

Two-Stage Penalized Composite Quantile Regression with Grouped Variables

Sungwan Bang^a, Myoungshic Jhun^{1,b}

^aDepartment of Mathematics, Korea Military Academy

^bDepartment of Statistics, Korea University

Abstract

This paper considers a penalized composite quantile regression (CQR) that performs a variable selection in the linear model with grouped variables. An adaptive sup-norm penalized CQR (ASCQR) is proposed to select variables in a grouped manner; in addition, the consistency and oracle property of the resulting estimator are also derived under some regularity conditions. To improve the efficiency of estimation and variable selection, this paper suggests the two-stage penalized CQR (TSCQR), which uses the ASCQR to select relevant groups in the first stage and the adaptive lasso penalized CQR to select important variables in the second stage. Simulation studies are conducted to illustrate the finite sample performance of the proposed methods.

Keywords: Composite quantile regression, factor selection, penalization, sup-norm, variable selection.

1. Introduction

Consider the estimation of the unknown regression coefficient vector $\boldsymbol{\beta}^* = (\beta_1^*, \dots, \beta_p^*)^T$ in the linear regression model

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta}^* + \varepsilon_i, \quad i = 1, \dots, n, \quad (1.1)$$

where y_i is the response of interest, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ is a set of predictor variables, and $\{\varepsilon_i\}$ are unobservable, independent, and identically distributed random errors from a distribution F with mean zero. Without a loss of generality, we center the response and each predictor variable. The least squares regression (LSR) is a popular method to estimate $\boldsymbol{\beta}^*$ with its mathematical beauty and its estimator is of full efficiency when the error term follows a normal distribution. However, it is well known that the traditional ordinary least squares may fail to produce a reliable estimator for data subject to heavy-tailed errors or outliers. As an alternative to the LSR method, Koenker (1984) proposed the composite quantile regression (CQR) to improve the efficiency of estimation. By using equal weights for the K different quantiles, the CQR method estimates $\boldsymbol{\beta}^*$ by

$$\left(\hat{b}_{\tau_1}, \dots, \hat{b}_{\tau_K}, \hat{\boldsymbol{\beta}}^{CQR} \right) = \arg \min_{b_1, \dots, b_K, \boldsymbol{\beta}} \sum_{k=1}^K \sum_{i=1}^n \rho_{\tau_k} (y_i - b_k - \mathbf{x}_i^T \boldsymbol{\beta}), \quad (1.2)$$

Bang's research was supported by 2013 Korea Military Academy's research funds (20130512). Jhun's research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0009204).

¹ Corresponding author: Professor, Department of Statistics, Korea University, Seoul 136-701, Republic of Korea.
E-mail: jhun@korea.ac.kr

where the loss function $\rho_{\tau_k}(\cdot)$, called the check function (Koenker and Bassett, 1978), is defined by $\rho_{\tau_k}(t) = t(\tau_k - I(t < 0))$ and $\{\tau_k, k = 1, \dots, K\}$ is a given sequence of quantiles. Note that the regression coefficients in the linear model (1.1) are identical for the K distinct quantile regressions. Therefore, the CQR method has the capability to share the strength among multiple conditional quantiles.

In a regression problem, variable selection plays an important role in model building because the irrelevant predictors selected in the final model make it difficult to interpret the resultant model and may decrease its predictive ability. In the context of mean regression, several penalization methods have been developed to automate variable selection that include the lasso (Tibshirani, 1996), the SCAD (Fan and Li, 2001), and the adaptive lasso (Zou, 2006); under the proper regularity conditions, both the SCAD and the adaptive lasso estimators possess the oracle property to work as well as if the correct submodel were known. By adopting the idea of the adaptive lasso, Zou and Yuan (2008a) suggested an adaptive lasso penalized CQR (ALCQR) method, and established the oracle property of its estimator $\hat{\boldsymbol{\beta}}^{ALCQR}$ obtained from the following model fitting:

$$\left(\hat{b}_{\tau_1}, \dots, \hat{b}_{\tau_K}, \hat{\boldsymbol{\beta}}^{ALCQR}\right) = \arg \min_{b_1, \dots, b_K, \boldsymbol{\beta}} \sum_{k=1}^K \sum_{i=1}^n \rho_{\tau_k}(y_i - b_k - \mathbf{x}_i^T \boldsymbol{\beta}) + \lambda \sum_{j=1}^p \frac{1}{|\hat{\beta}_j^{CQR}|^\gamma} |\beta_j|, \quad (1.3)$$

where $\hat{\beta}_j^{CQR}$ ($j = 1, \dots, p$) is the unpenalized CQR estimator and $\gamma > 0$ is some pre-specified positive number. When $K = 1$, the ALCQR reduces to the adaptive lasso penalized quantile regression (Wu and Liu, 2009).

As described in (1.3), the ALCQR method was invented to automate individual variable selection during the CQR model fitting. In many regression problems, however, it is often more meaningful to identify significant factors rather than individual predictors, where each factor is represented by a group of predictor variables (Yuan and Lin, 2006). In practice, such a problem arises naturally when a set of dummy variables are used to represent a categorical variable and/or when a set of basis functions of a continuous variable is included in the predictor set. In these situations, it is desirable to select predictor variables in a grouped manner. To satisfy this purpose to estimate the regression coefficients, we first propose the adaptive sup-norm penalized CQR (ASCQR) method that penalizes the CQR loss function by the sum of factor-wise adaptive sup-norm penalties. The proposed ASCQR method is a generalized form of ALCQR and possesses the oracle property under some mild regularity conditions as the ALCQR dose. To improve the efficiency of estimation and variable selection, furthermore, we suggest a two-stage penalized CQR (TSCQR) method that uses the ASCQR method to select relevant groups in the first stage and the ALCQR method to select important predictor variables in the second stage.

The rest of this paper is organized as follows. Section 2 introduces the ASCQR method and establishes the oracle property of its estimator. Section 3 focuses on an explanation of the TSCQR method. Simulation results are reported in Section 4. Section 5 contains a few concluding remarks and the technical proofs of theorems are presented in Section 6.

2. Adaptive Sup-Norm Penalized Composite Quantile Regression (ASCQR)

2.1. The ASCQR estimator

Suppose that in the regression model (1.1), the predictors $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$ are grouped into G factors as $(\mathbf{x}_{i(1)}^T, \dots, \mathbf{x}_{i(G)}^T)$, where $\mathbf{x}_{i(j)}^T = (x_{i,j1}, \dots, x_{i,jp_j})$ is a group of p_j predictor variables for $j = 1, \dots, G$

and $\sum_{j=1}^G p_j = p$. Then the regression model (1.1) can be represented by

$$y_i = \sum_{j=1}^G \mathbf{x}_{i,(j)}^T \boldsymbol{\beta}_{(j)}^* + \varepsilon_i, \quad i = 1, \dots, n, \quad (2.1)$$

where $\boldsymbol{\beta}_{(j)}^* = (\beta_{j1}^*, \dots, \beta_{jp_j}^*)^T \in R^{p_j}$ is the regression coefficient vector associated with the j^{th} factor. In this paper, we use the terms ‘‘factor’’ and ‘‘group’’ interchangeably to indicate the grouping of variables. For example, when we want to produce more flexible functions other than linear models and/or include categorical variables, we add new variables generated from original ones using appropriate transformation such as polynomials or dummy variables. In such a situation, it is more reasonable to automatically and simultaneously select significant factors, rather than individual derived variables.

In the context of mean regression, several penalization methods have been developed for simultaneous estimation of regression coefficients and factor selection that include the group lasso (Yuan and Lin, 2006), the adaptive group lasso (Wang and Leng, 2008), and the penalized LSR using composite absolute penalties (CAP) (Zhao *et al.*, 2009). However, these methods would be extremely sensitive to outlying observations and inefficient for heavy tailed error distributions due to the property of the quadratic loss function. To remedy such a limitation, we suggest using the CQR loss in (1.2) as an alternative to the quadratic loss, and propose the following adaptive sup-norm penalized CQR (ASCQR) model fitting:

$$\left(\hat{b}_{\tau_1}, \dots, \hat{b}_{\tau_K}, \hat{\boldsymbol{\beta}}^{ASCQR} \right) = \arg \min_{b_1, \dots, b_K, \boldsymbol{\beta}} \sum_{k=1}^K \sum_{i=1}^n \rho_{\tau_k} \left(y_i - b_k - \sum_{j=1}^G \mathbf{x}_{i,(j)}^T \boldsymbol{\beta}_{(j)} \right) + n\lambda \sum_{j=1}^G \frac{1}{\left\| \hat{\boldsymbol{\beta}}_{(j)}^{CQR} \right\|_{\infty}^{\gamma}} \left\| \boldsymbol{\beta}_{(j)} \right\|_{\infty}, \quad (2.2)$$

where the sup-norm penalty is defined by

$$\left\| \boldsymbol{\beta}_{(j)} \right\|_{\infty} = \max \left\{ \left| \beta_{j1} \right|, \dots, \left| \beta_{jp_j} \right| \right\}, \quad j = 1, \dots, G, \quad (2.3)$$

$\hat{\boldsymbol{\beta}}_{(j)}^{CQR}$ is the unpenalized CQR estimator associated with the j^{th} factor, and $\gamma > 0$ is some pre-specified number ($\gamma = 1$ was used for our simulation study). As suggested in Zou and Yuan (2008a), we also use the equally spaced quantiles

$$\tau_k = \frac{k}{K+1}, \quad k = 1, \dots, K. \quad (2.4)$$

When each individual predictor is considered as a factor, that is, $p_1 = \dots = p_G = 1$, the ASCQR reduces to the ALCQR described in (1.3). Therefore, the ASCQR method is a generalization of the ALCQR; the ASCQR method has the capability to automate factor selection in the model fitting, whereas the ALCQR contain no information on the factors. Instead of the sup-norm, we can also employ the L_2 -norm $\| \cdot \|_2$ for the penalty term in (2.2). In this paper, we focus on the sup-norm penalty because its use makes the computation efficient in the model fitting. With its computational efficiency, the sup-norm penalty (2.3) has also been considered for the simultaneous factor selection in the support vector machine (Zou and Yuan, 2008b) and in the quantile regression (Bang and Jhun, 2012).

2.2. Theoretical properties

In this section, we establish the oracle property of the ASCQR estimator. Without loss of generality, we assume that only the first $g \leq G$ factors are relevant and that $\sum_{j=1}^g p_j = p_0$. That is,

$$A = \{j : \boldsymbol{\beta}_{(j)}^* \neq \mathbf{0}\} = \{1, 2, \dots, g\} \quad \text{and} \quad A^c = \{j : \boldsymbol{\beta}_{(j)}^* = \mathbf{0}\} = \{g+1, \dots, G\}. \quad (2.5)$$

It is further assumed that all predictor variables associated with each relevant factor have nonzero coefficients, which implies

$$\beta_{jl} \neq 0, \quad \text{for } j = 1, \dots, g \text{ and } l = 1, \dots, p_j. \quad (2.6)$$

For convenience, let $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_A^{*T}, \boldsymbol{\beta}_{A^c}^{*T})^T$, where $\boldsymbol{\beta}_A^* = (\boldsymbol{\beta}_{(1)}^{*T}, \dots, \boldsymbol{\beta}_{(g)}^{*T})^T$ and $\boldsymbol{\beta}_{A^c}^* = (\boldsymbol{\beta}_{(g+1)}^{*T}, \dots, \boldsymbol{\beta}_{(G)}^{*T})^T$. For our theoretical results, we assume the following regularity conditions:

Condition 1. $\{\varepsilon_i\}$ are i.i.d. random errors with the τ_k^{th} quantile $b_{\tau_k}^* = F^{-1}(\tau_k)$ and a continuous, positive density $f(\cdot)$ in a neighborhood of $b_{\tau_k}^*$ for $k = 1, \dots, K$.

Condition 2. There exists a $p \times p$ positive definite matrix Σ such that $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T = \Sigma$.

Note that Condition 1 and Condition 2 are basically the same conditions for establishing the asymptotic normality of a single quantile regression estimator (Koenker, 2005). Under these regularity conditions, we have the following two theorems.

Theorem 1. (Root- n consistency) Assume that the regularity Condition 1 and Condition 2 hold. If $\sqrt{n}\lambda = O_p(1)$, then $\|\hat{\boldsymbol{\beta}}_{(j)}^{\text{ASCQR}} - \boldsymbol{\beta}^*\| = O_p(n^{-1/2})$.

Theorem 2. (Oracle property) Assume that the regularity Condition 1 and Condition 2 hold. If $\sqrt{n}\lambda \rightarrow 0$ and $n^{(\gamma+1)/2}\lambda \rightarrow \infty$, then the ASCQR estimator must satisfy:

(a) Sparsity: $P(\hat{\boldsymbol{\beta}}_{A^c}^{\text{ASCQR}} = \mathbf{0}) \rightarrow 1$.

(b) Asymptotic normality: $\sqrt{n}(\hat{\boldsymbol{\beta}}_{A^c}^{\text{ASCQR}} - \boldsymbol{\beta}_A^*) \rightarrow_d N(\mathbf{0}, \sigma_{\text{CQR}}^2 \Sigma_{AA}^{-1})$,

where Σ_{AA} is the top-left $p_0 \times p_0$ submatrix of Σ and $\sigma_{\text{CQR}}^2 = (\sum_{k=1}^K f(b_{\tau_k}^*))^{-2} (\sum_{k,k'=1}^K (\tau_k \wedge \tau_{k'} - \tau_k \tau_{k'}))$.

If a tuning parameter λ is appropriately chosen, then Theorem 2 implies that the ASCQR estimator is as efficient as oracle. The proofs of Theorem 1 and Theorem 2 given in Section 6 are similar to the findings of prior research (Koenker, 1984; Zou and Yuan, 2008a; Bang and Jhun, 2012).

2.3. Computing algorithm

The optimization problem of the ASCQR method can be formulated as a linear programming (LP) problem. To derive the LP formulation, we introduce $(2n)K$ slack variables $\{(u_{ik}, v_{ik}), i = 1, \dots, n, k = 1, \dots, K\}$ that satisfy the equality constraints $y_i - b_k - \sum_{j=1}^G \mathbf{x}_{i,(j)}^T \boldsymbol{\beta}_{(j)} = u_{ik} - v_{ik}$, where $u_{ik} \geq 0$ and $v_{ik} \geq 0$ for $i = 1, \dots, n$ and $k = 1, \dots, K$. The slack variables u_{ik} and v_{ik} represent the positive and negative parts of the residuals for the i^{th} observation of the τ_k^{th} regression quantile, respectively. Let the new variable $M_j = \|\boldsymbol{\beta}_{(j)}\|_\infty$ for $j = 1, \dots, G$ and write b_k as $b_k^+ - b_k^-$ and $\boldsymbol{\beta}_{(j)}$ as $\boldsymbol{\beta}_{(j)}^+ - \boldsymbol{\beta}_{(j)}^-$, where

$b_k^+ \geq 0, b_k^- \geq 0, \boldsymbol{\beta}_{(j)}^+ = (\beta_{j1}^+, \dots, \beta_{jp_j}^+)^T \geq \mathbf{0}$, and $\boldsymbol{\beta}_{(j)}^- = (\beta_{j1}^-, \dots, \beta_{jp_j}^-)^T \geq \mathbf{0}$. Then the adaptive sup-norm penalty term can be formulated linearly with some linear inequality constraints. By using these variables and notations, the ASCQR estimator in (2.2) can be obtained by minimizing

$$\sum_{k=1}^K \sum_{i=1}^n (\tau_k u_{ik} + (1 - \tau_k) v_{ik}) + n\lambda \sum_{j=1}^G \|\boldsymbol{\beta}_{(j)}^{CQR}\|_{\infty}^{-\gamma} M_j, \quad (2.7)$$

subject to

$$b_k^+ - b_k^- + \sum_{j=1}^G \mathbf{x}_{i,(j)}^T (\boldsymbol{\beta}_{(j)}^+ - \boldsymbol{\beta}_{(j)}^-) + u_{ik} - v_{ik} = y_i,$$

$$M_j \geq \beta_{jl}^+ + \beta_{jl}^-,$$

$$u_{ik} \geq 0, v_{ik} \geq 0, b_k^+ \geq 0, b_k^- \geq 0, \beta_{jl}^+ \geq 0, \beta_{jl}^- \geq 0, \text{ for all } 1 \leq i \leq n, 1 \leq j \leq p, 1 \leq k \leq K, 1 \leq l \leq p_j.$$

In this study, we used the lpSolve package provided in R to implement the above LP problem (R code is available for interested readers upon request).

3. Two-Stage Penalized Composite Quantile Regression (TSCQR)

To make the oracle property of the ASCQR estimator hold, it must be satisfied that all the predictor variables within each relevant group have nonzero coefficients as described in (2.6). However, it is common in practice that not all the predictors are relevant to the response within one group. The sup-norm penalty in (2.3) tends to select either all variables or no variable in one group. When one predictor variable in a group is selected, all other variables in the same group are also selected. Thus, the ASCQR method does not perform a variable selection within an identified group and there is no guarantee that the ASCQR estimator is as efficient as the oracle estimator in such situations. Nonetheless, the proposed ASCQR method can consistently identify irrelevant groups as described in Theorem 2(a) even when the assumption (2.6) is violated. In these regards, we suggest a two-stage penalized CQR (TSCQR) method that consists of a group selection stage and a variable selection stage; irrelevant groups are excluded in the first stage by using the ASCQR, and then in the second stage, the ALCQR is used to perform individual variable selection with retained variables.

Suppose that we first fit the ASCQR estimator by using all the groups of predictor variables. Let \hat{A} and \hat{A}^c be the estimated set of indices for the relevant groups and the irrelevant groups, respectively, which implies

$$\hat{A} = \left\{ j : \hat{\boldsymbol{\beta}}_{(j)}^{ASCQR} \neq \mathbf{0} \right\} \quad \text{and} \quad \hat{A}^c = \left\{ j : \hat{\boldsymbol{\beta}}_{(j)}^{ASCQR} = \mathbf{0} \right\}. \quad (3.1)$$

We use $\boldsymbol{\beta}_{\hat{A}}^*$ and $\boldsymbol{\beta}_{\hat{A}^c}^*$ to denote true regression coefficient vectors corresponding to the set \hat{A} and \hat{A}^c , respectively, and write the TSCQR estimators of $\boldsymbol{\beta}_{\hat{A}}^*$ and $\boldsymbol{\beta}_{\hat{A}^c}^*$ as $\hat{\boldsymbol{\beta}}_{\hat{A}}^{TSCQR}$ and $\hat{\boldsymbol{\beta}}_{\hat{A}^c}^{TSCQR}$, respectively. Based on the theoretical properties of the ASCQR estimator in Theorem 2(a), the estimated irrelevant groups corresponding to the set \hat{A}^c could be excluded in the first stage, that is, we estimate $\boldsymbol{\beta}_{\hat{A}^c}^*$ by $\hat{\boldsymbol{\beta}}_{\hat{A}^c}^{TSCQR} = \mathbf{0}$. Then in the second stage, we estimate the regression coefficients of retained variables $\boldsymbol{\beta}_{\hat{A}}^*$ by

$$\left(\hat{b}_{\tau_1}, \dots, \hat{b}_{\tau_K}, \hat{\boldsymbol{\beta}}_{\hat{A}}^{TSCQR} \right) = \arg \min_{b_1, \dots, b_K, \boldsymbol{\beta}_{\hat{A}}} \sum_{k=1}^K \sum_{i=1}^n \rho_{\tau_k} \left(y_i - b_k - \sum_{j \in \hat{A}} \mathbf{x}_{i,(j)}^T \boldsymbol{\beta}_{(j)} \right) + n\lambda \sum_{j \in \hat{A}} \sum_{l=1}^{p_j} \frac{1}{|\hat{\boldsymbol{\beta}}_{\hat{A}}^{ASCQR}|^{\gamma}} |\beta_{jl}|, \quad (3.2)$$

where β_{jl}^{ASCQR} is the ASCQR estimator associated with the l^{th} variable in the j^{th} group.

4. Simulation Study

In this section, simulation studies were conducted to evaluate the finite sample performance of the proposed ASCQR and TSCQR methods with the number of quantile $K = 5$. For comparison purpose, we also included the ALCQR method in the simulation. The two simulated examples were considered are similar to those in Yuan and Lin (2006) and Bang and Jhun (2012). For each of the simulated examples, five different error distributions were considered ranging from symmetric to asymmetric ones: normal $N(0, 1)$, t distribution with degree of freedoms 3 $T(3)$, a mixture of normals (MN) $0.1N(0, 25) + 0.9N(0, 1)$, lognormal $LN(0, 1)$, and exponential distribution $EXP(1)$. To maintain a fair comparison and to satisfy the assumption of mean zero error terms, we first standardized the error distributions to have a mean zero and variance one. In our simulated examples, we generated $n = 100$ training observations, along with 1,000 independent validation observations. The models were fitted on training data only and the validation set was used to select the tuning parameter λ . To evaluate the prediction accuracy for each method, the model error is computed by

$$ME = (\hat{\beta} - \beta^*)^T E(\mathbf{X}^T \mathbf{X}) (\hat{\beta} - \beta^*), \quad (4.1)$$

where \mathbf{X} is a design matrix. We used the model error of the LSR and CQR oracles as the benchmark. The model selection performance was measured by the number of correctly selected factors and variables (NC), the number of incorrectly selected factors and variables (NIC), and the number of times that the true model is correctly identified. To assess the sampling variability, this procedure was repeated 100 times independently. In the tables, the model error, NC and NIC, were reported on averages of over 100 runs. The numbers given in parentheses are the standard deviations of the model errors.

4.1. Example 1

This example concerns an additive model with continuous factors. Ten random variables Z_1, \dots, Z_9 and W were independently generated from a standard normal distribution. Then we considered the following model

$$Y = X_1 + X_1^2 + X_1^3 + \frac{2}{3}X_3 - X_3^2 + \frac{1}{3}X_3^3 + \varepsilon, \quad (4.2)$$

where the factors are defined as $X_j = (Z_j + W)/\sqrt{2}$ for $j = 1, \dots, 9$. This model has 9 factors (groups of variables), and each factor is represented by a third-order polynomial. Thus this model has 27 predictor variables and 9 groups. The 2 factors associated with X_1 and X_3 have nonzero coefficients. Table 1 summarizes the performance of each method for the above simulated model (4.2) in terms of model error and model selection. As expected, the LSR oracle was the best when the random error follows a normal distribution. However, for the non-normal distributions, the CQR oracle outperformed the LSR oracle in terms of model error. These results are consistent with the findings of Koenker (1984) and Zou and Yuan (2008a). The proposed ASCQR were superior to the ALCQR for each error distribution in terms of model error and factor selection performance. The ASCQR included a larger number of predictor variables than ALCQR due to its block-inclusion nature; however, we can see that ASCQR identified the relevant groups efficiently by retaining all important variables. The proposed TSCQR could produce more accurate and sparser model than the others because of the characteristics of ASCQR.

Table 1: Simulation results for Example 1

Error Dist.	Method	Mean absolute error	No. of factors selected		No. of variables selected		No. of times for true model
			NC	NIC	NC	NIC	
$N(0, 1)$	LSR-oracle	0.0920 (0.0624)	2	0	6	0	100
	CQR-oracle	0.1109 (0.0836)	2	0	6	0	100
	ALCQR	0.1718 (0.1307)	2	2.15	5.89	2.33	18
	ASCQR	0.1117 (0.0846)	2	1.84	6	5.52	23
	TSCQR	0.1071 (0.0748)	2	0.44	6	0.47	65
MN	LSR-oracle	0.0831 (0.0756)	2	0	6	0	100
	CQR-oracle	0.0529 (0.0550)	2	0	6	0	100
	ALCQR	0.0728 (0.0724)	2	1.76	6	1.99	18
	ASCQR	0.0509 (0.0490)	2	1.46	6	4.38	27
	TSCQR	0.0542 (0.0524)	2	0.25	6	0.27	78
$T(3)$	LSR-oracle	0.0683 (0.0656)	2	0	6	0	100
	CQR-oracle	0.0534 (0.0561)	2	0	6	0	100
	ALCQR	0.0810 (0.1055)	2	1.70	5.90	1.85	12
	ASCQR	0.0573 (0.0519)	2	1.66	6	4.98	15
	TSCQR	0.0554 (0.0583)	2	0.31	6	0.34	77
$LN(0, 1)$	LSR-oracle	0.0798 (0.0833)	2	0	6	0	100
	CQR-oracle	0.0228 (0.0231)	2	0	6	0	100
	ALCQR	0.0314 (0.0285)	2	1.20	6	1.31	40
	ASCQR	0.0236 (0.0243)	2	1.24	6	3.72	46
	TSCQR	0.0239 (0.0229)	2	0.11	6	0.12	91
EXP(1)	LSR-oracle	0.0879 (0.0776)	2	0	6	0	100
	CQR-oracle	0.0544 (0.0561)	2	0	6	0	100
	ALCQR	0.0878 (0.0775)	2	1.89	5.95	2.12	13
	ASCQR	0.0688 (0.0588)	2	1.51	6	4.53	26
	TSCQR	0.0578 (0.0479)	2	0.29	6	0.32	75

The numbers in parentheses are standard deviations.

4.2. Example 2

In this example, 10 factors $(X_1, \dots, X_{10})^T$ were generated from a multivariate normal distribution $N(\mathbf{0}, \Sigma)$, whose covariance matrix $\Sigma_{ij} = 0.5^{|i-j|}$ for $0 \leq i, j \leq 10$. Then the last 5 factors $(X_6, \dots, X_{10})^T$ were trichotomized as 0, 1, or 2 if it was smaller than $\Phi^{-1}(1/3)$, larger than $\Phi^{-1}(2/3)$, or somewhere in between. The true regression model is given by

$$Y = X_1 + X_1^2 + \frac{2}{3}X_3 + \frac{1}{3}X_3^3 + 2I(X_6 = 0) + I(X_6 = 1) + \varepsilon. \quad (4.3)$$

This model is an additive model involving 5 continuous factors and 5 categorical factors, where each continuous factor is represented by a third-order polynomial and each categorical factor with 3 levels is represented by 2 dummy variables; subsequently, this model has 25 predictor variables and 10 groups. The two groups associated with continuous factors X_1 and X_3 are related to the response, but not all the predictor variables within those groups are important, such as the coefficients of the predictor variables X_1^3 and X_3^2 are zero. Table 2 reports the performance of each method in terms of model error and model selection; in addition, the conclusions are similar to those of Example 1. ASCQR did not perform as well as the oracle since the simulated model (4.3) violated the assumption (2.6). However, we can see that by using ASCQR for the group selection in the first stage, TSCQR provided the efficient performance in terms of the model error as well as model selection.

5. Concluding Remarks

In this paper, we have developed the adaptive sup-norm penalized CQR (ASCQR) method, which

Table 2: Simulation results for Example 2

Error Dist.	Method	Mean absolute error	No. of factors selected		No. of variables selected		No. of times for true model
			NC	NIC	NC	NIC	
N(0, 1)	LSR-oracle	0.0708 (0.0486)	3	0	6	0	100
	CQR-oracle	0.0968 (0.0774)	3	0	6	0	100
	ALCQR	0.2094 (0.1681)	3	2.07	5.94	2.54	3
	ASCQR	0.1980 (0.1353)	3	2.09	6	7.69	0
	TSCQR	0.1184 (0.0881)	3	0.59	6	1.14	24
MN	LSR-oracle	0.0791 (0.0672)	3	0	6	0	100
	CQR-oracle	0.0547 (0.0506)	3	0	6	0	100
	ALCQR	0.0779 (0.0743)	3	1.59	6	1.92	14
	ASCQR	0.0828 (0.0609)	3	2.04	6	7.62	0
	TSCQR	0.0552 (0.0525)	3	0.45	6	1.06	40
T(3)	LSR-oracle	0.0616 (0.0462)	3	0	6	0	100
	CQR-oracle	0.0528 (0.0391)	3	0	6	0	100
	ALCQR	0.0849 (0.0593)	3	1.74	5.89	2.15	15
	ASCQR	0.1030 (0.0797)	3	1.69	6	6.74	0
	TSCQR	0.0588 (0.0453)	3	0.45	6	0.98	34
LN(0, 1)	LSR-oracle	0.0693 (0.0720)	3	0	6	0	100
	CQR-oracle	0.0186 (0.0138)	3	0	6	0	100
	ALCQR	0.0276 (0.0273)	3	1.11	5.92	1.30	29
	ASCQR	0.0401 (0.0350)	3	1.69	6	6.64	0
	TSCQR	0.0200 (0.0185)	3	0.22	6	0.52	63
EXP(1)	LSR-oracle	0.0685 (0.0591)	3	0	6	0	100
	CQR-oracle	0.0524 (0.0603)	3	0	6	0	100
	ALCQR	0.0916 (0.1052)	3	1.68	6	2.08	16
	ASCQR	0.0995 (0.0636)	3	1.74	6	6.75	0
	TSCQR	0.0563 (0.0627)	3	0.32	6	0.72	46

The numbers in parentheses are standard deviations.

performs variable selection in a grouped manner. We established the consistency and oracle property of the proposed ASCQR estimator. Furthermore, we have proposed a two-stage penalized CQR (TSCQR) method to improve the efficiency of estimation and variable selection. The simulation results showed that the proposed methods result in a higher efficiency than the ALCQR does, especially when the assumption (2.6) is satisfied. In this paper, we used the equal weights $w_1 = \dots = w_K = 1$ in the following CQR loss function:

$$\sum_{k=1}^K w_k \sum_{i=1}^n \rho_{\tau_k} \left(y_i - b_k - \sum_{j=1}^G \mathbf{x}_{i,(j)}^T \boldsymbol{\beta}_{(j)} \right). \quad (5.1)$$

Koenker (1984) discussed the optimal values of weights w_k ($k = 1, \dots, K$) theoretically. Bradic *et al.* (2011) suggested a data-driven method for estimating the optimal weights and showed that the efficiency of the CQR using the equal weights can be significantly improved by employing proper weights. In this regard, we expect that the proposed methods can provide a more efficient estimator if a reliable estimation method of the optimal weights is developed in future research.

6. Proofs of the Theorems

We now provide the proofs of Theorem 1 and Theorem 2.

Proof of Theorem 1:

Let $Q(\mathbf{b}_\tau, \boldsymbol{\beta}) = \sum_{k=1}^K \sum_{i=1}^n \rho_{\tau_k}(y_i - b_{\tau_k} - \mathbf{x}'_i \boldsymbol{\beta}) + n\lambda \sum_{j=1}^G \|\hat{\boldsymbol{\beta}}_{(j)}^{CQR}\|_\infty^{-\gamma} \|\boldsymbol{\beta}_{(j)}\|_\infty$ and $L_n(\mathbf{v}, \mathbf{u}) = Q(\mathbf{b}_\tau^* + n^{-1/2}\mathbf{v}, \boldsymbol{\beta}^* + n^{-1/2}\mathbf{u}) - Q(\mathbf{b}_\tau^*, \boldsymbol{\beta}^*)$, where $\mathbf{b}_\tau = (b_{\tau_1}, \dots, b_{\tau_K})' \in R^K$, $\mathbf{v} = (v_1, \dots, v_K)' \in R^K$, and $\mathbf{u} =$

$(\mathbf{u}'_{(1)}, \dots, \mathbf{u}'_{(G)})' \in R^p$ with $\mathbf{u}_{(j)} = (u_{j1}, \dots, u_{jp_j})' \in R^{p_j}$ for $j = 1, \dots, G$. Following a similar strategy as in Fan and Li (2001), then it is enough to show that for any given $\varepsilon > 0$, there exists a large constant C such that

$$P \left\{ \inf_{\|\mathbf{z}\|=C} Q(\mathbf{b}_\tau^* + n^{-\frac{1}{2}}\mathbf{v}, \boldsymbol{\beta}^* + n^{-\frac{1}{2}}\mathbf{u}) > Q(\mathbf{b}_\tau^*, \boldsymbol{\beta}^*) \right\} \geq 1 - \varepsilon, \quad (6.1)$$

where $\mathbf{z} = (\mathbf{v}', \mathbf{u}')' \in R^{K+p}$ satisfying $\|\mathbf{z}\| = C$.

Let $Z_n(\mathbf{v}, \mathbf{u}) = \sum_{k=1}^K \sum_{i=1}^n \{\rho_{\tau_k}(\varepsilon_i - b_{\tau_k}^* - n^{-1/2}(v_k + \mathbf{x}'_i \mathbf{u})) - \rho_{\tau_k}(\varepsilon_i - b_{\tau_k}^*)\}$. Then it follows that

$$\begin{aligned} L_n(\mathbf{v}, \mathbf{u}) &= Q(\mathbf{b}_\tau^* + n^{-\frac{1}{2}}\mathbf{v}, \boldsymbol{\beta}^* + n^{-\frac{1}{2}}\mathbf{u}) - Q(\mathbf{b}_\tau^*, \boldsymbol{\beta}^*) \\ &= Z_n(\mathbf{v}, \mathbf{u}) + n\lambda \sum_{j=1}^G \left\| \hat{\boldsymbol{\beta}}_{(j)}^{CQR} \right\|_\infty^{-\gamma} \left\{ \left\| \boldsymbol{\beta}_{(j)}^* + \frac{\mathbf{u}_{(j)}}{\sqrt{n}} \right\|_\infty - \left\| \boldsymbol{\beta}_{(j)}^* \right\|_\infty \right\} \end{aligned} \quad (6.2)$$

$$\geq Z_n(\mathbf{v}, \mathbf{u}) - \sqrt{n}\lambda \sum_{j=1}^g \left\| \hat{\boldsymbol{\beta}}_{(j)}^{CQR} \right\|_\infty^{-\gamma} \left\| \mathbf{u}_{(j)} \right\|_\infty. \quad (6.3)$$

Using Knight (1998)'s identity,

$$\rho_\tau(u - v) - \rho_\tau(u) = -v\psi_\tau(u) + \int_0^v \{I(u \leq s) - I(u \leq 0)\} ds \quad \text{with } \psi_\tau(u) = \tau - I(u \leq 0),$$

we can write

$$Z_n(\mathbf{v}, \mathbf{u}) = Z_n^{(1)}(\mathbf{v}, \mathbf{u}) + Z_n^{(2)}(\mathbf{v}, \mathbf{u}),$$

where

$$\begin{aligned} Z_n^{(1)}(\mathbf{v}, \mathbf{u}) &= - \sum_{k=1}^K v_k \sum_{i=1}^n \frac{1}{\sqrt{n}} (\tau_k - I(\varepsilon_i \leq b_{\tau_k}^*)) - \mathbf{u}' \sum_{i=1}^n \frac{1}{\sqrt{n}} \mathbf{x}_i \sum_{k=1}^K (\tau_k - I(\varepsilon_i \leq b_{\tau_k}^*)), \\ Z_n^{(2)}(\mathbf{v}, \mathbf{u}) &= - \sum_{k=1}^K \sum_{i=1}^n \int_0^{\frac{v_k + \mathbf{x}'_i \mathbf{u}}{\sqrt{n}}} \{I(\varepsilon_i \leq b_{\tau_k}^* + s) - I(\varepsilon_i \leq b_{\tau_k}^*)\} ds. \end{aligned}$$

It follows from the Lindeberg-Feller CLT and Cramer-Wald device that

$$Z_n^{(1)}(\mathbf{v}, \mathbf{u}) \rightarrow_d - \sum_{k=1}^K v_k w_k - \mathbf{u}' W, \quad (6.4)$$

where $w_k \sim N(0, \tau_k(1 - \tau_k))$ and $W \sim N_p(\mathbf{0}, (\sum_{k,k'=1}^K (\tau_k \wedge \tau_{k'} - \tau_k \tau_{k'})) \boldsymbol{\Sigma})$.

Letting $Z_{n,k}^{(2)}(v_k, \mathbf{u}) = \sum_{i=1}^n \int_0^{(v_k + \mathbf{x}'_i \mathbf{u})/\sqrt{n}} \{I(\varepsilon_i \leq b_{\tau_k}^* + s) - I(\varepsilon_i \leq b_{\tau_k}^*)\} ds$, we can write

$$Z_{n,k}^{(2)}(v_k, \mathbf{u}) = \sum_{i=1}^n EZ_{n,k,i}^{(2)}(v_k, \mathbf{u}) + \sum_{i=1}^n \{Z_{n,k,i}^{(2)}(v_k, \mathbf{u}) - EZ_{n,k,i}^{(2)}(v_k, \mathbf{u})\}.$$

Without loss of generality, we assume that the predictors are centered to have mean zero, *i.e.*, $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \mathbf{x}_i = \mathbf{0}$. Then we have

$$\begin{aligned} E\left(Z_{n,k}^{(2)}(v_k, \mathbf{u})\right) &= \sum_{i=1}^n EZ_{n,k,i}^{(2)}(v_k, \mathbf{u}) = \sum_{i=1}^n E\left\{\int_0^{\frac{v_k + \mathbf{x}'_i \mathbf{u}}{\sqrt{n}}} \left\{I(\varepsilon_i \leq b_{\tau_k}^* + s) - I(\varepsilon_i \leq b_{\tau_k}^*)\right\} ds\right\} \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^{v_k + \mathbf{x}'_i \mathbf{u}} \sqrt{n} \left\{F\left(b_{\tau_k}^* + \frac{s}{\sqrt{n}}\right) - F(b_{\tau_k}^*)\right\} ds \rightarrow \frac{1}{2} f(b_{\tau_k}^*) \begin{bmatrix} v_k & \mathbf{u}' \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0}' & \Sigma \end{bmatrix} \begin{bmatrix} v_k \\ \mathbf{u} \end{bmatrix}. \end{aligned}$$

Moreover, we have

$$\begin{aligned} \text{Var}\left(Z_{n,k}^{(2)}(v_k, \mathbf{u})\right) &= \sum_{i=1}^n E\left\{\left(Z_{n,k,i}^{(2)}(v_k, \mathbf{u}) - EZ_{n,k,i}^{(2)}(v_k, \mathbf{u})\right)^2\right\} \leq \sum_{i=1}^n \frac{1}{\sqrt{n}} |v_k + \mathbf{x}'_i \mathbf{u}| E\left(Z_{n,k,i}^{(2)}(v_k, \mathbf{u})\right) \\ &\leq \max_{1 \leq i \leq n} \frac{1}{\sqrt{n}} |v_k + \mathbf{x}'_i \mathbf{u}| E\left(Z_{n,k}^{(2)}(v_k, \mathbf{u})\right) \rightarrow 0. \end{aligned}$$

Thus, it follows that

$$Z_n^{(2)}(\mathbf{v}, \mathbf{u}) = \sum_{k=1}^K Z_{n,k}^{(2)}(v_k, \mathbf{u}) \rightarrow_p \sum_{k=1}^K \frac{1}{2} f(b_{\tau_k}^*) \begin{bmatrix} v_k & \mathbf{u}' \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0}' & \Sigma \end{bmatrix} \begin{bmatrix} v_k \\ \mathbf{u} \end{bmatrix}. \quad (6.5)$$

Following the results of (6.4) and (6.5), we can show that $Z_n^{(2)}(\mathbf{v}, \mathbf{u})$ dominates $Z_n^{(1)}(\mathbf{v}, \mathbf{u})$ uniformly in $\|\mathbf{z}\| = C$ when a sufficiently large C is chosen. From the assumption $\sqrt{n}\lambda = O_p(1)$ and $\|\hat{\boldsymbol{\beta}}_{(j)}^{CQR}\|_{\infty}^{-\gamma} \rightarrow_p \|\boldsymbol{\beta}_{(j)}^*\|_{\infty}^{-\gamma}$ ($j = 1, \dots, g$), furthermore, the second term in (5.4) is also dominated by $Z_n^{(2)}(\mathbf{v}, \mathbf{u})$. Hence, by choosing a sufficiently large C , (6.1) holds. \square

Proof of Theorem 2:

We first prove the asymptotic normality part. For the second term in the right hand side of (6.2), we have $\|\hat{\boldsymbol{\beta}}_{(j)}^{CQR}\|_{\infty}^{-\gamma} \rightarrow_p \|\boldsymbol{\beta}_{(j)}^*\|_{\infty}^{-\gamma}$ and $|\sqrt{n}\{\|\boldsymbol{\beta}_{(j)}^* + \mathbf{u}_{(j)} / \sqrt{n}\|_{\infty} - \|\boldsymbol{\beta}_{(j)}^*\|_{\infty}\}| \leq \|\mathbf{u}_{(j)}\|_{\infty}$ for $j = 1, \dots, g$. Hence, from the assumption $\sqrt{n}\lambda \rightarrow 0$, we have $n\lambda \|\hat{\boldsymbol{\beta}}_{(j)}^{CQR}\|_{\infty}^{-\gamma} \{\|\boldsymbol{\beta}_{(j)}^* + \mathbf{u}_{(j)} / \sqrt{n}\|_{\infty} - \|\boldsymbol{\beta}_{(j)}^*\|_{\infty}\} \rightarrow_p 0$ for $j = 1, \dots, g$. However, for $j = g+1, \dots, G$, we have $\sqrt{n}\lambda \|\hat{\boldsymbol{\beta}}_{(j)}^{CQR}\|_{\infty}^{-\gamma} = n^{(1+\gamma)/2} \lambda \|\sqrt{n} \hat{\boldsymbol{\beta}}_{(j)}^{CQR}\|_{\infty}^{-\gamma} \rightarrow_p \infty$ with $\sqrt{n} \hat{\boldsymbol{\beta}}_{(j)}^{CQR} = O_p(1)$ and the assumption $n^{(1+\gamma)/2} \lambda \rightarrow \infty$. Since $\sqrt{n}\{\|\boldsymbol{\beta}_{(j)}^* + \mathbf{u}_{(j)} / \sqrt{n}\|_{\infty} - \|\boldsymbol{\beta}_{(j)}^*\|_{\infty}\} = \|\mathbf{u}_{(j)}\|_{\infty}$ for $j = g+1, \dots, G$, it follows that $n\lambda \|\hat{\boldsymbol{\beta}}_{(j)}^{CQR}\|_{\infty}^{-\gamma} \{\|\boldsymbol{\beta}_{(j)}^* + \mathbf{u}_{(j)} / \sqrt{n}\|_{\infty} - \|\boldsymbol{\beta}_{(j)}^*\|_{\infty}\} \rightarrow_p \infty$ if $\mathbf{u}_{(j)} \neq \mathbf{0}$ and $= 0$ otherwise. These facts and the results of (6.4) and (6.5) in the proof of Theorem 1 imply that

$$L_n(\mathbf{v}, \mathbf{u}) \rightarrow_d L_0(\mathbf{v}, \mathbf{u})$$

$$= \begin{cases} -\sum_{k=1}^K v_k w_k - \mathbf{u}'_A W_A + \sum_{k=1}^K \frac{1}{2} f(b_{\tau_k}^*) v_k^2 + \frac{1}{2} \left(\sum_{k=1}^K f(b_{\tau_k}^*)\right) \mathbf{u}'_A \Sigma_{AA} \mathbf{u}_A, & \text{if } \mathbf{u}_{(j)} = \mathbf{0}, \text{ for } j \geq g+1, \\ \infty, & \text{otherwise,} \end{cases}$$

where $\mathbf{u}_A = (\mathbf{u}'_{(1)}, \dots, \mathbf{u}'_{(g)})' \in R^{p_0}$ and $W_A \sim N_p(\mathbf{0}, (\sum_{k,k'=1}^K (\tau_k \wedge \tau_{k'} - \tau_k \tau_{k'})) \Sigma_{AA})$. The function $L_n(\mathbf{v}, \mathbf{u})$ is convex and $L_0(\mathbf{v}, \mathbf{u})$ has a unique minimizer. Thus, following the same arguments in Knight (1998) and Koenker (2005), we have

$$\sqrt{n} \left(\hat{\boldsymbol{\beta}}_A^{ASCQR} - \boldsymbol{\beta}_A^* \right) \rightarrow_d \left(\sum_{k=1}^K f(b_{\tau_k}^*) \right)^{-1} \Sigma_{AA}^{-1} W_A \sim N(\mathbf{0}, \sigma_{CQR}^2 \Sigma_{AA}^{-1}),$$

where $\sigma_{CQR}^2 = (\sum_{k=1}^K f(b_{\tau_k}^*))^{-2} (\sum_{k,k'=1}^K (\tau_k \wedge \tau_{k'} - \tau_k \tau_{k'}))$. Therefore, the asymptotic normality part is proven.

Now, we show the consistency property of model selection. For any given \mathbf{b}_τ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_A^T, \boldsymbol{\beta}_{A^c}^T)^T$ satisfying that $\|\mathbf{b}_\tau - \mathbf{b}_\tau^*\| = O_p(n^{-1/2})$, $\|\boldsymbol{\beta}_A - \boldsymbol{\beta}_A^*\| = O_p(n^{-1/2})$, and $0 < \|\boldsymbol{\beta}_{A^c}\| \leq Cn^{-1/2}$, we have

$$\begin{aligned} & \mathcal{Q}(\mathbf{b}_\tau, (\boldsymbol{\beta}_A^T, \mathbf{0}^T)^T) - \mathcal{Q}(\mathbf{b}_\tau, (\boldsymbol{\beta}_A^T, \boldsymbol{\beta}_{A^c}^T)^T) \\ &= \left\{ \mathcal{Q}(\mathbf{b}_\tau, (\boldsymbol{\beta}_A^T, \mathbf{0}^T)^T) - \mathcal{Q}(\mathbf{b}_\tau^*, (\boldsymbol{\beta}_A^{*T}, \mathbf{0}^T)^T) \right\} - \left\{ \mathcal{Q}(\mathbf{b}_\tau, (\boldsymbol{\beta}_A^T, \boldsymbol{\beta}_{A^c}^T)^T) - \mathcal{Q}(\mathbf{b}_\tau^*, (\boldsymbol{\beta}_A^{*T}, \mathbf{0}^T)^T) \right\} \\ &= Z_n \left(\sqrt{n}(\mathbf{b}_\tau - \mathbf{b}_\tau^*), \sqrt{n}((\boldsymbol{\beta}_A - \boldsymbol{\beta}_A^*)^T, \mathbf{0}^T)^T \right) - Z_n \left(\sqrt{n}(\mathbf{b}_\tau - \mathbf{b}_\tau^*), \sqrt{n}((\boldsymbol{\beta}_A - \boldsymbol{\beta}_A^*)^T, \boldsymbol{\beta}_{A^c}^T)^T \right) \\ &\quad - n\lambda \sum_{j=g+1}^G \|\hat{\boldsymbol{\beta}}_{(j)}^{CQR}\|_\infty^{-\gamma} \|\boldsymbol{\beta}_{(j)}\|_\infty \end{aligned} \quad (6.6)$$

and the last term in (6.6) follows that

$$n\lambda \sum_{j=g+1}^G \|\hat{\boldsymbol{\beta}}_{(j)}^{CQR}\|_\infty^{-\gamma} \|\boldsymbol{\beta}_{(j)}\|_\infty = \left(n^{\frac{1+\gamma}{2}} \lambda \right) \sqrt{n} \sum_{j=g+1}^G \|\sqrt{n} \hat{\boldsymbol{\beta}}_{(j)}^{CQR}\|_\infty^{-\gamma} \|\boldsymbol{\beta}_{(j)}\|_\infty \rightarrow \infty$$

Hence, the assumption $n^{(1+\gamma)/2} \lambda \rightarrow \infty$ implies that $n\lambda \sum_{j=g+1}^G \|\hat{\boldsymbol{\beta}}_{(j)}^{CQR}\|_\infty^{-\gamma} \|\boldsymbol{\beta}_{(j)}\|_\infty$ is of higher order than any other terms in (6.6) and dominates as a result. This in turn implies that $\mathcal{Q}(\mathbf{b}_\tau, (\boldsymbol{\beta}_A^T, \mathbf{0}^T)^T) - \mathcal{Q}(\mathbf{b}_\tau, (\boldsymbol{\beta}_A^T, \boldsymbol{\beta}_{A^c}^T)^T) < 0$ for large n . This proves the consistency of the model selection. \square

References

- Bang, S. and Jhun, M. (2012). Simultaneous estimation and factor selection in quantile regression via adaptive sup-norm regularization, *Computational Statistics and Data Analysis*, **56**, 813–826.
- Bradic, J., Fan, J. and Wang, W. (2011). Penalized composite quasi-likelihood for ultrahigh dimensional variable selection, *Journal of the Royal Statistical Society, Series B*, **73**, 325–349.
- Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association*, **96**, 1348–1360.
- Knight, K. (1998). Limiting distributions for L1 regression estimators under general conditions, *The Annals of Statistics*, **26**, 755–770.
- Koenker, R. (1984). A note on L-estimates for linear models, *Statistics and Probability Letters*, **2**, 323–325.
- Koenker, R. (2005). *Quantile Regression*, Cambridge University Press, New York.
- Koenker, R. and Bassett, G. (1978). Regression quantiles, *Econometrica*, **4**, 33–50.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society, Series B*, **58**, 267–288.
- Wang, H. and Leng, C. (2008). A note on adaptive group lasso, *Computational Statistics and Data Analysis*, **52**, 5277–5286.
- Wu, Y. and Liu, Y. (2009). Variable selection in quantile regression, *Statistica Sinica*, **19**, 801–817.
- Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society, Series B*, **68**, 49–67.
- Zhao, P., Rocha, G. and Yu, B. (2009). The composite absolute penalties family for grouped and hierarchical variable selection, *The Annals of Statistics*, **37**, 3468–3497.

- Zou, H. (2006). The adaptive lasso and its oracle properties, *Journal of the American Statistical Association*, **101**, 1418–1429.
- Zou, H. and Yuan, M. (2008a). Composite quantile regression and the oracle model selection theory, *The Annals of Statistics*, **36**, 1108–1126.
- Zou, H. and Yuan, M. (2008b). The F_∞ -norm support vector machine, *Statistica Sinica*, **18**, 379–398.

Received February 19, 2013; Revised May 2, 2013; Accepted June 12, 2013