Routledge
Taylor & Francis Group

Economic Research-Ekonomska Istraživanja

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rero20

A sparse approach for high-dimensional data with heavy-tailed noise

Yafen Ye, Yuanhai Shao & Chunna Li

To cite this article: Yafen Ye, Yuanhai Shao & Chunna Li (2022) A sparse approach for high-dimensional data with heavy-tailed noise, Economic Research-Ekonomska Istraživanja, 35:1, 2764-2780, DOI: 10.1080/1331677X.2021.1978306

To link to this article: https://doi.org/10.1080/1331677X.2021.1978306

Published online: 21 Sep 2021.

ECONOMIC RESEARCH-EKONOMSKA ISTRAŽIVANJA
2022, VOL. 35, NO. 1, 2764–2780
https://doi.org/10.1080/1331677X.2021.1978306

Routledge
Taylor & Francis Group

OPEN ACCESS

A sparse approach for high-dimensional data with heavy-tailed noise

Yafen Ye[a], Yuanhai Shao[b] and Chunna Li[b]

[a]School of Economics, Zhejiang University of Technology, Hangzhou, P.R. China; [b]Management School, Hainan University, Haikou, P.R. China

ABSTRACT

High-dimensional data have commonly emerged in diverse fields, such as economics, finance, genetics, medicine, machine learning, and so on. In this paper, we consider the sparse quantile regression problem of high-dimensional data with heavy-tailed noise, especially when the number of regressors is much larger than the sample size. We bring the spirit of L_p-norm support vector regression into quantile regression and propose a robust L_p-norm support vector quantile regression for high-dimensional data with heavy-tailed noise. The proposed method achieves robustness against heavy-tailed noise due to its use of the pinball loss function. Furthermore, L_p-norm support vector quantile regression ensures that the most representative variables are selected automatically by using a sparse parameter. We use a simulation study to test the variable selection performance of L_p-norm support vector quantile regression, where the number of explanatory variables greatly exceeds the sample size. The simulation study confirms that L_p-norm support vector quantile regression is not only robust against heavy-tailed noise but also selects representative variables. We further apply the proposed method to solve the variable selection problem of index construction, which also confirms the robustness and sparseness of L_p-norm support vector quantile regression.

ARTICLE HISTORY
Received 15 May 2021
Accepted 5 September 2021

KEYWORDS
High-dimensional data; heavy-tailed noise; $\boldsymbol{L_p}$-norm support vector quantile regression; variable selection

JEL CODES
B16; B23; C02; C61

1. Introduction

The development of regression technology presents several challenges to modern data. The first challenge comes from the dimensionality of data. High-dimensional data, where the number of explanatory variables (K) greatly exceeds the sample size (N), vary greatly across different fields. For example, large panels of home-price data are high dimensionality (Huang et al., 2020). To consider the cross-sectional effects, the house price in one city depends on several other cities, most likely its geographic neighbors (Fan et al., 2011). Another example of high-dimensional data is in the finance field (Wang et al., 2020; Zhou et al., 2020). Portfolio allocation with a few

CONTACT Chunna Li ✉ na1013na@163.com

thousand stocks involves over one million explanatory variables (Fan & Lv, 2010). High-dimensional data have commonly emerged in other fields, such as genetics (Algamal & Lee, 2019), medicine (Dondelinger et al., 2020), and machine learning (Ye et al., 2017a, 2017b). In gene expression studies, for instance, one is able to collect far fewer observations than the total number of genes assayed (Clarke et al., 2008). A high-dimensional data set needs a sparse technique that can select the representative variables and discard the redundant variables. The second challenge comes from heavy-tailed noise, which exists in practice (Chen et al., 2020; Fan et al., 2017; Hsu & Sabato, 2016; Zhou et al., 2018). Modern data with heavy-tailed noise require robust techniques. The main purpose of this paper is to propose a new sparse regression method for modern high-dimensional data with heavy-tailed noise.

Although ordinary least squares (OLS) regression (Dempster et al., 1977; Greene, 1981; Hutcheson, 2011; Rzhetsky & Nei, 1992) is one of the commonly used methods for estimating conditional mean functions because its estimators have the smallest variance among the class of linear unbiased estimators (Berk & Hwang, 1989), OLS estimation exhibits some drawbacks when used for modern high-dimensional data with heavy-tailed noise. High-dimensional data require a sparse technique for selecting the representative variables and discarding the redundant variables. OLS estimation may suffer from the presence of redundant variables since the process utilizes all variables without discrimination. Heavy-tailed noise requires a robust regression technique. The use of sum of squared residuals makes OLS regression sensitive to noise in heavy-tailed situations (Koenker & Bassett, 1978).

In the field of statistics, the least absolute shrinkage and selection operator (Lasso) (Bertsimas et al., 2016; Kim et al., 2019; Liang & Jacobucci, 2020; Tibshirani, 1996; Wang et al., 2007; Zou, 2006) is a very popular sparse method for high-dimensional data since it shrinks some coefficients of the regression estimators toward 0. According to the regression estimators, the contribution of each explanatory variable to the final decision function can be judged, and then the representative explanatory variables are selected, while the redundant variables are discarded. However, Lasso lacks robustness against heavy-tailed noise. Koenker and Bassett (1978) proposed quantile regression (QR) and effectively dealt with a regression problem involving heavy-tailed noise. QR is robust against noise in heavy-tailed situations since the quantile estimators as a class of empirical 'location' measures for the dependent variable, are based on pinball loss rather than least squares loss (Newey & Powell, 1987). Although QR is robust against heavy-tailed noise (He et al., 2020), it does not focus on the variable selection problem of high-dimensional data.

Li and Zhu (2008) brought the spirit of the Lasso approach into quantile regression, and proposed L_1-norm regularized quantile regression. Thereafter, L_1-penalized quantile regression (Belloni & Chernozhukov, 2011; Peng & Wang, 2015; Yu et al., 2017) and generalized L_1-penalized quantile regression (Liu et al., 2020) were proposed. In addition, Li et al. (2010) studied regularization in quantile regression from a Bayesian perspective. The L_1-norm regularization term in the quantile regression methods has variable selection ability since some coefficients of the estimator are driven towards zero. Therefore, these L_1-norm regularized quantile regression methods conduct estimation and variable selection simultaneously.

In the field of machine learning, support vector regression (SVR) (Drucker et al., 1996; Smola & Schölkopf, 2004) in the framework of statistical learning theory, or Vapnik-Chervonenkis theory, is an effective method for addressing the $K \gg N$ regression problem. Sparse SVR, such as L_1-norm support vector regression (L_1-SVR) (Peng & Xu, 2013; Ye et al., 2017a, 2017b) and L_p-norm support vector regression ($L_\mathrm{p}-$SVR) ($0<p<1$) (Ye et al., 2015, 2017a, 2017b; Zhang et al., 2013), has been proven to be an effective variable selection tool for high-dimensional data. $L_\mathrm{p}-$SVR is much sparser than L_1-SVR since the L_p-norm regularization term in support vector regression shrinks some coefficients of an estimator towards 0, and some coefficients are shrunk to exactly 0, leading to some redundant variables being discarded and some representative variables remaining. Moreover, L_p-norm support vector regression can realize estimation and variable selection simultaneously.

Takeuchi et al. (2006) introduced the spirit of QR to support vector regression and proposed nonparametric quantile regression, which minimizes the pinball loss and the regularization term. Thereafter, support vector censored quantile regression (Shim & Hwang, 2009), semiparametric support vector quantile regression (Shim et al., 2012), and support vector quantile regression (Anand et al., 2020) were proposed. By using the quantile parameter, support vector quantile regression is robust against heavy-tailed noise. In addition, the L_2-norm regularization term in the support vector quantile regression effectively solves the $K \gg N$ estimation problem. However, support vector quantile regression may suffer from the presence of redundant variables in high-dimensional data since the estimator of the L_2-norm regularization term lacks sparseness. Thus, support vector quantile regression methods maintain the advantages of quantile regression and support vector regression but they may use all variables without discrimination in the estimation process for high-dimensional data.

To solve the sparse regression problem of high-dimensional data with heavy-tailed noise, we propose L_p-norm support vector quantile regression (L_p-SVQR). Because we use the quantile parameter in the pinball loss function, L_p-SVQR is robust to heavy-tailed noise. The L_p-norm regularization term in L_p-SVQR solves the $K \gg N$ estimation problem. Moreover, by using the sparse parameter p, L_p-SVQR automatically conducts variable selection and effectively improves the regression results simultaneously. We adopt a convergent successive linear algorithm (SLA) to obtain an approximate local solution of L_p-SVQR. Compared with L$_1$-norm regularized quantile regression (L_1-QR) (Li & Zhu, 2008) and ε-support vector quantile regression (ε−SVQR) (Anand et al., 2020), the simulation results show that L_p-SVQR selects sparser variables but with smaller estimation errors than those of ε−SVQR and L_1-QR, and this means that L_p-SVQR not only selects fewer representative variables but also has good regression effectiveness. To further test the sparseness of L_p-SVQR, we discuss the variable selection problem with regard to index construction. The real-world variable selection analysis of an innovation and entrepreneurship index also shows the sparseness of L_p-SVQR. The contributions of this paper are summarized as follows:

1. During the high-dimensional regression process of L_p-SVQR, useful variables are retained, and irrelevant variables are discarded.

2. By using the quantile parameter in the pinball loss function, L_p-SVQR is robust against heavy-tailed noise.
3. Simulation results indicate that L_p-SVQR outperforms the other two methods with better sparseness and robustness.

The remainder of this paper is organized as follows. In Sec. 2, we propose L_p-SVQR and present the properties of the solution path. The simulation study is shown in Sec. 3. Section 4 provides a real-world variable selection analysis. Section 5 concludes this paper.

2. A high-dimensional sparse regression model

In this section, we introduce the spirit of L_{p}-norm support vector regression into quantile regression, propose L_p-norm support vector quantile regression (L_p-SVQR), and then derive an efficient algorithm that computes the exact solution path for the parameter β.

Suppose that $Y=(y_1, \ldots, y_N)'$ is the response variable, X is a known $N\times K$ design matrix of covariates, and $x_i=(x_{i1}, \ldots, x_{iK})'$ is the K-dimensional explanatory variable. β_0 and $\beta=(\beta_1, \ldots, \beta_K)'$ are the unknown parameters that need to be estimated. Consider the following linear regression model:

$$Y=\beta_0+X\beta+\varepsilon \tag{1}$$

where ε is the random error.

2.1. L_p-norm support vector quantile regression

The estimator of the linear regression (1) can be defined as the solution to the proposed L_p-SVQR optimization problem:

$$\begin{aligned} \min_{\beta_0, \beta, \xi, \xi^*} \quad & \lambda|\beta|_p^p+\left[\tau e^T\xi+(1-\tau)e^T\xi^*\right] \\ s.t. \quad & Y-\beta_0-X\beta\leq\xi, \xi\geq 0, \\ & \beta_0+X\beta-Y\leq\xi^*, \xi^*\geq 0. \end{aligned} \tag{2}$$

where $\tau(0<\tau<1)$ is the quantile level, $\lambda(\lambda\geq 0)$ is a tuning parameter balancing the quantile loss, ξ and ξ^* are slack variables, and e is a vector of ones with appropriate dimensions. We penalize the model's complexity by using the L_p-norm regularization term, $|\beta|_p^p$. Generally, $|\beta|_p^p$ results in a much sparser estimator than that obtained with the L_1-norm regularization term. Furthermore, L_p-SVQR has an adaptive property since the optimal value of p is automatically chosen by the data set. Therefore, by using the parameters λ and p, L_p-SVQR can achieve a sparser estimator for selecting the representative variables and discarding the redundant variables.

Regression problem (2) is not differentiable because of the L_p-norm regularization term. To make it smooth, we introduce the upper bound variable $\upsilon=([\upsilon]_1, \ldots, [\upsilon]_K)^T$, and thus problem (2) can be written as:

$$\begin{aligned} \min_{\beta_0,\beta,\upsilon,\xi,\xi^*} \quad & \lambda\sum_{\mathrm{i}=1}^{\mathrm{K}}[\upsilon]_{\mathrm{i}}^{\mathrm{p}} + \tau e^T\xi + (1-\tau)e^T\xi^* \\ s.t. \quad & Y-\beta_0-X\beta \leq \xi, \xi \geq 0, \\ & \beta_0 + X\beta - Y \leq \xi^*, \xi^* \geq 0, \\ & |\beta| \leq \upsilon, \upsilon \geq 0. \end{aligned} \tag{3}$$

It should be noted that problem (3) is differentiable and can be solved using a successive linear algorithm (SLA) (Mangasarian, 2007). An SLA starts with a random initial point $\hat{\beta}_0^0, \hat{\beta}^0, \hat{\upsilon}^0, \hat{\xi}^0, \hat{\xi}^{*0}$, and the $k-th$ iteration,$k = 1, 2, \ldots,$ obtains $\hat{\beta}_0^k, \hat{\beta}^k, \hat{\upsilon}^k, \hat{\xi}^k, \hat{\xi}^{*k}$ by solving the following problem:

$$\begin{aligned} \min_{\beta_0,\beta,\upsilon,\xi,\xi^*} \quad & \lambda\sum_{\mathrm{i}=1}^{\mathrm{K}}[\upsilon^{k-1}]_{\mathrm{i}}^{\mathrm{p}-1}[\upsilon]_i + \tau\sum_{i=1}^{N}[\xi]_i + (1-\tau)\sum_{i=1}^{N}[\xi^*]_i \\ s.t. \quad & Y-\beta_0-X\beta \leq \xi, \xi \geq 0, \\ & \beta_0 + X\beta - Y \leq \xi^*, \xi^* \geq 0, \\ & |\beta| \leq \upsilon, \upsilon \geq 0. \end{aligned} \tag{4}$$

where $\hat{\upsilon}^{k-1} = ([\hat{\upsilon}^{k-1}]_1, \ldots, [\hat{\upsilon}^{k-1}]_K)^T$. The proof of convergence of the SLA is omitted here since it can be easily obtained from adaptations of the proof in Ye et al. (2015).

2.2. Properties of solution path

Similar to Li and Zhu (2008), we shed light on the properties of the whole solution path $\hat{\beta}$. Problem (4) can be rewritten as an equivalent optimization problem:

$$\begin{aligned} \min_{\beta_0,\beta,\upsilon,\xi,\xi^*} \quad & \tau\sum_{i=1}^{N}[\xi]_i + (1-\tau)\sum_{i=1}^{N}[\xi^*]_i \\ s.t. \quad & Y-\beta_0-X\beta \leq \xi, \xi \geq 0, \\ & \beta_0 + X\beta - Y \leq \xi^*, \xi^* \geq 0, \\ & |\beta| \leq \upsilon, \upsilon \geq 0, \\ & \sum_{\mathrm{i}=1}^{\mathrm{K}}[\upsilon^{k-1}]_i^{p-1}[\upsilon]_i \leq s, \end{aligned} \tag{5}$$

where s is the regularization parameter, which plays the same role as that of λ. From the KKT conditions in (5), we can obtain $0 \leq \alpha_i \leq \tau$ and $0 \leq \alpha_i^* \leq 1-\tau$, where α_i and α_i^* are Lagrangian multipliers. Furthermore, we are able to obtain the following relationships:

1. If $y_i-\beta_0-\beta^Tx_i>0$, then $\xi_i>0$. This implies $\alpha_i = \tau$, $\alpha_i^* = 0$, and $\xi_i^* = 0$.
2. If $y_i-\beta_0-\beta^Tx_i<0$, then $\xi_i^*>0$. This implies $\alpha_i^* = 1-\tau$, $\alpha_i = 0$, and $\xi_i = 0$.
3. If $y_i-\beta_0-\beta^Tx_i = 0$, then $\xi_i = 0$ and $\xi_i^* = 0$. This implies $\alpha_i \in [0,\tau]$ and $\alpha_i^* \in [0, 1-\tau]$.

Like in the condition stated by Li and Zhu (2008), the samples in the training set can be classified into the following three subsets:

$\varepsilon = \{i : y_i - \beta_0 - \beta^T x_i = 0,\ -(1-\tau) \le \alpha_i - \alpha_i^* \le \tau\}$(Elbow)
$L = \{i : y_i - \beta_0 - \beta^T x_i < 0,\ \alpha_i - \alpha_i^* = -(1-\tau)\}$(Left of the elbow)
$R = \{i : y_i - \beta_0 - \beta^T x_i > 0,\ \alpha_i - \alpha_i^* = \tau\}$(Right of the elbow)
$\nu = \{j : \beta_j \ne 0\}$(Active set)

We further discuss the solution path $\hat{\beta}$ as a function of s. We define the following events as s increases:

- Either a data point hits the elbow, that is, a residual $y_i - \hat{\beta}_0 - \hat{\beta}^T x_i$ changes from nonzero to zero, or a coefficient$\hat{\beta}_j$ changes from nonzero to zero.
- Either a data point hits left of the elbow or right of the elbow, that is, a residual $y_i - \hat{\beta}_0 - \hat{\beta}^T x_i$ changes from zero to nonzero, or a coefficient $\hat{\beta}_j$ changes from zero to nonzero.

If an event occurs, we need to update ε, L, R, and ν accordingly, as these are the index sets associated with $\hat{\beta}_j$. Moreover, nonzero $\hat{\beta}_j$ satisfies:

$$y_i - \hat{\beta}_0 - \sum_{j\in\nu} \hat{\beta}_j x_{ij} = 0,\ \text{for } i \in \varepsilon \tag{6}$$

since the number of observations in the elbow is equal to the number of variables in the active set, that is, $|\varepsilon| = |\nu|$.

2.3. Degrees of freedom

Degrees of freedom (df), measuring the effective dimensionality of the fitted model, play an important role in model assessment and variable selection. Suppose that y follows a distribution $y \sim (\mu(x), \sigma^2)$, where μ is the true mean and σ^2 is the variance. The 'degrees of freedom' is defined as:

$$df(\hat{f}) = \frac{1}{\sigma^2} \sum_{i=1}^{N} \mathrm{cov}(\hat{f}(x_i), y_i) \tag{7}$$

where$\hat{f}(x)$ is a fitted model. Stein (1981) showed that the number of degrees freedom for a fitted model $\hat{f}(x)$ can be calculated as:

$$df(\hat{f}) = \sum_{i=1}^{N} E(\frac{\partial \hat{f}(x_i)}{\partial y_i}) \tag{8}$$

There exists a set of regularization parameters $0 = s_0 < s_1 < s_2 \cdots < s_L = \infty$, such that in the interior of any interval (s_l, s_{l+1}), the sets ε, L, R and ν are constant with respect to s. These sets only change during each event. Similar to Li and Zhu (2008), we first list the useful lemmas below, and then we apply Stein's theory (Stein, 1981) to derive an expression for the number of degrees of freedom.

Lemma 1. For any fixed $s > 0$, there exists a set $y = (y_1, \ldots, y_N)^T$ such that s is a finite collection of hyperplanes in $\mathbb{R}^N$ and this set is denote as N_s.

Lemma 2. For any fixed $s>0$, $\hat{\beta}_0(y)$ and $\hat{\beta}(y)$ are continuous function of y, where $\hat{\beta}_0(y)$ and $\hat{\beta}(y)$ are the fitted intercept and coefficient vectors, respectively, when the response vector is y.

Lemma 3. For any fixed $s>0$ and any $y \in \mathbb{R}^N \backslash \mathrm{N}_s$, the sets ε, L and$R$ are locally constant with respect to y.

Theorem 1. For any fixed $s>0$ and any $y \in \mathbb{R}^N \backslash \mathrm{N}_s$, we have the following divergence formula:

$$\sum_{i=1}^{N} \frac{\partial \hat{f}(x_i)}{\partial y_i} = |\varepsilon| \tag{9}$$

The proofs of these results are omitted here since they can be easily obtained from the proofs by Li and Zhu (2008).

An important application of degrees of freedom is the selection of the regularization parameter s. Two commonly used criteria in the quantile regression literature are the Schwarz information criterion (SIC) (Schwarz 1978) and generalized approximate cross-validation criterion (GACV) (Yuan, 2006). Here, we chose to minimize the following normalized mean squared error (NMSE) and sum absolute estimation error (SAEE) to select the optimal s. They are defined as follows:

$$NMSE(s) = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 / \sum_{i=1}^{N} (y_i - \bar{y})^2 \tag{10}$$

$$SAEE(s) = \sum_{i=1}^{K} \left| \hat{\beta}_i - \beta_i \right| \tag{11}$$

where $\bar{y}$ is the average value of $y_1, \ldots, y_{N.}$

3. Simulation study

In this section, we conduct a simulation study to evaluate the variable selection performance of L_p-SVQR by comparing the obtained results with those of L_1-norm regularized quantile regression (L_1-QR) (Li & Zhu, 2008) and ε-support vector quantile regression ($\varepsilon-$SVQR) (Anand et al., 2020).

3.1. Simulation setup

In our simulation study, the parameter s is searched from the set $\{2^{(-10)}, 2^{(-9)}, \ldots, 2^9, 2^{10}\}$, and the parameter p is chosen from 0 to 1 with a fixed step size of 0.1. The optimal values of the parameters in the simulation study are obtained by utilizing the grid-search method.

Let n be the number of samples, $\hat{y}_i$ be the prediction value of y_i, and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ be the average value of $y_1, \ldots, y_n$. We use the following evaluation criteria to evaluate the variable selection ability of each model. Similar to the model setup in Peng and Wang (2015), we generate $(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_K)^T$ from $\mathrm{N}(0, \Sigma)$, where Σ is the covariance matrix with elements $\sigma_{ij} = 0.5^{|i-j|}$, $1 \le i, j \le K$. Then, we set $x_1 = F(\tilde{x}_1)$ and $x_k = \tilde{x}_k$ for $k = 2, 3, \ldots, K$, where $F(\cdot)$ is the cumulate distribution function of the standard normal distribution. Next we set $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_K)^T$ as the estimate of β. We use the following criteria to evaluate the variable selection ability of linear L_p-SVQR :

P$_1$: The proportion of simulation runs nonzero coefficients are selected.

P$_2$: The proportion of simulation runs is x_1 selected.

AEE: The absolute estimation error defined as $\left|\hat{\beta}_i - \beta_i\right|$, $i = 1, 2, \ldots, K$.

NMSE: The normalized mean squared error (NMSE) defined as:

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 / \sum_{i=1}^{n} (y_i - \bar{y})^2. \tag{12}$$

R2: The coefficient of determination R^2 defined as:

$$R^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 / \sum_{i=1}^{n} (y_i - \bar{y})^2. \tag{13}$$

We consider two simulation cases with five different values of τ : 0.1, 0.3, 0.5, 0.7, and 0.9. We set $K = 500$ and $N = 100$, and generate the first regression model as follows:

$$\text{Type A}: \mathrm{y} = x_{50} + x_{100} + x_{150} + x_{200} + x_{250} + x_{300} + x_{350} + x_{400} + x_{450} + x_{500} + x_1\varepsilon \tag{14}$$

where the random error $\varepsilon \sim \mathrm{N}(0, 1)$ is independent of the covariates.

The second regression function is as follows:

Type B : $y = x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10} + x_{11} + x_1\varepsilon$ (15)where the random error ε that is independent of the covariates follows a Cauchy distribution. Specifically, we set $K = 2000, N = 500$, and $\beta = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, \ldots, 0)$, where most elements are zeros and only the first ten components contain nonzero values.

Therefore, these two cases compose the $K \gg N$ estimation problem. The optimal parameter selection is based on 10-fold cross validation. For each case, a total of 100 simulation iterations are conducted to evaluate the parameter β. We chose minimized NMSE and SAEE to select the optimal parameters.

3.2. Sparseness and robustness analysis

Tables 1 and 2 list the variable selection results of L_p-SVQR, $\varepsilon-$SVQR, and L_1-QR. It is observed that L_p-SVQR selects fewer features than $\varepsilon-$SVQR and L_1-QR. The main reason is that the L_p- norm can find sparser estimations than those of the L_1- norm and L_2- norm. L_p-SVQR selects sparser variables than those of $\varepsilon-$SVQR and

Table 1. Type A simulation results of L_p-SVQR, $\varepsilon-$SVQR and L_1-QR.

τ	Regressor	df	P_1(%)	P_2(%)	NMSE	R^2
0.1	L_p-SVQR	**53.6**	**10.72**	**0**	**0.10(0.01)**	**0.80(0.03)**
	$\varepsilon-$SVQR	500	100	100	0.73(0.03)	0.24(0.01)
	L_1-QR	500	100	100	0.72(0.04)	0.26(0.01)
00.3	L_p-SVQR	**12.7**	**2.54**	**0**	**0.07(0.02)**	**0.92(0.03)**
	$\varepsilon-$SVQR	500	100	100	0.73(0.03)	0.24(0.01)
	L_1-QR	500	100	100	0.72(0.04)	0.26(0.01)
0.5	L_p-SVQR	**12.2**	**2.44**	**0**	**0.05(0.01)**	**0.94(0.02)**
	$\varepsilon-$SVQR	500	100	100	0.74(0.03)	0.25(0.01)
	L_1-QR	500	100	100	0.72(0.04)	0.26(0.02)
0.7	L_p-SVQR	**11**	**2.2**	**0**	**0.01(0.01)**	**0.97(0.01)**
	$\varepsilon-$SVQR	500	100	100	0.73(0.04)	0.25(0.02)
	L_1-QR	500	100	100	0.72(0.04)	0.26(0.02)
0.9	L_p-SVQR	**42.8**	**8.56**	**0**	**0.09(0.01)**	**0.83(0.02)**
	$\varepsilon-$SVQR	500	100	100	0.74(0.04)	0.25(0.02)
	L_1-QR	500	100	100	0.72(0.04)	0.26(0.02)

Source: Authors' calculations.

Table 2. Type B simulation results of L_p-SVQR, $\varepsilon-$SVQR and L_1-QR.

τ	Regressor	df	P_1(%)	P_2(%)	NMSE	R^2
0.1	L_p-SVQR	**360.1**	**18**	**10**	**0.02(0.01)**	**0.96(0.01)**
	$\varepsilon-$SVQR	2000	100	100	0.42(0.01)	0.34(0.01)
	L_1-QR	2000	100	100	0.44(0.01)	0.36(0.01)
00.3	L_p-SVQR	**480**	**24**	**10**	**0.03(0.01)**	**0.94(0.01)**
	$\varepsilon-$SVQR	2000	100	100	0.43(0.01)	0.34(0.01)
	L_1-QR	2000	100	100	0.44(0.01)	0.36(0.01)
0.5	L_p-SVQR	**491.6**	**24.6**	**10**	**0.03(0.01)**	**0.93(0.01)**
	$\varepsilon-$SVQR	2000	100	100	0.43(0.01)	0.34(0.01)
	L_1-QR	2000	100	100	0.44(0.01)	0.36(0.01)
0.7	L_p-SVQR	**487.75**	**24.4**	**10**	**0.03(0.01)**	**0.93(0.01)**
	$\varepsilon-$SVQR	2000	100	100	0.43(0.01)	0.34(0.01)
	L_1-QR	2000	100	100	0.44(0.01)	0.36(0.01)
0.9	L_p-SVQR	**399.7**	**19.98**	**10**	**0.03(0.01)**	**0.95(0.01)**
	$\varepsilon-$SVQR	2000	100	100	0.43(0.01)	0.34(0.01)
	L_1-QR	2000	100	100	0.44(0.01)	0.36(0.01)

Source: Authors' calculations.

L_1-QR but with a smaller estimation error. Moreover, L_p-SVQR drives a larger R^2 and smaller NMSE than those of $\varepsilon-$SVQR and L_1-QR. It is clear that L_p-SVQR is more robust to heavy-tailed noise than $\varepsilon-$SVQR and L_1-QR. The main reason is that L_p-SVQR adopts quantile loss and the sparse parameter p. In terms of the running times, the training speed of $\varepsilon-$SVQR is significantly faster than those of L_1-QR and L_p-SVQR.

Figures 1–3 show the absolute estimation error for each case when the values of τ are 0.3, 0.5, and 0.7, respectively. We can see that the regression estimates $\hat{\beta}$ obtained from L_p-SVQR are close to the real values of the regression coefficients β. From Figures 1–3, we can see that the red lines, representing the absolute estimation error of L_p-SVQR in different cases, fluctuate slightly around 0, which indicates that L_p-SVQR selects very few useful variables and captures the statistical information in the test data sets. It can be seen that L_p-SVQR realizes variable selection and regression simultaneously due to its inherent variable selection property. The blue lines, representing the absolute estimation error of L_1-QR in different cases, fluctuate a lot, especially when $\beta_i = 1 \ (i \in \nu)$, which shows that L_1-QR lacks the ability to select useful variables. The absolute estimation

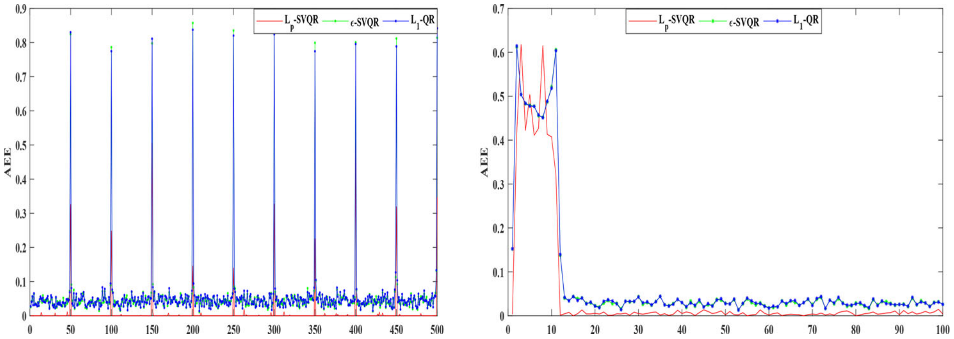
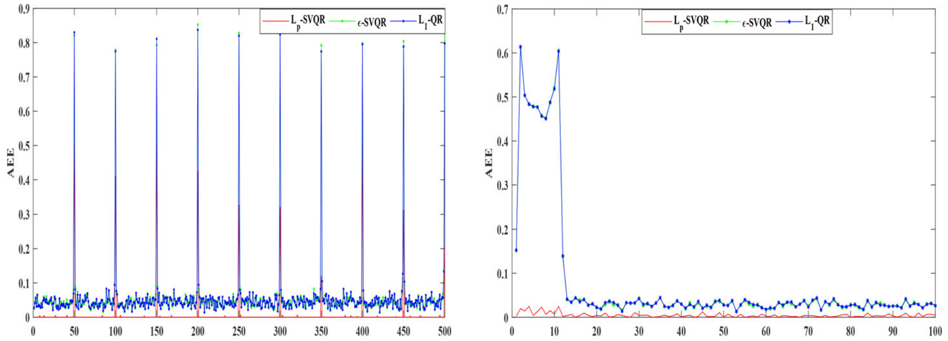


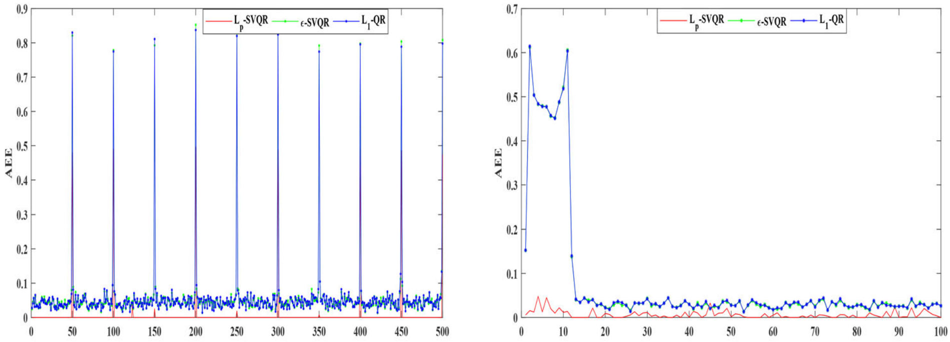
Figure 1. Absolute estimation error simulation with $\tau = 0.3$.
Source: Authors' calculations.



Type A

Type B

Figure 2. Absolute estimation error for simulation with $\tau = 0.5$.
Source: Authors' calculations.



Type A

Type B

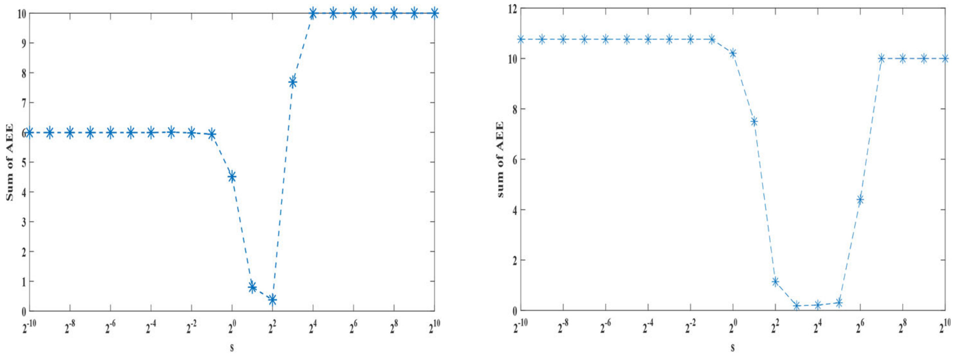
Figure 3. Absolute estimation error for simulation with $\tau = 0.7$.
Source: Authors' calculations.

errors of $\varepsilon-$SVQR have a similar trend to those of L_1-QR, thereby indicating that $\varepsilon-$SVQR lacks sparseness. Therefore, the variable selection results of L_p-SVQR perform significantly better than those of $\varepsilon-$SVQR and L_1-QR.

3.3. Solution path

The solution path of simulation treats $\hat{\beta}$ as a function of s and characterizes how $\hat{\beta}$ changes when s changes. We first fix the parameters τ and p as their optimal values from the experiments. Then, we investigate the influence of the regularization parameter s on the absolute estimation error. Figure 4 shows the sum of absolute estimation error (SAEE) as a function of s. In Figure 4 of the type A, we find that when s changes from 2^{-10} to 2^{-1}, SAEE remains unchanged. When s changes from 2^{-1} to 2^2, the SAEE sharply decreases. When the corresponding value of s is 2^2, the SAEE reaches a minimum value of 0.01. As the regularization parameter s increases from 2^2 to 2^4, the SAEE becomes larger. When s is larger than 2^4, the SAEE reaches a maximum value of 10.

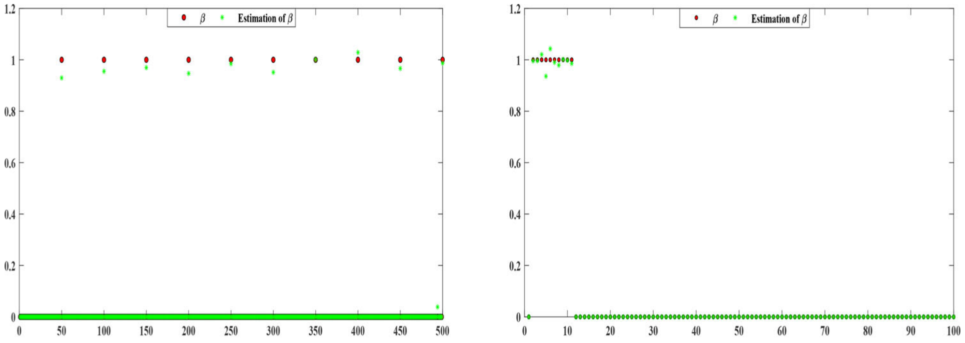
In the type A case, we fix the parameters τ, p, and s to 0.5, 0.3, and 2^2, respectively, and the fitted coefficients are shown in Figure 5 and the corresponding values of $\left|\beta_i - \hat{\beta}_i\right|$ are shown in Figure 6. From Figure 5 of the type A, we find that most fitted coefficients are zero and only 11 fitted coefficients are nonzero, and these



Type A

Type B

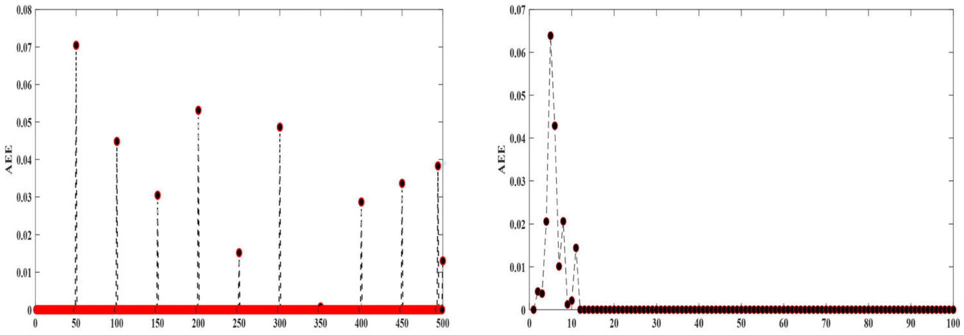
Figure 4. The sum of the absolute estimation error as a function of s.
Source: Authors' calculations.



Type A

Type B

Figure 5. Estimation results ($\hat{\beta}$) with parameters τ, p, and s fixed as the optimal values.
Source: Authors' calculations.



Type A

Type B

Figure 6. Absolute estimation error with the parameters τ, p, and s fixed as the optimal values.
Source: Authors' calculations.

fluctuate slightly around the real value of β_i. From Figure 6 of the type A, we see that most absolute estimation errors are zero and only 10 absolute estimation errors are nonzero. The absolute estimation error reaches a maximum value of 0.07.

The type B solution path treats $\hat{\beta}$ as a function of s. In Figure 4 of the type B, we observe that the SAEE has a similar trend as that of the type A. The SAEE reaches a minimum value of 0.18, where the corresponding value of s is 2^3. Figure 5 of the type B shows the estimation results $\hat{\beta}$ for the type B simulation with the parameters τ, p, and s fixed to 0.5, 0.5, and 2^3, respectively. We observe that most fitted coefficients are zeros and only the first 10 fitted coefficients are nonzero, and these fluctuate slightly around the real value of 1. In Figure 6 of the type B, we see that only $\left|\beta_i - \hat{\beta}_i\right| (i \in \{2, 3, \ldots, 11\})$ are nonzeros and the others are zero. Moreover, the absolute estimation error reaches a maximum value of 0.06.

4. Real data analysis

To further test the sparseness of L_p-SVQR, we discuss the variable selection problem with regard to index construction. Selecting the representative variables from the vast number of potential candidates in a system is a crucial process for index construction since representative variables are usually capable of providing the most important information for management to make decisions. Lasso (Tibshirani, 1996) is a very popular method for the variable selection problem since it shrinks some coefficients of the estimators toward 0. However, Lasso has some drawbacks in terms of the variable selection problem. Fan and Li (2001) found that Lasso uses the same tuning parameters for the regression coefficients, resulting in Lasso suffering an appreciable bias. In addition, Lasso may suffer from the influence of heavy-tailed noise since its least squares loss function is sensitive to noise.

L_p-SVQR can overcome these drawbacks of Lasso. In addition, by using the sparse parameter p, L_p-SVQR is much sparser than the L_1-norm regularization term in Lasso. Thus, L_p-SVQR is an effective method for solving the variable selection problem of index construction. Here, we discuss the variable selection problem of an

Table 3. The regression results in terms of the NMSE for L_p-SVQR, $\varepsilon-$SVQR, and L_1-QR.

Regressor	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
L_p-SVQR	1.82(0.12)	1.87(0.21)	1.73(0.19)	1.83(0.17)	1.75(0.19)
$\varepsilon-$SVQR	1.04(0.05)	1.03(0.03)	1.04(0.04)	1.03(0.03)	1.04(0.03)
L_1-QR	1.20(0.05)	1.06(0.07)	1.02(0.08)	0.82(0.05)	0.96(0.03)

Source: Authors' calculations.

Table 4. Variable selection results.

Group	Variables	Selected(√) or not(×)
Input factors of innovation and entrepreneurship	Fiscal expenditure of science and technology	×
	Funding of science and technology	×
	Loan balance of non-financial institutions	√
	Loan of small enterprise	√
	Loan of science and technology	×
	Investment of technology	×
	Number of talent	√
	Number of new employment	√
	Willingness index of innovation and entrepreneurship	×
	Tax revenue	×
Output factors of innovation and entrepreneurship	Output rate of new industrial products	√
	Ratio of high technology output	×
	Profit ratio of business	√
	Number of new invention patents	√
Vitality of innovation and entrepreneurship	Number of new enterprise	√
	Number of college student employment and entrepreneurship	√
	New registered capital	×
	Number of new registration of trademark	√
	Number of national high-technology enterprises	×
	Number of provincial high-technological enterprises	×
	Market value of listed companies	×
Environment of innovation and entrepreneurship	Amount of innovative vocabulary search	√
	Amount of entrepreneurship vocabulary search	√
	News of future science and technology town	×
	Satisfaction of innovation and entrepreneurship policy	√

Source: Authors' calculations.

innovation and entrepreneurship index. The innovation and entrepreneurship index summarizes information about the state level of innovation and entrepreneurship.

We collected 25 variables with data ranging from the first quarter of 2014 to the second quarter of 2017, all of which were obtained from the Hangzhou Bureau of Statistics (http://tjj.zj.gov.cn/col/col1525563/index.html). The sample size is 14, which is smaller than the number of explanatory variables. Gross domestic product (GDP) is used as a monitor for supervising the selection process. The experimental results indicate that $\varepsilon-$SVQR and L_1-QR select 25 variables, and L_p-SVQR only selects 13 variables when $\tau = 0.1, 0.3, 0.5, 0.7$ and 0.9. Table 3 shows the regression results in terms of the NMSE for L_p-SVQR, $\varepsilon-$SVQR, and L_1-QR. Although L_p-SVQR selects fewer variables than $\varepsilon-$SVQR and L_1-QR, L_p-SVQR obtains regression results that are comparable to those of the other methods. These results prove that L_p-SVQR is a much sparser method than $\varepsilon-$SVQR and L_1-QR.

The variable selection results of L_p-SVQR are shown in Table 4. We find that the loan balance of nonfinancial institutions, loans of small enterprises, number of talents, and number of new employments, as important input factors of innovation and entrepreneurship, are selected. The output rates of the number of new industrial products, profit ratio of business, number of new invention patents, and number of new invention patents, as output factors of innovation and entrepreneurship, are selected. The number of new enterprises, number of college student employments and entrepreneurships, and number of new trademark registrations, representing the vitality of innovation and entrepreneurship, are selected. The amount of innovative vocabulary searches, amount of entrepreneurship vocabulary searches, and satisfaction of innovation and entrepreneurship policies, representing the environment of innovation and entrepreneurship, are selected.

5. Conclusion

Our work focused on the sparse regression problem of high-dimensional data with heavy-tailed noise, especially when the dimensionality of the regressors is larger than the sample size. We proposed L_p-norm support vector quantile regression, which can be considered an extension of the L_1-norm regularized quantile regression method discussed by Li and Zhu (2008), to solve this problem. L_p-norm support vector quantile regression was robust against heavy-tailed noise due to its use of the pinball loss function. The L_p-norm regularization term and the supervised selection process of L_p-SVQR ensured that the representative variables were selected and the redundant variables were discarded automatically. The variable selection performance of L_p-SVQR in the simulation analysis and real data analysis in the $K \gg N$ setting confirmed its robustness and sparseness.

One of the important remaining problems is the application of the proposed high-dimensional sparse method in the real world where high-dimensional data sets are available, such as in economics (Queirós et al., 2019), finance (Bhat et al., 2020) and other fields (Medase & Abdul-Basit, 2020). By adjusting the sparse parameter p, the proposed high-dimensional sparse method ensures that the representative variables are selected. Moreover, using the quantile parameter, the proposed high-dimensional sparse method is robust against heavy-tailed noise. Therefore, how to apply the proposed high-dimensional sparse method to high- dimensional variable selection problem in the real world is our future work. For example, multi-factor quantitative stock selection is a typical high-dimensional research problem, and it is interesting to apply the proposed method to select the principal factors of stock return.

Funding

This work was supported by Philosophy and Social Sciences Leading Talent Training Project of Zhejiang Province (No. 21YJRC07-1YB).

ORCID

Yuanhai Shao http://orcid.org/0000-0002-1628-6133

References

Algamal, Z. Y., & Lee, M. H. (2019). A two-stage sparse logistic regression for optimal gene selection in high-dimensional microarray data classification. *Advances in Data Analysis and Classification*, *13*(3), 753–771. https://doi.org/10.1007/s11634-018-0334-1

Anand, P., Rastogi, R., & Chandra, S. (2020). A new asymmetric ϵ-insensitive pinball loss function based support vector quantile regression model. *Applied Soft Computing*, *94*, 106473. https://doi.org/10.1016/j.asoc.2020.106473

Belloni, A., & Chernozhukov, V. (2011). L_1-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics*, *39*(1), 82–130. https://doi.org/10.1214/10-AOS827

Berk, R., & Hwang, J. T. (1989). Optimality of the least squares estimator. *Journal of Multivariate Analysis*, *30*(2), 245–254. https://doi.org/10.1016/0047-259X(89)90038-9

Bertsimas, D., King, A., & Mazumder, R. (2016). Best subset selection via a modern optimization lens. *The Annals of Statistics*, *44*(2), 813–852.

Bhat, K. U., Chen, S., Chen, Y., & Jebran, K. (2020). Debt capacity, debt choice, and underinvestment problem: Evidence from China. *Economic Research-Ekonomska Istraživanja*, *33*(1), 267–287. https://doi.org/10.1080/1331677X.2019.1699438

Chen, X., Liu, W., Mao, X., & Yang, Z. (2020). Distributed high-dimensional regression under a quantile loss function. *Journal of Machine Learning Research*, *21*(182), 1–43.

Clarke, R., Ressom, H. W., Wang, A., Xuan, J., Liu, M. C., Gehan, E. A., & Wang, Y. (2008). The properties of high-dimensional data spaces: Implications for exploring gene and protein expression data. *Nature Reviews. Cancer*, *8*(1), 37–49.

Dempster, A. P., Schatzoff, M., & Wermuth, N. (1977). A simulation study of alternatives to ordinary least squares. *Journal of the American Statistical Association*, *72*(357), 77–91. https://doi.org/10.1080/01621459.1977.10479910

Dondelinger, F., Mukherjee, S., & Alzheimer's Disease Neuroimaging Initiative. (2020). The joint lasso: High-dimensional regression for group structured data. *Biostatistics*, *21*(2), 219–235. https://doi.org/10.1093/biostatistics/kxy035

Drucker, H., Burges, C. J., Kaufman, L., Smola, A., & Vapnik, V. (1996). Support vector regression machines. *Advances in Neural Information Processing Systems*, *9*, 155–161.

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, *96*(456), 1348–1360. https://doi.org/10.1198/016214501753382273

Fan, J., & Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, *20*(1), 101–148.

Fan, J., Li, Q., & Wang, Y. (2017). Estimation of high dimensional mean regression in the absence of symmetry and light tail assumptions. *Journal of the Royal Statistical Society, Series B*, *79*(1), 247–265.

Fan, J., Lv, J., & Qi, L. (2011). Sparse high-dimensional models in economics. *Annual Review of Economics*, *3*, 291–317. [22022635]

Greene, W. H. (1981). On the asymptotic bias of the ordinary least squares estimator of the Tobit model. *Econometrica*, *49*(2), 505–513. https://doi.org/10.2307/1913323

He, Q., Xu, L., & Men, Y. (2020). Composition effect matters: Decomposing the gender pay gap in Chinese university graduates. *Economic Research-Ekonomska Istraživanja*, *33*(1), 847–864. https://doi.org/10.1080/1331677X.2020.1734850

Hsu, D., & Sabato, S. (2016). Loss minimization and parameter estimation with heavy tails. *The Journal of Machine Learning Research*, *17*(1), 543–582.

Huang, X., Li, G., Zhang, J., Li, L., & Xu, X. (2020). Rise in house prices and industrial growth: Evidence from city-level data of China. *Economic Research-Ekonomska Istraživanja*, *34*(1), 1–19.

Hutcheson, G. D. (2011). Ordinary least-squares regression. *L. Moutinho and GD Hutcheson, the SAGE Dictionary of Quantitative Management Research*, 224–228.

Kim, Y., Hao, J., Mallavarapu, T., Park, J., & Kang, M. (2019). Hi-LASSO: High-dimensional LASSO. *IEEE Access.*, *7*, 44562–44573. https://doi.org/10.1109/ACCESS.2019.2909071

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*(1), 33–50. https://doi.org/10.2307/1913643

Li, Q., Xi, R., & Lin, N. (2010). Bayesian regularized quantile regression. *Bayesian Analysis*, *5*(3), 533–556. https://doi.org/10.1214/10-BA521

Li, Y., & Zhu, J. (2008). L_1-norm quantile regression. *Journal of Computational and Graphical Statistics*, *17*(1), 163–185. https://doi.org/10.1198/106186008X289155

Liang, X., & Jacobucci, R. (2020). Regularized structural equation modeling to detect measurement bias: Evaluation of lasso, adaptive lasso, and elastic net. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(5), 722–734. https://doi.org/10.1080/10705511.2019.1693273

Liu, Y., Zeng, P., & Lin, L. (2020). Generalized L_1-penalized quantile regression with linear constraints. *Computational Statistics & Data Analysis*, *142*, 106819. https://doi.org/10.1016/j.csda.2019.106819

Mangasarian, O. L. (2007). Absolute value programming. *Computational Optimization and Applications*, *36*(1), 43–53. https://doi.org/10.1007/s10589-006-0395-5

Medase, S. K., & Abdul-Basit, S. (2020). External knowledge modes and firm-level innovation performance: Empirical evidence from sub-Saharan Africa. *Journal of Innovation & Knowledge*, *5*(2), 81–95. https://doi.org/10.1016/j.jik.2019.08.001

Newey, W. K., & Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, *55*(4), 819–847. https://doi.org/10.2307/1911031

Peng, B., & Wang, L. (2015). An iterative coordinate descent algorithm for high-dimensional nonconvex penalized quantile regression. *Journal of Computational and Graphical Statistics*, *24*(3), 676–694. https://doi.org/10.1080/10618600.2014.913516

Peng, X., & Xu, D. (2013). A local information-based feature-selection algorithm for data regression. *Pattern Recognition*, *46*(9), 2519–2530. https://doi.org/10.1016/j.patcog.2013.02.010

Queirós, M., Braga, V., & Correia, A. (2019). Cross-country analysis to high-growth business: Unveiling its determinants. *Journal of Innovation & Knowledge*, *4*(3), 146–153. https://doi.org/10.1016/j.jik.2018.03.006

Rzhetsky, A., & Nei, M. (1992). Statistical properties of the ordinary least-squares, generalized least-squares, and minimum-evolution methods of phylogenetic inference. *Journal of Molecular Evolution*, *35*(4), 367–375. https://doi.org/10.1007/BF00161174

Shim, J., & Hwang, C. (2009). Support vector censored quantile regression under random censoring. *Computational Statistics & Data Analysis*, *53*(4), 912–919. https://doi.org/10.1016/j.csda.2008.10.037

Shim, J., Kim, Y., Lee, J., & Hwang, C. (2012). Estimating value at risk with semiparametric support vector quantile regression. *Computational Statistics*, *27*(4), 685–700. https://doi.org/10.1007/s00180-011-0283-z

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*(3), 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, *9*(6), 1135–1151. https://doi.org/10.1214/aos/1176345632

Takeuchi, I., Le, Q. V., Sears, T. D., & Smola, A. J. (2006). Nonparametric quantile estimation. *Journal of Machine Learning Research*, *7*(July), 1231–1264.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Wang, H., Li, G., & Jiang, G. (2007). Robust regression shrinkage and consistent variable selection through the LAD-Lasso. *Journal of Business & Economic Statistics*, *25*(3), 347–355. https://doi.org/10.1198/073500106000000251

Wang, M., Kang, X., & Tian, G. L. (2020). Modified adaptive group lasso for high-dimensional varying coefficient models. *Communications in Statistics-Simulation and Computation*, 1–16.

Ye, Y. F., Jiang, Y. X., Shao, Y. H., & Li, C. N. (2015). Financial conditions index construction through weighted L_p-norm support vector regression. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, *19*(3), 397–406. https://doi.org/10.20965/jaciii.2015.p0397

Ye, Y. F., Shao, Y. H., Deng, N. Y., Li, C. N., & Hua, X. Y. (2017a). Robust L_p-norm least squares support vector regression with feature selection. *Applied Mathematics and Computation*, *305*, 32–52. https://doi.org/10.1016/j.amc.2017.01.062

Ye, Y. F., Ying, C., Jiang, Y. X., & Li, C. N. (2017b). L_1-norm least squares support vector regression via the alternating direction method of multipliers. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, *21*(6), 1017–1025. https://doi.org/10.20965/jaciii.2017.p1017

Yu, L., Lin, N., & Wang, L. (2017). A parallel algorithm for large-scale nonconvex penalized quantile regression. *Journal of Computational and Graphical Statistics*, *26*(4), 935–939. https://doi.org/10.1080/10618600.2017.1328366

Yuan, M. (2006). GACV for quantile smoothing splines. *Computational Statistics & Data Analysis*, *50*(3), 813–829. https://doi.org/10.1016/j.csda.2004.10.008

Zhang, C., Li, D., & Tan, J. (2013). The support vector regression with adaptive norms. *Procedia Computer Science*, *18*, 1730–1736. https://doi.org/10.1016/j.procs.2013.05.341

Zhou, W. X., Bose, K., Fan, J., & Liu, H. (2018). A new perspective on robust M-estimation: Finite sample theory and applications to dependence-adjusted multiple testing. *Annals of Statistics*, *46*(5), 1904.

Zhou, Y., Uddin, M. S., Habib, T., Chi, G., & Yuan, K. (2020). Feature selection in credit risk modeling: An international evidence. *Economic Research-Ekonomska Istraživanja*, 1–31. https://doi.org/10.1080/1331677X.2020.1842225

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, *101*(476), 1418–1429. https://doi.org/10.1198/016214506000000735