

# A One-Step Variable Selection Procedure for SCAD Penalized Quantile Regression Models

Jan G. De Gooijer

Amsterdam School of Economics, University of Amsterdam, Amsterdam, Netherlands

Email: [j.degooijer@contact.uva.nl](mailto:j.degooijer@contact.uva.nl)

**How to cite this paper:** De Gooijer, J. G. (2026). A One-Step Variable Selection Procedure for SCAD Penalized Quantile Regression Models. *Theoretical Economics Letters*, 16, 57-71.

<https://doi.org/10.4236/tel.2026.161004>

**Received:** October 26, 2025

**Accepted:** January 6, 2026

**Published:** January 9, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

Variable selection using penalized estimation methods in quantile regression models is an important step in screening for relevant covariates. In this paper, we present a one-step estimation procedure for variable selection in sparse, linear additive quantile regression models, using the SCAD penalty. The main idea of the proposed procedure is that the usual  $L_1$ -norm objective function in quantile regression estimation is replaced by a smooth parametric approximation of this function, via iterative least squares computations. We conduct a simulation study and a real data analysis to check the finite sample performance of the one-step estimator. The results reveal that the one-step quantile SCAD method identifies relevant variables efficiently even when the process under study has heavy tails or if the process is contaminated with outliers.

## Keywords

High-Dimensional Data, Iterative Least Squares, Quantile Regression, Variable Selection

---

## 1. Introduction

Analyzing high-dimensional data has become an important task in diverse research fields such as biology, genetics, economics, and finance<sup>1</sup>. It has gained increasing attention in the application of emerging statistical methods that are proficient in analyzing and interpreting large datasets. Selection of the best representative covariate of the response variable is a key part of this process, but the traditional variable (or component) selection methods like subset selection, step-

---

<sup>1</sup>Throughout the paper the number of variables is denoted by  $p$  and the number of measurements (observations) by  $n$ . With high-dimensional data, we refer to the case that  $p$  is assumed to be large, possibly larger than  $n$ .

wise selection, backward and forward selection are not stable and accurate when the size of covariates is near or equal to the size of observations. These procedures are also not applicable when the number of observations is less than the number of covariates.

Multiple studies on penalized regression methods have addressed the above drawbacks. For example, the LASSO (least absolute shrinkage and selection operator) penalty was proposed by Tibshirani (1996), the adaptive LASSO by Zou (2006), the elastic net by Zou and Hastie (2005), and the smoothly clipped absolute deviation (SCAD) penalty by Fan and Li (2001). These methods are good with respect to prediction accuracy and efficiency. They also have the oracle property, but are limited to the small size paradigm of high-dimensional data (Fan & Li, 2001)<sup>2</sup>. Also, these methods work well with a fixed number of covariates. However, in the case of high-dimensional data, both classical and penalized techniques face problems like speed, stability and accuracy (Fan & Lv, 2008).

To overcome these issues, various variable selection methods in the context of quantile regression (QR) have been proposed. These include penalized quantile regression approaches using the LASSO penalty, and those using a non-convex penalty such as SCAD; see, e.g., Wang et al. (2012), Sherwood and Wang (2016), Khan et al. (2019) and the references therein. More recently, various penalized quantile regression approaches have been developed. For instance, Sherwood and Maidman (2022) proposed a method for simultaneous estimation and variable selection of an additive nonlinear quantile regression model in a high-dimensional setting. The R-rqPen package of Sherwood et al. (2025) readily solves the problem. Song et al. (2022) consider, as a generalized regression model, the single-index regression model with the SCAD and Laplace error penalty. Additionally, Chen and Lee (2023) addressed sparse linear quantile regression through the so-called  $\ell_0$  approach, which focuses on selecting a subset of covariates; see, e.g., Mai (2024) for an extensive literature review.

Using the SCAD penalty, the present paper introduces a computationally feasible and efficient approach, called one-step quantile estimation procedure, for selecting the nonzero QR parameters. The proposed method is similar in spirit to the one-step local linear approximation estimator of Zou and Li (2008). Given ordinary least squares estimates as initial values, these authors obtain a one-step estimator for the nonzero linear regression parameters by maximizing a penalized likelihood function. We use the same idea within the context of a penalized additive QR model<sup>3</sup>. However, we replace the usual  $L_1$ -norm objective function by a smooth parametric approximation of this function, via iterative least squares com-

<sup>2</sup>The oracle property of a method means that it can correctly select the nonzero parameters with probability converging to one, and that the estimators of the nonzero parameters are normal with the same mean and covariance that they would have if there is *a priori* knowledge about the active covariates.

<sup>3</sup>The concept of additivity, as a simplifying structure, is basic to many proposed models in science. Within economics, Deaton and Muelbauer (1980) provide many microeconomic examples in which the structure is convenient for analysis and important for interpretability. Indeed, additivity is widely used in parametric, semi- and nonparametric analysis of economic data.

putations. This approach offers an efficient way to obtain conditional quantile estimates in high-dimensions.

The rest of the paper is organized as follows. Section 2 briefly describes the standard QR minimization problem. Section 3 first introduces an initial QR estimator without variable selection. Next, a final QR estimator with variable selection is proposed. Finally, we discuss SCAD consistency in variable selection in this section. In Section 4, we assess the finite-sample properties of the estimator through a Monte Carlo simulation study. This section also contains a comparison with a similar one-step LASSO estimator. Section 5 contains a real data illustration. Section 6 provides some summary remarks.

## 2. Preliminaries

Let  $Y_i \in \mathbb{R}$  be a response variable which depends on a  $p_n \times 1$  vector of random covariates  $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,p_n})^T \in \mathbb{R}^{p_n}$ , ( $p_n > 2$ ) of the  $i$ th observation,  $i = 1, \dots, n$ . In the “traditional”, additive QR framework, interest is focussed on the  $\tau$ th ( $0 < \tau < 1$ ) conditional quantile of  $Y_i$  given  $\mathbf{X}_i = \mathbf{x}_i$ , i.e.,

$$Q_{Y_i}(\tau | \mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta}_0(\tau), i = 1, \dots, n, \tag{1}$$

where  $\boldsymbol{\beta}_0(\tau) = (\beta_{1,0}(\tau), \dots, \beta_{p_n,0}(\tau))^T$  is an unknown  $p_n \times 1$  parameter vector, including an intercept, depending on  $\tau$ . Note that for including an intercept term, the corresponding covariate needs to take value 1 for each observation.

For a fixed value of  $\tau$ , the parameter vector  $\boldsymbol{\beta}_0(\tau)$  can be estimated by minimizing the objective function

$$\mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^n \rho_\tau(r_i(\boldsymbol{\beta})), \tag{2}$$

where  $r_i(\boldsymbol{\beta}) = Y_i - \mathbf{X}_i^T \boldsymbol{\beta}$ ,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ ,  $\rho_\tau(u) = (\tau - I(u \leq 0))u$  is the quantile check function, and  $I(\cdot)$  is the usual indicator function. To simplify the notation from now on, we suppress the argument  $\boldsymbol{\beta}$  of  $r_i(\boldsymbol{\beta})$ . We also omit the subscript “ $n$ ” from  $p_n$  whenever no confusion is caused.

Linear programming algorithms can be employed to solve Equation (2). To describe the basic solution (feasible or not) of the QR minimization problem, we first introduce the following additional notations. In particular, let  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_p)$ , and let  $h = (i_1, \dots, i_p)$  denote a  $p$ -dimensional index, which identifies proper  $p$  observations in the sample of size  $n$ . Further, let  $\mathbf{X}(h)$  denote the  $p \times p$  submatrix of  $\mathbf{X}$  with rows  $\{\mathbf{X}_i, i \in h\}$ . Likewise, let  $\mathbf{Y}(h)$  be a  $p$ -dimensional vector with coordinates  $\{Y_i, i \in h\}$ . Then the basic solution (Abdelmalek, 1974; Koenker & Bassett, 1978) of the QR minimization problem may be expressed as

$$\hat{\boldsymbol{\beta}}(\tau) = \mathbf{X}(h)^{-1} \mathbf{Y}(h). \tag{3}$$

Clearly, only the  $p$  data points identified by  $h$  supply information about the QR estimator of  $\boldsymbol{\beta}_0(\tau)$ , although the whole observed sample is typically used to find the “optimal”  $h$ . Observe that the design matrix  $\mathbf{X}$  is partitioned into a critical active dataset, whose cardinality is  $p$ , and a non-active complement,

yielding  $Y(h) - X(h)\hat{\beta}(h) = \mathbf{0}_p$  with  $\mathbf{0}_p$  a  $p$ -dimensional zero vector.

### 3. Description of the QR Estimators

It is well known that conditional QR estimates are severely affected by the so-called curse-of-dimensionality when  $p$  is large. To tackle this issue, several authors have proposed penalized QR methods which identify relevant and irrelevant components consistently. In this section, first we discuss an initial estimator for  $\beta_0(\tau)$  based on a parametric smooth approximation of the objective function (2). Then we will explain how to use it in variable selection when constructing a one-step sparse conditional quantile estimator.

#### 3.1. An Initial QR Estimator without Variable Selection

Muggeo et al. (2011) proposed an exact path-following algorithm leading to the QR solution (3). For a fixed value of  $\tau$ , they approximate  $\rho_\tau(r_i)$  by an asymmetric “bent arc” expressed as a piecewise polynomial denoted by  $S_{i,\tau}(\beta; c)$ , where  $c$  is a small but known positive constant. Specifically, let  $\mathcal{A}(c) = \{i : -c\tau < r_i < c(1-\tau)\}$ ,  $i = 1, \dots, n$ , be a partition of the sample units. Then the approximation  $S_{i,\tau}(\beta; c)$  is defined as follows

$$S_{i,\tau}(\beta; c) = \frac{\omega_{i,2}(\beta; c)r_i^2}{2c} + \omega_{i,1}(\beta; c)r_i + I(i \in \mathcal{A}(c))\frac{c\tau(1-\tau)}{2}, \quad (4)$$

where  $\omega_{i,u}(\beta; c)$ ,  $u = 1, 2$ , are weights given by

$$\begin{aligned} \omega_{i,1}(\beta; c) &= (\tau - 1)v_{i,1}(\beta; c) + \tau v_{i,4}(\beta; c), \\ \omega_{i,2}(\beta; c) &= \frac{(1-\tau)^2 v_{i,2}(\beta; c) + \tau^2 v_{i,3}(\beta; c)}{\tau(1-\tau)}, \end{aligned} \quad (5)$$

and where  $v_{i,g}$ ,  $g = 1, 2, 3, 4$ , are the following indicator functions

$$\begin{aligned} v_{i,1}(\beta; c) &= I(r_i \leq -c\tau), \quad v_{i,2}(\beta; c) = I(-c\tau < r_i \leq 0), \\ v_{i,3}(\beta; c) &= I(0 < r_i < c(1-\tau)), \quad v_{i,4}(\beta; c) = I(r_i \geq c(1-\tau)). \end{aligned}$$

Note that we suppressed the dependence of  $\omega_{i,u}$  and  $v_{i,g}$  on  $\tau$ . This simplification in notation also holds in the rest of the paper.

The smooth objective function takes the form

$$\begin{aligned} S_\tau(\beta; c) &= \sum_{i=1}^n S_{i,\tau}(\beta; c) \\ &= \frac{1}{2c} \sum_{i \in \mathcal{A}(c)} \omega_{i,2}(\beta; c)r_i^2 + \sum_{i \in (1-\mathcal{A}(c))} \omega_{i,1}(\beta; c)r_i + |\mathcal{A}(c)|\frac{c\tau(1-\tau)}{2}, \end{aligned} \quad (6)$$

where  $\sum_i I(i \in \mathcal{A}(c)) = |\mathcal{A}(c)|$ , i.e., the cardinality of the set  $\mathcal{A}(c)$ . Now for minimizing Equation (6), the idea is to define a piecewise linear solution using a decreasing sequence of values, say  $0 < c_j < \dots < c_1$ , such that  $\mathcal{A}(c_{j+1}) \subseteq \mathcal{A}_j$  and  $|\mathcal{A}(c_j)| = p$ . Additionally, let  $c_j > 0$  be a fixed value and  $\beta(c_j)$  the corresponding minimizer of (6), with  $c_j$  chosen such that  $|\mathcal{A}(c_j)| > p$ . Given this setup, the set of constant weight indicator functions is given by

$\mathcal{D}(c_j) = \{(\boldsymbol{\beta}; c) \in \mathbb{R}^p \times \mathbb{R}^+ \mid \forall i, g : v_{i,g}(\boldsymbol{\beta}; c) = v_{i,g}(c_j)\}$ . Then the  $p$ -dimensional gradient vector of  $S_\tau(\boldsymbol{\beta}; c)$  can be written as follows

$$\frac{\partial S_\tau(\boldsymbol{\beta}; c)}{\partial \boldsymbol{\beta}} = \frac{1}{c} \mathbf{X}^T \mathbf{W}_2(c_j) (\mathbf{Y} - \mathbf{X} \boldsymbol{\beta}) + \mathbf{X}^T \mathbf{W}_1(c_j) \mathbf{1}_n \quad \forall (\boldsymbol{\beta}, c) \in \mathcal{D}(c_j), \quad (7)$$

where  $\mathbf{W}_u(c_j) = \text{diag}(\omega_{1,u}(c_j), \dots, \omega_{n,u}(c_j))$ ,  $j = 1, \dots, J$ ,  $u = 1, 2$ . Here  $\mathbf{Y} = (Y_1, \dots, Y_n)^T$  and  $\mathbf{1}_n$  is an  $n$ -dimensional vector of ones. Solving

$\partial S_\tau(\boldsymbol{\beta}; c) / \partial \boldsymbol{\beta} = \mathbf{0}_p$  gives the  $p$ -dimensional weighted least squares estimator

$$\begin{aligned} \hat{\boldsymbol{\beta}}_\tau(c) &= (\mathbf{X}^T \mathbf{W}_2(c_j) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_2(c_j) \mathbf{Y} + c (\mathbf{X}^T \mathbf{W}_2(c_j) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_1(c_j) \mathbf{1}_n \\ &= \hat{\boldsymbol{\beta}}_Y(c_j) + c \mathbf{v}(c_j), \end{aligned} \quad (8)$$

where

$$\begin{aligned} \hat{\boldsymbol{\beta}}_Y(c_j) &= (\mathbf{X}^T \mathbf{W}_2(c_j) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_2(c_j) \mathbf{Y}, \\ \mathbf{v}(c_j) &= (\mathbf{X}^T \mathbf{W}_2(c_j) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_1(c_j) \mathbf{1}_n. \end{aligned} \quad (9)$$

For every new value of  $c_j$ , information from the previous solution is used. Then, it is not hard to show that

$$c_{j+1} = \max_i \left\{ \frac{\hat{r}_{i,Y}(c_j)}{\mathbf{X}_i^T \mathbf{v}(c_j) - \tau}; \frac{\hat{r}_{i,Y}(c_j)}{\mathbf{X}_i^T \mathbf{v}(c_j)}; \frac{\hat{r}_{i,Y}(c_j)}{\mathbf{X}_i^T \mathbf{v}(c_j) - \tau + 1} \right\}_{(0,c_j)}, \quad (10)$$

where  $\hat{r}_{i,Y}(c_j) = Y_i - \mathbf{X}_i^T \hat{\boldsymbol{\beta}}_Y(c_j)$  is an estimate of the random error  $r_{i,Y}(\tau) = Y_i - \mathbf{X}_i^T \boldsymbol{\beta}_0(\tau)$ . Thus if  $|\mathcal{A}(c_{j+1})| = p$ , then  $\hat{\boldsymbol{\beta}}_\tau(c) = \mathbf{X}^{-1}(h) \mathbf{Y}(h)$ , with  $h \equiv |\mathcal{A}(c_j)|$ , is an iteratively weighted least squares (IWLS) estimator of  $\boldsymbol{\beta}_0(\tau)$ .

Suppose that we have  $\mathbf{x} = (x_1, x_2, \dots)^T$  with sign vector  $\text{sign}(\mathbf{x}) = (\text{sign}(x_1), \text{sign}(x_2), \dots)^T$ . The proposed path-following algorithm is based on the following theorem.

**Theorem 1.** (Muggeo et al., 2011) Let  $c_j > 0$  be a fixed value such that  $|\mathcal{A}(c_j)| = p$ . Then under the assumption that for any  $c \in [0, c_j]$ ,  $\text{sign}(r_\ell(\hat{\boldsymbol{\beta}}_\tau(c))) = \text{sign}(r_\ell(\hat{\boldsymbol{\beta}}_\tau(c_j)))$ ,  $\forall \ell \in 1 - |\mathcal{A}(c_j)|$ , the solution curve evaluated at  $0_p$  is equal to the solution of (2), namely  $\hat{\boldsymbol{\beta}}_\tau(0) = \hat{\boldsymbol{\beta}}(\tau)$ .

### 3.2. A Final QR Estimator with Variable Selection

It is well known that conditional QR estimates are severely affected by the so-called curse-of-dimensionality when  $p$  is large. To tackle this issue, several authors have proposed dimension reduction methods. The general framework for these methods can be characterized by a set of nonzero coefficients

$A = \{j : \beta_{j,0}(\tau) \neq 0, j = 1, \dots, p\}$ , and  $q = |A|$  is the cardinality of  $A$ . The set  $A$  is unknown and can be estimated by various (non)parametric approaches such as kernel-based smoothing or B-spline basis functions. Suppose that the true parameter vector  $\boldsymbol{\beta}_0(\tau)$  is sparse. Hence, we can write  $\boldsymbol{\beta}_0(\tau) = (\boldsymbol{\beta}_{1,0}^T(\tau), \boldsymbol{\beta}_{2,0}^T(\tau))^T$ , where  $\boldsymbol{\beta}_{1,0}(\tau)$  is a  $q \times 1$  vector and  $\boldsymbol{\beta}_{2,0}(\tau) = \mathbf{0}_{p-q}$ . Moreover, it is unknown which coefficients are nonzero and which are zero. To this end, one can use a penalty function  $p_\lambda(\cdot)$  with tuning parameter  $\lambda$  which depends on  $n$ , and

where for ease of presentation, we set  $\lambda \equiv \lambda_n$ .

In this paper, we consider the SCAD penalty which is defined on  $\mathbb{R}^+$  by its first derivatives as

$$p'_\lambda(x) = \lambda \left\{ I(x \leq \lambda) + \frac{(a\lambda - x)_+}{a-1} I(x > \lambda) \right\}, \tag{11}$$

where  $a > 2$  and  $\lambda > 0$  are tuning parameters, and where  $(x)_+ = \max(x, 0)$ . Note that (11) is symmetric, non-convex on  $[0, \infty)$ , and singular at the origin. Its derivative vanishes outside  $[-a\lambda, a\lambda]$ . Following the recommendation of [Fan and Li \(2001\)](#), we use  $a = 3.7$  in the numerical applications. The parameter  $\lambda$  controls the complexity of the selected model, and  $\lambda \rightarrow 0$  as  $n \rightarrow \infty$ . The function (11) overcomes some of the undesirable properties of the LASSO penalty, where  $p_\lambda(x) = \lambda|x|$ .

For a fixed  $c$ , let  $\hat{\beta}_{j,\tau}(c)$ ,  $j = 1, \dots, p$ , denote the univariate conditional (marginal) quantile estimator of  $\beta_{j,0}(\tau)$  obtained through the IWLS method of Section 3.1. Then, for given values of  $\lambda$  and  $\tau$ , the proposed final one-step conditional QR estimator  $\hat{\beta}_{\lambda,\tau}(c)$  of  $\beta_0(\tau)$  can be obtained by minimizing the following penalized objective function

$$Q_{\lambda,\tau}(\beta; c) = \frac{1}{2n} \sum_{i=1}^n \rho_\tau(Y_i - X_i^T \beta) + \sum_{j=1}^p p'_\lambda \left( \left| \hat{\beta}_{j,\tau}(c) \right| \right) |\beta_j|. \tag{12}$$

Thus, we use  $\hat{\beta}_{j,\tau}(c)$  as an initial estimate to determine the penalty weights for individual coefficients in QR. This adaptive approach is a natural extension of the local linear approximation as proposed by [Zou and Li \(2008\)](#) and briefly discussed in Section 1. The approach allows the procedure to apply smaller penalties to coefficients with strong evidence of being non-zero, thereby improving selection accuracy<sup>4</sup>.

Let  $|\hat{A}_{\lambda,\tau}|$  be the number of nonzero estimates. Then, following [Sherwood and Wang \(2016\)](#), we choose the value of  $\lambda$  that minimizes the following high-dimensional Bayesian-type information criterion (BIC):

$$\text{QBIC}_\tau(\lambda) = \log \left( \sum_{i=1}^n \rho_\tau(Y_i - X_i^T \hat{\beta}_{\lambda,\tau}(c)) \right) + |\hat{A}_{\lambda,\tau}| \frac{\log(p) \log(\log(n))}{2n}. \tag{13}$$

In the case  $\hat{\beta}_{j,\tau}(c)$  in (12) is replaced by an initial estimator not depending on  $c$ , simulation results by [Lee et al. \(2014\)](#) show that if the choice of  $\lambda$  is based on (13), all relevant components are selected consistently as  $n \rightarrow \infty$ . Further, similar to Theorem 1, it follows that  $\hat{\beta}_{\lambda,\tau}(c)$  is equal in sign to the true  $\beta_0(\tau)$  (denoted by  $\hat{\beta}_{\lambda,\tau}(c) =_s \beta_0(\tau)$ ) if and only if  $\text{sign}(\hat{\beta}_{\lambda,\tau}(c)) = \text{sign}(\beta_0(\tau))$ .

### 3.3. Asymptotic Properties

Due to the non-smoothness and non-convexity of the SCAD penalty function (11), the classical Karush-Kuhn-Tucker optimality condition is not applicable to study

<sup>4</sup>Note that  $|\beta_j|$  in (12) can be written as  $\rho_\tau(\beta_j) + \rho_\tau(-\beta_j)$ , the above minimization problem can be restated as an unpenalized weighted QR problem; see, e.g., [Sherwood and Wang \(2016\)](#).

the asymptotic properties of the final estimator  $\hat{\beta}_{\lambda,\tau}(c)$ . One common used technique to get around it in QR is to decompose a non-convex objective function as the difference of two convex (check loss) functions; see, e.g., [Wu and Liu \(2009\)](#) and [Sherwood and Wang \(2016\)](#). Using this technique to decompose the SCAD function, and assuming that certain regularity conditions hold, the following theorem can be proven.

**Theorem 2.** (Consistency) Assume that  $p'_\lambda(\cdot) = \lambda p'(\cdot)$ . Suppose  $p'(\cdot)$  is a continuous function on  $(0, \infty)$  and there is some  $k > 0$  such that  $p'(\theta) = O(\theta^{-k})$  as  $\theta \rightarrow 0^+$ , and  $q^{1/2}n^{-k} \rightarrow 0$ . In addition, let  $p'_\lambda(\cdot)$  be the SCAD function and  $\lambda n^k \rightarrow \infty$ , then we have  $\mathbb{P}(\hat{\beta}_{\lambda,\tau}(c) =_s \beta_0(\tau)) \rightarrow 1$ .

The result follows from similar arguments as in the proof of Theorem 4.2(i) in [Noh and Lee \(2014\)](#) and, hence, has been omitted from the text. [Zhao and Yu \(2006\)](#) study the consistency of LASSO variable selection in a high-dimensional linear regression setting.

**Remark** (Oracle property) Under certain regularity conditions on the design of the model, [Amin et al. \(2015\)](#), Theorem 3.2, proved sparsity and asymptotic normality of the SCAD-penalized quantile estimator  $\hat{\beta}_{\lambda,\tau}(c)$ .

### 3.4. Selecting $c$

Following [Muggeo et al. \(2011\)](#), we now present an algorithm for selecting  $c$ . Ideally,  $c$  should be small enough to get a good approximate solution to the objective function (12). However, for small values of  $c$  estimation will be more difficult as the objective function  $S_\tau(\beta; c)$  approaches  $\mathcal{L}(\beta)$  having a step-function gradient. The proposed pseudo code is given by [Algorithm 1](#). High values of  $P$  lead to high  $c^*$  and then to better convergence performance but to more biased estimates; small values of  $P$  are likely to cause failures in the estimating algorithm. Throughout the next two sections, we set  $c_{\text{start}} = 0.5$  and  $P = 0.7$ . These choices are recommended by [Muggeo et al. \(2011\)](#). In particular, these authors provide guidance on selecting  $c$  and  $p$  when  $n$  is large ( $n \geq 10000$ ). Among other things, their results indicate that it is possible to reduce  $P$  (or equivalently  $c$ ) without affecting the convergence performance of the IWLS.

---

```

Choose  $c_{\text{start}} > 0$ , and use this value to compute the starting value of the least square estimator (8), say
 $\hat{\beta}_{\text{start}} = (\hat{\beta}_{1,\text{start}}, \dots, \hat{\beta}_{p,\text{start}})^T$ .
while  $h < \text{maximum iteration (say, 100)}$  do
  for  $i = 1, \dots, n$  do
    • Calculate  $(h + 1)$ th iteration of the quantile corresponding to the  $P$ th percentile of the positive
      (negative) current set of residuals  $r_i(\hat{\beta}_{\text{start}}) = Y_i - \mathbf{X}_i^T \hat{\beta}_{\text{start}}$  when  $\tau = 0.5$  ( $\tau > 0.5$ ), and denote
      this value by  $c^*$ .
    • Calculate the new indicator functions  $v_{i,g}(\beta, c^*)$ , ( $g = 1, \dots, 4$ ), and a new estimate of  $\hat{\beta}_\tau(c)$ , say
       $\hat{\beta}_\tau^{(h+1)}(c^*)$ .
  end for
  if
     $\max_{1 \leq j \leq p} (\|(\hat{\beta}_{j,\text{start}}^{(h+1)} - \hat{\beta}_{j,\tau}^{(h)}(c^*)) / \hat{\beta}_{j,\text{start}}^{(h)}\|_\infty < \delta$ , where  $\delta = e^{-6}$  is a pre-set tolerance value. then
    Stop. At convergence this results in the final estimate of  $c$ .
  end if
end while

```

---

**Algorithm 1.** An adaptive algorithm for selecting  $c$ .

## 4. Numerical Studies

### 4.1. Simulation

In this example, we simulate 100 datasets from the  $p$ -dimensional linear regression model

$$Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + \varepsilon_i, i = 1, \dots, n, \tag{14}$$

where  $(n, p) = (100, 200)$  and  $(100, 400)$ , and where  $\varepsilon_i$  is an independent and identically (i.i.d.) distributed random variable. The nonzero parameters are given by  $\beta_1 = 3$ ,  $\beta_2 = 1.5$ ,  $\beta_5 = 2$ , and the rest of the  $\beta_j$ 's are zero.

We generate the covariates  $\mathbf{X}_i$  from the  $\mathcal{N}(\mathbf{0}_p, \boldsymbol{\Sigma})$  multivariate normal distribution, where  $\boldsymbol{\Sigma} = (\sigma_{u,v})_{p \times p}$  with  $\sigma_{u,v} = r^{|u-v|}$  and  $r = 0.5$ . For the distribution of the error term, we consider the following three scenarios: 1)

$\varepsilon_i \sim \overset{\text{i.i.d.}}{\mathcal{N}}(0, 1); 2) \varepsilon_i \sim \overset{\text{i.i.d.}}{t}_3$ , i.e., the Student  $t$  distribution with 3 degrees of freedom; and 3) a contaminated  $\mathcal{N}(0, 1)$  distribution having 10% outliers from a Cauchy distribution, denoted by  $C_{10}\text{-}\mathcal{N}$ . We show results for quantile levels  $\tau = 0.25, 0.50$ , and  $0.75$ . We focus on the one-step quantile regression method SCAD penalty (denoted by QS), and also consider the one-step quantile regression method with LASSO penalty (denoted by QL) as a ‘‘benchmark’’<sup>5</sup>. As a second benchmark for comparison, we compute the Oracle estimator, which is the least squares estimator of the true model.

To assess the performance of both quantile regression methods, we adopt the following criteria: 1) NN: average number of nonzero covariates; 2) True Positives (TP): average number of covariates that are correctly included in the model; 3) False Positives (FP): average number of covariates that are incorrectly included in the model; 4) AE: average of the absolute value of the estimation error which is defined as  $\sum_{j=1}^{p_n} |\hat{\beta}_{j,\tau}(c^*) - \beta_{j,0}(\tau)|$ .

**Table 1** summarizes the simulation results. The variable selection performance of QS is quite descent for all values of  $\tau$  and  $p_n$ . Indeed, except for one case (QL with  $\tau = 0.75$ ,  $p_n = 400$ , and  $\varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ ), all computed TP values are equal to the true active model size of 3, with standard errors equal to zero. Hence, to keep the paper clear and readable, all computed TP values have been omitted from **Table 1**. The QL method tends to have substantially higher AE values than QS at different quantile levels for almost all  $p_n$  values and error distributions. In the case  $\varepsilon_i \overset{\text{i.i.d.}}{\sim} t_3$ , the QS method is near the oracle method in terms of AE, but the performance of QL is relatively poor. Also, the advantage of the QS method over the QL method can be seen by its lower AE values when the distribution of  $\varepsilon_i$  is heavy-tailed. In general, the proposed QS method tends to be efficient for the three error distributions under study.

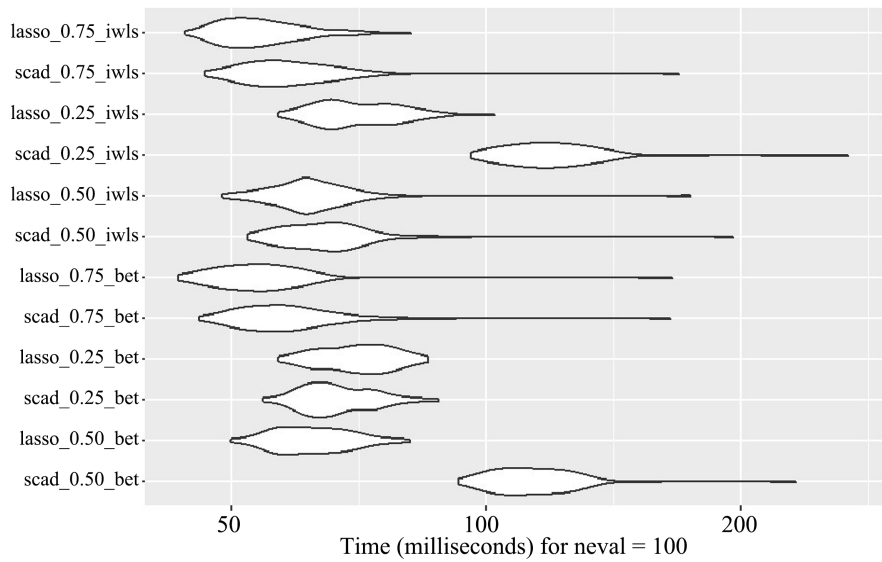
<sup>5</sup>The OL method is also a one-step adaptive procedure using the same IWLS initial estimates as the proposed QS method to ensure the comparison is on equal footing.

**Table 1.** Variable selection performance of one-step quantile regression with LASSO (QL) and SCAD (QS) type penalties for  $n = 100$ . Standard errors are in parentheses. NN = Average number of nonzero covariates; FP = False positives; AE = Average of the absolute values of the estimation error.

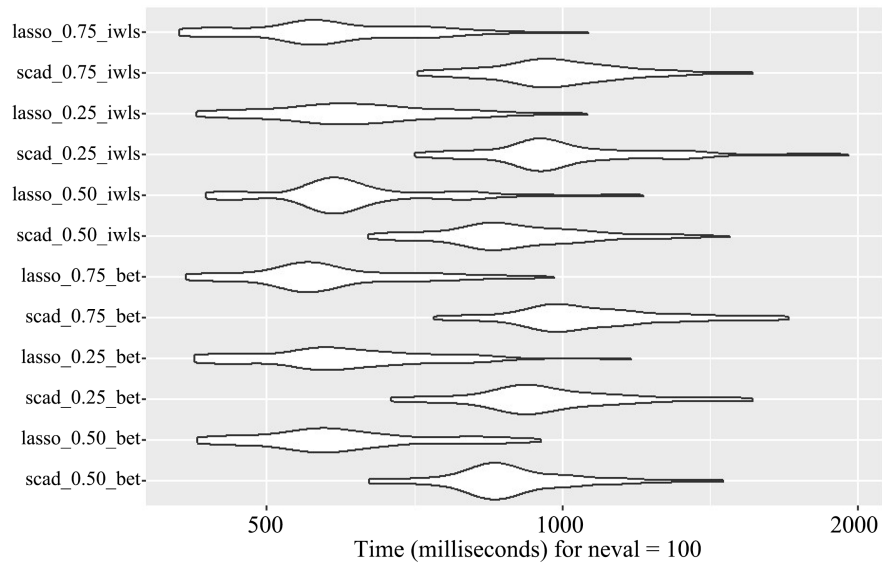
Method	$p_n$	$\varepsilon_i$	$\tau = 0.25$			$\tau = 0.5$			$\tau = 0.75$		
			NN	FP	AE	NN	FP	AE	NN	FP	AE
QL	200	$\mathcal{N}(0,1)$	4.05	1.05	1.16	3.75	0.75	1.16	4.25	1.25	1.15
			(1.07)	(1.07)	(0.41)	(0.87)	(0.87)	(0.34)	(1.17)	(1.17)	(0.40)
			3.27	0.27	0.48	3.29	0.29	0.44	3.25	0.58	0.47
QS	200		(0.57)	(0.57)	(0.26)	(0.57)	(0.57)	(0.24)	(0.52)	(0.52)	(0.25)
			3.00	0.37	0.00	3.00	0.31	0.00	3.00	0.34	0.00
			(0.00)	(0.17)	(0.00)	(0.00)	(0.14)	(0.00)	(0.00)	(0.17)	(0.00)
Oracle	200		4.11	1.10	1.30	3.69	0.68	1.15	3.89	0.90	1.36
			(1.14)	(1.13)	(0.41)	(0.83)	(0.83)	(0.34)	(1.09)	(1.09)	(0.41)
			3.26	0.26	0.50	6.14	3.10	0.80	3.27	0.27	0.49
QL	400	(0.54)	(0.54)	(0.25)	(16.60)	(16.37)	(1.98)	(0.53)	(0.53)	(0.25)	
		3.00	0.35	0.00	3.00	0.31	0.00	3.00	0.36	0.00	
		(0.00)	(0.18)	(0.00)	(0.00)	(0.16)	(0.00)	(0.00)	(0.20)	(0.00)	
Oracle	400	3.87	0.87	1.37	3.55	0.55	1.20	3.92	0.92	1.32	
		(0.98)	(0.98)	(0.55)	(0.69)	(0.69)	(0.38)	(0.93)	(0.93)	(0.42)	
		3.17	0.17	0.52	3.17	0.17	0.48	3.16	0.16	0.53	
QL	200	(0.40)	(0.40)	(0.52)	(0.38)	(0.38)	(0.31)	(0.37)	(0.37)	(0.31)	
		3.00	0.40	0.00	3.00	0.35	0.00	3.00	0.36	0.00	
		(0.00)	(0.22)	(0.00)	(0.00)	(0.20)	(0.00)	(0.00)	(0.23)	(0.00)	
Oracle	200	4.19	1.17	1.63	3.60	0.59	1.39	4.11	1.11	1.57	
		(1.31)	(1.28)	(0.58)	(0.83)	(0.83)	(0.43)	(1.18)	(1.17)	(0.56)	
		3.20	0.20	0.51	6.13	3.11	0.87	3.20	0.20	0.50	
QL	400	(0.45)	(0.45)	(0.34)	(16.60)	(16.48)	(2.50)	(0.47)	(0.47)	(0.29)	
		3.00	0.41	0.00	3.00	0.35	0.00	3.00	0.43	0.00	
		(0.00)	(0.18)	(0.00)	(0.00)	(0.17)	(0.00)	(0.00)	(0.21)	(0.00)	
Oracle	400	3.96	0.96	1.21	3.55	0.55	1.10	3.94	0.94	1.21	
		(0.86)	(0.86)	(0.37)	(0.73)	(0.73)	(0.34)	(1.08)	(0.94)	(0.39)	
		3.44	0.60	0.97	3.53	0.68	0.95	3.48	0.64	1.03	
QL	200	(0.80)	(0.99)	(1.71)	(1.26)	(1.49)	(1.84)	(1.14)	(1.36)	(1.92)	
		3.00	0.32	0.00	3.00	0.34	0.00	3.00	0.37	0.00	
		(0.00)	(0.18)	(0.00)	(0.00)	(0.17)	(0.00)	(0.00)	(0.16)	(0.00)	
Oracle	200	4.07	1.07	1.43	3.56	0.56	1.21	3.88	0.88	1.36	
		(1.13)	(1.13)	(0.45)	(0.81)	(0.81)	(0.36)	(1.06)	(1.06)	(0.45)	
		5.14	2.29	3.33	4.30	1.42	1.00	3.39	0.53	0.91	
QL	400	(13.63)	(13.77)	(17.24)	(9.71)	(9.53)	(2.04)	(0.99)	(1.30)	(0.91)	
		3.00	0.36	0.00	3.00	0.31	0.00	3.00	0.35	0.00	
		(0.00)	(0.21)	(0.00)	(0.00)	(0.18)	(0.00)	(0.00)	(0.24)	(0.00)	
Oracle	400										

### 4.2. Comparing Execution Time

We next compare the execution time of the QL and QS methods with initial quantile parameter estimates obtained by two methods: 1) the IWLS method of Section 3.1, and by 2) the well-known quantile regression method `rq` in the R-quant package and denoted here by the short-hand notation “bet”. The simulation setting is the same as in Section 4.1 with Scenario 1. However, in this case  $n = 80$ ,  $n = 1000$ , and  $p_n = 100$ . Using the R-package microbenchmark, **Figure 1(a)** and **Figure 1(b)** show plots of the microbenchmark timings for three different values of  $\tau$ , with “neval” denoting the number of replicates.



(a)



(b)

**Figure 1.** Autoplots of microbenchmark timings: plot (a) is for  $(n = 80, p_n = 100)$ , and plot (b) is for  $(n = 1000, p_n = 100)$ .

**Figure 1(a)** shows that there is hardly any difference between quantile estimates obtained with the QS-IWLS method and the QL-bet method for most values of  $\tau$ . However, we see from **Figure 1(b)** that the difference between both methods is more noteworthy for the case  $(n, p_n) = (1000, 100)$ . Indeed, QL is about 2 times faster than QS, for all values of  $\tau$ . This difference in computing time was also noted for simulation results obtained with the Student  $t_3$  error distribution and with the contaminated  $\mathcal{N}(0,1)$  error distribution, as specified earlier in Section 4.1. Thus, in terms of computing time, there is a price to be paid for using the proposed one-step QS-IWLS estimation method when  $n$  is large.

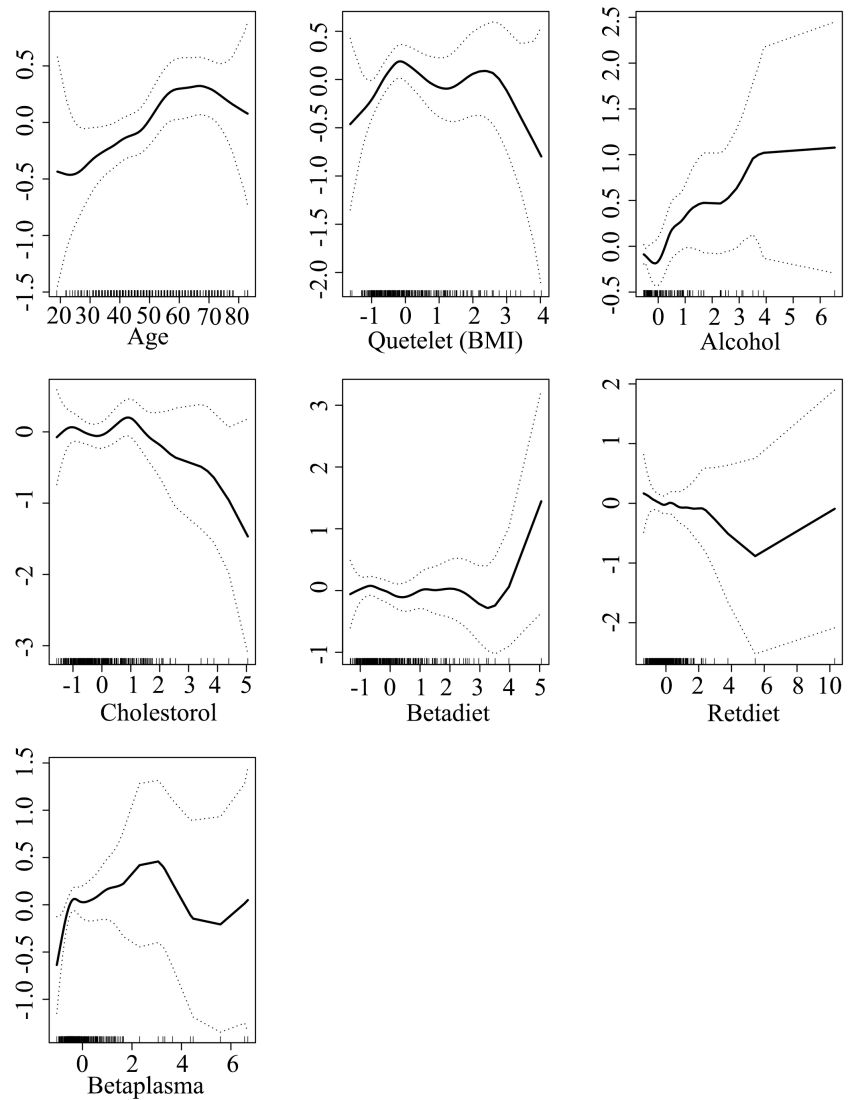
## 5. An Illustrative Application

In this section, we illustrate the one-step SCAD penalized QR estimator using a real dataset corresponding to beta-carotene levels (ng/mL) in 315 patients. The dataset is available online at [http://lib.stat.cmu.edu/datasets/Plasma\\_Retinol](http://lib.stat.cmu.edu/datasets/Plasma_Retinol). Also it is included in the R-Lock5Data package under the name NutritionStudy. The dataset contains observations on the following 14 variables: Age (years), Cholesterol (mg per day), Fiber (Grams consumed per day), Sex (1 = Female, 0 = Male), Smoking Status (1 = Never, 2 = Former, 3 = Current Smoker), Quetelet (Weight/Height<sup>2</sup>), Vitamin Use (1 = Yes, fairly often, 2 = Yes, not often, 3 = No), Calories (Number consumed per day), Fat (Grams consumed per day), Alcohol (Number of drinks consumed per week), Betadiet (Dietary beta-carotene consumed (mcg) per day), Retdiet (Dietary retinol consumed (mcg per day)), Betaplasma (Plasma beta-carotene (ng/mL)), and Retplasma (Plasma retinol (ng/mL)). Study subjects were patients who had an elective surgical procedure during a three-year period to biopsy or remove a lesion of the lung, colon, breast, skin, ovary or uterus that was found to be non-cancerous. The medical interest of the work was to investigate the relation between personal characteristics and diet factors on the one hand, and plasma concentrations of retinol, beta-carotene, or other carotenoids may be as associated with greater risk of developing certain types of cancer.

A boxplot of the response variable (Retplasma) indicated several outliers. In addition, the Shapiro-Wilk test showed that this variable is far from normally distributed with a  $p$ -value of 0.00. There was one extremely high (203) leverage point in the series "Alcohol" that was deleted prior to the analysis. Thus the actual set of observations consists of  $n = 314$  values. All series are standardized, except for the integer series Age, Sex, Smoking Status, and Vitamin Use. **Figure 2** shows the estimated additive fits of seven covariates. It appears that the relationship between the response variable and each of these covariates is nonlinear. The remaining six covariates seem to be linearly related to Retplasma. So, it is natural to analyze the complete dataset via penalized QR.

### 5.1. Low Dimension

To assess the performance of the one-step variable selection method, we randomly selected 200 observations from the complete dataset. We used these observations



**Figure 2.** Plots of the estimated additive functions for seven predictors, with associated standard errors (dotted lines), using the R-gam function.

for model fitting and variable selection. The remaining 114 observations are used for checking the prediction performance of the fitted model. This procedure is repeated 100 times through the bootstrapping technique. Given this setup, and to guard against false positives, we declare a covariate to be relevant when it is selected 60% of the time.

Panel I of **Table 2** contains the variable selection results for QL and QS, respectively. For  $\tau = 0.25$ , QL identifies different covariates than QS. However, for  $\tau = 0.5$  both methods select the variables Age and Alcohol as relevant covariates. On the other hand, when  $\tau = 0.75$  we see that the variables Age and Smoke1 are selected frequently. In terms of median absolute prediction error (MAPE) computed across all covariates, we have  $MAPE(QS) = 0.27$  and  $MAPE(QL) = 0.37$  for  $\tau = 0.25$ . When  $\tau = 0.5$ ,  $MAPE(QL) = MAPE(QS) = 0.36$ , and when  $\tau = 0.75$  the MAPEs are equal to 0.36 for both methods.

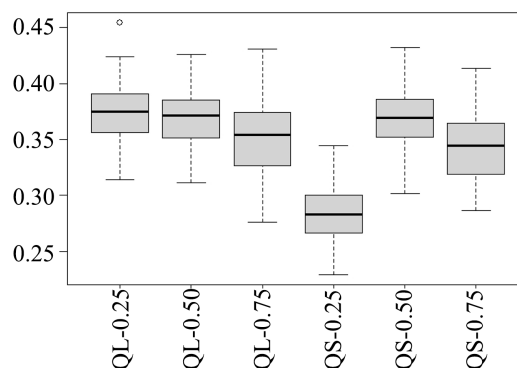
**Table 2.** Frequency ( $\geq 60\%$ ) of covariates selected by QL and QS at three different quantiles based on 100 bootstrap replicates. Selected frequencies are in parentheses.

Method	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
<b>I. Low dimension (Original dataset), <math>(n, p_n) = (315, 15)</math></b>			
QL	Age (61), Alcohol (95) Betaplasma (97)	Age (99), Alcohol (94)	Age (99), Alcohol (76) Sex (85), Smoke1 (71)
QS	Sex (100)	Age (90), Alcohol (62) Sex (96)	Age (78), Smoke1 (70)
<b>II. High dimension (Augmented dataset), <math>(n, p_n) = (315, 215)</math></b>			
QL	Alcohol (84), Betaplasma (70)	Age (98), Alcohol (89)	
QS		Age (85)	

### 5.2. High Dimension

To investigate the QL and QS methods in high dimensions, we add 200 irrelevant covariates to the dataset. These new variables are randomly drawn from a  $U(0, 1)$  distribution. Hence, the augmented dataset consists of 215 covariates (200 irrelevant and 15 relevant ones). Using the same procedure as above, Panel II of **Table 2** summarizes the variable selection results. For  $\tau = 0.25$  and  $\tau = 0.75$ , no covariates are selected by QS while QL selects covariates at these quantiles. Clearly, QS tends to produce more sparse models than QL. Further, in terms of MAPE, the QS method performs better than QL at  $\tau = 0.25$ . Specifically,  $MAPE(QS) = 0.28$  and  $MAPE(QL) = 0.37$ . For  $\tau = 0.5$   $MAPE(QL) = MAPE(QS) = 0.36$ , while for  $\tau = 0.75$   $MAPE(QL) = 0.35$  and  $MAPE(QS) = 0.34$ .

**Figure 3** presents a boxplot of the prediction errors of QL and QS at different quantiles using the augmented plasma dataset, and based on 100 bootstrapped replications. The QS method at  $\tau = 0.25$  provides a good prediction, followed by the same method at  $\tau = 0.5$  and  $\tau = 0.75$ . Overall, the one-step QS method is found to be a good variable selection procedure in detecting the significant variables, even when some irrelevant variables are included in the dataset under study. The prediction errors of the proposed method are also smaller than those using the one-step QL method.



**Figure 3.** Boxplot of the prediction errors for the augmented plasma dataset.

## 6. Summary

In this paper, we propose a one-step estimation procedure for variable selection in sparse, high-dimensional, linear QR models. In particular, we used the SCAD penalty with an iteratively weighted least squares estimator to determine the penalty weights. In a simulation study, we have shown that the resulting one-step QS method performs better than a one-step QL method for three different scenarios representing the error distribution. Additionally, we illustrated the merit of our method via a real data example.

A limitation of the proposed procedure is that there is a computational speed trade-off, as briefly discussed in Section 4.2. However, this result is outweighed by an increase in estimation accuracy when the sample size and/or the number of covariates increases. Another issue worth mentioning is the practical implication of the IWLS method. For instance, in specific fields like genomics or financial modeling, analyzing a large number of observations, our method's robustness to heavy-tailed distributions and outliers would be particularly advantageous.

Several possible extensions to this paper can be pursued. For instance, one may wish to estimate nonlinear regressions via the QS method. Perhaps, in that case, efficiency may be further enhanced by a more prudent choice of the initial estimator than using the IWLS method. It may also be of interest to assess the robustness of the IWLS method by changing the level of temporal correlation among the covariates. Selecting the optimal SCAD tuning parameter  $\lambda$  by a different data-driven method (e.g., AIC) and then investigating its impact on the QS estimates is another fruitful future research direction.

## Conflicts of Interest

The author reports no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

## References

- Abdelmalek, N. N. (1974). On the Discrete Linear L1 Approximation and L1 Solutions of Overdetermined Linear Equations. *Journal of Approximation Theory*, 11, 38-53. [https://doi.org/10.1016/0021-9045\(74\)90037-9](https://doi.org/10.1016/0021-9045(74)90037-9)
- Amin, M., Song, L., Thorlie, M. A., & Wang, X. (2015). SCAD-Penalized Quantile Regression for High-Dimensional Data Analysis and Variable Selection. *Statistica Neerlandica*, 69, 212-235. <https://doi.org/10.1111/stan.12056>
- Chen, L., & Lee, S. (2023). Sparse Quantile Regression. *Journal of Econometrics*, 235, 2195-2217. <https://doi.org/10.1016/j.jeconom.2023.02.014>
- Deaton, A., & Muellbauer, J. (1980). *Economics and Consumer Behavior*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511805653>
- Fan, J., & Li, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, 96, 1348-1360. <https://doi.org/10.1198/016214501753382273>
- Fan, J., & Lv, J. (2008). Sure Independence Screening for Ultrahigh Dimensional Feature Space. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 70, 849-911. <https://doi.org/10.1111/j.1467-9868.2008.00674.x>

- Khan, D. M., Yaqoob, A., Iqbal, N., Wahid, A., Khalil, U., Khan, M. et al. (2019). Variable Selection via SCAD-Penalized Quantile Regression for High-Dimensional Count Data. *IEEE Access*, 7, 153205-153216. <https://doi.org/10.1109/access.2019.2948278>
- Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica*, 46, 33-50. <https://doi.org/10.2307/1913643>
- Lee, E. R., Noh, H., & Park, B. U. (2014). Model Selection via Bayesian Information Criterion for Quantile Regression Models. *Journal of the American Statistical Association*, 109, 216-229. <https://doi.org/10.1080/01621459.2013.836975>
- Mai, T. T. (2024). A Sparse Pac-Bayesian Approach for High-Dimensional Quantile Prediction. *Statistics and Computing*, 35, Article No. 93. <https://doi.org/10.1007/s11222-025-10617-7>
- Muggeo, V. M. R., Sciandra, M., & Augugliaro, L. (2011). Quantile Regression via Iterative Least Squares Computations. *Journal of Statistical Computation and Simulation*, 82, 1557-1569. <https://doi.org/10.1080/00949655.2011.583650>
- Noh, H., & Lee, E. R. (2014). Component Selection in Additive Quantile Regression Models. *Journal of the Korean Statistical Society*, 43, 439-452. <https://doi.org/10.1016/j.jkss.2014.01.002>
- Sherwood, B., & Maidman, A. (2022). Additive Nonlinear Quantile Regression in Ultra-High Dimension. *Journal of Machine Learning Research*, 23, 1-47.
- Sherwood, B., & Wang, L. (2016). Partially Linear Additive Quantile Regression in Ultra-High Dimension. *The Annals of Statistics*, 44, 288-317. <https://doi.org/10.1214/15-aos1367>
- Sherwood, B., Li, S., & Maidman, A. (2025). *rqPen: Penalized Quantile Regression*. R Package Version 4.1.3. <https://github.com/bssherwood/rqpen>
- Song, Y., Li, Z., & Fang, M. (2022). Robust Variable Selection Based on Penalized Composite Quantile Regression for High-Dimensional Single-Index Models. *Mathematics*, 10, Article 2000. <https://doi.org/10.3390/math10122000>
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58, 267-288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Wang, L., Wu, Y., & Li, R. (2012). Quantile Regression for Analyzing Heterogeneity in Ultra-High Dimension. *Journal of the American Statistical Association*, 107, 214-222. <https://doi.org/10.1080/01621459.2012.656014>
- Wu, Y., & Liu, Y. (2009). Variable Selection in Quantile Regression. *Statistica Sinica*, 19, 801-817. <https://www.jstor.org/stable/24308857>
- Zhao, P., & Yu, B. (2006). On Model Selection Consistency of Lasso. *The Journal of Machine Learning Research*, 7, 2541-2563.
- Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, 101, 1418-1429. <https://doi.org/10.1198/016214506000000735>
- Zou, H., & Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67, 301-320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>
- Zou, H., & Li, R. (2008). One-Step Sparse Estimates in Nonconcave Penalized Likelihood Models. *The Annals of Statistics*, 36, 1509-1533. <https://doi.org/10.1214/009053607000000802>