Quantile correlation-based variable selection

Journal:	*Journal of Business & Economic Statistics*
Manuscript ID	Draft
Manuscript Type:	Paper
Keywords:	Quantile correlation, Variable selection, High dimensionality, False discovery rate

Quantile correlation-based variable selection

Abstract: This paper is concerned with identifying important features in high dimensional data analysis, especially when there are complex relationships among predictors. Without any specification of an actual model, we first introduce a multiple testing procedure based on the quantile correlation to select important predictors in high dimensionality. The quantile-correlation statistic is able to capture a wide range of dependence. A stepwise procedure is studied for further identifying important variables. Moreover, a sure independent screening based on the quantile correlation is developed in handling ultrahigh dimensional data. It is computationally efficient and easy to implement. We establish the theoretical properties under mild conditions. Numerical studies including simulation studies and real data analysis contain supporting evidence that the proposal performs reasonably well in practical settings.

Keywords: Quantile correlation; Variable selection; High dimensionality; False discovery rate.

1 Introduction

Many contemporary data arise in various scientific fields, such as finance, genomics, neuroimaging and social network, are of high dimensionality. In such type of data, the number of predictors can be much larger than the sample size, which poses unprecedent challenges for statistical analysis and numerical computation (Fan *et al.*, 2009). State-of-the-art statistical methodologies have been proposed for simultaneously selecting important variables and estimating the unknown parameters for various high dimensional regression models, including Lasso (Tibshirani, 1996), SCAD (Fan and Li, 2001), group Lasso (Yuan and Lin, 2006), adaptive Lasso (Zou, 2006), MCP (Zhang, 2010) and their variants. The optimization problems associated with the penalization methods can be solved effectively with moderate dimensionality. When the number of the predictors grows exponentially fast with the sample size, penalized methods encounter computational complication. Variable screening methods are particularly designed and examined to be effective to reduce the high dimensionality to a moderate scale, so that classical statistical inference methods can be applied to the reduced models. In particular, Fan and Lv (2008) proposed the sure independence screening (SIS) for linear regression based on marginal Pearson's correlation coefficient. Fan and Song (2010), Fan *et al.* (2011) and Ma *et al.* (2017) extended feature screening methods to generalized linear model, the additive model and the quantile linear model with heavy-tailed data. These approaches are model-based methods.

It is known that the relationship between predictors and the response is hard to specify for ultra-high dimensional data in many scientific applications. This poses a great deal of challenges to identify important effects, especially the important effects appear in some complex form. There has been a growing literature on model-free screening methods recently. Zhu *et al.* (2011) studied a sure independence ranking and screening procedure to detect important predictors. Li *et al.* (2012) proposed to use the Kendall's rank correlation as a robust ranking utility. A novel sure screening procedure based on the distance correlation was introduced by Li *et al.* (2012). He *et al.* (2013) developed a quantile-adaptive model-free variable screening for high dimensional heterogeneous data. The Kolmogorov-Smirnov distance was introduced by Mai and Zou (2013) to handle binary classification problems and was extended to deal with continuous response in Mai and Zou (2015). Cui *et al.* (2015)

advocated a marginal feature screening procedure for ultra-high dimensional discriminant analysis with a possibly diverging number of classes. Zhou *et al.* (2019) proposed a forward screening procedure based on a new metric named cumulative divergence.

Overall, there are three kinds of variable selection results reported in the literature, depending on the signal-to-noise level (Hao and Zhang, 2017). When the signal is strong enough, one can achieve model selection consistency under certain conditions on the design matrix (Zhao and Yu, 2006; Zhang, 2010; Fan and Lv, 2011; among many others). When the signal is somewhat strong but there are a huge number of noise variables, the sure screening property or screening consistency, a weaker result of the model selection consistency, can be established (Fan and Lv, 2008; Huang, Li and Wang, 2014; Chu *et al.*, 2016; etc). Nevertheless, for datasets arised in many modern scientific studies, some signals could be weak or appear only in the interaction terms. It would be generally difficult to distinguish them from noise variables. For such a case, one may consider to allow some false discoveries. In this regard, a practical approach is to conduct multiple testing so that the false discovery rate (FDR) can be controlled.

For illustration, let Y be the response and $\boldsymbol{X} = (X_1, \ldots, X_p)^\top$ be a p-dimensional vector of covariates. One may consider to simultaneously test

$$\mathbb{H}_{0,k} : X_k \perp\!\!\!\perp Y \quad \text{versus} \quad \mathbb{H}_{1,k} : X_k \not\!\perp\!\!\!\perp Y$$

for $1 \leq k \leq p$, where $\perp\!\!\!\perp$ represents independence. Thus, X_k is regarded as an influential predictor if and only if $\mathbb{H}_{0,k}$ is rejected. For multiple testing, the false discovery rate (FDR) is defined to be the expected proportion of false positives among all rejections. When the test statistics are independent, the multiple-testing procedure by Benjamini and Hochberg (1995) reported elegant results. Other novel advancements can be found in, for example, Storey (2002), Efron (2004), Sun and Cai (2007), among many others. The multiple testing problems are increasingly complicated when the test statistics are not independent; see Benjamini and Yekutieli (2001), Storey *et al.* (2004), Efron (2007), Sun and Cai (2009) and Cai and Liu (2016), Xie and Li (2019).

In this paper, without any parametric assumption or specification of a regression model, we propose a robust multiple testing procedure with false discovery rate control to detect influential predictors. A stepwise procedure is developed to further control the empirical false discovery rate. In contrast to existing variable screening methods, the proposed method has several distinctive features: the proposed procedure is based on a quantile correlation-based test statistic, which is shown to be asymptotically chi-square distribution under the null hypothesis; with the availability of the asymptotic distribution, we prove rigorously that the FDR control for high dimensionality is valid in theory; the stepwise procedure is computationally efficient in identifying important variables, as no optimization or resampling is involved; it is able to detect relevant variables with linear or nonlinear effects; last but not least, the proposed procedure is model-free in the sense that its validity does not rely on any specific assumption on the functional relationship between the predictors and the responses.

The rest part of this paper is organized as follows. Section 2 contains a detailed description of the proposed procedure to detect influential predictors. In section 3, we investigate its theoretical properties of the proposed procedure under some regularity conditions. In section 4, we evaluate the performance of the proposed procedure via extensive simulation studies. Section 5 reports a real data example. Technical proofs are given in the Appendix.

2 Methodologies and main results

2.1 A quantile correlation-based test statistic

Let $(Y, \boldsymbol{X}^\top)$ be a pair of scalar response and p-dimensional vector of covariates. In this paper, we consider continuous response variable Y. Write $\boldsymbol{X} = (X_1, \ldots, X_p)^\top$. The observations $\{(Y_i, \boldsymbol{X}_i^\top), i = 1, \ldots, n\}$ are independent and identical copies of $(Y, \boldsymbol{X}^\top)$. two sets of indices are defined, i.e, $\mathcal{S} = \{k : X_k \not\perp\!\!\!\perp Y, 1 \leq k \leq p\}$ consists of predictors that are relevant to Y and $\mathcal{S}^c$ contains all redundant predictors or noise predictors. Clearly, $\mathcal{S} \cup \mathcal{S}^c = \{1, 2, \ldots, p\}$. Those predictors in $\mathcal{S}$ are also called active or important predictors. Without specific assumption on the functional relationship between the predictors and the response, we consider the problem of detecting important predictors in high dimensional setting. To achieve this, we consider to simultaneously test the marginal independence between each X_k and Y for $1 \leq k \leq p$:

$$\mathbb{H}_{0,k} : X_k \perp\!\!\!\perp Y \quad \text{versus} \quad \mathbb{H}_{1,k} : X_k \not\perp\!\!\!\perp Y. \tag{2.1}$$

Recall that the classical Pearson's Chi-square test is for testing the independence of two categorical variables. Motivated by the Pearson's Chi-square test, we introduce a test statistic based on quantile correlation for testing (2.1) with high dimensionality without assuming any actual regression model between Y and $\boldsymbol{X}$. Let $0 = \gamma_0 < \gamma_1 < \gamma_2 < \ldots < \gamma_{D_1} = 1$ and $0 = \rho_0 < \rho_1 < \rho_2 < \ldots < \rho_{D_2} = 1$ be two sequences of quantile grid points, where D_1 and D_2 are pre-specified positive integers. Denote the γ_s-th ($1 \leq s \leq D_1$) quantile of X_k as $Q_{k,s}$ and the ρ_t-th ($1 \leq t \leq D_2$) quantile of Y as Q_t^*. It is known that theoretical quantiles can be estimated consistently by the respective sample quantiles. Following the definition of sample quantile in Rob and Fan (1996), $\hat{Q}_{k,s} = (1-\gamma_s)X_{k,j} + \gamma_s X_{k,j+1}$ is the γ_s-th ($1 \leq s \leq D_1 - 1$) sample quantile of X_k, where $j = \lfloor n\gamma_s \rfloor$ and $\lfloor x \rfloor$ is the integer part of x. Similarly, $\hat{Q}_t^* = (1-\rho_t)Y_j + \rho_t Y_{j+1}$ is the ρ_t-th ($1 \leq t \leq D_2 - 1$) sample quantile of Y with $j = \lfloor n\rho_t \rfloor$. For convenience, let $\hat{Q}_{k,0} = \hat{Q}_0^* = -\infty$ and $\hat{Q}_{k,D_1} = \hat{Q}_{D_2}^* = +\infty$.

For each $k = 1, \ldots, p$, in order to capture the relationship of Y and X_k, we first consider to construct a contingency table with D_1 rows and D_2 columns, in which the (s,t)-th cell is

$$T_{k,st} = \sum_{i=1}^{n} I\{\hat{Q}_{k,s-1} < X_{k,i} \leq \hat{Q}_{k,s}\} I\{\hat{Q}_{t-1}^* < Y_i \leq \hat{Q}_t^*\}. \tag{2.2}$$

Obviously, under $\mathbb{H}_{0,k}$,

$$E(T_{k,st}) = E_{st} = n\nu_s\nu_t,$$

where $\nu_s = \gamma_s - \gamma_{s-1}$ and $\nu_t = \rho_t - \rho_{t-1}$ are the difference between two consecutive quantile levels. Thus, for each $k = 1, \ldots, p$, we propose the following quantile correlation-based (QC) statistic for testing (2.1):

$$\tau_k = \sum_{s=1}^{D_1} \sum_{t=1}^{D_2} \frac{(T_{k,st} - E_{st})^2}{E_{st}}. \tag{2.3}$$

Since the indicator function in $T_{k,st}$ is bounded, the proposed statistic τ_k is robust to outliers of either Y or X_k. It can be shown later that under $\mathbb{H}_{0,k}$, τ_k converges to $\chi^2\{(D_1-1)(D_2-1)\}$ distribution asymptotically under mild conditions. With the availability of the asymptotic distribution, we can carry out a multiple testing procedure to test $\mathbb{H}_{0,k}$ for $k = 1, \ldots, p$ simultaneously. For brevity, the proposed statistic is referred as QC.

Remark 1. Similar to that of Huang *et al.* (2014), the proposed test statistic τ_k in (2.3) is a Chi-square type statistic. Nevertheless, there is essential difference in the sense that, the Pearson Chi-square ranking statistic in Huang *et al.* (2014) deals with high dimensional categorical data and hence the cell boundaries of the contingency table is fixed. In contrast, we deal with continuous response in this paper, and the cell boundaries of our proposed τ_k are data-driven and may vary across samples.

Remark 2. The definition of $T_{k,st}$ in (2.2) is for continuous or ordinal X_k. We can extend this definition to accommodate categorical X_k. For instance, when some X_k is categorical, one may consider a modified version of $T_{k,st}$:

$$\widetilde{T}_{k,mt} = \sum_{i=1}^{n} I\{X_{k,i} = m\} I\{\hat{Q}^*_{t-1} < Y_i \le \hat{Q}^*_t\},$$

where $m \in \mathcal{M}$ and $\mathcal{M}$ is the collection of the class labels of X_k. Thus, the proposed statistic τ_k for categorical X_k is

$$\tau_k = \sum_{m\in M} \sum_{t=1}^{D_2} \frac{(\widetilde{T}_{k,mt} - \pi_m v_t)^2}{\pi_m v_t}$$

where $\pi_m = \sum_{j\in M} I(X_j = m)/|\mathcal{M}|$ and $|\mathcal{M}|$ is the cardinality of the set $\mathcal{M}$.

Remark 3. Different from the Pearson Chi-square test statistic, we allow the number of quantile grid points D_1 and D_2 to diverge with the sample size n and the dimensionality p in theory. Heuristically, with certain proper choice of D_1 and D_2, the proposed test statistic τ_k is able to capture some delicate association of X_k and Y that only exists in some small cells.

2.2 Detecting active predictors with false discovery rate control

Without assuming an actual regression model, to effectively detect active predictors, we next introduce a false discovery rate (FDR) control procedure for simultaneously test $\mathbb{H}_{0,k}$ for $k = 1, \ldots, p$. First, we define the null index set $\mathbb{H}_0 = \{k : 1 \le k \le p,\ \mathbb{H}_{0,k} \text{ is true}\}$ and the full set $\mathbb{H} = \{k : 1 \le k \le p\}$. In fact, the null set $\mathbb{H}_0$ is the $\mathcal{S}^c$ defined in section 2.1. With the proposed QC test statistic in section 2.1, the false discovery proportion is

$$\mathrm{FDP}_t = \frac{\sum_{k\in\mathbb{H}_0} I(\tau_k \ge t)}{\max\{\sum_{k\in\mathbb{H}} I(\tau_k \ge t), 1\}},$$

and the false discovery rate is $\mathrm{FDR}_t = E(\mathrm{FDP}_t)$ for any given t. By Theorem 1 in section 3, under $\mathbb{H}_{0,k}$, each τ_k converges to $\chi^2\{(D_1 - 1)(D_2 - 1)\}$ in distribution asymptotically under certain regularity conditions. Let q be the cardinality of $\mathbb{H}_0$. Intuitively, for any given t, under assumption that $q/p \to 1$ as $p \to \infty$, one can estimate the FDR_t by

$$\frac{\sum_{k\in\mathbb{H}_0} I(\tau_k \ge t)/q}{\max\{\sum_{k\in\mathbb{H}} I(\tau_k \ge t), 1\}/p}.$$

Nevertheless, the null set $\mathbb{H}_0$ and q are unknown in practical situations. To circumvent the problem, we propose to estimate the FDR_t by replacing $\sum_{k\in\mathbb{H}_0} I(\tau_k \ge t)/q$ by $S_{D_1\times D_2}(t)$, where $S_{D_1\times D_2}(t)$ is the survival function of the Chi-square distribution with $(D_1-1)(D_2-1)$ degrees of freedom. Hence, for any given t, we define the estimated FDR_t as

$$\widehat{\mathrm{FDR}}_t = \frac{pS_{D_1\times D_2}(t)}{\max\{\sum_{k\in\mathbb{H}} I(\tau_k \ge t), 1\}}.$$

Consequently, following the procedure of Benjamini and Hochberg (1995) to control the false discovery rate at a pre-specified level $\alpha \in (0,1)$, we propose to determine the threshold t by

$$\hat{t} = \inf\left\{0 \le t \le t_0 : \widehat{\mathrm{FDR}}_t \le \alpha\right\} \tag{2.4}$$

for some constant t_0 that is allowed to depend on n and p, whose order is given in Theorem 2. In real implementation, we compute $\widehat{\mathrm{FDR}}_t$ for t taking each value of $\tau_1, \ldots, \tau_p$. As a result, the selected set is defined as

$$\widehat{\mathcal{S}}_\alpha \equiv \{k : \widehat{\mathrm{FDR}}_{\tau_k} \le \alpha, 1 \le k \le p\}.$$

Define $\tau_l \equiv \mathrm{argmax}_{k\in\widehat{\mathcal{S}}_\alpha} \widehat{\mathrm{FDR}}_{\tau_k}$. In other words, τ_l is the t such that $\widehat{\mathrm{FDR}}_t$ is maximized subject to $\widehat{\mathrm{FDR}}_t \le \alpha$. As a result, the respective estimated FDR is $\widehat{\mathrm{FDR}}_{\tau_l}$. The rundown of the proposed false discovery control procedure can be summarized as follows:

Step 1. Given D_1 and D_2, calculate $\tau_1, \cdots, \tau_p$;

Step 2. Compute each $\widehat{\mathrm{FDR}}_t$ for t taking each value of $\tau_1, \ldots, \tau_p$;

Step 3. Given α, search for the set $\widehat{\mathcal{S}}_\alpha \equiv \{k : \widehat{\mathrm{FDR}}_{\tau_k} \le \alpha, 1 \le k \le p\}$;

Step 4. Find $\tau_l \equiv \mathrm{argmax}_{\mathrm{k}\in\widehat{\mathcal{S}}_\alpha} \widehat{\mathrm{FDR}}_{\tau_\mathrm{k}}$ and let $\hat{t} = \tau_l$.

Based on the above procedure, the computational cost is at the order of $O(p)$. We refer the proposed FDR control procedure as QCS-FDR. The QCS-FDR is computationally efficient and its validity to detect important predictors is guaranteed by the main theorems in section 3.

2.3 A stepwise procedure based on the quantile correlation

In practice, when some inactive predictors are highly correlated with certain active predictors, the empirical FDR will be inflated. We next introduce a stepwise procedure to circumvent the problem. Let $\mathcal{C}$ be an arbitrary index set of predictors. And $\boldsymbol{X}_{\mathcal{C}} = (X_j : j \in \mathcal{C}) \in \mathbb{R}^{n\times|\mathcal{C}|}$ is the design matrix indexed by $\mathcal{C}$. Here $|\mathcal{C}|$ is the cardinality of $\mathcal{C}$. For any $k \notin \mathcal{C}$, we define

$$\tau_{k|\mathcal{C}} \equiv \sum_{s=1}^{D_1}\sum_{t=1}^{D_2} \frac{(\widetilde{T}_{k,st} - E_{st})^2}{E_{st}},$$

where $\widetilde{T}_{k,st} = \sum_{i=1}^{n} I\{\hat{Q}_{k,s-1} < \widetilde{X}_{k,i} \leq \hat{Q}_{k,s}\} I\{\hat{Q}^*_{t-1} < Y_i \leq \hat{Q}^*_t\}$ and $\widetilde{X}_k$ is the linear projection of X_k onto the orthogonal complement space spanned by $\boldsymbol{X}_{\mathcal{C}}$. For a special case that $\mathcal{C}$ is a null set, $\tau_{k|\mathcal{C}}$ reduced to τ_k in (2.3).

More notations are needed for presenting the stepwise procedure. Recall that $\widehat{\mathcal{S}}_\alpha$ is the selected set by the multiple testing procedure in section 2.2. Let $\mathcal{A}$ be the current selected set, which could change in different deletion or addition step. Define $\mathcal{A}^c \equiv \widehat{\mathcal{S}}_\alpha/\mathcal{A}$, $\mathcal{A}_{-j} \equiv \mathcal{A} - \{j\}$ and $\mathcal{A}_{+j} \equiv \mathcal{A} + \{j\}$. Initiate $\mathcal{A} = \widehat{\mathcal{S}}_\alpha$. We select relevant predictors by iterating the following deletion and addition steps until no new deletion or addition occurs.

Deletion step: Search for $j_1 \equiv \arg\min_{j\in\mathcal{A}} \tau_{j|\mathcal{A}_{-j}}$. If $\min_{j\in\mathcal{A}} \tau_{j|\mathcal{A}_{-j}} < \gamma_1$, delete j_1 from $\mathcal{A}$, i.e., let $\mathcal{A} = \mathcal{A}_{-j_1}$;

Addition step: Find $j_2 \equiv \arg\max_{j\in\mathcal{A}^c} \tau_{j|\mathcal{A}}$. If $\max_{j\in\mathcal{A}^c} \tau_{j|\mathcal{A}} > \gamma_2$, add j_2 to $\mathcal{A}$, i.e., let $\mathcal{A} = \mathcal{A}_{+j_2}$.

The two constants γ_1 and γ_2 are two pre-specified thresholds that will be discussed later in Theorem 3. The final selected set is $\mathcal{A}$ when the iterated procedure ceases. We refer the proposed stepwise procedure as QCS-S.

2.4 Sure independence screening via quantile correlation

In practical applications, when the dimensionality p is extremely high, the performance of the proposed procedure in section 2.2 might be compromised. By this consideration, as a preliminary step prior to the QCS-FDR procedure, we introduce an independence screening method to reduce ultrahigh dimensionality to a moderate scale at the first stage. To capture the dependence between Y and each X_k, $k = 1, \ldots, p$, in the (s,t)-th cell in the contingency table, we define $\eta_{k,st} = E\left[\{I(Q_{k,s-1} < X_{ik} \le Q_{k,s}) - v_s\}\{I(Q^*_{t-1} < Y_i \le Q^*_t) - v_t\}\right]/(v_s v_t)^{1/2}$. Intuitively, $\eta_{k,st} = 0$ when X_k and Y are independent; otherwise, when X_k and Y are dependent, there exists some (s,t)-th cell such that $\eta_{k,st} \neq 0$. This motivates us to propose the following screening utility to measure the dependence between Y and each X_k, $k = 1, \ldots, p$, which is defined as

$$\hat{\eta}_k = \sum_{s=1}^{D_1}\sum_{t=1}^{D_2}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\{I(\hat{Q}^*_{t-1} < Y_i \le \hat{Q}^*_t) - v_t\}\{I(\hat{Q}_{k,s-1} < X_{k,i} \le \hat{Q}_{k,s}) - v_s\}}{(v_s v_t)^{1/2}}\right]^2. \quad (2.5)$$

As a result, we estimate the true model $\mathcal{S}$ by

$$\widehat{\mathcal{M}} = \{1 \le k \le p : \hat{\eta}_k \ge C_0' n^{-\varrho}\},$$

where C_0' and ϱ are pre-determined thresholds that will be discussed in Condition (C4*) in the next section. In practice, we select a reduced model

$$\widehat{\mathcal{M}}^* = \{1 \le j \le p : \hat{\tau}_k \text{ is among the top } d \text{ largest of all}\},$$

where d is a pre-determined size. One can set $d = a\lfloor n/\log(n)\rfloor$, where a is some constant and $\lfloor n/\log(n)\rfloor$ is the integer part of $n/\log(n)$. We refer this sure independence screening method based on the QC statistic as QCS.

Remark 4. The numbers of quantile levels D_1 and D_2 play similar role as the number of slices in Li (1991) and Zhu and Ng (1995). Similar to Mai and Zou (2015), our numerical experience showed that the performance of the proposed variable screening is not sensitive to the choices of D_1 and D_2 when the sample size is relatively large. More theoretical discussions on the choices of D_1 and D_2 can be found in the next section. Alternatively, in finite-sample experiments, one may also consider a refined version of τ_k that might be more stable to the choice of D_1 and D_2. Suppose that there are B different partitions of the interval $(0,1)$, denoted by $D_{1j} \times D_{2j}$ for $j = 1, \ldots, B$. For the j-th partition, with a slight abuse of notation, we still write $0 = \gamma_0 < \gamma_1 < \gamma_2 < \cdots < \gamma_{D_{1j}} = 1$ and $0 = \rho_0 < \rho_1 < \rho_2 < \cdots < \rho_{D_{2j}} = 1$.

Inspired by Cook and Zhang (2014), one may consider a fused version of τ_k:

$$\hat{\eta}_k^B = \sum_{j=1}^{B}\sum_{s=1}^{D_{1j}}\sum_{t=1}^{D_{2j}}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\{I(\hat{Q}^*_{t_j-1} < Y_i \le \hat{Q}^*_{t_j}) - v_{t_j}\}\{I(\hat{Q}_{k,s_j-1} < X_{k,i} \le \hat{Q}_{k,s_j}) - v_{s_j}\}}{(v_{t_j}v_{s_j})^{1/2}}\right]^2 \tag{2.6}$$

to measure the dependence of X_k and Y for each $k = 1,\dots,p$, where $v_{t_j} = \rho_{t_j} - \rho_{t_j-1}$ and $v_{s_j} = \gamma_{s_j} - \gamma_{s_j-1}$.

3 Theoretical properties

Without loss of generality, let $D = \max(D_1, D_2)$ and the following conditions are needed to establish the asymptotic properties:

(C1) There exists a constant C such that $\max_{1\le k\le p} f(Y|X_k) < C < \infty$ where $f(Y|X_k)$ is the conditional probability density function of Y given X_k, $k = 1,\dots,p$.

(C2) The dimensionality $p = O(n^r)$ for some $r > 0$ and $D = O(\log n)$.

(C2*) The dimensionality p satisfies $\log(p) = O(n^\zeta)$ for some $\zeta > 0$.

(C3) There exist positive constants $v_{\min}$ and $v_{\max}$ such that $0 < v_{\min} \le \min_{1\le s\le D_1} v_s \le \max_{1\le s\le D_1} v_s \le v_{\max} < 1$. Likewise, there exist positive constants $u_{\min}$ and $u_{\max}$ such that $0 < u_{\min} \le \min_{1\le t\le D_2} v_t \le \max_{1\le t\le D_2} v_t \le u_{\max} < 1$.

(C4) There exist some positive constant C_0', $0 < \varrho < 1/2$ and (s,t)-th cell such that $\min_{k\in\mathcal{S}} |\eta_{k,st}| \ge \sqrt{2C_0' n^{-\varrho}}$.

(C5) The cardinality of $\mathcal{S}$ satisfies $|\mathcal{S}| = O(n^\xi)$ for $0 < \xi < 1/2$.

(C6) $E(X_k|\boldsymbol{X}_{\mathcal{S}})$ is linear in $\boldsymbol{X}_{\mathcal{S}}$, for any $k \in \mathcal{S}^c$.

Condition (C1) requires that the conditional density function of Y given each X_k is bounded, which is satisfied by many commonly-used distributions. Condition (C2) requires the dimensionality p to be at a polynomial order of n, but p could grow faster than n. Condition (C2) is only needed to establish the moderate deviation result in Theorem 1. In addition, D is allowed to diverge with n at the order of $O(\log n)$; see the Appendix. In practice, one may set $D = a\log(n)$ for some pre-specified constant a. An alternative condition (C2*) is imposed to establish the sure screening properties in Theorem 3 for the preliminary screening. We allow p to be at an exponential rate of n under condition (C2*). This is crucial to capture those delicate quantile correlation in some small or local regions when the sample size n is large. Condition (C3) requires the quantile grid points are bounded away from 0 and 1, which is imposed to avoid technicality arises at the tail area. Condition (C4) assumes that the minimum true signal vanishes to zero in order of $n^{-\varrho/2}$ as the sample size goes to infinity. Such a condition is typical in the feature screening literature, such as Fan and Lv (2008), He *et al.* (2013) and Cui *et al.* (2015). Condition (C5) requires the number of true important variables is at the order of $o(n^{1/2})$, indicating that the true model may not be very sparse. Condition (C6) assumes that the projection of X_k onto the orthogonal complement space spanned by $\boldsymbol{X}_{\mathcal{S}}$ is linear, which is satisfied when $\boldsymbol{X}$ follows some commonly-used distributions, e.g., the multivariate normal distribution. Similar condition can be found in Zhong *et al.* (2012) and Jiang and Liu (2014).

We next present an important property of the proposed test statistic in Theorem 1.

THEOREM 1. *Suppose Conditions (C1),(C2), (C3) and (C5) hold and* $n^{-1/2} < \Delta_n \leq (\log n^{\beta}/n)^{1/2}$ *for any* $\beta > 0$. *Then, for any positive integers* $D_1 > 1$, $D_2 > 1$ *and any constant* $C_1 > 0$,

$$\sup_{0\leq t\leq C_1 n\Delta_n^2} \sup_{k\in\mathbb{H}_0} \frac{\Pr(\tau_k \geq t)}{S_{D_1\times D_2}(t)} \to 1$$

as $n \to \infty$, *where* $S_{D_1\times D_2}(t)$ *is the survival function of Chi-square distribution with* $(D_1 - 1)(D_2 - 1)$ *degrees of freedom.*

Theorem 1 is a moderate deviation result, that is stronger than the convergence in distribution. It concludes that τ_k converges to a chi-square random variable with $(D_1-1)(D_2-1)$ degrees of freedom in distribution. This theorem indicates that the convergence rate of τ_k is faster than the decaying rate of the survival function itself at the tail area. The proof of Theorem 1 is deferred to the Appendix.

THEOREM 2. *Suppose Conditions (C1), (C2) and (C3)-(C5) hold and* $n^{-1/2} < \Delta_n \leq$

$(\log n^{\beta}/n)^{1/2}$ *for any* $\beta > 0$*. Then, for any* $C_2 > 0$*,*

$$\lim_{n\to\infty} \Pr\Big\{\inf_{k\in\mathbb{H}_1} \tau_k > C_2 n\Delta_n^2\Big\} = 1.$$

Moreover, when $\Delta_n = (\log n/n)^{1/2}$*,*

$$\lim_{n\to\infty} \Pr\Big\{\inf_{k\in\mathbb{H}_1} \tau_k > C_2 \log n\Big\} = 1.$$

Under regularity conditions, for those k in the true active set, Theorem 2 ensures that τ_k has a lower bound with high probability. And the null and the alternative distribution of the test statistic can be well separated, which further implies

$$\lim_{n\to\infty} \Pr\{\tau_k > C_0 n\Delta_n^2 | k \in \mathbb{H}_0\} = 0,$$

$$\lim_{n\to\infty} \Pr\{\tau_k > C_0 n\Delta_n^2 | k \in \mathbb{H}_1\} = 1.$$

Based on Theorem 1 and Theorem 2, the next corollary justifies that the false discovery rate can be controlled by the proposed QCS-FDR procedure in theory.

COROLLARY 1. *Suppose Conditions (C1), (C2) and (C3)-(C5) hold. For a pre-specified level* α*, the FDR of the proposed multiple testing procedure satisfies*

$$\lim_{n\to\infty} \frac{\widehat{\mathrm{FDR}}_{\hat{t}}}{\alpha} = 1.$$

Corollary 1 ensures that the FDR of our procedure can be controlled by a pre-specified level α theoretically.

Remark 5. A common result on controlling falsely discoveries for screening methods is to control the size of the selected model such that the number of variables retained is negligible compared with p, hence the percentage of falsely discoveries is also negligible. Contrary to existing FDR results in the screening literature such as Zhu *et. al* (2011), our proposed procedure can control the FDR at the quantitative level α with rigorous statistical guarantee.

THEOREM 3. *Suppose Conditions (C1), (C2) and (C3)-(C6) hold and* $n^{-1/2} < \Delta_n \leq (\log n^{\beta}/n)^{1/2}$ *for any* $\beta > 0$*. Let* $\mathcal{C}$ *be an arbitrary given index set of predictors. Then, for any positive integers* $D_1 > 1$*,* $D_2 > 1$*,* $C_3 > 0$ *and any* $k \notin \mathcal{C}$*, we have*

$$\sup_{0\le t\le C_3 n\Delta_n^2} \sup_{k\in\mathbb{H}_0} \frac{\Pr(\tau_{k|\mathcal{C}} \geq t)}{S_{D_1\times D_2}(t)} \to 1$$

as $n \to \infty$, *where* $S_{D_1 \times D_2}(t)$ *is the survival function of Chi-square distribution with* $(D_1 - 1)(D_2 - 1)$ *degrees of freedom.*

Based on Theorem 3, the thresholds γ_1 and γ_2 in the stepwise algorithm can be chosen as quantiles of the Chi-square distribution with $(D_1 - 1)(D_2 - 1)$ degrees of freedom.

When p grows at an exponential rate of n and only relatively small number of variables are relevant, a preliminary feature screening prior to the multiple testing procedure will be performed. The following theorem presents the sure screening property of the proposed QCS screening procedure.

THEOREM 4. *If* $D_1 = O(n^{\kappa})$ *and* $D_2 = O(n^{\xi})$, *where* $\kappa \geq 0, \xi \geq 0$ *and* $\kappa + \xi + \varrho < 1/2$, *under Conditions (C1), (C2*), (C3)-(C5), we have*

$$\Pr\left(\mathcal{S} \subset \widehat{\mathcal{M}}\right) \geq 1 - O\left(p \exp\{-bn^{1-2\varrho-2\kappa-2\xi} + (\kappa + \xi)\log(n)\}\right),$$

where b *is a positive constant. Thus, if* $\log(p) = O(n^{\zeta})$ *and* $\zeta < 1 - 2\varrho - 2\kappa - 2\xi$, *QCS enjoys the sure screening property.*

According to Theorem 4, one can apply the QCS screening method together with the proposed QC-FDR procedure to select important variables when the dimensionality is ultra high. It also indicates that, when n is sufficiently large, D_1 and D_2 could be at the order of $O(n^{\kappa})$ and $O(n^{\xi})$ with $\kappa, \xi \geq 0$ satisfying $\zeta < 1 - 2\varrho - 2\kappa - 2\xi$, so that the sure screening property still holds.

4 Simulation studies

In this section, we demonstrate the performance of the proposed procedure via several simulated examples. In practice, to avoid mathematical challenges caused by the reuse of the sample, the sample splitting idea (Hartigan, 1969; Cox, 1975) is adopted. Let $\{(Y_i^{(1)}, \boldsymbol{X}_i^{(1)\top}), i = 1, \ldots, n_1\}$, $\{(Y_i^{(2)}, \boldsymbol{X}_i^{(2)\top}), i = 1, \ldots, n_2\}$ and $\{(Y_i^{(3)}, \boldsymbol{X}_i^{(3)\top}), i = 1, \ldots, n_3\}$ be a random disjoint partition of $\{(Y_i, \boldsymbol{X}_i^{\top}), i = 1, \ldots, n\}$. Our proposed procedure consists of three steps: QCS, to screen active predictors; QCS-FDR, to control the FDR; QCS-S, to identify true active predictors. The three steps are implemented in the following way:

(1) QCS: The p predictors are ranked in a descending order according to (2.5) based on $\{(Y_i^{(1)}, \boldsymbol{X}_i^{(1)\top}),\ i = 1, \ldots, n_1\}$ and the top $d = o(n)$ predictors are selected, denoted by $\widehat{\mathcal{M}}^*$.

(2) QCS-FDR: Given a FDR level α, the threshold $\hat{t}$ are determined according to (2.4) based on $\{(Y_i^{(2)}, \boldsymbol{X}_i^{(2)\top}), i = 1, \ldots, n_2\}$ and the selected set is $\widehat{\mathcal{S}}_\alpha \equiv \{k : \tau_k \geq \hat{t}, k \in \widehat{\mathcal{M}}^*\}$.

(3) QCS-S: Iterate the deletion and addition steps on $\widehat{\mathcal{S}}_\alpha$ in Step 2 based on $\{(Y_i^{(3)}, \boldsymbol{X}_i^{(3)\top}),\ i = 1, \ldots, n_3\}$ until no deletion and addition occurs.

4.1 Performance of QCS

In this subsection, we first compare the variable screening performance of our proposed QCS with the sure independence screening (SIS) (Fan and Lv, 2008), the distance correlation based screening (DC-SIS) (Li *et al.,* 2012) and the sliced inverse regression via inverse modeling (SIRI) (Jiang and Liu, 2014). Notice that SIRI is an iterative procedure and others are non-iterative ones. For a fair comparison, we adopt the initial screening step described in Section 2.3 of Jiang and Liu (2014) to implement SIRI in a non-iterative fashion. We evaluate the performance of each method via 5%, 25%, 50%, 75% and 95% quantiles of the minimum model size that all relevant predictors are included based on 100 replications, with closer to the true model size indicating better performance in variable screening.

In the simulation, the predictors $\boldsymbol{X} = (X_1, \ldots, X_p)^\top$ are generated from a p-variate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})_{p\times p}$, where $\sigma_{ij} = \rho^{|i-j|}$. We set $\rho = 0$, 0.5 and 0.9. The number of quantile grid points $D_1 = D_2 = 8, 9$ or 10. To simulate ultra-high dimensional scenario, we set $n = 500$ and $p = 1000, 5000$ for each scenario. The response variable is generated from the following models:

Scenario 1.1: $Y = X_1 + X_2 + X_{100} + \boldsymbol{\varepsilon}$;

Scenario 1.2: $Y = 3X_1 + 4X_2^2 + 2\tan(\pi X_{100}/2) + \boldsymbol{\varepsilon}$;

Scenario 1.3: $Y = 3\exp(3X_1) + 4\sin(\pi X_2/2) + 5X_{100}I(X_{100} > 0) + \boldsymbol{\varepsilon}$;

Scenario 1.4: $Y = 1 - 2(X_1 + X_2)^{-3}\exp\{1 + 3\sin(\pi X_{100}/2)\} + \boldsymbol{\varepsilon}$,

where $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, 1)$ independent of $\boldsymbol{X}$.

The quantiles of the minimum model size that includes all three active predictors are reported in Tables 1-2. Under Scenario 1.1 with linear model, all four competitors perform well. Under Scenario 1.2 with additive model, SIRI and SIS screening procedures fail to detect the active predictors. Under Scenarios 1.3-1.4 with nonlinear relationship between the response and predictors, SIS and DC-SIS screening procedures behave poor. In contrast, the proposed QCS procedure works reasonably well in all scenarios and outperforms the other three competitors. The performance of the four methods are only sightly discounted when p increases from 1000 to 5000.

4.2 Performance of QCS-FDR

In this subsection, some scenarios are simulated to examine the FDR control as well as the sure screening property of the proposed procedure. Consider the following three regression models with the covariates $\boldsymbol{X}$ generated from $\mathcal{N}(0, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = (\rho^{|i-j|})_{p\times p}$,

Scenario 2.1: $Y = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $\rho = 0.5$, $\boldsymbol{\beta} = (\mathbf{1.5}_{s_0}^\top, \mathbf{0}_{p-s_0}^\top)^\top$ with $s_0 = 10$;

Scenario 2.2: $Y = \exp(\boldsymbol{X}\boldsymbol{\beta}) + \boldsymbol{\varepsilon}$, $\rho = 0.5$, $\boldsymbol{\beta} = (\mathbf{1.5}_{s_0}^\top, \mathbf{0}_{p-s_0}^\top)^\top$ with $s_0 = 10$;

Scenario 2.3: $Y = \frac{\sum_{j=1}^{10} X_j}{0.5+(1.5+\sum_{j=2}^{4} X_j)^2} + 0.1\boldsymbol{\varepsilon}$, $\rho = 0.0$,

where $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, 1)$ independent of $\boldsymbol{X}$.

In this example, we set $n = 1000$ and $p = 1000, 5000$. The number of quantile grid points $D_1 = D_2 = 3, 4$ or 5. In each scenario, the full sample are randomly divided into two non-overlapping sub-samples. The sample size for QCS $n_1 = 250$ and $d = \lfloor n/\log(n)\rfloor$. The sample size for QCS-FDR $n_2 = 750$. The nominal false discovery rate $\alpha = 0.05$. Based on 100 replications, we evaluate the performance based on the following criteria:

- X_j: the probability that the active predictors X_j is selected;
- $|\widehat{\mathcal{S}}_\alpha|$: the average number of selected predictors;
- $\widehat{\mathrm{FDR}}$: the average of empirical false discovery proportion.

The results of QCS-FDR procedure are summarized in Table 3. One can see that the proposed QCS-FDR procedure controls the empirical FDR under the pre-specified level α for most scenarios. The proposed procedure also possesses the sure screening property reasonably well in all scenarios based on FDR control. Table 3 also reports that the empirical FDR are not sensitive to different D_1 and D_2, indicating that the QCS-FDR procedure is robust to the choice of D_1 and D_2. The proposed QCS-FDR procedure works reasonably well as p increases from 1000 to 5000. In summary, the proposed procedure performs well in various practical settings.

4.3 Performance of QCS-S

In this subsection, we further compare the performance of the proposed stepwise procedure with the sliced inverse regression via inverse modeling (SIRI) (Jiang and Liu, 2014). Several models with linear, nonlinear or higher-order interaction effects are considered. The predictors $\boldsymbol{X} = (X_1, \dots, X_p)^\top$ are generated from a multivariate normal distribution with mean $\mathbf{0}$ and covariances $\text{Cov}(X_i, X_j) = \rho^{|i-j|}$ for $1 \leq i, j \leq p$. The response variable is simulated from the following models:

Scenario 3.1: $Y = X_2 - 0.5X_1 + 0.5X_{100} + 0.2\boldsymbol{\varepsilon}$, $\rho = 0.9$;

Scenario 3.2: The same model in Scenario 1.2 with $\rho = 0.5$;

Scenario 3.3: The same model in Scenario 1.3 with $\rho = 0.5$;

Scenario 3.4: The same as in Scenario 2.1 except for $\rho = 0$ and $\boldsymbol{\beta} = (\mathbf{1.5}_5^\top, \mathbf{0}_{p-5}^\top)^\top$;

Scenario 3.5: $Y = 2X_1 + 2X_2 + 3X_2X_9 + \boldsymbol{\varepsilon}$, $\rho = 0$;

Scenario 3.6: $Y = 2X_1 + 4\tan(\pi X_3 X_8/2) + \boldsymbol{\varepsilon}$, $\rho = 0$;

Scenario 3.7: $Y = 6X_1/(1 + 2X_1X_3 + 2X_1X_5) + \boldsymbol{\varepsilon}$, $\rho = 0$;

Scenario 3.8: $Y = 4X_5X_{10}X_{50} + \boldsymbol{\varepsilon}$, $\rho = 0$,

where $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, 1)$ independent of $\boldsymbol{X}$.

In this example, we set $n = 1500$ and $p = 1000, 5000$. The number of quantile grid points $D_1 = D_2 = 8, 9$ or 10. In each scenario, we randomly divide the sample into three non-overlapping sub-samples. Set $n_1 = 500$, $d = \lfloor n/\log(n) \rfloor$, $n_2 = 500$ and $n_3 = 500$. The nominal false discovery rate $\alpha = 0.05$. The thresholds in the deletion and addition steps are set to be $\gamma_1 = \chi^2(1 - \alpha - 0.2, (D_1 - 1)(D_2 - 1))$ and $\gamma_2 = \chi^2(1 - \alpha, (D_1 - 1)(D_2 - 1))$, where $\chi^2(\gamma, m)$ is the 100γ-th quantile of the Chi-square distribution with m degrees of freedom. Based on 100 replications, two quantities are used to measure the variable selection performance of each method:

- FN: average number of true active predictors falsely excluded as irrelevant predictors;
- FP: average number of irrelevant predictors falsely selected as true active predictors.

The results are reported in Table 4. Our proposed QCS-S is able to detect most of the relevant predictors with a comparable number of false positives under all scenarios, while SIRI is able to detect all relevant predictors (FN=0.00) with a large number of false positives under Scenarios 3.1-3.8 except Scenarios 3.2 and 3.6. Under Scenarios 3.2 and 3.4

with additive models, SIRI missed one of relevant predictor (FN$\approx$ 1.00) that appeared in the additive term most of the time. Overall, the proposed QCS-S procedure works reasonably well in all scenarios and outperforms the SIRI under these settings.

5 Applications

In this section, we will illustrate the performance of the proposed procedures with a rat eye microarray expression dataset collected by Scheetz *et al.* (2006), which is available at Gene Expression Omnibus (`http://www.ncbi.nlm.nih.gov/geo`) with GEO accession number GSE5680. This experiment was designed to investigate gene regulation in the mammalian eye and identify genetic variation relevant to human eye disease. In this dataset, 120 12-week old male offspring of rats, containing $31,042$ different probe sets, were selected for tissue harvesting from the eyes and for microarray analysis. Following Huang *et al.* (2008), we selected $18,976$ gene probes that exhibited significant signal for reliable analysis in the mammalian eye. According to Chiang *et al.* (2006) and Huang *et al.* (2008), the probe from TRIM32 whose probe number is 1389163_at, was recently found to causes Bardet-Biedl syndrome. The main interest of this study is to find the genes that are correlated with the gene TRIM32. Therefore, the probe of TRIM32 is regarded as response Y. In this case, the sample size $n = 120$ and the number of probe is $p = 18,975$. For our analysis, all $18,976$ probes were analyzed on a logarithmic scale and scaled to have zero mean and unit variance.

To evaluate the performance of different methods, we randomly split the 120 samples into training and testing data with $n_{tra} = 80$ and $n_{tes} = 40$. We first set $D = 4$ and apply our proposed screening procedure (QCS) to the training data set and retain the top $\lfloor n_{tra}/\log(n_{tra})\rfloor = 18$ probes. We then apply the QCS-FDR and get the selected probe set $\widehat{\mathcal{S}}$. Considering that some important probes might appear in the pairwise interaction effects, we consider a set of all possible pairwise interaction terms, denoted by $\widehat{\mathcal{I}} = \{(i,j) : i \in \widehat{\mathcal{S}}, j \in \widehat{\mathcal{S}}\}$. We next fit a regularized linear model with all variables in $\widehat{\mathcal{S}}$ and all pairwise interaction effects in $\widehat{\mathcal{I}}$ to the training data using lasso penalty. To implement this, the R package program RAMP (Hao *et al.* 2018) was used. Lastly, we evaluate the predictive performance of the resulting model by calculating the mean square prediction errors (MSPE) with the testing data set:

$$\text{MSPE} = \frac{1}{n_{tes}} \sum_{i \in \mathcal{T}} (\hat{Y}_i - Y_i)^2,$$

where $\mathcal{T} = \{i\text{: the } i\text{th sample belongs to testing data}\}$ and $\hat{Y}_i$ is the fitted value of Y_i.

We repeat the above randomly partition of training data and testing data for 50 times. For comparison, we also apply the sliced inverse regression via inverse modeling (SIRI) (Jiang and Liu, 2014), the interaction pursuit (IP) (Fan *et al.* 2016) and the interaction pursuit via distance correlation (IPDC) (Kong *et al.* 2017) to this data set under the same settings. The result is shown in Table 4. It can be seen that the proposed method exhibits superior predictive performance, namely, the mean of MSPE of QCS-FDR is the smallest among that of all methods and the variance is relatively small.

6 Concluding remarks

In this paper, we advocate a multiple testing procedure with FDR control to detect important variables. The multiple testing procedure can be applied together with the QCS screening method when the dimensionality is ultra-high. The proposed procedure is built on the quantile-correlation (QC) statistic, which depends on the partition of the sample space. Generally speaking, if D grows faster than n and p, the QC statistic can capture more subtle associations. But large D will slow down the convergence of the asymptotic null distribution. In the simulation studies, we set different values of D_1 and D_2 to examine the performance of our procedure. Nevertheless, it would be of interest to study a data-driven way to select D_1 and D_2. We leave space here for future research.

Appendix

Define $\hat{I}_{k,i,s} = I(\hat{Q}_{k,s-1} < X_{ik} \le \hat{Q}_{k,s})$, $I_{k,i,s} = I(Q_{k,s-1} < X_{ik} \le Q_{k,s})$, $\hat{I}^*_{i,t} = I(\hat{Q}^*_{t-1} < Y_i \le \hat{Q}^*_t)$ and $I^*_{i,t} = I(Q^*_{t-1} < Y_i \le Q^*_t)$. Write $\tau_{k,st} = E[\{I(Q_{k,s-1} < X_{ik} \le Q_{k,s}) - v_s\}\{I(Q^*_{t-1} < Y_i \le Q^*_t) - v_t\}/(v_s v_t)^{1/2}]$.

LEMMA 1. *Suppose*

$$\delta_{k,st} = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \frac{\hat{I}_{k,i,s}\hat{I}^*_{i,t} - v_s v_t}{(v_s v_t)^{1/2}} - \frac{1}{n^{1/2}} \sum_{i=1}^{n} \frac{(I_{k,i,s} - v_s)(I^*_{i,t} - v_t)}{(v_s v_t)^{1/2}}.$$

Then, under Conditions (C1), (C2) and (C3), for all constant $C_i > 0, (i = 1, 2, 3, 4)$

$$\Pr\Big(\sup_{k \in \mathbb{H}_1} \sup_{1 \le s \le D_1, 1 \le t \le D_2} |\delta_{k,st}| > C_1 \sqrt{n}\Delta_n\Big) \le C_0 D_1 D_2 p \exp(-C_2 n \Delta_n^2),$$

$$\sup_{k \in H_0} \sup_{1 \le s \le D_1, 1 \le t \le D_2} E(\delta_{k,st}) \le C_3 D_1 D_2 \sqrt{n} \Delta_n^2,$$

$$\sup_{k \in H_0} \sup_{1 \le s \le D_1, 1 \le t \le D_2} E(|\delta_{k,st}|^m) \le C_4 D_1 D_2 (\sqrt{n}\Delta_n)^m,$$

where $n^{-1/2} < \Delta_n \le (\log n^\beta / n)^{1/2}$ with $\beta > 0$ and $C_1,\ C_2,\ C_3,\ C_4$ are different constants irrelevant to n.

Proof. Note that

$$\begin{aligned}\delta_{k,st} =& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(I_{k,i,s} - v_s)(\hat{I}^*_{i,t} - I^*_{i,t})}{(v_s v_t)^{1/2}} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(I^*_{i,t} - v_t)(\hat{I}_{k,i,s} - I_{k,i,s})}{(v_s v_t)^{1/2}} \\ &+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(\hat{I}_{k,i,s} - I_{k,i,s})(\hat{I}^*_{i,t} - I^*_{i,t})}{(v_s v_t)^{1/2}} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\hat{I}_{k,i,s} v_t + \hat{I}^*_{i,t} v_s - 2 v_s v_t}{(v_s v_t)^{1/2}} \\ =& \delta_{1,k,st} + \delta_{2,k,st} + \delta_{3,k,st} + \delta_{4,k,st}.\end{aligned}$$

First, we consider $\delta_{1,k,st}$. For simplicity, we set $D_1 = D_2 = D$. Under Condition (C1), we have

$$\sup_{1 \le s,t \le D} |\delta_{1,k,st}| = \sup_{1 \le s,t \le D} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(I_{k,i,s} - v_s)(\hat{I}^*_{i,t} - I^*_{i,t})}{(v_s v_t)^{1/2}} \Big| \le \frac{1 - v_{min}}{v_{min}} \frac{1}{\sqrt{n}} \sup_{1 \le t \le D} \sum_{i=1}^{n} |\hat{I}^*_{i,t} - I^*_{i,t}|. \tag{A.1}$$

By the Bahadur representation of sample quantiles (Hesse *et al.*, 1990),

$$\hat{Q}^*_t - Q^*_t = \frac{F(Q^*_t) - F_n(Q^*_t)}{f(Q^*_t)} + R_n,$$

where $F_n(x)$ is the empirical distribution function and $R_n = O\{n^{-3/4}(\log n)^{1/2}(\log\log n)^{1/4}\}$. Let $n^{-1/2} < \Delta_n \leq (\log n^{\beta}/n)^{1/2}$ with $\beta > 0$. Since $R_n = O\{n^{-3/4}(\log n)^{1/2}(\log\log n)^{1/4}\} = o(\Delta_n)$, by the Hoeffding's inequality, for any t and certain constant $C > 0$,

$$\begin{aligned}&\Pr(\hat{Q}_t^* - Q_t^* > \Delta_n)\\ =&\Pr\Big\{\frac{F(Q_t^*) - F_n(Q_t^*)}{f(Q_t^*)} + R_n > \Delta_n\Big\}\\ =&\Pr\{F(Q_t^*) - F_n(Q_t^*) > \Delta_n f(Q_t^*)\}\\ \leq&\exp(-2nC\Delta_n^2).\end{aligned}$$

Then,

$$\Pr\Big\{\sup_{t=1}^{D}(\hat{Q}_t^* - Q_t^*) > \Delta_n\Big\} \leq D\exp(-2Cn\Delta_n^2).$$

Similarly, for X_k, $k = 1, \ldots, p$,

$$\Pr\Big\{\sup_{t=1}^{D}(\hat{Q}_{k,s} - Q_{k,s}) > \Delta_n\Big\} \leq D\exp(-2Cn\Delta_n^2).$$

Define $\mathcal{X} = \Big\{(X_{ik}, Y_i)_{i=1}^n : \sup_{t=1}^{D}(\hat{Q}_t^* - Q_t^*) > \Delta_n, \sup_{s=1}^{D}(\hat{Q}_{k,s} - Q_{k,s}) > \Delta_n\Big\}$. On the space $\mathcal{X}$, for sufficiently large n, we have

$$\sup_{1\leq t\leq D}|\hat{I}_{i,t}^* - I_{i,t}^*| \leq \sup_{1\leq t\leq D} I(Y_i \text{ is between } Q_t^* \pm \Delta_n). \tag{A.2}$$

Then, in view of (A.1),

$$\sup_{1\leq s,t\leq D}|\delta_{1,k,st}| \leq \frac{1-v_{\min}}{v_{\min}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sup_{1\leq t\leq D} I(Y_i \text{ is between } Q_t^* \pm \Delta_n). \tag{A.3}$$

Note that $\Pi = (1-v_{min})/v_{min}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sup_{1\leq t\leq D} I(Y_i \text{ is between } Q_{i,t}^* \pm \Delta_n)$. Then,

$$E(\Pi) \leq \frac{1-v_{\min}}{v_{\min}}\sqrt{n}\sup_{1\leq t\leq D}\Pr(Y_i \text{ is between } Q_{i,t}^* \pm \Delta_n) \leq 2\sqrt{n}\frac{1-v_{min}}{v_{min}}C_1\Delta_n,$$

where C_1 is some positive constant. By the Azuma's inequality,

$$\Pr\left\{|\Pi - E(\Pi)| > 4\sqrt{n}\frac{1-v_{min}}{v_{min}}C_1\Delta_n\right\} \leq 2\exp(-C_2 n\Delta_n^2),$$

where C_2 is some positive constant. There exists $C_n = O(\sqrt{n}\Delta_n)$ for n sufficiently large such that

$$\begin{aligned}
&\Pr\Big(\sup_{k\in\mathbb{H}_1}\sup_{1\le t\le D}|\delta_{1,k,st}| > C_n\Big)\\
\le&\Pr\{\sup_{k\in\mathbb{H}_1}|\Pi - E(\Pi)| > C_n - E(\Pi)\}\\
=&\Pr\{\sup_{k\in\mathbb{H}_1}|(\Pi) - E(\Pi)| > C_n - 2\sqrt{n}\frac{1-u_0}{u_0}C_1\Delta_n\}\\
\le&2C_3 D\exp(-C_4 n\Delta_n^2)
\end{aligned}$$

for some positive constant C_3 and C_4.

Next, for $\delta_{2,k,st}$, it can be easily checked that

$$\Pr\Big(\sup_{k\in\mathbb{H}_1}\sup_{1\le t\le D}|\delta_{2,k,st}| > C_n\Big) \le 2C_5 Dp\exp(-C_6 n\Delta_n^2).$$

For $\delta_{3,k,st}$, we have

$$\sup_{1\le s,t\le D}|\delta_{3,k,st}| = \sup_{1\le s,t\le D}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(\hat{I}_{k,i,s} - I_{k,i,s})(\hat{I}^*_{i,t} - I^*_{i,t})}{(v_s v_t)^{1/2}}\Big| \le \frac{2}{u_0}\frac{1}{\sqrt{n}}\sup_{1\le t\le D}\sum_{i=1}^{n}|\hat{I}^*_{i,t} - I^*_{i,t}|.$$

It can be shown in a similar fashion that

$$\Pr\Big(\sup_{k\in\mathbb{H}_1}\sup_{1\le t\le D}|\delta_{2,k,st}| > C_n\Big) \le 2C_5 Dp\exp(-C_6 n\Delta_n^2), \tag{A.4}$$

and

$$\Pr\Big(\sup_{k\in\mathbb{H}_1}\sup_{1\le s,t\le D}|\delta_{3,k,st}| > C_n\Big) \le 2C_7 D\exp(-C_8 n\Delta_n^2). \tag{A.5}$$

Furthermore,

$$\begin{aligned}
\hat{I}_{k,i,s} =&I(\hat{Q}_{k,s-1} < X_{ik} \le \hat{Q}_{k,s})\\
=&I(\hat{Q}_{k,s-1} < X_{k,i} < Q_{k,s-1}) - I(\hat{Q}_{k,s} < X_{k,i} \le Q_{k,s}) + I(Q_{k,s-1} < X_{k,i} \le Q_{k,s}).
\end{aligned}$$

Thus, for $r_n = o(n^{-1/2})$, there exists a postive constant C_9 such that,

$$\sup_{1\le s,t\le D}\Big|\frac{1}{n}\sum_{i=1}^{n}\hat{I}_{k,i,s} - v_s\Big| = \sup_{1\le s,t\le D}E|\hat{I}_{k,i,s} - I_{k,i,s}| + r_n$$

$$
\begin{aligned}
&= \sup_{1\le s,t\le D} E|I(\hat{Q}_{k,i,s-1} < X_{k,i} < Q_{k,i,s-1}) - I(\hat{Q}_{k,i,s} < X_{k,i} < Q_{k,i,s})| + r_n \\
&\le \sup_{1\le s,t\le D} \Pr\{\mathrm{X_{k,i}\ is\ between\ Q_{k,s} \pm \Delta_n}\} \\
&\le C_9\Delta_n.
\end{aligned}
$$

Therefore, for $\varepsilon_n = O(n\Delta_n)$,

$$
v_s n - \varepsilon_n \le \sum_{i=1}^{n} \hat{I}_{k,i,s} \le v_s n + \varepsilon_n, \quad v_s n - \varepsilon_n \le \sum_{i=1}^{n} \hat{I}^*_{k,i,s} \le v_s n + \varepsilon_n. \tag{A.6}
$$

Combining (A.3) and (A.6), we have

$$
\delta_{4,k,st} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{v_t(\hat{I}_{k,i,s} - v_s) + v_s(\hat{I}^*_{i,t} - v_t)}{(v_s v_t)^{1/2}} = O(\sqrt{n}\Delta_n).
$$

Then, for some postive constant C_0, C and $C_n = O(\sqrt{n}\Delta_n)$,

$$
\Pr\Big(\sup_{k\in\mathbb{H}_1}\sup_{1\le s,t\le D} |\delta_{k,st}| > C_n\Big) \le C_0 D p \exp(-Cn\Delta_n^2). \tag{A.7}
$$

Under $\mathbb{H}_{0,k}$, by (A.6), for some constant C_{10},

$$
\begin{aligned}
E(\delta_{k,st}) \le& E\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{(\hat{I}_{k,i,s}\hat{I}^*_{i,t} - v_s v_t)}{(v_s v_t)^{1/2}}\right\} \\
\le& C_{10}\sqrt{n}(E\hat{I}_{k,i,s}\hat{I}_{k,t}) \\
\le& C_{10}\sqrt{n}\Delta_n^2.
\end{aligned}
$$

When m is a positive even number, as $E|\delta^m_{k,st}| \le C\sum_{l=1}^{3} E|\delta^m_{l,k,st}|$,

$$
E\delta^m_{1,k,st} = \sum_{\sum_{l=1}^{L} m_l = m} C_m n^{-m/2} E\Big\{\prod_{l=1}^{L}(I_{k_l,i,s} - v_s)^{m_l}\Big\} E\Big\{\prod_{l=1}^{L}(\hat{I}^*_{k_l,t} - I^*_{k_l,t})^{m_l}\Big\}.
$$

Then, there exists $m_l = 1$ such that $E\{\prod_{l=1}^{L}(I_{k_l,i,s} - v_s)\} = 0$. And by (A.3),

$$
\sup_{1\le t\le D} \Big|E\left\{\prod_{l=1}^{L}(\hat{I}^*_{k_l,t} - I^*_{k_l,t})^{m_l}\right\}\Big| \le C(\sqrt{n}\Delta_n)^L. \tag{A.8}
$$

As a result,

$$
\sup_{1\le k\le p}\sup_{1\le s,t\le D} E|\delta_{1,k,st}|^m \le CD^2(\sqrt{n}\Delta_n)^m.
$$

Similarly, we can show $\sup_{1\le t\le D}|\delta_{2,k,st}|^m \le C(\sqrt{n}\Delta_n)^m$. Finally,

$$E\delta_{3,k,st}^m = \sum_{\sum_{l=1}^L m_l=m} n^{-m/2}E\{\prod_{l=1}^{L}(\hat{I}_{k_l,i,t}-I_{k_l,i,t})^{m_l}\}E\{\prod_{l=1}^{L}(\hat{I}^*_{k_l,t}-I^*_{k_l,t})^{m_l}\}.$$

It follows directly that

$$\sup_{1\le k\le p}\sup_{1\le s,t\le D} E\{|\delta_{3,k,st}|^m\} \le C_4D^2(\sqrt{n}\Delta_n)^m$$

for some positive constant C_4. The proof of Lemma 1 is complete. □

LEMMA 2. *Suppose X is a $\chi^2(D)$ random variable with D degrees of freedom. Then,*

$$\lim_{t\to+\infty}\frac{P(X>t)}{\{\Gamma(D/2)\}^{-1}(t/2)^{D/2-1}e^{-t/2}} = 1.$$

LEMMA 3. *Let $\tilde{\tau}_i = \sum_{s=1}^D\sum_{t=1}^D\{n^{-1/2}\sum_{k=1}^n(I_{k,i,s}-v_s)(I_{k,t}-v_t)/(v_sv_t^*)^{1/2}\}^2$. Then,*

$$\sup_{k\in\mathbb{H}_0}\left|\frac{P(\tilde{\tau}_k\ge t)}{S_{D_1\times D_2}(t)}-1\right| \le CD^6(1+t)^{3/2}n^{-1/2},$$

for $t=o(n^{1/3}D^{-4})$ and some postive constant C.

The proofs of Lemma 2 and Lemma 3 can be referred to Xie and Li (2018). It is omitted here.

LEMMA 4 (Hoeffding's inequality)**.** *Let $X_1,\ldots,X_n$ be independent random variables. Assume that $\Pr(X_i\in[a_i,b_i])=1$ for $i=1,\ldots,n$ where a_i and b_i are constants. Let $\bar{X}=n^{-1}\sum_{i=1}^n X_i$. Then the following inequality holds,*

$$\Pr\{|\bar{X}-E(\bar{X})|\ge t\} \le 2\exp\left\{-\frac{2nt^2}{\sum_{i=1}^n(b_i-a_i)^2}\right\},$$

where t is a positive constant and $E(\bar{X})$ is the expected value of $\bar{X}$.

The proof of Lemma 4 can be found in Hoeffding *et al.* (1963). We omit it here.

Proof of Theorem 1. For convenience, we set $D_1=D_2=D$. Note that

$$L_{k,st} = n^{-1/2}\sum_{i=1}^n\frac{(I_{k,i,s}-v_s)(I^*_{i,t}-v^*_t)}{(v_sv^*_t)^{1/2}},$$

$$\delta_{k,st} = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \frac{\hat{I}_{k,i,s}\hat{I}^*_{i,t} - v_s v^*_t}{(v_s v^*_t)^{1/2}} - \frac{1}{n^{1/2}} \sum_{i=1}^{n} \frac{(I_{k,i,s} - v_s)(I^*_{i,t} - v^*_t)}{(v_s v^*_t)^{1/2}}.$$

Let $\tilde{\tau}_k = \sum_{s=1}^{D}\sum_{t=1}^{D} L^2_{k,st}$ and

$$R_k = 2\sum_{s=1}^{D}\sum_{t=1}^{D} L_{k,st}\delta_{k,st} + \sum_{s=1}^{D}\sum_{t=1}^{D} \delta^2_{k,st}.$$

Then,

$$\tilde{\tau}_k + R_k = \sum_{s=1}^{D}\sum_{t=1}^{D}(L_{k,st} + \delta_{k,st})^2 = \sum_{s=1}^{D}\sum_{t=1}^{D}\left\{\frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{\hat{I}_{k,i,s}\hat{I}_{i,t} - v_s v^*_t}{(v_s v^*_t)^{1/2}}\right\}^2 = \tau_k. \tag{A.9}$$

Suppose there exists a constant $\varepsilon > 0$ such that

$$\begin{aligned}\frac{P(\tau_k > t)}{S_{D_1\times D_2}(t)} &= \frac{P(\tilde{\tau}_k + R_k > t - \varepsilon + \varepsilon)}{S_{D_1\times D_2}(t)} \\ &\leq \frac{P(\tilde{\tau}_k > t - \varepsilon) + P(R_k > \varepsilon)}{S_{D_1\times D_2}(t)} \\ &= \frac{S(t-\varepsilon)}{S_{D_1\times D_2}(t)}\frac{P(\tilde{\tau}_k > t - \varepsilon)}{S_{D_1\times D_2}(t-\varepsilon)} + \frac{P(R_k > \varepsilon)}{G_D(t)}.\end{aligned}$$

Therefore, for any positive constant ε,

$$\sup_{0\leq t\leq C_0 n\Delta_n^2}\ \sup_{k\in\mathbb{H}_0} \frac{P(\tau_k > t)}{S_{D_1\times D_2}(t)} \leq \sup_{0\leq t\leq C_0 n\Delta_n^2}\ \sup_{(k)\in\mathbb{H}_0}\left\{\frac{S_{D_1\times D_2}(t-\varepsilon)}{S_{D_1\times D_2}(t)}\frac{P(\tilde{\tau}_k > t-\varepsilon)}{S(t-\varepsilon)} + \frac{P(R_k > \varepsilon)}{S_{D_1\times D_2}(t)}\right\}.$$

To establish the asymptotic distribution of $\tilde{\tau}_k$ for $k = 1, \ldots, p$, we write $\boldsymbol{L}_k$=$(L_{k,11}, L_{k,12}, ..., L_{k,DD})$. Under $\mathbb{H}_0$,

$$E(L_{k,st}) = n^{-1/2}\sum_{i=1}^{n}\frac{\{E(I_{k,i,s}) - v_s\}\{E(I_{i,t})^* - v^*_t\}}{(v_s v^*_t)^{1/2}} = 0.$$

By the central limit theorem, $\boldsymbol{L}_k$ converges to the multivariate normal distribution $N(\mu, \Sigma)$ in distribution, where $\mu = E(\boldsymbol{L}_k) = \boldsymbol{0}$, and $\Sigma = Var(\boldsymbol{L}_k) = \Sigma_1 \bigotimes \Sigma_1$, $\Sigma_1 = I_D - \sqrt{\boldsymbol{v}}\sqrt{\boldsymbol{v}}^\top$ and $\boldsymbol{v} = (v_1, ..., v_D)$. $\bigotimes$ is the note for outer product. Let $\Sigma_1 = (\sigma_{ij})_{D*D}$. Clearly, Σ_1 is idempotent and symmetric as $\sqrt{\boldsymbol{v}}^\top\sqrt{\boldsymbol{v}}$=$\sum_{i=1}^{D} v_i = 1$. Thus, Σ is also symmetric and idempotent. This is because $\Sigma^2 = (\delta_{ij}\Sigma_1) \times (\sigma_{ij}\Sigma_1) = (\sum_{k=1}^{D}\sigma_{ik}\sigma_{kj}\Sigma_1^2)_{D^2\times D^2} = \Sigma$ and $\sum_{k=1}^{D}\sigma_{ik}\sigma_{kj} = \sigma_{ij}$. Therefore, $\tilde{\tau}_i = \boldsymbol{L}_i'\boldsymbol{L}_i$ converges to $\chi^2_{(D-1)^2}$ in distribution.

For $t = O(n\Delta_n^2)$ and ε sufficiently small, since $S_{D_1\times D_2}(t)$ is survival function of $\chi^2_{(D-1)^2}$ distribution, there exists $\varepsilon_n > 0$ such that

$$\left|\frac{S_{D_1\times D_2}(t-\varepsilon)}{S_{D_1\times D_2}(t)} - 1\right| \le \varepsilon_n. \tag{A.10}$$

By lemma 2 and under the condition that $t = o(n^{1/3}D^{-4})$, for certain postive constant C,

$$\left|\frac{P(\tilde{\tau}_k > t-\varepsilon)}{S_{D_1\times D_2}(t-\varepsilon)} - 1\right| \le C(\varepsilon)^{3/2}n^{-1/2} = o(1). \tag{A.11}$$

Given $t \in [0, C_0 n\Delta_n^2]$, we have $\Delta_n = o(n^{-1/3})$. Under $\mathbb{H}_0$, $E(L_{k,st}) = 0$ and

$$\sup_{1\le s,t\le D} E(L_{k,st}^m) = \sum_{1\le L\le m, \sum_{l=1}^L m_l = m} Cn^{-m/2}E\{\prod_{l=1}^L (I_{k,i_l,s} - v_s)^{m_l}\}E\{\prod_{l=1}^L (I_{i_l,t} - v_t)^{m_l}\}.$$

To prove $\sup_{1\le s,t\le D} E(L_{k,st}^m) \le C$, applying (A.8) in Lemma 1, we have

$$\sup_{1\le s,t\le D} E(L_{k,st}^m) \le m!$$

and

$$\sup_{1\le s,t\le D} E(\delta_{k,st}^2) \le C(\sqrt{n}\Delta_n)^2.$$

Next, by the binomial expansion,

$$\begin{aligned}
E|R_i|^2 &= CE\big(2\sum_{s=1}^D\sum_{t=1}^D L_{k,st}\delta_{k,st} + \sum_{s=1}^D\sum_{t=1}^D \delta_{k,st}^2\big)^2 \\
&\le 2CD^4(EL_{k,st}^2)(E\delta_{k,st}^2) + E(\delta_{k,st}^4) + 4(EL_{k,st})(E\delta_{k,st}^3) \\
&\le CD^4\Delta_n^4 n^2.
\end{aligned}$$

By Markov inequality, for any $\varepsilon_0 > 0$,

$$P(|R_k| > \varepsilon_0) \le \frac{E|R_k|^2}{\varepsilon_0^2} \le CD^4\Delta_n^4 n^2\varepsilon_0^{-2}. \tag{A.12}$$

For some large M such that $S_{D_1\times D_2}(t) \ge \Delta_n$ for $0 < t < M$,

$$P(|R_k| > \varepsilon_0)/S_{D_1\times D_2}(t) \le CD^4\Delta_n^4 n^2\varepsilon_0^{-2}.$$

When C_D is a constant related to D,

$$\frac{P(R_k > \varepsilon_0)}{S_{D_1\times D_2}(t)} \le C_D\Delta_n^4 n^2\varepsilon_0^{-2}.$$

Under condition (C3) that $\Delta_n = o(n^{-1/3})$, $C_D(\Delta_n)^4 n^2\varepsilon^{-2} = o(1)$.

On the other hand, for $M < t < Cn\Delta_n^2$, we have $S_{D_1\times D_2}(t) < C\Delta_n$. Then by Lemma 2 and Condition (C2), $D \le C\log n$,

$$\begin{aligned}\frac{P(R_k > \varepsilon)}{S_{D_1\times D_2}(t)} &\le \frac{C_D\Delta_n^4 n^2\varepsilon^{-2}}{\Gamma\{(D-1)^2/2\}^{-1}M^{(D-1)^2/2-1}\exp(-n\Delta_n^2/2)}\\ &\le C_D\Delta_n^4 n^2\varepsilon^{-2}\exp(n\Delta_n^2/2).\end{aligned}$$

As a result, for $n^{-\frac{1}{2}} < \Delta_n \le (\log n^\beta/n)^{1/2}$ and $\beta > 0$, there exists $\epsilon = o(1)$ such that

$$\sup_{0\le t\le Cn\Delta_n^2}\sup_{k\in\mathbb{H}_0}\frac{P(\tau_k > t)}{S_{D_1\times D_2}(t)} \le 1+\epsilon. \tag{A.13}$$

Similarly, for any $t > 0$, there exists a positive constant c such that

$$\sup_{0\le t\le Cn\Delta_n^2}\sup_{k\in\mathbb{H}_0}\frac{P(\tau_k > t)}{S_{D_1\times D_2}(t)} \ge \sup_{i\in\mathbb{H}_0}\frac{P(\tilde\tau_k \le t+c) - P(R_i > c)}{S_{D_1\times D_2}(t)}.$$

By Lemma 2, $P(\tilde\tau_k \le t+c)/G_D(t) \le 1-\epsilon$. By $(A.13)$, $P(R_k > c)G_D(t) \le \varepsilon$. Therefore, for any $\varepsilon > 0$,

$$\sup_{0\le t\le Cn\Delta_n^2}\sup_{k\in\mathbb{H}_0}\frac{P(\tau_k > t)}{S_{D_1\times D_2}(t)} \ge 1-2\varepsilon.$$

□

Proof of Theorem 2. Write $Z_{k,st} = (I_{k,i,s}-v_s)(I^*_{i,t}-v_t)/(v_sv_t)^{1/2} - E[(I_{k,i,s}-v_s)(I^*_{i,t}-v_t)/(v_sv_t)^{1/2}]$. Since

$$|Z_{k,st}| \le \left|\frac{(I_{k,i,s}-v_s)(I^*_{i,t}-v_t)}{(v_sv_t)^{1/2}} - \eta_{k,st}\right| \le \frac{1-v_{min}}{v_{min}}$$

is bounded under Condition (C3), by Hoeffding's inequality, there exists postive constants C and C_1 such that

$$\begin{aligned}&\Pr\left\{\sup_{1\le s,t\le D}\left|n^{-1/2}\sum_{i=1}^n Z_{k,st} - n^{-1/2}\sum_{i=1}^n E(Z_{k,st})\right| > M_2\sqrt{n}\Delta_n\right\}\\ =&\Pr\left(\sup_{1\le s,t\le D}\left|L_{k,st} - n^{1/2}\eta_{k,st}\right| > M_2\sqrt{n}\Delta_n\right)\\ \le& 2CD^2\exp(-C_1n\Delta_n^2).\end{aligned}$$

Then,

$$\Pr\left(\sup_{1\le k\le p}\sup_{1\le s,t\le D}|L_{k,st}-\sqrt{n}\eta_{k,st}|>M_2\sqrt{n}\Delta_n\right)\le 2CD^2p\exp(-C_1n\Delta_n^2).$$

Note that

$$\tau_k=\sum_{s=1}^{D}\sum_{t=1}^{D}(\sqrt{n}\tau_{k,st}+L_{k,st}-\sqrt{n}\tau_{k,st}+\delta_{k,st})^2.$$

Under Condition (C4), since $\min_{k\in\mathcal{S}}|\eta_{k,st}|\ge\sqrt{2C_0'n^{-\varrho}}$, $k\in\mathbb{H}_1$, subsequently $\eta_{k,st}\ge C\Delta_n$. Under Conditions (C1)-(C2), it follows from Lemma 1 that

$$\begin{aligned}&\Pr\Big(\inf_{k\in\mathbb{H}_1}\tau_k>C_1n\Delta_n^2\Big)\\ \ge&1-\Pr\Big(\sup_{k\in\mathbb{H}_1}\sup_{1\le s,t\le D}|\delta_{k,st}|>M_1\sqrt{n}\Delta_n\Big)-\Pr\Big(\sup_{k\in\mathbb{H}_1}\sup_{1\le s,t\le D}|L_{k,st}-n^{1/2}\tau_{k,st}|\ge M_2\sqrt{n}\Delta_n\Big).\\ \ge&1-(p-q)\exp(-C_1n\Delta_n^2),\end{aligned}$$

where M_1, M_2, C_1 and C_2 are some postive constants.

Consequently, under Condition (C2), $\Pr\Big(\inf_{k\in\mathbb{H}_1}\tau_k>C_1n\Delta_n^2\Big)\to 1$ as $n\to\infty$. The proof of Theorem 1 is complete.

□

Proof of Theorem 3. Define $J_{k,i}=(J_{k,i,11},\ldots,J_{k,i,st},\ldots,J_{k,ij,(D_1-1)(D_2-1)})^\top$, where $J_{k,i,st}=\{I(\tilde{Q}_{k,s-1}<\widetilde{X}_{i,k}<\tilde{Q}_{k,s})-v_s\}\{I(\hat{Q}^*_{t-1}<Y_i<\hat{Q}^*_t)-v_t\}/(v_sv_t)^{1/2}$ and $\tilde{Q}_{k,s}$ is the s-th empirical quantile for $\widetilde{X}_{i,k}$. Denote $I(\tilde{Q}_{k,s-1}<\widetilde{X}_{i,k}<\tilde{Q}_{k,s})$ as $\tilde{I}_{i,k,s}$. Let $J_k=n^{-1/2}\sum_{i=1}^n J_{k,i}$ and $V_k=cov(J_{k,i})$. Obviously, V_k is a positive definite matrix. Therefore we have

$$\begin{aligned}\tau_{k|\mathcal{C}}&=\sum_{s=1}^{D_1}\sum_{t=1}^{D_2}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{(\hat{I}^*_{it}-v_t)(\tilde{I}_{i,k,s}-v_s)}{(v_sv_t)^{1/2}}\right]^2\\&=(H_kU_kV_k^{-1/2}n^{-1/2}\sum_{i=1}^n J_{k,i})^\top(H_kU_kV_k^{-1/2}n^{-1/2}\sum_{i=1}^n J_{k,i})\\&=(n^{-1/2}\sum_{i=1}^n J_{k,i})^\top V_k^{-1}(n^{-1/2}\sum_{i=1}^n J_{k,i}),\end{aligned}$$

where U_k is an orthogonal matrix and H_k is an tranformation matrix that satisfies $\tilde{I}_{i,k,s}-v_s=H_k(\hat{I}_{i,k,s})$. Following the definition of order statistics, $I(\tilde{Q}_{k,s-1}<\widetilde{X}_{i,k}<\tilde{Q}_{k,s})=I(s-1<r(\widetilde{X}_{i,k})<s)$, where $r(\widetilde{X}_{i,k})$ is the order for $\widetilde{X}_{i,k}$ among n observations of $\widetilde{X}_k$. The

order for $\tilde{I}_{i,k,s}$ doesn't change when permutation is conducted twice. Then $H_k^2 = I$. Also $P(\tau_{k|\mathcal{C}} \le t) = P(J_k \in \varepsilon_{k,t})$, where $\varepsilon_{k,t}$ is an $(D_1-1)(D_2-1)$ dimensional ellipsoid.

When $J_{k,i}$ follows a non-lattice distribution, by Theorem 19.2 in Bhattacharya and Rao (2010), a bounded continous density of $n^{-1/2}\sum_{i=1}^n J_{k,i}$ exists and approximate to multivariate Gaussian density function ϕ_v with mean 0 and covariance V. By Cramer-Edgeworth decomposition, it is easy to see that

$$|p_n(u) - \phi_v(u)| \le Cn^{-1/2}P_{1,J}(u)\phi_v(u),$$

where p_n is a bounded continuous density of $n^{-1/2}\sum_{i=1}^n J_{k,i}$ and $P_{1,J}(u)$ is first Cramer-Edgeworth polynomial. Since the relationship between Gaussian and Chi-square distribution, $\int_{\varepsilon_{k,t}^c}\phi_v(u)du = G_{(D_1-1)\times(D_2-1)}(t)$. Therefore, the result of Lemma 3 holds, that is

$$\left|\int_{\varepsilon_{k,t}^c} P_{1,J}(u)\phi_v(u)du\right| \le C(1+t^{3/2})(D_2-1)^3(D_1-1)^3G_{(D_1-1)\times(D_2-1)}(t).$$

Then Theorem 1 holds and under H_0, $\tau_{k|\mathcal{C}}$ follows chi-square distribution. The proof of Theorem 3 is complete. $\square$

Proof of Corollary 1. It suffices to show that

$$\widehat{\mathrm{FDR}}_{\hat{t}} = \frac{pS_{D_1\times D_2}(\hat{t})}{\max\left\{\sum\limits_{k\in\mathbb{H}} I(\tau_k > \hat{t}), 1\right\}} = \alpha. \tag{A.14}$$

Since $\sum\limits_{1\le k\le p} I(\tau_k > t)$ is monotone in t and $S_{D_1\times D_2}(t)$ is continuous, there exists a constant $0 < \hat{t} \le Cn\Delta_n^2$ such that (A.14) holds. By the definition of $\hat{t}$ in (4), $\hat{t}$ is chosen for controlling the false discovery rate at α. Under the assumption that $q/p \to 1$ as $p \to \infty$,

$$\begin{aligned}\mathrm{FDR_t} &= \frac{\sum_{k\in\mathbb{H}_0} I(\tau_k > t)}{\max\{\sum_{1\le k\le p} I(\tau_k > t), 1\}} \\ &= \frac{\sum_{k\in\mathbb{H}_0} I(\tau_k > t)/q}{\max\{\sum_{k\in\mathbb{H}} I(\tau_k > t), 1\}/p}.\end{aligned}$$

Thus, in order to prove $\widehat{\mathrm{FDR}}_{\hat{t}} \to \alpha$, it suffices to show

$$\sum_{k\in\mathbb{H}_0} I(\tau_k > \hat{t})/pS_{D_1\times D_2}(t) \to 1$$

in probability. For any positive constant ε, by Markov inequality,

$$\Pr\left\{\left|\frac{\sum_{k\in\mathbb{H}_0} I(\tau_k > t)}{pS_{D_1\times D_2}(t)} - 1\right| > \varepsilon\right\} \le \varepsilon^{-2} E\left\{\frac{\sum_{k\in\mathbb{H}_0} I(\tau_k > t)}{pS_{D_1\times D_2}(t)} - 1\right\}^2.$$

By Theorem 1, $\Pr(\tau_k > t)/S_{D_1\times D_2} \to 1$. Then,

$$\begin{aligned}
&\Pr\left\{\left|\frac{\sum_{k\in\mathbb{H}_0} I(\tau_k > t)}{pS_{D_1\times D_2}(t)} - 1\right| > \varepsilon\right\}\\
\le&\varepsilon^{-2}\frac{Var[\sum_{k\in\mathbb{H}_0} I(\tau_k > t)]}{p^2 S^2_{D_1\times D_2}(t)}\\
\le&\{\varepsilon p S_{D_1\times D_2}(t)\}^{-2}(t)\left[E\sum_{k\in\mathbb{H}_0}\sum_{j\in\mathbb{H}_0} I(\tau_k > t)I(\tau_j > t) - E\left\{\sum_{k\in\mathbb{H}_0} I(\tau_k > 0)\right\}E\left\{\sum_{j\in\mathbb{H}_0} I(\tau_j > t)\right\}\right]\\
\le&\varepsilon^{-2}p^{-2}S^{-2}_{D_1\times D_2}(t)\left[\sum_{k\in\mathbb{H}_0}\sum_{j\in\mathbb{H}_0}\Pr\{\min(\tau_k,\tau_j) > t\} - \sum_{k\in\mathbb{H}_0}\Pr(\tau_k > t)\sum_{j\in\mathbb{H}_0}\Pr(\tau_j > t)\right]\\
\le&\varepsilon^{-2}p^{-1}S^{-2}_{D_1\times D_2}(t)\sum_{k\in\mathbb{H}_0}\Pr(\tau_k > t).
\end{aligned}$$

As a result, for $0 < t < Cn\Delta_n^2$, $S_{D_1\times D_2}(t) > c_p$, where c_p is a positive constant related to p. Therefore,

$$\sum_{k\in\mathbb{H}_0} I(\tau_k > t)/pS_{D_1\times D_2}(t) \to 1.$$

The proof of Corollary 1 is complete. $\square$

Proof of Theorem 4. Define

$$\tilde{\eta}_k = \sum_{s=1}^{D_1}\sum_{t=1}^{D_2}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{(I^*_{it} - v_t)(I_{k,i,s} - v_s)}{(v_s v_t)^{1/2}}\right\}^2.$$

Recall the definition of $\hat{\eta}_k$, we have

$$\begin{aligned}
|\hat{\eta}_k - \tilde{\eta}_k| =& \left|\sum_{s=1}^{D_1}\sum_{t=1}^{D_2}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{(\hat{I}^*_{it} - v_t)(\hat{I}_{k,i,s} - v_s)}{(v_s v_t)^{1/2}}\right\}^2 - \sum_{s=1}^{D_1}\sum_{t=1}^{D_2}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{(I^*_{it} - v_t)(I_{k,i,s} - v_s)}{(v_s v_t)^{1/2}}\right\}^2\right|\\
&\le \sum_{s=1}^{D_1}\sum_{t=1}^{D_2}\frac{1}{n^2 v_s v_t}\left|\left\{\sum_{i=1}^{n}(\hat{I}^*_{it} - v_t)(\hat{I}_{k,i,s} - v_s)\right\}^2 - \left\{\sum_{i=1}^{n}(I^*_{it} - v_t)(I_{k,i,s} - v_s)\right\}^2\right|\\
&\le \sum_{s=1}^{D_1}\sum_{t=1}^{D_2}\frac{8}{n v_s v_t}\left|\sum_{i=1}^{n}\{\hat{I}^*_{i,t}\hat{I}_{k,i,s} - I^*_{i,t}I_{k,i,s} - v_s(\hat{I}^*_{i,t} - I^*_{i,t}) - v_t(\hat{I}_{k,i,s} - I_{k,i,s})\}\right|\\
&\le \sum_{s=1}^{D_1}\sum_{t=1}^{D_2}\frac{16}{n v_{\min} u_{\min}}\sum_{i=1}^{n}|\hat{I}^*_{i,t} - I^*_{i,t}| + \sum_{s=1}^{D_1}\sum_{t=1}^{D_2}\frac{16}{n v_{\min} u_{\min}}\sum_{i=1}^{n}|\hat{I}_{k,i,s} - I_{k,i,s}| = J_1 + J_2,
\end{aligned}$$

where the second inequality holds because $|\hat{I}^*_{i,t}\hat{I}_{k,i,s}|$, $|I^*_{i,t}I_{k,i,s}|$, $|\hat{I}^*_{i,t}v_s|$, $|\hat{I}_{k,i,s}v_t|$, $|I^*_{i,t}v_s|$, $|I_{k,i,s}v_t|$ and $|v_sv_t|$ all have upper bounds, the last inequality holds due to the Condition (C3). Then, by Condition (C4*), we have

$$\begin{aligned}
\Pr\left(\mathcal{S}\subset\widehat{\mathcal{M}}\right) \geq & \Pr\left(\left|\hat{\eta}_k-\sum_{s=1}^{D_1}\sum_{t=1}^{D_2}\eta^2_{k,st}\right|\leq C_0'n^{-\varrho}, \forall k\in\mathcal{S}\right)\\
\geq & \Pr\left(\max_{1\leq k\leq p}\left|\hat{\eta}_k-\sum_{s=1}^{D_1}\sum_{t=1}^{D_2}\eta^2_{k,st}\right|\leq C_0'n^{-\varrho}\right)\\
\geq & 1-p\Pr\left(\left|\hat{\eta}_k-\sum_{s=1}^{D_1}\sum_{t=1}^{D_2}\eta^2_{k,st}\right|\geq C_0'n^{-\varrho}\right)\\
\geq & 1-p\Pr\left(|\hat{\eta}_k-\tilde{\eta}_k|\geq C_0'n^{-\varrho}/2\right)-p\Pr\left(\left|\tilde{\eta}_k-\sum_{s=1}^{D_1}\sum_{t=1}^{D_2}\eta^2_{k,st}\right|\geq C_0'n^{-\varrho}/2\right)\\
\geq & 1-\Pr\left(J_1+J_2\geq C_0'n^{-\varrho}/2\right)-p\Pr\left(\left|\tilde{\eta}_k-\sum_{s=1}^{D_1}\sum_{t=1}^{D_2}\eta^2_{k,st}\right|\geq C_0'n^{-\varrho}/2\right)\\
\geq & 1-p\Pr\left(\max_{s,t}\frac{1}{n}\sum_{i=1}^{n}\left|\hat{I}^*_{i,t}-I^*_{i,t}\right|\geq\frac{C_0'n^{-\varrho}v_{\min}u_{\min}}{64D_1D_2}\right)\\
& -p\Pr\left(\max_{s,t}\frac{1}{n}\sum_{i=1}^{n}\left|\hat{I}_{k,i,s}-I_{k,i,s}\right|\geq\frac{C_0'n^{-\varrho}v_{\min}u_{\min}}{64D_1D_2}\right)\\
& -p\Pr\left[\max_{s,t}\left|\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{(I^*_{it}-v_t)(I_{k,i,s}-v_s)}{(v_sv_t)^{1/2}}\right\}^2-\eta^2_{k,st}\right|\geq\frac{C_0'n^{-\varrho}}{2D_1D_2}\right]\\
\geq & 1-pD_1D_2\Pr\left(\frac{1}{n}\sum_{i=1}^{n}\left|\hat{I}^*_{i,t}-I^*_{i,t}\right|\geq\frac{C_0'n^{-\varrho}v_{\min}u_{\min}}{64D_1D_2}\right)\\
& -pD_1D_2\Pr\left(\frac{1}{n}\sum_{i=1}^{n}\left|\hat{I}_{k,i,s}-I_{k,i,s}\right|\geq\frac{C_0'n^{-\varrho}v_{\min}u_{\min}}{64D_1D_2}\right)\\
& -pD_1D_2\Pr\left\{\left|\frac{1}{n}\sum_{i=1}^{n}(I^*_{it}-v_t)(I_{k,i,s}-v_s)-\sqrt{v_sv_t}\eta_{k,st}\right|\geq\frac{C_0'n^{-\varrho}v_{\min}u_{\min}}{16D_1D_2}\right\}.
\end{aligned}$$

Combining the proof of Lemma 1 and Hoeffding's inequality, there exist some positive constants C_1' and C_2' such that

$$pD_1D_2\Pr\left(\frac{1}{n}\sum_{i=1}^{n}\left|\hat{I}^*_{i,t}-I^*_{i,t}\right|\geq\frac{C_0'n^{-\varrho}v_{\min}u_{\min}}{64D_1D_2}\right)\leq pD_1D_2\exp\{-C_1'n^{1-2\varrho}/(D_1^2D_2^2)\},$$

$$pD_1D_2\mathrm{Pr}\left(\frac{1}{n}\sum_{i=1}^{n}\left|\hat{I}_{k,i,s}-I_{k,i,s}\right|\geq\frac{C_0'n^{-\varrho}v_{\min}u_{\min}}{64D_1D_2}\right)\leq pD_1D_2\exp\{-C_2'n^{1-2\varrho}/(D_1^2D_2^2)\}$$

Since $|(I_{it}^*-v_t)(I_{k,i,s}-v_s)|\leq 1$, we use the Hoeffding's inequality to obtain

$$\begin{aligned}&pD_1D_2\mathrm{Pr}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}(I_{it}^*-v_t)(I_{k,i,s}-v_s)-\sqrt{v_sv_t}\eta_{k,st}\right|\geq\frac{C_0'n^{-\varrho}v_{\min}u_{\min}}{16D_1D_2}\right\}\\&\leq 2pD_1D_2\exp\left(\frac{C_0'^2n^{1-2\varrho}v_{\min}^2u_{\min}^2}{128D_1^2D_2^2}\right).\end{aligned}$$

Therefore, there exists a positive constant C' such that

$$\begin{aligned}\mathrm{Pr}\left(\mathcal{S}\subset\widehat{\mathcal{M}}\right)\geq&1-O(pD_1D_2)\exp\{-C'n^{1-2\varrho}/(D_1^2D_2^2)\}\\\geq&1-O(pn^{\kappa+\xi})\exp(-bn^{1-2\varrho-2\kappa-2\xi})\\\geq&1-O(p\exp\{-bn^{1-2\varrho-2\kappa-2\xi}+(\kappa+\xi)\log(n)\}),\end{aligned}$$

where b is a positive constant. We have completed the proof of Theorem 4. □

Table 1

The quantiles of minimum model size for $p = 1000$ under Scenarios 1.1-1.4 in Section 4.1.

	$\rho = 0$					$\rho = 0.5$					$\rho = 0.9$				
Method	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%
Scenario 1.1															
QCS(8)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	6.0	8.0	9.0	10.0	12.0
QCS(9)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	7.0	8.0	9.0	10.0	12.0
QCS(10)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	7.0	8.0	9.0	10.0	13.0
FQCS	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	7.0	8.0	8.0	10.0	12.0
SIRI	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	7.0	8.0	8.0	9.0	11.0
DC-SIS	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	7.0	8.0	8.0	9.0	10.0
SIS	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	7.0	8.0	8.0	9.0	11.0
Scenario 1.2															
QCS(8)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QCS(9)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QCS(10)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
FQCS	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
SIRI	40.9	182.3	438.0	735.8	945.2	43.2	262.5	454.5	718.5	913.5	70.1	218.8	484.5	816.3	958.4
DC-SIS	3.0	3.0	3.0	4.0	6.0	3.0	3.0	3.0	3.0	5.1	5.0	7.0	8.0	9.0	12.1
SIS	190.6	450.5	729.0	852.5	972.3	291.7	466.5	692.5	830.0	971.2	123.6	363.8	615.0	762.0	968.2
Scenario 1.3															
QCS(8)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	7.0	8.8	9.0	10.0	11.0
QCS(9)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	7.0	8.0	9.0	10.0	12.0
QCS(10)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	7.0	8.0	9.0	10.0	12.1
FQCS	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	8.0	8.0	9.0	10.0	11.0
SIRI	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	8.0	9.0	10.0	10.0	12.0
DC-SIS	116.0	248.3	362.0	530.3	713.1	16.0	109.0	256.0	419.0	617.3	42.0	87.0	193.0	307.0	491.8
SIS	276.0	504.0	689.0	852.0	962.1	64.3	240.5	517.5	742.3	942.4	123.1	327.0	548.0	750.3	951.8
Scenario 1.4															
QCS(8)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	4.0	5.0	5.0	6.0
QCS(9)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	4.0	5.0	5.0	6.0
QCS(10)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	4.0	5.0	5.0	6.0
FQCS	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	4.0	5.0	5.0	6.0
SIRI	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	5.0	6.0	7.0	8.0	11.1
DC-SIS	201.8	391.3	565.0	703.8	904.2	96.2	278.5	384.0	674.8	866.6	34.8	121.5	254.0	490.0	885.5
SIS	374.6	560.0	719.0	883.3	970.2	536.4	734.8	828.0	913.3	982.3	746.9	880.0	930.0	971.3	992.0

Note: QCS(8), QCS(9) and QCS(10), our proposed method defined in (5) with different quantile grid points ($D_1 = D_2 = 8, 9, 10$); Fused, our proposed method defined in (7); SIS, the sure independence screening (Fan and Lv 2008); DC-SIS, the distance correlation based screening, (Li et al. 2012); SIRI, the sliced inverse regression via inverse modeling method (Jiang and Liu 2014).

Table 2

The quantiles of minimum model size for $p = 5000$ under Scenarios 1.1-1.4 in Section 4.1.

	$\rho = 0$					$\rho = 0.5$					$\rho = 0.9$				
Method	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%
								Scenario 1.1							
QCS(8)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	7.0	8.0	9.0	10.0	14.0
QCS(9)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	7.0	8.0	9.0	10.0	13.0
QCS(10)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	7.0	8.0	9.0	10.0	13.1
FQCS	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	7.0	8.0	9.0	10.0	12.1
SIRI	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	7.0	8.0	9.0	10.0	12.0
DC-SIS	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	7.0	8.0	8.0	10.0	12.0
SIS	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.1	7.0	8.0	8.5	10.0	12.0
								Scenario 1.2							
QCS(8)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QCS(9)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QCS(10)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
FQCS	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
SIRI	140.6	1058.5	2551.5	3760.8	4813.9	315.9	1019.8	2014.5	3470.3	4692.5	103.4	746.8	1936.5	3402.8	4709.4
DC-SIS	3.0	3.0	3.0	5.0	14.1	3.0	3.0	3.0	4.0	10.1	5.0	6.0	7.0	9.0	13.2
SIS	1464.2	2763.3	3613.5	4456.3	4874.5	1098.1	1872.3	3225.5	4471.3	4939.7	521.6	1667.0	2756.5	3730.8	4800.2
								Scenario 1.3							
QCS(8)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	7.0	8.0	10.0	11.0	13.0
QCS(9)	3.0	3.0	3.0	3.0	3.1	3.0	3.0	3.0	3.0	4.0	7.0	8.0	9.0	10.0	13.0
QCS(10)	3.0	3.0	3.0	3.0	4.0	3.0	3.0	3.0	3.0	7.1	7.0	8.0	9.0	10.0	12.1
FQCS	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.1	8.0	8.0	9.0	10.0	12.0
SIRI	3.0	3.0	3.0	3.0	3.1	3.0	3.0	3.0	3.0	4.0	8.0	9.0	10.0	11.0	12.0
DC-SIS	322.8	863.0	1644.5	2642.8	3811.0	103.7	413.8	1024.5	2077.8	3770.0	166.2	538.0	1165.0	2570.3	3506.6
SIS	1112.0	2315.3	3530.0	4277.5	4697.1	251.8	1147.3	2460.0	4001.3	4685.2	566.2	1798.8	2969.0	3928.3	4803.8
								Scenario 1.4							
QCS(8)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	4.0	5.0	5.0	6.0
QCS(9)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	4.0	5.0	5.0	6.0
QCS(10)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	4.0	5.0	5.0	6.0
FQCS	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	4.0	5.0	5.0	6.0
SIRI	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	5.0	6.0	7.0	8.0	12.0
DC-SIS	414.0	1551.8	2230.5	3353.0	4377.7	493.0	1215.3	1979.0	2724.8	4053.0	84.7	646.5	1660.0	2915.3	4505.3
SIS	1480.7	2649.5	3739.5	4374.3	4926.2	1897.6	3124.3	3966.5	4488.5	4930.8	3390.2	4156.3	4584.5	4858.3	4980.1

Note: All notations are the same as those of Table 1.

Table 3

The simulation results of the proposed FDR control procedure under Scenarios 2.1-2.3 in Section 4.2.

p	$D_1 = D_2$	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	$\lvert\widehat{\mathcal{S}}_\alpha\rvert$	$\widehat{\mathrm{FDR}}$
						Scenario 2.1							
1000	3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	10.17	0.03
	4	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	10.14	0.01
	5	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	10.12	0.01
						Scenario 2.2							
	3	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	10.23	0.03
	4	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	10.10	0.02
	5	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	10.05	0.02
						Scenario 2.3							
	3	0.95	0.95	0.94	0.97	0.94	0.95	0.96	0.96	0.95	0.96	8.76	0.06
	4	0.96	0.99	0.99	0.99	0.99	0.97	0.95	0.95	0.97	0.98	9.64	0.06
	5	0.98	0.99	1.00	0.99	0.93	0.92	0.95	0.96	0.97	0.95	9.50	0.06
						Scenario 2.1							
5000	3	0.97	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.92	9.98	0.01
	4	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	10.11	0.01
	5	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	9.99	0.01
						Scenario 2.2							
	3	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.95	9.98	0.02
	4	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.95	10.02	0.01
	5	0.98	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.93	9.93	0.01
						Scenario 2.3							
	3	0.83	0.84	0.79	0.75	0.72	0.82	0.82	0.84	0.79	0.82	7.35	0.02
	4	0.87	0.98	0.97	0.96	0.74	0.83	0.82	0.87	0.86	0.83	8.63	0.01
	5	0.81	0.97	0.96	0.98	0.74	0.84	0.83	0.84	0.77	0.74	8.35	0.02

Note: X_j: probability that the active predictors X_j is selected; $|\widehat{\mathcal{S}}_\alpha|$: average number of selected predictors; $\widehat{\mathrm{FDR}}$: average of empirical false discovery proportion.

Table 4

The simulation results of the proposed stepwise procedure under Scenarios 3.1-3.8 in Section 4.3.

		Scenario 3.1		Scenario 3.2		Scenario 3.3		Scenario 3.4	
p	Method	FN	FP	FN	FP	FN	FP	FN	FP
1000	QCS-S(8)	0.12(0.33)	0.09(0.29)	0.00(0.00)	0.00(0.00)	0.08(0.27)	0.01(0.10)	0.06(0.24)	0.00(0.00)
	QCS-S(9)	0.20(0.40)	0.12(0.36)	0.00(0.00)	0.00(0.00)	0.11(0.31)	0.00(0.00)	0.21(0.41)	0.00(0.00)
	QCS-S(10)	0.33(0.47)	0.08(0.27)	0.04(0.20)	0.00(0.00)	0.24(0.43)	0.00(0.00)	0.46(0.63)	0.00(0.00)
	SIRI	0.00(0.00)	3.35(1.71)	0.99(0.10)	7.34(1.97)	0.00(0.00)	6.60(2,01)	0.00(0.00)	5.35(1.86)
		Scenario 3.5		Scenario 3.6		Scenario 3.7		Scenario 3.8	
		FN	FP	FN	FP	FN	FP	FN	FP
1000	QCS-S(8)	0.25(0.44)	0.00(0.00)	0.25(0.46)	0.00(0.00)	0.30(0.56)	0.00(0.00)	0.32(0.53)	0.00(0.00)
	QCS-S(9)	0.25(0.44)	0.00(0.00)	0.49(0.59)	0.00(0.00)	0.54(0.61)	0.00(0.00)	0.41(0.62)	0.00(0.00)
	QCS-S(10)	0.29(0.46)	0.00(0.00)	0.65(0.66)	0.00(0.00)	0.76(0.74)	0.00(0.00)	0.50(0.64)	0.00(0.00)
	SIRI	0.00(0.00)	6.75(1.94)	0.79(0.78)	7.32(2.14)	0.00(0.00)	6.99(1.89)	0.24(0.43)	7.18(2.11)
		Scenario 3.1		Scenario 3.2		Scenario 3.3		Scenario 3.4	
p	Method	FN	FP	FN	FP	FN	FP	FN	FP
5000	QCS-S(8)	0.13(0.34)	0.06(0.24)	0.00(0.00)	0.00(0.00)	0.09(0.29)	0.00(0.00)	0.09(0.29)	0.00(0.00)
	QCS-S(9)	0.26(0.44)	0.10(0.33)	0.00(0.00)	0.00(0.00)	0.13(0.34)	0.00(0.00)	0.30(0.58)	0.00(0.00)
	QCS-S(10)	0.34(0.52)	0.07(0.26)	0.03(0.17)	0.00(0.00)	0.25(0.44)	0.00(0.00)	0.49(0.72)	0.00(0.00)
	SIRI	0.00(0.00)	6.25(1.75)	1.00(0.00)	10.33(1.86)	0.00(0.00)	9.26(1.75)	0.00 (0.00)	7.53(1.81)
		Scenario 3.5		Scenario 3.6		Scenario 3.7		Scenario 3.8	
		FN	FP	FN	FP	FN	FP	FN	FP
5000	QCS-S(8)	0.24(0.43)	0.00(0.00)	0.34(0.55)	0.00(0.00)	0.28(0.51)	0.00(0.00)	0.34(0.59)	0.00(0.00)
	QCS-S(9)	0.22(0.42)	0.00(0.00)	0.44(0.62)	0.00(0.00)	0.41(0.59)	0.00(0.00)	0.32(0.53)	0.00(0.00)
	QCS-S(10)	0.26(0.44)	0.00(0.00)	0.61(0.75)	0.00(0.00)	0.72(0.68)	0.00(0.00)	0.47(0.66)	0.00(0.00)
	SIRI	0.00(0.00)	9.55(1.94)	1.08(0.72)	10.68(2.14)	0.01(0.1)	9.45(1.86)	0.45(0.58)	10.36(1.87)

Note: QCS-S(8), QCS-S(9) and QCS-S(10), our proposed procedure defined in Section 4.3 with different quantile grid points ($D_1 = D_2 = 8, 9, 10$); SIRI, the sliced inverse regression via inverse modeling method (Jiang and Liu 2014).FN: average number of true active predictors falsely excluded as irrelevant predictors; FP: average number of irrelevant predictors falsely selected as true active predictors; The number in brackets are the standard deviation of distribution for FN and FP, respectively.

Table 5

Mean square prediction errors (MSPE) for the rat eye data with different methods.

	IP	IPDC	SIRI	QCS-FDR
Mean	1.41	1.46	1.55	1.40
Sd	0.55	0.73	0.97	0.66

Note: Mean, the mean of MSPE; Sd, the standard deviation of MSPE.

References

[1] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc.* B **57**, 289–300.

[2] Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29**, 1165–1188.

[3] Cai, T. T. and Liu, W. D. (2016). Large scale multiple testing of correlations. *Ann. Appl. Statist.* **10**, 596–617.

[4] Chiang, A. P., Beck, J. S., Yen, H. J., Tayeh, M. K., Scheetz, T. E., Swiderski, R. E., Nishimura, D. Y., Braun, T. A., Kim, Kim, K.-Y., Huang, J., Elbedour, K., Carmi, R., Slusarski, D. C., Casavant, T. L., Stone, E. M. and Sheffield, V. C. (2006). Homozygosity mapping with SNP arrays identifies a novel gene for Bardet-Biedl syndrome gene (bbs11). *Proc. Natl. Acad. Sci.* **103**, 6287–6292.

[6] Chu, W., Li, R. and Reimherr, M. (2016). Feature screening for time-varying coefficient models with ultrahigh dimensional longitudinal data. *J. Am. Statist. Assoc.* **109**, 815–827.

[6] Cook, R. D. and Zhang, X. (2014). Fused estimators of the central subspace in sufficient dimension reduction. *J. Am. Statist. Assoc.* **109**, 815–827.

[7] Cui, H., Li, R. and Zhong, W. (2015). Model-free feature screening for ultrahigh dimensional discriminant analysis. *J. Am. Statist. Assoc.* **110**, 630–641.

[8] Efron, B. (2007). Correlation and large-scale simultaneous significance testing. *J. Am. Statist. Assoc.* **102**, 93–103.

[9] Efron, B. (2004). Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *J. Am. Statist. Assoc.* **99**, 96–104.

[10] Jerome H. Friedman (1989). Regularized Discriminant Analysis. *J. Am. Statist. Assoc.* **84**, 165–175.

[11] Golub, T.R., Slonim D.K., Tamayo, P., Huard, C., Gaasenbeek, M., J.P, Mesirov, H, Coller, M.L, Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield and E. S. Lander (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531–537.

[12] Fan, J., Feng, Y. and Song, R (2011). Nonparametric independence screening in sparse ultrahigh-dimensional additive models. *J. Am. Statist. Assoc.* **106**, 544–557.

[13] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Statist. Assoc.* **96**, 1348–1360.

[14] Fan, J. and Lv, J (2008). Sure independence screening for ultrahigh dimensional feature space. *J. R. Statist. Soc.* B **57**, 5467–5484.

[15] Fan, J. and Lv, J (2011). Nonconcave penalized likelihood with NP-dimensionality. *IEEE Trans. Inf. Theory* **70**, 849–911.

[16] Fan, J. and Samworth, R. (2009). Ultrahigh dimensional variable selection: beyond the linear model. *J. Mach. Learn. Res.* **10**, 1829–1853.

[17] Fan, J. and Song, R. (2010). Sure independence screening in generalized linear models with NP-dimensionality. *Ann. Statist.* **38**, 3567–3604.

[18] Hall, P. and Xue, J. H. (2014). On selecting interacting features from high-dimensional data. *Comput. Stat. Data. Anal.* **71**, 694–708.

[19] Hao, N., Feng, Y. and Zhang, H. H. (2018). Model selection for high dimensional quadratic regression via regularization. *J. Am. Statist. Assoc.* **113**, 615–625.

[20] Hao, N. and Zhang, H. H. (2014). Interaction screening for ultra-high dimensional data. *J. Am. Statist. Assoc.* **109**, 1285–1301.

[21] He, X., Wang, L. and Hong, H. G. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *Ann. Statist.* **41**, 342–369.

[22] Huang, D., Li, R. and Wang, H. (2014). Feature screening for ultrahigh dimensional categorical data with applications. *J. Bus. Econ. Stat.* **32**, 237–244.

[23] Huang, J., Ma, S. and Zhang, C.H . (2014). Adaptive lasso for sparse high-dimensional regression models. *Stat. Sin.* **18**, 1603–1618.

[24] Hsing, T. and Carroll, R. J. (2014). An asymptotic theory for sliced inverse regression.. *Ann. Statist.* **20**, 1040-1061.

[25] Jiang, B. and Liu, J. S. (2014). Variable selection for general index models via sliced inverse regression. *Ann. Statist.* **42**, 1751–1786.

[26] Li, G., Peng, H., Zhang, J. and Zhu, L. (2012). Robust rank correlation based screening. *Ann. Statist.* **40**, 1846–1877.

[27] Li, K. C. (1991). Sliced inverse regression for dimension reduction. *J. Am. Statist. Assoc.* **86**, 316–327.

[28] Li, R., Zhong, W. and Zhu, L. (2012). Feature screening via distance correlation learning. *J. Am. Statist. Assoc.* **107**, 1129–1139.

[29] Ma, S. J., Li, R. and Tsai, C.L. (2017). Variable screening via quantile partial correlation. *J. Am. Statist. Assoc.* **112**, 650–663.

[30] Mai, Q. and Zou, H. (2015). The kolmogorov filter for variable screening in high-dimensional binary classification. *Biometrika* **100**, 229–234.

[31] Mai, Q. and Zou, H. (2015). The Fused Kolmogorov Filter: a nonparametric model-free screening method. *Ann. Statist.* **43**, 1471–1497.

[32] R. J. Hyndman, A. and Y. Fan (1996). Sample Quantiles in Statistical Packages. *Am. Stat.* **50**, 361–365.

[33] Storey, J. D. (2002). A direct approach to false discovery rates. *J. R. Statist. Soc.* B **64**, 479–498.

[34] Sun, W. and Cai, T. T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *J. Am. Statist. Assoc.* **102**, 901–912.

[35] Sun, W. and Cai, T. T. (2009). Large-scale multiple testing under dependence. *J. R. Statist. Soc.* B **71**, 393–424.

[36] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Statist. Soc.* B **58**, 267–288.

[37] Xie, J. and Li, R. (2019). False discovery rate control for high-dimensional networks of quantile associations conditioning on covariates. *J. R. Statist. Soc.* B (in press, https://doi.org/10.1111/rssb.12288).

[38] Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Statist. Soc.* B **68**, 49–67.

[39] Zhao, P. and Yu, B. (2006). On model selection consistency of lasso. *J. Mach. Learn. Res.* **7**, 2541–2563.

[40] Zhong, W., Zhang, T, Zhu, Y. and Liu, J. S. (2012). Correlation pursuit: forward stepwise variable selection for index models. *J. R. Statist. Soc.* B **74**, 849–870.

[41] Zhou, T., Zhu, L., Xu, C. and Li, R. (2019). Model-free forward screening via vumulative divergence. *J. Am. Statist. Assoc.*, 1–13. doi:10.1080/01621459.2019.1632078

[43] Zhu, L. P., Li, L., Li, R. and Zhu, L. X. (2011). Model-free feature screening for ultrahigh-dimensional data. *J. Am. Statist. Assoc.* **106**, 1464–1475.

[43] Zhu, L. X. and Ng, K. W. (1995). Asymptotics of sliced inverse regression. *Stat. Sin.* **5**, 727–736.

[44] Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Am. Statist. Assoc.* **101**, 1418–1429.