Scientific Research

Penalized Flexible Bayesian Quantile Regression

Ali Alkenani[1,2*], Rahim Alhamzawi[1,2], Keming Yu[1]

[1]Mathematical Science Department, School of Information Systems, Computing and Mathematics,
Brunel University, Uxbridge, UK
[2]Statistics Department, College of Administration and Economics, Al-Qadisiyah University, Al Diwaniyah, Iraq
Email: [*]mapgaja@brunel.ac.uk

ABSTRACT

The selection of predictors plays a crucial role in building a multiple regression model. Indeed, the choice of a suitable subset of predictors can help to improve prediction accuracy and interpretation. In this paper, we propose a flexible Bayesian Lasso and adaptive Lasso quantile regression by introducing a hierarchical model framework approach to enable exact inference and shrinkage of an unimportant coefficient to zero. The error distribution is assumed to be an infinite mixture of Gaussian densities. We have theoretically investigated and numerically compared our proposed methods with Flexible Bayesian quantile regression (FBQR), Lasso quantile regression (LQR) and quantile regression (QR) methods. Simulations and real data studies are conducted under different settings to assess the performance of the proposed methods. The proposed methods perform well in comparison to the other methods in terms of median mean squared error, mean and variance of the absolute correlation criterions. We believe that the proposed methods are useful practically.

Keywords: Adaptive Lasso; Lasso; Mixture of Gaussian Densities; Prior Distribution; Quantile Regression

1. Introduction

Quantile regression has become a widely used technique to describe the distribution of a response variable given a set of explanatory variables. It provides a more complete statistical analysis of the stochastic relationships among random variables. It has been applied in many different areas such as finance, microarrays, medical and agricultural studies—see Koenker [1] and Yu *et al.* [2] for more details.

Let y_i be a response variable and x_i a $k \times 1$ vector of covariates for the ith obsevation, x_i' is the ith row in the X matrix, $q_\tau(x_i)$ is the inverse cumulative distribution function of y_i given x_i. Then, the relationship between $q_\tau(x_i)$ and x_i can be modelled as $q_\tau(x_i) = x_i'\beta_\tau$, where β_τ is a vector of k unknown parameters of interest and τ determines the quantile level.

Koenker and Bassett [3] demonstrated that the regression coefficients β_τ can be estimated by

$$\min_{\beta_\tau} \sum_{i=1}^n \rho_\tau(y_i - x_i'\beta_\tau). \tag{1}$$

where $\rho_\tau(\cdot)$ is the check function defined by

$$\rho_\tau(u) = \tau u I_{[0,\infty)}(u) - (1-\tau) u I_{(-\infty,0)}(u). \tag{2}$$

As a parametric possible link with minimizing the check function (1), Koenker and Machado [4] showed that maximum likelihood solution of the asymmetric Laplace (ASL) distribution is equivalent to the minimization of an objective function in (1). The idea of Koenker and Machado [4] was exploited by Yu and Moyeed [5] to propose Bayesian quantile regression. Yu and Moyeed [5] considered Markov Chain Monte Carlo (MCMC) methods for posterior inference. Recently, Bayesian approach for quantile regression has attracted much interest in literature. For example, Tsionas [6] developed Gibbs sampling algorithm for the quantile regression model, while Yu and Standard [7] proposed Bayesian Tobit quantile regression. Additionally, Geraci and Bottai [8] considered Bayesian quantile regression for longitudinal data using asymmetric Laplace distribution. Likewise, Kozumi and Kobayashi [9] and Reed and Yu [10] developed Gibbs sampling algorithms based on a location-scale mixture representation of the (ASL) distribution while Benoit and Poel [11] proposed Bayesian binary quantile regression.

Some researchers suggested nonparametric approaches to avoid the restrictive assumptions of the parametric approaches. See for example Walker and Mallick [12], Kottas and Gelfand [13], Hanson and Johnson [14], Hjort [15], Hjort and Petrone [16], Taddy and Kottas [17] and Kottas and Krnjajic [18]. Recently, Reich *et al.* [19] pro-

---

[*]Corresponding author.

posed a flexible Bayesian quantile regression model for independent and clustered data. The authors assumed that the error distribution is an infinite mixture of Gaussian densities. They called their method "flexible" because it does not impose a parametric assumptions (e.g., asymmetric Laplace) or shape restrictions on the residual distribution (e.g., mode at the quantile of interest) as with other approaches (personal communication with Reich).

The selection of predictors plays a crucial role in building a multiple regression model. The choice of a suitable subset of predictors can help to improve prediction accuracy. Also, in practice, the interpretation of a smaller subset of predictors is easier than a large set of predictors (Li *et al.* [20]). Variable selection by penalizing the classical least squares has attracted much research interest. See for example least absolute shrinkage and selection operator (Lasso) (Tibshirani [21]), smoothly clipped absolute deviation (SCAD) (Fan and Li [22]), Adaptive Lasso (Zou [23]) and the Bayesian approach of Park and Casella [24]. Although the classical least squares is popular for its mathematical beauty, it is not robust to outliers (Bradic *et al.* [25]; Koenker and Bassett [3]). For this reason, robust variable selection can be achieved using a rigorous method such as quantile regression.

The first use of regularization in quantile regression is made by Koenker [26]. The author put the Lasso penalty on the random effects in a mixed-effect quantile regression model to shrink individual effects towards a common value. Yuan and Yin [27] proposed Bayesian approach to shrink the random effects towards a common value by introducing l_2 penalty in the usual quantile regression check function. In addition, Wang *et al.* [28] 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 [29] developed the piecewise linear solution path of the l_1 penalized quantile regression. Moreover, Wu and Liu [30] considered penalized quantile regression with the SCAD and Adaptive Lasso penalties. Li *et al.* [20] suggested Bayesian regularized quantile regression. The authors proposed different penalties including Lasso, group Lasso and elastic net penalties. Alhamzawi *et al.* [31] extended the Bayesian Lasso quantile regression reported in Li *et al.* [20] by allowing different penalization parameters for different regression coefficients.

In this paper, we develop a flexible Bayesian framework for regularization in quantile regression model. Similar to Reich *et al.* [19], we assume the error distribution to be an infinite mixture of Gaussian densities. This work is quite different from Bayesian Lasso quantile regression employing asymmetric Laplace error distribution. In fact, the use of the asymmetric Laplace distribution is unattractive due to the lack of coherence (Kottas

and Krnjajić [18]). For example, for different τ we have different distribution for the y_i's and it is difficult to reconcile these differences. Our motivating example is an analysis of body fat data which is previously analyzed by Johnson [32] and available in the package "mfp". This study had a total body measurements of 252 men. The objective of this study is to investigate the relationship between the percentage body fat and 13 simple body measurements. In this paper we are interested in selecting the most significant simple body measurements for the quantile regression model, relating to the percentage body fat. Certain correlation is present between the predictors in the body fat data. For example, the correlation coefficient is 0.943 between the weight and the hip circumference, 0.916 between the chest circumference and the abdomen circumference, 0.894 between the hip circumference and the thigh circumference, 0.894 between the weight and the chest circumference, 0.887 between the weight and the abdomen circumference, 0.874 between the abdomen circumference and the hip circumference, and so on. The selection of variables is important in this application, in order to know which predictors have coefficients that vary among subjects. The high 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.

The remainder of this paper is organized as follows. A brief review of the flexible Bayesian quantile regression model for independent data is given in Section 2. Penalized flexible Bayesian quantile regression with Lasso and adaptive Lasso are proposed in Section 3 and Section 4, respectively. The experimental results are reported in Section 5. Finally, the conclusions are summarized in Section 6.

2. Flexible Bayesian Quantile Regression (FBQR)

Following He [33], Reich *et al.* [19] considered the heteroskedastic linear regression model

$$y_i = x_i'\beta + x_i'\gamma\varepsilon_i \qquad (3)$$

where $x_i'\gamma > 0$ for all x_i and the residuals ε_i are independent and identically distributed. The authors rewrote the above model as quantile regression model:

$$y_i = x_i'\beta_{(\tau)} + x_i'\gamma_{(\tau)}\varepsilon_{i(\tau)} \qquad (4)$$

where $\varepsilon_{i(\tau)} = \varepsilon_i - q_\tau(\varepsilon)$ has τth quantile equal to zero, $q_\tau(\varepsilon)$ is the inverse cumulative distribution function of ε_i. To analyze y_i's τth quantile, $x_i'\beta_{(\tau)}$, the authors considered only distributions for the residual term with τth quantile equal to zero. Also, they fixed the element of $\gamma_{(\tau)}$ corresponding to the intercept at 1 to separate out the scale of the errors from $\gamma_{(\tau)}$. For simplicity of nota-

tion, we will omit the subscript τ in the remainder of the paper. From a Bayesian approach view, Reich *et al.* [19] assumed that the error term ε_i has a normal distribution with mean $\mu_{H_i G_i}$ and variance $\sigma^2_{H_i G_i}$, where $G_i \in \{1, 2, \cdots\}$ and $H_i \in \{1, 2\}$ are latent variables introduced to indicate the mixture component from which the *i*th observation is drawn. Under these assumptions, the conditional distribution of y_i given $\mu_{H_i G_i}$ and $\sigma^2_{H_i G_i}$ is given by:

$$f\left(y_i \mid \mu_{H_i G_i}, \sigma^2_{H_i G_i}\right) \propto \left(x_i' \gamma \sigma_{H_i G_i}\right)^{-1}$$
$$\exp\left\{-\frac{\left(y_i - x_i'\beta - x_i'\gamma \mu_{H_i G_i}\right)^2}{2\left(x_i'\gamma \sigma_{H_i G_i}\right)^2}\right\}, \quad (5)$$

where $\sigma_{1m}, \sigma_{2m} \sim \text{uniform}(0, c_1)$ and $\mu_{1m}, \mu_{2m} \sim ALD(0, \theta, \tau)$, where the parameters are the location, scale and the skewness. To specify a prior for p_m, where the p_m are the mixture proportions with $\sum_{m=1}^{\infty} p_m = 1$, Reich *et al.* [19] defined the proportions $p_m, (m = 1, 2, \cdots)$ through the latent variables V_m which are independently identically distributed from $\text{beta}(1, D)$ where D controls the strength of the prior for p_m. The first proportion is $p_1 = V_1$ and the others are given by $p_m = V_m \prod_{d < m} (1 - V_m)$ where $1 - \sum_{d < m} p_j = \prod_{d < m} (1 - V_m)$.

The priors for the regression coefficients and scale parameters are $\beta_j \sim N(0, c_2)$ for large c_2 and $\theta \sim \text{Gamma}(0.1, 0.1)$. The prior for γ_j is vague normal prior subject to $x_i'\gamma > 0$ for all x_i and the first element of γ corresponding to the intercept to be 1 to identify the scale of the residuals. As shown in Reich *et al.* [19], the performance of the above method is better than the frequentist method.

3. Flexible Bayesian Quantile Regression with Lasso Penalty (FBLQR)

Tibshirani [21] proposed the Lasso for simultaneous variable selection and parameter estimation. The Lasso, formulated in the penalized likelihood framework, minimizes the residual sum of squares with a constraint on the l_1 norm of β. The author stated the Lasso estimator can be interpreted as the posterior mode using normal likelihood and iid Laplace prior for β. As extension to Lasso Tibshirani [21], Li and Zhu [29] suggested Lasso quantile regression for simultaneous estimation and variable selection in quantile regression models, and it is given by:

$$\min_{\beta} \sum_{i=1}^{n} \rho_\tau \left(y_i - x_i'\beta\right) + \lambda \sum_{j=1}^{k} |\beta_j|. \quad (6)$$

where $\lambda > 0$ is the tuning parameter controlling the amount of penalty. The second term in (6) is the so-

called l_1 penalty, which is essential for the success of the Lasso. In this paper we consider a fully Bayesian approach to enable exact inference and shrinkage of an unimportant coefficient to zero. We propose flexible Bayesian Lasso quantile regression which solves the following:

$$\min_{\beta} \left(u - x_i'\beta\right)' W^{-1} \left(u - x_i'\beta\right) + \lambda \sum_{j=1}^{k} |\beta_j| \quad (7)$$

where $u = y_i - x_i'\gamma \mu_{H_i G_i}$ and W is a diagonal matrix with the element $\left(x_i'\gamma \sigma_{H_i G_i}\right)^2$ on the diagonal i. The second term of (7) can be represented as a mixture of normals (Andrews and Mallows [34])

$$\frac{\lambda}{2} e^{-\lambda|\tau|} = \int_0^{\infty} \frac{1}{\sqrt{2\pi s}} e^{-z^2/(2s)} \frac{\lambda^2}{2} e^{-\lambda^2 s/2} ds, \lambda > 0$$ We further

put gamma priors, $f(\lambda^2) = \lambda^{2(c_1 - 1)} e^{(-c_2 \lambda^2)}$, on the parameter λ^2 (not λ). Then, we have the following hierarchical model:

$$y_i \big/ \mu_{H_i G_i}, \sigma^2_{H_i G_i} \sim N\left(x_i'\beta + x_i'\gamma \mu_{H_i G_i}, \left(x_i'\gamma \sigma_{H_i G_i}\right)^2\right),$$

$$\beta \big/ \lambda^2 \sim \int_0^{\infty} \frac{1}{\sqrt{2\pi s}} e^{-\beta^2/(2s)} \frac{\lambda^2}{2} e^{-\lambda^2 s/2} ds$$

$$f(\lambda^2) = \lambda^{2(c_1 - 1)} e^{(-c_2 \lambda^2)} \mu_{1m}, \mu_{2m} \sim ALD(0, \theta, \tau),$$

$$\sigma_{1m}, \sigma_{2m} \sim \text{Uniform}(0, c_1),$$

$$G_i \sim \text{Categorical}(p_1, p_2, \cdots),$$

$$H_i \sim \text{Categorical}(q_{G_i}, 1 - q_{G_i})$$

$$q_m = \frac{\tau - \Phi(-\mu_{2m}/\sigma_{2m})}{\Phi(-\mu_{1m}/\sigma_{1m}) - \Phi(-\mu_{2m}/\sigma_{2m})}$$

$$p_1 = V_1, p_m = V_m \prod_{d < m} (1 - V_m) \quad \text{where}$$

$1 - \sum_{d < m} p_d = \prod_{d < m} (1 - V_m)$, the latent variables V_m which are independently identically distributed from beta $(1, D)$. The details of the Gibbs sampler are given in the appendix.

4. Flexible Bayesian Quantile Regression with Adaptive Lasso Penalty (FBALQR)

Fan and Li [22] studied a class of penalization methods including the Lasso. The authors showed that the Lasso can perform automatic variable selection because the l_1 penalty is singular at the origin. On the other hand, the Lasso shrinkage produces biased estimates for the large coefficients, and thus it could be suboptimal in terms of estimation risk. Fan and Li [22] conjectured that the oracle properties do not hold for the Lasso. The Adaptive Lasso can be viewed as a generalization of the Lasso penalty. Basically the idea is to penalize the coefficients

of different covariates at a different level by using adaptive weights. In the case of least squares regression, Zou [23] proposed the Adaptive Lasso in which adaptive weights are used to penalize different coefficients in the l_1 penalty. The author showed that the adaptive Lasso has the oracle properties that Lasso does not have. We propose Flexible Bayesian adaptive Lasso quantile regression which solves the following

$$\min_{\beta}(u-x_i'\beta)'W^{-1}(u-x_i'\beta)+\sum_{j=1}^{k}\lambda_j|\beta_j| \quad (8)$$

The second term of (8) can be represented as a mixture of normals (Andrews and Mallows [34])

$$\frac{\lambda_j}{2}e^{-\lambda_j|z_j|}=\int_0^{\infty}\frac{1}{\sqrt{2\pi s_j}}e^{-z_j^2/(2s_j)}\frac{\lambda_j^2}{2}e^{-\lambda_j^2 s_j/2}\mathrm{d}s_j,\ \lambda_j^2>0 \quad \text{for}$$

each $\lambda_j, j=1,\cdots,k$, we assume gamma prior, $f(\lambda_j^2)=\lambda_j^{2(c_1-1)}e^{(-c_2\lambda_j^2)}$, on the parameter λ^2. Then we have the following hierarchical model:

$$y_i\big/\mu_{H_iG_i},\sigma_{H_iG_i}^2 \sim N\left(x_i'\beta+x_i'\gamma\mu_{H_iG_i},\left(x_i'\gamma\sigma_{H_iG_i}\right)^2\right),$$

$$\beta_j\big/\lambda_j^2 \sim \int_0^{\infty}\frac{1}{\sqrt{2\pi s_j}}e^{-\beta_j^2/(2s)}\frac{\lambda_j^2}{2}e^{-\lambda_j^2 s_j/2}\mathrm{d}s_j$$

$$f(\lambda_j^2)=\lambda_j^{2(c_1-1)}e^{(-c_2\lambda_j^2)}\mu_{1m},\mu_{2m}\sim ALD(0,\theta,\tau),$$

$$\sigma_{1m},\sigma_{2m}\sim \text{Uniform}(0,c_1),$$

$$G_i\sim \text{Categorical}(p_1,p_2,\cdots),$$

$$H_i\sim \text{Categorical}(q_{G_i},1-q_{G_i})$$

$$q_m=\frac{\tau-\Phi(-\mu_{2m}/\sigma_{2m})}{\Phi(-\mu_{1m}/\sigma_{1m})-\Phi(-\mu_{2m}/\sigma_{2m})}$$

$$p_1=V_1, p_m=V_m\prod_{d<m}(1-V_m)$$

where $1-\sum_{d<m}p_j=\prod_{d<m}(1-V_m)$, the latent variables V_m

which are independently identically distributed from beta $(1,D)$.

5. The Experimental Results

5.1. A Simulation Study

A numerical study was conducted to assess the performance of the proposed methods. We generated $R=200$ data-sets with size $n=300$ observations from $y_i=x_i'\beta+x_i'\gamma\varepsilon_i$, where x_i are generated as independently and identically distributed standard normals. We simulated the error ε_i from three possible error distributions: standard normal, a $t_{(3)}$ distribution with three

degrees of freedom and Chi-squared distribution with three degrees of freedom $(\chi_{(3)}^2)$. We use the following designs for the vector β:

Design 1: $\beta=(1;-1;-1;0;0;0;0;0;0;0;0;0;0;0)'$

Design 2: $\beta=(3;-3;-3;0;0;0;0;0;0;0;0;0;0;0)'$

Design 3: $\beta=(1;-1;-1;0;0;0;0;0;0;0;0;0;1;-1;-1)'$

Design 4: $\beta=(3;-3;-3;0;0;0;0;0;0;0;0;0;3;-3;-3)'$.

where the first element in β corresponding to the intercept.

Each simulated data set is analysis using five methods. We use our methods flexible Bayesian quantile regression with Lasso penalty (FBLQR) and flexible Bayesian quantile regression with Adaptive Lasso penalty (FBALQR) which are proposed in Sections 3 and 4 respectively. We compare the proposed methods with Lasso quantile regression (LQR) and the standard frequentist quantile regression (QR) using the "quantreg" package in R. Also, the proposed methods are compared with flexible Bayesian quantile regression (FBQR) (Reich *et al.* [19]).

The results of the simulation are presented in **Tables 1-4** and **Figures 1-4**. From **Tables 1-4** and **Figures 1-4**, according to the absolute correlation value of $X\hat{\beta}$ and $X\beta(|r|)$ and the median of the mean squared error (MMSE) for $X\hat{\beta}$, it can be seen that the proposed methods (FBALQR and FBLQR) are better than the other three methods for all the distributions under consideration and for all the τ values. This indicates that the proposed methods give precise estimates even when the error distribution is an asymmetric. Most noticeably, when $\tau=0.25$ and $\tau=0.75$ the FBALQR and FBLQR are more significantly efficient than the other methods. In addition, we can observe that the worst estimators for all the τ values are QR.

5.2. Body Fat Data

Percentage of body fat is one important measure of health, which can be accurately estimated by underwater weighing techniques. These techniques often require special equipment and are sometimes not easily achieved, thus fitting percentage body fat to simple body measurements is a convenient way to predict body fat. Johnson [32] introduced a data set in which percentage body fat and 13 simple body measurements (such as weight, height and abdomen circumference) are recorded for 252 men. The data set is available at the package ("mfp"). The response variable *Y* is the percent boday fat (%). The $k=13$ predictor variables x are, respectively, the age (years) x_1, the weight (pounds) x_2, the height (inches) x_3, the neck circumference (cm) x_4, the chest circumference (cm) x_5, the abdomen circumference (cm) x_6,

Table 1. Simulation results for FBALQR, FBLQR, FBQR, LQR and QR based on the linear model $y_i = x_i\beta + x_i\gamma\varepsilon_i$, design 1.
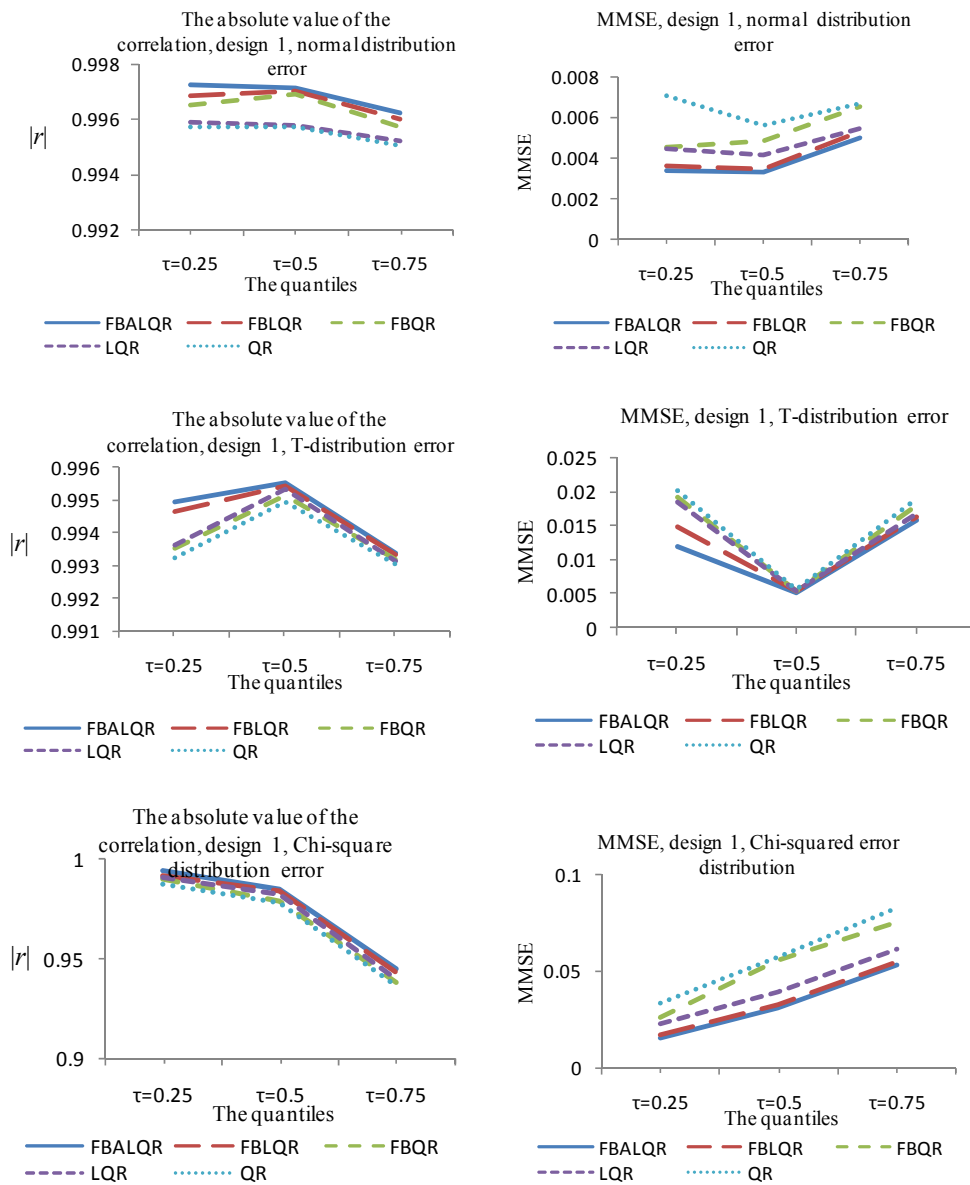
τ	Method	Error distribution								
		Standard normal			$t_{(3)}$			$\chi^2_{(3)}$		
		$\|r\|$ mean	$\|r\|$ SD	MMSE	$\|r\|$ mean	$\|r\|$ SD	MMSE	$\|r\|$ mean	$\|r\|$ SD	MMSE
	FBALQR	0.99724	0.00131	0.00341	0.99494	0.00224	0.01182	0.99460	0.00172	0.01544
	FBLQR	0.99684	0.00120	0.00363	0.99464	0.00241	0.01495	0.99210	0.00155	0.01673
$\tau = 0.25$	FBQR	0.99655	0.00142	0.00452	0.99352	0.00251	0.01910	0.99053	0.00141	0.02620
	LQR	0.99591	0.00141	0.00445	0.99361	0.00385	0.01852	0.99075	0.00143	0.02261
	QR	0.99572	0.00143	0.00703	0.99322	0.00372	0.02010	0.98805	0.00196	0.03304
	FBALQR	0.99714	0.00130	0.00331	0.99550	0.00113	0.00527	0.98537	0.00824	0.03151
	FBLQR	0.99703	0.00131	0.00344	0.99543	0.00111	0.00554	0.98461	0.00726	0.03264
$\tau = 0.5$	FBQR	0.99693	0.00144	0.00480	0.99514	0.00243	0.00535	0.97965	0.00733	0.05517
	LQR	0.99580	0.00176	0.00412	0.99532	0.00213	0.00531	0.98221	0.00601	0.03926
	QR	0.99571	0.00182	0.00562	0.99492	0.00127	0.00560	0.97846	0.00657	0.05702
	FBALQR	0.99623	0.00103	0.00503	0.99337	0.00251	0.01572	0.94525	0.01794	0.05344
	FBLQR	0.99602	0.00101	0.00530	0.99331	0.00244	0.01632	0.94325	0.01664	0.05443
$\tau = 0.75$	FBQR	0.99575	0.00087	0.00653	0.99316	0.00315	0.01791	0.93853	0.01761	0.07512
	LQR	0.99521	0.00143	0.00549	0.99314	0.00337	0.01689	0.93979	0.01783	0.06122
	QR	0.99504	0.00139	0.00665	0.99302	0.00311	0.01895	0.93623	0.02290	0.08290

Table 2. Simulation results for FBALQR, FBLQR, FBQR, LQR and QR based on the linear model $y_i = x_i\beta + x_i\gamma\varepsilon_i$, design 2.
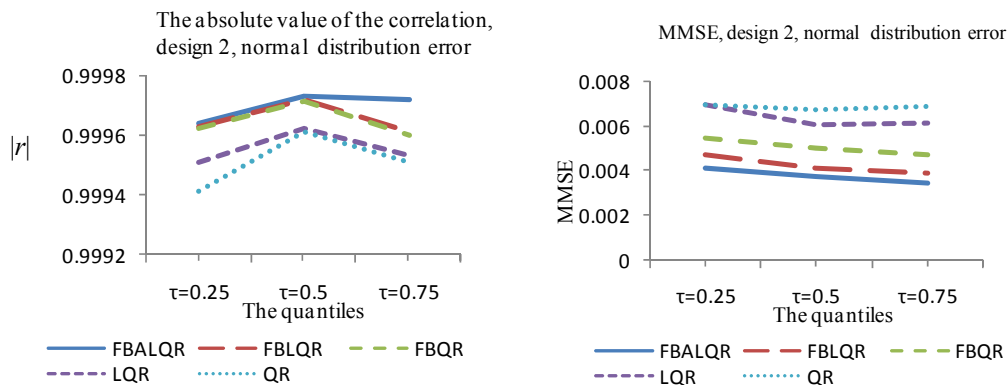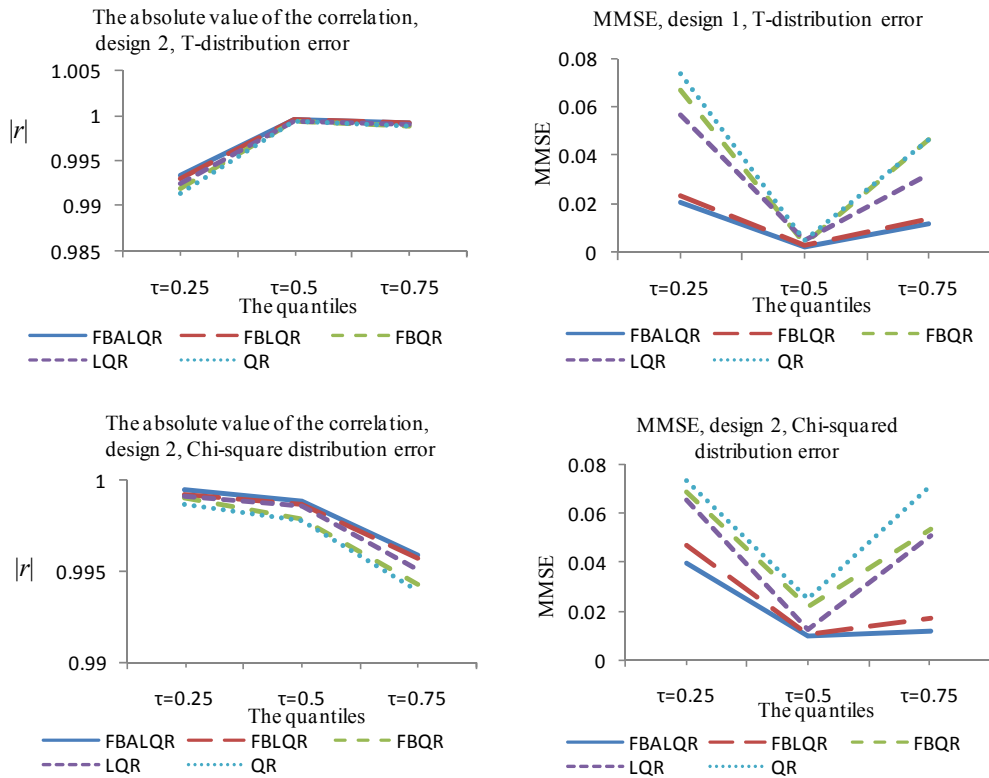
τ	Method	Error distribution								
		Standard normal			$t_{(3)}$			$\chi^2_{(3)}$		
		$\|r\|$ mean	$\|r\|$ SD	MMSE	$\|r\|$ mean	$\|r\|$ SD	MMSE	$\|r\|$ mean	$\|r\|$ SD	MMSE
	FBALQR	0.99964	0.00021	0.00415	0.99335	0.00264	0.02084	0.99949	0.00035	0.03962
	FBLQR	0.99963	0.00013	0.00473	0.99293	0.00271	0.02313	0.99926	0.00055	0.04691
$\tau = 0.25$	FBQR	0.99962	0.00022	0.00544	0.99190	0.00233	0.06672	0.99903	0.00036	0.06922
	LQR	0.99951	0.00021	0.00693	0.99241	0.00210	0.05682	0.99913	0.00043	0.06550
	QR	0.99941	0.00023	0.00694	0.99134	0.00312	0.07380	0.99874	0.00043	0.07351
	FBALQR	0.99973	0.00013	0.00377	0.99952	0.00024	0.00254	0.99885	0.00061	0.01026
	FBLQR	0.99972	0.00014	0.00412	0.99951	0.00024	0.00272	0.99873	0.00062	0.01094
$\tau = 0.5$	FBQR	0.99971	0.00016	0.00503	0.99943	0.00025	0.00355	0.99790	0.00070	0.02183
	LQR	0.99962	0.00021	0.00605	0.99942	0.00030	0.00471	0.99861	0.00095	0.01263
	QR	0.99961	0.00020	0.00675	0.99941	0.00050	0.00493	0.99781	0.00095	0.02541
	FBALQR	0.99972	0.00022	0.00342	0.99923	0.00033	0.01215	0.99589	0.00212	0.01207
	FBLQR	0.99961	0.00022	0.00391	0.99915	0.00033	0.01342	0.99574	0.00157	0.01765
$\tau = 0.75$	FBQR	0.99960	0.00024	0.00471	0.99882	0.00036	0.04641	0.99432	0.00243	0.05400
	LQR	0.99953	0.00022	0.00614	0.99911	0.00050	0.03240	0.99515	0.00291	0.05099
	QR	0.99951	0.00023	0.00690	0.99878	0.00051	0.04671	0.99391	0.00292	0.07122

Table 3. Simulation results for FBALQR, FBLQR, FBQR, LQR and QR based on the linear model $y_i = x_i\beta + x_i\gamma\varepsilon_i$, design 3.

τ	Method	Error distribution								
		Standard normal			$t_{(3)}$			$\chi^2_{(3)}$		
		$|r|$ mean	$|r|$ SD	MMSE	$|r|$ mean	$|r|$ SD	MMSE	$|r|$ mean	$|r|$ SD	MMSE
	FBALQR	0.99868	0.00042	0.00602	0.99755	0.00128	0.01001	0.99699	0.00066	0.01391
	FBLQR	0.99855	0.00044	0.00603	0.99744	0.00123	0.01043	0.99654	0.00091	0.01661
$\tau = 0.25$	FBQR	0.99754	0.00060	0.00624	0.99713	0.00132	0.01180	0.99603	0.00163	0.02303
	LQR	0.99853	0.00051	0.00613	0.99731	0.00130	0.01122	0.99615	0.00100	0.02275
	QR	0.99732	0.00063	0.00643	0.99701	0.00142	0.01263	0.99586	0.00157	0.02370
	FBALQR	0.99875	0.00052	0.00435	0.99852	0.00072	0.00593	0.99541	0.00173	0.01253
	FBLQR	0.99872	0.00054	0.00484	0.99824	0.00074	0.00664	0.99511	0.00180	0.01727
$\tau = 0.5$	FBQR	0.99823	0.00056	0.00543	0.99811	0.00077	0.00690	0.99270	0.00196	0.02465
	LQR	0.99863	0.00055	0.00495	0.99815	0.00076	0.00685	0.99491	0.00190	0.01965
	QR	0.99821	0.00059	0.00563	0.99802	0.00079	0.00688	0.99225	0.00201	0.03368
	FBALQR	0.99835	0.00052	0.00434	0.99765	0.00080	0.01025	0.98729	0.00471	0.04346
	FBLQR	0.99833	0.00055	0.00430	0.99757	0.00100	0.01044	0.98661	0.00505	0.04612
$\tau = 0.75$	FBQR	0.99812	0.00060	0.00476	0.99752	0.00112	0.01699	0.98354	0.00547	0.05250
	LQR	0.99815	0.00058	0.00441	0.99755	0.00111	0.01652	0.98453	0.00521	0.04792
	QR	0.99803	0.00061	0.00478	0.99754	0.00115	0.01910	0.98255	0.00574	0.05293

Table 4. Simulation results for FBALQR, FBLQR, FBQR, LQR and QR based on the linear model $y_i = x_i\beta + x_i\gamma\varepsilon_i$, design 4.

τ	Method	Error distribution								
		Standard normal			$t_{(3)}$			$\chi^2_{(3)}$		
		$|r|$ mean	$|r|$ SD	MMSE	$|r|$ mean	$|r|$ SD	MMSE	$|r|$ mean	$|r|$ SD	MMSE
	FBALQR	0.99994	0.00011	0.00435	0.99977	0.00010	0.02063	0.99978	0.00011	0.01440
	FBLQR	0.99991	0.00014	0.00482	0.99975	0.00012	0.03141	0.99976	0.00013	0.01622
$\tau = 0.25$	FBQR	0.99983	0.00018	0.00510	0.99973	0.00017	0.03340	0.99973	0.00022	0.01678
	LQR	0.99985	0.00015	0.00495	0.99974	0.00014	0.03272	0.99974	0.00021	0.01673
	QR	0.99982	0.00022	0.00530	0.99969	0.00019	0.04525	0.99963	0.00023	0.01701
	FBALQR	0.99996	0.00011	0.00114	0.99989	0.00012	0.00844	0.99962	0.00024	0.02352
	FBLQR	0.99993	0.00013	0.00122	0.99987	0.00014	0.00864	0.99941	0.00031	0.02796
$\tau = 0.5$	FBQR	0.99984	0.00022	0.00262	0.99982	0.00017	0.01826	0.99921	0.00036	0.03619
	LQR	0.99992	0.00017	0.00127	0.99985	0.00015	0.00958	0.99933	0.00034	0.02842
	QR	0.99982	0.00023	0.00300	0.99980	0.00021	0.02028	0.99906	0.00037	0.03653
	FBALQR	0.99987	0.00009	0.00286	0.99978	0.00012	0.02270	0.99876	0.00050	0.04221
	FBLQR	0.99986	0.00011	0.00445	0.99976	0.00013	0.02281	0.99852	0.00062	0.04271
$\tau = 0.75$	FBQR	0.99980	0.00020	0.00563	0.99970	0.00017	0.02403	0.99810	0.00073	0.04899
	LQR	0.99983	0.00014	0.00460	0.99972	0.00015	0.02322	0.99821	0.00065	0.04754
	QR	0.99977	0.00024	0.00670	0.99964	0.00020	0.02470	0.99791	0.00076	0.05685

The absolute value of the correlation, design 1, normal distribution error

MMSE, design 1, normal distribution error

The absolute value of the correlation, design 1, T-distribution error

MMSE, design 1, T-distribution error

The absolute value of the correlation, design 1, Chi-square distribution error
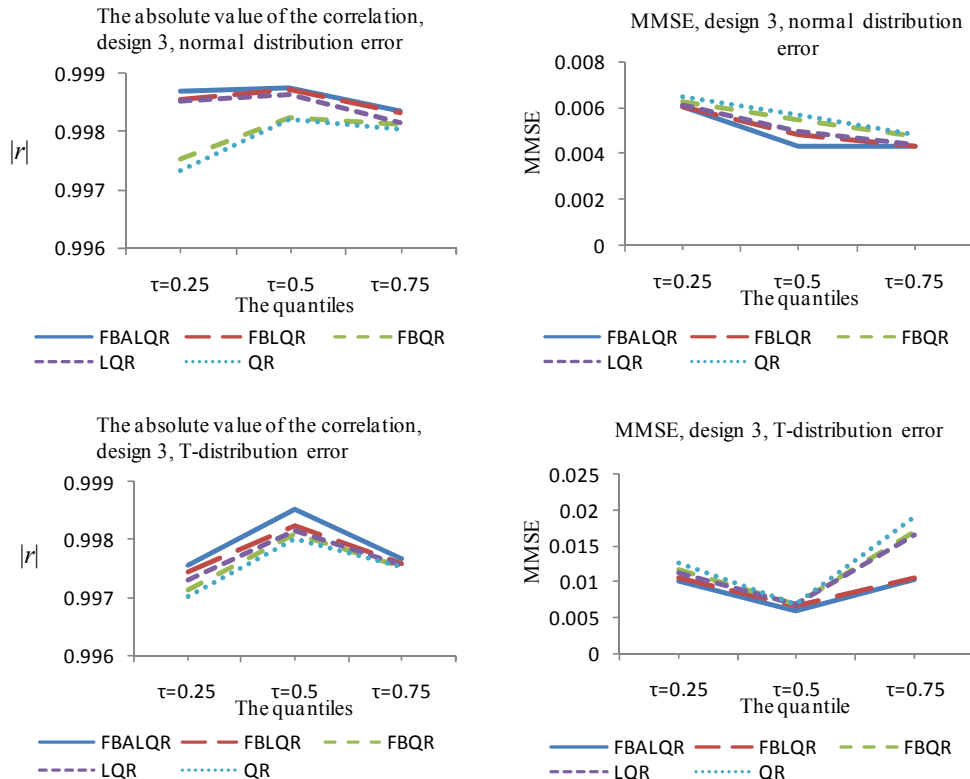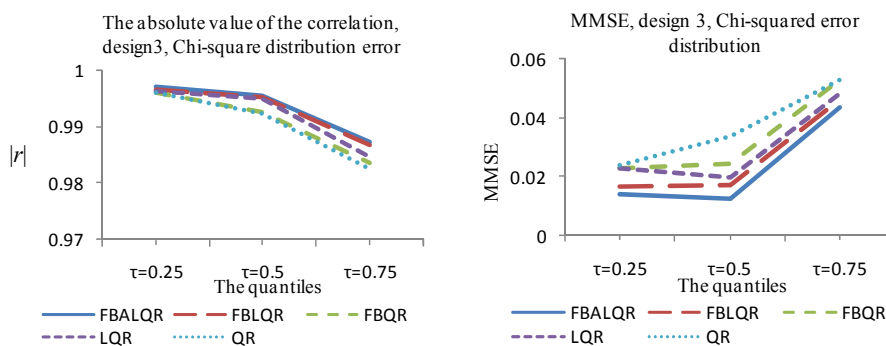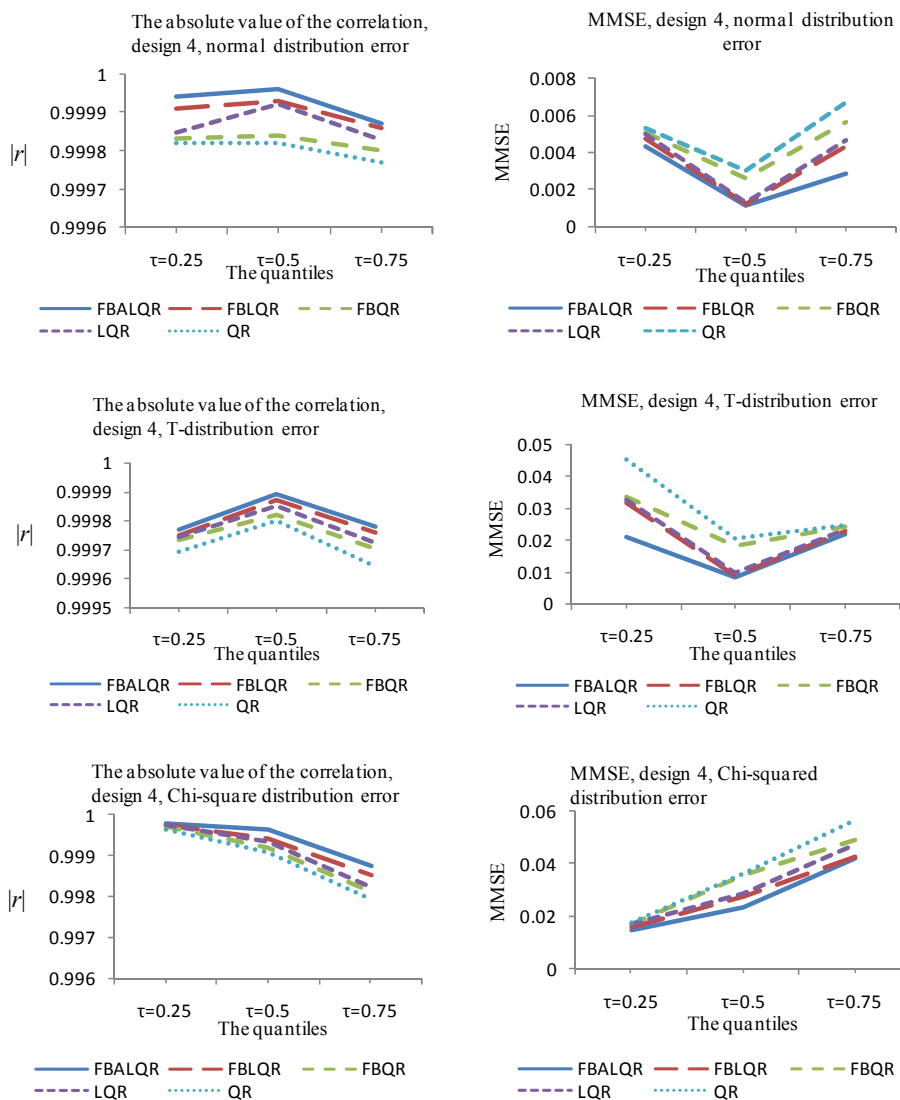
MMSE, design 1, Chi-squared error distribution

Figure 1. The left column explains the plots for the absolute value of the correlation between $X\hat{\beta}$ and $X\beta$, design 1, and the error distributions are normal, T and Chi-square respectively. The right column explains the plots for the MMSE of $X\hat{\beta}$, design 1, and the error distributions are normal, T and Chi-square respectively.

The absolute value of the correlation, design 2, normal distribution error

MMSE, design 2, normal distribution error

Figure 2. The left column explains the plots for the absolute value of the correlation between $X\hat{\beta}$ and β, model 2, and the error distributions are normal, T and Chi-square respectively. The right column explains the plots for the MMSE of $X\hat{\beta}$, model 2, and the error distributions are normal, T and Chi-square respectively.

Figure 3. The left column explains the plots for the absolute value of the correlation between $X\hat{\beta}$ and $X\beta$, model 3, and the error distributions are normal, T and Chi-square respectively. The right column explains the plots for the MMSE of $X\hat{\beta}$, model 3, and the error distributions are normal, T and Chi-square respectively.

Figure 4. The left column explains the plots for the absolute value of the correlation between $X\hat{\beta}$ and $X\beta$, model 4, and the error distributions are normal, T and Chi-square respectively. The right column explains the plots for the MMSE of $X\hat{\beta}$, model 4, and the error distributions are normal, T and Chi-square respectively.

The hip circumference (cm) x_7, the thigh circumference (cm) x_8, the knee circumference (cm) x_9, the ankle circumference (cm) x_{10}, the extended biceps circumference x_{11}, the forearm circumference (cm) x_{12} and the wrist circumference (cm) x_{13}.

Hoeting *et al*. [35] and Leng *et al*. [36] analyzed this data set. Previous diagnostic checking (Hoeting *et al*. [35]; Leng *et al*. [36]) showed that it is reasonable to assume a linear regression model.

The results of the body fat data analysis are presented in **Tables 5** and **6** and **Figures 5** and **6**. From **Table 5** and **Figure 5**, we have the following observations. According to the mean squared error criterion, it can be seen that the performance of the proposed methods (FBALQR and FBLQR) is better than the performance of other

methods. It is clear that the proposed methods give precise estimates. Again, we can see that when $\tau = 0.25$ and $\tau = 0.75$ the proposed methods are significantly more efficient than the other methods. The results of the simulation studies and real data example suggested that the proposed methods perform well.

6. Conclusion

In this study, we have proposed a flexible Bayesian Lasso and adaptive Lasso quantile regression by proposing a hierarchical model framework. We have theoretically investigated and numerically compared our proposed methods with the FBQR, Lasso quantile regression (LQR) and quantile regression (QR) methods. In order to

Table 5. MSE for $\hat{\beta}X$ which is estimated by FBALQR, FBLQR, FBQR, LQR and QR based on body fat data for $\tau = 0.25$, $\tau = 0.5$ and $\tau = 0.75$.

The methods	MSE		
	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
FBALQR	0.1084258	1.133324e–05	0.1370678
FBLQR	0.1241451	7.436673e–05	0.1423978
FBQR	0.1377442	9.326444e–05	0.1477935
LQR	0.1270168	0.001221321	0.1497241
QR	0.1414865	0.0009612835	0.1639030

Table 6. The estimated $\hat{\beta}$ which is estimated by FBALQR, FBLQR, FBQR, LQR and QR based on body fat data for $\tau = 0.25$, $\tau = 0.5$ and $\tau = 0.75$.

	$\tau = 0.25$					$\tau = 0.5$					$\tau = 0.75$				
	FBALQR	FBLQR	FBQR	LQR	QR	FBALQR	FBLQR	FBQR	LQR	QR	FBALQR	FBLQR	FBQR	LQR	QR
$\hat{\beta}_0$	–0.329	–0.352	–0.371	–0.356	–0.376	–0.003	–0.009	–0.009	–0.035	–0.031	0.370	0.377	0.384	0.387	0.405
$\hat{\beta}_1$	0.145	0.088	0.067	0.089	0.071	0.111	0.105	0.108	0.095	0.106	0.123	0.124	0.145	0.099	0.127
$\hat{\beta}_2$	–0.001	–0.058	–0.047	–0.324	–0.018	–0.119	–0.099	–0.105	–0.001	0.127	–0.097	–0.163	–0.120	–0.307	–0.726
$\hat{\beta}_3$	–0.060	–0.028	–0.027	–0.007	–0.028	–0.030	–0.039	–0.037	–0.039	–0.042	–0.056	–0.039	–0.043	0.010	0.091
$\hat{\beta}_4$	–0.129	–0.202	–0.225	–0.199	–0.245	–0.160	–0.155	–0.171	–0.110	–0.147	–0.082	–0.095	–0.094	–0.018	–0.062
$\hat{\beta}_5$	0.022	0.040	0.024	0.128	0.017	–0.037	–0.020	–0.046	–0.089	–0.110	–0.076	–0.104	–0.122	–0.112	0.034
$\hat{\beta}_6$	1.010	1.085	1.149	1.169	1.162	1.163	1.122	1.167	1.190	1.181	1.089	1.175	1.168	1.180	1.353
$\hat{\beta}_7$	–0.176	–0.237	–0.324	–0.236	–0.370	–0.211	–0.172	–0.235	–0.296	–0.348	–0.072	–0.101	–0.123	–0.039	–0.003
$\hat{\beta}_8$	0.140	0.159	0.189	0.233	0.208	0.165	0.136	0.170	0.224	0.212	0.078	0.089	0.132	0.074	0.129
$\hat{\beta}_9$	0.001	0.054	0.059	–0.008	0.072	–0.009	–0.004	0.004	–0.043	–0.057	–0.028	–0.045	–0.065	–0.066	–0.129
$\hat{\beta}_{10}$	–0.001	–0.076	–0.08	–0.025	–0.090	–0.003	–0.011	–0.004	–0.011	–0.011	0.054	0.077	0.085	0.110	0.122
$\hat{\beta}_{11}$	0.018	0.018	0.024	–0.001	0.019	0.052	0.051	0.062	0.016	0.034	0.082	0.106	0.101	0.043	0.141
$\hat{\beta}_{12}$	0.132	0.084	0.079	0.207	0.091	0.081	0.064	0.075	0.105	0.105	0.047	0.049	0.054	0.074	0.078
$\hat{\beta}_{13}$	–0.244	–0.137	–0.116	–0.183	–0.118	–0.174	–0.169	–0.172	–0.193	–0.210	–0.213	–0.215	–0.225	–0.216	–0.204

MSE for different methods, τ=0.25

MSE for different methods, τ=0.5
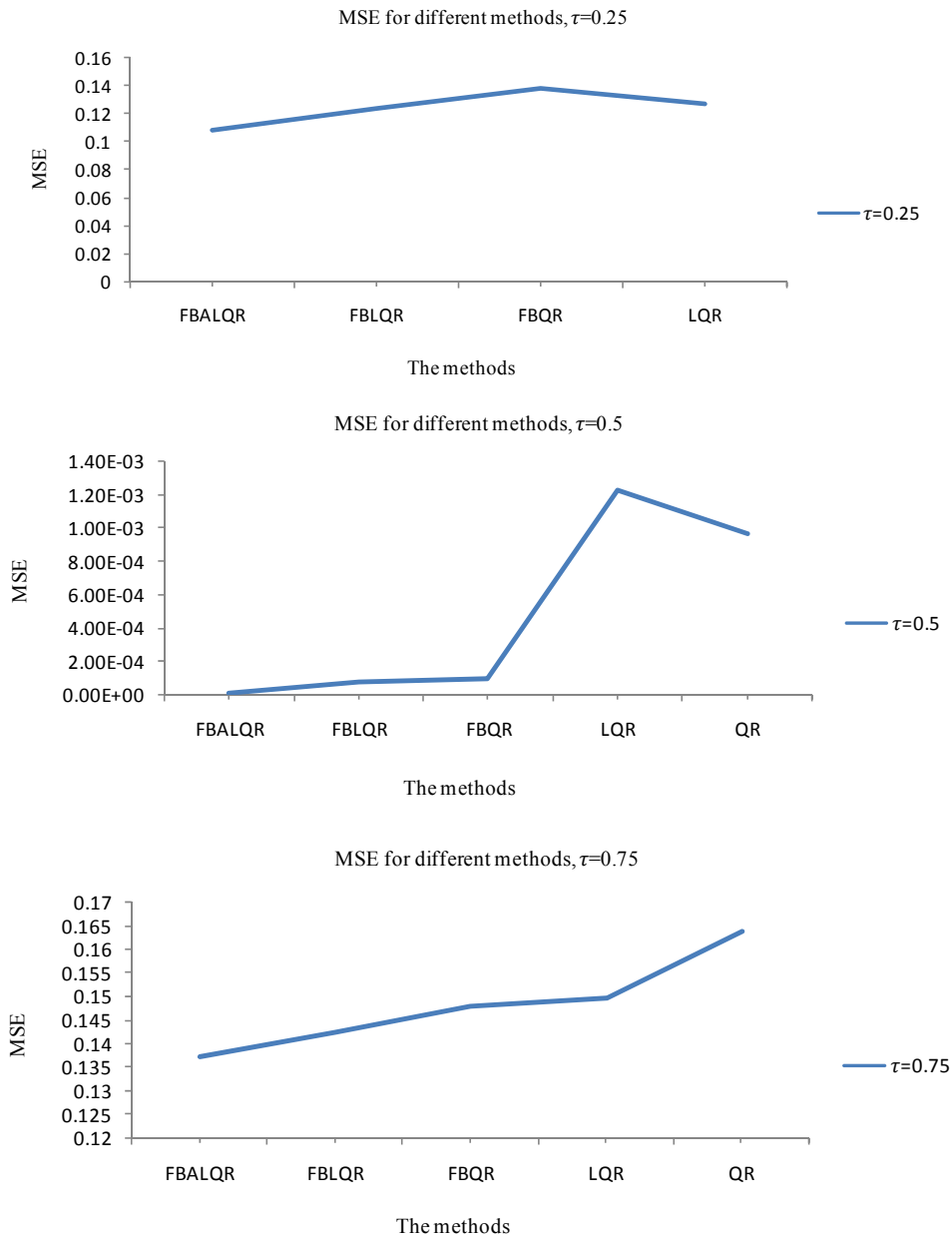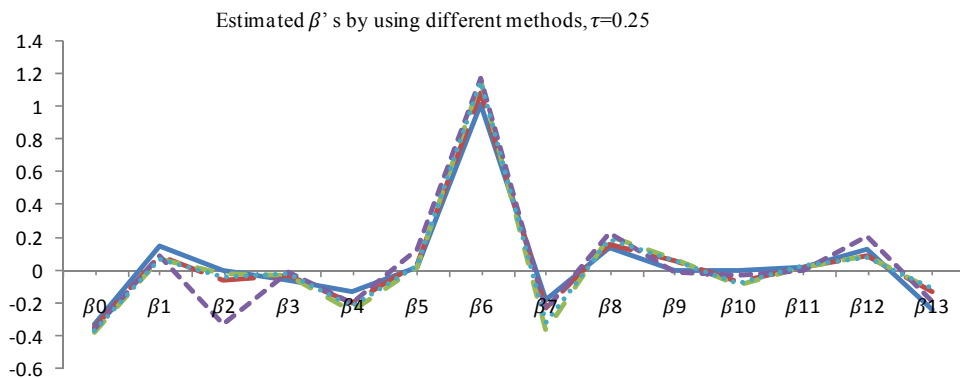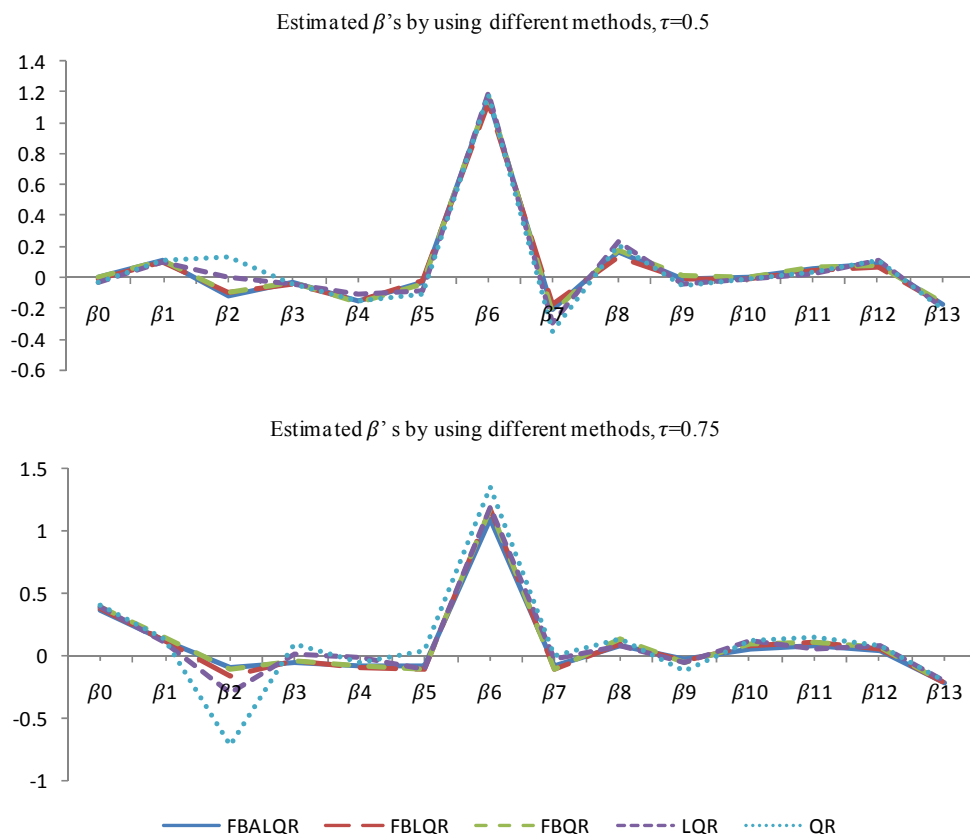
MSE for different methods, τ=0.75

Figure 5. plots explain MSE for $\hat{\beta}X$ which is estimated by FBALQR, FBLQR, FBQR, LQR and QR based on body fat data for $\tau = 0.25$, $\tau = 0.5$ and $\tau = 0.75$.

Estimated β's by using different methods, τ=0.25

Figure 6. Plots explain the estimated $\hat{\beta}$ which is estimated by FBALQR, FBLQR, FBQR, LQR and QR based on body fat data for $\tau = 0.25$, $\tau = 0.5$ and $\tau = 0.75$.

assess the numerical performance, we have conducted a simulation study based on the model $y_i = x_i'\beta + x_i'\gamma\varepsilon_i$ as described in Section 5 with error ε_i from three possible error distributions: standard normal, a $t_{(3)}$ distribution with three degrees of freedom and Chi-squared distribution with three degrees of freedom $\left(\chi^2_{(3)}\right)$ and five designs for the vector β. From the simulation study and real data, we can conclude that the proposed methods perform well in comparison to the other methods and thus we believe that the proposed methods are practically useful.

REFERENCES

[1] R. Koenker, "Quantile Regression," Cambridge University Press, Cambridge, 2005. doi:10.1017/CBO9780511754098

[2] K. Yu, Z. Lu and J. Stander, "Quantile Regression: Applications and Current Research Areas," *The Statistician*, Vol. 52, No. 3, 2003, pp. 331-350. doi:10.1111/1467-9884.00363

[3] R. Koenker and G. Bassett Jr., "Regression Quantiles," *Econometrica*, Vol. 46, No. 1, 1978, pp. 33-50. doi:10.2307/1913643

[4] R. Koenker and J. Machado, "Goodness of Fit and Re- lated Inference Processes for Quantile Regression," *Journal of the American Statistical Association*, Vol. 94, No. 448, 1999, pp. 1296-1310. doi:10.1080/01621459.1999.10473882

[5] K. Yu and R. A. Moyeed, "Bayesian Quantile Regres- sion," *Statistics and Probability Letters*, Vol. 54, No. 2, 2001, pp. 437-447. doi:10.1016/S0167-7152(01)00124-9

[6] E. Tsionas, "Bayesian Quantile Inference," *Journal of Statistical Computation and Simulation*, Vol. 73, No. 9, 2003, pp. 659-674. doi:10.1080/0094965031000064463

[7] K. Yu and J. Stander, "Bayesian Analysis of a Tobit Quantile Regression Model," *Journal of Econometrics*, Vol. 137, No. 1, 2007, pp. 260-276. doi:10.1016/j.jeconom.2005.10.002

[8] M. Geraci and M. Bottai, "Quantile Regression for Lon- gitudinal Data Using the Asymmetric Laplace Distribu- tion," *Biostatistics*, Vol. 8, No. 1, 2007, pp. 140-154. doi:10.1093/biostatistics/kxj039

[9] H. Kozumi and G. Kobayashi, "Gibbs Sampling Methods for Bayesian Quantile Regression," Technical Report, Kobe University, Kobe, 2009.

[10] C. Reed and K. Yu, "An Efficient Gibbs Sampler for Bayesian Quantile Regression," Technical Report, Brunel University, Uxbridge, 2009.

[11] D. F. Benoit and D. Van den Poel, "Binary Quantile Re- gression: A Bayesian Approach Based on the Asymmetric

　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　AM

Laplace Distribution," *Journal of Applied Econometrics*, Vol. 27, No. 7, 2012, pp. 1174-1188. doi:10.1002/jae.1216

[12] S. Walker and B. Mallick, "A Bayesian Semi-Parametric Accelerated Failure Time Model," *Biometrics*, Vol. 55, No. 2, 1999, pp. 477-483. doi:10.1111/j.0006-341X.1999.00477.x

[13] A. Kottas and A. Gelfand, "Bayesian Semi-Parametric Median Regression Modelling," *Journal of the American Statistical Association*, Vol. 96, No. 465, 2001, pp. 1458-1468. doi:10.1198/016214501753382363

[14] T. Hanson and W. O. Johnson, "Modeling Regression Error with a Mixture of Polya Trees," *Journal of the American Statistical Association*, Vol. 97, No. 460, 2002, pp. 1020-1033. doi:10.1198/016214502388618843

[15] N. L. Hjort, "Topics in Nonparametric Bayesian Statistics," In: P. J. Green, S. Richardson and N. L. Hjort, Eds., *Highly Structured Stochastic Systems*, Oxford University Press, Oxford, 2003, pp. 455-475.

[16] N. L. Hjort and S. Petrone, "Nonparametric Quantile Inference Using Dirichlet Processes," In: V. Nair, Ed., *Advances in Statistical Modeling and Inference*: *Essays in Honor of Kjell A. Doksum*, World Scientific, Singapore City, 2007, pp. 463-492. doi:10.1142/9789812708298_0023

[17] M. Taddy and A. Kottas, "A Nonparametric Model-Based Approach to Inference for Quantile Regression," Technical Report, UCSC Department of Applied Math and Statistics, 2007.

[18] A. Kottas and C. M. Krnjajić, "Bayesian Nonparametric Modelling in Quantile Regression," *Scandinavian Journal of Statistics*, Vol. 36, No. 3, 2009, pp. 297-319.

[19] B. Reich, H. Bondell and H. Wang, "Flexible Bayesian Quantile Regression for Independent and Clustered Data," *Biostatistics*, Vol. 11, No. 2, 2010, pp. 337-352. doi:10.1093/biostatistics/kxp049

[20] Q. Li, R. Xi and N. Lin, "Bayesian Regularized Quantile Regression," *Bayesian Analysis*, Vol. 5, No. 3, 2010, pp. 1-24. doi:10.1214/10-BA521

[21] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society Series B*, Vol. 58, No. 1, 1996, pp. 267-288.

[22] J. Fan and R. Li, "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties," *Journal of the American Statistical Association*, Vol. 96, No. 456, 2001, pp. 1348-1360. doi:10.1198/016214501753382273

[23] H. Zou, "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, Vol. 101,

No. 476, 2006, pp. 1418-1429. doi:10.1198/016214506000000735

[24] T. Park and G. Casella, "The Bayesian Lasso," *Journal of the American Statistical Association*, Vol. 103, No. 482, 2008, pp. 681-686. doi:10.1198/016214508000000337

[25] J. Bradic, J. Fan and W. Wang, "Penalized Composite Quasi-Likelihood for Ultrahigh-Dimensional Variable Selection," *Journal of Royal Statistics Society Series B*, Vol. 73, No. 3, 2010, pp. 325-349. doi:10.1111/j.1467-9868.2010.00764.x

[26] R. Koenker, "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, Vol. 91, No. 1, 2004, pp. 74-89. doi:10.1016/j.jmva.2004.05.006

[27] Y. Yuan and G. Yin, "Bayesian Quantile Regression for Longitudinal Studies with Non-Ignorable Missing Data," *Biometrics*, Vol. 66, No. 1, 2010, pp. 105-114. doi:10.1111/j.1541-0420.2009.01269.x

[28] H. Wang, G. Li and G. Jiang, "Robust Regression Shrinkage and Consistent Variable Selection through the LAD," *Journal of Business and Economic Statistics*, Vol. 25, No. 3, 2007, pp. 347-355. doi:10.1198/073500106000000251

[29] Y. Li and J. Zhu, "L_1-Norm Quantile Regressions," *Journal of Computational and Graphical Statistics*, Vol. 17, No. 1, 2008, pp. 1-23.

[30] Y. Wu and Y. Liu, "Variable Selection in Quantile Regression," *Statistica Sinica*, Vol. 19, No. 2, 2009, pp. 801-817.

[31] R. Alhamzawi, K. Yu and D. Benoit, "Bayesian Adaptive Lasso Quantile Regression," *Statistical Modelling*, Vol. 12, No. 3, 2012, pp. 279-297. doi:10.1177/1471082X1101200304

[32] R. W. Johnson, "Fitting Percentage of Body Fat to Simple Body Measurements," *Journal of Statistics Education*, Vol. 4, No. 1, 1996, pp. 236-237.

[33] X. He, "Quantile Curves without Crossing," *The American Statistician*, Vol. 51, No. 2, 1997, pp. 186-192. doi:10.1080/00031305.1997.10473959

[34] D. F. Andrews and C. L. Mallows, "Scale Mixtures of Normal Distributions," *Journal of Royal Statistics Society Series B*, Vol. 36, No. 1, 1974, pp. 99-102.

[35] J. A. Hoeting, D. Madigan, A. E. Raftery and C. T. Volinsky, "Bayesian Model Averaging: A Tutorial," *Statistical Science*, Vol. 14, No. 4, 1999, pp. 382-417.

[36] C. Leng, M. N. Tran and D. Nott, "Bayesian Adaptive Lasso," Technical Report, 2010. http://arxiv.org/PScache/arxiv/pdf/1009/1009.2300v1.pdf

Appendix

The details of Gibbs sampler for the penalized flexible Bayesian Lasso quantile regression are given as follows:

1) The full conditional distribution of β_j is multivariate normal distribution $\left(\bar{\beta}_j, \hat{\sigma}_j^2\right)$,

where $\hat{\sigma}_j^2 = \left(\sum_{i=1}^{n} x_{ij}^2 \left(x_i' \boldsymbol{\gamma} \sigma_{H_l G_l} \right)^{-2} + s_j^{-1} \right)^{-1}$, and

$$\bar{\beta}_j = \hat{\sigma}_j^2 \sum_{i=1}^{n} x_{ij} \left(x_i' \boldsymbol{\gamma} \sigma_{H_l G_l} \right)^{-2} \left(yi - x_i' \beta - x_i' \boldsymbol{\gamma} \mu_{H_i G_i} \right).$$

2) The full conditional distribution of s_j is inverse Gaussian (μ', λ'), $j = 1, \cdots, k$ where $\mu' = \sqrt{\beta_j^2 \lambda^2}$ and $\lambda' = \beta_j^2$.

3) The full conditional distribution of λ^2 is

$$\text{Gamma}\left(k + c_1, \sum_{j=1}^{k} s_j + c_2 \right).$$

Given $N = \max\{G_1, \cdots, G_n\}$ and $m = 1, \cdots, N$, the parameters $\mu_{1m}, \mu_{2m}, V_k, \sigma_{1m}, \sigma_{2m}$ and the standard deviation parameters can be updated using Gaussian distribution. The group indicators G_i are also updated using Metropolis-Hasting sampling (see Reich *et al.* [19] for more details). The full conditional distribution for all parameters in the penalized flexible Bayesian Adaptive Lasso quantile regression is similar to the above description except for the full conditional distributions for s_j and λ_j^2, $j = 1, \cdots, k$ which are given by

1) The full conditional distribution of s_j is inverse Gaussian (μ', λ'), $j = 1, \cdots, k$ where

$\mu' = \sqrt{\beta_j^2 \lambda_j^2}$ and $\lambda' = \beta_j^2$.

2) The full conditional distribution of λ_j^2 is Gamma $\left(1 + c_1, s_j + c_2\right)$.