UNIVERSITEIT GENT

FACULTEIT ECONOMIE EN BEDRIJFSKUNDE

TWEEKERKENSTRAAT 2
B-9000 GENT
Tel. : 32 - (0)9 – 264.34.61
Fax. : 32 - (0)9 – 264.35.92

WORKING PAPER

Bayesian adaptive Lasso quantile regression

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

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

Dries F. Benoit, Department of Marketing, Faculty of Economics and Business Administration, Ghent University, Ghent, Belgium

July 2011

2011/728

D/2011/7012/33

Bayesian adaptive Lasso quantile regression

Rahim Alhamzawi[†*], Keming Yu[†⋆], Dries F. Benoit [‡]

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

⋆ *Business School, Shihezi University, China*

‡ *Department of Marketing, Faculty of Economics and Business Administration, Ghent University, Ghent, Belgium*

Abstract

Recently, variable selection by penalized likelihood has attracted much research interest. In this paper, we propose adaptive Lasso quantile regression (BALQR) from a Bayesian perspective. The method extends the Bayesian Lasso quantile regression by allowing different penalization parameters for different regression coefficients. Inverse gamma prior distributions are placed on the penalty parameters. We treat the hyperparameters of the inverse gamma prior as unknowns and estimate them along with the other parameters. A Gibbs sampler is developed to simulate the parameters from the posterior distributions. Through simulation studies and analysis of a prostate cancer data set, we compare the performance of the BALQR method proposed with six existing Bayesian and non-Bayesian methods. The simulation studies and the prostate cancer data analysis indicate that the BALQR method performs well in comparision to the other approaches.

Keywords: Gibbs sampler, Lasso, Quantile regression, Skewed Laplace distribution.

1. Introduction

Variable selection plays an important role in building a multiple regression model. In particular, the selection process provides a good tool for estimating the parameters, a good

*Corresponding author.
Email address: rahim.al-hamzawi@brunel.ac.uk (Rahim Alhamzawi[†])

prediction as well as identification of significant variables (Griffin and Brown, 2010). However, classical variable selection methods are often highly time consuming and maybe suffer from instability (Breiman, 1996). Bayesian methods for subset selection implemented using stochastic search variable selection (SSVS) algorithms have become widely used in linear regression, generalized linear models and other modeling frameworks (Mitchell and Beauchamp, 1988; George and McCulloch, 1997; Fahrmeir, Kneib and Konrath, 2010). However, SSVS is computationally very demanding when the number of variables is greater than 10,000 (Griffin and Brown, 2010). Variable selection by penalized likelihood has attracted much interest recently; see for example, Lasso (Tibshirani, 1996), SCAD (Fan and Li, 2001), LARS (Efron, Hastie, Johnstone and Tibshirani 2004) and adaptive Lasso (Zou, 2006) who extended the Lasso approach proposed by (Tibshirani, 1996) allowing different penalization parameters for different regression coefficients. Similarly, from a Bayesian point of view, Lasso-based models were proposed by Park and Casella (2008), Hans (2009); Bayesian adaptive Lasso, iterative adaptive Lasso (Sun, Ibrahim and Zou, 2010); and global-local shrinkage approach (Polsen and Scott, 2011).

Quantile regression models are rapidly gaining popularity, particularly in econometrics, social sciences, medicine and public health. A comprehensive account of these recent developments can be found in Koenker (2005). Like standard mean regression models, dealing with parameter and model uncertainty as well as the updating of information is of great importance for quantile regression and its applications.

Koenker (2004) developed an l_1-regularization quantile regression method to shrink individual effects towards a common value. Additionally, Wang, Li and Jiang (2007) proposed the LAD-Lasso method which combines the idea of least absolute deviance (LAD) and Lasso for robust regression shrinkage and selection. Li and Zhu (2008) developed the solution path of the l_1 penalized quantile regression and Wu and Liu (2009) studied penalized quantile regression with the SCAD and the adaptive-Lasso penalties. Recently, Li, Xi and Lin (2010)

proposed Bayesian regularization quantile regression approaches including Lasso (BQR.L), group Lasso and elastic net penalties (BQR.N).

With regards to Lasso regression, Zou (2006) proved that the adaptive Lasso regression enjoys oracle properties reported in Fan and Li (2001) and Fan and Peng (2004) that Lasso does not have, i.e., adaptive Lasso chooses the true model of nonzero coefficients with probability tending to one. Huang et al. (2008) showed that under a partial orthogonality condition in which the predictors with zero coefficients (unimportant predictors) are weakly correlated with the predictors with nonzero coefficients (important predictors), adaptive Lasso has the oracle property even when there are far more predictors than the sample size. However, in many real world applications, the unimportant predictors are often highly correlated with some important predictors (Sun, Ibrahim and Zou, 2010). In this paper, we focus on one such application: prostate cancer data analysis. Prostate specific antigen (PSA) is a protein produced by cells and has been routinely used as a biomarker for screening prostate cancer. Nowadays, significant effort is made in finding candidate predictors that relate to prostate cancer. Certain correlation is present between the predictors in the prostate cancer data set. For example, the correlation coefficient is 0.752 between Gleason score and percentage Gleason scores 4 or 5, 0.675 between cancer volume and capsular penetration, 0.673 between seminal vesicle invasion and capsular penetration, and so on. The correlation between the predictors is an argument to use the adaptive Lasso because the procedure deals with correlated predictors by using adaptive weights for the different predictors. It could be expected that the conditional mean function is inaccurate in representing the relationship between the predictors and the level of prostate specific antigen.

Therefore, we propose Bayesian adaptive Lasso quantile regression (BALQR). In particular, we extend the Bayesian Lasso quantile regression reported in Li, Xi and Lin (2010) by allowing different penalization parameters for different regression coefficients. Inverse gamma prior distributions are placed on the penalty parameters. Similar to Yi and Xu (2008) and

Sun et al. (2010), we treat the hyperparameters of the inverse gamma prior as unknowns and estimate them along with the other parameters. We present a Gibbs sampler for the BALQR that is based on a theoretic derivation of the skewed Laplace distribution as a scale mixture of normal distributions. Using both simulation studies and real data we compare the performance of the BALQR method with six existing Bayesian and non-Bayesian methods. These methods encompass Bayesian regularized quantile regression with the Lasso penalty (BQR.L) and the elastic net penalty (BQR.N). Also, non-Bayesian methods including the Lasso (lasso), the elastic net (EN), the standard quantile regression (QR) and regularized quantile regression with Lasso penalty (QR-L) are used. Both our simulation studies and data analysis show that BALQR performs well and this method may be preferred over most existing methods it is compared against.

The rest of the paper is organized as follows. In Section 2, we present Bayesian quantile regression with adaptive Lasso penalty as well as an outline of the Gibbs sampler estimation procedure. In Section 3, we carry out simulation studies to examine the performance of the method proposed and in Section 4, we illustrate the performance of our method via analysis of the prostate cancer data set. We conclude with a brief conclusions in Section 5.

2. Methods

2.1. Quantile Regression

The simple linear quantile regression model is given by,

$$y_i = \beta_0 + \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i, \tag{1}$$

where $\{(x_i, y_i), i = 1, 2, ..., n\}$ is a sample of independent observations, y_i is the response variable, $\boldsymbol{x}_i' = (x_{i1}, x_{i2}, ..., x_{ik})$ represents the k known predictors, β_0 is the intercept, $\boldsymbol{\beta}$

is a $k \times 1$ vector of regression coefficients and $\varepsilon_i, i = 1, ..., n$ represent error terms. The distribution of the error is assumed as unknown, and for $0 < p < 1$, it is restricted to have the pth quantile equal to zero. The pth quantile regression model takes the form

$$Q_{y_i}(p|\boldsymbol{x}_i) = \beta_0 + \boldsymbol{x}_i'\boldsymbol{\beta}, \tag{2}$$

where $Q_{y_i}(p|\boldsymbol{x}_i)$ is the inverse cumulative distribution function of y_i given x_i evaluated at p. The regression coefficients can be estimated consistently as the solution to the following minimization problem

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_p(y_i - \beta_0 - \boldsymbol{x}_i'\boldsymbol{\beta}), \quad \text{where} \quad \rho_p(u) = u\{p - I(u < 0)\}. \tag{3}$$

A possible parametric link between the minimization problem in (3) and maximum likelihood theory is given by the skewed Laplace distribution; see Koenker and Machado (1999) and Yu and Moyeed (2001). The density function of a skewed Laplace distribution is given by

$$f(y|\mu, \sigma, p) = \sigma p(1-p) \exp\{-\sigma\rho_p(y - \mu)\}, \tag{4}$$

where μ is the location parameter and σ is the scale parameter. The minimization problem given by (3) is equivalent to maximizing the likelihood function of y_i by assuming y_i's are random variables from a skewed Laplace distribution with $\mu = \beta_0 + \boldsymbol{x}_i'\boldsymbol{\beta}$ and $\sigma = 1$. From a Bayesian framework, the skewed Laplace distribution link has been exploited by a number of authors. In particular, Yu and Moyeed (2001) implemented Bayesian inference for quantile regression, Yu and Stander (2007) developed Bayesian estimation procedure for the Tobit quantile regression, while Benoit and Van den Poel (2011) propose a methodology for binary quantile regression. A new Gibbs sampler for Bayesian analysis of quantile regression model

based on a theoretic derivation of skewed Laplace distribution was introduced by Reed and Yu (2009), Reed. Dunson and Yu (2009) developed Bayesian methods for variable selection with a simple and efficient stochastic search variable selection (SSVS) algorithm proposed for posterior computation and Alhamzawi et al. (2011) introduced methods for conducting Bayesian quantile analysis of an allometric model that includes random effects. Furthermore, another attractive property of the skewed Laplace distribution is that it can be represented as a scale mixture of normal distributions (Tsionas, 2003; Kozumi and Kobayashi (2009); Rue and Held, 2005; Reed and Yu, 2009):

$$W =^{d} \theta z + \phi\xi\sqrt{\sigma^{-1}z},$$

where

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

The random variables z and ξ are mutually independent and follow an exponential distribution with mean (σ^{-1}) and a standard normal distribution, respectively. This mixture representation allow us to express a quantile regression model as a normal regression model. In addition, it provides an easy way to construct a Gibbs sampler as well as saving time in sampling the regression coefficients. Recently, this property appeared in papers by Li, Xi and Lin (2010), Yue and Rue (2010) and Alhamzawi and Yu (2011) to conduct Bayesian quantile regression via Gibbs sampler.

2.2. Quantile regression with Lasso penalty

Lasso quantile regression (Li and Zhu, 2008) is a regularization technique for simultaneous estimation and variable selection. The Lasso quantile regression (Li and Zhu, 2008)

estimates are defined as

$$\min_{\beta_0,\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_p(y_i - \beta_0 - \boldsymbol{x}_i'\boldsymbol{\beta}) + \lambda\|\boldsymbol{\beta}\|_1, \tag{5}$$

where λ is a nonnegative regularization parameter. The second term in (5) is the so-called l_1 penalty quantile regression that is crucial for the success of the Lasso method. As λ increases the Lasso continuously shrinks quantile regression coefficients towards zero.

Li, Xi and Lin (2010) employ a Laplace prior $p(\beta_j|\sigma,\lambda) = \sigma\lambda/2\exp\{-\sigma\lambda|\beta_j|\}$ on $\beta_j,\ \beta_j \in \boldsymbol{\beta}$ and assumed that the residuals ε_i come from the skewed Laplace distribution (4). Specifically, Laplace prior distributions are placed on the k regression coefficients.

In this paper, we extend this idea to Bayesian adaptive Lasso quantile regression (BALQR). We put different penalization parameters on the different regression coefficients. Thus, we propose a Laplace prior on β_j taking the form

$$p(\beta_j|\sigma,\lambda_j) = \frac{\sigma^{1/2}}{2\lambda_j}\exp\{-\frac{\sigma^{1/2}|\beta_j|}{\lambda_j}\}, \tag{6}$$

which can be represented as a scale mixture of normals with an exponential mixing density (Andrews and Mallows, 1974)

$$\frac{\nu}{2}\exp\{-\nu|t|\} = \int_0^\infty \frac{1}{\sqrt{2\pi s}}\exp\{-t^2/2s\}\frac{\nu^2}{2}\exp\{-\nu^2 s/2\}ds, \quad \nu > 0. \tag{7}$$

Let $\nu_j = \sigma^{1/2}/\lambda_j$. Then, the proposed prior can be written as

$$p(\beta_j|\sigma,\lambda_j) \quad = \frac{\nu_j}{2}\exp\{-\nu_j|\beta_j|\}$$

$$= \int_0^\infty \frac{1}{\sqrt{2\pi s_j}} \exp\{-\beta_j^2/2s_j\} \frac{\nu_j^2}{2} \exp\{-\nu_j^2 s_j/2\} ds_j.$$

Then, we have

$$p(\beta_j|\sigma, \lambda_j^2) = \int_0^\infty \frac{1}{\sqrt{2\pi s_j}} \exp\{-\beta_j^2/2s_j\} \frac{\sigma}{2\lambda_j^2} \exp\{-\sigma s_j/2\lambda_j^2\} ds_j \tag{8}$$

This motivates us to consider the class of inverse gamma priors on λ_j^2 (not λ_j) of the form

$$p(\lambda_j^2|\delta, \tau) = \frac{\tau^\delta}{\Gamma(\delta)} (\lambda_j^2)^{-1-\delta} \exp\{-\frac{\tau}{\lambda_j^2}\}, \tag{9}$$

where $\delta > 0$ and $\tau > 0$ are two hyperparameters. The posterior density function of λ_j^2, combining the prior (9) with (8), is inverse gamma with shape parameter $1 + \delta$ and rate parameter $\sigma s_j/2 + \tau$. The amount of shrinkage in the prior (9) depends on the values of the hyperparameters τ and δ (Yi and Xu, 2008). Because smaller τ and larger δ lead to bigger penalization, it is important to treat τ and δ as unknown parameters to avoid enforcing specific values that affect the estimates of the regression coefficients (Yi and Xu, 2008 and Sun et al., 2010). This procedure is quite different from Bayesian Lasso quantile regression reported in Li, Xi and Lin (2010). Bayesian adaptive Lasso quantile regression uses a Laplace prior for β_j such that each β_j has a Lasso-type of penalization parameter $\sigma^{1/2}/\lambda_j$, as in the adaptive Lasso. For the moment, the parameters σ and λ_j^2 are considered to be known, however this assumption is relaxed later.

2.3. Bayesian Quantile regression with adaptive Lasso penalty

Bayesian Adaptive Lasso Quantile regression is a Bayesian hierarchical model given by

$$y_i = \beta_0 + \boldsymbol{x}_i'\boldsymbol{\beta} + \theta z_i + \phi\xi_i\sqrt{\sigma^{-1}z_i},$$

$$
\begin{aligned}
&p(\beta_0) \propto 1,\\
&p(\xi_i) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{\xi_i^2}{2}\},\\
&p(z_i|\sigma) = \sigma \exp\{-\sigma z_i\},\\
&p(\beta_j, s_j|\sigma, \lambda_j^2) = \frac{1}{\sqrt{2\pi s_j}} \exp\{-\beta_j^2/2s_j\} \frac{\sigma}{2\lambda_j^2} \exp\{-\sigma s_j/2\lambda_j^2\},\\
&p(\lambda_j^2|\delta, \tau) = \frac{\tau^\delta}{\Gamma(\delta)} (\lambda_j^2)^{-1-\delta} \exp\{-\frac{\tau}{\lambda_j^2}\},\\
&p(\sigma) = \sigma^{a-1} \exp\{-b\sigma\},\\
&p(\tau, \delta) = \tau^{-1}.
\end{aligned}
\tag{10}
$$

The posterior distribution of all parameters is given by

$$
\begin{aligned}
&p(\beta_0, \boldsymbol{\beta}, \mathbf{z}, \mathbf{s}, \sigma, \lambda_1, ..., \lambda_k|\mathbf{y}, \mathbf{X})\\
&\quad\propto p(\mathbf{y}|\beta_0, \boldsymbol{\beta}, \mathbf{z}, \sigma, \mathbf{X}) \prod_{i=1}^{n} p(z_i|\sigma)\\
&\quad\times \prod_{j=1}^{k} p(\beta_j, s_j|\sigma, \lambda_j^2) p(\lambda_j^2|\tau, \delta) p(\sigma) p(\tau, \delta),\\
&\quad\propto \prod_{i=1}^{n} \frac{\sigma}{\sqrt{\sigma^{-1}\phi^2 z_i}} \exp\{-\frac{\sigma(y_i - \beta_0 - x_i'\boldsymbol{\beta} - \theta z_i)^2}{2\phi^2 z_i} - \sigma z_i\}\\
&\quad\times \prod_{j=1}^{k} \frac{1}{\sqrt{2\pi s_j}} \exp\{-\beta_j^2/2s_j\} \frac{\sigma}{2\lambda_j^2} \exp\{-\sigma s_j/2\lambda_j^2\} \frac{\tau^\delta}{\Gamma(\delta)} (\lambda_j^2)^{-1-\delta} \exp\{-\frac{\tau}{\lambda_j^2}\}\\
&\quad\times \tau^{-1} \sigma^{a-1} \exp\{-b\sigma\},
\end{aligned}
\tag{11}
$$

where $\mathbf{y} = (y_1, ..., y_n)$, $\mathbf{X} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)$, $\mathbf{z} = (z_1, ..., z_n)$ and $\mathbf{s} = (s_1, ..., s_k)$. The expression (11) yields a tractable and efficient Gibbs sampler that works as follows:

1- Fix the value of p so that the pth quantile is modelled.

2- Simulate $\beta_0|\cdot \sim N(\bar{\beta}_0, s^2_{\beta_0})$, where $\bar{\beta}_0 = (n^{-1})\sum_{i=1}^n (y_i - \beta_0 - \boldsymbol{x}_i'\boldsymbol{\beta} - \theta z_i)$, and $s^2_{\beta_0} = (\sigma\phi^2/n^2)\sum_{i=1}^n z_i$

3- Simulate $z_i^{-1}|\cdot \sim$ Inverse Gaussian $(\mu', \lambda'),\ i = 1,\ \ldots, n$, where

$$\mu' = \sqrt{\frac{\theta^2 + 2\phi^2}{(y_i - \beta_0 - \boldsymbol{x}_i'\boldsymbol{\beta})^2}} \quad \text{and} \quad \lambda' = \frac{\sigma(\theta^2 + 2\phi^2)}{\phi^2}, \tag{12}$$

in the parameterization of inverse Gaussian density given by

$$f(x|\lambda', \mu') = \sqrt{\frac{\lambda'}{2\pi}} x^{-3/2} \exp\{\frac{-\lambda'(x-\mu')^2}{2(\mu')^2 x}\}, \quad x > 0; \tag{13}$$

see, e.g., Chhikara and Folks (1989).

4- Simulate $\beta_j|\cdot \sim$ normal distribution $N(\bar{\beta}_j, \hat{\sigma}_j^2)$, where

$$\hat{\sigma}_j^2 = (\sigma\phi^{-2}\sum_{i=1}^n x_{ij}^2 z_i^{-1} + s_j^{-1})^{-1}, \quad \text{and} \quad \bar{\beta}_j = \hat{\sigma}_j^2 \sigma\phi^{-2}\sum_{i=1}^n x_{ij} z_i^{-1}(y_i - \beta_0 - \sum_{l\neq j} x_{il}\beta_l - \theta z_i)$$

5- Simulate $s_j|\cdot \sim$Inverse Gaussian $(\mu', \lambda'),\ i = 1,\ \ldots, k$, where

$$\mu' = \sqrt{\frac{\beta_j^2\lambda_j^2}{\sigma}} \quad \text{and} \quad \lambda' = \beta_j^2, \tag{14}$$

6- Simulate $\sigma|\cdot \sim$ Gamma (a_1, a_2), where

$$a_1 = 3n/2 + k + a, \quad \text{and} \quad a_2 = \{\sum_{i=1}^n (\frac{(y_i - \beta_0 - \boldsymbol{x}_i'\boldsymbol{\beta} - \theta z_i)^2}{2\phi^2 z_i} + z_i) + \sum_{j=1}^k \frac{s_j}{2\lambda_j^2} + b\}.$$

7- Simulate $\lambda_j^2|\cdot \sim$ Inverse Gamma $(1+\delta,\ \sigma s_j/2+\tau)$.

8- Simulate $\tau|\cdot \sim$ Gamma $(k\delta,\ \Sigma_{j=1}^{k}\lambda_j^{-2})$.

9- Simulate $\delta|\cdot$: The conditional posterior distribution of δ is

$$p(\delta|\cdot) \propto (\Gamma(\delta))^{-k}\tau^{k\delta}\prod_{j=1}^{k}\lambda_j^{-2\delta}. \tag{15}$$

Although the full conditional posterior distribution of δ does not have a closed form, it is log-concave. The adaptive rejection sampling algorithm (Gilks, 1992) is used to sample from this distribution.

3. Simulation studies

In this section, we compare our method with several Bayesian and non-Bayesian methods. The Bayesian methods include Bayesian regularized quantile regression with Lasso penalty (BQR.L) and elastic net penalty (BQR.N) (Li, Xi and Lin, 2010). The non-Bayesian methods include Lasso (lasso), elastic net (EN), quantile regression with Lasso penalty (QR-L) and standard quantile regression (QR).

3.1. i.i.d. random errors

The simulation setup is similar to the simulation studies 1, 2 and 3 in Li, Xi and Lin (2010) with different parameter values for the error distributions. In addition, we further test the methodology with two alternative error distributions. Specifically, we simulate 20 training observations, 20 validation observations and 200 testing observations from the model $y = x_i'\boldsymbol{\beta} + \varepsilon_i$ where the true value for the $\boldsymbol{\beta}$'s are set as follows:

Simulation 1: $\boldsymbol{\beta} = (3, 1.5, 0, 0, 2, 0, 0, 0)'$,

Simulation 2: $\boldsymbol{\beta} = (0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85)'$,

Simulation 3: $\boldsymbol{\beta} = (5, 0, 0, 0, 0, 0, 0, 0)'$,

Simulation 4: $\boldsymbol{\beta} = (5, 5, 5, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)'$.

Simulation studies 1, 2, and 3 correspond to the sparse, dense and very sparse cases respectively. We have set up Simulation 4 as a sparse recovery problem in which $k = 18$ with most coefficients set to zero except $\beta_j = 5,\ j = 1, 2, 3$. We fit the model using the generated data set and choose the penalty parameter in lasso (λ), EN (λ_1 and λ_2) and QR-L (λ) via an independent validation set. Note that the standard quantile regression (QR) is not a regularization method and the estimate of the penalty parameter is automatically in BALQR, BQR.L and BQR.EN. These methods do not need a validation set. Therefore, we estimate the $\boldsymbol{\beta}$ by using the validation sets as an additional data set. The rows of X follow a multivariate normal distribution $N(0, \Sigma)$ with $(\Sigma)_{ij} = 0.5^{|i-j|}$. In each simulation study, we consider the following six error distributions so that the pth quantile is 0:

1: The distribution of the error is a normal distribution: $N(\mu, 1)$.

2: The distribution of the error is a mixture of two normal distributions: $0.1N(\mu, 1) + 0.9N(\mu, 9)$.

3: The distribution of the error is a Laplace distribution: Laplace$(\mu, 1)$.

4: The distribution of the error is a mixture of two Laplace distributions: $0.1\text{Laplace}(\mu, 1) + 0.9\text{Laplace}(\mu, 3)$.

5: The distribution of the error is a t distribution with three degrees of freedom, $t_{(3)}$.

6: The distribution of the error is χ^2 distribution with three degrees of freedom, $\chi^2_{(3)}$.

For each simulation study and for each $p \in (0.5, 0.75, 0.95)$, we run 150 simulations. In BALQR, the parameters a and b of the Gamma prior for σ are set to be 0.1. Since the true

model is known, we can compute the median of mean absolute deviations (MMAD), that is, median $(1/200\Sigma_{i=1}^{200}(|x_i'\hat{\boldsymbol{\beta}} - x_i'\boldsymbol{\beta}^{true}|))$, where the median is taken over the 150 simulations.

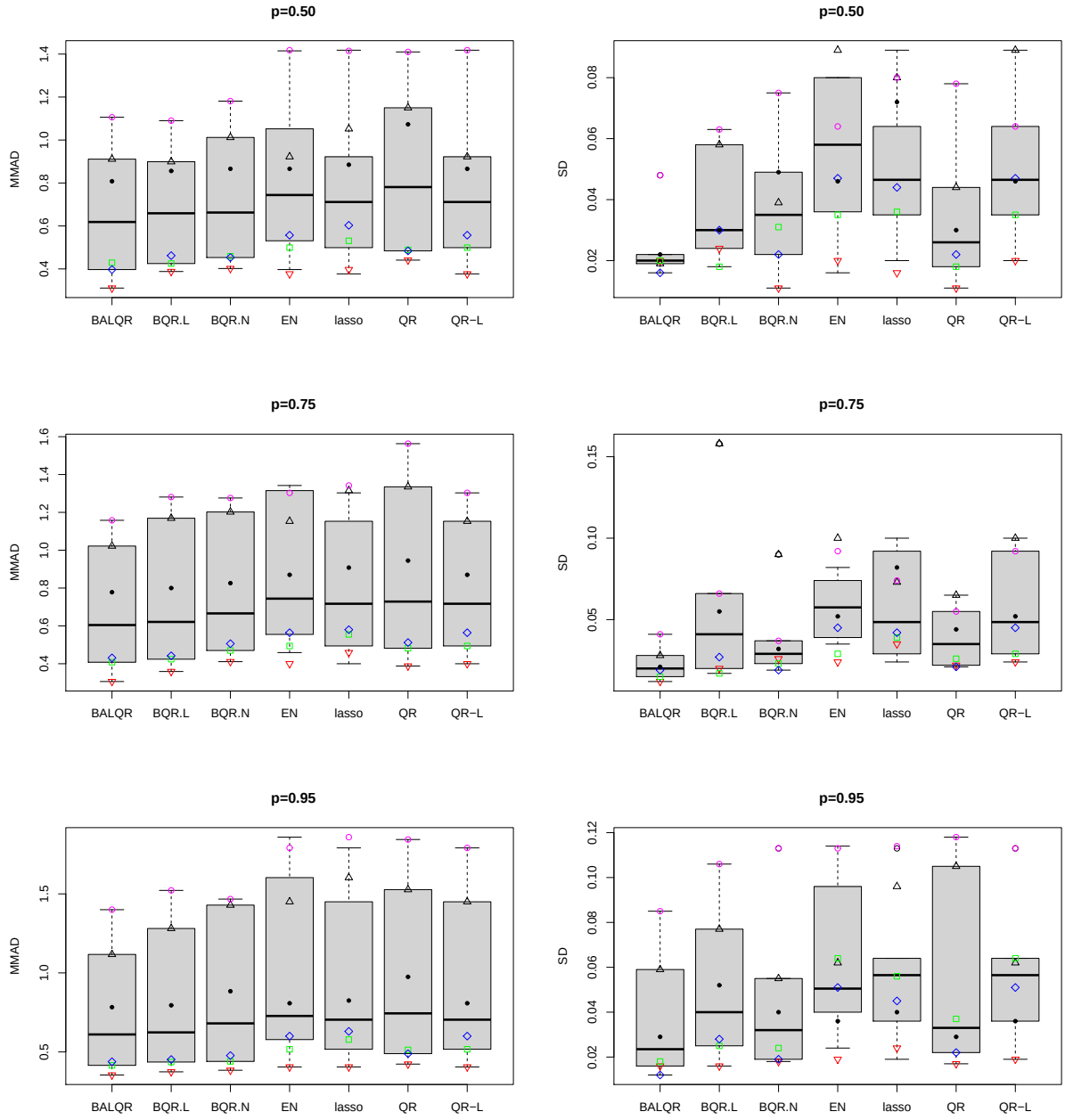

Figure 1: Boxplots summarizing the MMADs and the standard deviations of MMADs (SD) for the seven methods using the six error distributions in Simulation 1. Overlaid are the normal error ($\triangledown$), normal mixture ($\triangle$), Laplace ($\square$), Laplace mixture ($\circ$), t_3 ($\diamond$) and χ_3^2 ($\bullet$).

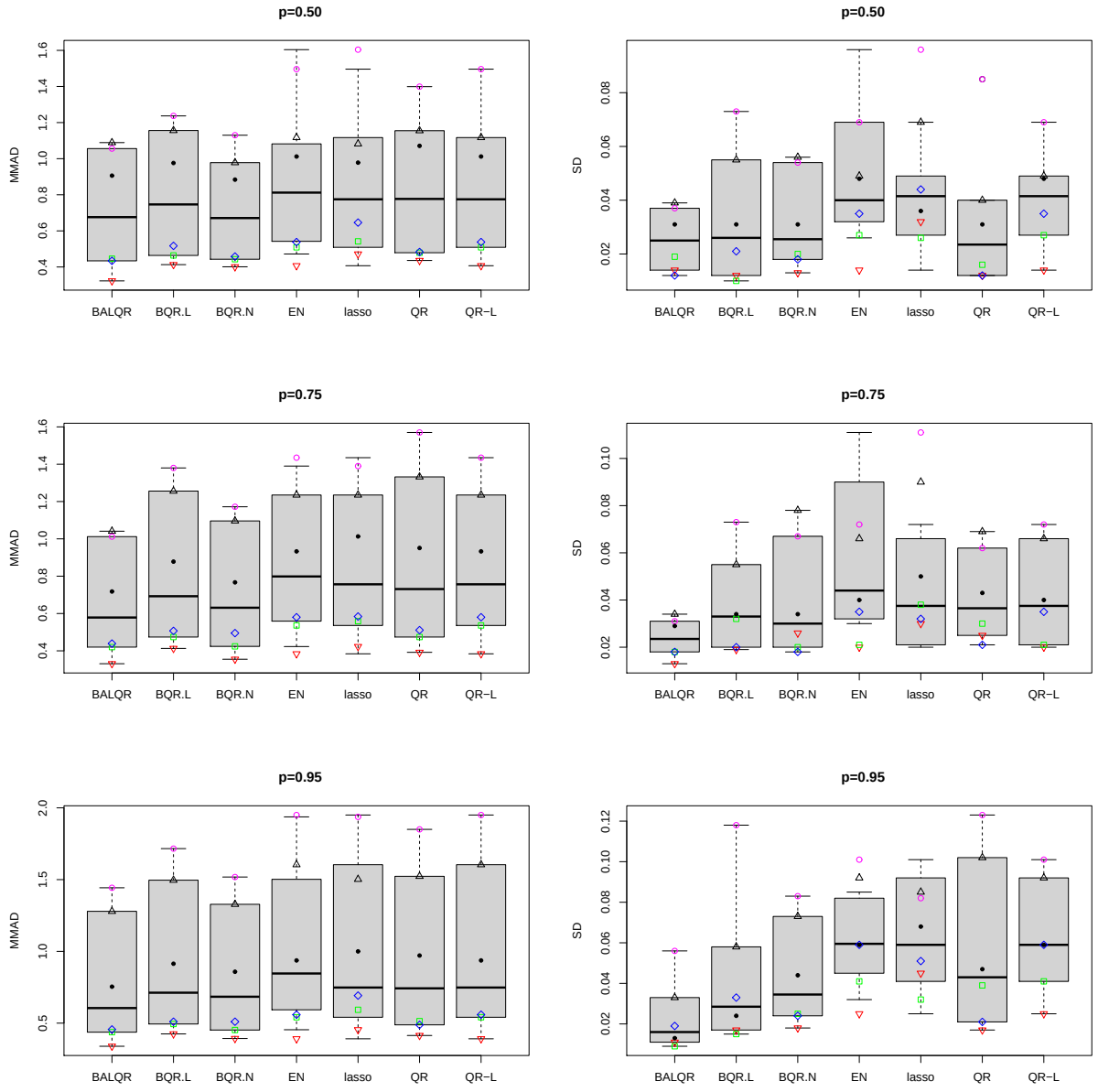

Figure 2: Boxplots summarizing the MMADs and the standard deviations of MMADs (SD) for the seven methods using the six error distributions in Simulation 2. Overlaid are the normal error ($\triangledown$), normal mixture ($\triangle$), Laplace ($\square$), Laplace mixture ($\circ$), t_3 ($\diamond$) and χ^2_3 ($\bullet$).

A number of observations can be done from Figures 1, 2, 3, and 4. For the MMAD and the standard deviations criteria, the proposed method (BALQR) generally performs better than the other six methods for all the distributions under consideration. Most noticeably,

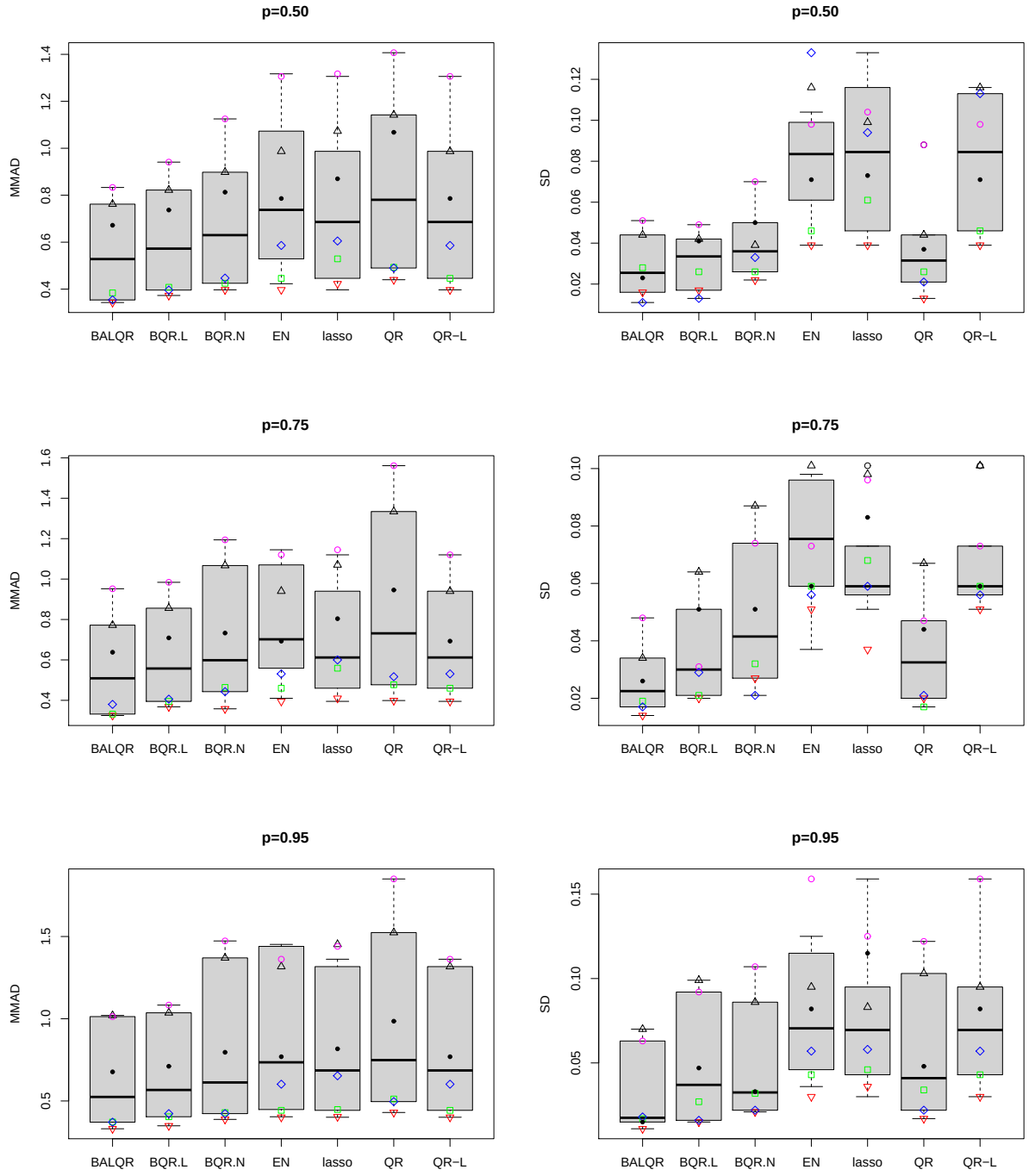

Figure 3: Boxplots summarizing the MMADs and the standard deviations of MMADs (SD) for the seven methods using the six error distributions in Simulation 3. Overlaid are the normal error ($\triangledown$), normal mixture ($\triangle$), Laplace ($\square$), Laplace mixture ($\circ$), t_3 ($\diamond$) and χ^2_3 ($\bullet$).

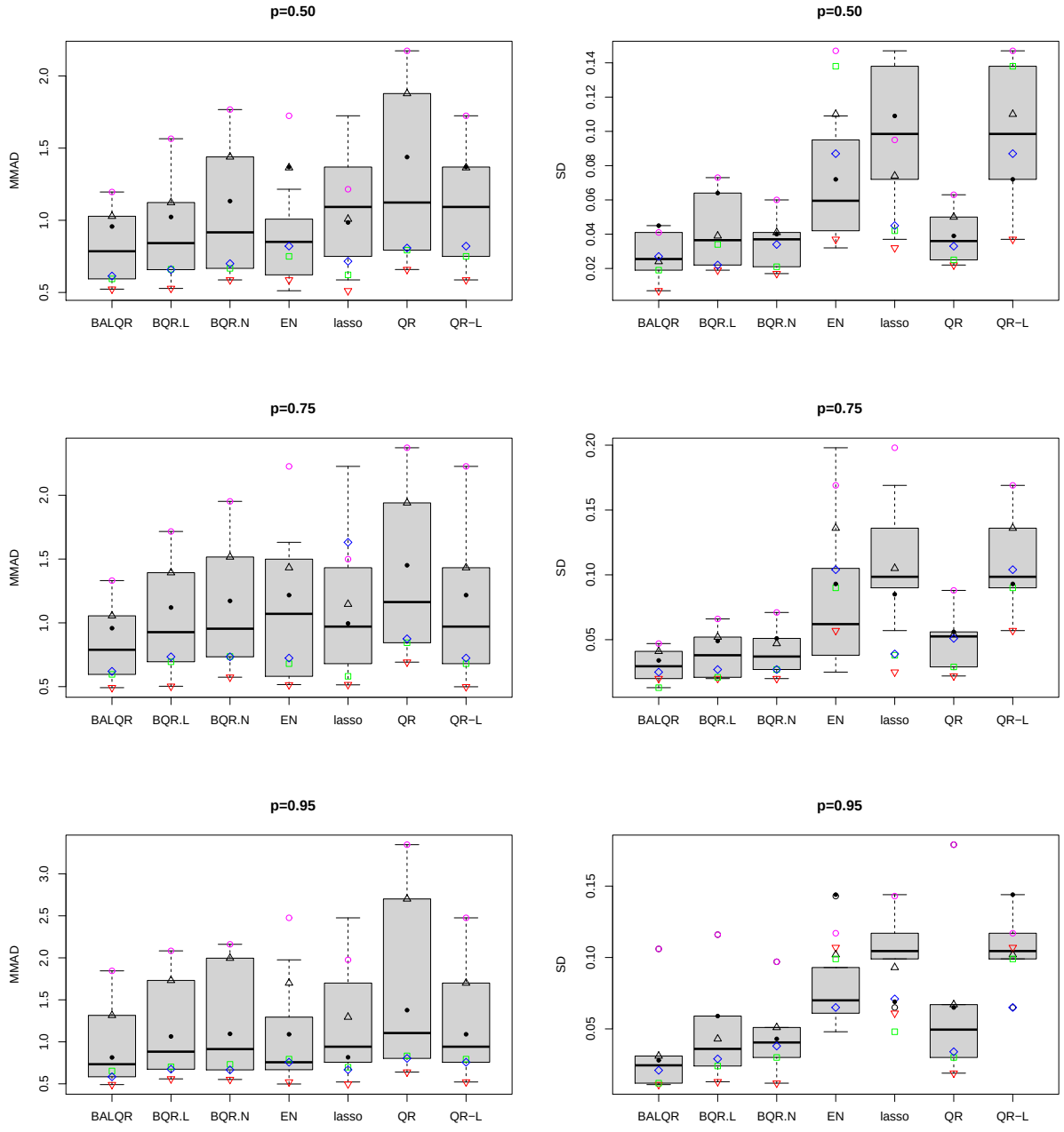

Figure 4: Boxplots summarizing the MMADs and the standard deviations of MMADs (SD) for the seven methods using the six error distributions in Simulation 4. Overlaid are the normal error ($\triangledown$), normal mixture ($\triangle$), Laplace ($\square$), Laplace mixture ($\circ$), t_3 ($\diamond$) and χ^2_3 ($\bullet$).

when $p = 0.75$ and $p = 0.95$ the proposed method was significantly efficient than the other six methods. Secondly, from Table 1 we see that, in general, the proposed method performs

Table 1: Posterior means for the simulated data when the error is normal and p=0.95.

Simulation Study	Method	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$
1	$\boldsymbol{\beta}^{true}$	3.000	1.500	0.000	0.000	2.000	0.000	0.000	0.000
	BALQR	2.988	1.469	0.002	0.013	1.994	0.001	-0.006	0.011
	BQR.L	2.937	1.446	-0.024	0.046	1.964	0.054	-0.019	0.037
	BQR.N	2.922	1.472	-0.030	0.049	1.957	0.040	-0.033	0.065
	lasso	2.811	1.357	0.000	0.000	1.782	0.000	0.000	0.000
	EN	2.796	1.453	0.000	0.000	1.774	0.000	0.000	0.000
	QR	2.960	1.462	-0.053	0.038	2.000	0.034	-0.049	0.057
	QR-L	2.915	1.392	0.000	0.000	1.810	0.000	-0.001	0.000
2	$\boldsymbol{\beta}^{true}$	0.850	0.850	0.850	0.850	0.850	0.850	0.850	0.850
	BALQR	0.835	0.852	0.849	0.865	0.858	0.860	0.846	0.863
	BQR.L	0.811	0.820	0.774	0.865	0.851	0.877	0.790	0.878
	BQR.N	0.800	0.831	0.775	0.872	0.854	0.879	0.783	0.887
	lasso	0.805	0.836	0.788	0.834	0.849	0.837	0.747	0.873
	EN	0.710	0.814	0.897	0.889	0.915	0.913	0.808	0.731
	QR	0.838	0.820	0.786	0.857	0.865	0.853	0.790	0.897
	QR-L	0.463	0.753	0.691	0.522	0.765	0.429	0.784	0.555
3	$\boldsymbol{\beta}^{true}$	5.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	BALQR	4.971	0.004	-0.024	0.008	0.012	0.007	-0.010	0.056
	BQR.L	4.883	0.010	-0.047	0.033	0.035	0.035	-0.031	0.058
	BQR.N	4.869	0.020	-0.055	0.045	0.042	0.048	-0.025	0.062
	lasso	4.591	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	EN	4.614	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	QR	4.934	-0.019	-0.053	0.018	0.026	0.054	-0.049	0.067
	QR-L	4.936	0.003	-0.020	0.000	0.000	0.000	0.000	0.001

well when comparing the estimates of β_j with the true values of β_j.

3.2. Non-i.i.d. random errors

In this part, we consider the case of non-i.i.d. random errors to demonstrate the performance of our method. The data was generated from model 2 of Kocherginsky, He and Mu (2005)

$$y = 1 + x_1 + x_2 + x_3 + (1 + x_3)\varepsilon, \tag{16}$$

where $x_1 \sim N(0,1)$, $x_3 \sim$ uniform$[0,1]$, $x_2 = x_1 + x_3 + z$, where $z \sim N(0,1)$ and $\varepsilon \sim N(0,1)$. This data generating process is often used in the context of variable selection (e.g. Wu and Liu, 2009; Li, Xi and Lin, 2010). In this example, we generated five additional mutually independent standard normal noise variables, $x_4, ..., x_8$. The results are summarized in Table 2 and are based on 150 repetitions, each with sample size n = 100.

Table 2: MMSEs, MMADs and test errors for the simulation with heterogeneous random errors.

p	Method	MMSE (SD)	MMAD (SD)	Test Error (SD)
0.50	BALQR	**0.0243** (0.0031)	**0.2472** (0.0074)	**0.5931** (**0.0029**)
	BQR.L	0.0307 (0.0033)	0.2764 (0.0086)	0.6152 (0.0040)
	BQR.N	0.0288 (**0.0021**)	0.2778 (0.0105)	0.6162 (0.0040)
	lasso	0.1423 (0.0057)	1.2185 (0.0359)	0.7898 (0.0129)
	EN	0.2066 (0.0202)	1.4075 (0.0570)	0.8476 (0.0193)
	QR	0.0301 (0.0042)	0.2984 (**0.0053**)	0.6151 (0.0041)
	QR-L	0.1208 (0.0103)	0.2902 (0.0058)	0.6122 (0.0031)
0.75	BALQR	**0.0213** (0.0021)	**0.2984** (0.0016)	**0.4233** (**0.0017**)
	BQR.L	0.0319 (0.0035)	0.3111 (0.0097)	0.4929 (0.0023)
	BQR.N	0.0289 (0.0023)	0.3018 (0.0086)	0.4927 (0.0019)
	lasso	0.2271 (0.0012)	1.4854 (0.0158)	0.7520 (0.0120)
	EN	0.2268 (**0.0006**)	1.5257 (**0.0012**)	0.7696 (0.0063)
	QR	0.0298 (0.0020)	0.3165 (0.0073)	0.4967 (0.0019)
	QR-L	0.1395 (0.0037)	0.3186 (0.0078)	0.4929 (0.0030)
0.95	BALQR	**0.0570** (0.0013)	**0.4041** (**0.0035**)	**0.1521** (**0.0008**)
	BQR.L	0.0780 (0.0094)	0.4640 (0.0143)	0.1664 (0.0018)
	BQR.N	0.0686 (0.0084)	0.4703 (0.0144)	0.1668 (0.0014)
	lasso	0.2361 (0.0015)	1.6075 (0.0109)	0.3277 (0.0073)
	EN	0.2304 (**0.0008**)	1.5890 (0.0066)	0.3167 (0.0041)
	QR	0.0860 (0.0060)	0.5135 (0.0173)	0.1706 (0.0020)
	QR-L	0.1371 (0.0063)	0.4390 (0.0209)	0.1657 (0.0013)

Table 2 lists the median of mean squared errors (MMSE), MMAD and test errors. It can be seen that the performance of our method is quite good compared to the other six methods. Most noticeably, when $p = 0.95$, our method behaves significantly better than the other methods. In addition, we can see that the Bayesian Lasso performs better than its non-Bayesian counterparts. This indicates that the model in (4) is merely a *working* model , in which the skewed Laplace distribution assumptions imposed on y_i are essentially artificial (Ying and Yin, 2010).

4. Prostate cancer data analysis

In this section, we analyze a real dataset to demonstrate the performance of the method proposed. For our real data example we use the data on prostate cancer reported by Stamey et al. (1989) and analyzed by Tibshirani (1996) and Yuan and Lin (2005), among others. This data set consists of the medical records of 97 male patients who were about to receive a radical prostatectomy and is available in the "`bayesQR`" R-package (Benoit et al., 2011). The response variable is the level of prostate antigen (lpsa) and there are eight predictors. These predictors are log cancer volume (lcavol), log prostate weight (lweight), age, log of the amount of benign prostatic hyperplasia (lbph), seminal vesicle invasion (svi), log of capsular penetration (lcp), Gleason score (gleason) and percentage of Gleason scores 4 or 5 (pgg45). We estimate a quantile regression model between the response lpsa and the 8 predictors without intercept. Similar to Section (3), we analyze three different quantiles, $p = 0.50, 0.75$ and 0.95. The shrinkage parameters in lasso (λ), EN (λ_1 and λ_2) and QR-L (λ) are tuned by 5-fold cross-validation.

Table 3 summarizes the results of 5-fold cross-validation technique for all methods. The results show that the BQR.N outperforms the other methods when $p = 0.50$. However, the performance of the method proposed in this study is very close to the performance of the BQR.N method. Moreover, for the quantiles $p = 0.75$ and $p = 0.95$, the method proposed

performs better than the other six methods. Also, the results show that the Bayesian lasso method has a poor performance when p = 0.50 and p = 0.95 due to the high pairwise correlations between some of variables. Thus, the proposed method attempts to remedy the shortcomings of Bayesian Lasso by using adaptive weights for different predictors.

Figure 5 depicts the posterior estimates for the prostate cancer data set using different methods for $p = 0.95$. To increase the readibility of the plot, we add a slight horizontal shift to the estimators. We can see that our method gives very similar posterior mean estimates compared to the other Bayesian methods. However, more importantly, it can be observed that the credible intervals for our approach are narrower than the alternative Bayesian methods. Although the BALQR credible intervals are narrower, it is observed that the estimates of the other Bayesian methods still lie inside the BALQR credible intervals. Hence, the analysis show strong support for the use of the proposed method to inference for quantile regression.

Table 3: Cross validation results for the prostate cancer data analysis. Standard errors are in parentheses.

	Test error		
Method	$p = 0.50$	$p = 0.75$	$p = 0.95$
BALQR	0.26754 (0.05448)	**0.26722** (**0.04798**)	**0.26743** (**0.04857**)
BQR.L	0.29061 (0.05952)	0.26979 (0.05803)	0.28289 (0.07249)
BQR.N	**0.26416** (**0.05214**)	0.28537 (0.07039)	0.27455 (0.05701)
lasso	0.27990 (0.05902)	0.27719 (0.06380)	0.27719 (0.06380)
EN	0.27938 (0.05897)	0.27876 (0.06002)	0.27876 (0.06002)
QR	0.27618 (0.05218)	0.27618 (0.05218)	0.27618 (0.05218)
QR-L	0.30146 (0.06471)	0.28493 (0.07208)	0.29032 (0.07216)

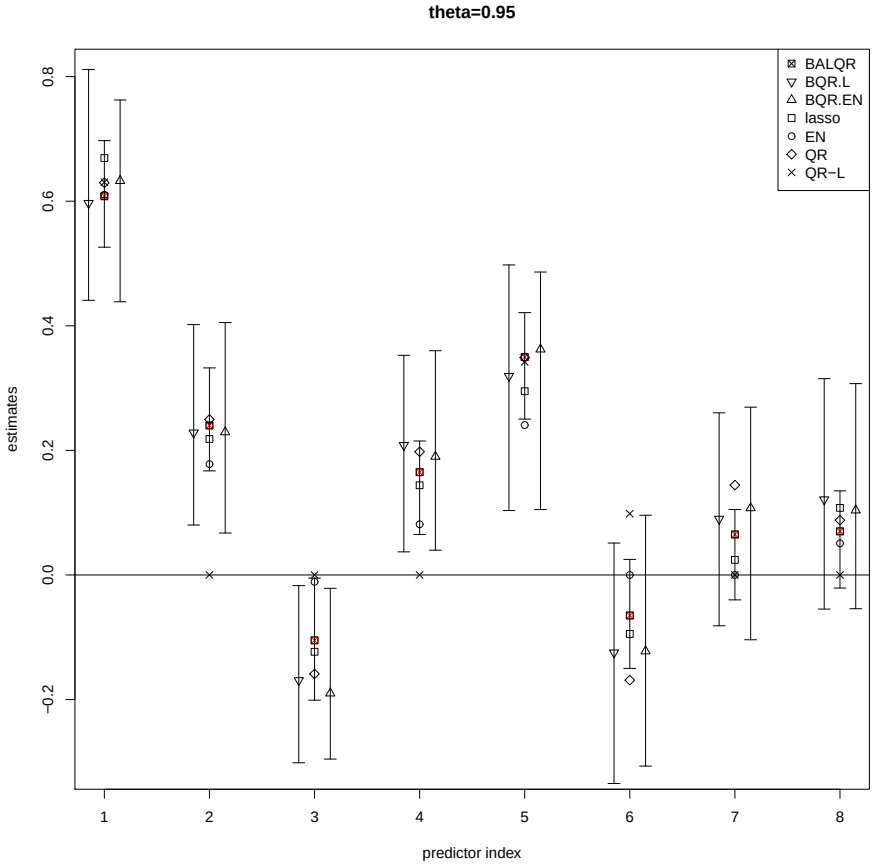

Figure 5: Posterior estimates for the prostate cancer data set using different methods with $p = 0.95$. The figure also shows 95% credible intervals for the Bayesian methods.

5. Conclusion

In this paper, we propose Bayesian adaptive Lasso quantile regression for variable selection and estimation. This method extends Bayesian quantile regression with Lasso penalty by allowing different penalization parameters for different regression coefficients. Independent inverse gamma priors are put on the penalty parameters. A novel aspect of the Bayesian adaptive Lasso quantile regression is to treat the hyperparameters of the inverse gamma priors as unknowns and let the data estimate them along with other parameters. This procedure allows us to control the amount of shrinkage in the inverse gamma priors. We developed

Bayesian hierarchical models for Bayesian adaptive Lasso quantile regression as well as introduced a Gibbs sampler for Bayesian adaptive Lasso quantile regression. This Gibbs sampler is based on a theoretic derivation of the skewed Laplace distribution as a scale mixture of normal distributions. The simulation studies and data analyses both show that the Bayesian adaptive Lasso quantile regression performs well and may be preferred over current existing Bayesian and non-Bayesian methods. The R-package "`bayesQR`" (Benoit et al., 2011) contains R function that implements the methodology proposed in this study. We hope that by making the code of our method available, we will lower the barrier for other researcher to use the Bayesian adaptive Lasso for quantile regression in their studies.

6. Acknowledgements

The authors wish to thank the Editor, the Associate Editor and the referees for their helpful comments and suggestions which have led to an improvement of this paper.

References

Alhamzawi, R., Yu, K. (2011). Power Prior Elicitation in Bayesian Quantile Regression, *Journal of Probability and Statistics*, vol. 2011, Article ID 874907, 16 pages. doi:10.1155/2011/874907

Alhamzawi, R., Yu, K., Vinciotti, V. and Tucker, A. (2011). Prior elicitation for mixed quantile regression with an allometric model. *Environmetrics*, 22: n/a. doi: 10.1002/env.1118

Andrews, D. F. and Mallows, C. L. (1974). Scale Mixtures of Normal Distributions. *Journal of the Royal Statistical Society, Ser. B* **36**, 99–102.

Benoit, D. F. and Van den Poel, D. (2011). Binary quantile regression: A Bayesian approach based on the asymmetric Laplace distribution. *Journal of Applied Econometrics* **26**, n/a. doi: 10.1002/jae.1216.

Benoit, D. F., Alhamzawi, R., Yu, K. and Van den Poel, D. (2011). bayesQR: Bayesian Quantile Regression. Available on CRAN R website: *http://cran.r-project.org/web/packages/bayesQ*

Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *Annals of Statistics* **24**, 2350-238.

Chhikara, R.S. and Folks, L. (1989). *The Inverse Gaussian Distribution: Theory, Methodology, and Applications.* Marcel Dekker, New York.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression (with discussion). *Annals of Statistics* **32**, 407-451.

Fahrmeir, L., Kneib, T. and Konrath, S. (2010). Bayesian regularisation in structured additive regression: a unifying perspective on shrinkage, smoothing and predictor selection. *Statistics and Computing* **20**, 203-219.

Fan, J. and Li, R. Z. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**, 1348-1360.

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

Gilks, W. R., (1992). Derivative-free adaptive rejection sampling for gibbs sampling. In Bernardo, J., Berger, J., Dawid, A. P., and Smith, A. F. M., editors, Bayesian Statistics 4, pages 641-649. Oxford University Press.

Griffin, J. E. and Brown P. J. Brown(2010). Bayesian adaptive Lassos with non-convex penalization. Technical report. Institute of Mathematics, Statistics and Actuarial Science, University of Kent.

Hans, C. (2009). Bayesian lasso regression. *Biometrika* **96**, 835–845.

Huang, J., Ma, S., and Zhang, C.-H. (2008). Adaptive Lasso for sparse high-dimensional regression models. Statistica Sinica **18**, 1603–1618.

Kocherginsky, M., He, X. and Mu, Y. (2005). Practical confidence intervals for regression quantiles. *Journal of Computational and Graphical Statistics* **14**, 41-55.

Koenker, R. (2005). *Quantile Regression.* New York: Cambridge University Press.

Koenker, R. (2004). Quantile Regression for Longitudinal Data. *Journal of Multivariate Analysis* **91**, 74-89.

Koenker, R. and Machado, J. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association* **94**, 1296-1310.

Kozumi, H. and Kobayashi, G. (2009). Gibbs Sampling Methods for Bayesian Quantile Regression. Technical report, Graduate School of Business Administration, Kobe University.

Li, Q., Xi, R. and Lin, N. (2010). Bayesian Regularized Quantile Regression. *Bayesian Analysis* **5**, 1-24.

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

Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression (with Discussion). *Journal of the American Statistical Association* **83**, 1023-1036.

Park, T. and Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association* **103**, 681–686.

Polsen, N. G. and Scott, J. G. (2011): Shrink Globally, act locally: Sparse Bayesian regularization and prediction, Proceedings of Valencia International Conference 2010, forthcoming.

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

Reed, C. and Yu, K. (2009). An efficient Gibbs sampler for Bayesian quantile regression. Technical report, Department of Mathematical Sciences, Brunel University.

Rue, H., Held, L. (2005). Gaussian Markov Random Fields: Theory and Applications. In *Monographs on Statistics and Applied Probability* **104**. Chapman and Hall, London.

Stamey, T., Kabalin, J., McNeal, J., Johnstone, I., Freiha, F., Redwine, E. and Yang, N. (1989). Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate, ii: Radical prostatectomy treated patients. *J. Urol* **16,** 1076–1083.

Sun, W., Ibrahim, J. G., and Zou, Fei. (2010). Genomewide multiple-loci mapping in experimental crosses by iterative adaptive penalized regression. *Genetics Society of America* DOI: 10.1534/genetics.110.114280

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Ser. B* **58,** 267-288.

Tsionas, E. G. (2003). Bayesian quantile inference. *Journal of Statistical Computation and Simulation.* **73**, 659-674.

Wang, H., Li, G., and Jiang, G. (2007). Robust Regression Shrinkage and Consistent Variable Selection Through the LAD-Lasso. *Journal of Business and Economic Statistics* **25**, 347-355.

Wu, Y. and Liu, Y. (2009). Variable Selection in Quantile Regression. *Statistica Sinica.* **19**, 801-817.

Yi, N. and Xu, S., (2008). Bayesian LASSO for Quantitative Trait Loci Mapping. *Genetics* **179**, 1045-1055.

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

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

Yuan, M. and Lin, Y. (2005). Efficient empirical Bayes variable selection and estimation in linear models. *J. Amer. Statist. Assoc.* **100,** 1215–1224.

Yuan, Y. and Yin, G. (2010). Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics* **66**, 105-114.

Yue, Y. R. and Rue, H. (2010). Bayesian inference for additive mixed quantile regression models. *Computational Statistics and Data Analysis.* To appear.

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