Quantile Regression Models, (2011) 1103.1458.
[19] J. Bradic, J. Fan, W. Wang, Penalized composite quasi-likelihood for ultrahigh dimensional variable selection, J. R. Stat. Soc. Ser. B 73 (3) (2011) 325–349.
[20] X. Jiang, J. Jiang, X. Song, Oracle model selection for nonlinear models based on weighted composite quantile regression, Stat. Sin. (2012) 1479–1506.
[21] G. Ciuperca, Quantile Regression in High-Dimension with Breaking, (2013) 1302.4244.
[22] Q. Li, R. Xi, N. Lin, Bayesian regularized quantile regression, Bayesian Anal. 5 (3) (2010) 533–556.
[23] R. Alhamzawi, K. Yu, D.F. Benoit, Bayesian adaptive lasso quantile regression, Stat. Model. Int. J. 12 (3) (2012) 279–297.
[24] D.F. Andrews, C.L. Mallows, Scale mixtures of normal distributions, J. R. Stat. Soc. Ser. B (1974) 99–102.
[25] R. Alhamzawi, Bayesian elastic net tobit quantile regression, Commun. Stat. Simulat. Comput. 45 (7) (2014) 2409–2427.
[26] W.-h. Zhao, R.-q. Zhang, Y.-z. Lü, J.-c. Liu, Bayesian regularized regression based on composite quantile method, Acta Math. Appl. Sin. Engl. Ser. 32 (2) (2016) 495–512.
[27] R.W. Koenker, V. D'Orey, Algorithm as 229: computing regression quantiles, J. R. Stat. Soc. Ser. C Appl. Stat. 36 (3) (1987) 383–393.
[28] R. Koenker, J.A. Machado, Goodness of fit and related inference processes for quantile regression, J. Am. Stat. Assoc. 94 (448) (1999) 1296–1310.
[29] K. Yu, R.A. Moyeed, Bayesian quantile regression, Stat. Probab. Lett. 54 (4) (2001) 437–447.
[30] K. Yu, J. Zhang, A three-parameter asymmetric laplace distribution and its extension, Commun. Stat. Theor. Methods 34 (9–10) (2005) 1867–1879.
[31] M. Geraci, M. Bottai, Quantile regression for longitudinal data using the asymmetric laplace distribution, Biostatistics 8 (1) (2006) 140–154.
[32] H. Kozumi, G. Kobayashi, Gibbs sampling methods for Bayesian quantile regression, J. Stat. Comput. Simul. 81 (11) (2011) 1565–1578.
[33] S.P. Brooks, A. Gelman, General methods for monitoring convergence of iterative simulations, J. Comput. Graph. Stat. 7 (4) (1998) 434–455.
[34] M. Kocherginsky, X. He, Y. Mu, Practical confidence intervals for regression quantiles, J. Comput. Graph. Stat. 14 (1) (2005) 41–55.
[35] D. Harrison, D.L. Rubinfeld, Hedonic housing prices and the demand for clean air, J. Environ. Econ. Manag. 5 (1) (1978) 81–102.
[36] T.A. Stamey, J.N. Kabalin, J.E. McNeal, I.M. Johnstone, F. Freiha, E.A. Redwine, N. Yang, Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate. ii. radical prostatectomy treated patients, J. Urol. 141 (5) (1989) 1076–1083.
[37] F. Leisch, E. Dimitriadou, Machine Learning Benchmark Problems. R Package Version 2, (2010) 1-1.
[38] D. Benoit, R. Al-Hamzawi, K. Yu, D. Van den Poel, bayesQR: Bayesian Quantile Regression. R Package Version 2, (2011) 1-1.
[39] B. Jorgensen, Statistical Properties of the Generalized Inverse Gaussian Distribution vol. 9, Springer Science & Business Media, 2012.
[40] W. Gilk, Derivative-free Adaptive Rejection Sampling for Gibbs Sampling, (1992) Bayesian statistics 4.