Pak. J. Statist.
2015 Vol. 31(1), 49-70

COMBINED PENALIZED QUANTILE REGRESSION IN HIGH DIMENSIONAL MODELS

Muhammad Amin[§], Lixin Song[§], Milton Abdul Thorlie and **Xiaoguang Wang**
School of Mathematical Sciences, Dalian University of Technology
Dalian, 116023, P.R. China
[§]Corresponding authors Email: aminkanju@gmail.com, lxsongdl@163.com

ABSTRACT

The quantile regression technique is considered as an alternative to the classical ordinary least squares (OLS) regression in case of outliers and heavy tailed errors existing in linear models. In this work, the consistency, asymptotic normality, and oracle property are established for sparse quantile regression with a diverging number of parameters. The rate of convergence of the combined penalized estimator is also established. Furthermore, the rank correlation screening (RCS) method is applied to deal with an ultrahigh dimensional data. The simulation studies, the analysis of hedonic housing prices and the demand for clean air dataset are conducted to illustrate the finite sample performance of the proposed method.

KEYWORDS

Combined penalization; Ridge-SCAD; Variable selection; Quantile regression.

1. INTRODUCTION

The existence of heavy tailed error or outliers (response or predictors) in linear models restrict the use of classical regression techniques. The present variable selection methods of linear regression models based on Ridge-SCAD regression are only applicable for the finite number of predictors and most of them lack the oracle property associated with the estimator. The traditional variable selection methods have several drawbacks, the most important is that they are unstable due to their inherent discreteness (Breiman, 1996). To overcome such drawbacks some shrinkage procedures based on the penalized function have been proposed, which include the bridge regression (Frank and Friedman, 1993) and the smoothly clipped absolute deviation (SCAD) (Fan and Li, 2001). Various penalized approaches have been proposed for exploring the selection of variables and explaining the statistical properties of high dimensional data. Huang et al. (Huang, Horowitz and Ma, 2008) studied variable selection in the accelerated failure time model via the bridge penalty. Ma and Du (2012) studied the variable selection in the partially linear model with high dimensional covariates. The L_1 penalty that yields the soft threshold rule (Donoho and Johnstone, 1994) and leads to the least absolute shrinkage and selection operator (Lasso) which was introduced by Tibshirani (1996), the same was further studied by Efron et al. (2004). The L_2 penalty which resulted in the ridge regression was discussed by Hoerl and Kennard (1970). Zou and Hastie (2005)

proposed the elastic net (Enet) method, which is a combination of Lasso and ridge penalties.

The quantile regression method proposed by Koenker and Bassett (1978) has attracted much attention as an alternative approach to least square regression. A comprehensive introduction and recent important developments are studied by Koenker (2005). It is a flexible technique in assessing the effect of predictors on different locations of the response distribution. He and Shao (2000) established the asymptotic theory for high dimensional M-regression with non-smooth objective function. The results of which can be applied to quantile regression without assuming sparseness assumption.

Wang et al. (2010) proposed a new combined penalization in the linear regression which outperforms the SCAD penalty especially when the correlation among the predictors is high. In this article, the methodology and the theory of quantile regression for variable selection via Ridge-SCAD regression considering high dimensional regression setting in which the number of covariates "p" grows at an increasing rate of the sample size "n" is further extended.

Furthermore, we demonstrated that under certain asymptotic conditions, the combined penalized quantile estimator with properly selected tuning parameters satisfies the oracle property. The rank correlation screening (RCS) method proposed by Li et al. (2012) is required to deal with ultrahigh dimensional data.

The rest of this article is organized as following. Section 2 introduces the estimation and the combined penalty quantile selection procedure in a high-dimension and it also presents the theoretical results. The simulation studies and the numerical comparisons are given in Section 3. All the technical proofs of theorems are given in Section 4. Section 5 concludes the article with a discussion.

2. COMBINED PENALIZED QUANTILE REGRESSION

2.1 Ridge-SCAD Estimation and Variable Selection Procedure

If we consider linear regression model

$$Y_i = \beta_0^{\mathrm{T}} x_i + \varepsilon_i,\ i = 1, \cdots, n \tag{1}$$

where Y_i is the response variable, β_0 is the vector of regression coefficients, x_i is a random vector of predictors p_n and ε_i is the random error with median zero. Without loss of generality, it is assumed that the outcome is centered and the predictors are standardized. Theerefore, interncept β_0 is excluded in the regression model. The model is sparse given by $\beta_0 = \left(\beta_{10}^{\mathrm{T}}, \beta_{20}^{\mathrm{T}}\right)^{\mathrm{T}}$, in which $\beta_{10} \neq 0$ is a $k_n \times 1$ vector and $\beta_{20} = 0$ is a $m_n \times 1$ vector. For $p_n = k_n + m_n$, where k_n & m_n are the significant and non significant predictors accordingly. The error term satisfies $p\left(\varepsilon_i \geq 0 \mid x_i\right) = \tau$ for $0 < \tau \leq 1$. Let $A = \left\{1 \leq j \leq p_n, \beta_{j0} \neq 0\right\}$ is a set of nonzero coefficients. The oracle estimator is defined

as $\hat{\beta}_0 = \left(\hat{\beta}_{10}^{\mathrm{T}}, \hat{\beta}_{20}^{\mathrm{T}}\right)^{\mathrm{T}}$, where $\hat{\beta}_{10}$ is the combined quantile regression estimator when the model is fitted using covariants in set A. The recent literature reveals that various versions of penalized functions are proposed for high dimensional data. Belloni and Chernozhukov (2011) derived an error bound for quantile regression corresponding to L_1-penalty to reduce the quantile regression bias but it lacks the oracle property. The bridge penalty was proposed by Frank and Friedman (1993) regarding L_q-penalty defined as $p_\lambda(|\theta|) = \lambda|\theta|^q$, $0 < q < 1$. Knight and Fu (2000) investigated the asymptotic nature of bridge estimators for finite number of covariates. Fan and Li (2001) discussed the SCAD penalty, given by

$$p_\lambda(|\theta|) = \lambda|\theta| I(0 \le |\theta| < \lambda) + \frac{(a^2-1)\lambda^2 - (|\theta| - a\lambda)^2}{2(a-1)} I(\lambda \le |\theta| < a\lambda) + \frac{(a+1)\lambda^2}{2} I(|\theta| \ge a\lambda) \quad (2)$$

where $a = 3.7$ and $\lambda > 0$ is the tuning parameter. It is continuous and differentiable on $(-\infty, 0) \bigcup (0, \infty)$ but not differentiable at zero. It's derivative vanishes outside $[-a\lambda, a\lambda]$. An alternative combined quantile (ridge and SCAD) strategy is proposed. For any fixed non-negative value of γ and λ, the combined penalty can be written as:

$$J_{\lambda,\gamma}(\theta) = \frac{\gamma}{2}(\theta)^2 + P_\lambda(\theta),\ \theta > 0.$$

If we consider a combined penalized quantile high dimensional regression model given by

$$Q_n(\beta_n) = \frac{1}{n}\sum_{i=1}^{n} \rho_{\tau_n}\left(Y_i - \beta_n^{\mathrm{T}} x_i\right) + \sum_{j=1}^{p_n}\left[\frac{\gamma_n}{2}\beta_{nj}^2 + P_{\lambda_n}(|\beta_{nj}|)\right], \quad (3)$$

where $\rho_{\tau(m)} = m\left(\tau - I(m < 0)\right)$ is a quantile check function. $P_{\lambda_n}(.)$ and $\gamma_n(.)$ are the functions of SCAD and ridge penalization. The λ_n and γ_n are non-negative tuning parameters of SCAD and ridge penalties and are responsible to control the model complexity with given rates. Combined penalized estimator can be minimized as

$$\frac{1}{n}\sum_{i=1}^{n} \rho_{\tau_n}\left(Y_i - \beta_n^{\mathrm{T}} x_i\right) + \sum_{j=1}^{p_n}\left[\frac{\gamma_n}{2}\beta_{nj}^2 + P_{\lambda_n}(|\beta_{nj}|)\right],$$

This minimization problem can be expressed as a constraints to smoothen the optimization problem. In optimization, these constraints will lead to asymptotic aspect of the Ridge-SCAD penalized regression under the given conditions. In order to significantly improve computational efficiency of the model, we consider the objective function given by

$$Q_n(\beta_n) = \frac{1}{n}\sum_{i=1}^{n}\rho_{\tau_n}\left(Y_i - \beta_n^{\mathrm{T}} x_i\right) + \sum_{j=1}^{p_n}\left[\gamma_n\beta_{nj}^0 + P'_{\lambda_n}\left(|\beta_{nj}^0|\right)|\beta_{nj}|\right], \text{ for } \beta_{nj} > 0 \qquad (4)$$

where $J'_{\lambda,\gamma}(\beta_{nj}) = \gamma_n\beta_{nj}^0 + P'_{\lambda_n}\left(|\beta_{nj}^0|\right)|\beta_{nj}|$, is a first order derivative of Ridge-SCAD penalization. The $\beta_n^0 = \left(\beta_{n1}^0,\cdots,\beta_{np_n}^0\right)^{\mathrm{T}}$ is an initial estimator, which is usually set to be the unpenalized quantile regression estimator, as followed by He and Shao (2000) it can be defined as $\left\|\beta_n^0 - \beta_0\right\| = Op\left(\sqrt{p_n/n}\right)$. The value of $\hat{\beta}_n$ which minimizes (4) is called Ridge-SCAD estimator, as obtained by He and Shao (2000). Given the notation $A_n = Op(B_n)$, where A_n is a sequence order less than or equal to B_n in probability. By partitioning the parameter vector $\beta_n = \left(\beta_{1n}^{\mathrm{T}}, \beta_{2n}^{\mathrm{T}}\right)^{\mathrm{T}}$ in the same fashion as β_0, defined by the initial estimator β_n^0. The population covariant vector $x = \left(w^{\mathrm{T}}, z^{\mathrm{T}}\right)^{\mathrm{T}}$ can be partitioned with the corresponding samples given by $x_i = \left(w_i^{\mathrm{T}}, z_i^{\mathrm{T}}\right)^{\mathrm{T}}$, where $w_i = \left(x_{i1},\ldots,x_{ik_n}\right)^{\mathrm{T}}$ and $z_i = \left(x_{i(k_n+1)},\ldots,x_{ip_n}\right)^{\mathrm{T}}$. Suppose ρ_{n1} & ρ_{n2} and τ_{n1} & τ_{n2} are the smallest and largest eigenvalues of the matrix $E\left(xx^{\mathrm{T}}\right)$ and $E\left(ww^{\mathrm{T}}\right)$ respectively.

2.2 Asymptotic Properties

Let $x_i^{\mathrm{T}} = \left(w_i^{\mathrm{T}}, z_i^{\mathrm{T}}\right)$, where $w_i = \left(x_{i1},\ldots,x_{ik_n}\right)^{\mathrm{T}}$ and $z_i = \left(x_{i(k_n+1)},\ldots,x_{ip_n}\right)^{\mathrm{T}}$. Given the penalized check function $F = \left\{f(x): f(x) = g(x) - h(x)\right\}$, where h, g are convex. Let $g = \left\{x : g(x) < \infty\right\}$ be the domain with respect to g and also let $\partial g(x_0) = \left\{t : g(x) \geq g(x_0) + (x - x_0)^{\mathrm{T}} t, \forall x\right\}$ is a sub-differential function of $g(x)$ at point x_0.

C1) The error ε_i are independent and identically distributed with τ^{th} quantile zero and has a continuous and positive density $f(.)$ at the origin. The density function $f(.)$ has a finite derivative in a neighborhood around 0.

C2) There exists a positive constant $M < \infty$ such that $\max_{1\leq i\leq n, 1\leq j\leq p_n} |x_{ij}| \leq M$.

C3) Given $p_n^3/n \to 0$ as $n \to \infty$ then there exist a positive constant M such that $M_1 \leq \lambda_{\min}\left\{I_n(\beta_n)\right\} \leq \lambda_{\max}\left\{I_n(\beta_n)\right\} \leq M_2 < \infty$, where $\lambda_{min}, \lambda_{max}$ are the smallest

and largest eigenvalues respectively. Given $max_{1\le i\le n} \| z_i \| = O_p\left(\sqrt{p_n / n} + a_n\right)$, where $a_n = \max\left\{\lambda_{nj}, \gamma_{nj}, 1 \le j \le p_n\right\}$ λ_j, γ_j are the functions of n.

C4) There exist constants $0 < \rho_1 < \rho_2 < \infty$ and $0 < \tau_1 < \tau_2 < \infty$ such that $\rho_1 \le \rho_{n1} \le \rho_{n2} \le \rho_2$ and $\tau_1 \le \tau_{n1} \le \tau_{n2} \le \tau_2$.

C5) The true model dimension satisfies $\lambda_n \to 0$ for $\sqrt{n / p_n}\lambda_n \to \infty, \sqrt{n}\lambda_n \to \infty$ and $\sqrt{n}\gamma_n \to 0, as\ n \to \infty$.

The condition (C1) is typical and widely employed in the literature (Knight, 1998; Pollard, 1991; Wu and Liu, 2009), (C2) explains the restriction on covariates and defines the properties of consistency and asymptotic normality. It requires design matrix corresponding to the true model if it is well behaved. The same one was used by Huang and Xie (2007). The (C3) express that $p_n = o\left(n^{1/3}\right)$ and the same was taken by Huber (1973), (C4) assume the matrices $E\left(xx^T\right)$ and $E\left(ww^T\right)$ are positive definite and is identical to the condition given in Huang et al. (2008) and Li et al. (2011). In particular they are similar to Kim et al. (2008). Condition (C5) ensures the estimator with sparsity property by Fan and Peng (2004).

Theorem 2.2.1 (Consistency)

Under conditions (C1) to (C5), $\left\|\hat{\beta}_n - \beta_0\right\| = O_P\left(\sqrt{p_n / n} + a_n\right)$, it shows the consistent estimator for true parameter β_0 with optimal convergence rate regarding diverging number of parameters.

We partition $\hat{\beta}_n = \left(\hat{\beta}_{1n}^T, \hat{\beta}_{2n}^T\right)^T$, given β_0. Let $\theta_n = \beta_{1n} - \beta_{10}$ and $\hat{\theta}_n = \hat{\beta}_{1n} - \beta_{10}$, we have a new function, $Q_{n\tau}^*\left(\theta_n\right) = \frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(\varepsilon_i - \theta_n^T w_i\right) + \sum_{j=1}^{k_n}\left(\gamma_n\beta_{nj} + P'_{\lambda_n}\left(|\beta_{nj}^0|\right)\right)|\theta_{nj} + \beta_{10j}|$. Let $g\left(\theta_n\right) = \left(\varepsilon - \theta_n^T w\right)$, then linear approximation near 0 is

$$g\left(\theta_n\right) = g(0) + \rho_\tau\theta_n^{\mathrm{T}} D(0) + \left\|\theta_n\right\| r\left(\theta_n\right),$$

where $D(0) = -\left[I\left(\varepsilon > 0\right) - I\left(\varepsilon < 0\right)\right]w$, $I(\cdot)$ is the indicator function, and $r\left(\theta_n\right)$ is the remainder term. Let $S_n\left(\theta_n\right) = \frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(\varepsilon_i - \theta_n^{\mathrm{T}} w_i\right)$ and $S\left(\theta_n\right) = \rho_\tau E\left(\varepsilon - \theta_n^{\mathrm{T}} w\right)$. Then

$$S\left(\theta_n\right) = E\left[E\left(\varepsilon - \theta_n^{\mathrm{T}} w\right) | w\right] \underline{\underline{\Delta}} E\left[H\left(\theta_n^{\mathrm{T}} w\right)\right].$$

Given

$$\rho_\tau\left(\varepsilon_i - \theta_n' w_i\right) = 0$$

where $w_i = \left|\beta_{\tau_j}\right|^{-1}$, $1 \le j \le p$. $\dfrac{\partial Q_n(\beta_n)}{\partial \beta_{nj}}|_\beta = -\sum_{i=1}^{n}\rho_\tau\left(Y_i - \beta_n^T x_i\right)x_{ij} + \lambda_n w_j sgn\left(\beta_{nj}\right)$

Suppose $H(\cdot)$ has a finite third order derivative, then the empirical process E_n defined by $E_n(\cdot) = \sqrt{n}\left(S_n(\cdot) - S(\cdot)\right)$ given $\rho_\tau(.)$ therefore $\rho_\tau(t) = \tau_i\left((t \ge 0) - (1-\tau)\,|\,(t < 0)\right)$ and let $E\left(ww^{\mathrm{T}}\right) \underline{\Delta} \Sigma_{n,11}$.

Theorem 2.2.2 (Oracle Property)

Suppose that $k_n^3 p_n^3 / n \to 0$, $E_n\left(r\left(\hat{\theta}_n\right)\right) = o_P\left(1/\sqrt{p_n}\right)$, $\sqrt{p_n/n}/\lambda_n \to 0$, and $\sqrt{n}\gamma_n \to 0$.

If the conditions (C1)-(C5) hold, the Ridge-SCAD estimator $\hat{\beta}_n = \left(\hat{\beta}_{1n}^{\mathrm{T}}, \hat{\beta}_{2n}^{\mathrm{T}}\right)^{\mathrm{T}}$ satisfies

(1) Sparsity: $\Pr\left(\hat{\beta}_{2n} = 0\right) \to 1$ as $n \to \infty$.

(2) Asymptotic normality: $n^{1/2}\alpha^{\mathrm{T}}\rho_\tau\Sigma_{n,11}^{1/2}\left(\hat{\beta}_{1n} - \beta_{10}\right) \xrightarrow{D} N\left(0, \tau(1-\tau)\left(f(0)\right)^{-2}\right)$,

where α is an arbitrary $k_n \times 1$ vector with $\|\alpha\| = 1$, $\xrightarrow{D}$ means convergence in distribution.

2.3 Computation and Selection of Tuning Parameters

By following the similar idea of computation proposed by Wang et al. (2012), the objection function in (4) can be minimized as

$$\min_{\beta}\left\{\frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(Y_i - \beta_{nj}^T x_i\right) + \sum_{j=1}^{p_n} w_j^{(t-1)}\beta_j\right\}$$

where $w_j^{(t-1)} = \left[\gamma_n\tilde{\beta}_j^{(t-1)} + P'\lambda_n\left(\tilde{\beta}_j^{(t-1)}\right)\right] > 0$. Given λ_n, γ_n and $\tilde{\beta}_j$, and with the aid of slack variables, the convex optimization problem above can also be casted into a constrained linear programming problem as follows

$$\min_{\xi,\zeta}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\tau\xi_i^+ + (1-\tau)\xi_i^-\right) + \sum_{j=1}^{p_n} w_j^{(t-1)}\zeta_j\right\}$$

subject to $\xi_i^+ - \xi_i^- = Y_i - \beta_{nj}^T x_i$; $i = 1, 2, \ldots, n$,

$\xi_i^+ \ge 0, \xi_i^- \ge 0$; $i = 1, 2, \ldots, n$,

$\zeta_j \ge \beta_j, \zeta_j \ge -\beta_j$; $j = 1, 2, \ldots, p$.

This minimization can be achieved by using any existing linear programming software, we use the R function $rq.fit.fn$ in the package *quantreg*.

In order to deal with the ultrahigh dimensional case, we first reduce the dimensionality from $p_n >> n$ to $p_n < n$ and applied RCS method originally proposed by Li et al. (2012). Let $w = \left(w_1, \ldots, w_{p_n}\right)^{\mathrm{T}}$ be a p_n -vector, where

$$w_k = \frac{1}{n(n-1)} \sum_{i \neq j}^{n} I\left(x_{ik} < x_{jk}\right) I\left(y_i < y_j\right) - \frac{1}{4},\ k = 1, \ldots, p_n.$$

where $I(.)$ denotes the indictor function, and w is the marginal rank correlation coefficient. The p_n which is a magnitudes of the vector w is sorted in a decreasing order to define the submodel,

$$\mathcal{M} = \left\{1 \leq k \leq p_n : |\, w_k \,| \text{ is among the first } d_n \text{ largest of all}\right\},$$

the method is relatively simple, since it shrinks the full model $\{1, \ldots, p_n\}$ to a submodel $\mathcal{M}$ with size $d_n < n$.

Several selection criterias such as cross validation (CV), generalized cross validation (GCV), Akaike information criterion (AIC) and Bayesian information criterion (BIC) can be used to choose proper tuning parameters. BIC-type criterion was used that is given in the following formula which was also used in Jiang et al. (2014).

$$BIC(\lambda) = \log\left(\frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(Y_i - \hat{\beta}^{\mathrm{T}}\mathrm{x}_i\right)\right) + \frac{\log(n)}{n} edf(\lambda),$$

where the first term measures the quadratic loss and *edf* as the number of nonzero coefficients.

3. NUMERICAL STUDIES

This section illustrates the finite sample performance of the proposed method. It mainly focuses on comparing the performance of the combined Ridge-SCAD (CP) method with Lasso, Enet and SCAD. A real data set is used for further demonstration.

Example 3.1

The data simulated from linear model is,

$$Y = \beta^{\mathrm{T}}\mathrm{x} + \varepsilon,$$

with "n" observations. β is a $p_n \times 1$ vector with $\beta_1 = 3, \beta_2 = 1.5, \beta_5 = 2$ and the other β_j's being 0. The vector of covariates "x" follows a multivariate normal distribution $N\left(0, \Sigma_x\right)$, $\left(\Sigma_x\right)_{ij} = r^{|i-j|}$ for $1 \leq i, j \leq p_n$ with r=0.5. To emphasize the dependency of the

number of parameters on the sample size, we considered $n = 100$ with $p_n = \lceil 10n^{1/4} \rceil$. For comparison, three error distributions are studied: the standard normal distribution, the t-distribution with 3 d.f, and the contaminated standard normal (CN) distribution with 10% outliers from the standard cauchy distribution. With model error, the finite performance of Qr.Lasso, Qr.Enet, Qr.SCAD, quantile Combined Penalization (Qr.CP) and Oracle compared. The oracle estimator is computed by using the true model when the zero coefficients are known, in practice, it cannot be obtained. It can only be used here as a benchmark for comparison. The model error is computed by $ME = \left(\hat{\beta}_n - \beta\right)^{\mathrm{T}} \Sigma_{\mathrm{x}} \left(\hat{\beta}_n - \beta\right)$. The results of the variable selection and the model errors are summarized in Table 1. The column "MRME" stands for the median of the relative model error, which is defined as the ratio of ME to MEL, where ME is the model error of a selected model, and MEL is the model error of the proposed estimate under the full model. "C" denotes the average number of zero coefficients correctly set to zero, and "IC" gives the average number of nonzero coefficients incorrectly set to zero. Table 1 shows that Qr.CP and Qr.SCAD are very close to oracle in terms of MRME in all error distributions, but the others performed worse. The Qr.CP and Qr.Enet are better than Qr.SCAD and Qr. Lasso, respectively. Overall Qr.CP and Qr.SCAD are one of the best variable selectors among the penalization methods in this example.

Table 1

Simulation Results with 1000 Data Sets $\left(n = 100,\ p_n = 31\right)$

τ	**Methods**	$\varepsilon \sim N(0,1)$			$\varepsilon \sim t(3)$			$\varepsilon \sim CN$		
		C	**IC**	**MRME**	**C**	**IC**	**MRME**	**C**	**IC**	**MRME**
0.3	Qr.Lasso	26.43	0.000	0.2998	26.70	0.016	0.2696	26.76	0.012	0.3244
	Qr.Enet	26.45	0.000	0.2988	26.79	0.013	0.2674	26.80	0.012	0.3199
	Qr.SCAD	27.88	0.000	0.0664	27.82	0.016	0.0639	27.93	0.012	0.0545
	Qr.CP	27.88	0.000	0.0663	27.83	0.012	0.0629	27.95	0.003	0.0541
	Oracle	28.00	0.000	0.0551	28.00	0.000	0.0439	28.00	0.000	0.0507
0.5	Qr.Lasso	26.69	0.000	0.3247	27.01	0.004	0.3082	26.98	0.006	0.3339
	Qr.Enet	26.73	0.000	0.3247	27.07	0.005	0.3059	27.01	0.006	0.3328
	Qr.SCAD	27.92	0.000	0.0664	27.91	0.012	0.0582	27.97	0.001	0.0537
	Qr.CP	27.92	0.000	0.0662	27.92	0.007	0.0572	27.97	0.007	0.0542
	Oracle	28.00	0.000	0.0598	28.00	0.000	0.0467	28.00	0.000	0.0519
0.7	Qr.Lasso	26.46	0.002	0.2999	26.57	0.012	0.2621	26.70	0.012	0.3129
	Qr.Enet	26.46	0.002	0.2946	26.65	0.008	0.2619	26.76	0.012	0.3098
	Qr.SCAD	27.85	0.000	0.0686	27.81	0.018	0.0601	27.93	0.01	0.0579
	Qr.CP	27.86	0.000	0.0685	27.82	0.022	0.0599	27.95	0.003	0.0574
	Oracle	28.00	0.000	0.0567	28.00	0.000	0.0444	28.00	0.000	0.0552

Example 3.2

Consider the same linear model as it is in Example 3.1, 200 datasets are simulated with $(n, p_n) = (100, 500)$. The RCS method was used to reduce the dimensionality p_n to d_n, where $d_n = \lceil 4n/\log(n) \rceil$. Simulation results are displayed in Table 2. The column "MEE" stands for the median of the estimation error, defined as $\| \hat{\beta} - \beta \|$. This example is set to see the performance of the proposed procedure in an ultrahigh dimensional case. We can see that the Ridge-SCAD estimator performs as well as the oracle estimator in terms of estimation accuracy and model complexity for 3 error distributions than other competitors.

Table 2

Simulation Results with 200 Data Sets $(n = 100,\ p_n = 500)$

τ	**Methods**	$\varepsilon \sim N(0,1)$			$\varepsilon \sim t(3)$			$\varepsilon \sim CN$		
		C	**IC**	**MEE**	**C**	**IC**	**MEE**	**C**	**IC**	**MEE**
0.3	Qr.Lasso	492.28	0.005	0.7667	493.18	0.035	0.9742	493.43	0.055	0.8680
	Qr.Enet	492.39	0.005	0.7704	493.20	0.035	0.9740	493.66	0.060	0.8551
	Qr.SCAD	496.69	0.010	0.2285	495.18	0.205	0.5003	494.42	0.220	0.3607
	Qr.CP	496.76	0.010	0.2269	495.39	0.140	0.4744	496.24	0.225	0.3606
	Oracle	497.00	0.000	0.2056	497.00	0.000	0.2185	497.00	0.000	0.2122
0.5	Qr.Lasso	493.42	0.005	0.7248	493.82	0.030	0.8443	494.39	0.020	0.8435
	Qr.Enet	493.31	0.005	0.7197	494.05	0.030	0.8368	494.47	0.020	0.8416
	Qr.SCAD	496.79	0.005	0.2134	495.54	0.095	0.3964	496.08	0.160	0.2903
	Qr.CP	496.80	0.005	0.2124	495.78	0.055	0.3813	496.41	0.090	0.2872
	Oracle	497.00	0.000	0.1950	497.00	0.000	0.2145	497.00	0.000	0.1944
0.7	Qr.Lasso	492.63	0.010	0.7301	492.41	0.055	1.0121	492.92	0.035	0.8477
	Qr.Enet	492.74	0.010	0.7177	492.69	0.065	0.9566	492.85	0.035	0.8445
	Qr.SCAD	496.64	0.010	0.2453	495.40	0.185	0.3215	494.00	0.145	0.3874
	Qr.CP	496.68	0.010	0.2445	495.76	0.080	0.3012	495.42	0.105	0.3770
	Oracle	497.00	0.000	0.2152	497.00	0.000	0.2417	497.00	0.000	0.2455

Example 3.3

Considering other setting like Example 3.1, we generate X_j to be i.i.d. and $X_j \sim N(0, n),\ j = 1, ..., p_n$. When n=100 and $p_n = 31$, for $j = 6, ..., 8$, and each X_j is replaced by $X_{j+3} + \eta_j$, where η_j are i.i.d. and $\eta_j \sim N(0, 0.01)$. Therefore, each significant variable X_j, $j = 9, ..., 31$, has another strongly correlated variable X_{j-3}. This example is designed to observe how the penalization strategy performs when there are ultra-high correlations existing among predictors. Table 3 demonstrates the simulation results of Example 3.3, The Qr.CP outperforms the Qr.SCAD in both aspects "MRME" and "C", because of noise predictors with strongly correlated groups of pairs. The Qr.SCAD is more worse than the others in this situation.

Table 3
Simulation Results with 200 Data Sets $(n=100,\ p_n=31)$

Methods	$\tau=0.3$			$\tau=0.5$			$\tau=0.7$		
	C	**IC**	**MRME**	**C**	**IC**	**MRME**	**C**	**IC**	**MRME**
Qr.Lasso	26.17	0	0.0018	26.61	0	0.0026	26.02	0	0.0020
Qr.Enet	26.26	0	0.0016	26.77	0	0.0023	26.19	0	0.0019
Qr.SCAD	22.39	0	0.6072	22.47	0	0.7593	22.24	0	0.6456
Qr.CP	27.55	0	0.0002	27.56	0	0.0002	27.46	0	0.0002
Oracle	28.00	0	0.0001	28.00	0	0.0001	28.00	0	0.0001

Example 3.4

To examine the usefulness of proposed method, It was applied to analyze a data set of the Boston housing data to examine the correlation between clean air and housing prices (Harrison and Rubinfeld, 1978). There are 506 observations, 13 independent factors, and a response variable LMV, which is the logarithm of the median value of owner-occupied homes. Details are available online at *http://lib.stat.cmu.edu/datasets/dostoncorrected.txt*. Linear model to the LMV after first standardizing the predictors is fitted. The Q-Q plot and box plot showed that the response variable LMV contains many obvious outliers. Therefore, application of the penalized quantile regression is much better than penalized OLS approach. We split 506 observations into first 400 observations as a training data set to select and fit the model, and rest as a testing data set to evaluate the prediction ability of the selected model. Then was calculated the median absolute prediction error (MAPE) $\left(\text{median}\left\{|\ y_i-\hat{y}_i\ |, i=1,...,106\right\}\right)$ using the testing data. The performance of the penalized quantile regression with different penalties and different quantiles are summarized in Table 4. The results indicate that Qr.CP and Qr.SCAD select the simplest model, while Qr.Lasso and Qr.Enet include extra variables.

Table 4
Results of the Boston House Price Data

Methods	$\tau=0.3$		$\tau=0.5$		$\tau=0.7$	
	No. Zeros	**MAPE**	**No. Zeros**	**MAPE**	**No. Zeros**	**MAPE**
OLS	0	5.3147	0	5.3147	0	5.3147
QR.	0	3.0460	0	3.0835	0	3.2543
Qr.Lasso	5	3.3552	5	3.1569	5	3.1896
Qr.Enet	5	3.2641	6	3.1684	5	3.1248
Qr.SCAD	10	3.0339	8	3.0506	6	3.0521
Qr.CP	10	3.0213	8	3.0578	6	3.0124

4. PROOF OF THEOREMS

Proof of Theorem 2.2.1:

With any given $\varepsilon > 0$, there exist a constant C such that

$$\Pr\left\{\inf_{\|u\|=C} Q_n\left(\beta_0 + \alpha_n u\right) > Q_n\left(\beta_0\right)\right\} \geq 1-\varepsilon, \tag{5}$$

where $\alpha_n \triangleq \sqrt{p_n / n} + a_n$ which implies that probability at least $1-\varepsilon$, there exist a local minimum in the ball $\{\beta_0 + \alpha_n u : \| u \| \leq C\}$, where u is a $p_n \times 1$ vector, that is there exists a local minimizer such that $\| \hat{\beta}_n - \beta_0 \| = O_P\left(\sqrt{p_n / n} + a_n\right)$. Let $D_n(u) = Q_n\left(\beta_0 + \alpha_n u\right) - Q_n\left(\beta_0\right)$, then

$$\begin{aligned}
D_n(u) &= \left[\frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(Y_i - \left(\beta_0 + \alpha_n \mathrm{u}\right)^{\mathrm{T}} x_i - \frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(Y_i - \beta_0^{\mathrm{T}} x_i\right)\right)\right] \\
&\quad + \sum_{j=1}^{p_n}\left[\gamma_n\left(\beta_{0j} + \alpha_n \mathrm{u}_j\right) - \left(\beta_{0j}\right)\right] \\
&\quad - \sum_{j=1}^{p_n}\left[p'_{\lambda_n}\left(|\beta_{nj}^0|\right)\left(\beta_{0j} + \alpha_n u_j\right) - p'_{\lambda_n}\left(|\beta_{nj}|\right)\right] \\
&= \left[\frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(Y_i - \left(\beta_0 + \alpha_n \mathrm{u}\right)^{\mathrm{T}} x_i\right]\right. \\
&\quad - \sum_{j=1}^{p_n}\left[\gamma_n\left(\beta_{0j} + \alpha_n \mathrm{u}_j\right) + p'_{\lambda_n}\left(|\beta_{nj}^0|\right) + \left(|\beta_{0j} + \alpha_n \mathrm{u}_j|\right)\right] \\
&\quad \left. - \left[\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \beta_0\right)^{\mathrm{T}} x_i - \sum_{j=1}^{p_n}\left[\gamma_n\left(\beta_{0j}\right) + p'_{\lambda_n}\left(|\beta_{nj}^0|\right)|\beta_{0j}|\right]\right]\right] \\
&\geq \left[\frac{1}{n}\sum_{i=1}^{n}\left[\rho_\tau\left(\varepsilon_i - \alpha_n \mathrm{u}^{\mathrm{T}} x_i\right) - \varepsilon_i\right]\right] - \sum_{j=1}^{k_n}\gamma_n\left(\alpha_n \mathrm{u}_j\right) - \sum_{j=1}^{k_n}\left[p'_{\lambda_n}\left(|\beta_{nj}^0|\right)|\alpha_n \mathrm{u}_j|\right] \\
&\triangleq I_{n1} + I_{n2} + I_{n3}.
\end{aligned}$$

According to Knight (1998), it holds that for $x \neq 0$ we have

$$|x-y| - |x| = -y\left[I(x>0) - I(x<0)\right] + 2\int_0^y \left[I(x \leq s) - I(x \leq 0)\right] ds,$$

Applying this equation, we can expand

$$|x-y| - |x| = -y\left\{\left[I(x>0) - I(x<0)\right] + 2\int_0^{\left(\sqrt{p_n/n}+a_n\right)\rho_\tau u^{\mathrm{T}} x_i}\left[I(x \leq s) - I(x \leq 0)\right]\right\} ds$$

First, we consider the term I_{n1}, we have

$$I_{n1} = -\frac{1}{n}\sum_{i=1}^{n}\rho_\tau\alpha_n u^{\mathrm{T}}x_i\left[I\left(\varepsilon_i>0\right)-I\left(\varepsilon_i<0\right)\right]$$
$$+\frac{2}{n}\sum_{i=1}^{n}\int_0^{\left(\sqrt{p_n/n}+a_n\right)\rho_\tau u^{\mathrm{T}}x_i}\left[I\left(\varepsilon_i\le s\right)-I\left(\varepsilon_i\le 0\right)\right]ds$$
$$\triangleq \mathrm{I}_{n_{11}}+\mathrm{I}_{n_{12}}$$

By condition C(4) using central limit theorem, $I_{n_{11}}$ converges in distribution then $u^{\mathrm{T}}x_i$, x_i is a p-dimensional random vector with mean zero and $Cov(x_i)=\sum$ and if I_{n3} converges to real function u in probability denoted by cumulative distribution function of ε_i then, we obain

$$\int_0^{\rho_\tau\alpha_n\mathrm{u}^{\mathrm{T}}x_i}\left[I\left(\varepsilon_i>0\right)-I\left(\varepsilon_i<0\right)\right]ds,$$

and

$$Var\left(\mathrm{I}_{n_{11}}\right)=\frac{\alpha_n^2}{n^2}\rho_\tau^2\sum_{i=1}^{n}\mathrm{u}^{\mathrm{T}}E\left(x_ix_i^{\mathrm{T}}\right)\mathrm{u}E\left[I\left(\varepsilon_i>0\right)-I\left(\varepsilon_i<0\right)\right]^2$$
$$=\frac{\alpha_n^2}{n}\rho_\tau^2\mathrm{u}^{\mathrm{T}}E\left(x_ix_i^{\mathrm{T}}\right)\mathrm{u}\le\frac{\alpha_n^2}{n}\rho_\tau^2\rho_{n2}\parallel\mathrm{u}\parallel^2.$$

By Markov Inequality, $\Pr\left(\mid \mathrm{I}_{n_{11}}\mid\ge K\alpha_n^2\rho_\tau^2\right)\le\frac{E\left(\mathrm{I}_{n_{11}}^2\right)}{K^2\rho_\tau^4\alpha_n^4}\le\frac{\left(\alpha_n^2\rho_\tau^2/n\right)\rho_{n2}\parallel\mathrm{u}\parallel^2}{K^2\rho_\tau^4\alpha_n^4}\to 0,$

which implies $\mathrm{I}_{n_{11}}=o_P\left(\alpha_n^2\right)\rho_\tau^2$. Considering the second term of I_{n1} which gives $I_{n_{12}}$, if its converges to a real function u then we get,

$$\int_0^{\rho_\tau\alpha_n\mathrm{u}^{\mathrm{T}}x_i}\left[I\left(\varepsilon_i>0\right)-I\left(\varepsilon_i<0\right)\right]ds \quad by \quad Z_{ni}(\mathrm{u}u), \text{ hence,}$$
$$\mathrm{I}_{n_{12}}\triangleq\sum_{i=1}^{n}Z_{ni}=\sum_{i=1}^{n}\left(Z_{ni}-E\left(Z_{ni}\right)\right)+\sum_{i=1}^{n}E\left(Z_{ni}\right)\triangleq\mathrm{I}_{n_{13}}+\mathrm{I}_{n_{14}}$$

Since

$$\frac{4\rho_\tau}{n^2}E\left[Z_{ni}^2(u)\left[I\left(\varepsilon_i>0\right)-I\left(\varepsilon_i<0\right)ds\right]^2\right]\le\frac{4}{n}E\left[\int_0^{\rho_\tau\alpha_n|u^{\mathrm{T}}x_i|}ds\right]^2$$
$$=\frac{4}{n}E\left[\rho_\tau^2\alpha_n^2\mid\mathrm{u}^{\mathrm{T}}x_i\mid^2\right]$$
$$\le\rho_\tau^2\alpha_n^2\frac{4}{n}\mathrm{u}^{\mathrm{T}}E\left(x_ix_i^{\mathrm{T}}\right)\mathrm{u}\le\rho_\tau^2\alpha_n^2\frac{4}{n}\rho_{n2}\parallel\mathrm{u}\parallel^2,$$

Given that there exist a constant function f, $\forall\eta>0$ and $0<k<\infty$, therefore $\sup_{|x|<\eta}f(x)<f(0)+k$, then we have

$$= \frac{4}{n}\eta E\left[Z_{ni}^2(\mathrm{u}) I\left(\rho_\tau \alpha_n \mid \mathrm{u}^{\mathrm{T}} x_i \mid\right) < \eta \right]$$
$$\leq \frac{4}{n}\eta \left\{ \int_0^{\rho_\tau \alpha_n \mid \mathrm{u}^{\mathrm{T}} x_i \mid} \left[I\left(\varepsilon_i \leq s\right) - I\left(\varepsilon_i \leq 0\right) \right] ds \; I\left(\rho_\tau \alpha_n \mid u^{\mathrm{T}} x_i \mid\right) < \eta \right\}$$
$$\leq \frac{4}{n}\eta \left[f(0) + k \right] E\left\{ \int_0^{\rho_\tau \alpha_n \mid \mathrm{u}^{\mathrm{T}} x_i \mid} s ds \; I\left(\rho_\tau \alpha_n \mid \mathrm{u}^{\mathrm{T}} x_i \mid\right) < \eta \right\}$$
$$\leq \frac{4}{n}\eta \left[f(0) + k \right] E\left[\rho_\tau^2 \alpha_n^2 \mid \mathrm{u}^{\mathrm{T}} x_i \mid^2 \right]$$

From the dominated convergence theorem, when given expression converges to zero as $\eta \to 0$ then it follows that as $n \to \infty$, $Var\left(\sum_{i=1}^n Z_{ni}\right) = \sum_{i=1}^n Var\left(Z_{ni}\right) \leq \frac{4}{n} Z_{ni}^2(\mathrm{u}) \to 0$, we have, $\sum_{i=1}^n \left[Z_{ni}(\mathrm{u}) - E\left(Z_{ni}(\mathrm{u})\right) \right] = op\left(\alpha_n^2\right)\rho_\tau^2$, then $\mathrm{I}_{n_{14}} = op\left(\alpha_n^2\right)\rho_\tau^2$.

By Markov Inequality, from conditions C(2)-C(3), we obtain

$$\Pr\left(\max_{1 \leq i \leq n} \left(\rho_\tau \alpha_n \mid \mathrm{u}^{\mathrm{T}} x_i \mid\right) > \eta^* \right) = \Pr\left(\bigcup_{i=1}^n \left(\rho_\tau \alpha_n \mid \mathrm{u}^{\mathrm{T}} x_i \mid > \eta^*\right) \right)$$
$$\leq n \Pr\left(\rho_\tau \alpha_n \mid \mathrm{u}^{\mathrm{T}} x_1 \mid > \eta^*\right) \leq n \frac{\alpha_n^6}{\eta_*^6} \rho_\tau^6 E\left(\mathrm{u}^{\mathrm{T}} x_1\right)^6 \leq \frac{p_n^3}{n^2} \frac{1}{\eta_*^6} \rho_\tau^6 C^6 p_n^3 \to 0,$$

which gives us, $\max_{1 \leq i \leq n}\left(\rho_\tau \alpha_n \mid \mathrm{u}^{\mathrm{T}} x_i \mid\right) = o_P(1)$.

$$\mathrm{I}_{n_{15}} = nZ_{ni} = 2E \int_0^{\rho_\tau \alpha_n u^{\mathrm{T}} x_i} \left[I\left(\varepsilon_i \leq s\right) - I\left(\varepsilon_i \leq 0\right) \right] ds$$
$$= 2E\left[E\left\{ \int_0^{\rho_\tau \alpha_n^{\mathrm{T}} x_i} \left[F(s) - F(0) \right] ds \right\} \right] = f(0) E\left(\left(1 + o(1)\right) \alpha_n^2 \rho_\tau \mathrm{u}^{\mathrm{T}} E\left(x_i x_i^{\mathrm{T}} \right) \right) \mathrm{u}.$$

Consider I_{n2}, $\| u \|$ is large enough, $I_{n1} < 0$, we have

$$\mid \mathrm{I}_{n2} \mid = \mid \sum_{j=1}^{k_n} \gamma_n \alpha_n u_j \mid = \alpha_n^2 \frac{\gamma_n}{\alpha_n} \| \mathrm{u} \| \leq Op\left(\alpha_n^2\right) \| \mathrm{u} \|, \text{ for } \mathrm{I}_{n3}, \text{ we have}$$
$$\mid \mathrm{I}_{n3} \mid = \mid \sum_{j=1}^{k_n} \alpha_n p'_{\lambda_n}\left(\mid \beta_{nj}^0 \mid \right) sgn\left(\mid \beta_{nj} \mid \right) \alpha_n \mathrm{u}_j \mid \triangleq \sum_{j=1}^{k_n} \mid \alpha_n p'_{\lambda_n}\left(\mid \beta_{nj}^0 \mid \right) sqn\left(\mid \beta_{nj} \mid \right) \alpha_n \mathrm{u}_j \mid$$
$$\geq \sqrt{k_n} \alpha_n a_n \| \mathrm{u} \| \geq \alpha_n^2 \| \mathrm{u} \|$$

From condition C(5), we have $\mathrm{I}_{n3} = Op\left(\alpha_n^2\right) \| \mathrm{u} \|$. It can be followed from the law of large numbers that $\sum_{i=1}^n Z_{ni}(\mathrm{u}) \to p$, $" \to p"$ convergence in probability then the right

hand side of I_{n1} converges to $f(0)(1+o(1))\alpha_n^2\rho_\tau^2\sum u$ in probability. By condition C(5), we get

$$\Pr\left(\sqrt{n}\alpha_n p'_{\lambda_n}\left(|\beta_{nj}^0| sgn\left(\beta_{nj}\right)u_j\right) > \eta^*\right) \le \Pr\left(|\beta_{nj}^0| I\left(|\beta_{nj}|\right)u_j < a\lambda_n\right),$$

for $\forall\eta^* > 0$, we can have $\sqrt{n}p'_{\lambda_n}\left(|\beta_{nj}^0| u_j\right) = op(1)$. By condition C(1), choosing large value C, which is uniform then $\|u\| = C$ and $D_n(u)$ converges to 0, hence the terms I_{n1} and I_{n2} are dominated by I_{n3} which is positive, therefore this satisfies our proof.

To facilitate the proof of the Theorem 2.2.2, the following Lemma 4.1 is needed, which indicates that under certain conditions, the proposed estimator has the sparsity property, i.e. the insignificant variables can exactly be estimated by zero with probability approaches to one.

Lemma 4.1:

Under conditions (C1) to (C5), the estimator $\hat{\beta}_n = \left(\hat{\beta}_{1n}^T, \hat{\beta}_{2n}^T\right)^T$ satisfies $\Pr\left(\hat{\beta}_{2n} = 0\right) \to 1.$

Proof:

By following Theorem 2.2.1, for a sufficiently large C, $\hat{\beta}_n$ lies in the ball $\left\{\beta : \|\beta - \beta_0\| \le \alpha_n C\right\}$ with probability converging to 1, where $\alpha_n = \sqrt{p_n/n} + a_n$. Let $\beta_{1n} = \beta_{10} + \alpha_n u_1$ and $\beta_{2n} = \beta_{20} + \alpha_n u_2 = \alpha_n u_2$, where $\|u\|^2 = \|u_1\|^2 + \|u_2\|^2 \le C^2$, we obtain $S_j\left(\beta_n\right) = -\frac{1}{n}\sum_{i=1}^n X_{ij}\rho_\tau\left(Y_i - \beta_0^T x_i\right)$ for $j \le p_n$, where $S_j\left(\hat{\beta}_n\right) = 0$ satisfies the oracle estimator. $\Pr\left(|\beta_{nj}^0| \ge a\lambda_n\right)$ for $\left(j = 1, 2, \cdots, p_n\right) \to 1$. Let $H_n\left(u_1, u_2\right)$, given $\|u\| \le C$ and if $\|u_2\| > 0$ then the minimizer $H_n\left(u_1, u_2\right)$ over $\|u\| \le C$ converges to 0, hence $H_n\left(u_1, u_2\right) - H_n\left(u_1, 0\right) > 0$ with probability converging to 1. Note That,

$$\begin{aligned}
&H_n\left(u_1, u_2\right) - H_n\left(u_1, 0\right) \\
&= Q_n\left(\beta_{1n}, \beta_{2n}\right) - Q_n\left(\beta_{10}, 0\right) - \left(Q_n\left(\beta_{1n}, 0\right) - Q_n\left(\beta_{10}, 0\right)\right) \\
&= \left\{\frac{1}{n}\sum_{i=1}^n \rho_\tau\left(\varepsilon_i - \alpha_n u_1^T w_i - \alpha_n u_2^T z_i\right) - \frac{1}{n}\sum_{i=1}^n \rho_\tau\left(\varepsilon_i\right)\right\} \\
&\quad - \left\{\frac{1}{n}\sum_{i=1}^n \rho_\tau\left(\varepsilon_i - \alpha_n u_1^T w_i\right) - \frac{1}{n}\sum_{i=1}^n \rho_\tau\left(\varepsilon_i\right)\right\} + \sum_{j=1}^{m_n}\left[\gamma_n\beta_{nj} + P'_{\lambda_n}\left(|\beta_{nj}^0|\right)\right]|\alpha_n u_{2j}|.
\end{aligned}$$

By condition C(4), we have $\Pr\left(|\,z_i\,|>t\right)=op\left(\alpha_n^2\right)\rho_\tau^2$ for $\eta>0$ hence $\Pr\left(|\,z_i\,|>\eta\left(\sqrt{p_n/n}+a_n\right)\right)$, for some, $i=1,2,\cdots,p_n$

$$\le\sum_{i=1}^{p_n}\left(|\,z_i\,|>\eta\left(\alpha_n^2\right)\rho_\tau^2\right)=\frac{1}{\eta}\,p_n\left(\alpha_n^2\right)\rho_\tau^2\le\frac{1}{\eta}\lambda_n^2\rho_\tau^2\to 0$$

$$\left\{\frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(\varepsilon_i-\alpha_n u_1^{\mathrm{T}}w_i-\alpha_n u_2^{\mathrm{T}}z_i\right)-\frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(\varepsilon_i\right)\right\}$$

$$-\left\{\frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(\varepsilon_i-\alpha_n u_1^{\mathrm{T}}w_i\right)-\frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(\varepsilon_i\right)\right\}$$

$$\ge O_P\left(\alpha_n^2\right)\rho_\tau+\frac{f(0)}{2}\alpha_n^2\rho_\tau\rho_{n1}\,\|\,u\,\|^2+O_P(\alpha_n^2)\rho_\tau-\frac{3f(0)}{2}\alpha_n^2\rho_\tau\tau_{n2}\,\|\,u_1\,\|^2$$

Note that

$$\frac{\gamma_n\beta_{nj}+P'_{\lambda_n}\left(|\,\beta_{nj}^0\,|\right)}{\gamma_n\lambda_n}$$

$$=\left(\frac{\gamma_n\beta_{nj}+P'_{\lambda_n}\left(|\,\beta_{nj}^0\,|\right)}{\gamma_n\lambda_n}I_n\left(|\,\beta_{nj}^0\,|\right)\le\gamma_n\lambda_n\right)$$

$$+\left(\frac{\gamma_n\beta_{nj}+P'_{\lambda_n}\left(|\,\beta_{nj}^0\,|\right)}{\gamma_n\lambda_n}I_n\left(|\,\beta_{nj}^0\,|\right)>\gamma_n\lambda_n\right)$$

$$=1+\left(\frac{\gamma_n\beta_{nj}+P'_{\lambda_n}\left(|\,\beta_{nj}^0\,|\right)}{\gamma_n\lambda_n}I_n\left(|\,\beta_{nj}^0\,|\right)>\gamma_n\lambda_n\right).$$

For $j=k_n+1,\cdots,p_n$ we have $\sqrt{p_n/n}+a_n\,|\,\beta_{nj}^0\,|=Op(1)$. By condition C(5), and for $\forall\eta>0$,

$$\Pr\left(\frac{\gamma_n\beta_{nj}+P'_{\lambda_n}\left(|\,\beta_{nj}^0\,|\right)}{\gamma_n\lambda_n}-1>\eta\right)\le\Pr\left(\left(\beta_{nj}+|\,\beta_{nj}^0\,|\right)>\gamma_n\lambda_n\right)$$

$$=\Pr\left(\left(\sqrt{p_n/n}+a_n\right)\left(\beta_{nj}+|\,\beta_{nj}^0\,|\right)>\sqrt{p_n/n}\gamma_n\lambda_n\right)\to 0.$$

Namely, $\dfrac{\gamma_n\beta_{nj}+P'_{\lambda_n}\left(|\,\beta_{nj}^0\,|\right)}{\gamma_n\lambda_n}\xrightarrow{\;p\;}1.$

Hence, we have

$$H_n(u_1,u_2)-H_n(u_1,0)\geq \lambda_n\gamma_n\alpha_n\left[op\left(\frac{\sqrt{p_n/n}}{\lambda_n\gamma_n}\right)\right.$$
$$\left.+\frac{\sqrt{p_n/n}}{\lambda_n\gamma_n}f(0)\left(\frac{\rho_1}{2}-\frac{3\tau_2}{2}\right)\|\mathrm{u}\|^2+1+op(1)\sum_{j=1}^{m_n}|\mathrm{u}_{2j}|\right].$$

By conditions C1 and C5, the results follows.

Proof of Theorem 2.2.2:

Part (1) holds through above Lemma 4.1, to prove part (2), assume that mean function $F(u)$ is a two order continuous and derivable function and a sequence $\{E_n r(.,\beta)\}$ is stochastically continuous at β_0, if the Ridge-SCAD estimator posses sparsity property and the same asymptotic distribution then $\{E_n r(.,\beta)\} = E_n r\left(\hat{\theta}_n\right) = op\left(\frac{1}{\sqrt{p_n}}\right)$.

Given $S_n(\beta_n)=\frac{1}{n}\sum_{i=1}^{n}\left\{\tau\left(Y_i-F\left(\beta_n^{\mathrm{T}}\mathrm{x}_i\right)\right)+(1-\tau)\left(Y_i-F\left(\beta_n^{\mathrm{T}}\mathrm{x}_i\right)\right)\right\}$, it has its minimum $\hat{\beta}_n$ and the vector function $\Delta(.)$ component that belongs to a given penalty, we have

$$\Delta(.)=\left\{\tau\left[\mathrm{sgn}(\varepsilon)>0\right]+(1-\tau)\left[\mathrm{sgn}(\varepsilon)<0\right]\right\}F'\left(\beta_n^{\mathrm{T}}x_i\right)x_i, \tag{6}$$

since β_0 is the unique minimal value point of $S_n(\beta_n)$, therefore $H'(0)=0$ and $H''(0)>0$. If β and x are bounded, $\exists$ positive constant M such that $\beta_n^{\mathrm{T}}\mathrm{x}_i\in[-M,M]$, $F(u)$ is bounded on $[-M,M]$ by dominated convergence theorem, we differentiate under integral sign for $S_n(\beta_n)$, if the regression error $\{\varepsilon_i\}$ are independent and identically distributed with τ^{th} quantile zero and continuous positive density $f(.)$ in a neighborhood zero and that there exist a finite third derivative in relation to dominated convergence theorem, we have

$$S'(\theta_n)=-E\left[H'\left(\frac{F\left(\beta_n^{\mathrm{T}}x\right)-F\left(\beta_0^{\mathrm{T}}x\right)}{V(\beta_0^{\mathrm{T}}x)}\right)F'\left((\beta_n^{\mathrm{T}}x\right)x\right]$$
$$=-E\left(H'\left(\theta_n^{\mathrm{T}}\right)\mathrm{w}\right),\ S'(\theta_0)=E\left[H'(0)F'\left(\beta_0^{\mathrm{T}}x\right)x\right]=Op\times 1,$$

$$S''(\theta_n)=E\left((H''\left(\theta_n^{\mathrm{T}}w\right)ww^{\mathrm{T}}\right)\text{ and }\frac{\partial S^{(3)}(\theta_n)}{\partial\theta_{ni}\partial\theta_{nj}\partial\theta_{nk}}=E\left(H^{(3)}\left(\theta_n^{\mathrm{T}}w\right)w_iw_jw_k\right).$$

Some simple calculations show that $S'(0) = 0$ and $S''(0) = 2f(0)\Sigma_{n,11}$. By the Taylor expansion, we have

$$S\left(\hat{\theta}_n\right) = S(0) + S'(0)^{\mathrm{T}}\hat{\theta}_n + \frac{1}{2}\hat{\theta}_n^T S''(0)\hat{\theta}_n + \frac{1}{6}\sum_{i=1}^{k_n}\sum_{j=1}^{k_n}\sum_{k=1}^{k_n} E\left(H^{(3)}\left(\theta_n^{*\mathrm{T}} w\right) w_i w_j w_k\right)\hat{\theta}_{ni}\hat{\theta}_{nj}\hat{\theta}_{nk}. \quad (7)$$

where θ_n^* is between 0 and $\hat{\theta}_n$. Consider the fourth term of (7) being Z, we show that $Z = o_P\left(\frac{1}{n}\right)$. By Caucky-Schwarz inequality and conditions (C2) and (C3), we obtain

$$Z^2 \le \left[\sum_{i=1}^{k_n}\sum_{j=1}^{k_n}\sum_{k=1}^{k_n}\left(E\left(H^{(3)}\left(\theta_n^{*\mathrm{T}}\mathrm{w}\right)\mathrm{w}_i\mathrm{w}_j\mathrm{w}_k\right)\right)^2\right] \times \sum_{i=1}^{k_n}\sum_{j=1}^{k_n}\sum_{k=1}^{k_n}\left(\hat{\theta}_{ni}\hat{\theta}_{nj}\hat{\theta}_{nk}\right)^2$$

$$\le Mk_n^3 \| \hat{\theta}_n \|^6 \le O_P\left(\frac{k_n^3 p_n^3}{n}\cdot\frac{1}{n^2}\right) = o_P\left(\frac{1}{n^2}\right).$$

If V is an identity matrix then, we have

$$V = S''(\beta_0) = H''(0)E\left[\frac{\left(F'\left(\beta_0^{\mathrm{T}}x\right)\right)^2}{V\left(\beta_0^{\mathrm{T}}x\right)}xx^{\mathrm{T}}\right] = f_\varepsilon(0)E\left[\left(\frac{F'\left(\beta_0^{\mathrm{T}}x\right)}{\sqrt{V\left(\beta_0^{\mathrm{T}}x\right)}}x\right)\left(\frac{F'\left(\beta_0^{\mathrm{T}}x\right)}{\sqrt{V\left(\beta_0^{\mathrm{T}}x\right)}}x\right)^{\mathrm{T}}\right],$$

where $f_\varepsilon(0) > 0,$ we justify that

$$E_n(\Delta(.)) = E\left[\left\{\tau\left[\mathrm{sgn}(\varepsilon) > 0\right] + (1-\tau)\left[\mathrm{sgn}(\varepsilon) < 0\right]\right\}F'\left(\beta_n^{\mathrm{T}}x_i\right)x_i\right]$$

$$= E\left[F'\left(\beta_0^{\mathrm{T}}x\right)xE\left\{\tau\left[\mathrm{sgn}(\varepsilon) > 0\right] + (1-\tau)\left[\mathrm{sgn}(\varepsilon) < 0\right]\right\}\right] \quad (8)$$

If the density function of error term ε is f then $f_\varepsilon(\mathrm{u})$, we have

$$F_n(\mathrm{u}) - F_n(0) = \Pr\left(f(.,\mathrm{u}) - f(.,0)\right) + n^{-\frac{1}{2}}E_n\Delta(.)\left[f(.,\mathrm{u}) - f(.,0)\right]$$

$$= \frac{1}{2}|\mathrm{u}|^2 + \rho\left(|\mathrm{u}|^2\right) + n^{-\frac{1}{2}}\mathrm{u}^{\mathrm{T}}E_n\Delta(.) + n^{-\frac{1}{2}}|\mathrm{u}|S_n\left(\beta_n\right)$$

if $\{\tau_n\}$ is a sequence of random vectors converging in probability to the value u then $F_n(\tau_n)$ comes within $Op\left(n^{-1}\right)$ then we have $Op\left(n^{-1}\right) \ge F_n(\tau_n) - F_n(0)$,

$$= \frac{1}{2}|\tau_n|^2 + Op\left(|\tau_n|^2\right) + n^{-\frac{1}{2}}\tau_n E_n\Delta(.) + Op\left(n^{-\frac{1}{2}}|\tau_n|\right)$$

Given the empirical process $E_n(.) = \sqrt{n}\left(S_n(.) - S(.)\right)$, for convenience given that

$$\theta_n = \beta_{1n} - \beta_{10} = \frac{F\left(\beta_n^{\mathrm{T}} x\right) - F\left(\beta_0^{\mathrm{T}} x\right)}{V\left(\beta_0^{\mathrm{T}} x\right)} \text{ and } \hat{\theta}_n = \hat{\beta}_{1n} - \hat{\beta}_{10} = \hat{\beta}_n^{\mathrm{T}} x - \hat{\beta}_0^{\mathrm{T}} x,$$

As per Lemma 4.1, we can conclude that with probability tending to 1, $Q_n\left(\hat{\beta}_n\right) = Q_n^*\left(\hat{\theta}_n\right)$. Note that $g\left(\theta_n\right) = \left(Y_i - \beta_n^T x\right)$, then the linear approximation of $g\left(\theta_n\right)$ near 0 is $g\left(\theta_n\right) = g(0) + \theta_n^{\mathrm{T}} E_n \mathbb{D}(0) + \| \theta_n \| r\left(\theta_n\right)$, where $r\left(\theta_n\right)$ is the remainder term.

$$\begin{aligned} Q_n^*(\hat{\theta}_n) - Q_n^*(0) &= S_n\left(\hat{\theta}_n\right) - S_n(0) + \sum_{j=1}^{k_n}\left(\gamma_n \beta_{nj} + P'_{\lambda_n}\left(|\beta_{nj}^0|\right)\right)\left(|\left(\hat{\theta}_{nj} + \beta_{10j}\right) - \beta_{10j}|\right) \\ &= \frac{1}{\sqrt{n}} \tau_n E_n \Delta(0)\left(g\left(\hat{\theta}_n\right) - g(0)\right) + \left(S\left(\hat{\theta}_n\right) - S(0)\right) \\ &\quad + \sum_{j=1}^{k_n}\left(\gamma_n \beta_{nj} + P'_{\lambda_n}\left(|\beta_{nj}^0|\right)\right)\left(|\left(\hat{\theta}_{nj} + \beta_{10j}\right) - \beta_{10j}|\right) \end{aligned} \tag{9}$$

By considering the third term of (9), using the differential mean value theorem, we get

$$\begin{aligned} &\left| \sum_{j=1}^{k_n}\left[p'_{\lambda_n}\left(|\beta_{nj}^0|\right)|\beta_{nj}| + \gamma_n \beta_{nj}\right]\left(\left(\hat{\theta}_{nj} + \beta_{10j}\right) - \beta_{10j}\right)\right| \\ &\quad \le \max_{1 \le j \le k_n}\left(p'_{\lambda_n}\left(|\beta_{nj}^0|\right) + \gamma_n \beta_{nj}\right)\left|\sum_{j=1}^{k_n} \operatorname{sgn}\left(\xi_i\right)\hat{\theta}_{nj}\right| \\ &\quad \le \max_{1 \le j \le k_n}\left(p'_{\lambda_n}\left(|\beta_{nj}^0|\right) + \gamma_n \beta_{nj}\right)\sqrt{k_n}\left\|\hat{\theta}_{nj}\right\| = o_P\left(n^{-\frac{1}{2}}\right) \end{aligned}$$

If $\frac{p'\lambda_n(\theta)}{\lambda_n} > 0$ as $n \to \infty$, and $\sqrt{n}\lambda_n \to \infty$ then for any β that satisfies the $\| \beta - \beta_0 \| = o\left(n^{-\frac{1}{2}}\right)$, where ξ_i is between β_{10j} and $\hat{\theta}_{nj} + \beta_{10j}$, from (7), we have

$$E\left[\left\{\tau\left[sgn(\varepsilon) > 0\right] + (1-\tau)\left[sgn(\varepsilon) < 0\right]\right\}\right] = \int_{-\infty}^{0}(\tau - 1) f_\varepsilon(u) d(u) + \tau\int_0^{+\infty} f_\varepsilon(u) d(u) = 0$$

hence $E_n \Delta(.) = Op \times (1)$, assumes that $E_n\left(r\left(\hat{\theta}_n\right)\right) = Op\left(\frac{1}{\sqrt{p_n}}\right)$, from (9), we have

$$Q_n^*\left(\hat{\theta}_n\right)-Q_n^*(0) = \frac{1}{\sqrt{n}}\hat{\theta}_n^{\mathrm{T}}\rho_\tau E_n\left(\Delta(0)\right)+\frac{1}{\sqrt{n}}\|\hat{\theta}_n\|\rho_\tau E_n\left(r(\hat{\theta}_n)\right)$$
$$+\frac{1}{2}\hat{\theta}_n^{\mathrm{T}}\rho_\tau\left(2f(0)\Sigma_{n,11}\right)\hat{\theta}_n+o_P\left(n^{-\frac{1}{2}}\right)+o_P\left(n^{-\frac{1}{2}}\right)$$
$$=\frac{1}{2}\left(2f(0)\right)\|\Sigma_{n,11}^{1/2}\hat{\theta}_n\rho_\tau+\frac{1}{\sqrt{n}}\left(2f(0)\right)^{-1}\Sigma_{n,11}^{-1/2}\rho_\tau E_n\left(\Delta(0)\right)\|^2$$
$$-\frac{1}{2}\left(2f(0)\right)\|\frac{1}{\sqrt{n}}\left(2f(0)\right)^{-1}\Sigma_{n,11}^{-1/2}\rho_\tau E_n\left(\Delta(0)\right)\|^2+o_P\left(n^{-\frac{1}{2}}\right). \quad (10)$$

In particular,

$$Q_n^*\left(-\frac{1}{\sqrt{n}}\left(2f(0)\right)^{-1}\Sigma_{n,11}^{-1}\rho_\tau E_n\left(\Delta(0)\right)\right)-Q_n^*(0)$$
$$=-\frac{1}{2}\left(2f(0)\right)\|\frac{1}{\sqrt{n}}\left(2f(0)\right)^{-1}\rho_\tau\Sigma_{n,11}^{-1/2}E_n\left(\Delta(0)\right)\|^2+o_P\left(n^{-\frac{1}{2}}\right). \quad (11)$$

Here we need to justify $\|\frac{1}{\sqrt{n}}\left(2f(0)\right)^{-1}\Sigma_{n,11}^{-1}\rho_\tau E_n\left(\Delta(0)\right)\|=O_P\left(\sqrt{p_n}/\sqrt{n}\right)$, in fact,

$$\Pr\left(\|\frac{1}{\sqrt{n}}\left(2f(0)\right)^{-1}\Sigma_{n,11}^{-1}\rho_\tau E_n\left(\Delta(0)\right)\|\geq L\sqrt{p_n}/\sqrt{n}\right)$$
$$\leq\frac{(2f(0)\tau_{n1})^{-2}E\|\rho_\tau E_n(\Delta(0))\|^2}{L^2p_n}\leq\frac{m_n(2f(0)\tau_1)^{-2}\rho_\tau K}{L^2p_n}\to 0.$$

Subtracting (11) from (10), we have

$$\|\Sigma_{n,11}^{1/2}\rho_\tau\hat{\theta}_n+\frac{1}{\sqrt{n}}\left(2f(0)\right)^{-1}\Sigma_{n,11}^{-1/2}\rho_\tau E_n\left(\Delta(0)\right)\|^2=o_P\left(n^{-\frac{1}{2}}\right).$$

From Cauchy-Schwarz inequality and conditions $(C1)$ and $(C3)$,

$$\tau_n E_n\Delta(0)\geq-|\tau_n|Op\left(n^{-\frac{1}{2}}\right) \text{ where}$$

$$Op\left(n^{-\frac{1}{2}}\right)\geq\left[\frac{1}{2}-Op(1)\right]|\tau_n|^2-n^{-\frac{1}{2}}|\tau_n|Op(1)-Op\left(n^{-\frac{1}{2}}\right)|\tau_n|$$
$$=\left[\frac{1}{2}-Op(1)\right]\left[|\tau_n|-Op\left(n^{-\frac{1}{2}}\right)\right]^2-Op\left(n^{-\frac{1}{2}}\right),$$

It follows that the square term is at $Op\left(n^{-\frac{1}{2}}\right)$ and hence $\tau_n = Op\left(n^{-\frac{1}{2}}\right)$. From conditions $C(1)-C(3)$, we have,

$$
\begin{aligned}
&n^{1/2}\alpha^{\mathrm{T}}\rho_{\tau}\Sigma_{n,11}^{1/2}\left(\hat{\beta}_{1n}-\beta_{10}\right)\\
&=E\left[\left\{\tau\left[\operatorname{sgn}(\varepsilon)>0\right]+(1-\tau)\left(\operatorname{sgn}(\varepsilon)<0\right)\right\}\right]^2\left(\left[F'\left(\beta_0^{\mathrm{T}}x\right)\right]^2 xx^{\mathrm{T}}\right)\\
&=E\left[\tau^2\left[\operatorname{sgn}(\varepsilon)>0\right]^2\right]\left(\left[F'\left(\beta_0^{\mathrm{T}}x\right)\right]^2 xx^{\mathrm{T}}\right)\\
&\qquad+E\left[(1-\tau)^2\left[\operatorname{sgn}(\varepsilon)<0\right]^2\right]\left(\left[F'\left(\beta_0^{\mathrm{T}}x\right)\right]^2 xx^{\mathrm{T}}\right)\\
&=E\left\{\tau^2\left[F'\left(\beta_0^{\mathrm{T}}x\right)\right]^2 xx^{\mathrm{T}}\int_0^{+\infty}f_{\varepsilon}(u)du\right\}\\
&\qquad+E\left\{(1-\tau)^2\left[F'\left(\beta_0^{\mathrm{T}}x\right)\right]^2 xx^{\mathrm{T}}\int_{-\infty}^{0}f_{\varepsilon}(u)du\right\}\\
&=E\left\{\tau^2\left[F'\left(\beta_0^{\mathrm{T}}x\right)\right]^2 xx^{\mathrm{T}}\left[1-F_{\varepsilon}(0)\right]\right\}+E\left\{(1-\tau)^2\left[F'\left(\beta_0^{\mathrm{T}}x\right)\right]^2 xx^{\mathrm{T}}\left[1-F_{\varepsilon}(0)\right]\right\}\\
&=\left\{\tau^2(1-\tau)+(1-\tau)^2\tau\right\}E\left\{\left[F'\left(\beta_0^{\mathrm{T}}x\right)x\right]\left[F'\left(\beta_0^{\mathrm{T}}x\right)x\right]^{\mathrm{T}}\right\}\\
&=\tau(1-\tau)E\left\{\left[F'\left(\beta_0^{\mathrm{T}}x\right)x\right]\left[F'\left(\beta_0^{\mathrm{T}}x\right)x\right]^{\mathrm{T}}\right\},
\end{aligned}
$$

Therefore it is given, $E\left\{\left[F'\left(\beta_0^{\mathrm{T}}x\right)x\right]\left[F'\left(\beta_0^{\mathrm{T}}x\right)x\right]^{\mathrm{T}}\right\}=v$, thus $\Delta(0)$ belongs to given penalty, so the condition hold, we obtain

$$
n^{1/2}\alpha^{\mathrm{T}}\rho_{\tau}\Sigma_{n,11}^{1/2}\left(\hat{\beta}_{1n}-\beta_{10}\right)=\tau(1-\tau)\mathrm{v}^{-1}=\tau(1-\tau)\left(f_{\varepsilon}(0)\right)^{-1}=\tau(1-\tau)\left(f^2(0)\right)^{-1}
$$

Hence, we have

$$
n^{1/2}\alpha^{\mathrm{T}}\rho_{\tau}\Sigma_{n,11}^{1/2}\left(\hat{\beta}_{1n}-\beta_{10}\right)\xrightarrow{D}N\left(0,\tau(1-\tau)\left(f(0)\right)^{-2}\right).
$$

5. DISCUSSION

In this paper, we study Ridge-SCAD combined penalization technique for variable selection, coefficient estimation and establish the asymptotic properties of its estimator when the number of covariates increases to infinity as $n\to 0$. Our theory also suggest

that noncovex combined penalized regression with diverging number of parameters enjoys oracle property under regularity conditions. Furthermore, we used RCS method to handle ultrahigh dimensional data. Simulation studies reveal that the performance of Qr.CP is good when there is high correlations among predictors. The real data analysis also demonstrated the practical aspects of this method.

REFERENCES

1. Belloni, A. and Chernozhukov, V. (2011). ℓ1-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics*, 39(1), 82-130.
2. Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *The Annals of Statistics*, 24(6), 2350-2383.
3. Donoho, D.L. and Johnstone, J.M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3), 425-455.
4. Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32(2), 407-499.
5. Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.*, 96(456), 1348-1360.
6. Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *The Annals of Statistics*, 32(3), 928-961.
7. Frank, L.E. and Friedman, J.H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, 35(2), 109-135.
8. Harrison Jr, D. and Rubinfeld, D.L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5(1), 81-102.
9. He, X. and Shao, Q.M. (2000). On parameters of increasing dimensions. *Journal of Multivariate Analysis*, 73(1), 120-135.
10. Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: Biased estimation for non-orthogonal problems. *Technometrics*, 12(1), 55-67.
11. Huang, J. and Xie, H. (2007). Asymptotic oracle properties of SCAD-penalized least squares estimators. *Lecture Notes-Monograph Series*, 55, 149-166.
12. Huang, J., Horowitz, J.L. and Ma, S. (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *The Annals of Statistics*, 36(2), 587-613.
13. Huber, P.J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *The Annals of Statistics*, 1, 799-821.
14. Jiang, L., Bondell, H.D. and Wang, H.J. (2014). Interquantile shrinkage and variable selection in quantile regression. *Computational Statistics and Data Analysis*, 69, 208-219.
15. Kim, Y., Choi, H. and Oh, H.-S. (2008). Smoothly clipped absolute deviation on high dimensions. *J. Amer.Statist. Assoc*., 103(484), 1665-1673.
16. Knight, K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *The Annals of Statistics*, 26(2), 755-770.
17. Knight, K. and Fu, W. (2000). Asymptotics for lasso-type estimators. *Annals of Statistics*, 28(5), 1356-1378.
18. Koenker, R. (2005). *Quantile Regression*: Cambridge university press.
19. Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, 46, 33-50.

20. Li, G., Peng, H. and Zhu, L. (2011). Nonconcave penalized M-estimation with a diverging number of parameters. *Statistica Sinica*, 21(1), 391.
21. Li, G., Peng, H., Zhang, J. and Zhu, L. (2012). Robust rank correlation based screening. *The Annals of Statistics*, 40(3), 1846-1877.
22. Ma, S. and Du, P. (2012). Variable selection in partly linear regression model with diverging dimensions for right censored data. *Statistica Sinica*, 22(3), 1003-1020.
23. Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, 7(02), 186-199.
24. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc.. Series B (Methodological)*, 58(1), 267-288.
25. Wang, L., Wu, Y. and Li, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *J. Amer. Statist. Assoc.*, 107(497), 214-222.
26. Wang, X., Park, T. and Carriere, K. (2010). Variable selection via combined penalization for high-dimensional data analysis. *Computational Statistics and Data Analysis*, 54(10), 2230-2243.
27. Wu, Y. and Liu, Y. (2009). Variable selection in quantile regression. *Statistica Sinica*, 19(2), 801.
28. Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. Roy. Statist. Soc.: Series B (Statistical Methodology)*, 67(2), 301-320.