Journal of Statistical Planning and Inference 143 (2013) 1029–1038

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

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

Adaptive penalized quantile regression for high dimensional data

Qi Zheng*, Colin Gallagher, K.B. Kulasekera

Department of Mathematical Sciences, Clemson University, Clemson, SC 29634-0975, United States

ARTICLE INFO

Article history:
Received 12 August 2012
Received in revised form
5 December 2012
Accepted 20 December 2012
Available online 2 January 2013

Keywords:
Adaptive
Quantile regression
Oracle rate
Asymptotic normality
Variable selection

ABSTRACT

We propose a new adaptive L_1 penalized quantile regression estimator for high-dimensional sparse regression models with heterogeneous error sequences. We show that under weaker conditions compared with alternative procedures, the adaptive L_1 quantile regression selects the true underlying model with probability converging to one, and the unique estimates of nonzero coefficients it provides have the same asymptotic normal distribution as the quantile estimator which uses only the covariates with non-zero impact on the response. Thus, the adaptive L_1 quantile regression enjoys oracle properties. We propose a completely data driven choice of the penalty level λ_n, which ensures good performance of the adaptive L_1 quantile regression. Extensive Monte Carlo simulation studies have been conducted to demonstrate the finite sample performance of the proposed method.

1. Introduction

Consider the high dimensional sparse regression model

$$y_i = \beta_0^* + \beta_1^* z_{i1} + \cdots + \beta_p^* z_{ip} + \epsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where $\{y_i\}$'s are random variables, $\{\mathbf{z}_i\}$'s are $p \times 1$ independent random covariate vectors, and $\{\epsilon_i\}$ are independent random error terms with $P(\epsilon_i \le 0 | \mathbf{z}_i) = \tau$ for some quantile index τ. We allow the dimension of the covariate vector to be very large, possibly of order $O(\exp(n^\alpha))$, for some constant $0 < \alpha < 1$; but the regression parameter β^* is sparse in the sense that only $s \ll p$ of its components are non-zero. Of interest is to identify the nonzero regressors and estimate their regression coefficients as well. Such models have attracted great attention due to the demand for data analysis created by many new applications arising in genetics, signal processing, machine learning, climate change point detection and other fields with high-dimensional data sets available.

Various methods have been developed to identify the unknown model and estimate the corresponding coefficients simultaneously for the high dimensional sparse model (see Fan and Peng, 2004; Huang et al., 2008a, 2008b), which mostly focus on the penalized least squares regression. Although some of them enjoy desirable oracle properties (Fan and Li, 2001), they generally require stringent moment assumptions (Cramér condition) on the unobservable homoscedastic random errors, $\{\epsilon_i\}$. Therefore, they are not robust and may not be applicable in practice. Compared with least squares, another important statistical method, quantile regression (Koenker and Basset, 1978), is robust and allows relaxation of moment conditions on the heterogeneous error sequence. The advantage of quantile regression goes beyond that: it can provide a more complete model of the relationship between predictors and response variables. (e.g. Koenker, 2005), it

* Corresponding author. Tel.: +1 8646562192.
E-mail addresses: qiz@clemson.edu (Q. Zheng), cgallag@clemson.edu (C. Gallagher), kk@clemson.edu (K.B. Kulasekera).

0378-3758/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2012.12.009

owns excellent computational properties. (e.g. Portnoy and Koenker, 1997), and it has widespread applications (e.g. Yu et al., 2003; Chernozhukov, 2005). Belloni and Chernozhukov (2011) integrate general quantile regression into an L_1 penalty framework for the high-dimensional sparse model. Another interesting estimator, the Dantzig selector, considered by Candes and Tao (2007), can be considered as a penalized median regression. However, both of these estimators achieve the $\sqrt{n/(s\log(p))}$ consistency rate, which is slower than the oracle rate $\sqrt{n/s}$ from He and Shao (2000). Wang et al. (2012) proposed a quantile regression with SCAD penalty. Since the objective function is not convex, the solutions are not unique. To our best knowledge, the desirable oracle properties have not been achieved by any penalized quantile regression for the high-dimensional sparse model.

In this paper we attempt to overcome the limitations of the existing quantile regression techniques by combining quantile regression with a fully adaptive L_1 penalty function to produce adaptive L_1 quantile regression, which can simultaneously select the model and provide a robust estimator possessing oracle properties. Exploiting the ideas of Wang et al. (2007) and Zou and Yuan (2006), we use the consistent estimator from Belloni and Chernozhukov (2011) to determine adaptive weights. Since we are using quantile loss functions, we do not require the Cramér condition on the error sequence. This paper's contributions are summarized as follows:

- First, we show that under mild conditions, the adaptive L_1 quantile regression will select the correct model with probability converging to 1, and for any quantile index in a compact set in (0, 1), the unique adaptive L_1 quantile regression estimates are consistent with the oracle rate $\sqrt{n/s}$. This is an advancement from the existing quantile regression methods for the high-dimensional sparse model.
- Second, any linear combination of the estimates is asymptotically normal with the same asymptotic variance as that of the oracle estimator.
- Third, in deriving the aforementioned oracle properties, we propose a new data-driven procedure to select the penalty level and show that it satisfies the requirements to achieve the oracle rate.

The rest of the paper is organized as follows. In Section 2, we define the adaptive L_1 quantile regression procedure. In Section 3, we study the asymptotic properties of the L_1 quantile regression estimator and discuss the choice of penalty level λ_n. Numerical studies are presented in Section 4. We give concluding remarks in Section 5, and relegate the technical proofs to Appendix.

2. The adaptive L_1 quantile regression

We start with introducing notations. We implicitly index all parameter values by the sample size n, but we omit the index whenever this does not cause confusion. We use the notation $a \vee b = \max\{a,b\}$ and $a \wedge b = \min\{a,b\}$. We denote the l_2-norm by $\|\cdot\|$, and the l_0-"norm" (the number of nonzero components) by $\|\cdot\|_0$. Given a vector $\delta \in \mathbb{R}^{p+1}$, and a set of indices $T \subset \{0,1,\ldots,p\}$, we denote by δ_T the vector in which $\delta_{Tj} = \delta_j$ if $j \in T$, $\delta_{Tj} = 0$ if $j \notin T$. And q^* is the τth quantile of ϵ.

In order to define the adaptive L_1 quantile regression, let us briefly review quantile regression and L_1 penalized quantile regression. Let $\mathbf{x}_i = (1,\mathbf{z}_i^T)^T$. Quantile regression estimator of β^* can be obtained by solving

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i^T\beta), \tag{2}$$

where $\rho_\tau(t) = \tau 1(t>0)t - (1-\tau)1(t \le 0)t$ is the check function.

Without loss of generality, we assume that the first $s+1$ elements of β^* are nonzero, and the rest are zero. For simplicity, write $\beta^* = (\beta_a^{*T}, \beta_b^{*T})^T$, where β_a^* is a $(s+1)\times 1$ vector and β_b^* is a $(p-s)\times 1$ vector of zeroes. Similarly, we decompose $\mathbf{x}_i$ as $(\mathbf{x}_{ia}^T, \mathbf{x}_{ib}^T)^T$.

Belloni and Chernozhukov (2011) proposed a penalized L_1 quantile regression estimator $\tilde{\beta}$, which minimizes

$$\tilde{Q}_\tau(\beta) = \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i^T\beta) + \frac{\lambda_n\sqrt{\tau(1-\tau)}}{n} \sum_{i=1}^{p} \hat{\sigma}_j|\beta_j|, \tag{3}$$

where $\hat{\sigma}_j = \sum_{i=1}^{n} x_{ij}^2/n, j = 1,\ldots,p$ and obeys $P(\max_{1\le j\le p}|\hat{\sigma}_j - 1| \le 1/2) \ge 1-\alpha \to 1$. Here λ_n is the penalty parameter. Ideally, a penalty function should be adaptive in the sense that it penalizes insignificant variables enough to force estimates of their regression coefficients to be zero, but does not overpenalize significant variables, so that the correct model can be identified and hence oracle properties can be attained. However, it can be seen that the penalty for each variable in (3) is of the same order, λ_n/n, and hence not quite adaptive. A similar issue appears in the estimator proposed by Candes and Tao (2007).

To improve the quantile regression for the high-dimensional sparse model, we attempt to assign fully adaptive weights to different variables and propose the adaptive L_1 quantile regression estimator β, which is a minimizer of the objective function

$$Q_\tau(\beta) = \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i^T\beta) + \lambda_n \sum_{j=1}^{p} \omega_j|\beta_j|, \tag{4}$$

where $\boldsymbol{\omega} \in \mathbb{R}^p$ is weights vector chosen to be $|\tilde{\beta}|^{-1} \wedge \sqrt{n}$, for any $\sqrt{n/(s\log(n \vee p))}$-consistent estimator $\tilde{\beta}$ of β^*. For example, we can take the estimator from Belloni and Chernozhukov (2011) as β, which under conditions A1–A3 given below will converge at a sufficiently fast rate. The formulation (4) includes the LAD-Lasso proposed by Wang et al. (2007) as a special case that the dimensionality p is fixed.

3. Asymptotic properties

In this section, we state primitive regularity conditions and then establish the asymptotic properties of the adaptive L_1 quantile regression estimator.

3.1. Regularity conditions

The following regularity conditions are assumed throughout the rest of this paper.

A1 (Sampling and smoothness). For any value $\mathbf{x}$ in the support of $\mathbf{x}_i$, the conditional density $f_{\epsilon|\mathbf{z}}(\epsilon|\mathbf{z})$ is continuously differentiable at each $y \in \mathbb{R}$, and $f_{\epsilon|\mathbf{x}}(\epsilon|\mathbf{x})$ and $\partial/\partial\epsilon f_{\epsilon|\mathbf{x}}(\epsilon|\mathbf{x})$ are bounded in absolute value by constants $\overline{f}$ and $\overline{f'}$ uniformly in $\epsilon \in \mathbb{R}$ and $\mathbf{x}$ in the support of $\mathbf{x}_i$. Moreover, the conditional density of $\epsilon|\mathbf{x}$ evaluated at the conditional quantile $q^*_{\mathbf{x}}$ is bounded away from 0 uniformly for any $\mathbf{x}$ in the support of $\mathbf{x}_i$. That is, there exists a constant $\underline{f}$, such that $f_{\epsilon|\mathbf{x}}(q^*_{\mathbf{x}}|\mathbf{x}) > \underline{f} > 0$

A2 (Restricted identifiability and nonlinearity). Define $T = \{0,1,\ldots,s\}$, and $\overline{T}(\delta,m) \subset \{0,1,\ldots,p\} \backslash T$ as the support of the m largest in absolute value components of the vector. For some constants $m \geq 0$ and $c \geq 0$, the matrix $E[\mathbf{x}_i\mathbf{x}_i{}']$ satisfies

$$\kappa_m^2 := \inf_{\delta \in A_{\mho}, \delta \neq 0} \frac{\delta' E[\mathbf{x}_i \mathbf{x}_i{}']\delta}{\|\delta_{T \bigcup \overline{T}(\delta,m)}\|^2} > 0,$$

where $A := \{\delta \in \mathbb{R}^{p+1} : \|\delta_{T^c}\| \leq c_0 \|\delta_T\|, \|\delta_{T^c}\|_0 \leq n\}$ and $\kappa_0^2 \leq C_f$ for some constant C_f. Moreover,

$$q := \frac{3}{8} \frac{\underline{f}^{3/2}}{\overline{f'}} \inf_{\delta \in A, \delta \neq 0} \frac{E[|\mathbf{x}_i^T \delta|^2]^{3/2}}{E[|\mathbf{x}_i^T \delta|^3]} > 0.$$

A3 (Growth rate of covariates). The growth rate of significant variables and all variables allowed is assumed to satisfy $s^3(\log(n \vee p))^{2+\gamma}/n \to 0$, for some $\gamma > 0$.

A4 (Moments of covariate). Covariates satisfy the Cramér condition $E[|z_{ij}|^k] \leq 0.5 C_m M^{k-2} k!$ for some constant C_m, M, all $k \geq 2$ and all $j = 1, \ldots, p$.

A5 (Well separated regression coefficients). We assume that there exists a $b_0 > 0$, such that for all $j \leq s$, $|\beta_j^*| > b_0$. We note b_0 could still be unknown to us.

Conditions A1–A5 are commonly assumed in the literature (see e.g. Fan and Peng, 2004; Huang et al., 2008a, 2008b; Belloni and Chernozhukov, 2011). Condition A1 is slightly different from Condition D.1 in Belloni and Chernozhukov (2011). The assumption D.1 in Belloni and Chernozhukov (2011), requiring the conditional density at the conditional quantile is uniformly bounded away from 0, can be replaced by a more general condition. In fact, we only need that the conditional density is nonvanishing. Condition A2 requires that there exists a constant C_f, such that $\kappa_0^2 \leq C_f$. This along with the fact that κ_m^2 is nonincreasing in m, immediately entails that the smallest eigenvalue of the covariance matrix $\Sigma_s := E[\mathbf{x}_{ia}\mathbf{x}_{ia}{}']$ is finite and bounded away from 0.

Condition A3 seems to be a strong assumption at first glance, because it limits the size of significant variables to be less than $n^{1/3}$, rather than $n^{2/3}$ as shown in Portnoy (1984). However, this assumption is in accord with Welsh (1989), in which the author showed that if the score function is discontinuous, the growth rate for covariates, $p^3(\log(n))^{2+\gamma}/n \to 0$ is sufficient to obtain the consistency and asymptotic normality under the full model. Since we deal with the high-dimensional sparse model, the growth rate would be expected to obey $s^3(\log(n \vee p))^{2+\gamma}/n \to 0$. Condition A4 is important for us to apply Bernstein's inequality, and hence to establish the sparsity property of the adaptive L_1 quantile estimator. In addition, A5 also implies $\sum_{i=1}^n E\|\mathbf{x}_{ia}\|^2 \sim O(ns)$, which is essential for establishing the oracle consistency property. Condition A5 is also required in Huang et al. (2008b). It assumes that the nonzero coefficients are uniformly bounded away from 0; in other words, the parameter values of the true model are well separated from zero. This assumption can be relaxed to that $\min_{j \leq s}|\beta_j^*|$ goes to 0 at a suitable rate, at the cost of more complicated technical proofs.

3.2. Oracle properties

We show that the adaptive L_1 quantile regression estimator enjoys oracle properties.

Theorem 3.1. *Suppose that assumptions A1–A5 are satisfied. Furthermore, if λ_n satisfies $\lambda_n s/\sqrt{n} \to 0$ and $\lambda_n/(\sqrt{s}\log(n \vee p)) \to \infty$, then the adaptive L_1 quantile regression estimator $\hat{\beta}$ must satisfy the following three properties*:

1. *Variable selection consistency*:
$$P(\hat{\beta}_b = 0) \geq 1 - 6\exp\left\{-\frac{\log(n \vee p)}{4}\right\}.$$
2. *Estimation consistency*:
$$\|\hat{\beta} - \beta^*\| = O_p\left(\sqrt{\frac{s}{n}}\right).$$
3. *Asymptotic normality*: *Let $u_s^2 = \alpha^T \Sigma_s \alpha$ for any vector $\alpha \in \mathbb{R}^s$ satisfying $\|\alpha\| < \infty$. Then*
$$n^{1/2} u_s^{-1} \alpha^T (\hat{\beta}_a - \beta_a^*) \xrightarrow{D} N\left(0, \frac{\tau(1-\tau)}{f^2(q^*)}\right).$$

Remark 3.1. $\tilde{\beta}$ must be at least $\sqrt{n/(s\log(n \vee p))}$-consistent. If $\tilde{\beta}$ is a consistent estimator of β^* with some faster rate, that is, there is a sequence of a_n such that $a_n\|\tilde{\beta} - \beta^*\| \sim O_p(1)$ and $\sqrt{n/(s\log(n \vee p))} \sim o(a_n)$, the oracle properties can still be achieved if $\lambda_n s/\sqrt{n} \to 0$ and $\lambda_n a_n/\sqrt{n\log(n \vee p)} \to \infty$.

Remark 3.2. The asymptotic normality of any linear combination $u_s^{-1}\alpha(\hat{\beta}_a - \beta_a^*)$ is a substitute for the traditional asymptotic normality. Convergence of the finite-dimensional distributions ensures convergence in sequence space. In practice, hypothesis tests and confidence intervals would be constructed using linear combinations.

3.3. *The choice of λ_n*

The regularization parameter, λ_n, plays a crucial role for the adaptive L_1 quantile estimator. It controls the overall magnitude of the adaptive weights and should be chosen so that insignificant variables' regression coefficient estimates shrink to zero, while significant variables are not overpenalized.

Procedures, which are commonly used to select λ_n, such as k-fold cross-validation, generalized cross-validation (Tibshirani, 1996; Fan and Li, 2001), and so on, can be applied to choose λ_n with some appropriate modification. However, using them may have several drawbacks. First, p, the number of variables in the full model, is increasing as the sample size grows. This factor results in an unpleasant issue in that the number of potential models goes to infinity very quickly, which makes computation much too expensive. Second, their statistical properties are not clearly understood for (ultra)high-dimensional regression. For example, there is no guarantee that K-fold cross-validation would provide a choice of λ_n with a proper rate. Third, their statistical properties are still uncharted under the heavy-tailed errors, where quantile regressions are often applied.

Wang and Leng (2007) developed a BIC criterion to select the tuning parameter λ_n for least square approximation (LSA) procedure, and its theoretical model selection consistency property has been demonstrated in Wang et al. (2007) for fixed dimensionality and in Wang et al. (2009) for high-dimensional regression. However, two limitations make such a BIC criterion less favorable in this ultra-high dimensional problem. The first limitation is that one of the requirements in Wang et al. (2009) is $p < n$, which may not be satisfied in the ultra-high dimensional problem. The other limitation is that there is no efficient path-finding algorithm for quantile regression. Thus, we need to search all possible subsets to find the minimum BIC. This could potentially exhaust our computation. One might be able to use the LSA to approximate the quantile regression, and then implement least angle regression slicing (LARS) algorithm to find a solution path in an easier manner, as pointed out in Wang and Leng (2007). However, this would require obtaining a reliable estimate of the inverse of the covariance matrix (see Wang and Leng, 2007), which is a difficult problem in the ultra-high dimensional case. Instead we consider an alternative method for selecting λ_n.

According to Theorem 3.1, a proper λ_n must satisfy two conditions: $\lambda_n s/\sqrt{n} \to 0$ and $\lambda_n/(\sqrt{s}\log(n \vee p)) \to \infty$. We can see that $O(\sqrt{s}\log(n \vee p)(\log n)^{\gamma/2})$ is a suitable choice of λ_n under the condition A5. However, the obstacle is that we do not know the true dimension s. Hence, a natural problem is can we find a good estimate of s, or at least get a quantity of order $O(s)$? Belloni and Chernozhukov (2011) show that their estimator $\|\tilde{\beta}_\tau\|_0 \sim O_p(s)$. If the parameter values of the minimal true model are well separated from zero as condition A7 assumes, then $\|\tilde{\beta}\|_0 \sim O_p(s)$. Since $\tilde{\beta}$ is consistent, $\|\tilde{\beta}_\tau\|_0$ is of order s with a large probability. Therefore, we can use $\tilde{\beta}_\tau$ not only to adjust weights for each regression coefficient, but also to get a quantity used to construct a good choice of λ_n. In practice, we choose $\lambda_n = 0.25\sqrt{\|\tilde{\beta}\|_0}\log(n \vee p)(\log n)^{0.1/2}$ and it works well in our simulation studies.

4. Numerical analysis

To evaluate the finite sample performance of the proposed estimator, we conducted Monte Carlo simulations. We compare the performance of the oracle quantile estimator, the L_1 penalized, post L_1 penalized quantile estimators (Belloni and Chernozhukov, 2011), and the proposed adaptive estimator. The post L_1 penalized quantile estimator is obtained by applying ordinary quantile regression to the model selected by the L_1 penalized quantile regression.

We adopt the simulation settings used in Belloni and Chernozhukov (2011). Consider the regression model 1

$$y_i = \mathbf{x}_i^T \beta + \epsilon,$$

where $\beta = (1,1,1/2,1/3,1/4,1/5,0,\ldots,0)^T$ and $\mathbf{x}_i = (1,\mathbf{z}_i^T)^T$ consists of an intercept and covariates $\mathbf{z}_i \sim \mathrm{N}(0,\Sigma)$, and the errors ϵ are independently and identically distributed $\epsilon \sim \mathrm{N}(0,\sigma^2)$. The dimension p of covariate is 500, and the true dimensional s is 6. The regressors are correlated with $\Sigma_{ij} = \rho^{|i-j|}$ and $\rho = 0.5$. We apply the median regression and choose $\lambda_n = 0.25\sqrt{\|\tilde{\beta}\|_0 \log(n \vee p)(\log n)^{0.1/2}}$. We consider three levels of noise $\sigma = 1, 0.5$ and 0.1. 100 training data sets are generated, each consisting of 100 observations.

We assess model selection by calculating N1: the number of covariates selected by each estimator $\hat{\beta}$, N2: the correct number of covariates selected by each estimator, and the percentage of underfitted, correctly fitted, and overfitted. We evaluate the estimation accuracy by computing the norm of the bias and the empirical risk $[E[\mathbf{x}_i^T(\hat{\beta}-\beta)]^2]^{1/2}$. The results are summarized in Table 1. We can see that although the proposed estimator may still fail to select some significant variables when σ is large due to the ultra-high dimensionality, it significantly improves the performance of quantile regression in both model selection and estimation, compared with the L_1 penalized, post L_1 penalized quantile estimators. Notice that the proposed estimator does not necessarily treat 0 as an absorbing status even when the initial L_1 penalized estimator provides a zero estimate. This is the advantage of using $\omega_j = |\tilde{\beta}|^{-1} \wedge \sqrt{n}$, which provides another opportunity to select the significant regressors, and hence provides better results.

Table 1
Simulation results for model 1.

	Average N1	Average N2	Underfitted	Correctly fitted	Overfitted	Bias	Empirical risk
$\sigma = 1$							
Oracle	6	6	0	1	0	0.03	0.31
L_1	3.21	3.21	1	0	0	0.77	1.09
Post L_1	3.21	3.21	1	0	0	0.30	0.59
Adaptive	4.04	4.04	1	0	0	0.22	0.43
$\sigma = 0.5$							
Oracle	6	6	0	1	0	0.02	0.15
L_1	4.41	4.40	0.98	0.02	0	0.49	0.69
Post L_1	4.41	4.40	0.98	0.02	0	0.21	0.31
Adaptive	5.05	5.04	0.73	0.26	0.01	0.16	0.25
$\sigma = 0.1$							
Oracle	6	6	0	1	0	0	0.03
L_1	5.93	5.93	0.07	0.93	0	0.15	0.20
Post L_1	5.93	5.93	0.07	0.93	0	0.01	0.04
Adaptive	6.05	5.99	0.01	0.95	0.04	0.01	0.03

Table 2
Simulation results for model 2.

	Average N1	Average N2	Underfitted	Correctly fitted	Overfitted	Bias	Empirical risk
$\sigma = 1$							
Oracle	6	6	0	1	0	0.02	0.11
L_1	4.36	4.35	0.96	0.04	0	0.53	0.74
Post L_1	4.36	4.35	0.96	0.04	0	0.20	0.31
Adaptive	5.08	5.06	0.75	0.25	0	0.14	0.22
$\sigma = 0.5$							
Oracle	6	6	0	1	0	0	0.05
L_1	5.35	5.34	0.62	0.38	0	0.33	0.46
Post L_1	5.35	5.34	0.62	0.38	0	0.12	0.15
Adaptive	5.88	5.85	0.15	0.85	0	0.05	0.08

Following Wang et al. (2012), we consider model 2, which is a heterogenous version model 1.

$$y_i = \mathbf{x}_i^T \beta + \Phi(x_{i2})\epsilon,$$

where $\Phi(\cdot)$ is the standard normal cumulative density function. We consider $\sigma = 1$ and $\sigma = 0.5$. And the results are presented in Table 2. Similar conclusions can be drawn from Table 2. All three methods are able to work for regression models with heterogenous errors. However, as observed from Table 2, the adaptive penalized quantile regression drastically outperformed the L_1 penalized, post L_1 penalized quantile estimators in both model selection and estimation.

5. Conclusion

In this paper, the adaptive L_1 quantile regression is introduced for high-dimensional sparse models. It is shown that such an adaptive robust estimator enjoys the oracle properties. In the case of quantile regression we can relax the moment conditions and the constant variance assumption on the error sequence from those used to prove oracle properties of penalized least squares loss methods for high-dimensional data. Our simulation results demonstrate that the proposed estimator owns satisfactory finite sample performances. Although the oracle properties of a single quantile index τ are presented here, the result can be easily extended to a finite composite quantile regression (Zou and Yuan, 2006).

Appendix A. Consistency and sparsity

Define the score function of $\rho_\tau(\cdot)$ by $\varphi_\tau(\cdot)$, i.e. $\varphi_\tau(t) = \tau 1(t \geq 0) - (1-\tau)1(t < 0)$. $\hat{\beta}_\tau$ is the minimizer of the objective function

$$Q_\tau(\beta) = \sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}_i^T \beta) + \lambda_n \sum_{j=0}^p \omega_j |\beta_j|.$$

Throughout $\tilde{\beta}$ is a $\sqrt{n/(s\log(n \vee p))}$-consistent estimator of β^*.

Lemma A1. *Under assumptions A1–A5, if* $\lambda_n/(\sqrt{s}\log(n \vee p)) \to \infty$ *and* $\omega_i = |\tilde{\beta}_{\tau j}|^{-1}$ *for* $1 \leq j \leq p$, *then the adaptive* L_1 *quantile regression estimator* β_τ *satisfies* $\beta_{\tau b} = 0$ *with probability tending to*1.

Proof. It can be seen that the objective function $Q_\tau(\beta)$ is piecewise linear. According to Theorem 1 in Bloomfield and Steiger (1983, p. 7), the minimum of $Q_\tau(\beta)$ can be achieved at some breaking point β, where $\rho_\tau(y_i - \mathbf{x}_i^T \beta) = 0$ for some values of $i = 1, \ldots, n$.

Take the first derivative of $Q(\beta)$ at any differential point $\check{\beta} \in R^{p+1}$ with respect to $\beta_j, j = s+1, \ldots, p$, and we obtain that

$$\frac{\partial Q(\beta)}{\partial \beta_j}\Big|_{\check{\beta}} = -\sum_{i=1}^n \varphi(y_i - \mathbf{x}_i^T \check{\beta}) x_{ij} + \lambda_n \omega_j \,\mathrm{sgn}(\check{\beta}_j). \tag{A.1}$$

Let

$$D(\check{\beta}, \beta^*) = \sum_{i=1}^n \varphi(y_i - \mathbf{x}_i^T \check{\beta}) x_{ij} - \sum_{i=1}^n \varphi(y_i - \mathbf{x}_i^T \beta^*) x_{ij}.$$

Note that,

$$D(\check{\beta}, \beta^*) = \sum_{\epsilon_i \geq q^*_{\mathbf{x}_i}, \epsilon_i \geq q^*_{\mathbf{x}_i} + \mathbf{x}_i^T(\check{\beta} - \beta^*)} [\tau x_{ij} - \tau x_{ij}] + \sum_{\epsilon_i \geq q^*_{\mathbf{x}_i}, \epsilon_i < q^*_{\mathbf{x}_i} + \mathbf{x}_i^T(\check{\beta} - \beta^*)} [-(1-\tau)x_{ij} - \tau x_{ij}]$$
$$+ \sum_{\epsilon_i < q^*_{\mathbf{x}_i}, \epsilon_i \geq q^*_{\mathbf{x}_i} + \mathbf{x}_i^T(\check{\beta} - \beta^*)} [\tau x_{ij} + (1-\tau)x_{ij}] + \sum_{\epsilon_i < q^*_{\mathbf{x}_i}, \epsilon_i < q^*_{\mathbf{x}_i} + \mathbf{x}_i^T(\check{\beta} - \beta^*)} [-(1-\tau)x_{ij} + (1-\tau)x_{ij}],$$

where $q^*_{\mathbf{x}_i}$ is the conditional τth quantile of $\epsilon_i | \mathbf{x}_i$. For $K_1 = \{i : q^*_{\mathbf{x}_i} \leq \epsilon_i < q^*_{\mathbf{x}_i} + \mathbf{x}_i^T(\check{\beta} - \beta^*)\}$ and $K_2 = \{i : q^*_{\mathbf{x}_i} > \epsilon_i \geq q^*_{\mathbf{x}_i} + \mathbf{x}_i^T(\check{\beta} - \beta^*)\}$,

$$D(\check{\beta}, \beta^*) = -\sum_{K_1} x_{ij} + \sum_{K_2} x_{ij}.$$

Hence,

$$\left|\sum_{i=1}^n \varphi(y_i - \mathbf{x}_i^T \check{\beta}) x_{ij}\right| = \left|\sum_{i=1}^n \varphi(y_i - \mathbf{x}_i^T \beta^*) x_{ij} + D(\check{\beta}, \beta^*)\right| \leq \left|\sum_{i=1}^n \varphi(y_i - \mathbf{x}_i^T \beta^*) x_{ij}\right| + \left|\sum_{K_1} x_{ij}\right| + \left|\sum_{K_2} x_{ij}\right| =: I_1 + I_2 + I_3.$$

Consider I_1 first. Let $\xi_i=\varphi(y_i-\mathbf{x}_i^T\beta^*)=\tau 1(\epsilon_i \geq q^*_{\mathbf{x}_i})-(1-\tau)1(\epsilon_i < q^*_{\mathbf{x}_i})$. Conditional on $\mathbf{x}_i$, it is easy to verify that $E[\xi_i x_{ij}]=0$ and $\xi_i x_{ij}, i=1,\dots,n$ satisfy the Cramér condition. As a result, applying Bernstein's inequality yields

$$P\left(\left|\sum_{i=1}^n \xi_i x_{ij}\right| > \sqrt{5C_m n\log(n\vee p)}\right) \leq 2\exp\left\{-\frac{5C_m\log(n\vee p)}{2\left[C_m+M\sqrt{5C_m}\frac{\sqrt{\log(n\vee p)}}{\sqrt{n}}\right]}\right\} \leq 2\exp\left\{-\frac{5\log(n\vee p)}{4}\right\}.$$

Let

$$\Omega_1=\left\{\max_{s+1\leq j\leq p}\left|\sum_{i=1}^n \xi_i x_{ij}\right| \leq \sqrt{5C_m n\log(n\vee p)}\right\}.$$

Then

$$P(\Omega_1)\geq 1-2\exp\left\{\log(p-s)-\frac{5\log(n\vee p)}{4}\right\}\geq 1-v_1,$$

where $v_1=2\exp\{-\log(n\vee p)/4\}\to 0$ as $n\to\infty$. Applying Bernstein's inequality to I_2 yields

$$P\left(\left|\sum_{K_1} x_{ij}\right| > \sqrt{5C_m\log(n\vee p)}\right) \leq 2\exp\left\{-\frac{5C_m\log(n\vee p)}{2\left[\frac{|K_1|C_m}{n}+M\sqrt{5C_m}\frac{\sqrt{\log(n\vee p)}}{\sqrt{n}}\right]}\right\}.$$

Define

$$\Omega_2=\left\{\max_{s+1\leq j\leq p}\left|\sum_{i\in K_1} x_{ij}\right| \leq \sqrt{5C_m n\log(n\vee p)}\right\}.$$

We obtain $P(\Omega_2)\geq 1-v_1$. A similar argument will show that $P(\Omega_3)\geq 1-v_1$, where

$$\Omega_3=\left\{\max_{s+1\leq j\leq p}\left|\sum_{i\in K_2} x_{ij}\right| \leq \sqrt{5C_m n\log(n\vee p)}\right\}.$$

Note that $\Omega_1\bigcup\Omega_2\bigcup\Omega_3\subset\{|\varphi(y_i-\mathbf{x}_i^T\check{\beta})x_{ij}|\leq 3\sqrt{5C_m n\log(n\vee p)}\}$. Therefore,

$$P(|\varphi(y_i-\mathbf{x}_i^T\check{\beta})x_{ij}|\leq 3\sqrt{5C_m n\log(n\vee p)})\geq 1-3v_1.$$

Since $\|\tilde{\beta}\|\sim O_p(\sqrt{s\log(n\vee p)/n})$, for n sufficiently large with probability approaching 1,

$$\frac{\lambda_n\omega_j}{3\sqrt{5C_m n\log(n\vee p)}}>1.$$

With probability at least $1-3v_1$, we have

$$\frac{|\varphi(y_i-\mathbf{x}_i^T\check{\beta})x_{ij}|}{3\sqrt{5C_m n\log(n\vee p)}}\leq 1<\frac{\lambda_n\omega_j}{3\sqrt{5C_m n\log(n\vee p)}}$$

for all $j>s$. This implies that with probability tending to 1

$$\frac{\partial Q(\beta)}{\partial\beta_j}\Big|_{\check{\beta}}=\begin{cases}>0 & \text{if } \check{\beta}_j>0\\ <0 & \text{if } \check{\beta}_j<0\end{cases}.$$

Since $Q(\beta)$ is a continuous function, $\hat{\beta}$, the minimizer of $Q(\beta)$ must satisfy $\hat{\beta}_b=0$. □

Lemma A2. *Under the assumptions A1–A5, if $\lambda_n s/\sqrt{n}\to 0$ and $\omega_i=|\tilde{\beta}_{\tau j}|^{-1}$ for $0\leq j\leq p$, then the adaptive L_1 quantile regression estimator is $\sqrt{n/s}$-consistent.*

Proof. We want to show that for any $\epsilon>0$, there exists a sufficiently large constant, such that

$$P\left\{\inf_{\|\boldsymbol{\delta}_{\mathbf{a}}\|=C} Q_a\left(\beta_a^*+\sqrt{\frac{s}{n}}\delta_a\right)>Q_a(\beta_a^*)\right\}>1-\epsilon \tag{A.2}$$

where $Q_a(\cdot)$ is the objective function restricted to the true underlying model, $\delta_a\in\mathbb{R}^s$ and $\|\boldsymbol{\delta}\|=C$. Since the objective function $Q_a(\beta_a)$ is strictly convex, the inequality (A.2) implies, with probability at least $1-\epsilon$, the oracle quantile estimator lies in the shrinking ball $\{\beta^*+\sqrt{s/n}\delta_a : \delta_a \in \mathbb{R}^{s+1}, \|\boldsymbol{\delta}_a\| \le C\}$. This provides the consistency result immediately.

$$Q_a\left(\beta_a^*+\sqrt{\frac{s}{n}}\delta_a\right)-Q_a(\beta_a^*)=\sum_{i=1}^{n}\rho\left(y_i-\mathbf{x}_{ia}^T\left(\beta_a^*+\sqrt{\frac{s}{n}}\delta_a\right)\right)-\rho(y_i-\mathbf{x}_{ia}^T\beta_a^*)+\lambda_n\sum_{j=0}^{s}\omega_j\left(|\beta_{\tau j}^*+\sqrt{\frac{s}{n}}\delta_{aj}|-|\beta_{\tau j}^*|\right) \tag{A.3}$$

According to Knight (1998), for any $x\neq 0$, we have

$$|x-y|-|x| = -y[1(x>0)-1(x<0)]+2\int_0^y [1(x<t)-1(x<0)]\,dt$$

Then we have

$$\rho(x-y)-\rho(x) = y[1(x<0)-\tau]+2\int_0^y [1(x<t)-1(x<0)]\,dt$$

Hence, (A.3) can be written as

$$\sqrt{\frac{s}{n}}\sum_{i=1}^{n}\mathbf{x}_{ia}^T\delta_a[1(y_i-\mathbf{x}_{ia}^T\beta_a^*<0)-\tau]+\sum_{i=1}^{n}\int_0^{(\sqrt{s/n})\mathbf{x}_{ia}^T\delta_a}[1(y_i-\mathbf{x}_{ia}^T\beta_a^*<t)-1(y_i-\mathbf{x}_{ia}^T\beta_a^*<0)]\,dt+\lambda_n\sum_{j=0}^{s}\omega_j\left(|\beta_{\tau j}^*+\sqrt{\frac{s}{n}}\delta_{aj}|-|\beta_{\tau j}^*|\right)$$

$$:= \sqrt{\frac{s}{n}}T_1+T_2+T_3$$

Using independence and the Cauchy–Schwarz inequality,

$$E[T_1^2]=E\left[\left(\sum_{i=1}^{n}\mathbf{x}_{ia}^T\delta_a[1(y_i-\mathbf{x}_{ia}^T\beta_a^*<0)-\tau]\right)^2\right]=E\left[\sum_{i=1}^{n}(\mathbf{x}_{ia}^T\delta_a[1(y_i-\mathbf{x}_{ia}^T\beta_a^*<0)-\tau])^2\right]\le n\tau(1-\tau)E[\|\mathbf{x}_{ia}\|^2\|\delta_a\|^2]$$
$$\le ns\tau(1-\tau)C_mC^2.$$

Using Chebychev's inequality, we see that for any constant k

$$P\left(\sqrt{\frac{s}{n}}T_1>ksC^2\right)\le\frac{\tau(1-\tau)C_m}{C^2}. \tag{A.4}$$

Next, we deal with T_2. The goal is to show that $T_2 \overset{p}{\ge} 0.5s\underline{f}\kappa_0^2C^2$. Using independence and the fact that $V(X)\le EX^2$,

$$V[T_2]=V\left[\sum_{i=1}^{n}\int_0^{(\sqrt{s/n})\mathbf{x}_{ia}^T\delta_a}[1(y_i-\mathbf{x}_{ia}^T\beta_a^*<t)-1(y_i-\mathbf{x}_{ia}^T\beta_a^*<0)]\,dt\right]\le nE\left[\int_0^{(\sqrt{s/n})\mathbf{x}_{ia}^T\delta_a}[1(y_i-\mathbf{x}_{ia}^T\beta_a^*<t)-1(y_i-\mathbf{x}_{ia}^T\beta_a^*<0)]\,dt\right]^2$$

Given an $\eta>0$ we have

$$nE\left[\left(\int_0^{(\sqrt{s/n})\mathbf{x}_{ia}^T\delta_a}[1(y_i-\mathbf{x}_{ia}^T\beta_a^*<t)-1(y_i-\mathbf{x}_{ia}^T\beta_a^*<0)]\,dt\right)^2 1\left(\sqrt{\frac{s}{n}}|\mathbf{x}_{ia}^T\delta_a|>\eta\right)\right]\le 4sE\left[(\mathbf{x}_{ia}^T\delta_a)^2 1\left(\sqrt{\frac{s}{n}}|\mathbf{x}_{ia}^T\delta_a|>\eta\right)\right]$$
$$\le 4sE[|\mathbf{x}_{ia}^T\delta_a|^3]^{2/3}\left(P\left(\sqrt{\frac{s}{n}}|\mathbf{x}_{ia}^T\delta_a|>\eta\right)\right)^{1/3}, \tag{A.5}$$

where the last line follows from Holder's inequality. Under condition A4,

$$E[|\mathbf{x}_{ia}^T\delta_a|^3]\le\frac{3}{8}\frac{\underline{f}^{3/2}}{\overline{f}'}\frac{E[|\mathbf{x}_{ia}^T\delta_a|^2]^{3/2}}{q}. \tag{A.6}$$

Applying Bernstein's inequality (Lemma 2.2.11 of Van Der Vaart and Wellner 1996),

$$P\left(|\mathbf{x}_{ia}^T\delta_a|>\eta\frac{\sqrt{n}}{\sqrt{s}}\right)\le 2\exp\left\{\frac{-\eta^2 n}{2s\left(C^2C_m+MC\eta\frac{\sqrt{n}}{\sqrt{s}}\right)}\right\}. \tag{A.7}$$

Combining bounds (A.6) and (A.7) yields

$$\text{RHS of (A.5)}\le 4s\left(\frac{3}{8}\frac{\underline{f}^{3/2}}{\overline{f}'}\frac{E[|\mathbf{x}_{ia}^T\delta_a|^2]^{3/2}}{q}\right)^{2/3}\left(2\exp\left\{\frac{-\eta\sqrt{n}}{2MC\sqrt{s}}\right\}\right)^{1/3}\le 3^{2/3}2^{1/3}\frac{\underline{f}}{(\overline{f}'q)^{2/3}}C_mC^2s^2\exp\left\{\frac{-\eta\sqrt{n}}{6MC\sqrt{s}}\right\}$$
$$=3^{2/3}2^{1/3}\frac{\underline{f}}{(\overline{f}'q)^{2/3}}C_mC^2\exp\left\{2\log(s)-\frac{\eta\sqrt{n}}{6MC\sqrt{s}}\right\},$$

which converges to 0 if η satisfies (C1): $\log(s) \sim o(\eta\sqrt{n}/(12MC\sqrt{s}))$ and (C2): $\eta\sqrt{n}/\sqrt{s} \to \infty$. On the other hand,

$$\begin{aligned} & nE\left[\left(\int_0^{\sqrt{\frac{s}{n}}\mathbf{x}_{ia}^T\delta_a}[1(y_i - \mathbf{x}_{ia}^T\beta_a^* < t) - 1(y_i - \mathbf{x}_{ia}^T\beta_a^* < 0)]\,dt\right)^2 1\left(\sqrt{\frac{s}{n}}|\mathbf{x}_{ia}^T\delta_a| \le \eta\right)\right] \\ & \le 2n\eta E\left[\left(\int_0^{\sqrt{\frac{s}{n}}|\mathbf{x}_{ia}^T\delta_a|}[1(y_i - \mathbf{x}_{ia}^T\beta_a^* < t) - 1(y_i - \mathbf{x}_{ia}^T\beta_a^* < 0)]\,dt\right) 1\left(\sqrt{\frac{s}{n}}|\mathbf{x}_{ia}^T\delta_a| < \eta\right)\right] \\ & = 2n\eta E\left[\left(\int_0^{\sqrt{s/n}|\mathbf{x}_{ia}^T\delta_a|}[F_{\epsilon|\mathbf{x}_i}(q_{\mathbf{x}_i}^* + t) - F_{\epsilon|\mathbf{x}_i}(q_{\mathbf{x}_i}^*)]\,dt\right) 1\left(\sqrt{\frac{s}{n}}|\mathbf{x}_{ia}^T\delta_a| < \eta\right)\right] \end{aligned} \tag{A.8}$$

If η is close to 0, then $F(t) - F(0) \le \bar{f}t, \forall |t| < \eta$. Thus, we obtain

$$(\text{A.7}) \le \bar{f}t\eta nE\left[\left(\int_0^{\sqrt{s/n}|x_{ia}^T\delta_a|} t\,dt\right) 1\left(\sqrt{\frac{s}{n}}|x_{ia}^T\delta_a| < \eta\right)\right] \le \bar{f}t\eta^3 n$$

which converges to 0, if η satisfies (C3): $\eta^3 n \to 0$. If η satisfies conditions C1, C2 and C3, then as $n \to \infty$ $V(T_2) \to 0$. By Chebyshev's inequality, we have

$$T_2 - nE\left\{\int_0^{\sqrt{\frac{s}{n}}\mathbf{x}_{ia}^T\delta_a}[1(y_i - \mathbf{x}_{ia}^T\beta_a^* < t) - 1(y_i - \mathbf{x}_{ia}^T\beta_a^* < 0)]\,dt\right\} \xrightarrow{p} p$$

Using Cauchy–Schwartz inequality and a similar argument as in the proof of $V(T_2) \to 0$, we can show that for n sufficiently large

$$nE\left\{\int_0^{\sqrt{\frac{s}{n}}\mathbf{x}_{ia}^T\delta_a}[1(y_i - \mathbf{x}_{ia}^T\beta_a^* < t) - 1(y_i - \mathbf{x}_{ia}^T\beta_a^* < 0)]\,dt\right\} \ge \frac{1}{2}\underline{f}\kappa_0^2 C^2 s$$

Finally for T_3, we have

$$\left|\lambda_n \sum_{j=0}^{s} \omega_j \left(\left|\beta_j^* + \sqrt{\frac{s}{n}}\delta_{aj}\right| - |\beta_j^*|\right)\right| \le \lambda_n \sum_{j=1}^{s} \omega_j \sqrt{\frac{s}{n}}|\delta_{aj}| \le \lambda_n \frac{s}{\sqrt{n}} \max_{1 \le j \le s} \frac{1}{|\beta_j^*|} C \to 0$$

Combining the fact that T_3 converges to zero in probability with (A.4), we see that for sufficiently large C, (A.3) is positive with probability at least $1-\epsilon$ and (A.2) is satisfied. □

Appendix B. Asymptotic normality

Proof of Theorem 3.1. As in the foregoing proofs, we see that with probability at least $1-3v_1$, $\hat{\beta} = \check{\beta}$. Therefore, properties (1) and (2) are achieved automatically. We know that $\check{\beta} = ((\beta^* + \sqrt{s/n}\check{\delta}_a)^T, 0)^T$ where $\sqrt{s/n}\check{\delta}_a$ is the minimizer of the following function:

$$\begin{aligned} Q_a\left(\beta_a^* + \sqrt{\frac{s}{n}}\delta_a\right) - Q_a(\beta_a^*) &= \sqrt{\frac{s}{n}}\sum_{i=1}^n \mathbf{x}_{ia}^T\delta_a[1(\epsilon_i < q_{\mathbf{x}_i}^*) - \tau] \\ &\quad + \sum_{i=1}^n \int_0^{(\sqrt{s/n})\mathbf{x}_{ia}^T\delta_a}[1(\epsilon_i < q_{\mathbf{x}_i}^* + t) - 1(\epsilon_i < q_{\mathbf{x}_i}^*)]\,dt + \lambda_n \sum_{j=0}^{s}\omega_j\left(|\beta_j^* + \sqrt{\frac{s}{n}}\delta_{aj}| - |\beta_j^*|\right) \\ &:= J_1 + J_2 + J_3. \end{aligned}$$

And with probability at least $1-\epsilon$, $\check{\delta}_a$ locates in a ball $B_\epsilon := \{\delta_a : \|\delta\| \le C\}$ for some constant C that implicitly depends on ϵ. For any $\delta_a \in B_\epsilon$, using the argument as in the proof of consistency, we can show that

$$E|J_1/s|^2 \le C_m\|\delta_a\|^2, \quad J_2 \xrightarrow{p} \frac{1}{2}f(q^*)s\delta_a^T\Sigma_S\delta_a,$$

and

$$|J_3| \le \|\delta_a\| O(\sqrt{s}(\log(n))^{\gamma/2}\log(n \vee p)) \frac{s}{\sqrt{n}} \max_{1\, j \le s} \frac{1}{|\tilde{\beta}_j|} = o(1).$$

Thus, with probability at least $1-3v_1-\epsilon$, minimizing $Q_a(\beta_a^* + \sqrt{s/n}\delta_a) - Q_a(\beta_a^*)$ is equivalent to minimizing

$$\sqrt{\frac{s}{n}}\sum_{i=1}^n \mathbf{x}_{ia}^T\delta_a[1(\epsilon_i < q_{\mathbf{x}_i}^*) - \tau] + \frac{1}{2}f(q^*)s\delta_a^T\Sigma_S\delta_a,$$

which provides

$$\check{\delta}_a = \frac{\sum_{i=1}^{n} \Sigma_s^{-1} \mathbf{x}_{ia}[1(\epsilon_i < q^*_{\mathbf{x}_i}) - \tau]}{f(q^*)\sqrt{ns}}$$

Therefore, with probability at least $1-3v_1-\epsilon$

$$\sqrt{n} u_s^{-1} \alpha^T (\hat{\beta}_a - \beta^*_a) = \sqrt{n} \frac{\sum_{i=1}^{n} u_s^{-1} \alpha^T \Sigma_s^{-1} \mathbf{x}_{ia}[1(\epsilon_i < q^*_{\mathbf{x}_i}) - \tau]}{f(q^*)n}.$$

Denote ζ_i by $u_s^{-1} \alpha^T \Sigma_s^{-1} \mathbf{x}_{ia}[1(\epsilon_i < q^*_{\mathbf{x}_i}) - \tau]$ for $i = 1, \ldots, n$. Then $E[\zeta_i] = 0$ and $\text{Var}[\zeta_i] = \tau(1-\tau)$. Therefore, we have

$$\sqrt{n} \frac{\sum_{i=1}^{n} \zeta_i}{f(q^*)n} \xrightarrow{d} \text{N}\left(0, \frac{\tau(1-\tau)}{f^2(q^*)}\right),$$

which completes the proof. □

References

Belloni, A., Chernozhukov, V., 2011. l_1 penalized quantile regression in high-dimensional sparse models. The Annals of Statistics 39, 82–130.
Bloomfield, P., Steiger, W.L., 1983. Least Absolute Deviation: Theory, Applications and Algorithms. Birkhauser, Boston.
Candes, E., Tao, T., 2007. The Dantzig selector: statistical estimation when p is much larger than n. The Annals of Statistics 35, 2313–2351.
Chernozhukov, V., 2005. Extremal quantile regression. The Annals of Statistics 33, 806–839.
Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association 96, 1348–1360.
Fan, J., Peng, H., 2004. Nonconcave penalized likelihood with a diverging number of parameters. The Annals of Statistics 32, 928–961.
He, X., Shao, Q., 2000. On parameters of increasing dimensions. Journal of Multivariate Analysis 73, 120–135.
Huang, J., Horowitz, J.L., Ma, S., 2008a. Asymptotic properties of bridge estimators in sparse high-dimensional regression models. The Annals of Statistics 36, 587–613.
Huang, J., Ma, S., Zhang, C., 2008b. Adaptive lasso for sparse high-dimensional regression models. Statistica Sinica 18, 1603–1618.
Knight, K., 1998. Limiting distributions for L1 regression estimators under general conditions. The Annals of Statistics 26, 755–770.
Koenker, R., Basset, G., 1978. Regression quantiles. Econometrica 46, 33–50.
Koenker, R., 2005. Regression Quantiles. Cambridge University Press, Cambridge.
Portnoy, S., Koenker, R., 1997. The Gaussian Hare and the Laplacian tortoise: computability of square-error versus absolute-error estimators. Statistical Science 12, 279–300.
Portnoy, S., 1984. Asymptotic behavior of M-estimatiors of p regression parameter when p2/n is large I. Consistency. The Annals of Statistics 13, 1402–1417.
Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society, Series B 58, 267–288.
Wang, H., Leng, C., 2007. Unified lasso estimation via least square approximation. Journal of American Statistical Association 102, 1039–1048.
Wang, H., Li, G., Jiang, G., 2007. Robust regression shrinkage and consistent variable selection via the LAD-Lasso. Journal of Business and Economic Statistics 25, 347–355.
Wang, H., Li, B., Leng, C., 2009. Shrinkage tuning parameter selection with a diverging number of parameters. Journal of the Royal Statistical Society, Series B 71, 671–683.
Wang, K., Wu, Y., Li, R., 2012. Quantile regression: applications and current research areas. Journal of the Royal Statistical Society, Series D 52, 331–350.
Welsh, A.H., 1989. On M-processes and M-estimation. The Annals of Statistics 17, 337–361.
Yu, K., Liu, Z., Stander, J., 2003. Quantile regression: applications and current research areas. Journal of the Royal Statistical Society, Series D 52, 331–350.
Zou, H., Yuan, M., 2006. Composite quantile regression and the oracle model selection theory. The Annals of Statistics 36, 1108–1126.