




Variable Screening via Quantile Partial Correlation

Shujie Ma, Runze Li & Chih-Ling Tsai


To cite this article: Shujie Ma, Runze Li & Chih-Ling Tsai (2017) Variable Screening via Quantile Partial Correlation, Journal of the American Statistical Association, 112:518, 650-663, DOI: 10.1080/01621459.2016.1156545

To link to this article: <http://dx.doi.org/10.1080/01621459.2016.1156545>

 View supplementary material [↗](#)

 Accepted author version posted online: 22 Mar 2016.
Published online: 30 Mar 2017.

 Submit your article to this journal [↗](#)

 Article views: 643

 View Crossmark data [↗](#)



Variable Screening via Quantile Partial Correlation

Shujie Ma^a, Runze Li^b, and Chih-Ling Tsai^c

^aDepartment of Statistics, University of California-Riverside, Riverside, CA; ^bDepartment of Statistics, The Pennsylvania State University, University Park, PA; ^cGraduate School of Management, University of California-Davis, Davis, CA

ABSTRACT

In quantile linear regression with ultrahigh-dimensional data, we propose an algorithm for screening all candidate variables and subsequently selecting relevant predictors. Specifically, we first employ quantile partial correlation for screening, and then we apply the extended Bayesian information criterion (EBIC) for best subset selection. Our proposed method can successfully select predictors when the variables are highly correlated, and it can also identify variables that make a contribution to the conditional quantiles but are marginally uncorrelated or weakly correlated with the response. Theoretical results show that the proposed algorithm can yield the sure screening set. By controlling the false selection rate, model selection consistency can be achieved theoretically. In practice, we proposed using EBIC for best subset selection so that the resulting model is screening consistent. Simulation studies demonstrate that the proposed algorithm performs well, and an empirical example is presented. Supplementary materials for this article are available online.

ARTICLE HISTORY

Received April 2015
Revised February 2016

KEYWORDS

Quantile correlation;
Quantile partial correlation;
Screening; Variable selection

1. Introduction

Advances in modern technology have enabled people to collect massive data with a large number of variables, many of which may be irrelevant to the response variable. Examples can be found in gene expression microarray data, single nucleotide polymorphism (SNP) data, imaging data, high-frequency financial data, and others. Hence, extracting useful variables for the prediction of the response in the high-dimensional data has become a focal research area in the past two decades. Apparently, the traditional variable selection methods such as best subset selection and backward elimination become computationally infeasible when the number of predictors is large. As a result, a variety of penalization methods have been developed. These methods include, but are not limited to, Lasso and Adaptive Lasso (Tibshirani 1996; Zou 2006; Huang, Ma, and Zhang 2008a), bridge regression (Huang, Horowitz, and Ma 2008b), SCAD (Fan and Li 2001), elastic net (Zou and Hastie 2005), and MCP (Zhang 2010).

When the dimension is much larger than the sample size, penalized estimation can perform poorly or even become infeasible (Fan and Lv 2010). Then the variable screening method becomes a natural way to consider in this context, which assumes that the relevant features lie in a low-dimensional manifold. Thus, the ultrahigh-dimensional problem can be greatly simplified into a low-dimensional one. Recently, Fan and Lv (2008) introduced the marginal screening method (the sure independent screening, SIS) to select relevant variables based on the marginal correlation of each variable and the response. Its good numerical performance and novel theoretical properties have made SIS popular in ultrahigh-dimensional analysis. As a result, SIS and its extensions have been applied to many

important settings including generalized linear model (Fan and Song 2010), multi-index semiparametric models (Zhu et al. 2011), nonparametric regression (Fan, Feng, and Song 2011; Liu, Li, and Wu 2014), quantile regression (He, Wang, and Hong 2013; Wu and Yin 2015), and so forth.

The marginal screening method employs the marginal correlation to measure the strength of association between predictors and the response. Hence, it can miss some relevant variables that are associated with the response conditionally but not marginally. Furthermore, the marginal correlation can be misleading when there exist nonnegligible correlations among the predictors. As a result, an irrelevant variable can be selected prior to relevant variables. Moreover the issue of collinearity can yield spurious phenomena in the high-dimensional data as demonstrated by Fan and Lv (2008).

To handle the problem of high correlations between predictors, several methods have been developed in the literature. Bühlmann, Kalisch, and Maathuis (2009) proposed the PC-Simple algorithm, which uses partial correlation as a criterion to measure the association of each predictor with the response. Wang (2009) applied the forward selection method in the ultrahigh-dimensional setting and developed a forward regression (FR) algorithm to select the most relevant variable in each step sequentially by removing the confounding effects of the selected variables from the previous steps. It can be shown that Wang's algorithm is also based on the partial correlation measure. Moreover, Cho and Fryzlewicz (2012) proposed a "tilted" correlation to measure the contribution of each predictor to the response. Based on simulation studies, they found that their proposed tilted correlation screening (TCS2) algorithm performs well. The above studies demonstrate that the partial correlation plays an important role in the screening process.

In the area of quantile linear regression with low-dimensional data, theoretical properties and practical applications have been well developed; see Koenker (2005). For high-dimensional data, however, it is far from complete. Recently, Wang, Wu, and Li (2012) and Lee, Noh, and Park (2014), respectively, extended the penalized approach and Bayesian selection method from the classical mean regression model to quantile regression model. Their generalizations motivate us to propose a screening method for high-dimensional quantile regression model. Specifically, we adopt Li et al.'s (2015) quantile partial correlation (QPCOR) as a criterion to measure the association of each predictor with the response at each quantile, and then introduce a new screening procedure by using the sample QPCOR. Our goal is to identify a sparse set of ultrahigh-dimensional variables $\mathbf{X} = (X_1, \dots, X_p)^T$ that are relevant for modeling the conditional quantile of the response Y .

To employ QPCOR, we transform each predictor X_j by projecting it onto a set of variables, denoted by S_j , which is either the union of its related variables and the previously selected variables or only the previously selected variables. We then introduce an adaptive approach to choose the subset of related variables by adopting a sequential testing method based on the partial correlations of related variables. It is worth noting that the size of S_j cannot be too large since it would distort the association between X_j and Y . To this end, we suggest a hard threshold for determining which variables are related to X_j , and then obtain the upper bound of the maximal cardinality of the subsets S_j . In addition, we derive a uniform bound of the difference between the sample QPCOR and the population QPCOR, and subsequently establish the sure screening property of the proposed procedure as needed in screening methods (Fan and Lv 2008; Fan, Feng, and Song 2011; He, Wang, and Hong 2013). Moreover, we generalize Wang's (2009) FR algorithm and Cho and Fryzlewicz's (2012) TCS2 algorithm to the quantile regression model. After the screening procedure, we apply the extended Bayesian information criterion (EBIC) (Chen and Chen 2008; Wang, Li, and Leng 2009; Lee, Noh, and Park 2014) for best subset selection. Consequently, our proposed approach not only selects relevant variables when the variables are highly correlated, but also identifies the variables that are marginally uncorrelated or weakly correlated with the response.

The article is organized as follows. Section 2 introduces quantile partial correlation. Section 3 provides the theoretical properties of the quantile screening procedure including the sure screening property. Section 4 presents three algorithms, which consist of our proposed algorithm and the quantile version of the forward regression and tilted correlation screening algorithms. We also introduce the extended BIC criterion for best subset selection. Section 5 conducts simulation studies, while Section 6 illustrates the usefulness of the proposed method through the analysis of gene expression data. A discussion is given in Section 7. All technical proofs are relegated to the Appendix and online supplementary materials, and additional simulation results are presented in the online supplementary materials.

2. Quantile Partial Correlation (QPCOR)

Before we present the QPCOR, we review the quantile correlation (QCOR) and its connection to regression coefficients in the linear quantile regression model.

Quantile correlation. For mean regression models with ultrahigh-dimensional covariates, Fan and Lv (2008) proposed the SIS procedure to select variables according to the magnitudes of their marginal Pearson correlations associated with the response. Analogously, in the quantile regression context, we introduce Li et al.'s (2015) quantile correlation of Y and X_j :

$$\begin{aligned} \text{qcor}_\tau\{Y, X_j\} &= \frac{\text{qcov}_\tau\{Y, X_j\}}{\sqrt{\text{var}\{\psi_\tau(Y - Q_\tau(Y))\}\text{var}(X_j)}} \\ &= \frac{E\{\psi_\tau(Y - Q_\tau(Y))(X_j - EX_j)\}}{\sqrt{(\tau - \tau^2)\sigma_{X_j}^2}}, \end{aligned} \quad (1)$$

for $1 \leq j \leq p$, where $Q_\tau(Y)$ is the τ th unconditional quantile of Y such that $P(Y < Q_\tau(Y)) = \tau$ and $\psi_\tau(w) = \tau - I(w < 0)$. As a result, $-1 \leq \text{qcor}_\tau\{Y, X_j\} \leq 1$. As shown by Li, Li, and Tsai (2015), there is a nice relationship between the quantile correlation (1) and the slope of the τ th quantile linear regression line with Y and X_j being the response and predictor, respectively. Consider the following minimizers:

$$(b_{0\tau}^0, b_{j\tau}^0) = \arg \min_{b_{0\tau}, b_{j\tau}} E[\rho_\tau(Y - b_{0\tau} - b_{j\tau}X_j)], \quad (2)$$

where $\rho_\tau(w) = w\tau - wI(w < 0)$ is the quantile loss function (see Koenker 2005). Then $\text{qcor}_\tau\{Y, X_j\} = \varrho(b_{j\tau}^0)$, where $\varrho(b_{j\tau}^0)$ is a continuous and increasing function, and $\varrho(b_{j\tau}^0) = 0$ if and only if $b_{j\tau}^0 = 0$. Accordingly, we can adopt the SIS procedure of Fan and Lv (2008) to rank the significance of predictors on the quantile of Y via the marginal quantile correlation $\text{qcor}_\tau\{Y, X_j\}$. However, this marginal approach ignores possible effects from other variables and may yield misleading results when the predictors are correlated. To illustrate this phenomenon, we first introduce the quantile multiple regression model and its associated estimators given below.

Let Y and $\mathbf{X} = (X_1, \dots, X_p)^T$ be the response and predictors, respectively. Consider a linear quantile model

$$Y = \beta_{0\tau}^0 + \beta_{1\tau}^0 X_1 + \dots + \beta_{p\tau}^0 X_p + \varepsilon, \quad (3)$$

where the error term satisfies $P(\varepsilon < 0|\mathbf{X}) = \tau$. Then, the τ th conditional quantile of Y given \mathbf{X} is $Q_\tau(Y|\mathbf{X}) = \beta_{0\tau}^0 + \beta_{1\tau}^0 X_1 + \dots + \beta_{p\tau}^0 X_p$. Without loss of generality, we assume that $E(X_j) = 0$ and $\text{var}(X_j) = 1$ for all $j = 1, \dots, p$. Furthermore, denote $f_\varepsilon(u|\mathbf{x})$ and $f_Y(y|\mathbf{x})$ as the conditional density of ε and Y given $\mathbf{X} = \mathbf{x}$, respectively.

Assuming that the conditional density $f_Y(y|\mathbf{x})$ exists, we can follow the same procedure as given in theorem 2 of Angrist, Chernozhukov, and Fernández-Val (2006) and obtain the coefficient $b_{j\tau}^0$ that is the minimizer of the weighted least squares:

$$b_{j\tau}^0 = \arg \min_{b_{j\tau} \in \mathbb{R}} E[\tilde{w}_\tau(\mathbf{X}) \cdot (Q_\tau(Y|\mathbf{X}) - b_{j\tau}X_j)^2],$$

where $\tilde{w}_\tau(\mathbf{X}) = \frac{1}{2} \int_0^1 f_\varepsilon(u \cdot \Delta_\tau(X_j, b_{j\tau}^0)|\mathbf{X}) du$, and $\Delta_\tau(X_j, b_{j\tau}^0) = b_{0\tau}^0 + b_{j\tau}^0 X_j - Q_\tau(Y|\mathbf{X})$ for $j = 1, \dots, p$. As a result,

$$b_{j\tau}^0 = \{E(\tilde{w}_\tau(\mathbf{X})X_j^2)\}^{-1} E(\tilde{w}_\tau(\mathbf{X})X_jQ_\tau(Y|\mathbf{X})) = \beta_{j\tau}^0 + d_{j\tau},$$

where

$$d_{j\tau} = \sum_{k \neq j} \beta_{k\tau}^0 \{E(\tilde{w}_\tau(\mathbf{X})X_j^2)\}^{-1} E(\tilde{w}_\tau(\mathbf{X})X_jX_k).$$

The term $d_{j\tau}$ can be viewed as the "bias" of the quantile estimator. It can be considerably large when the components

$E(\tilde{w}_\tau(\mathbf{X})X_jX_k)$ are nonnegligible. Thus, QCOR may lead to inaccurate screening results. This motivates us to propose a screening procedure, based on the quantile partial correlation, to reduce the confounding effects from other predictors that are highly related to X_j .

Quantile partial correlation. To reduce the confounding effects, consider $\mathbf{X}_{-j} = (1, \{X_k, k \neq j\}^T)^T$. Then, let $\boldsymbol{\theta}_j^0 = \arg \min_{\boldsymbol{\theta}_j \in \mathbb{R}^p} E[(X_j - \mathbf{X}_{-j}^T \boldsymbol{\theta}_j)^2]$ and $\boldsymbol{\alpha}_j^0 = \arg \min_{\boldsymbol{\alpha}_j \in \mathbb{R}^p} E[\rho_\tau(Y - \mathbf{X}_{-j}^T \boldsymbol{\alpha}_j)]$ so that $E[\psi_\tau(Y - \mathbf{X}_{-j}^T \boldsymbol{\alpha}_j^0) \mathbf{X}_{-j}] = \mathbf{0}$. Adopting Li et al.'s (2015) approach, we define the quantile partial correlation (QPCOR) as follows:

$$\begin{aligned} \text{qpcor}_\tau\{Y, X_j | \mathbf{X}_{-j}\} &= \frac{\text{cov}\{\psi_\tau(Y - \mathbf{X}_{-j}^T \boldsymbol{\alpha}_j^0), X_j - \mathbf{X}_{-j}^T \boldsymbol{\theta}_j^0\}}{\sqrt{\text{var}\{\psi_\tau(Y - \mathbf{X}_{-j}^T \boldsymbol{\alpha}_j^0)\} \text{var}(X_j - \mathbf{X}_{-j}^T \boldsymbol{\theta}_j^0)}} \\ &= \frac{E\{\psi_\tau(Y - \mathbf{X}_{-j}^T \boldsymbol{\alpha}_j^0)(X_j - \mathbf{X}_{-j}^T \boldsymbol{\theta}_j^0)\}}{\sqrt{(\tau - \tau^2)\sigma_j^2}}, \quad (4) \end{aligned}$$

where $\sigma_j^2 = \text{var}(X_j - \mathbf{X}_{-j}^T \boldsymbol{\theta}_j^0)$. Based on the result after Equation (2.2) on page 247 of Li et al. (2015), we have that $\text{qpcor}_\tau\{Y, X_j | \mathbf{X}_{-j}\} = \varrho(\beta_{j\tau}^*)$, where $\varrho(\beta_{j\tau}^*)$ is a continuous and increasing function of $\beta_{j\tau}^*$, and $\varrho(\beta_{j\tau}^*) = 0$ if and only if $\beta_{j\tau}^* = 0$, where

$$(\beta_{0\tau}^*, \beta_{j\tau}^*) = \arg \min_{\beta_{0\tau}, \beta_{j\tau}} E[\rho_\tau(Y - \mathbf{X}_{-j}^T \boldsymbol{\alpha}_j^0 - \beta_{0\tau} - X_j \beta_{j\tau})]. \quad (5)$$

In the general situation, the coefficients $\beta_{j\tau}^*$ and $\beta_{0\tau}^*$ in models (5) and (3) are not equal to each other. However, Lemma A.1 in the Appendix shows that $\beta_{j\tau}^0 = 0$ if and only if $\beta_{j\tau}^* = 0$. Hence, $\beta_{j\tau}^0 = 0$ if and only if $\varrho(\beta_{j\tau}^*) = 0$. Therefore, we use the QPCOR to select relevant variables in our screening procedure.

In general, the estimates of $\boldsymbol{\theta}_j^0$ and $\boldsymbol{\alpha}_j^0$ cannot be obtained when the dimension of \mathbf{X}_{-j} is high. To this end, we remove the confounding effects from X_j that are induced by a subset of $\{k : k \neq j\}$, and we denote the resulting set by \mathcal{S}_j and name it the conditional set. Then, we propose a screening method via the sequential procedure. In each sequential step, let \mathcal{S}_j contain either the union of the previously selected variables and the variables related to X_j or only the previously selected variables, which will be discussed in Section 4. For any arbitrary subset $\mathcal{S} \subset \{1, \dots, p\}$, we denote $\mathbf{X}_{\mathcal{S}}$ the subvector of \mathbf{X} associated with \mathcal{S} . Accordingly, $\mathbf{X}_{\mathcal{S}_j} = (X_0, \{X_k, k \in \mathcal{S}_j\}^T)^T$ with $X_0 = 1$. For the sake of screening, we modify QPCOR given in (4) as

$$\text{qpcor}_\tau\{Y, X_j | \mathbf{X}_{\mathcal{S}_j}\} = \frac{E\{\psi_\tau(Y - \mathbf{X}_{\mathcal{S}_j}^T \boldsymbol{\pi}_j^0)(X_j - \mathbf{X}_{\mathcal{S}_j}^T \boldsymbol{\vartheta}_j^0)\}}{\sqrt{(\tau - \tau^2)\sigma_j^2}}, \quad (6)$$

where $\boldsymbol{\vartheta}_j^0 = \arg \min_{\boldsymbol{\vartheta}_j \in \mathbb{R}^{|\mathcal{S}_j|}} E(X_j - \mathbf{X}_{\mathcal{S}_j}^T \boldsymbol{\vartheta}_j)^2$, $\boldsymbol{\pi}_j^0 = \arg \min_{\boldsymbol{\pi}_j \in \mathbb{R}^{|\mathcal{S}_j|}} E[\rho_\tau(Y - \mathbf{X}_{\mathcal{S}_j}^T \boldsymbol{\pi}_j)]$, $\sigma_j^2 = \text{var}(X_j - \mathbf{X}_{\mathcal{S}_j}^T \boldsymbol{\vartheta}_j^0)$, and $|\mathcal{S}|$ denotes the cardinality of a set \mathcal{S} .

In practice, QCOR and QPCOR are unknown, and we employ the sample estimates of QCOR and QPCOR to study the screening process. These sample estimates are defined as follows. Let $\{(Y_i, \mathbf{X}_i^T)^T, i = 1, \dots, n\}$ be a dataset of n random samples from the distribution of $(Y, \mathbf{X}^T)^T$, where $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})^T$. In this article, we focus on the scenario in which $p \gg n$ and we sometimes denote p by p_n since it can be a function of n . In addition, let $\mathbf{X}_{i,\mathcal{S}}$ be the subvector of \mathbf{X}_i for any subset \mathcal{S} . The sample

estimate of QCOR in (1) is defined as

$$\widehat{\text{qcor}}_\tau\{Y, X_j\} = \frac{n^{-1} \sum_{i=1}^n \{\psi_\tau(Y_i - \widehat{Q}_\tau(Y))(X_{ij} - \bar{X}_j)\}}{\sqrt{(\tau - \tau^2)\widehat{\sigma}_{X_j}^2}}, \quad (7)$$

where $\widehat{Q}_\tau(Y) = \inf\{y : F_n(y) \geq \tau\}$ is the sample τ th quantile of Y_1, \dots, Y_n . Additionally, $F_n(y) = n^{-1} \sum_{i=1}^n I(Y_i \leq y)$ is the empirical distribution function, $\bar{X}_j = n^{-1} \sum_{i=1}^n X_{ij}$, and $\widehat{\sigma}_{X_j}^2 = n^{-1} \sum_{i=1}^n (X_{ij} - \bar{X}_j)^2$. The sample estimate of QPCOR in (6) is given as

$$\begin{aligned} \widehat{\text{qpcor}}_\tau\{Y, X_j | \mathbf{X}_{\mathcal{S}_j}\} &= \frac{n^{-1} \sum_{i=1}^n \{\psi_\tau(Y_i - \mathbf{X}_{i,\mathcal{S}_j}^T \widehat{\boldsymbol{\pi}}_j)(X_{ij} - \mathbf{X}_{i,\mathcal{S}_j}^T \widehat{\boldsymbol{\vartheta}}_j)\}}{\sqrt{(\tau - \tau^2)\widehat{\sigma}_j^2}}, \quad (8) \end{aligned}$$

where $\widehat{\boldsymbol{\vartheta}}_j = \arg \min_{\boldsymbol{\vartheta}_j \in \mathbb{R}^{|\mathcal{S}_j|}} n^{-1} \sum_{i=1}^n (X_{ij} - \mathbf{X}_{i,\mathcal{S}_j}^T \boldsymbol{\vartheta}_j)^2$, $\widehat{\boldsymbol{\pi}}_j = \arg \min_{\boldsymbol{\pi}_j \in \mathbb{R}^{|\mathcal{S}_j|}} n^{-1} \sum_{i=1}^n [\rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{S}_j}^T \boldsymbol{\pi}_j)]$, and $\widehat{\sigma}_j^2 = n^{-1} \sum_{i=1}^n (X_{ij} - \mathbf{X}_{i,\mathcal{S}_j}^T \widehat{\boldsymbol{\vartheta}}_j)^2$. We next study the asymptotic property of the sample estimate of QPCOR and the screening property of the selected variables via this estimate.

3. Theoretical Properties

To use the sample estimate of QPCOR, $\widehat{\text{qpcor}}_\tau\{Y, X_j | \mathbf{X}_{\mathcal{S}_j}\}$, given in (8) as a criterion to identify important variables sequentially, we need to establish the uniform convergence of the sample QPCOR to its population counterpart $\text{qpcor}_\tau\{Y, X_j | \mathbf{X}_{\mathcal{S}_j}\}$. Let $r_n = \max_{1 \leq j \leq p} |\mathcal{S}_j|$ be the maximal cardinality of the subsets \mathcal{S}_j ($j = 1, \dots, p$) given in the screening procedure, and allow r_n to increase with the sample size n . Note that \mathcal{S}_j represents every possible conditional set. In addition, let $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ be the largest and smallest eigenvalues of the symmetric matrix \mathbf{A} , respectively, and let $\|\mathbf{a}\|$ denote the L_2 norm for any vector $\mathbf{a} = (a_1, \dots, a_p)^T$. Then, we make the following assumptions to facilitate the technical proofs, while these assumptions may not be the weakest ones.

- (C1) The conditional density $f_{Y|X=\mathbf{x}}(y)$ of y given $\mathbf{X} = \mathbf{x}$ satisfies the Lipschitz condition of order 1 and $f_{Y|X=\mathbf{x}}(y) > 0$ for any y in a neighborhood of $\mathbf{X}_{\mathcal{S}_j}^T \boldsymbol{\pi}_j^0 = \mathbf{x}_{\mathcal{S}_j}^T \boldsymbol{\pi}_j^0$, for $1 \leq j \leq p$.
- (C2) The predictors satisfy: (i) $\sup_{i,j} |X_{ij}| \leq M_1$, $\sup_{i,j} |\mathbf{X}_{i,\mathcal{S}_j}^T \boldsymbol{\vartheta}_j^0| \leq M_2$, $\sup_{i,j} |\mathbf{X}_{i,\mathcal{S}_j}^T \boldsymbol{\pi}_j^0| \leq M_3$, and $\sup_j \|n^{-1} \sum_{i=1}^n \mathbf{X}_{i,\mathcal{S}_j} X_{ij}\| \leq M_4$ for some positive finite constants M_1, M_2, M_3 , and M_4 ;
(ii) For $1 \leq j \leq p$, there exist positive finite constants m and M such that

$$m \leq \lambda_{\min}(E(\mathbf{X}_{\mathcal{S}_j} \mathbf{X}_{\mathcal{S}_j}^T)) \leq \lambda_{\max}(E(\mathbf{X}_{\mathcal{S}_j} \mathbf{X}_{\mathcal{S}_j}^T)) \leq M.$$

Condition (C1) is a standard condition in the literature on quantile regression. Condition (i) in (C2) assumes that the absolute values of the predictors are bounded, which is commonly assumed in high-dimensional analysis; see Wang, Wu, and Li (2012) and Lee, Noh, and Park (2014). This assumption can be relaxed to the moment condition given in Li, Zhong, and Zhu (2012) and Zhu et al. (2011) that there exists a positive constant t_0 such that $\max_{1 \leq j \leq p} E\{\exp(tX_j)\} < \infty$ for $0 < t \leq t_0$. In this case, our theoretical results still hold with some modification to the proofs. To mitigate notational complexity and

facilitate mathematical derivations, we assume that covariates are bounded. We also assume that, for each subset \mathcal{S}_j used as a conditional set for X_j , the L_2 norm of the correlation vector is bounded. Condition (ii) in (C2) is the sparse Riesz condition (Chen and Chen 2008; Lee, Noh, and Park 2014), which is used for dealing with a large number of regressors. We next demonstrate the uniform convergence of $\widehat{\text{qpcor}}_\tau\{Y, X_j|\mathbf{X}_{\mathcal{S}_j}\}$ to its population counterpart, $\text{qpcor}_\tau\{Y, X_j|\mathbf{X}_{\mathcal{S}_j}\}$.

Theorem 1. Under Conditions (C1) and (C2), for any $C_1 > 0$, there exist some positive constants $C_2, C_3, C_2^*,$ and C_3^* such that, for $0 < \kappa < 1/2$ and $r_n = Cn^\omega$ for some $0 \leq \omega < \min((1 - 2\kappa), 2\kappa)$ and a positive constant C , we have

$$P\left(\sup_{1 \leq j \leq p_n} |\widehat{\text{qpcor}}_\tau\{Y, X_j|\mathbf{X}_{\mathcal{S}_j}\} - \text{qpcor}_\tau\{Y, X_j|\mathbf{X}_{\mathcal{S}_j}\}| \geq C_1 r_n^{1/2} n^{-\kappa}\right) \leq p_n\{C_2^* r_n^2 \exp(-C_2 r_n^{-1} n^{1-2\kappa}) + C_3^* r_n^2 \exp(-C_3 n r_n^{-2})\}, \quad (9)$$

when n is sufficiently large.

Remark 1. To handle ultrahigh-dimensional data, [Theorem 1](#) indicates that we need to have $\log(p_n) = o(r_n^{-1} n^{1-2\kappa} + n r_n^{-2})$. Accordingly, p_n grows with the sample size n at an exponential rate.

To study the screening property via the quantile partial correlation $\widehat{\text{qpcor}}_\tau$, we consider $\mathcal{M}_* = \{j : \beta_{j\tau}^0 > 0 \text{ for } 1 \leq j \leq p\}$, which is the set of indices associated with the nonzero coefficients in the true sparse model (3) with nonsparsity size $s_n = |\mathcal{M}_*|$. Furthermore, we assume that the population QPCORs with nonzero coefficients in \mathcal{M}_* satisfy the following condition:

$$(C3) \min_{j \in \mathcal{M}_*} |\text{qpcor}_\tau\{Y, X_j|\mathbf{X}_{\mathcal{S}_j}\}| \geq C_0 r_n^{1/2} n^{-\kappa} \text{ for some } 0 < \kappa < 1/2 \text{ and } C_0 > 0.$$

In our proposed algorithm, we select variables sequentially by finding the variable with the maximal sample QPCOR and then adding it to the selected active set in each step. Let the resulting active set via the screening procedure be $\widehat{\mathcal{M}}_{v_n}$ such that the sample QPCORs of the selected variables in $\widehat{\mathcal{M}}_{v_n}$ are greater than a threshold. That is,

$$\widehat{\mathcal{M}}_{v_n} = \{j : \widehat{\text{qpcor}}_\tau\{Y, X_j|\mathbf{X}_{\mathcal{S}_j}\} \geq v_n \text{ for } 1 \leq j \leq p\},$$

where v_n is a threshold value. The theorem below presents the sure screening property.

Theorem 2. Under the conditions in [Theorem 1](#) and Condition (C3), taking C_2, C_3, C_2^*, C_3^* , and κ as given in [Theorem 1](#) and letting $v_n = C_4 r_n^{1/2} n^{-\kappa}$ with $C_4 \leq C_0/2$, we have that

$$P(\mathcal{M}_* \subset \widehat{\mathcal{M}}_{v_n}) \geq 1 - s_n\{C_2^* r_n^2 \exp(-C_2 r_n^{-1} n^{1-2\kappa}) + C_3^* r_n^2 \exp(-C_3 n r_n^{-2})\},$$

when n is sufficiently large.

It is worth noting that [Theorem 2](#) indicates that the probabilistic bound for the sure screening property depends on the number of nonzero coefficients s_n , but not on the number of covariates p_n . It also depends on r_n .

In addition to ensuring that relevant variables are selected, controlling the false selection rate is also critical. Ideally, we could assume that $\sup_{j \notin \mathcal{M}_*} |\text{qpcor}_\tau\{Y, X_j|\mathbf{X}_{\mathcal{S}_j}\}| = o(r_n^{1/2} n^{-\kappa})$ and then employ [Theorem 1](#), to find that, with probability tending to one, $\sup_{j \notin \mathcal{M}_*} |\widehat{\text{qpcor}}_\tau\{Y, X_j|\mathbf{X}_{\mathcal{S}_j}\}| \leq C_1 r_n^{1/2} n^{-\kappa}$ for any constant $C_1 > 0$. Accordingly, by the choice of v_n given in [Theorem 2](#), we obtain model selection consistency,

$$P(\mathcal{M}_* = \widehat{\mathcal{M}}_{v_n}) \rightarrow 1.$$

However, this ideal assumption may not be met in general. Hence, we consider a more practical assumption, $\sum_{j=1}^p |\text{qpcor}_\tau\{Y, X_j|\mathbf{X}_{\mathcal{S}_j}\}| = O(n^\zeta)$ for some $\zeta > 0$. Under this assumption, for any $c > 0$, the cardinality of $\{j : |\text{qpcor}_\tau\{Y, X_j|\mathbf{X}_{\mathcal{S}_j}\}| \geq c r_n^{1/2} n^{-\kappa}\}$ is no greater than $O(r_n^{-1/2} n^{\zeta+\kappa}) = O(n^{\zeta+\kappa-\omega/2})$ and $0 \leq \omega < \min((1 - 2\kappa), 2\kappa)$. Furthermore, on the set

$$\Omega_n = \left\{ \sup_{1 \leq j \leq p_n} |\widehat{\text{qpcor}}_\tau\{Y, X_j|\mathbf{X}_{\mathcal{S}_j}\} - \text{qpcor}_\tau\{Y, X_j|\mathbf{X}_{\mathcal{S}_j}\}| \leq c r_n^{1/2} n^{-\kappa} \right\},$$

we have

$$\begin{aligned} &|\{j : |\widehat{\text{qpcor}}_\tau\{Y, X_j|\mathbf{X}_{\mathcal{S}_j}\}| \geq 2c r_n^{1/2} n^{-\kappa}\}| \\ &\leq |\{j : |\text{qpcor}_\tau\{Y, X_j|\mathbf{X}_{\mathcal{S}_j}\}| \geq c r_n^{1/2} n^{-\kappa}\}| \leq C^* n^{\zeta+\kappa-\omega/2} \end{aligned} \quad (10)$$

for some constant $0 < C^* < \infty$. As a result, we obtain the following property that is used to control the size of the selected model.

Proposition 1. Under the conditions in [Theorem 1](#) and Condition (C3), taking C_2, C_3, C_2^*, C_3^* , and κ as given in [Theorem 1](#), letting $v_n = C_4 r_n^{1/2} n^{-\kappa}$ with $C_4 \leq C_0/2$, and assuming $\sum_{j=1}^p |\text{qpcor}_\tau\{Y, X_j|\mathbf{X}_{\mathcal{S}_j}\}| = O(n^\zeta)$ for some $\zeta > 0$, we have that

$$\begin{aligned} &P(|\widehat{\mathcal{M}}_{v_n}| \leq C^* n^{\zeta+\kappa-\omega/2}) \\ &\geq 1 - p_n\{C_2^* r_n^2 \exp(-C_2 r_n^{-1} n^{1-2\kappa}) + C_3^* r_n^2 \exp(-C_3 n r_n^{-2})\}, \end{aligned}$$

for some constant $0 < C^* < \infty$, when n is sufficiently large.

Remark 2. [Proposition 1](#) indicates that the proposed screening procedure via the quantile partial correlation can reduce the ultra-high dimensionality of the original model to the selected model size with a polynomial order of n . This proposition and [Theorem 2](#) imply that if we choose the first d variables sequentially based on the sample QPCOR with $d = \lceil n^{\varkappa+\zeta+\kappa-\omega/2} / \log(n) \rceil$ for some $\varkappa > 0$, then all relevant variables will be selected with high probability. Note that $\lfloor b \rfloor$ stands for the integer part of b . By assuming $\zeta < 1 + \omega/2 - \kappa$ and letting $\varkappa = 1 + \omega/2 - \kappa - \zeta$, we also have $d = \lceil n / \log(n) \rceil$, which is used in our numerical analysis and is commonly accepted in screening procedures (Fan and Lv 2008; He, Wang, and Hong 2013).

4. QPCOR and Selection

Applying quantile partial correlation, we first introduce the three screening algorithms. Based on the set of candidate models obtained via the screening procedure, we subsequently use the Bayesian information criterion to select the best model.

4.1. Screening Algorithms

In this section, we employ QPCOR to propose a quantile screening procedure (QPCS) for selecting variables. For the sake of comparison, we also generalize Cho and Fryzlewicz's (2012) TCS2 algorithm and Wang's (2009) FR algorithm from classical mean regression model to quantile regression model. We name them QTCS and QFR, respectively. In developing QPCS, we need to remove the confounding effect from the target variable that is induced by its correlated variables in each step. To this end, we consider a sequential test to identify a confounding subset for each X_j ($j = 1, \dots, p$). Let ρ_{jk} be the sample correlation coefficient of X_j and X_k . Then, define

$$S_j^v(m_j) = \{k \neq j \text{ and } k \in S_j : |\rho_{jk}| \text{ is among the first } m_j \text{ largest correlations of all possible correlations}\},$$

and name it the confounding set. A careful choice of m_j is important in the high-dimensional setting. For example, if m_j is too large, then any vector in R^n may be well approximated by some X_k with $k \in S_j$. We next consider a sequential testing procedure based on the partial correlations along the path $\{S_j^v(m_j)\}_{m_j \geq 1}$ to select m_j . This allows us to find the smallest subset so that all covariates not in this subset will have a zero partial correlation with X_j .

Let $\mathbb{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)^T$ be the design matrix and denote $\mathcal{Q}_S = \mathbf{I} - \mathbb{X}_S(\mathbb{X}_S^T \mathbb{X}_S)^{-1} \mathbb{X}_S^T \in R^{n \times n}$, where \mathbb{X}_S is any submatrix of \mathbb{X} . For $m_j \geq 1$, define the partial correlation as $\varrho_{jk}(S_j^v(m_j)) = c_{jk}^{-1} \mathbb{X}_j^T \mathcal{Q}_{S_j^v(m_j)} \mathbb{X}_k / n$, where $c_{jk} = \|\mathcal{Q}_{S_j^v(m_j)} \mathbb{X}_k\| / \|\mathcal{Q}_{S_j^v(m_j)} \mathbb{X}_j\| / n$. As for $m_j = 0$, $S_j^v(m_j)$ is an empty set and $\varrho_{jk}(S_j^v(m_j)) = \mathbb{X}_j^T \mathbb{X}_k / n$. Furthermore, let

$$F_{jk}(S_j^v(m_j)) = 0.5 \log\{[1 + \varrho_{jk}(S_j^v(m_j))] / [1 - \varrho_{jk}(S_j^v(m_j))]\},$$

which is the Fisher's Z-transformation considered in Kalisch and Bühlmann (2007) for identifying nodes connected to the variable X_j conditional on a set of other nodes in a Gaussian graph. Then, sequentially select the smallest size, m_j^* , that satisfies $(n - |S_j^v(m_j^*)| - 3)^{1/2} \max_{k \notin S_j^v(m_j^*)} |F_{jk}(S_j^v(m_j^*))| < z_{1-\alpha/2}$, where $z_{1-\alpha/2}$ is the threshold of z values with a prespecified significance level α and $S_j^{v,+}(m_j^*) = \{j\} \cup S_j^v(m_j^*)$. The resulting m_j^* can help us to determine the size of the selected confounding set, denoted by \widehat{m}_j . It is worth noting that based on our theoretical condition given in Theorem 1, we have $r_n = o(n^{1/2})$. Thus, $\widehat{m}_j \leq r_n = o(n^{1/2})$. We then let \widehat{m}_j be bounded by $c\{n/\log(n)\}^{1/2}$ for some constant $c > 0$. Afterwards, let $\widehat{m}_j = m_j^*$ if $m_j^* < c\{n/\log(n)\}^{1/2}$; $\widehat{m}_j = c\{n/\log(n)\}^{1/2}$, otherwise. In practice, the constant c should not be too large so that the resulting size \widehat{m}_j is under control. In our numerical analysis, we choose $c = 1$. Denote the selected confounding set as $S_j^v = S_j^v(\widehat{m}_j)$. The above procedure allows us to find the confounding subset of the j th variable, and we will use it in the screening algorithm given below.

Algorithm 1 (QPCS). Start with an empty active set $\mathcal{A}^{(0)} = \emptyset$.

Step 1. In the k th step, for given $\mathcal{A}^{(k-1)}$, we update $S_j = \mathcal{A}^{(k-1)} \cup S_j^v$. Then, employ the maximal sample QPCOR to find the variable index j^* that satisfies $j^* = \arg \max_{j \notin \mathcal{A}^{(k-1)}} |\widehat{\text{qpcor}}_\tau\{Y, X_j | \mathbf{X}_{S_j}\}|$. Update $\mathcal{A}^{(k)} = \mathcal{A}^{(k-1)} \cup \{j^*\}$.

Step 2. Repeat Step 1 until the cardinality of active set $|\mathcal{A}^{(d^*)}|$ reaches a prespecified d^* .

Step 3. Starting from the $k = (d^* + 1)$ th step, we set $S_j = \mathcal{A}^{(d^*)} \cup S_j^v$. In the k th step, find $j^* = \arg \max_{j \notin \mathcal{A}^{(k-1)}} |\widehat{\text{qpcor}}_\tau\{Y, X_j | \mathbf{X}_{S_j}\}|$. Update $\mathcal{A}^{(k)} = \mathcal{A}^{(k-1)} \cup \{j^*\}$.

Step 4. Repeat Step 3 until the cardinality of active set $|\mathcal{A}^{(d)}|$ reaches a prespecified value $d < n$.

In the above algorithm, the conditioning set S_j contains the selected variables up to step d^* and a subset of variables with nonnegligible correlations identified by the sequential testing procedure.

In linear regression modeling, Cho and Fryzlewicz (2012) proposed the TCS2 algorithm and then demonstrated that it usually performs well in comparison with LASSO, SCAD, MCP, FR, and iterative SIS (ISIS, see Fan and Lv 2008). This inspires us to extend their TCS2 to quantile regression, and we name it QTCS.

Algorithm 2 (QTCS). Start with an empty active set $\mathcal{A}^{(0)} = \emptyset$.

Step 1. In the k th step, for given $\mathcal{A}^{(k-1)}$, let $S_j = \mathcal{A}^{(k-1)}$. Then, find the variable index that has the maximal sample QPCOR such that $j^* = \arg \max_{j \notin \mathcal{A}^{(k-1)}} |\widehat{\text{qpcor}}_\tau\{Y, X_j | \mathbf{X}_{S_j}\}|$. If $S_j^v / \mathcal{A}^{(k-1)} = \emptyset$, let $j^* = j'$ and go to Step 3.

Step 2. If $S_j^v / \mathcal{A}^{(k-1)} \neq \emptyset$, then screen the sample QPCOR for all X_j in which $j \in S_j^v \cup \{j'\}$. Let $S_j = \mathcal{A}^{(k-1)} \cup S_j^v$, and find $j^* = \arg \max_{j \in S_j^v \cup \{j'\}} |\widehat{\text{qpcor}}_\tau\{Y, X_j | \mathbf{X}_{S_j}\}|$.

Step 3. Update $\mathcal{A}^{(k)} = \mathcal{A}^{(k-1)} \cup \{j^*\}$.

Step 4. Repeat Steps 1–3 until the cardinality of active set, $|\mathcal{A}^{(d^*)}|$, reaches a prespecified $d^* = \lceil C^* \{n/\log(n)\}^{1/2} \rceil$ for some constant $C^* > 0$.

Step 5. Starting from the $k = (d^* + 1)$ th step, repeat Steps 1–3 by letting $S_j = \mathcal{A}^{(d^*)}$ and $S_j = \mathcal{A}^{(d^*)} \cup S_j^v$ in Steps 1 and 2, respectively. Repeat the procedure until the cardinality of active set, $|\mathcal{A}^{(d)}|$, reaches a prespecified value $d < n$.

Based on extensive simulation studies in linear model settings, Wang (2009) indicated that FR is a promising method for variable screening by comparing with LASSO, SCAD, SIS, and ISIS. This motivates us to generalize his forward selection screening algorithm to quantile regression, and name it QFR.

Algorithm 3 (QFR). Start with an empty active set $\mathcal{A}^{(0)} = \emptyset$.

Step 1. In the k th step, for given $\mathcal{A}^{(k-1)}$, let $S_j = \mathcal{A}^{(k-1)}$ for $k \leq d^*$ and $S_j = \mathcal{A}^{(d^*)}$ for $k > d^*$. Then, find the variable index that has the maximal sample QPCOR such that $j^* = \arg \max_{j \notin \mathcal{A}^{(k-1)}} |\widehat{\text{qpcor}}_\tau\{Y, X_j, \mathbf{X}_{\mathcal{A}^{(k-1)}}\}|$. Update $\mathcal{A}^{(k)} = \mathcal{A}^{(k-1)} \cup \{j^*\}$.

Step 2. Repeat Step 2 until the cardinality of active set, $|\mathcal{A}^{(k)}|$, reaches a prespecified value $d < n$.

For the sake of comparison, Table 1 summarizes the three algorithms.

Without the thresholding step, by replacing the sample QPCOR with the tilted correlation defined in Cho and Fryzlewicz (2012) and the residual sum of squares given in Wang (2009), respectively, the QTCS and QFR become TCS2 and FR.

Table 1. Comparison of the QPCS, QTCS, and QFR algorithms.

Initialization	QPCS $A^{(0)} = \emptyset$	QTCS $A^{(0)} = \emptyset$	QFR $A^{(0)} = \emptyset$
Action	One variable is selected	One variable is selected	One variable is selected
Conditional set S_j for $k \leq d^*$	$A^{(k-1)} \cup S_j^v$	$A^{(k-1)}$ or $A^{(k-1)} \cup S_j^v$	current set $A^{(k-1)}$
Conditional set S_j for $k > d^*$	$A^{(d^*)} \cup S_j^v$	$A^{(d^*)}$ or $A^{(d^*)} \cup S_j^v$	$A^{(d^*)}$

To use the above three algorithms, we need to specify d^* and d . It is worth noting that the thresholding size d^* needs to satisfy $d^* \leq r_n = o(n^{1/2})$ (see Algorithm 1). Hence, we consider $d^* = \lceil C^* \{n/\log(n)\}^{1/2} \rceil$ for some $C^* > 0$. In addition, choosing a value of C^* needs to assure that d^* does not exceed d due to the requirement of the screening algorithms. With Remark 3, we have set $d = \lfloor n/\log(n) \rfloor$. To meet the requirement $d^* < d$ with $n = 200$ and $n = 120$ used in our simulation and empirical examples, respectively, we choose $C^* = 2$, which yields good performance in our numerical studies. However, this does not exclude other possible choices that also satisfy this requirement.

Since $|S_j^v| = m_j \leq \lceil c\{n/\log(n)\}^{1/2} \rceil$, we have $|S_j| \leq |S_j^v| + d^* \leq \tilde{C}\{n/\log(n)\}^{1/2}$ for some constant $0 < \tilde{C} < \infty$ and $j = 1, \dots, p$. This provides an upper bound for the conditional set of each variable, which is not very large. Otherwise, any vector in R^n can be well approximated by the variables in this set. A similar consideration can be found in Cho and Fryzlewicz (2012) when they discussed their “conditioning set” C_j . Note that C_j is our confounding set S_j^v , which is different from the conditional set S_j . In their assumption 3 of page 598, they consider a bound for the size of the conditioning set C_j such that $|C_j| \leq Cn^\xi$ with $\xi \in [0, 2(\gamma - \delta))$ for $\delta \in [0, 1/2)$ and $\gamma \in (\delta, 1/2)$. Thus, $|C_j| = o(n^{1-2\delta})$ for $\delta \in [0, 1/2)$. It is of interest to note that, based on the condition given in Theorem 1, our conditional set $|S_j| \leq r_n = o(n^{1-2\kappa})$ for $0 < \kappa < 1/2$.

Remark 3. From Table 1 and the above discussion, we find that both QPCS and QTCS can prevent more overfitting than QFR. This is because they consider the confounding effect of explanatory variables, while QFR does not take it into account. Although $|S_j|$ in QPCS and $|C_j|$ (i.e., $|S_j^v|$) in QTCS have the same order, the confounding set S_j^v in QPCS is always included in the conditional set S_j in every screening step, while C_j in QTCS may not be always included in S_j . Accordingly, QPCS is likely to reduce more overfitting than QTCS.

Based on the quantile partial correlation, we have introduced three screening algorithms. Although the quantile correlation is not the focus of our article, one can employ it to propose the sure independent screening procedure for quantile regression. Specifically, we adopt the SIS method of Fan and Lv (2008) by replacing their Pearson correlation with the quantile correlation QCOR. The resulting selected model is

$$\widehat{\mathcal{M}}_{\text{QCOR}} = \{j : |\widehat{\text{qcor}}_\tau\{Y, X_j\}| \text{ is among the first } d \text{ largest quantile correlations} \}.$$

He, Wang, and Hong (2013) also applied the SIS method for model selection in nonparametric quantile regression. In classical mean regression, Fan and Lv (2008) also introduced ISIS

for selecting variables. As aforementioned, Wang (2009) and Cho and Fryzlewicz (2012) have demonstrated that their FR and TCS2 algorithms perform well in comparison with ISIS, respectively. Thus, we will focus our numerical comparison of the newly proposed procedure with the corresponding procedures proposed in Wang (2009) and Cho and Fryzlewicz (2012).

4.2. Best Subset Selection

In the previous section, the proposed QPCS algorithm generates a solution path $\mathbb{A} = \{\mathcal{A}^{(k)}, 1 \leq k \leq d\}$, which includes the d selected models $\mathcal{A}^{(1)} \subset \mathcal{A}^{(2)} \subset \dots \subset \mathcal{A}^{(d)}$. To find the best model among them, we consider an extended Bayesian information criterion (EBIC) for best subset selection. This criterion has been used for classical mean regression model in high-dimensional data analysis (e.g., see Chen and Chen 2008; Wang 2009). As for our quantile regression model, we follow the approach of Lee, Noh, and Park (2014) and adopt the criterion

$$\text{EBIC}(\mathcal{A}) = \log \left(n^{-1} \sum_{i=1}^n \rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{A}}^T \widehat{\boldsymbol{\beta}}_{\mathcal{A}}) \right) + |\mathcal{A}| \frac{\log n}{2n} C_n, \tag{11}$$

where C_n is a positive constant that diverges along with the sample size n , and

$$\widehat{\boldsymbol{\beta}}_{\mathcal{A}} = \arg \min_{\boldsymbol{\beta}_{\mathcal{A}} \in R^{|\mathcal{A}|}} n^{-1} \sum_{i=1}^n \rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{A}}^T \boldsymbol{\beta}_{\mathcal{A}}).$$

Let $\widehat{k} = \arg \min_{1 \leq k \leq d} \text{EBIC}(\mathcal{A}^{(k)})$, and denote the resulting best model $\widehat{\mathcal{M}}_{\text{EBIC}} = \mathcal{A}^{(\widehat{k})}$. We make the following condition, which corresponds to condition (A2)(ii) in Lee, Noh, and Park (2014) and is needed for establishing the consistency of EBIC.

(C4) There exist positive finite constants m' and M' such that

$$m' \leq \lambda_{\min}(E(\mathbf{X}_S \mathbf{X}_S^T)) \leq \lambda_{\max}(E(\mathbf{X}_S \mathbf{X}_S^T)) \leq M'$$

uniformly for any subset $S \subset \{1, \dots, p\}$ satisfying $|S| \leq C^* n^{\varsigma + \kappa - \omega/2}$.

We next establish the screening consistency of the best model selected by EBIC.

Theorem 3. Under the conditions given in Proposition 1 and Condition (C4), and assuming that $\varsigma < 1/2 + \omega/2 - \kappa$, $C_n^{-1} = o(1)$, $C_n \log(n) n^{(\varsigma + \kappa - \omega/2) - 1} = o(1)$, and $E|\varepsilon| < \infty$, we have $P(\mathcal{M}_* \subset \widehat{\mathcal{M}}_{\text{EBIC}}) \rightarrow 1$ as $n \rightarrow \infty$.

When $C_n = 1$, EBIC reduces to the classical BIC (Schwarz 1978). Recently, Wang, Li, and Leng (2009) and Lee, Noh, and Park (2014) used $C_n = \log(\log(d))$ and $C_n = \log(d)$, respectively, in their simulation studies when the number of predictors diverged along with the sample size. We use both approaches in our numerical studies. In addition, EBIC can be applied not only to QPCS, but also to QTCS and QFR for best subset selection. It is worth noting that our proposed QPCS algorithm yields a family of nested candidate models, $\mathcal{A}^{(1)} \subset \mathcal{A}^{(2)} \subset \dots \subset \mathcal{A}^{(d)}$. Thus, we propose using the model selection criterion EBIC to select the best model. On the other hand, the screening procedure of SIS only produces a single final model, followed by the SCAD or other penalized methods for variable selections.

In addition to classical model selection, another popular variable selection approach is penalization. In other words, by using

Table 2. The average rank of the relevant predictors R_j and the average number of the minimum size of the selected model M with $p = 1000$ and $\rho = 0.95$ in Example 1.

τ	Method	Standard normal					Laplace distribution				
		R_1	R_2	R_3	R_4	M	R_1	R_2	R_3	R_4	M
0.2	QPCS	2.205	1.945	2.060	3.825	4.015	4.060	2.520	4.885	8.845	13.500
	QTCS	4.050	3.570	4.120	7.180	8.540	6.970	14.290	8.500	19.595	33.530
	QFR	5.215	4.150	4.560	275.955	276.490	9.600	8.980	12.505	480.455	483.040
	SIS	343.580	327.575	330.285	499.975	682.580	304.795	302.405	307.890	499.880	668.845
0.5	QPCS	2.125	2.140	2.090	3.795	4.095	2.275	3.375	3.565	4.040	6.725
	QTCS	4.270	3.875	3.825	17.550	18.615	4.340	5.405	4.730	36.810	39.565
	QFR	4.470	4.335	3.995	345.760	345.875	5.000	5.455	8.110	487.020	487.545
	SIS	345.635	345.250	337.305	500.985	691.675	320.695	310.385	319.025	510.840	685.620
0.8	QPCS	1.935	2.085	2.145	3.855	4.010	5.015	3.835	3.180	11.195	14.520
	QTCS	3.775	3.905	4.115	6.540	7.615	15.445	7.470	8.720	32.495	46.225
	QFR	4.410	4.375	4.150	252.240	252.315	9.425	7.945	12.175	459.700	459.715
	SIS	338.835	335.465	346.660	501.060	686.38	311.050	308.080	320.905	492.905	662.565

the selected model $\mathcal{A}^{(d)}$ with $d < n$ from the screening procedure, we can obtain the estimated parameters by minimizing

$$n^{-1} \sum_{i=1}^n \rho_{\tau}(Y_i - \mathbf{X}_{i, \mathcal{A}^{(d)}}^T \boldsymbol{\beta}_{\mathcal{A}^{(d)}}) + \sum_{j=1}^{|\mathcal{A}^{(d)}|} p_{\lambda}(|\beta_{j, \mathcal{A}^{(d)}}|),$$

with respect to parameters $\boldsymbol{\beta}_{\mathcal{A}^{(d)}} = (\beta_{1, \mathcal{A}^{(d)}}, \dots, \beta_{|\mathcal{A}^{(d)}|, \mathcal{A}^{(d)}})^T$, where $p_{\lambda}(\cdot)$ is a penalty function with a regularization parameter λ . In our numerical studies, we consider the LASSO penalty for demonstration, but other penalties such as SCAD and MCP can also be applied. It is worth noting that the penalization method only employs the largest selected model $\mathcal{A}^{(d)}$, while EBIC uses the entire solution path obtained from the screening procedure.

5. Simulation Studies

In this section, we conduct simulation studies to compare the finite sample performance of the four screening procedures QPCS, QTCS, QFR, and SIS. We further illustrate the extended BIC approach for best subset selection by using $C_n = \log(\log(d))$ and $C_n = \log(d)$, respectively, and denote the corresponding methods by EBIC1 and EBIC2. We also compare the EBIC method with the LASSO penalization method after the screening, where the tuning parameter for LASSO is selected by the BIC method. Moreover, we compare our method with the l_1 penalization of Belloni and Chernozhukov (2011) and the ISIS-SCAD method of Fan and Lv (2008) in the last example. The tuning parameter for the l_1 penalization method is selected by the approach described in sec. 2.3 of Belloni and Chernozhukov (2011), and the tuning parameter for the ISIS-SCAD method is selected by the extended BIC as given in the R package ‘‘SIS.’’

To demonstrate the performance of the QPCS, QTCS, QFR, and SIS screening procedures, we present two examples. We consider three quantiles $\tau = 0.2, 0.5, \text{ and } 0.8$, and all simulation results are based on 200 realizations with $n = 200$ and $p = 1000$. Moreover, seven measurements are used to assess the screening and selection performance: the rank of selected variables (see Liu, Li, and Wu 2014), the number of true positive and false positive selections (see Liu, Li, and Wu 2014), and the number of correct-fitting, overfitting, and underfitting selections (see Wang, Li, and Tsai 2007). We next describe these measures in detail.

1. R_j : the average rank of X_j ;
2. M : the average minimum size of the selected model that contains all the relevant (i.e., true) predictors;
3. TP : the average number of true positives (i.e., the average number of relevant predictors being correctly selected);
4. FP : the average number of false positives (i.e., the average number of irrelevant predictors being incorrectly selected);
5. C : the proportion in which exactly relevant predictors are selected;
6. O : the proportion in which all relevant predictors and some irrelevant predictors are selected;
7. I : the proportion in which some relevant predictors are not selected.

Note that the average or proportion used in the above measures is calculated from 200 realizations.

Example 1. We generate the response from Model D considered by Cho and Fryzlewicz (2012) and originally taken from Fan and Lv (2008):

$$Y = \beta X_1 + \beta X_2 + \beta X_3 - 3\beta\sqrt{\rho}X_4 + \varepsilon,$$

where the predictors \mathbf{X} are simulated from $N(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \{\sigma_{ij}\}$ is a $p \times p$ covariance matrix satisfying $\sigma_{ii} = 1$ and $\sigma_{ij} = \rho$, $j \neq i$, except that $\sigma_{4j} = \sigma_{i4} = \sqrt{\rho}$. Thus, X_4 is marginally uncorrelated with Y at the population level. To take quantiles into account in regression coefficients, we let $\beta = 2.5(1 + |\tau - 0.5|)$ rather than $\beta = 2.5$ given in Cho and Fryzlewicz (2012). The random error ε is generated according to the standard normal distribution and the Laplace distribution. We also let $\rho = 0.5$ and $\rho = 0.95$ represent a moderate correlation and a high correlation, respectively. For the sake of saving space, we report the results for $\rho = 0.5$ in Tables S1–S3 of the online supplementary materials.

Table 2 reports R_j ($j = 1, \dots, 4$) and M for $p = 1000$ and $\rho = 0.95$. When the predictors are highly correlated ($\rho = 0.95$), SIS cannot successfully identify all four relevant predictors. In Table S1 of the online supplementary materials, we find that even under moderate correlation ($\rho = 0.5$), the SIS approach fails to identify the fourth predictor, which is marginally uncorrelated with Y (see the large values of R_4 in Table S1 of the online supplementary materials).

Table 3. Variable selection results of TP and FP for the extended BIC and LASSO with $\rho = 1000$ and $\rho = 0.95$ in Example 1.

τ		QPCS			PTCS			QFR		
		EBIC1	EBIC2	LASSO	EBIC1	EBIC2	LASSO	EBIC1	EBIC2	LASSO
Standard normal										
0.2	TP	4.000	4.000	2.035	3.945	3.830	2.035	3.385	3.210	2.020
	FP	10.405	0.430	0.345	10.465	2.600	1.375	11.780	4.680	1.895
0.5	TP	4.000	3.980	2.490	3.790	3.200	2.500	3.265	2.470	2.330
	FP	9.605	0.145	2.300	10.365	1.895	3.555	11.105	2.680	4.750
0.8	TP	4.000	4.000	2.995	3.955	3.785	2.685	3.350	3.085	2.410
	FP	10.580	0.305	0.350	10.585	2.415	2.030	11.985	4.585	2.875
Laplace distribution										
0.2	TP	3.960	3.740	1.930	3.605	2.680	1.970	2.905	1.970	1.955
	FP	9.740	0.765	1.125	10.785	2.100	2.145	11.660	3.155	2.750
0.5	TP	3.960	3.830	2.460	3.580	2.515	2.335	3.010	2.455	2.430
	FP	6.625	0.150	2.650	8.685	1.360	2.865	10.215	1.675	3.250
0.8	TP	3.920	3.625	2.875	3.525	2.690	2.560	2.910	2.105	1.880
	FP	10.205	0.760	1.205	10.825	2.320	2.895	11.985	3.030	3.610

The aim of the QFR method is to remove the effect from the predictors identified in the previous steps. It performs reasonably well for the moderate correlation (see small values of R_j and M in Table S1 of the online supplementary materials). When $\rho = 0.95$, however, QFR is not capable of identifying the fourth predictor. This finding is not surprising since FR is not designed to remove high collinearity (i.e., confounding) effect. As for the QPCS and QTCS screening procedures, Table 2 indicates that both are able to control the effect of collinearity and identify relevant variables. However, QPCS is uniformly superior to QTCS in all measures. This is because QPCS can prevent more overfitting than QTCS by removing the confounding effect in every sequential step.

After examining screening performance, we next evaluate best subset selection. Since SIS does not show strong performance, we only consider variable selection via other three screening procedures. Table 3 reports TP and FP calculated under three selection methods, EBIC1, EBIC2, and LASSO, for $p = 1000$ and $\rho = 0.95$. Furthermore, Table 4 correspondingly presents the percentages of correct-fitting (C), overfitting (O), and incorrect-fitting (I) for $p = 1000$ and $\rho = 0.95$.

We observe that the TP values for the LASSO method are much smaller than 4, which is the number of true predictors in the model. In addition, the FP values are not large since LASSO tends to select a small number of variables among highly correlated covariates. Moreover, the proportion of incorrect-fitting is most often 1 for the LASSO method. Hence, LASSO is not an effective method for variable selection in this context. On the contrary, EBIC1 tends to exhibit overfitting, as evidenced by FPs twice as large as true number of predictors. In addition, the proportion of overfitting is often very large as shown in Table 4, except for QFR. This is because QFR tends to fit incorrectly under this scenario, which is consistent with the findings in Table 3. In comparison with EBIC1, EBIC2 yields much better FP with slightly weaker TP. In addition, EBIC2 is uniformly superior to LASSO in both TP and FP measures. Consequently, EBIC2 is a favorable choice. Moreover, Tables 3 and 4 indicate that QPCS-EBIC2 outperforms its competitors in best subset selection. Finally, it is not surprising that the performance of all screening and selection procedures deteriorates when ρ becomes large, or ε has a heavy-tailed distribution. To save

Table 4. Variable selection results of C, O, and I for the extended BIC and LASSO with $p = 1000$ and $\rho = 0.95$ in Example 1.

τ		QPCS			QTCS			QFR		
		EBIC1	EBIC2	LASSO	EBIC1	EBIC2	LASSO	EBIC1	EBIC2	LASSO
Standard normal										
0.2	C	0.000	0.695	0.000	0.000	0.125	0.000	0.000	0.000	0.000
	O	1.000	0.305	0.000	0.950	0.760	0.000	0.400	0.390	0.000
	I	0.000	0.000	1.000	0.050	0.115	1.000	0.600	0.610	1.000
0.5	C	0.000	0.845	0.000	0.000	0.155	0.000	0.000	0.000	0.000
	O	1.000	0.135	0.000	0.795	0.485	0.000	0.300	0.250	0.000
	I	0.000	0.020	1.000	0.205	0.360	1.000	0.700	0.750	1.000
0.8	C	0.000	0.765	0.000	0.000	0.130	0.000	0.000	0.000	0.000
	O	1.000	0.235	0.000	0.965	0.760	0.000	0.350	0.335	0.000
	I	0.000	0.000	1.000	0.035	0.110	1.000	0.650	0.665	1.000
Laplace distribution										
0.2	C	0.000	0.530	0.000	0.000	0.110	0.000	0.000	0.000	0.000
	O	0.960	0.290	0.000	0.715	0.340	0.000	0.050	0.045	0.000
	I	0.040	0.180	1.000	0.285	0.550	1.000	0.950	0.955	1.000
0.5	C	0.040	0.805	0.000	0.010	0.160	0.000	0.000	0.005	0.000
	O	0.930	0.085	0.000	0.605	0.195	0.000	0.080	0.020	0.000
	I	0.070	0.110	1.000	0.385	0.645	1.000	0.920	0.795	1.000
0.8	C	0.000	0.470	0.000	0.000	0.045	0.000	0.000	0.000	0.000
	O	0.935	0.340	0.005	0.650	0.370	0.000	0.080	0.080	0.000
	I	0.065	0.190	0.095	0.350	0.585	1.000	0.920	0.920	1.000

space, we report the additional simulation results for $p = 2000$ in Tables S4– S6 of the online supplementary materials. These simulations yield the same conclusion as that of $p = 1000$.

Example 2. We generate the response from the model

$$Y = \beta X_1 + \beta X_2 + \beta X_3 - 3\beta\sqrt{\rho}X_4 + 0.25\beta X_5 + \varepsilon,$$

where β , ρ , and \mathbf{X} are defined as in Example 1 except that $\sigma_{5j} = \sigma_{15} = 0$ such that X_5 is uncorrelated with X_j ($j \neq 5$). In addition, X_5 has a small contribution to Y . This model is also considered by Cho and Fryzlewicz (2012) and Fan and Lv (2008). To save space, we report the results for $\rho = 0.5$ in Tables S1– S3 of the online supplementary materials.

Tables 5 and S7 report R_j ($j = 1, \dots, 4$) and M for $\rho = 0.95$ and 0.5, respectively. From Table S7 in the online supplementary materials, we observe that SIS gives large values for R_4 , R_5 , and M for $\rho = 0.5$. Hence, SIS is not able to identify variables X_4 and X_5 in this case. When $\rho = 0.95$, SIS is able to identify X_5 due to its lack of correlation with other variables, whereas it fails to identify variables X_1 to X_4 since they are highly correlated with others. As a result, QFR, QPCS, and QTCS outperform SIS. For these three procedures, we obtain the same conclusion as in Example 1. Both QPCS and QTCS are superior to QFR, and QPCS performs the best. Tables 6 and 7 summarize the results of subset selection, by presenting TP, FP, and the proportions of correct-fitting (C), overfitting (O), and incorrect-fitting (I) calculated via EBIC1, EBIC2, and LASSO. Both tables show that QPCS-EBIC2 performs the best. The same findings emerge from Tables S8 and S9 of the online supplementary materials when $\rho = 0.5$.

Example 3. In the first two examples, we have demonstrated the performance of the proposed variable screening procedure. In this example, we compare our recommended method QPCS-EBIC2 with other two methods, the l_1 penalization and ISIS-SCAD. We use the “SIS” R package to implement ISIS-SCAD, in which it first implements the Iterative Sure Independence

Table 5. The average rank of the relevant predictors R_j and the average minimum size of the selected model M with $\rho = 1000$ and $\rho = 0.95$ in Example 2.

τ	Method	Standard normal					M	Laplace distribution					
		R_1	R_2	R_3	R_4	R_5		R_1	R_2	R_3	R_4	R_5	
$\rho = 0.95$													
0.2	QPCS	2.720	2.560	2.685	4.560	2.655	5.060	6.380	4.625	5.990	13.455	3.155	21.485
	QTCS	5.020	4.870	4.870	8.450	1.010	9.450	7.445	15.180	18.815	20.085	1.120	42.720
	QFR	5.775	5.175	5.340	235.085	1.010	235.185	10.605	8.685	14.225	444.770	1.145	446.630
0.5	SIS	388.100	373.320	377.390	491.820	1.015	700.930	330.035	330.145	340.550	509.575	1.830	689.145
	QPCS	2.765	2.805	2.825	4.510	2.495	5.19	2.950	4.540	2.970	5.140	2.580	7.420
	QTCS	4.580	5.655	5.235	20.290	1.045	21.415	6.765	12.340	6.870	26.250	1.010	35.080
0.8	QFR	5.650	5.315	6.165	358.660	1.045	358.77	9.715	6.085	9.905	474.955	1.010	475.385
	SIS	342.660	326.660	330.650	505.585	1.015	686.76	359.165	349.170	350.990	499.660	5.920	698.655
	QPCS	2.775	2.785	2.790	4.645	2.475	5.200	6.510	3.325	3.980	16.295	5.235	23.035
	QTCS	4.810	5.040	4.625	7.055	1.015	8.145	6.945	18.015	9.310	19.055	1.065	36.94
	QFR	5.755	5.740	5.290	214.880	1.015	214.965	15.905	15.115	10.305	461.625	1.095	468.02
	SIS	383.340	385.060	383.290	499.870	5.855	713.315	347.490	344.295	375.930	518.075	1.045	715.87

Table 6. Variable selection results of TP and FP for the extended BIC and LASSO with $\rho = 1000$ and $\rho = 0.95$ in Example 2.

τ		QPCS			QTCS			QFR		
		EBIC1	EBIC2	LASSO	EBIC1	EBIC2	LASSO	EBIC1	EBIC2	LASSO
Standard normal										
0.2	TP	5.000	5.000	2.995	4.920	4.785	2.995	4.340	4.135	2.995
	FP	9.420	0.450	0.580	10.085	2.765	2.040	10.905	4.680	2.825
0.5	TP	5.000	4.965	3.495	4.800	4.075	3.500	4.220	3.255	3.220
	FP	8.550	0.250	2.260	9.655	1.755	2.385	10.496	2.555	2.605
0.8	TP	5.000	4.995	3.970	4.975	4.810	3.995	4.370	4.120	3.985
	FP	9.380	0.525	0.610	9.735	2.535	2.425	10.995	4.445	3.310
Laplace distribution										
0.2	TP	4.910	4.495	2.890	4.635	3.685	2.855	3.885	3.030	2.860
	FP	9.550	0.885	1.480	9.935	1.865	2.645	11.080	2.995	3.500
0.5	TP	4.950	4.605	3.475	4.640	3.445	3.060	3.985	2.800	2.405
	FP	6.580	0.205	2.625	8.480	1.135	2.860	9.400	1.325	3.145
0.8	TP	4.905	4.570	3.845	4.600	3.850	3.825	3.875	3.205	2.825
	FP	9.570	1.025	3.975	9.770	2.045	3.975	11.145	2.945	3.915

Screening, and then fits the final regression model by using the SCAD penalty. In short, we denote these three methods as QPCS, l_1 , and ISIS, respectively. We simulate data from the same data generating process given in Example 1 with $\rho = 1000$. Since

Table 7. Variable selection results of C, O, and I for the extended BIC and LASSO with $\rho = 1000$ and $\rho = 0.95$ in Example 2.

τ		QPCS			QTCS			QFR		
		EBIC1	EBIC2	LASSO	EBIC1	EBIC2	LASSO	EBIC1	EBIC2	LASSO
Standard normal										
0.2	C	0.000	0.675	0.005	0.000	0.150	0.000	0.000	0.000	0.000
	O	1.000	0.325	0.000	0.925	0.720	0.000	0.355	0.335	0.000
	I	0.000	0.000	0.995	0.075	0.130	1.000	0.645	0.665	1.000
0.5	C	0.000	0.775	0.025	0.000	0.175	0.000	0.000	0.010	0.000
	O	1.000	0.190	0.165	0.815	0.455	0.000	0.265	0.210	0.000
	I	0.000	0.035	0.810	0.185	0.370	1.000	0.735	0.780	1.000
0.8	C	0.000	0.615	0.290	0.000	0.155	0.000	0.000	0.005	0.000
	O	1.000	0.380	0.060	0.975	0.745	0.000	0.390	0.365	0.000
	I	0.000	0.005	0.650	0.025	0.100	1.000	0.610	0.630	1.000
Laplace distribution										
0.2	C	0.000	0.420	0.000	0.000	0.125	0.000	0.000	0.005	0.000
	O	0.920	0.345	0.010	0.730	0.345	0.000	0.065	0.055	0.000
	I	0.080	0.235	0.990	0.270	0.530	1.000	0.935	0.940	1.000
0.5	C	0.000	0.735	0.000	0.000	0.165	0.000	0.000	0.020	0.000
	O	0.960	0.115	0.145	0.700	0.175	0.000	0.100	0.025	0.000
	I	0.040	0.150	0.855	0.300	0.660	1.000	0.900	0.955	1.000
0.8	C	0.000	0.350	0.110	0.000	0.115	0.000	0.000	0.000	0.000
	O	0.925	0.440	0.045	0.700	0.375	0.000	0.050	0.040	0.000
	I	0.075	0.210	0.845	0.300	0.510	1.000	0.950	0.960	1.000

Table 8. Variable selection results of TP and FP for the QPCS, l_1 , and ISIS methods with $\rho = 1000$ and $\tau = 0.5$ in Example 3.

	$\rho = 0.95$			$\rho = 0.50$			$\rho = 0.05$		
	QPCS	l_1	ISIS	QPCS	l_1	ISIS	QPCS	l_1	ISIS
Standard normal									
TP	3.980	0.425	3.825	4.000	3.065	4.000	4.000	4.000	4.000
FP	0.145	0.350	4.950	0.145	45.365	1.035	0.125	16.815	1.150
Laplace distribution									
TP	3.830	0.435	3.710	4.000	3.005	3.985	4.000	4.000	4.000
FP	0.150	0.280	4.905	0.015	39.110	1.860	0.015	16.585	1.765

ISIS is designed for mean regression models, we only consider $\tau = 0.5$ for fair comparison.

Table 8 reports TP and FP, and Table 9 presents the percentages of correct-fitting (C), overfitting (O), and incorrect-fitting (I). For the l_1 penalization method, we observed that both of the true positive and false positive values are very small when $\rho = 0.95$, since it only selects one or zero variable from highly correlated covariates. For $\rho = 0.5$ and 0.05 , however, it has very large false positives. It is also worth noting that l_1 has very large portions of incorrect-fitting even at the moderate correlation of $\rho = 0.5$. This is because it often has missed the fourth variable, X_4 , which is highly correlated with the other three variables. As a result, l_1 can be seriously affected by the correlation and its performance deteriorates as the correlation becomes larger. As for ISIS, its performance at $\rho = 0.95$ is worse than that at $\rho = 0.5$ and 0.05 . In addition, it has larger false positive values and its correct-fitting rates close to zero when $\rho = 0.95$. In contrast to

Table 9. Variable selection results of C, O, and I for the QPCS, l_1 , and ISIS methods with $\rho = 1000$ and $\tau = 0.5$ in Example 3.

	$\rho = 0.95$			$\rho = 0.50$			$\rho = 0.05$		
	QPCS	l_1	ISIS	QPCS	l_1	ISIS	QPCS	l_1	ISIS
Standard normal									
C	0.845	0.000	0.070	0.875	0.000	0.385	0.890	0.000	0.355
O	0.135	0.000	0.855	0.125	0.065	0.615	0.110	1.000	0.645
I	0.020	1.000	0.075	0.000	0.935	0.000	0.000	0.000	0.000
Laplace distribution									
C	0.805	0.000	0.055	0.985	0.000	0.195	0.965	0.000	0.180
O	0.085	0.000	0.870	0.015	0.005	0.800	0.035	1.000	0.820
I	0.110	1.000	0.075	0.000	0.995	0.005	0.000	0.000	0.000

l_1 and ISIS, Tables 8 and 9 indicate that QPCS has the best performance in all cases. Specifically, its correct-fitting percentages are more than 80% even at $\rho = 0.95$, the numbers of true positives are close to 4, and the numbers of false positives are small. It is of interest that QPCS performs slightly better for the Laplace error distribution than for the Normal error distribution when $\rho = 0.05$ and 0.5. This finding may be related to the fact that at $\tau = 0.5$ the parameter estimate from quantile regression is the MLE under the Laplace distribution.

Upon an anonymous referee's suggestion, we further compare QPCS with the SCAD penalization (Wang, Wu, and Li 2012) and l_1 when the error is the t -distribution with three degrees of freedom. The tuning parameter for the SCAD method is selected by PBIC as given in the R package "rqPen." The results demonstrate that QPCS outperforms l_1 and SCAD. To save space, we report the simulation results in Case 1 of Example S3 in the online supplementary materials. Inspired by an anonymous referee's comment, we have conducted a simulation experiment using a block diagonal covariance matrix. The detailed descriptions of simulation settings and the results are given in Case 2 of Example S3 in the online supplementary materials. The numerical results also demonstrate the superiority of QPCS in comparing with l_1 and ISIS.

Example 4. In our simulation studies, we assume that the covariance matrix of covariates is exchangeable (i.e., compound symmetry), except for Case 2 of Example 3. Hence, for a given correlation coefficient ρ and the subset S_j , the largest and the smallest eigenvalues of the exchangeable covariance matrix Σ_{S_j} are $(1 - \rho) + \rho|S_j|$ and $1 - \rho$, respectively. In our proposed algorithm, we require that $|S_j| = o(n^{1/2})$. As a result, the maximal eigenvalue of Σ_{S_j} does not satisfy the boundedness condition given in (C2)(ii), when $|S_j|$ is allowed to diverge with the sample size n . Although Condition (C2)(ii) is not satisfied in this scenario, our proposed method shows good performance because $|S_j|$ is often small (much smaller than n) in practice. It is worth noting that this can be an interesting future research subject to relax Condition (C2)(ii) so that the proposed method is applicable for a wide variety of covariance structures.

6. Application

In this section, we apply the proposed methods to gene expression data that was used by Scheetz et al. (2006) for investigating gene regulation in the mammalian eye and identifying genetic variations relevant to human eye disease. The dataset contains gene expression values of 31,042 probe sets on 120 rats. The expression levels of genes are analyzed on a log scale with base 2. The response variable is the expression of gene TRIM32 (probe 1389163_at), which is associated with human hereditary diseases of the retina. The purpose of this study is to analyze how the response variable depends on the expression of other genes. Before applying the screening method, we adopt the preprocessing procedure of Scheetz et al. (2006) to first remove each probe for which the maximum expression among the 120 rats is less than the 25th percentile of the entire sample of expression values, and then remove any probe for which the range of the expression among 120 rats is less than 2. As a result, there are 18,958 probes left in our analysis. Following the approach

of Wang, Wu, and Li (2012) and Lee, Noh, and Park (2014), we subsequently select 3000 genes with the largest variance in expression value, and then select the top 300 gene expression values in a ranking of their (absolute value) correlation with the response variable. For further illustration, we also consider the top 400, 500, and 800 gene expression values. Afterwards, we apply our proposed method to identify relevant genes for the response variable at quantiles $\tau = 0.3, 0.5$, and 0.7 as in Wang, Wu, and Li (2012). Note that Lee, Noh, and Park (2014) considered $\tau = 0.25, 0.5$, and 0.75.

To assess the finite sample performance, we consider 50 random partitions. For each partition, we randomly divide the data into a training dataset with 80 observations and a testing dataset with 40 observations. From the training dataset, we conduct screening and subset selection, and then fit the quantile regression model with the selected predictors. Subsequently, we employ the resulting quantile regression estimators and the testing data to compute the prediction error $\sum_{i=1}^{40} \rho_\tau(Y_i - \hat{Y}_i)$. A smaller value of the prediction error indicates better performance. In simulation studies, we find that SIS does not perform satisfactorily, and EBIC1 and LASSO are inferior to EBIC2. As a result, we only employ three proposed screening procedures, QPCS, QTCS, and QFR, to screen predictors and one selection criterion, EBIC2, for best subset selection. The resulting three methods are denoted QPCS-EBIC2, QTCS-EBIC2, and QFR-EBIC2, respectively.

For $p = 300$, Wang et al.'s (2012) methods revealed that the averaged number of relevant predictors ranges from 9.08 to 21.66 and the averaged prediction errors are between 1.30 and 1.82 across the three quantiles. Recently, Lee et al. (2014) found that their method not only obtains relevant predictor sizes of 2.24, 2.16, and 1.16 when $\tau = 0.25$, $\tau = 0.5$, and $\tau = 0.75$, respectively, but also yields comparable prediction errors of 1.42, 1.64, and 1.30 accordingly. By applying our proposed methods, Table 10 shows that the size selected via QPCS-EBIC2, QTCS-EBIC2, and QFR-EBIC2 ranges from 1.68 to 2.50 across the three quantiles. In addition, the resulting PEs are between 0.502 and 1.016, and they are smaller than those values obtained via the approaches of Wang et al. (2012) and Lee et al. (2014). Among our three proposed methods, QPCS-EBIC2 is superior to QTCS-EBIC2 and QFR-EBIC2 in terms of average prediction error measure, although its sizes are slightly larger than those of QTCS-EBIC2 when $\tau = 0.5$ and 0.7. As p increases to 400, 500, and 800, however, QPCS-EBIC2 has the smallest values in both size and PE across all three quantiles. This finding is consistent with simulation results. In sum, the proposed quantile partial correlation screening algorithm can be considered for quantile regression selection with high-dimensional data.

7. Discussion

In sparse ultrahigh-dimensional quantile regression, we introduce three algorithms, QPCS, QTCS, and QFR, that use quantile correlation and quantile partial correlation to screen explanatory variables. We then employ an extended BIC for model selection. Simulation results demonstrate that the QPCS algorithm supports our theoretical findings. In addition, we find that QPCS performs well in the following settings: (1) highly correlated covariates, (2) ultra-high dimension, (3) covariates

Table 10. The average number of selected variables (Size) and its standard error as shown in the parentheses and the average prediction errors (PE) and its standard error as shown in the parentheses for the three screening methods with EBIC2 at $\tau = 0.3, 0.5, 0.7$ and $p = 300, 400, 500, 800$ among the 50 random partitions.

		$\tau = 0.3$		$\tau = 0.5$		$\tau = 0.7$	
		Size	PE	Size	PE	Size	PE
$p = 300$	QPCS1-EBIC2	1.86(0.130)	0.502(0.053)	1.96(0.201)	0.966(0.095)	1.80(0.178)	0.845(0.091)
	QPCS2-EBIC2	2.38(0.156)	0.545(0.077)	1.92(0.156)	1.016(0.109)	1.68(0.135)	0.891(0.080)
	QFR-EBIC2	2.50(0.207)	0.507(0.053)	2.24(0.136)	0.985(0.113)	2.02(0.175)	0.877(0.108)
$p = 400$	QPCS1-EBIC2	1.64(0.120)	0.552(0.051)	1.98(0.177)	0.839(0.068)	1.70(0.137)	0.715(0.067)
	QPCS2-EBIC2	2.66(0.182)	0.571(0.038)	2.06(0.122)	0.877(0.069)	1.78(0.122)	0.957(0.094)
	QFR-MBIC2	2.94(0.190)	0.584(0.058)	2.52(0.146)	0.985(0.060)	2.52(0.135)	0.799(0.075)
$p = 500$	QPCS1-EBIC2	1.94(0.174)	0.558(0.051)	1.68(0.157)	0.684(0.055)	1.70(0.167)	0.811(0.114)
	QPCS2-EBIC2	2.38(0.140)	0.574(0.047)	2.26(0.158)	1.072(0.087)	1.94(0.174)	1.128(0.090)
	QFR-EBIC2	3.48(0.237)	0.600(0.061)	2.26(0.123)	0.887(0.069)	2.08(0.121)	1.028(0.094)
$p = 800$	QPCS1-EBIC2	1.92(0.219)	0.665(0.081)	1.64(0.145)	0.647(0.053)	1.70(0.194)	0.834(0.089)
	QPCS2-EBIC2	2.80(0.202)	0.670(0.081)	2.16(0.172)	1.020(0.070)	2.40(0.194)	1.099(0.118)
	QFR-EBIC2	3.62(0.269)	0.693(0.082)	2.80(0.162)	0.783(0.050)	2.54(0.132)	0.991(0.018)

being either marginally uncorrelated or weakly correlated with the response, and (4) heavy-tailed errors. Moreover, our simulation results show that it is superior to LASSO, SCAD, SIS, and ISIS-SCAD.

To broaden the usefulness of QPCS, we discuss some extensions for future research in variable screening. There are three possible avenues. The first one is to extend quantile correlation and quantile partial correlation to various quantile regression models such as single-index quantile regression. We have conducted simulation studies by changing the conditional quantile function in Example 1 to its exponential form. The results, not presented here, show that QPCS still performs well under this setting.

The second avenue is considering an alternative correlation measure used in the QPCS algorithm. In simple linear regression, it is known that the square of the coefficient of correlation is the same as the coefficient of determination. In multiple linear regression, Kutner et al. (2005) indicated that the square of the partial correlation is the same as the coefficient of partial determination. In addition, Nagelkerke (1991) proposed a general definition of the coefficient of determination via the log-likelihood function of the response variable. Accordingly, as long as the likelihood function (or its related version, such as partial likelihood or quasi-likelihood) is available for any specific regression model and the maximum likelihood estimators of the regression parameters are also attainable, one can replace the quantile correlation and quantile partial correlation used in the QPCS algorithm by their corresponding coefficient of determination and coefficient of partial determination. This approach can be used for various regression models, for example, generalized linear models, extreme value regression models, and parametric survival models.

Instead of the determination measure, the third avenue for future research considers the residual sum of squares. This is particularly useful for regression models that have no log-likelihood function. In linear regression, it can be easily shown that maximizing correlation is the same as minimizing the residual sum of squares. Analogously, maximizing partial correlation is the same as minimizing the partial residual sum of squares. It is known that the residual sum of squares is the objective function of regression estimators based on the L_2 -norm distance. As a result, the partial residual sum of squares is the difference between the two nested objective functions, which is namely

the partial objective function. In general, the objective function can be a distance metric such as the L_p -norm distance or another robust function. This motivates us to replace the quantile correlation and quantile partial correlation used in the QPCS algorithm by the objective function and partial objective function, respectively. This approach can be used for many regression models such as generalized additive models, semiparametric models, and robust regression models. In sum, the above three avenues can shed light on areas of future research that warrant thorough investigation and study.

Appendix

Before proving the three theorems and one proposition, we present the following five lemmas. Lemma A.1 shows the relationship between $\beta_{j\tau}^0$ and $\beta_{j\tau}^*$ for $j = 1, \dots, p$. Lemmas A.2 and A.3 are used in the proofs of Lemmas A.4 and A.5, while Lemmas A.4 and A.5 are needed in the proof of Theorem 1. The proofs of these five lemmas are given in the online supplementary materials. For the sake of convenience, we denote $\lim_{n \rightarrow \infty} a_n/b_n = c$ with $c > 0$ and $\lim_{n \rightarrow \infty} a_n/b_n = 1$ by $a_n \asymp b_n$ and $a_n \sim b_n$, respectively, for any positive series a_n and b_n . In addition, for any matrix $\mathbf{A} = (A_{ij})_{i=1, j=1}^{s, t}$, denote $|\mathbf{A}| = \max_{1 \leq i \leq s, 1 \leq j \leq t} |A_{ij}|$. Moreover, in the following lemmas, we assume that $0 < \kappa < 1/2$, and $r_n = Cn^\omega$ for some $0 \leq \omega < \min((1 - 2\kappa), 2\kappa)$ and a positive constant C , as stated in Theorem 1.

Lemma A.1. Assume that $\beta_\tau^0 = (\beta_{0\tau}^0, \dots, \beta_{p\tau}^0)^T$ is the unique minimizer of $E[\rho_\tau(Y - \beta_{0\tau} - \beta_{1\tau}X_1 - \dots - \beta_{p\tau}X_p)]$, and, for given $1 \leq j \leq p$, α_j^0 and $(\beta_{0\tau}^*, \beta_{j\tau}^*)^T$ are unique minimizers of $E[\rho_\tau(Y - \mathbf{X}_j^T \alpha_j)]$ and $E[\rho_\tau(Y - \beta_{0\tau} - X_j \beta_{j\tau})]$, respectively, where $Y^* = Y - \mathbf{X}_{-j}^T \alpha_j^0$. Then we have $\beta_{j\tau}^0 = 0$ if and only if $\beta_{j\tau}^* = 0$.

Lemma A.2. Under Condition (C2) and the assumption $n^{-1}\delta_n = O(1)$, for every $1 \leq j \leq p_n$ and $c_1 > 0$, there exist some positive constants c_2 and c_3 such that

$$P(\|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^0\| \geq c_1 r_n n^{-1} \delta_n) \leq 4r_n^2 \exp(-c_2 \delta_n^2 n^{-1}) + 2r_n^2 \exp(-c_3 n r_n^{-2}),$$

when n is sufficiently large.

Lemma A.3. Under Conditions (C1) and (C2), for every $1 \leq j \leq p_n$ and for any given constant $c_4 > 0$, there exists some positive constant c_5 such that

$$P(\|\widehat{\boldsymbol{\pi}}_j - \boldsymbol{\pi}_j^0\| \geq c_4 n^{-\kappa}) \leq 3 \exp(-c_5 r_n^{-1} n^{1-2\kappa}),$$

when n is sufficiently large.

Lemma A.4. Under Conditions (C1) and (C2), for every $1 \leq j \leq p_n$ and for any given constant $c_6 > 0$, there exist some positive constants c_7 and c_8 such that

$$\begin{aligned}
 &P \left[\left| n^{-1} \sum_{i=1}^n \{ \psi_\tau(Y_i - \mathbf{X}_{i,S_j}^T \hat{\boldsymbol{\pi}}_j)(X_{ij} - \mathbf{X}_{i,S_j}^T \hat{\boldsymbol{\theta}}_j) \right. \right. \\
 &\quad \left. \left. - E\{ \psi_\tau(Y - \mathbf{X}_{S_j}^T \boldsymbol{\pi}_j^0) X_j \} \right| \geq c_6 r_n^{1/2} n^{-\kappa} \right] \\
 &\leq 7 \exp(-c_7 r_n^{-1} n^{1-2\kappa}) + 6 r_n^2 \exp(-c_8 n r_n^{-2}),
 \end{aligned}
 \tag{A.7}$$

when n is sufficiently large.

Lemma A.5. Under Condition (C2) and the assumption of r_n in [Theorem 1](#), for every $1 \leq j \leq p_n$ and for any $c_9 > 0$, there exist some positive constants c_{10} and c_{11} such that

$$\begin{aligned}
 &P \left(|\hat{\sigma}_j^2 - \sigma_j^2| \geq c_9 r_n^{1/2} n^{-\kappa} \right) \leq (2 + 8 r_n^2) \exp(-c_{10} r_n^{-1} n^{1-2\kappa}) \\
 &\quad + (2 + 6 r_n^2) \exp(-c_{11} n r_n^{-2}),
 \end{aligned}
 \tag{A.1}$$

when n is sufficiently large. Note that $\hat{\sigma}_j^2$ and σ_j^2 have been defined after [Equations \(6\) and \(8\)](#), respectively. Moreover, for $a \in (0, 1)$,

$$\begin{aligned}
 &P \left(|\hat{\sigma}_j^2 - \sigma_j^2| \geq a \sigma_j^2 \right) \leq (2 + 8 r_n^2) \exp(-c_{10} r_n^{-1} n^{1-2\kappa}) \\
 &\quad + (2 + 6 r_n^2) \exp(-c_{11} n r_n^{-2}),
 \end{aligned}
 \tag{A.2}$$

when n is sufficiently large.

Proof of Theorem 1. Denote $\phi_{jn} = n^{-1} \sum_{i=1}^n \{ \psi_\tau(Y_i - \mathbf{X}_{i,S_j}^T \hat{\boldsymbol{\pi}}_j)(X_{ij} - \mathbf{X}_{i,S_j}^T \hat{\boldsymbol{\theta}}_j) \}$ and $\phi_j = E\{ \psi_\tau(Y - \mathbf{X}_{S_j}^T \boldsymbol{\pi}_j^0) X_j \}$. Then

$$\begin{aligned}
 &|\widehat{\text{qpcor}}_\tau\{Y, X_j, \mathbf{X}_{S_j}\} - \text{qpcor}_\tau\{Y, X_j, \mathbf{X}_{S_j}\}| \\
 &= |(\hat{\sigma}_j^2)^{-1}(\phi_{jn} - \phi_j) - (\sigma_j^2)^{-1}\phi_j(\hat{\sigma}_j^2 - \sigma_j^2)| \\
 &\leq (\hat{\sigma}_j^2)^{-1}|\phi_{jn} - \phi_j| + (\hat{\sigma}_j^2)^{-1}(\sigma_j^2)^{-1}|\phi_j||\hat{\sigma}_j^2 - \sigma_j^2|.
 \end{aligned}
 \tag{A.3}$$

After algebraic simplification, we have that, for any $a \in (0, 1)$, $|(\hat{\sigma}_j^2)^{-1} - (\sigma_j^2)^{-1}| \geq a^*(\sigma_j^2)^{-1}$ implies $|\hat{\sigma}_j^2 - \sigma_j^2| \geq a\sigma_j^2$, where $a^* = (1 - a)^{-1} - 1$. Hence, by [\(A.2\)](#) in [Lemma A.5](#) and $(\sigma_j^2)^{-1} \leq c_\sigma^{-1}$, we obtain

$$\begin{aligned}
 &P \left((\hat{\sigma}_j^2)^{-1} \geq (1 + a^*)c_\sigma^{-1} \right) \\
 &\leq P \left(|(\hat{\sigma}_j^2)^{-1} - (\sigma_j^2)^{-1}| \geq a^*(\sigma_j^2)^{-1} \right) \\
 &\leq P \left(|\hat{\sigma}_j^2 - \sigma_j^2| \geq a\sigma_j^2 \right) \\
 &\leq (2 + 8 r_n^2) \exp(-c_{10} r_n^{-1} n^{1-2\kappa}) + (2 + 6 r_n^2) \exp(-c_{11} n r_n^{-2}).
 \end{aligned}
 \tag{A.4}$$

This, in conjunction with [Lemma A.4](#), implies that for any $c_6 > 0$,

$$\begin{aligned}
 &P \left((\hat{\sigma}_j^2)^{-1} |\phi_{jn} - \phi_j| \geq (1 + a^*)c_\sigma^{-1} c_6 r_n^{1/2} n^{-\kappa} \right) \\
 &\leq (9 + 8 r_n^2) \exp(-c_1^{***} r_n^{-1} n^{1-2\kappa}) + (2 + 12 r_n^2) \exp(-c_1^{**} n r_n^{-2}),
 \end{aligned}
 \tag{A.5}$$

for some positive constants $0 < c_1^{***}, c_1^{**} < \infty$.

It is worth noting that $|\phi_j| \leq M_1$. Then, employing [\(A.1\)](#) in [Lemma A.5](#) and [\(A.4\)](#), we have that for any $c_9 > 0$,

$$\begin{aligned}
 &P \left((\hat{\sigma}_j^2)^{-1} (\sigma_j^2)^{-1} |\phi_j| |\hat{\sigma}_j^2 - \sigma_j^2| \geq (1 + a^*)c_\sigma^{-1} c_\sigma^{-1} M_1 c_9 r_n^{1/2} n^{-\kappa} \right) \\
 &\leq (4 + 16 r_n^2) \exp(-c_{10} r_n^{-1} n^{1-2\kappa}) + (4 + 12 r_n^2) \exp(-c_{11} n r_n^{-2}).
 \end{aligned}
 \tag{A.6}$$

By [\(A.3\)](#), [\(A.5\)](#), and [\(A.6\)](#), we have that, for any $C_1 > 0$, there exist some positive constants C_2 and C_3 such that

$$\begin{aligned}
 &P \left(|\widehat{\text{qpcor}}_\tau\{Y, X_j, \mathbf{X}_{S_j}\} - \text{qpcor}_\tau\{Y, X_j, \mathbf{X}_{S_j}\}| \geq C_1 r_n^{1/2} n^{-\kappa} \right) \\
 &\leq (13 + 24 r_n^2) \exp(-C_2 r_n^{-1} n^{1-2\kappa}) + (6 + 24 r_n^2) \exp(-C_3 n r_n^{-2}) \\
 &\leq C_2^* r_n^2 \exp(-C_2 r_n^{-1} n^{1-2\kappa}) + C_3^* r_n^2 \exp(-C_3 n r_n^{-2}),
 \end{aligned}
 \tag{A.7}$$

for some positive constants C_2^* and C_3^* . The last inequality follows from $13 + 24 r_n^2 \leq 37 r_n^2$ and $6 + 24 r_n^2 \leq 30 r_n^2$ with $C_2^* = 37$ and $C_3^* = 30$. This completes the proof. \square

Proof of Theorem 2. On the event

$$\begin{aligned}
 A_n &= \{ \sup_{j \in \mathcal{M}_*} |\widehat{\text{qpcor}}_\tau\{Y, X_j, \mathbf{X}_{S_j}\} - \text{qpcor}_\tau\{Y, X_j, \mathbf{X}_{S_j}\}| \\
 &\leq C_0 r_n^{1/2} n^{-\kappa} / 2 \},
 \end{aligned}$$

we apply [Condition \(C3\)](#) and obtain $|\widehat{\text{qpcor}}_\tau\{Y, X_j, \mathbf{X}_{S_j}\}| \geq C_0 r_n^{1/2} n^{-\kappa} / 2$. Hence, by the choice of $v_n = C_4 r_n^{1/2} n^{-\kappa}$ with $C_4 \leq C_0 / 2$, we have $\mathcal{M}_* \subset \bar{\mathcal{M}}_{v_n}$ on the event A_n . This, together with [\(A.7\)](#) and the union bound of probability, yields that

$$P(A_n^c) \leq s_n \{ C_2^* r_n^2 \exp(-C_2 r_n^{-1} n^{1-2\kappa}) + C_3^* r_n^2 \exp(-C_3 n r_n^{-2}) \},$$

which completes the proof. \square

Proof of Proposition 1. Employing [Equation \(10\)](#) by letting $2c = C_4$, we have that, on the event Ω_n , $|\bar{\mathcal{M}}_{v_n}| \leq C^* n^{\varsigma + \kappa - \omega/2}$, where $0 < C^* < \infty$. This, in conjunction with [Theorem 1](#), leads to

$$P(\Omega_n) \geq 1 - p_n \{ C_2^* r_n^2 \exp(-C_2 r_n^{-1} n^{1-2\kappa}) + C_3^* r_n^2 \exp(-C_3 n r_n^{-2}) \}.$$

Accordingly, [Proposition 1](#) follows. \square

Proof of Theorem 3. Define $k_{\min} = \min_{1 \leq k \leq d} \{ k : \mathcal{M}^* \subset \mathcal{A}^{(k)} \}$. By the assumption that $\varsigma + \kappa - \omega/2 < 1/2$ and the result in [Proposition 1](#), k_{\min} is well defined and satisfies $k_{\min} \leq C' n^{\varsigma + \kappa - \omega/2} = o(n^{1/2})$ for some constant $0 < C' < \infty$. For any $1 \leq k < k_{\min}$, $\mathcal{A}^{(k)}$ are underfitted models such that $\mathcal{M}^* \not\subset \mathcal{A}^{(k)}$ and $\mathcal{A}^{(k)}$ are nested. By [\(A.18\)](#) in the online supplementary materials of [Lee et al. \(2014\)](#), with probability approaching 1, we can choose a sequence of constants $\{L_n\}$ such that $L_n \rightarrow \infty$, $L_n/C_n \rightarrow 0$, and

$$\begin{aligned}
 &\left| n^{-1} \sum_{i=1}^n \{ \rho_\tau(Y_i - \mathbf{X}_{i, \mathcal{A}^{(k_{\min})}}^T \hat{\boldsymbol{\beta}}_{\mathcal{A}^{(k_{\min})}}) - \rho_\tau(\varepsilon_i) \} \right| \\
 &= \left| n^{-1} \sum_{i=1}^n \{ \rho_\tau(\varepsilon_i - \mathbf{X}_{i, \mathcal{A}^{(k_{\min})}}^T (\hat{\boldsymbol{\beta}}_{\mathcal{A}^{(k_{\min})}} - \boldsymbol{\beta}_{\mathcal{A}^{(k_{\min})}}^0)) - \rho_\tau(\varepsilon_i) \} \right| \\
 &\leq n^{-1} L_n k_{\min} \log(d) \leq C'' n^{-1} L_n n^{\varsigma + \kappa - \omega/2} \log(n)
 \end{aligned}$$

for some constant $0 < C'' < \infty$, where $\varepsilon_i = Y_i - \mathbf{X}_{i, \mathcal{A}^{(k_{\min})}}^T \boldsymbol{\beta}_{\mathcal{A}^{(k_{\min})}}^0$ and $\boldsymbol{\beta}_{\mathcal{A}^{(k_{\min})}}^0 = \{ \beta_{kt}^0 : k \in \mathcal{A}^{(k_{\min})} \}^T$. Under the assumption that $E|\varepsilon| < \infty$, we obtain that $n^{-1} \sum_{i=1}^n \rho_\tau(\varepsilon_i) \rightarrow_p E\rho_\tau(\varepsilon)$ and $c' \leq E\rho_\tau(\varepsilon) \leq c''$ for some constants $0 < c', c'' < \infty$. In addition, by assuming that $(n^{\varsigma + \kappa - \omega/2} n^{-1} \log n) C_n = o(1)$, we have $n^{-1} L_n n^{\varsigma + \kappa - \omega/2} \log(n) = o(1)$. The above results imply that, with probability approaching 1,

$$\begin{aligned}
 c' + o(1) &\leq n^{-1} \sum_{i=1}^n \{ \rho_\tau(\varepsilon_i - \mathbf{X}_{i, \mathcal{A}^{(k_{\min})}}^T (\hat{\boldsymbol{\beta}}_{\mathcal{A}^{(k_{\min})}} - \boldsymbol{\beta}_{\mathcal{A}^{(k_{\min})}}^0)) \} \\
 &\leq c'' + o(1).
 \end{aligned}
 \tag{A.8}$$

Moreover, by employing the same techniques as those used in the proof of [\(A.20\)](#) from the online supplementary materials of [Lee et al. \(2014\)](#), we have, with probability approaching 1,

$$n^{-1} \sum_{i=1}^n \rho_\tau((Y_i - \mathbf{X}_{i, \mathcal{A}^{(k)}}^T \hat{\boldsymbol{\beta}}_{\mathcal{A}^{(k)}}) - n^{-1} \sum_{i=1}^n$$

$$\times \rho_{\tau}(Y_i - \mathbf{X}_{i, \mathcal{A}^{(k_{\min})}}^T \widehat{\beta}_{\mathcal{A}^{(k_{\min})}}) \geq 2c''' \quad (\text{A.9})$$

for any $1 \leq k < k_{\min}$, for some constant $0 < c''' < \infty$. Then, we have, with probability approaching 1, as $n \rightarrow \infty$,

$$\begin{aligned} & \min_{1 \leq k < k_{\min}} \text{EBIC}(\mathcal{A}^{(k)}) - \text{EBIC}(\mathcal{A}^{(k_{\min})}) \\ &= \min_{1 \leq k < k_{\min}} \left\{ \log(1 \right. \\ & \quad \left. + \frac{n^{-1} \sum_{i=1}^n \rho_{\tau}((Y_i - \mathbf{X}_{i, \mathcal{A}^{(k)}}^T \widehat{\beta}_{\mathcal{A}^{(k)}}) - n^{-1} \sum_{i=1}^n \rho_{\tau}(Y_i - \mathbf{X}_{i, \mathcal{A}^{(k_{\min})}}^T \widehat{\beta}_{\mathcal{A}^{(k_{\min})}}))}{n^{-1} \sum_{i=1}^n \rho_{\tau}(Y_i - \mathbf{X}_{i, \mathcal{A}^{(k_{\min})}}^T \widehat{\beta}_{\mathcal{A}^{(k_{\min})}})} \right. \\ & \quad \left. + (k - k_{\min}) \frac{\log n}{2n} C_n \right\} \\ & \geq \min \left\{ \log 2, \frac{c'''}{n^{-1} \sum_{i=1}^n \rho_{\tau}(Y_i - \mathbf{X}_{i, \mathcal{A}^{(k_{\min})}}^T \widehat{\beta}_{\mathcal{A}^{(k_{\min})}})} \right\} - C' n^{\kappa + \omega/2} \frac{\log n}{2n} C_n \\ & \geq \min \left\{ \log 2, \frac{c'''}{c'' + o(1)} \right\} - C' n^{\kappa + \omega/2} \frac{\log n}{2n} C_n > 0, \end{aligned}$$

where the first inequality follows from the fact that $\log(1+x) \geq \min\{x/2, \log 2\}$ for any $x > 0$, (A.9) and $k_{\min} \leq C' n^{\kappa + \omega/2}$, and the second inequality follows from (A.8) and the assumption that $(n^{\kappa + \omega/2} n^{-1} \log n) C_n = o(1)$. The above result implies that $P(\widehat{k} \geq k_{\min}) \rightarrow 1$, as $n \rightarrow \infty$, which completes the proof. \square

Supplementary Materials

In the supplementary materials, we provide the proofs for Lemmas A.1–A.5, and present additional simulation results for Examples 1–3.

Acknowledgments

We are grateful to the editor, the associate editor, and three anonymous reviewers for their constructive comments that helped us improve the article substantially.

Funding

Ma's research was supported by NSF grant DMS 1306972 and Hellman Fellowship. Li's research was supported by NSF grant DMS 1512422 and NIDA, NIH grants P50 DA036107 and P50 DA039838. The content is solely the responsibility of the authors and does not necessarily represent the official views of NSF, NIDA, or NIH.

References

- Angrist, J., Chernozhukov, V., and Fernández-Val, I. (2006), "Quantile Regression Under Misspecification, With an Application to the U.S. Wage Structure," *Econometrica*, 74, 539–563. [651]
- Belloni, A., and Chernozhukov, V. (2011), " ℓ_1 -Penalized Quantile Regression in High-Dimensional Sparse Model," *Annals of Statistics*, 39, 759–771. [656]
- Bühlmann, P., Kalisch, M., and Maathuis, M. (2009), "Variable Selection for High-Dimensional Models: Partially Faithful Distributions and the PC-Simple Algorithm," *Biometrika*, 97, 1–19. [650]
- Chen, J., and Chen, Z. (2008), "Extended Bayesian Information Criteria for Model Selection With Large Model Spaces," *Biometrika*, 95, 759–771. [651,653,655]
- Cho, H., and Fryzlewicz, P. (2012), "High Dimensional Variable Selection via Tilting," *Journal of the Royal Statistical Society, Series B*, 74, 593–622. [650,651,654,655,656,657]
- Fan, J., Feng, Y., and Song, R. (2011), "Nonparametric Independence Screening in Sparse Ultra-High Dimensional Additive Models," *Journal of the American Statistical Association*, 106, 544–557. [650,651]

- Fan, J., and Li, R. (2001), "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties," *Journal of the American Statistical Association*, 96, 1348–1360. [650]
- Fan, J., and Lv, J. (2008), "Sure Independence Screening for Ultra-High Dimensional Feature Space" (with discussion), *Journal of the Royal Statistical Society, Series B*, 70, 849–911. [650,651,653,654,655,656,657]
- (2010), "A Selective Overview of Variable Selection in High Dimensional Feature Space," *Statistica Sinica*, 20, 101–148. [650]
- (2010), "Sure Independence Screening in Generalized Linear Models with NP-Dimensionality," *The Annals of Statistics*, 38, 3567–3604. [650]
- He, X., Wang, L., and Hong, H. G. (2013), "Quantile-Adaptive Model-Free Variable Screening for High-Dimensional Heterogeneous Data," *Annals of Statistics*, 41, 342–369. [650,651,653,655]
- Huang, J., Horowitz, J. L., and Ma, S. (2008b), "Asymptotic Properties of Bridge Estimators in Sparse High-Dimensional Regression Models," *Annals of Statistics*, 36, 587–613. [650]
- Huang, J., Ma, S. G., and Zhang, C. H. (2008a), "Adaptive Lasso for Sparse High-Dimensional Regression Models," *Statistica Sinica*, 18, 1603–1618. [650]
- Kalisch, M., and Bühlmann, P. (2007), "Estimating High-Dimensional Directed Acyclic Graphs With the PC-Algorithm," *Journal of Machine Learning Research*, 8, 613–636. [654]
- Koenker, R. (2005), *Quantile Regression*, New York: Cambridge University Press. [651]
- Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2005), *Applied Linear Statistical Models* (5th ed.), New York: McGraw-Hill and Irwin. [660]
- Lee, E. R., Noh, H., and Park, B. U. (2014), "Model Selection via Bayesian Information Criterion for Quantile Regression Models," *Journal of the American Statistical Association*, 109, 216–229. [651,652,655,659,661]
- Li, G., Li, Y., and Tsai, C. L. (2015), "Quantile Correlations and Quantile Autoregressive Modeling," *Journal of the American Statistical Association*, 110, 246–261. [651,652]
- Li, R., Zhong, W., and Zhu, L. (2012), "Feature Screening via Distance Correlation Learning," *Journal of American Statistical Association*, 107, 1129–1139. [652]
- Liu, J., Li, R., and Wu, R. (2014), "Feature Selection for Varying Coefficient Models With Ultrahigh Dimensional Covariates," *Journal of the American Statistical Association*, 109, 266–274. [650,656]
- Nagelkerke, N. J. D. (1991), "A Note on a General Definition of the Coefficient of Determination," *Biometrika*, 78, 691–692. [660]
- Scheetz, T. E., Kim, K.-Y. A., Swiderski, R. E., Philp, A. R., Braun, T. A., Knudtson, K. L., Dorrance, A. M., DiBona, G. F., Huang, J., Casavant, T. L., Sheffield, V. C., and Stone, E. M. (2006), "Regulation of Gene Expression in the Mammalian Eye and Its Relevance to Eye Disease," *Proceedings of the National Academy of Sciences of the United States of America*, 103, 14429–14434. [659]
- Schwarz, C. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464. [655]
- Tibshirani, R. (1996), "Regression Shrinkage and Selection via the LASSO," *Journal of the Royal Statistical Society, Series B*, 58, 267–288. [650]
- Wang, H. (2009), "Forward Regression for Ultra-High Dimensional Variable Screening," *Journal of the American Statistical Association*, 104, 1512–1524. [650,651,654,655]
- Wang, H., Li, B., and Leng, C. (2009), "Shrinkage Tuning Parameter Selection With a Diverging Number of Parameters," *Journal of Royal Statistical Society, Series B*, 71, 671–683. [651,655]
- Wang, H., Li, R., and Tsai, C.-L. (2007), "Tuning Parameter Selectors for the Smoothly Clipped Absolute Deviation Method," *Biometrika*, 94, 553–568. [656]
- Wang, L., Wu, Y., and Li, R. (2012), "Quantile Regression for Analyzing Heterogeneity in Ultra-High Dimension," *Journal of the American Statistical Association*, 107, 214–222. [651,652,659]
- Wu, Y., and Yin, G. (2015), "Conditional Quantile Screening in Ultrahigh-Dimensional Heterogeneous Data," *Biometrika*, 102, 65–76. [650]

- Zhang, C. H. (2010), "Nearly Unbiased Variable Selection Under Minimax Concave Penalty," *The Annals of Statistics*, 38, 894–942. [650]
- Zhu, L.-P., Li, L., Li, R., and Zhu, L.-X. (2011), "Model-Free Feature Screening for Ultrahigh-Dimensional Data," *Journal of the American Statistical Association*, 106, 1464–1475. [650,652]
- Zou, H. (2006), "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, 101, 1418–1429. [650]
- Zou, H., and Hastie, T. (2005), "Regression Shrinkage and Selection via the Elastic Net With Application to Microarrays," *Journal of the Royal Statistical Society, Series B*, 67, 301–320. [650]