

Article

Quantile-Adaptive Sufficient Variable Screening by Controlling False Discovery

Zihao Yuan ¹, Jiaqing Chen ^{1,*}, Han Qiu ¹ and Yangxin Huang ²¹ Department of Statistics, Wuhan University of Technology, Wuhan 430070, China² Department of Epidemiology and Biostatistics, University of South Florida, Tampa, FL 33612, USA

* Correspondence: jqchenwhut@163.com

Abstract: Sufficient variable screening rapidly reduces dimensionality with high probability in ultra-high dimensional modeling. To rapidly screen out the null predictors, a quantile-adaptive sufficient variable screening framework is developed by controlling the false discovery. Without any specification of an actual model, we first introduce a compound testing procedure based on the conditionally imputing marginal rank correlation at different quantile levels of response to select active predictors in high dimensionality. The testing statistic can capture sufficient dependence through two paths: one is to control false discovery adaptively and the other is to control the false discovery rate by giving a prespecified threshold. 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: high dimensionality; sufficient variable screening; false discovery controlling; quantile heterogeneity



Citation: Yuan, Z.; Chen, J.; Qiu, H.; Huang, Y. Quantile-Adaptive Sufficient Variable Screening by Controlling False Discovery. *Entropy* **2023**, *25*, 524. <https://doi.org/10.3390/e25030524>

Academic Editors: Diego Oliva and Ali Rıza Yıldız

Received: 29 December 2022

Revised: 6 March 2023

Accepted: 14 March 2023

Published: 17 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

When the dimension p grows exponentially with n , the unbearable computational cost of the classical variable selection incurred by the ultra-high dimensionality will not only heavily slow down the computing speed of the algorithm but also result in unstable solutions [1]. To rapidly screen out the inactive predictors, variable screening methods for ultra-high dimensional data coupled with response have been examined to reduce dimension and effectively retain all the active variables with high probability in the reduced variable space [2]. This is referred to as the sure screening property. Fan and Lv (2008) proposed the concept of sure independence screening (SIS) based on the marginal Pearson correlation coefficient in the linear regression model [1]. Since then, a series of variable screening methods have been proposed successively, such as variable screening frameworks based on the generalized linear model, additive model, and general models [3–6]. The above methods are based on some specific model assumptions. In many scientific applications, the correlation between predictors and the response is difficult to assume for ultra-high dimensional data [7]. Model-based screening procedures enjoy quick computational speed but suffer the risk of model misspecification [7,8].

To avoid the inconsistency between the assumptions of the regression model and the actual distribution of data, model-free variable screening methods have been initially designed for the continuous outcome variables. For example, Zhu et al. (2011) reported a screening procedure to detect active predictors named SIRS; Li et al. (2012) proposed a distance-based sure screening procedure called DC-SIS; He et al. (2013) introduced a quantile-adaptive model-free variable screening for high dimensional data; Lin et al. (2013) discussed a nonparametric ranking feature screening (NRS) through local information flows of the predictors, and Lu and Lin (2020) studied feature screening procedure based

on the unconditional model [7–11]. For ultra-high dimensional covariates coupled with the categorical response, Mai and Zou (2013) advocated the Kolmogorov–Smirnov distance for binary classification problems [12]. With a possibly diverging number of classes, the marginal feature screening procedure for ultra-high dimensional discriminant analysis was introduced by Huang et al. (2014) and Cui et al. (2015) [13,14]. Han (2019) researched a general and unified nonparametric screening framework under conditional strictly convex loss [15]. Zhou et al. (2020) established a forward screening procedure based on a new measure called cumulative divergence [16]. Xie et al. (2020) explored a category-adaptive screening procedure with ultrahigh dimensional heterogeneous categorical data [17].

As reported in Hao and Zhang (2017), the variable screening results depend on the signal-to-noise (SNR) level: when the signal is weak with massive noise variables, it could not be easy to detect the active variables from the noise variables, and the sure screening property may not be established [18]. In terms of this situation, one path considers controlling some false discoveries. In this regard, Tang et al. (2021) explored a quantile correlation-based screening framework (QCS), which could screen variables by conducting multiple testing to control the false discovery rate (FDR) [19]. Liu et al. (2022) posed a two-step approach to specify the threshold for feature screening with the help of knockoff features such that the FDR is controlled under a prespecified level [20]. For controlling the FDR, Guo et al. (2022) advocated a data-adaptive threshold selection procedure with FDR control based on sample-splitting [21].

However, most of the above sure screening methods are not sufficient variable screening (SVS) technology, which was first proposed in Cook (2004) and also discussed by Yin and Hilafu (2015) and Yuan et al. (2022) [22–24]. For illustration, consider a population with a response variable Y and a p -dimensional vector of predictors $X = (X_1, \dots, X_p)^T$, let $X_{\mathcal{A}}$ be a subset of X , and $X_{\mathcal{A}^c}$ denotes the orthogonal complement of $X_{\mathcal{A}}$. Based on the research of Yin and Hilafu (2015) and Yuan et al. (2022), sufficient variable screening means to find the smallest and unique active variable set $X_{\mathcal{A}}$ such that $Y \perp\!\!\!\perp X_{\mathcal{A}^c} \mid X_{\mathcal{A}}$, that is, given the set $X_{\mathcal{A}}$, Y is independent of $X_{\mathcal{A}^c}$ [23,24].

In this paper, without any specific regression or parametric assumptions, we advocate a new sufficient variable screening procedure by using a robust multiple testing procedure by controlling the false discovery to distinguish active variables by splitting the continuous response at different quantile levels. Thus, we achieve quantile-adaptive sufficient variable screening by controlling the false discovery (QA-SVS-FD). The proposed procedure is based on a one-versus-rest (OVR) test statistic with an asymptotic chi-square distribution under the null hypothesis. Thus, with the asymptotic distribution, the sufficient variables set could be estimated precisely by controlling the FDR accurately at a given level for high dimensionality or controlling the number of the FD adaptively with the error of 1. In addition, the proposed procedure is a model-free method for the measurement of independence without any specified distribution model; thus, it is robust to detect sufficient relevant variables against different model types.

The rest of this paper is organized as follows: Section 2 develops the sufficient variable screening testing statistic by using the conditionally imputing marginal rank correlation at different quantile levels of response. The false discovery controlling procedure under mild conditions will be studied in Section 3. Sections 4 and 5 evaluate the proposed procedure's performance via extensive numerical research, which contains simulation studies and two real data examples, which verify the robustness and flexibility of our methods. In Section 6, we shall give a short concluding discussion. All the theoretical properties are proved in Appendix A.

2. Sufficient Screening Utility

As the statement of Yuan et al. (2022), the analysis of sufficient variable screening includes the iterative two-step screening procedure, which contains complex computation [24]. This section will propose a novel sufficient variable screening statistic by the quantile-adaptive correlation test (QA-SVS).

2.1. A Quantile-Adaptive Correlation Test Statistic

Throughout this paper, denote (Y, \mathbf{X}^\top) as a pair of the continuous scalar of response and the p -dimensional vector of covariates, where $\mathbf{X}^\top = (X_1, \dots, X_p)^\top$. The complete observations $\{Y_i, \mathbf{X}_i^\top\}$, $i = 1, \dots, n$ are independent and identically distributed observation samples of (Y, \mathbf{X}^\top) . Based on the research of Yin and Hilafu (2015) [23] and Yuan et al. (2022) [24], define two indicant sets, i.e, sufficient active or active variables index set \mathcal{A} consists of variables that are relevant to Y , and \mathcal{A}^c contains all redundant predictors or noise predictors, where $\mathbf{X}_{\mathcal{A}}$ is a subset of \mathbf{X} , and $\mathbf{X}_{\mathcal{A}^c}$ denotes the orthogonal complement of $\mathbf{X}_{\mathcal{A}}$. Sufficient variable screening means to find the smallest and unique active variable set $\mathbf{X}_{\mathcal{A}}$ such that $Y \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}^c} \mid \mathbf{X}_{\mathcal{A}}$, that is, given the set $\mathbf{X}_{\mathcal{A}}$, Y is independent of $\mathbf{X}_{\mathcal{A}^c}$. Clearly, full index set $\mathcal{I} = \mathcal{A} \cup \mathcal{A}^c = \{1, 2, \dots, p\}$. Without any specific assumption on the functional correlation between the covariates and the response, inspired by Tang et al. (2021) [19], one may consider testing the sufficient independence simultaneously between each X_j and Y for $1 \leq j \leq p$ to detect important variables in the high dimensional setting

$$\mathbb{H}_{0,j} : Y \perp\!\!\!\perp X_j \mid \mathbf{X}_2 \quad \text{versus} \quad \mathbb{H}_{1,j} : Y \not\perp\!\!\!\perp X_j \mid \mathbf{X}_2 \tag{1}$$

for $X_j \in \mathbf{X}_1$, where $\perp\!\!\!\perp$ represents independence, and $(\mathbf{X}_1, \mathbf{X}_2)^\top$ is a split of \mathbf{X} . When the part \mathbf{X}_2 satisfies that $\mathbf{X}_{\mathcal{A}} \supseteq \mathbf{X}_2$, $X_j \in \mathbf{X}_1$ is regarded as a sufficient active predictor if and only if $H_{0,j}$ is rejected.

Recall that the category-adaptive screening procedure in Xie et al. (2020) [17] is to measure the marginal independence in high-dimensional heterogeneous data. Motivated by the marginal utility, a test statistic is introduced based on the quantile split response for testing (1) with high dimensionality without assuming any parametrical correlation between Y and \mathbf{X} . Let $0 = \gamma_0 < \gamma_1 < \gamma_2 < \dots < \gamma_K = 1$ be the sequence of quantile grid points of Y , where D is prespecified positive integers. Denote the γ_k -th ($1 \leq k \leq K$) quantile of Y as Q_k . The theoretical quantiles can be estimated consistently by the sample quantiles, which is that $\hat{Q}_k = (1 - \gamma_k)Y_{[n\gamma_k]} + \gamma_k Y_{[n\gamma_k]+1}$ is the γ_k -th ($1 \leq k \leq K - 1$) sample quantile of Y according to Hyndman and Fan (1996) [25]. For convenience, let $\hat{Q}_0 = -\infty$ and $\hat{Q}_K = +\infty$, denote that $G_k = [Q_{k-1}, Q_k)$ and $\hat{G}_k = [\hat{Q}_{k-1}, \hat{Q}_k)$. It is obvious that $p_k = \Pr(Y \in G_k) = \gamma_k - \gamma_{k-1}$.

To explore the nature and provide a complete picture of the conditional distribution of the outcome variable given the predictor vector, we assume the following assumptions:

- (I) (Quantile-Heterogeneity) the index set of sufficient active predictors satisfies that $\mathcal{A}_k = \{1 \leq j \leq p : \Pr(Y \in G_k \mid \mathbf{X}) \text{ functionally depends on } X_j\}$ may be different for different $k = 1, 2, \dots, K$;
- (II) (Sparsity) the dimensionality $p = o\{\exp(n^\alpha)\}$ for some constant $\alpha > 0$, but $|\mathcal{A}_k| = s_k = o(n)$, where $|\mathcal{A}_k|$ is the cardinality of \mathcal{A}_k , and n is the sample size.

Assumption (I) describes heterogeneity of the sufficient correlated predictors set at different level quantiles of the response, and Assumption (II) provides the high-dimensional settings.

Remark 1. The sufficient active variables index set at the γ_k -th quantile defined in Assumption (I) screens active variables sufficiently when the response belongs to the grid $[Q_{k-1}, Q_k)$. That is, $\mathcal{A}_k^c = \{1 \leq j \leq p : I(Y \in G_k) \not\perp\!\!\!\perp X_j \mid \mathcal{A}_k\}$, where \mathcal{A}_k^c is the orthogonal complement of \mathcal{A}_k . The proof of equivalence will be expanded in Appendix A.1. In other words, $\mathcal{A} = \{1 \leq j \leq p : \mathbb{H}_{0,j} \text{ is true}\} = \cup_{k=1}^K \mathcal{A}_k$.

Remark 2. If response $Y \in \{y_1, y_2, \dots, y_K\}$ is discrete, then \mathcal{A}_k in Assumption (I) should be rewritten as $\mathcal{A}_k = \{1 \leq j \leq p : \Pr(Y = y_k \mid \mathbf{X}) \text{ functionally depends on } X_j\}$.

Remark 3. For the special case of $K = 2$, by the definition of \mathcal{A}_k in Assumption (I), it is clear that $\mathcal{A}_1 = \mathcal{A}_2$, which reduces to the active set $\mathcal{A}_\gamma = \{1 \leq j \leq p : Q_\gamma(Y \mid \mathbf{X}) \text{ functionally depends on } X_j\}$.

$X_j\}$ defined in He et al. (2013). We do not elaborate on it again as we regard it as a special case of the proposed framework in this paper.

Lemma 1. For any $j = 1, \dots, p$ and $k = 1, \dots, K$, if $j \notin \mathcal{A}_k$, then $F_{jk}(x) = F_j(x)$, where $F_{jk}(x) = \Pr(X_j \leq x | Y \in G_k)$ and $F_j(x) = \Pr(X_j \leq x)$.

Lemma 1 will be proved in Appendix A.2. According to Yuan et al. (2022) [24], the sufficient active variable set is actually screened based on the structure of $(Y, \mathbf{X}_{\mathcal{A}_k}, \mathbf{X}_{\mathcal{A}_k^c})$. Lemma 1 provides the information that the structure of $(Y, \mathbf{X}_{\mathcal{A}_k}, \mathbf{X}_{\mathcal{A}_k^c})$ could be transformed into another structure based on the marginal structure of (Y, X_j) by judging the difference of $F_{jk}(x)$ and $F_j(x)$ for each $j = 1, \dots, p$ and $k = 1, \dots, K$.

In terms of the quantile-heterogeneity of response, consider a series of tests to detect sufficient active variables simultaneously at different quantile levels, that is,

$$\mathbb{H}_{0,j,k} : I(Y \in G_k) \perp\!\!\!\perp X_j \mid \mathbf{X}_{\mathcal{A}_k} \quad \text{versus} \quad \mathbb{H}_{1,j,k} : I(Y \in G_k) \not\perp\!\!\!\perp X_j \mid \mathbf{X}_{\mathcal{A}_k} \tag{2}$$

for $1 \leq j \leq p$ and $1 \leq k \leq K$. Rewrite the test in Equation (1) as

$$\mathbb{H}_{0,j,k} : Y \perp\!\!\!\perp X_j \mid \mathbf{X}_{\mathcal{A}} \quad \text{versus} \quad \mathbb{H}_{1,j,k} : Y \not\perp\!\!\!\perp X_j \mid \mathbf{X}_{\mathcal{A}}, \tag{3}$$

where $\mathcal{A} = \cup_{k=1}^K \mathcal{A}_k$. To investigate the difference of the conditional distribution of X_j ($j = 1, \dots, p$) different quantile levels, for given $k \in \{1, \dots, K\}$, a variable screening approach developed by capturing the dependence between $I(Y \in G_k)$ and x_j is specified as the following screening utility:

$$v_{jk} = 12 \cdot (n + 1) \cdot \frac{p_k}{1 - p_k} \cdot \tau_{jk}^2, \tag{4}$$

where $\tau_{jk} = \mathbb{E}_{X_j}\{F_{jk}(x) - F_j(x)\} = \mathbb{E}_x\{F_{jk}(x)\} - \frac{1}{2}$, and $F_{jk}(x) - F_j(x)$ reflects the difference between conditional cumulative distribution function (CDF) and marginal CDF of X_j at each quantile level. Actually, $v_{jk} = 12 \cdot (n + 1) \cdot \text{Var}_{\text{OVR}}\{\tau_{jk}\} = 12 \cdot (n + 1) \cdot (p_k \cdot \tau_{jk} + (1 - p_k) \cdot \tau_{jk})$, where $\tau_{j\bar{k}} = \mathbb{E}_{X_j}\{F_{j\bar{k}}(x) - F_j(x)\} = \mathbb{E}_x\{F_{j\bar{k}}(x)\} - \frac{1}{2}$, $F_{j\bar{k}}(x) = \Pr(X_j \leq x | Y \notin [Q_{k-1}, Q_k])$, and $\text{Var}_{\text{OVR}}\{\tau_{jk}\}$ represents the variance of τ_{jk} in the one-versus-rest test of k -th series of $\mathbb{H}_{0,j,k}$. As the definition in Equation (4), the higher v_{jk} represents the stronger correlation between the variable X_j and $I(Y \in G_k)$.

Then, under the complete sample (\mathbf{X}, Y) , a sample estimate of $\tau_{jk}, j = 1, \dots, p, k = 1, \dots, K$ is suggested as

$$\hat{\tau}_{jk} = \frac{1}{n + 1} \sum_{l=1}^n \left\{ \frac{1}{n} \sum_{i=1}^n \frac{I(X_{ij} \leq x_{lj}, Y_i \in \hat{G}_k)}{\hat{p}_k} \right\} - \frac{1}{2} = \frac{1}{(n + 1)n_k} \sum_{i=1}^n \eta_i \cdot \zeta_{ij} - \frac{1}{2},$$

where $\eta_{ik} = I(Y_i \in \hat{G}_k)$ represents whether the response of the i -th sample belongs to the k -th grid, $\hat{p}_k \triangleq n_k/n = n^{-1} \sum_{i=1}^n \eta_{ik}$, and $k = 1, \dots, K$. Correspondingly, the test statistic specifies the following conditional rank correlation, which is

$$\hat{v}_{jk} = 12 \cdot (n + 1) \cdot \frac{n_k}{n - n_k} \hat{\tau}_{jk}^2 = 12 \cdot (n + 1) \cdot \frac{n_k}{n - n_k} \left(\frac{1}{(n + 1)n_k} \sum_{i=1}^n \hat{\eta}_i \cdot \zeta_{ij} - \frac{1}{2} \right)^2. \tag{5}$$

2.2. Asymptotic Properties of the Test Statistic

According to approximation distribution for a sample sum in sampling without replacement from a finite population in Mohamed and Mirakhmedov (2016) [26], the asymptotic properties of $\hat{\tau}_{jk}$ and \hat{v}_{jk} would be obtained as follows.

Lemma 2 (Asymptotic Distribution of $\hat{\tau}_{jk}$). *if $\mathbb{H}_{0,j,k}$ is true for all $k = 1, 2, \dots, K$ and $j = 1, 2, \dots, p$, and $\lim_{n \rightarrow \infty} \hat{p}_k(1 - \hat{p}_k) > 0$, then we obtain the following asymptotic distribution:*

$$\hat{\tau}_{jk} \xrightarrow{L} \mathcal{N}\left(0, \frac{1 - \hat{p}_k}{12(n + 1)\hat{p}_k}\right).$$

Denote $\Phi(u) = 2\pi^{-1/2} \int_{-\infty}^u \exp\{-\frac{v^2}{2}\} dv$; then, $F_{\hat{\tau}_{jk}}(u) = \Pr\{\hat{\tau}_{jk} < u \sqrt{\frac{1 - \hat{p}_k}{12(n + 1)\hat{p}_k}}\}$ satisfies that

$$\frac{1 - F_{\hat{\tau}_{jk}}(u)}{1 - \Phi(u)} = \frac{F_{\hat{\tau}_{jk}}(-u)}{\Phi(-u)} = 1 + O(n^{-1/2}),$$

where $u = O(n^\beta) > 0$ and $\beta \leq 1/2$.

Corollary 1 (Asymptotic Distribution of \hat{v}_{jk}). *If $\mathbb{H}_{0,j,k}$ is true for any $k = 1, 2, \dots, K$ and $j = 1, 2, \dots, p$, and $\lim_{n \rightarrow \infty} \hat{p}_k(1 - \hat{p}_k) > 0$, we have $\hat{v}_{jk} \xrightarrow{D} \chi^2_1$, where χ^2_m follows the chi-square distribution with degree of freedom m . If $\mathbb{H}_{0,j}$ is true for any $j = 1, 2, \dots, p$, we obtain $\hat{v}_j \xrightarrow{D} \chi^2_{K-1}$, where $\hat{v}_j = (K - 1) / K \cdot \sum_{k=1}^K \hat{v}_{jk}$. Then, \hat{v}_{jk} and \hat{v}_j satisfies that*

$$\frac{F_{\hat{v}_{jk}}(u)}{F_{\chi^2_1}(u)} = 1 + O(n^{-1/2}), \quad \frac{F_{\hat{v}_j}(u)}{F_{\chi^2_{K-1}}(u)} = 1 + O(n^{-1/2}),$$

where $F_{\chi^2_{K-1}}(u) = \Pr(\chi^2_{K-1} \leq u)$ is the CDF of χ^2_{K-1} .

Lemma 2 and Corollary 1 will be proved in Appendices A.3 and A.4, respectively. According to Lemma 2, it could be found that the asymptotic normal distribution of $\hat{\tau}_{jk}$ depends on \hat{p}_k . Thus, to remove the influence of \hat{p}_k on asymptotic distribution and consider the composite hypothesis testing in (3), \hat{v}_{jk} is established in this paper.

When giving additional conditions as below, we can obtain Theorem 1 and 2.

- (C1) There exists constants $c_1 > 0$ and $c_2 > 0$, s.t. $c_1 / K \leq \min_{1 \leq k \leq K} p_k \leq \max_{1 \leq k \leq K} p_k \leq c_2 / K$;
- (C2) There exists $\rho_0 = O(1/p^2) > 0$, s.t. $\lim_{p \rightarrow \infty} \min_{j_1 \in \mathcal{A}_{k_1}} v_{j_1 k_1} \geq \rho_0 \geq \liminf_{p \rightarrow \infty} \max_{j_2 \notin \mathcal{A}_{k_2}} v_{j_2 k_2}$;
- (C3) The grids number of response satisfies $K = O(n^\zeta)$, where $\zeta > 0$ and $\kappa + \zeta < 1/2$.

Condition (C1) requires that the proportion of samples in each grid should not be too small or too large. Condition (C2) guarantees that there exist thresholds ρ_0 ensuring that the smaller value of $v_{jk} \leq \rho_0$ represents the weaker correlation. Condition (C3) allows the number of grids to diverge as n increases. This ensures the rationality of the series of hypothesis tests. Conditions of (C1)–(C3) provide concise foundations and do not specify any distribution models and moment assumptions of variables.

Theorem 1 (Sure Screening Property). *Suppose conditions of (C1) and (C3) hold. Then, for any constant $c_3 > 0$, we have*

$$\Pr(\max_{1 \leq j \leq p} |\hat{v}_{jk} - v_{jk}| \geq c_3 n^{-\kappa}) \leq 8p \exp(-c_4 n^{1-2\kappa-2\zeta}), \tag{6}$$

where $c_4 > 0$ and $k = 1, \dots, K$.

Theorem 2 (Ranking Consistency Property). *Suppose conditions of (C1) and (C2) hold. If $K \log(p) = o(np_0^2)$, then*

$$\liminf_{n \rightarrow \infty} \left\{ \min_{j_1 \in \mathcal{A}_{k_1}} \hat{v}_{j_1 k_1} - \max_{j_2 \notin \mathcal{A}_{k_2}} \hat{v}_{j_2 k_2} \right\} > 0. \tag{7}$$

We shall provide the proofs of Theorems 1 and 2 in Appendices A.5 and A.6, respectively. Note that Theorem 1 is established for the fixed number of variables p . As long as $4(n + 2)p \exp(-c_4 n^{1-2\kappa-\xi})$ to 0 asymptotically, the sure screening property of QA-SVS is robust to heavy-tailed distributions of the predictors and the presence of potential outliers. The ranking consistency property in Theorem 2 indicates that the values of \hat{v}_{jk} of sufficient active variables responding to the k -th grid can be ranked above that of all the inactive ones with a high probability, which implies that the QA-SVS can separate the active and inactive with a certain threshold. Theorems 1 and 2 mainly illustrate the properties of the marginal utility itself, and the estimation of the certain threshold for the partition of sufficient variable sets needs further work in Section 3.

3. False Discovery Control Model

Based on Theorems 1 and 2, we shall design two routes for screening sufficient active variables by considering the false discovery (FD): one is to control the cardinality of FD adaptively by detecting the outlier, and the other is to control false discovery rate (FDR) accurately by survival analysis function.

3.1. Adaptive False Discovery Control Model

Recall the property (iii) of controlling the false discovery in Xie et al. (2020)–[17], the number of screened variables is bounded with high probability. If we let ρ satisfy that $S_{\chi_1^2}(\rho) = 1 - F_{\chi_{k-1}^2}(\rho) = 1/p$, where ρ indicates the $1/p$ -th upper quantile of χ_1^2 , we can capture adaptive path to control the false discovery by the outlier method. Without assuming any actual distribution, using the proposed test statistic \hat{v}_{jk} , the false discovery is

$$FD_{k,\rho} = \sum_{j \in \mathcal{A}_k^c} I(\hat{v}_{jk} \geq \rho),$$

where the expectation of false discovery rate is $EFD_{k,\rho} = \mathbb{E}(FD_{k,\rho})$, and the variance of false discovery rate is $VFD_{k,\rho} = \text{Var}(FD_{k,\rho})$ for any given ρ . By Corollary 1 in Section 2.2, under $\mathbb{H}_{0,j,k}$, each \hat{v}_{jk} converges to distribution of χ_1^2 under certain mild conditions. Let $q_k = 1 - s_k$ be the cardinality of \mathcal{A}_k^c , for any given ρ , we have $s_r/p \rightarrow 0$ as $p \rightarrow \infty$. Intuitively, the $FDR_{k,\rho}$ could be estimated by

Theorem 3 (Adaptively Controlling False Discovery). *Suppose conditions of (C1)–(C3) hold. A fixed $\rho = \hat{\rho}_0$ satisfying $S_{\chi_1^2}(\hat{\rho}_0) = 1/p = O(n^{-2\beta})$, where $\beta > 1/2$, we obtain that*

$$\lim_{p \rightarrow \infty} Pr\{FD_{k,\hat{\rho}_0} > 0\} = 1 - e^{-1} \quad (k = 1, \dots, K), \tag{8}$$

$$\lim_{n \rightarrow \infty} EFD_{k,\hat{\rho}_0} = \lim_{n \rightarrow \infty} VFD_{k,\hat{\rho}_0} = 1, \tag{9}$$

where $EFD_{k,\hat{\rho}_0} = 1 + O(n^{-1/2})$ and $VFD_{k,\hat{\rho}_0} = 1 + O(n^{-1/2})$.

We shall prove this property in Appendix A.7. Theorem 3 implies that the adaptive threshold $\hat{\rho}_0$ can separate the active and inactive variables with a low false discovery in a high probability, which converges to $1 - e^{-1}$ as n increases. The expectation and variance of the FD number can be controlled at $1 + O(n^{-1/2})$, indicating that the number of the selected variables can be sufficiently controlled. The sufficient screened set is defined as

$$\hat{\mathcal{A}}_{k,\hat{\rho}_0} \equiv \{j : \hat{v}_{jk} \geq \hat{\rho}_0, 1 \leq j \leq p\}. \tag{10}$$

The definition of $\hat{\mathcal{A}}_{k,\hat{\rho}_0}$ satisfies the estimation of the smallest and unique active variable set $\mathbf{X}_{\hat{\mathcal{A}}}$ such that $Y \perp\!\!\!\perp \mathbf{X}_{\hat{\mathcal{A}}} \mid \mathbf{X}_{\mathcal{A}}$. Furthermore, we obtain the sufficient screening property of $\hat{\mathcal{A}}_{k,\hat{\rho}_0}$.

Corollary 2 (Sufficient Screening Property by AFD). *Supposing that conditions of (C1)–(C3) hold, we have*

$$\Pr(\mathcal{A}_k \subset \hat{\mathcal{A}}_{k,\hat{\rho}_0}) \geq 1 - 8s_k \exp\left(-c_6 n^{1-2\kappa-2\zeta}\right),$$

where c_6 is some positive constant, and $s_k = |\mathcal{A}_k|$ is the true model size, $k = 1, \dots, K$.

Corollary 2 will be proved in Appendix A.8. In fact, Corollary 2 also can be regarded as the sure screening property as in Fan and Lv (2008). Under the definition of the sufficient variable, the utility in this paper screening the sufficient variables by controlling the false discovery could lead to more precise results. Thus, rename the property in Corollary 2 as a sufficient screening property. We call the proposed AFD control procedure QA-SVS-AFD. The QA-SVS-AFD is computationally efficient and its validity to detect active variables is guaranteed by Corollary 2. A stock-in-trade in the existing screening methods such as Xie et al. (2020) [17] is to control the cardinality of the screened active variable set by setting a certain threshold, and reduce the number of screened variables to be negligible with the ultra-high dimensionality. However, the number of the FD is non-negligible. In this paper, the QA-SVS-AFD procedure can control false discovery precisely by sufficiently controlling the expectation and variance of false discovery to converge to 1.

The estimation of the certain threshold by Theorem 3 is to control the determination of the rejection region under the level of $O(1/p)$. In other words, we reject the null hypothesis $\mathbb{H}_{0,j,k}$ with the significant level around $1/p$. As a result, the maximum subset of variables in the rejection region is the estimation of the sufficient active variable set by the AFD procedure. The AFD control path can be summarized as the following Algorithm 1:

Algorithm 1 QA-SVS-AFD algorithm.

Input: Observation sample (X, Y) and the number of grids K

Output: The screened sufficient variable set $\hat{\mathcal{A}}_{k,\hat{\rho}_0}$ ($k = 1, \dots, K$)

Step 1 Calculate $\hat{v}_{k,1}, \dots, \hat{v}_{k,p}$ of Equation (5) for different $k = 1, \dots, K$;

Step 2 Compute $\hat{\rho}_0$ by $S_{\lambda_1^2}(\hat{\rho}_0) = 1/p$;

Step 3 Search for the screened sufficient active variable set $\hat{\mathcal{A}}_{k,\hat{\rho}_0}$ in Equation (10).

Remark 4. *Alternatively, if one focuses on selecting sufficient predictors relevant to the response Y , one can consider a refined version*

$$\hat{\mathcal{A}}_{\hat{\rho}_0} \equiv \{j : \hat{v}_j \geq \hat{\rho}_0^*, 1 \leq j \leq p\}$$

for testing the $\mathbb{H}_{0,j}$ in Equation (1), where $\hat{v}_j = (K-1)/K \cdot \sum_{k=1}^K \hat{v}_{jk}$, $j = 1, \dots, p$ and $\hat{\rho}_0^*$ satisfies that $S_{\lambda_{K-1}^2}(\hat{\rho}_0^*) = 1/p = O(n^{-2\beta})$. The estimations in Algorithm 1 are replaced by \hat{v}_j and $\hat{\mathcal{A}}_{\hat{\rho}_0}$. As a special case, when $K = 2$, we have $\hat{\mathcal{A}}_{\hat{\rho}_0} = \hat{\mathcal{A}}_{1,\hat{\rho}_0} = \hat{\mathcal{A}}_{2,\hat{\rho}_0}$. The result can be simply proved by Corollary 2, and we omit it.

3.2. False Discovery Rate Control Model

The adaptive error detection control model can adaptively set the rejection region, with the probability of rejection of the null hypothesis testing. It leads to a large type-II error in hypothesis testing (2). Therefore, similar to Tang et al. (2021) [19], considering the control of the type-I error in hypothesis testing (2), a false discovery rate (FDR) control procedure is developed for testing $\mathbb{H}_{0,j,k}$ simultaneously for $j = 1, \dots, p, k = 1, \dots, K$. Without assuming any prespecified distribution, to sufficiently detect active variables at different quantile levels, we provide a suitable estimation for the threshold ρ to separate sufficient active variables by controlling the FDR of each $\mathbb{H}_{0,j,k}$.

With the proposed test statistic \hat{v}_{jk} , the false discovery proportion is

$$FDP_{k,\rho} = \frac{FD_{k,\rho}}{\max\{\sum_{j \in \mathcal{I}} I(\hat{v}_{jk} \geq \rho), 1\}} = \frac{\sum_{j \in \mathcal{A}_k^c} I(\hat{v}_{jk} \geq \rho)}{\max\{\sum_{j \in \mathcal{I}} I(\hat{v}_{jk} \geq \rho), 1\}}$$

for any given ρ , and the false discovery rate is $FDR_{k,\rho} = \mathbb{E}(FDP_{k,\rho})$. By Corollary 1 in Section 2.2, under $\mathbb{H}_{0,j,k}$, each \hat{v}_{jk} converges to distribution of χ_1^2 under conditions of (C1)–(C3). Let $q_k = 1 - s_k$ be the cardinality of \mathcal{A}_k^c , for any given ρ , under the assumption that $s_r/p \rightarrow 0$ as $p \rightarrow \infty$. Intuitively, the estimation for the $FDR_{k,\rho}$ can use the equation that $\frac{FD_{k,\rho}/q_k}{\max\{\sum_{j \in \mathcal{I}} I(\hat{v}_{jk} \geq \rho), 1\}/p}$. However, the separation of the null set \mathcal{A}_k^c and q_k is still intractable. Thus, we attempt to estimate the FDR, by replacing $FD_{k,\rho}/q_k$ by $S_{\chi_1^2}(\rho) = 1 - F_{\chi_1^2}(\rho)$, the survival function of the distribution of χ_1^2 . Hence, for any given ρ , the estimated $FDR_{k,\rho}$ is defined as

$$\widehat{FDR}_{k,\rho} = \frac{pS_{\chi_1^2}(\rho)}{\max\{\sum_{k \in \mathcal{I}} I(\hat{v}_{jk} \geq \rho), 1\}}$$

Consequently, similar to the procedures of Benjamini and Hochberg (1995) [27] and Tang et al. (2021) [19] to control the FDR at a prespecified level $\alpha \in (0, 1)$, we suggest selecting the estimation of the threshold ρ for screening the sufficient active variables by

$$\hat{\rho}_k = \inf\{0 \leq \rho \leq \rho_0 : \widehat{FDR}_{k,\rho} \leq \alpha\} \tag{11}$$

for some constant ρ_0 given in Condition (C2). In practical implementation, adopt the appropriate value of $\hat{v}_{1k}, \dots, \hat{v}_{pk}$ as the estimation of $\widehat{FDR}_{k,\rho}$ for ρ . Thus, the screened set could be defined as

$$\hat{\mathcal{A}}_{k,\alpha} \equiv \{j : \widehat{FDR}_{k,\hat{v}_{jk}} \leq \alpha, 1 \leq j \leq p\}. \tag{12}$$

Define $\hat{v}_{lk} \equiv \arg \max_{k \in \hat{\mathcal{A}}_{k,\alpha}} \widehat{FDR}_{k,\hat{v}_{jk}}$. In other words, \hat{v}_{lk} is the threshold ρ such that $FDR_{k,\rho}$ is maximized subject to $\widehat{FDR}_{k,\rho} \leq \alpha$. Hence, the estimation of FDR is $\widehat{FDR}_{k,\hat{v}_{jk}}$. The proposed FDR control path can be summarized as the following Algorithm 2:

Algorithm 2 QA-SVS-FDR(K) algorithm.

- Input:** Observation sample (X, Y) , the number of grids K , and the prespecified level α
 - Output:** The screened sufficient variable sets $\hat{\mathcal{A}}_{k,\alpha}$ ($k = 1, \dots, K$)
 - Step 1** Calculate $\hat{v}_{k,1}, \dots, \hat{v}_{k,p}$ of Equation (5) for different $k = 1, \dots, K$;
 - Step 2** Compute each $\widehat{FDR}_{k,\rho}$ of Equation (11) for ρ by taking each value of $\hat{v}_{k,1}, \dots, \hat{v}_{k,p}$;
 - Step 3** For given α , search for the set $\hat{\mathcal{A}}_{k,\alpha} \equiv \{j : \widehat{FDR}_{k,\hat{v}_{jk}} \leq \alpha, 1 \leq j \leq p\}$ in Equation (12);
 - Step 4** Find $\hat{v}_{k,l} \equiv \arg \max_{l \in \hat{\mathcal{A}}_{k,\alpha}} \widehat{FDR}_{k,\hat{v}_{jk}}$ and let $\hat{\rho}_k = \hat{v}_{k,l}$;
 - Step 5** Separate the screened sufficient active set $\hat{\mathcal{A}}_{k,\alpha}$ of Equation (11) by $\hat{\rho}_k$.
-

We call the proposed FDR control path QA-SVS-FDR. The computational cost of the QA-SVS-FDR is on the order of $K \cdot O(p)$. The QA-SVS-FDR is also computationally efficient, and its validity to detect active variables is guaranteed by the following theorem.

Theorem 4 (Sufficient Screening Property by Controlling FDR). *Supposing conditions (C1)–(C3) hold, we obtain that*

$$Pr(\mathcal{A}_k \subset \hat{\mathcal{A}}_{k,\alpha}) \geq 1 - 8 \exp(-c_7 n^{1-2\kappa-\xi}),$$

where c_7 is some positive constant and $s_k = |\mathcal{A}_k|$ is the true model size, $k = 1, \dots, K$. For a prespecified level α , if $s_k = |\mathcal{A}_k| = O(n^\zeta)$ for some $\zeta < 1/2$, the FDR of the proposed multiple testing procedure satisfies

$$\lim_{n \rightarrow \infty} \frac{\widehat{FDR}_{\hat{\rho}_k}}{\alpha} = 1,$$

where $\hat{\rho}_k$ is given in Equation (11).

We shall prove Theorem 4 in Appendix A.9. Theorem 4 shows the sufficient screening property of the estimation by controlling FDR accurately. The result of the screened variable set by controlling the cardinality with an empirical threshold leads to the FDR being non-negligible. Hence, in terms of the asymptotic null distribution of the test statistic in Theorem 1, the FDR of the QA-SVS-FDR can be controlled accurately at a prespecified level α , as the estimation of FDR can be approximated sufficiently well by large n .

Remark 5. Alternatively, if focusing on selecting sufficient predictors relevant to the response Y by testing the $\mathbb{H}_{0,j}$ in Equation (1), one can consider a refined version that is

$$FDP_{\rho^*} = \frac{\sum_{j \in \mathcal{A}^c} I(\hat{\nu}_j \geq \rho^*)}{\max\{\sum_{j \in \mathcal{I}} I(\hat{\nu}_j \geq \rho^*), 1\}},$$

and the estimation of FDR_{ρ^*} as

$$\widehat{FDR}_{\rho^*} = \frac{pS_{\chi^2_{(K-1)}}(\rho^*)}{\max\{\sum_{k \in \mathcal{I}} I(\hat{\nu}_j \geq \rho^*), 1\}},$$

where $\hat{\nu}_j = (K - 1) / K \cdot \sum_{k=1}^K \hat{\nu}_{jk}$. Consequently, select the threshold ρ^* by

$$\hat{\rho}^* = \inf\{0 \leq \rho^* \leq \rho_0 : FDR_{k,\rho^*} \leq \alpha\}.$$

As a result, the screened sufficient active variable set is defined as

$$\hat{\mathcal{A}}_\alpha \equiv \{j : \widehat{FDR}_{\hat{\nu}_j} \leq \alpha, 1 \leq j \leq p\}.$$

Define $\hat{\nu}_1 \equiv \arg \max_{k \in \hat{\mathcal{A}}_\alpha} \widehat{FDR}_{\hat{\nu}_j}$. The estimation of the FDR is $\widehat{FDR}_{k,\hat{\nu}_k}$. The path of $\hat{\mathcal{A}}_\alpha$ is summarized in Algorithm 3. Under the given level α , the FDR of the testing (3) satisfies that $\lim_{n \rightarrow \infty} \frac{\widehat{FDR}_{\hat{\rho}^*}}{\alpha} = 1$. The conclusion can be simply proved by Corollary 2, and we omit it.

Algorithm 3 QA-SVS-FDR-S algorithm.

Input: Observation sample (X, Y) , the number of grids K , and the prespecified level α

Output: The screened sufficient variable sets $\hat{\mathcal{A}}_\alpha$

Step 1 Calculate $\hat{\nu}_1, \dots, \hat{\nu}_p$ in Remark 5;

Step 2 Compute each \widehat{FDR}_{ρ^*} in Remark 5 for ρ taking each value of $\hat{\nu}_1, \dots, \hat{\nu}_p$;

Step 3 For given α , separate for the set $\hat{\mathcal{A}}_\alpha$ in Remark 5;

Step 4 Find $\hat{\nu}_1 \equiv \arg \max_{k \in \hat{\mathcal{A}}_\alpha} \widehat{FDR}_{\hat{\nu}_j}$ and let $\hat{\rho}^* = \hat{\nu}_1$;

Step 5 Separate the screened sufficient active set $\hat{\mathcal{A}}_\alpha$ in Remark 5 by $\hat{\rho}^*$.

Thus far, we have completely shown the two paths of sufficient variable screening by controlling the false discovery. The two paths have different essential frameworks: one is to give the adaptive threshold and outlier detecting model to control the false discovery, and the other is to control the false discovery rate accurately by using the survival functions for

estimation under a given prespecified level α . These two paths both can control the false discovery to sufficiently screen active predictors, which is simply the two-step sufficient screening procedure in Yuan et al. (2022) [24].

4. Simulation Studies

In this section, the performance of the proposed procedure will be demonstrated via several simulated examples. In practice, the sample splitting idea is adopted to avoid mathematical challenges caused by the reuse of the sample. Let $\{(Y_i^{(1)}, \mathbf{X}_i^{(1)\top}), i = 1, \dots, n_1\}$ and $\{(Y_i^{(2)}, \mathbf{X}_i^{(2)\top}), i = 1, \dots, n_2\}$ be a random disjoint partition of $\{(Y_i, \mathbf{X}_i^\top), i = 1, \dots, n\}$. The proposed sufficient screening procedure consists of two steps: QA-SVS-SUP, to screen all active variables; QA-SVS-FD, to control the FD adaptively (QA-SVS-AFD) and to control the FDR accurately (QA-SVS-FDR). The two steps are specified as the following:

(1) QA-SVS-A: The p covariates are ranked in descending order according to Remark 5 based on $\{(Y_i^{(1)}, \mathbf{X}_i^{(1)\top}), i = 1, \dots, n_1\}$ and evaluate the minimum model size that all active variables are included.

(2) QA-SVS-FD: Based on $\{(Y_i^{(2)}, \mathbf{X}_i^{(2)\top}), i = 1, \dots, n_2\}$, (i) the sufficient predictors are screened according to Equation (10) at different quantile levels, denoted by $\hat{\mathcal{A}}_k^{\text{AFD}}$; (ii) Given an FDR level α , the threshold $\hat{\rho}_k$ is estimated by Equation (11), and the selected set $\hat{\mathcal{A}}_{k,\alpha}^{\text{AFD}}$ is defined by Equation (12).

4.1. Performance of QA-SVS-A

In this subsection, the variable screening performance of our proposed QA-SVS is compared with SIS (Fan and Lv, 2008) [1], the distance correlation-based screening (DC-SIS; Li et al., 2012) [8], the quantile-adaptive model-free sure independence screening (QA-SIS; He et al., 2013) [9], and the quantile-based correlation screening (QCS; Tang et al., 2013) [19]. The performance of each procedure is evaluated via 5%, 25%, 50%, 75%, and 95% quantiles of the minimum model size that all active variables belong to based on 100 replications. The size is closer to the true model size, which indicates the better performance of variable screening.

In the simulation, the predictors $\mathbf{X} = (X_1, \dots, X_p)^\top$ are generated from a p -variate normal distribution with mean $\mathbf{0}$ and covariance matrix $\Sigma = (\sigma_{ij})_{p \times p}$, where $\sigma_{ij} = \rho^{|i-j|}$. We set $\rho = 0$ and 0.5 . Let the number of quantile grid points $K = 5, 6, \dots, 11$. To simulate a high-dimensional scenario, we set $n = 500$ and $p = 1000$ or 5000 for each scenario. The response variable is sampled from the following models:

Scenario 1.1: $Z_1 = 0.5X_1 + 0.5X_2 + 0.5X_{101} + \varepsilon$;

Scenario 1.2: $Z_2 = 0.8X_3 + 0.5(X_4 + 1)^2 + 0.5 \tan(\pi(X_{102} + 1)/4) + \varepsilon$;

Scenario 1.3: $Z_3 = 0.5 \exp(3X_5) + \sin(\pi X_6/2) + 5X_{103} I\{X_{103} > Q(0.8, X_{103})\} + \varepsilon$;

Scenario 1.4: $Z_4 = (1 + X_7 + X_8)^{-3} \exp\{1 + 3 \sin(\pi X_{104}/2)\} + \varepsilon$;

Scenario 1.5: $Z_5 = 0.5X_9 + \tan(\pi(X_{10} - 1)(X_{105} + 1)/4) + \varepsilon$;

Scenario 1.6: $Z_6 = 2(X_{11} + 1)^2 X_{12} X_{106} I\{X_{12} > Q(0.5, X_{12}), X_{106} < Q(0.5, X_{106})\} + \varepsilon$.

The error term ε follows $\mathcal{N}(0, 1)$, independent of \mathbf{X} . The quantiles of the minimum model size in Scenario 1.1 and Scenario 1.2 that include all active variables with $p = 1000$ and $p = 5000$ are shown in Tables 1 and 2. Due to limited space, the simulation results of the rest Scenario are presented in Appendix B Tables A1–A4.

Table 1. The quantiles of minimum model size in Scenario 1.1 of Section 4.1.

Method	$\rho = 0$					$\rho = 0.5$				
	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%
	$p = 1000$									
QA-SVS-A(4)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QA-SVS-A(5)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QA-SVS-A(6)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QA-SVS-A(7)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QA-SVS-A(8)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QA-SVS-A(9)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QA-SVS-A(10)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QCS(4)	3.0	3.0	3.0	3.0	4.0	3.0	3.0	3.0	3.0	4.0
QCS(5)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	7.0
QCS(6)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	5.0
QCS(7)	3.0	3.0	3.0	3.0	4.0	3.0	3.0	3.0	3.0	5.0
QCS(8)	3.0	3.0	3.0	3.0	5.0	3.0	3.0	3.0	3.0	8.0
QCS(9)	3.0	3.0	3.0	3.0	10.5	3.0	3.0	3.0	3.0	6.5
QCS(10)	3.0	3.0	3.0	4.0	7.5	3.0	3.0	3.0	3.0	7.5
SIS	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
DC-SIS	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QA-SIS(0.1)	3.0	3.0	5.0	10.0	60.0	3.0	3.0	3.0	5.0	23.0
QA-SIS(0.3)	3.0	3.0	3.0	3.0	4.0	3.0	3.0	3.0	3.0	4.5
QA-SIS(0.5)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.5
QA-SIS(0.7)	3.0	3.0	3.0	3.0	4.0	3.0	3.0	3.0	3.0	5.0
QA-SIS(0.9)	3.0	3.0	4.0	12.0	56.0	3.0	3.0	3.0	7.5	38.5
	$p = 5000$									
QA-SVS-A(4)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.5
QA-SVS-A(5)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QA-SVS-A(6)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QA-SVS-A(7)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QA-SVS-A(8)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QA-SVS-A(9)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QA-SVS-A(10)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QCS(4)	3.0	3.0	3.0	3.0	5.0	3.0	3.0	3.0	3.0	7.0
QCS(5)	3.0	3.0	3.0	3.0	6.0	3.0	3.0	3.0	3.0	5.0
QCS(6)	3.0	3.0	3.0	3.0	5.0	3.0	3.0	3.0	3.0	8.5
QCS(7)	3.0	3.0	3.0	3.0	10.0	3.0	3.0	3.0	3.0	14.5
QCS(8)	3.0	3.0	3.0	3.0	30.0	3.0	3.0	3.0	3.0	24.0
QCS(9)	3.0	3.0	3.0	4.0	28.5	3.0	3.0	3.0	4.0	46.5
QCS(10)	3.0	3.0	3.0	5.0	36.5	3.0	3.0	3.0	5.5	71.5
SIS	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
DC-SIS	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QA-SIS(0.1)	3.0	4.0	11.5	56.5	241.0	3.0	3.0	5.0	18.0	254.0
QA-SIS(0.3)	3.0	3.0	3.0	3.0	8.5	3.0	3.0	3.0	4.0	6.5
QA-SIS(0.5)	3.0	3.0	3.0	3.0	4.0	3.0	3.0	3.0	3.0	4.5
QA-SIS(0.7)	3.0	3.0	3.0	3.0	14.0	3.0	3.0	3.0	3.0	9.0
QA-SIS(0.9)	3.0	4.0	13.0	55.0	160.0	3.0	3.0	6.0	14.0	222.5

Notes: QA-SVS-A(4), QA-SVS-A(5), ..., and QA-SVS-A(10), our proposed method defined in Remark 3 with different quantile grid points ($K = 4, \dots, 10$); QCS-A(4), QCS-A(5), ..., and QCS-A(10), the quantile correlation-based screening method (Tang et al., 2013) [19] with different quantile grid points ($K = 4, \dots, 10$); SIS, the sure independence screening (Fan and Lv, 2008) [1]; DC-SIS, the distance correlation-based screening (Li et al., 2012) [8]; QA-SIS(0.1), QA-SIS(0.3), ..., QA-SIS(0.9), the quantile-adaptive model-free sure independence screening (He et al., 2013) [9] at different quantile levels.

Table 2. The quantiles of minimum model size in Scenario 1.2 of Section 4.1.

Method	$\rho = 0$					$\rho = 0.5$				
	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%
$p = 1000$										
QA-SVS-A(4)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QA-SVS-A(5)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QA-SVS-A(6)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QA-SVS-A(7)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QA-SVS-A(8)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QA-SVS-A(9)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QA-SVS-A(10)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QCS(4)	3.0	3.0	3.0	3.0	11.0	3.0	3.0	3.0	3.0	4.0
QCS(5)	3.0	3.0	3.0	3.0	17.0	3.0	3.0	3.0	3.0	7.0
QCS(6)	3.0	3.0	3.0	4.0	12.5	3.0	3.0	3.0	3.0	5.0
QCS(7)	3.0	3.0	3.0	4.0	15.5	3.0	3.0	3.0	3.0	5.0
QCS(8)	3.0	3.0	3.0	5.0	35.0	3.0	3.0	3.0	3.0	8.0
QCS(9)	3.0	3.0	3.0	9.0	127.0	3.0	3.0	3.0	3.0	6.5
QCS(10)	3.0	3.0	3.0	6.0	32.5	3.0	3.0	3.0	3.0	7.5
SIS	287.0	486.5	697.5	870.0	986.5	127.5	330.5	573.5	824.0	971.0
DC-SIS	3.0	3.0	3.0	3.0	260.0	3.0	3.0	3.0	3.0	17.0
QA-SIS(0.1)	176.5	262.0	394.5	576.5	814.5	61.5	147.5	257.0	394.0	630.5
QA-SIS(0.3)	3.0	3.0	4.0	6.0	21.5	3.0	3.0	3.0	3.0	4.0
QA-SIS(0.5)	3.0	3.0	3.0	3.0	6.5	3.0	3.0	3.0	3.0	3.0
QA-SIS(0.7)	3.0	5.0	8.5	23.5	67.0	3.0	3.0	3.0	4.0	5.5
QA-SIS(0.9)	100.5	238.0	368.0	517.5	866.5	35.5	85.5	143.5	301.0	601.0
$p = 5000$										
QA-SVS-A(4)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(5)	3.0	3.0	3.0	3.0	5.0	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(6)	3.0	3.0	3.0	3.0	4.0	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(7)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(8)	3.0	3.0	3.0	3.0	4.0	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(9)	3.0	3.0	3.0	3.0	4.0	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(10)	3.0	3.0	3.0	3.0	4.0	3.0	3.0	3.0	3.0	3.0
QCS(4)	3.0	3.0	3.0	3.5	46.0	3.0	3.0	3.0	3.0	3.0
QCS(5)	3.0	3.0	3.0	8.0	72.5	3.0	3.0	3.0	3.0	3.0
QCS(6)	3.0	3.0	3.0	6.5	38.5	3.0	3.0	3.0	3.0	3.0
QCS(7)	3.0	3.0	3.0	10.5	171.5	3.0	3.0	3.0	3.0	3.0
QCS(8)	3.0	3.0	4.0	12.0	124.5	3.0	3.0	3.0	3.0	3.0
QCS(9)	3.0	3.0	4.0	15.5	353.5	3.0	3.0	3.0	3.0	3.0
QCS(10)	3.0	3.0	4.0	20.0	221.5	3.0	3.0	3.0	3.0	3.0
SIS	1391.0	2628.5	3487.0	4254.5	4810.5	1165.5	2144.0	3288.5	4218.5	4937.5
DC-SIS	3.0	3.0	3.0	4.0	1197.5	3.0	3.0	3.0	3.0	57.5
QA-SIS(0.1)	439.0	1062.5	1653.5	2543.5	3717.0	260.5	629.0	1142.0	1813.5	3317.0
QA-SIS(0.3)	4.0	6.0	9.5	20.5	110.5	3.0	3.0	4.0	5.0	9.5
QA-SIS(0.5)	3.0	3.0	3.0	5.0	13.0	3.0	3.0	3.0	3.0	3.0
QA-SIS(0.7)	5.0	11.0	25.0	61.0	377.0	3.0	3.0	4.0	5.0	9.5
QA-SIS(0.9)	625.0	1510.5	2279.0	3218.0	4569.0	175.0	500.0	868.0	1592.5	2403.0

Notes: All notations are the same as those in Table 1.

Under Scenario 1.1 with the linear model at a small signal-to-noise level, all five methods perform well. Under Scenario 1.2 with the additive model, QCS, DC-SIS and QA-SIS perform comparably to the proposed QA-SVS-S procedure, while SIS fails to detect the active predictors. Under Scenario 1.3 with a nonlinear relationship between the response and predictors, all methods perform well to effectively screen out the inactive predictors except for QA-SIS at high quantile level. Under Scenario 1.4 with a nonlinear relationship between the response and predictors, the proposed QA-SVS-S and QCS screening procedures behave effectively, QA-SIS behaves little weaker at the 0.5th quantile level, and the other screening procedures struggle to maintain a reasonable model size at all quantiles.

Under Scenarios 1.5 and 1.6 with interactions, the proposed QA-SVS-S and QCS perform relatively stable, while both of them behave a little poorly when there are higher-order effects. The QA-SIS in extremely low or high quantile level suffers a major setback, but the proposed QA-SVS-S screens robustly. In addition, the performance of the proposed QA-SVS-S is only discounted slightly when p increases from 1000 to 5000, but the other methods are not. Furthermore, the results of the proposed QA-SVS-S in all Scenarios under $\rho = 0.5$ indicate that the correlation of covariates provides sufficient screening relationships. Through the different settings of the number of grid points K , it shows that the QA-SVS-S will be more effective at detecting the active predictors as the increase of K , whereas QCS has the opposite trend.

4.2. Performance of QA-SVS-FD

In this subsection, some scenarios are simulated to examine the proposed QA-SVS-FD as well as the sufficient screening property of the proposed procedure. We compare the variable screening performance of our proposed QA-SVS-FD with the quantile-based correlation screening under controlling FDR (QCS-FDR; Tang et al., 2013) [19]. The predictors $\mathbf{X} = (X_1, \dots, X_p)^T$ are generated from a p -variate normal distribution with mean $\mathbf{0}$ and covariance matrix $\Sigma = (\sigma_{ij})_{p \times p}$. The response is generated from the following models:

Scenario 2.1: $Y = \sum_{j=1}^{10} X_j + \varepsilon$, $\Sigma = (\rho^{|i-j|})_{p \times p}$, and $\rho = 0.5$;

Scenario 2.2: $Y = \sum_{j=1}^{50} X_j + \varepsilon$, $\Sigma = (\rho^{|i-j|})_{p \times p}$, and $\rho = 0.5$;

Scenario 2.3: $Y = \exp(\sum_{j=1}^{10} X_j) + \varepsilon$, $\Sigma = (\rho^{|i-j|})_{p \times p}$, and $\rho = 0.5$;

Scenario 2.4: $Y = \exp(\sum_{j=1}^{50} X_j) + \varepsilon$, $\Sigma = (\rho^{|i-j|})_{p \times p}$, and $\rho = 0.5$;

Scenario 2.5: $Y = \sum_{j=1}^{10} (-1)^{j-1} X_j + \varepsilon$;

Scenario 2.6: $Y = \sum_{j=1}^{50} (-1)^{j-1} X_j + \varepsilon$;

Scenario 2.7: $Y = \sum_{j=1}^{10} X_j / \{0.5 + (1.5 + \sum_{j=2}^4 (-1)^{j-1} X_j)^2\} + 0.1\varepsilon$;

Scenario 2.8: $Y = \sum_{j=1}^{50} X_j / \{0.5 + (1.5 + \sum_{j=21}^{40} (-1)^{j-1} X_j)^2\} + 0.1\varepsilon$.

Σ in Scenarios 2.5–2.8 has diagonal element 1 and sub-diagonal element 0.2. The covariates are independent in Scenarios 2.1–2.4 and weakly dependent in Scenarios 2.5–2.8. We consider $n = 500$ and $p = 1000$ or 5000 for all scenarios. Set the number of quantile grid points $K = 2, 3, 4, 5, 6$. The nominal false discovery rate is $\alpha = 0.05$. We evaluate the performance based on the following criteria:

- $|\hat{\mathcal{A}}|$: the average number of screened variables;
- FDR: the average of empirical FDP;
- F_1 -score: the average of $2 \cdot |\{j : j \in \hat{\mathcal{A}}\}| / (|\{j : j \in \mathcal{A}\}| + |\{j : j \in \hat{\mathcal{A}}\}|)$.

Based on 100 replications, the results of the QA-SVS-FDR and the QCS procedure are stored in Table 3, and the results with $p = 5000$ are presented in Appendix B Table A5.

Under Scenarios 2.1–2.4, the proposed QA-SVS-FDR performs as well as QCS-FDR. The proposed QA-SVS-AFD has the same performance with a small K , whereas QA-SVS-AFD would miss some active predictors as the increase of K . It can be found that the three procedures control the empirical FDR under the prespecified level α for most scenarios. As the increase of the number of active predictors, F_1 -score of the proposed QA-SVS-FD (QA-SVS-AFD and QA-SVS-FDR) has a little improvement, such as 0.92 to 0.97. The QCS-FDR shows the opposite trend. Combined with the $|\hat{\mathcal{A}}|$, we obtain that our proposed method screens out the null predictors more accurately but will lose some active predictors. With sufficient screening by controlling FDR, our procedure can retain active predictors as much as possible. Under Scenarios 2.5 and 2.7, it can be seen that our method works slightly better than in QCS, especially the FDR and F_1 -score of QA-SVS-AFD reach 0 and 1, respectively. Under Scenarios 2.6 and 2.8, the proposed QA-SVS-AFD and QCS-FDR both fail. However, it is worth mentioning that QA-SVS-FD has larger $|\hat{\mathcal{A}}|$ and F_1 -score than QCS-FDR, which indicates that the performance of QA-SVS-FDR is more effective. In addition, our QA-SVS-FD procedure works reasonably well as p increases from 1000 to

5000, where QCS behaves slightly poorly. In summary, our proposed method performs almost as well and is more effective than QCS-FDR in various practical settings.

In terms of the highly sensitivity of the model-free method to some factors that can distort the underlying relationships between the covariates and the response, we suggest that one can reduce the sensitivity by using the QA-SVS procedure with different numbers of unfixing grid points. This can lead to different model complexity, where the large K can lead to overfitting, and the small K can lead to under-fitting.

Table 3. The result of criteria in all scenarios under $p = 1000$ with $\alpha = 0.05$ of Section 4.2.

Method	QA-SVS-AFD(K)					QA-SVS-FDR(K)					QCS-FDR(K)				
	K	2	3	4	5	6	2	3	4	5	6	2	3	4	5
Scenario 2.1															
$ \hat{\mathcal{A}} $	12.17	10.81	10.57	10.23	10.05	11.70	11.74	11.76	11.75	11.52	11.33	11.24	11.29	11.09	11.10
FDR	0.17	0.07	0.05	0.02	0.00	0.14	0.14	0.14	0.14	0.13	0.11	0.10	0.11	0.09	0.09
F ₁ -score	0.90	0.96	0.97	0.99	1.00	0.92	0.92	0.92	0.92	0.93	0.94	0.94	0.94	0.95	0.95
Scenario 2.2															
$ \hat{\mathcal{A}} $	50.40	48.14	44.87	37.95	26.82	52.16	52.29	52.28	52.21	52.35	50.53	50.52	50.16	49.59	48.59
FDR	0.02	0.00	0.00	0.00	0.00	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
F ₁ -score	0.98	0.98	0.95	0.86	0.69	0.97	0.97	0.97	0.97	0.97	0.96	0.95	0.95	0.94	0.94
Scenario 2.3															
$ \hat{\mathcal{A}} $	11.98	10.66	10.38	10.13	10.02	11.48	11.51	11.70	11.40	11.59	11.11	11.02	11.22	10.96	11.00
FDR	0.16	0.06	0.03	0.01	0.00	0.12	0.12	0.14	0.12	0.13	0.09	0.09	0.10	0.08	0.08
F ₁ -score	0.91	0.97	0.98	0.99	1.00	0.93	0.93	0.92	0.94	0.93	0.95	0.95	0.94	0.96	0.95
Scenario 2.4															
$ \hat{\mathcal{A}} $	50.31	46.07	39.30	26.54	14.55	52.22	52.26	51.75	51.88	51.77	49.77	47.84	46.93	44.97	43.69
FDR	0.02	0.00	0.00	0.00	0.00	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.05	0.05	0.05
F ₁ -score	0.98	0.96	0.88	0.69	0.45	0.97	0.97	0.97	0.97	0.97	0.95	0.93	0.92	0.90	0.89
Scenario 2.5															
$ \hat{\mathcal{A}} $	10.82	9.93	9.16	7.84	5.71	10.36	10.73	10.50	10.44	10.55	9.93	9.99	9.76	9.57	9.09
FDR	0.08	0.00	0.00	0.00	0.00	0.04	0.06	0.05	0.04	0.05	0.05	0.05	0.04	0.05	0.05
F ₁ -score	0.96	0.99	0.95	0.87	0.72	0.98	0.97	0.97	0.98	0.97	0.94	0.95	0.94	0.93	0.90
Scenario 2.6															
$ \hat{\mathcal{A}} $	14.87	3.63	0.53	0.07	0.01	12.86	15.04	12.49	10.79	9.23	1.54	1.35	1.07	0.61	0.48
FDR	0.06	NaN	NaN	NaN	NaN	0.05	0.05	0.03	NaN	NaN	NaN	NaN	NaN	NaN	NaN
F ₁ -score	0.43	0.13	0.02	0.00	0.00	0.38	0.43	0.38	0.33	0.29	0.05	0.04	0.03	0.02	0.02
Scenario 2.7															
$ \hat{\mathcal{A}} $	11.11	10.00	10.00	9.99	9.93	10.66	10.44	10.65	10.45	10.54	10.62	10.60	10.64	10.45	10.61
FDR	0.09	0.00	0.00	0.00	0.00	0.06	0.04	0.06	0.04	0.05	0.05	0.05	0.05	0.04	0.05
F ₁ -score	0.95	1.00	1.00	1.00	1.00	0.97	0.98	0.97	0.98	0.98	0.97	0.97	0.97	0.98	0.97
Scenario 2.8															
$ \hat{\mathcal{A}} $	38.86	19.20	5.93	1.26	0.29	43.22	42.34	40.35	37.54	36.33	23.00	16.93	11.91	8.62	6.20
FDR	0.03	0.00	0.00	NaN	NaN	0.05	0.05	0.05	0.04	0.05	0.05	0.04	0.04	0.05	NaN
F ₁ -score	0.85	0.55	0.21	0.05	0.01	0.88	0.87	0.85	0.82	0.80	0.59	0.48	0.36	0.27	0.20

Notes: QA-SVS-AFD(K), our proposed method by controlling FD adaptively defined in Remark 4 with different quantile grid points ($K = 2, \dots, 6$); QA-SVS-FDR(K), our proposed method by controlling FDR defined in Remark 5 with different quantile grid points ($K = 2, \dots, 6$); QCS-FDR(K), the quantile correlation-based screening method (Tang et al., 2013) [19] with different quantile grid points ($K = 2, \dots, 6$). $|\hat{\mathcal{A}}|$: the average number of selected predictors; FDR: the average of empirical false discovery proportion, where 'NaN' indicates the method loss validity; F₁-score: the average of F1-score.

5. Real Dataset Research

In the era of rapid development of machine learning and pattern recognition, some image recognition technologies are applied in the medical field. For example, through the processing of lung CT images, we can identify whether the lung has a disease. The following two methods are often used to quantitatively evaluate the severity of emphysema: one is

CT density measurement. Based on the pixel image of CT digital, calculate the average lung density of the patient, then establish the threshold, calculate the proportion of the area below the threshold, and evaluate the situation of emphysema. The other is the percentile density measurement (PD) technique. Analyze the attenuation distribution curve of lung density, give a percentile (commonly 5% and 95%), calculate the area below the percentile density curve, and evaluate the symptoms of emphysema [28]. In this section, we shall apply our proposed method to analyze the lung CT image dataset downloaded from Kaggle, which can segment lungs accurately.

There is a picture of a subject in Appendix C Figure A1. Among them, we regard 5% and 95% PD data as the corresponding continuous response variable, respectively. For smokers, these values are usually high, indicating that other substances in the lungs have accumulated. The data include 267 instances and 512*512 continuous covariates stretched by picture pixels.

By giving different values of quantile grid points $K = 2, 3, \dots, 6$ and considering the threshold of FDR under the given prespecified level $\alpha = 0.05$, we obtain the different segmentations and extractions. The numbers of selected picture pixels are displayed in Table 4. It is clear that QCS-FDR loses efficacy, and QA-SVS-AFD works when $K \leq 3$. Fortunately, QA-SVS-FDR works effectively under all values of $K = 2, 3, \dots, 6$. Compared with the QA-SVS-AFD(K) and the QA-SVS-FDR(K), under the hypothesis testing (2), it could be found that the screened active variable set is estimated by the rejection region of the QA-SVS-AFD(K) path, which controls the probability around $1 / (512 * 512)$ with not enough number of active variables. The QA-SVS-FDR(K) selects the active variable sufficiently by testing the null hypothesis of testing (2) under the given prespecified FDR level $\alpha = 0.5$. We illustrate the extraction by plotting the segmented lung CT with the average of the values of the selected predictors, which are presented in Appendix C, Figures A2–A6. These results may provide some information for measuring important clinical parameters (lung volume, PD, etc); considering the length of this paper, we do not go further.

Table 4. The numbers of selected picture pixels in applications of Section 5.

<i>K</i>	2	3	4	5	6
			5% PD		
QA-SVS-AFD(K)	20,152	2435	28	0	0
QA-SVS-FDR(K)	89,353	85,426	83,991	76,492	74,106
QCS-FDR(K)	262,144	262,144	262,144	262,144	262,144
			95% PD		
QA-SVS-AFD(K)	5151	502	15	0	0
QA-SVS-FDR(K)	76,800	76,442	75,863	78,157	66,473
QCS-FDR(K)	262,144	262,144	262,144	262,144	262,144

Note: The number of complete figure pixels is 262,144. QA-SVS-AFD(K), our proposed method by controlling FD adaptively defined in Remark 4 with different quantile grid points; QA-SVS-FDR(K), our proposed method by controlling FDR defined in Remark 5 with different quantile grid points; QCS-FDR(K), the quantile correlation-based screening method (Tang et al., 2013) [19] with different quantile grid points.

6. Conclusions

In this paper, we propose a multiple testing procedure with false discovery control to detect active variables sufficiently. The multiple testing procedure can be applied with the quantile-adaptive screening method when the dimensionality is ultra-high. Although the QA-SVS procedure is built on the quantile-adaptive marginal screening statistic, by controlling the FD of the marginal structural testings, the QA-SVS procedure can screen out sufficient variables through the precise separation of the sufficient variable set. As the results in this paper, if the grid points K grow faster than n and p , the QA-SVS statistic can capture more subtle values better than QC-SIS, which is in line with the definition of the sufficient variable. In addition, the convergent rate of the asymptotic null distribution of our proposed procedure is larger than the QCS under a large K . In the simulation studies,

we set different values of K to inspect the performance of the QA-SVS. Nevertheless, it would be of interest to study a data-driven way to select K . We leave some space here for the future research.

Author Contributions: Conceptualization, J.C.; Data curation, Z.Y.; Formal analysis, Z.Y.; Funding acquisition, J.C.; Investigation, Z.Y.; Methodology, Z.Y.; Project administration, H.Q. and Y.H.; Supervision, J.C.; visualization, H.Q.; Validation, Y.H.; Writing—original draft, Z.Y.; Writing—review and editing, Z.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China under Grant No. 81671633 to J.C.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data in this paper have been presented in the manuscript.

Acknowledgments: Many thanks to reviewers for their positive feedback, valuable comments, and constructive suggestions that helped improve the quality of this article. Many thanks to the editors' great help and coordination for the publication of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Main Proof

Appendix A.1. Proof of Remark 1

Proof. According to the definition of \mathcal{A}_k in Assumption (I), for any $\mathbf{x}_{\mathcal{A}_k^c} \in \mathbb{R}_{\mathcal{A}_k^c}$, we have

$$\Pr(Y \in G_k | \mathbf{X}) = \Pr(Y \in G_k | \mathbf{X}_{\mathcal{A}_k}). \quad (\text{A1})$$

Due to that $\mathbf{X} = (\mathbf{X}_{\mathcal{A}_k}, \mathbf{X}_{\mathcal{A}_k^c})$, for any $\mathbf{x}_{\mathcal{A}_k^c} \in \mathbb{R}_{\mathcal{A}_k^c}$, multiply $\Pr(\mathbf{X}_{\mathcal{A}_k^c} = \mathbf{x}_{\mathcal{A}_k^c} | \mathbf{X}_{\mathcal{A}_k})$ on both sides of the Equation (A1), we obtain that

$$\Pr(Y \in G_k, \mathbf{X}_{\mathcal{A}_k^c} = \mathbf{x}_{\mathcal{A}_k^c} | \mathbf{X}_{\mathcal{A}_k}) = \Pr(\mathbf{X}_{\mathcal{A}_k^c} = \mathbf{x}_{\mathcal{A}_k^c} | \mathbf{X}_{\mathcal{A}_k}) \cdot \Pr(Y \in G_k | \mathbf{X}_{\mathcal{A}_k}).$$

Thus, we obtain that

$$\Pr(Y \in G_k, \mathbf{X}_{\mathcal{A}_k^c} \leq \mathbf{x}_{\mathcal{A}_k^c} | \mathbf{X}_{\mathcal{A}_k}) = \Pr(\mathbf{X}_{\mathcal{A}_k^c} \leq \mathbf{x}_{\mathcal{A}_k^c} | \mathbf{X}_{\mathcal{A}_k}) \cdot \Pr(Y \in G_k | \mathbf{X}_{\mathcal{A}_k}).$$

Due to the multiplicative law of probability, it is clear that

$$I(Y \in G_k) \perp\!\!\!\perp \mathcal{A}_k^c | \mathcal{A}_k.$$

In terms of invertibility, we proved that

$$\begin{aligned} \mathcal{A}_k &= \{1 \leq j \leq p : \Pr(Y \in G_k | \mathbf{X}) \text{ functionally depends on } X_j\} \\ &= \{1 \leq j \leq p : I(Y \in G_k) \perp\!\!\!\perp X_j | \mathbf{X}_{\mathcal{A}_k}\}. \end{aligned}$$

Thus, \mathcal{A}_k is the sufficient screening variables index set. \square

Appendix A.2. Proof of Lemma 1

Proof. Note that $\mathcal{A}_k^c = \{1 \leq j \leq p : I(Y \in G_k) \perp\!\!\!\perp X_j | \mathbf{X}_{\mathcal{A}_k}\}$; this indicates that

$$\begin{aligned} \Pr(Y \in G_k, \mathbf{X}_{\mathcal{A}_k^c} \leq \mathbf{x}_{\mathcal{A}_k^c} | \mathbf{X}_{\mathcal{A}_k}) &= \Pr(Y \in G_k | \mathbf{X}_{\mathcal{A}_k}) \cdot \Pr(\mathbf{X}_{\mathcal{A}_k^c} \leq \mathbf{x}_{\mathcal{A}_k^c} | \mathbf{X}_{\mathcal{A}_k}) \\ \iff \Pr(\mathbf{X}_{\mathcal{A}_k^c} \leq \mathbf{x}_{\mathcal{A}_k^c} | Y \in G_k, \mathbf{X}_{\mathcal{A}_k}) &= \Pr(\mathbf{X}_{\mathcal{A}_k^c} \leq \mathbf{x}_{\mathcal{A}_k^c} | \mathbf{X}_{\mathcal{A}_k}) \\ \iff \mathbb{E}_{\mathbf{X}_{\mathcal{A}_k^c}} \left\{ \Pr(\mathbf{X}_{\mathcal{A}_k^c} \leq \mathbf{x}_{\mathcal{A}_k^c} | Y \in G_k, \mathbf{X}_{\mathcal{A}_k}) \right\} &= \mathbb{E}_{\mathbf{X}_{\mathcal{A}_k^c}} \left\{ \Pr(\mathbf{X}_{\mathcal{A}_k^c} \leq \mathbf{x}_{\mathcal{A}_k^c} | \mathbf{X}_{\mathcal{A}_k}) \right\} \\ \iff \Pr(\mathbf{X}_{\mathcal{A}_k^c} | Y \in G_k) &= \Pr(\mathbf{X}_{\mathcal{A}_k^c}). \end{aligned}$$

Thus, we obtain that $F_j(x | Y \in G_k) - F_j(x) = 0$ holds when $j \notin \mathcal{A}_k$. \square

Appendix A.3. Proof of Lemma 2

Proof. Note $\Omega = \{0, 1, \dots, n - 1\}$, and $\Omega_k = \{\xi_1, \xi_2, \dots, \xi_e\} (e = 1, \dots, n)$ is an e size sample set random sampled in equal probability without replacement from the population, where $\xi_i (i = 1, 2, \dots, e)$ represents the i -th random sample, which satisfies a discrete uniform distribution from 0 to $n - 1$, that is, $\xi_i \sim \text{DiscreteU}(0, n - 1)$. ξ_i has the following properties:

$$\mathbb{E}\xi_i = \frac{n - 1}{2}, \quad \mathbb{E}\xi_i^2 = \frac{(n - 1)(2n - 1)}{6}, \quad \text{Var}\xi_i = \frac{n^2 - 1}{12}.$$

In addition, for all $i_1 \neq i_2$, where $i_1, i_2 \in \{1, 2, \dots, n\}$,

$$\mathbb{E}\xi_{i_1}\xi_{i_2} = \frac{(n - 2)(3n - 1)}{12}, \quad \text{Cov}(\xi_{i_1}, \xi_{i_2}) = -\frac{n + 1}{12}.$$

For continuous variables subject to arbitrary distribution $X_j = \{x_{1j}, x_{2j}, \dots, x_{nj}\}$, let $\xi_{ij} = \sum_{k \neq i}^n I(X_{ij} \leq x_{kj})$. It is easy to find that the random variable $\xi_{ij} (i = 1, 2, \dots, n)$ is a special case in Mohamed and Mirakhmedov (2016) [26], where $e = n$. According to the definition of \hat{t}_{jk} and ξ_{ij} , it is obviously established that $\Pr(\xi_i = r) = 1/n (i = 1, 2, \dots, n; r = 0, 1, \dots, n - 1)$ and $\xi_{i_1} \neq \xi_{i_2} (\forall i_1 \neq i_2)$. By the definition of ξ_{ij} , for all $j = 1, 2, \dots, p, k = 1, 2, \dots, K$,

$$\begin{aligned} & \frac{1}{n + 1} \sum_{k=1}^n \frac{1}{n} \sum_{i=1}^n \frac{I(x_{ij} \leq x_{kj}, Y_i \in \hat{G}_k)}{\hat{p}_k} \\ &= \frac{1}{n(n + 1)\hat{p}_k} \sum_{k=1}^n \sum_{i=1}^n I(Y_i \in \hat{G}_k) I(x_{ij} \leq x_{kj}) \\ &= \frac{1}{n(n + 1)\hat{p}_k} \left[\sum_{k=1}^n \sum_{i \neq k}^n I(Y_i \in \hat{G}_k) I(x_{ij} \leq x_{kj}) + \sum_{k=1}^n I(Y_i \in \hat{G}_k) \right] \\ &= \frac{1}{n(n + 1)\hat{p}_k} \sum_{i=1}^n I(Y_i \in \hat{G}_k) \xi_{ij} + \frac{1}{n + 1}. \end{aligned}$$

If $\mathbb{H}_{0,j,k} (j = 1, 2, \dots, p; k = 1, 2, \dots, K)$ is true, the expectation and variance of \hat{t}_{jk} can be obtained as follows:

$$\begin{aligned} \mathbb{E}\{\hat{t}_{jk}\} &= \mathbb{E}\left\{ \frac{1}{n + 1} \sum_{k=1}^n \frac{1}{n} \sum_{i=1}^n \frac{I(x_{ij} \leq x_{kj}, Y_i \in \hat{G}_k)}{\hat{p}_k} - \frac{1}{2} \right\} \\ &= \mathbb{E}\left\{ \frac{1}{n(n + 1)\hat{p}_k} \sum_{i=1}^n I(Y_i \in \hat{G}_k) \xi_{ij} \right\} + \frac{1}{n + 1} - \frac{1}{2} \\ &= \frac{1}{n(n + 1)\hat{p}_k} \sum_{i=1}^n I(Y_i \in \hat{G}_k) \mathbb{E}\{\xi_{ij}\} + \frac{1}{n + 1} - \frac{1}{2} \\ &= \frac{1}{n(n + 1)\hat{p}_k} \frac{n - 1}{2} n \hat{p}_k + \frac{1}{n + 1} - \frac{1}{2} = 0 \end{aligned}$$

and

$$\begin{aligned} \text{Var}\{\hat{t}_{jk}\} &= \text{Var}\left\{ \frac{1}{n + 1} \sum_{k=1}^n \frac{1}{n} \sum_{i=1}^n \frac{I(x_{ij} \leq x_{kj}, Y_i \in \hat{G}_k)}{\hat{p}_k} - \frac{1}{2} \right\} \\ &= \text{Var}\left\{ \frac{1}{n(n + 1)\hat{p}_k} \sum_{i=1}^n I(Y_i \in \hat{G}_k) \xi_{ij} \right\} \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{n^2(n+1)^2\hat{p}_k^2} \left[\sum_{i=1}^n I(Y_i \in \hat{G}_k) \text{Var}\{\xi_{ij}\} \right. \\
 &\quad \left. + \sum_{i_1 \neq i_2}^n I(Y_{i_1} \in \hat{G}_k, Y_{i_2} \in \hat{G}_k) \text{Cov}(\xi_{i_1j}, \xi_{i_2j}) \right] \\
 &= \frac{1}{n^2(n+1)^2\hat{p}_k^2} \left\{ n\hat{p}_k \frac{n^2-1}{12} - n\hat{p}_k(n\hat{p}_k-1) \frac{n+1}{12} \right\} = \frac{1-\hat{p}_k}{12(n+1)\hat{p}_k}.
 \end{aligned}$$

Let $\Omega_n = (\zeta_{1n}, \dots, \zeta_{nn})$ be a random permutation of $\{0, 1, \dots, n-1\}$, and $r = (r_1, \dots, r_N)$ is a random vector independent of $\zeta_{1n}, \dots, \zeta_{nn}$ satisfying $P\{r_1 = k_1, \dots, r_n = k_n\} = 1/(n!)$, where $k = (k_1, \dots, k_n)$ is also a random permutation of $1, \dots, n$. Note that $S_{n\hat{p}_k, n} = \zeta_{r_1n} + \dots + \zeta_{r_{n\hat{p}_k}n}$ represents a sum of n random vector samples chosen at random without replacement from the population Ω_n . It can be expressed equivalently as $S_{n\hat{p}_k, n} = I\{\hat{Q}_{k-1} \leq Y_1 < \hat{Q}_k\} \zeta_{1N} + \dots + I\{\hat{Q}_{k-1} \leq Y_n < \hat{Q}_k\} \zeta_{nn}$, where $I(\cdot)$ represents indicative function. It is shown that $\sum_{i=1}^n I(Y_i \in \hat{G}_k) \xi_{ij}$ and $S_{n\hat{p}_k, n}$ have the same distribution. Denote

$$\begin{aligned}
 \tau &= \frac{n-1}{2}, \quad b^2 = \frac{n^2-1}{12}, \quad \hat{\zeta}_{mn} = \frac{(\zeta_{mn} - \tau)}{b}, \\
 \sigma^2 &= \frac{(1-\hat{p}_k)(n^2-1)}{12}, \quad \sigma^{*2} = \frac{(1-\hat{p}_k)(n+1)n}{12}, \\
 A_k &= \frac{1}{n} \sum_{m=1}^n \hat{\zeta}_{mn}^k, \quad B_k = \frac{1}{n} \sum_{m=1}^n |\hat{\zeta}_{mn}|^k.
 \end{aligned}$$

Let $F_n(u) = \Pr\{S_{n\hat{p}_k, n} < u\sigma^* \sqrt{n\hat{p}_k} + n\hat{p}_k\tau\}$. Based on the same distribution of $S_{n\hat{p}_k, n}$ and $\sum_{i=1}^n I(Y_i \in \hat{G}_k) \xi_{ij}$, for any $j \in \{1, \dots, p\}$, we have

$$\begin{aligned}
 F_n(u) &= \Pr\left\{ \sum_{i=1}^n I(Y_i \in \hat{G}_k) \xi_{ij} < u\sigma^* \sqrt{n\hat{p}_k} + n\hat{p}_k\tau \right\} \\
 &= \Pr\left\{ \frac{1}{n\hat{p}_k(n+1)} \left(\sum_{i=1}^n I(Y_i \in \hat{G}_k) \xi_{ij} - n\hat{p}_k \frac{n-1}{2} \right) < u \frac{\sigma^*}{\sqrt{n\hat{p}_k(n+1)}} \right\} \\
 &= \Pr\left\{ \hat{\tau}_{jk} < u \sqrt{\frac{1-\hat{p}_k}{12(n+1)\hat{p}_k}} \right\} := F_{\hat{\tau}_{jk}}(u).
 \end{aligned}$$

Considering

$$\begin{aligned}
 \hat{\tau}_{jk} &= \frac{1}{n+1} \sum_{k=1}^n \frac{1}{n} \sum_{i=1}^n \frac{I(x_{ij} \leq x_{kj}, Y_i \in \hat{G}_k)}{\hat{p}_k} - \frac{1}{2} \\
 &= \frac{1}{n(n+1)\hat{p}_k} \left[\sum_{k=1}^n \sum_{i \neq k}^n I(Y_i \in \hat{G}_k) I(x_{ij} \leq x_{kj}) + \sum_{k=1}^n I(Y_i \in \hat{G}_k) \right] - \frac{1}{2} \\
 &= \frac{1}{n(n+1)\hat{p}_k} \left[\sum_{k=1}^n \sum_{i \neq k}^n I(Y_i \in \hat{G}_k) I(x_{ij} \leq x_{kj}) \right] - \frac{n-1}{2(n+1)} \\
 &= \frac{n-1}{2(n+1)} - \frac{1}{n(n+1)\hat{p}_k} \left[\sum_{k=1}^n \sum_{i \neq k}^n I(Y_i \in \hat{G}_k) I(x_{ij} \geq x_{kj}) \right] \\
 &= - \left[\frac{1}{n(n+1)\hat{p}_k} \sum_{i=1}^n I(Y_i \in \hat{G}_k) \xi'_{ij} - \frac{n-1}{2(n+1)} \right],
 \end{aligned}$$

where continuous variable x_j satisfies $\Pr(x_{ij} = x_{kj}) = 0$. Define $\xi'_{ij} = \sum_{k \neq i}^n I(X_{ij} \geq X_{kj})$; then, ξ'_{ij} has the same distribution with $S_{n\hat{p}_k, n}$; as a result, ξ'_{ij} has the same distribution

with ξ_{ij} . In other words, $\hat{\tau}_{jk}$ and $-\hat{\tau}_{jk}$ have the same distribution, which is $F_{\hat{\tau}_{jk}}(u) = 1 - F_{\hat{\tau}_{jk}}(-u)$. If $\lim_{n \rightarrow \infty} \hat{p}_k(1 - \hat{p}_k) > 0$, we have the following relationship by using Theorem 3.4, Corollary 3.4, Corollary 3.5, and Corollary 3.6 of Mohamed and Mirakhmedov (2016) [26], and we can easily obtain that

- For all $u = o(\sqrt{n}) \geq 0$, we obtain that

$$\frac{1 - F_{\hat{\tau}_{jk}}(u)}{1 - \Phi(u)} = \frac{F_{\hat{\tau}_{jk}}(-u)}{\Phi(-u)} = \exp\left\{\frac{u^3}{\sqrt{n}}L_n\left(\frac{u}{\sqrt{n}}\right)\right\}\left(1 + O\left(\frac{u+1}{\sqrt{n}}\right)\right). \tag{A2}$$

where $L_n(v)$ is an odd power series that, for all sufficiently large N , is majorized by a power series with coefficients not depending on N , and is convergent in some discussions, and $L_n(v)$ converges uniformly in n for sufficiently small values of v , where $L_n(0) = 0$.

- For all $u = o(n^{1/4}) \geq 0$, we obtain that

$$\begin{aligned} \frac{1 - F_{\hat{\tau}_{jk}}(u)}{1 - \Phi(u)} &= \frac{F_{\hat{\tau}_{jk}}(-u)}{\Phi(-u)} = \exp\left\{\frac{u^3}{\sqrt{n}}O\left(\frac{u}{\sqrt{n}}\right)\right\}\left(1 + O\left(\frac{u+1}{\sqrt{n}}\right)\right) \\ &= \left(1 + O\left(\frac{u^4}{n}\right)\right)\left(1 + O\left(\frac{u+1}{\sqrt{n}}\right)\right). \end{aligned} \tag{A3}$$

- For all $u = o(n^{1/6}) \geq 0$, we obtain that

$$\frac{1 - F_{\hat{\tau}_{jk}}(u)}{1 - \Phi(u)} = \frac{F_{\hat{\tau}_{jk}}(-u)}{\Phi(-u)} = 1 + O\left(\frac{(u+1)^3}{\sqrt{n}}\right). \tag{A4}$$

- For all $0 \leq u \leq C \min\left(\sqrt{n^*\hat{q}}/\max|\hat{Y}_{mN}|, (n^*\hat{q})^{1/6}/B_3^{1/3}\right)$, we have

$$1 - F_{\hat{\tau}_{jk}}(u) = (1 - \Phi(u))\left(1 + O\left(\frac{(u+1)^3 B_3}{\sqrt{n^*\hat{q}}}\right)\right) \tag{A5}$$

Based on the above cases, using Taylor’s expansion of the Equations (A2)–(A5), the following conclusions can be obtained through the order correlation of u and n : if $\mathbb{H}_{0,j,k}(j = 1, 2, \dots, p; k = 1, 2, \dots, K)$ is true, then

$$\frac{1 - F_{\hat{\tau}_{jk}}(u)}{1 - \Phi(u)} = \frac{F_{\hat{\tau}_{jk}}(-u)}{\Phi(-u)} = 1 + O\left(n^{-1/2}\right),$$

where $\Phi(-u) = 1/2p = O(n^\alpha)$ and $\alpha \leq 1/2$. In other words, $\hat{\tau}_{jk}$ has the asymptotic normal distribution $\mathcal{N}(0, (1 - \hat{p}_k)/(12(n + 1)\hat{p}_k))$ for all $j \in \{1, 2, \dots, p\}$ and all $k \in \{1, 2, \dots, K\}$. \square

Appendix A.4. Proof of Corollary 1

Proof. By definition of \hat{v}_{jk} , we can easily find that $12(n + 1)\hat{v}_{jk}$ has the asymptotic distribution as χ_1^2 , where χ_1^2 is the chi-square distribution with degree of freedom 1. Since $\sum_{k=1}^K \hat{p}_k \hat{\tau}_{jk} = 0$, then $12(n + 1)\sum_{k=1}^K \hat{v}_{jk} = 12(n + 1)\hat{v}_j$ has the asymptotic distribution as χ_{K-1}^2 with degree of freedom $K - 1$. \square

Appendix A.5. Proof of Theorem 1

Proof. For some $j = 1, \dots, p$, let $\{X_{ij} : i = 1, \dots, n\}$ be a random sample of X_j . Some notations are employed. Let $p_k = \Pr(Y \in G_k)$ and $p_k^* = \Pr(Y \in \hat{G}_k)$. Write $W_k = \mathbb{I}(Y \in G_k)$,

$$W_k^* = I(Y \in \hat{G}_k), W_{ki} = I(Y_i \in \hat{G}_k), f_j(k, x) = I(X_j \leq x, Y \in \hat{G}_k), f_j^*(k, x) = I(X_j \leq x, Y \in \hat{G}_k), f_{ij}(k, x) = I(X_{ij} \leq x, Y_i \in \hat{G}_k), \zeta_j(k) = \mathbb{E}_X\{F_{jk}(x)\},$$

$$\begin{aligned} \zeta_j^*(k) &= \mathbb{E}_X\{F_{jk}^*(x)\} = \mathbb{E}_X\{\Pr(X_j \leq x \mid Y \in \hat{G}_k)\}, \\ \tilde{\zeta}_j(k) &= \frac{1}{n^2} \sum_{l=1}^n \sum_{i=1}^n \frac{I(X_{ij} \leq X_{lj}, Y_i \in \hat{G}_k)}{\hat{p}_k}, \end{aligned}$$

and

$$\hat{\zeta}_j(k) = \frac{1}{n(n+1)} \sum_{l=1}^n \sum_{i=1}^n \frac{I(X_{ij} \leq X_{lj}, Y_i \in \hat{G}_k)}{\hat{p}_k}.$$

Let $F_n(\cdot)$ be the empirical distribution function. By Hoeffding’s inequality, for any k and certain constant $c_8 > 0$,

$$\Pr\{|F(Q_k) - F_n(Q_k)| \geq \epsilon\} \leq \exp(-2nc_8\epsilon^2)$$

hold for any $\epsilon \in (0, 1)$, where \mathbb{R}_{x_j} is the support of a continuous variable $x_j, k = 1, \dots, K$ and $= 1, \dots, p$. Due to the fact that $W_k^* - W_k = I(\hat{Q}_{k-1} \leq Y < Q_{k-1}) + I(Q_k \leq Y < \hat{Q}_k)$, we obtain that

$$\begin{aligned} &\Pr\{\mathbb{E}(|W_k^* - W_k|) \geq 2\epsilon\} \\ &= 1 - \Pr\{\mathbb{E}(|W_k^* - W_k|) < 2\epsilon\} \\ &\leq 1 - \Pr\{|F(Q_k) - F_n(Q_k)| < \epsilon, |F(Q_{k-1}) - F_n(Q_{k-1})| < \epsilon\} \\ &\leq 1 - (1 - \exp(-2nc_8\epsilon^2))^2 \\ &\leq 2\exp(-2nc_8\epsilon^2) + o(\exp(-2nc_8\epsilon^2)) \\ &= 2\exp(-2nc_8\epsilon^2). \end{aligned} \tag{A6}$$

Similarly, due to the fact that $p_k^* - p_k = (F(Q_k) - F_n(Q_k)) - (F(Q_{k-1}) - F_n(Q_{k-1}))$, we have

$$\begin{aligned} &\Pr\{|p_k^* - p_k| \geq 2\epsilon\} = 1 - \Pr\{|p_k^* - p_k| < 2\epsilon\} \\ &\leq 1 - \Pr\{|F(Q_k) - F_n(Q_k)| < \epsilon, |F(Q_{k-1}) - F_n(Q_{k-1})| < \epsilon\} \\ &\leq 2\exp(-2nc_8\epsilon^2). \end{aligned} \tag{A7}$$

Note that

$$\begin{aligned} |\zeta_j^*(k) - \zeta_j(k)| &= \left| \frac{\mathbb{E}\{f_j^*(k, x)\}}{p_k^*} - \frac{\mathbb{E}\{f_j(k, x)\}}{p_k} \right| \\ &\leq \frac{\mathbb{E}\{f_j^*(k, x)\}|p_k^* - p_k|}{p_k p_k^*} + \frac{|\mathbb{E}\{f_j^*(k, x) - f_j(k, x)\}|}{p_k} \\ &\leq \sup_{x \in \mathbb{R}_{x_j}} \frac{\mathbb{E}\{f_j^*(k, x)\}|p_k^* - p_k|}{p_k p_k^*} + \sup_{x \in \mathbb{R}_{x_j}} \frac{|\mathbb{E}\{I(X_j \leq x) \cdot (W_j^*(k) - W_j(k))\}|}{p_k} \\ &= \frac{|p_k^* - p_k| + \mathbb{E}(|W_k^* - W_k|)}{p_k}, \end{aligned}$$

where the last equality holds due to the fact that

$$\sup_{x \in \mathbb{R}_{x_j}} \mathbb{E}\{f_j^*(k, x)\} = \sup_{x \in \mathbb{R}_{x_j}} \Pr(x_j \leq x, Y \in \hat{G}_k) = p_k^*$$

and

$$\begin{aligned} & \sup_{x \in \mathbb{R}_{x_j}} \mathbb{E} \left| \{f_j^*(k, x) - f_j(k, x)\} \right| = \sup_{x \in \mathbb{R}_{x_j}} \left| \mathbb{E} \{I(X_j \leq x) \cdot (W_j^*(k) - W_j(k))\} \right| \\ &= \sup_{x \in \mathbb{R}_{x_j}} \Pr(x_j \leq x, Y \in G_k, Y \notin \hat{G}_k) + \sup_{x \in \mathbb{R}_{x_j}} \Pr(x_j \leq x, Y \notin G_k, Y \in \hat{G}_k) \\ &= \mathbb{E}(|W_k^* - W_k|). \end{aligned}$$

According to Equations (A6) and (A8), we obtain that

$$\begin{aligned} & \Pr(|\zeta_j^*(k) - \zeta_j(k)| \geq \epsilon) \\ & \leq \Pr \left(\frac{|p_k^* - p_k| + \mathbb{E}(|W_k^* - W_k|)}{p_k} \geq \epsilon \right) \\ & \leq \Pr(|p_k^* - p_k| + \mathbb{E}(|W_k^* - W_k|) \geq \epsilon c_1 / 2K) \\ & \leq \Pr(|p_k^* - p_k| \geq \epsilon c_1 / 4K) + \Pr(\mathbb{E}(|W_k^* - W_k|) \geq \epsilon c_1 / 4K) \\ & \leq 4 \exp(-nc_9 \epsilon^2 / K^2) \end{aligned} \tag{A8}$$

hold for any $\epsilon \in (0, 1), k = 1, \dots, K$ and $= 1, \dots, p$. According to Lemmaa 1 and 2 of Xie et al. (2020) [17], under Conditions (C1) and (C3), for any $\epsilon \in (0, 1/2)$ and $j = 1, \dots, p$, there exists a positive constant c_3 , which satisfies that

$$\Pr \left\{ \left| \tilde{\zeta}_j(k) - \zeta_j^*(k) \right| \geq \epsilon \right\} \leq 4(n + 2) \exp(-c_3 n \epsilon^2 / R). \tag{A9}$$

Let

$$\hat{\zeta}_j(k) - \zeta_j^*(k) \equiv \frac{n}{n + 1} \left\{ \tilde{\zeta}_j(k) - \zeta_j^*(k) \right\} + \Delta,$$

where $\Delta = -\zeta_j(k)^* / (n + 1)$.

It follows from Condition (C3) and Equations (A5) and (A9) that $K = O(n^\xi)$ for $\xi + \kappa < 1/2$. By letting $\epsilon = 2 * c_3^* n^{-\kappa} = 2c_3^* n^{-1/2\beta}$ for $0 \leq \kappa < 1/4$ and $c_3^* > 0$, we have

$$\begin{aligned} & \Pr \left(\max_{1 \leq j \leq p} \left| \hat{\tau}_{jk} - \tau_{j,k} \right| \geq 2c_3^* n^{-\kappa} \right) \leq p \Pr \left(\left| \hat{\tau}_{jk} - \tau_{j,k} \right| \geq 2c_3^* n^{-\kappa} \right) \\ &= p \Pr \left\{ \left| \hat{\zeta}_j(k) - \zeta_j(k) \right| \geq 2c_3^* n^{-\kappa} \right\} \\ &= p \Pr \left\{ \left| (\hat{\zeta}_j(k) - \zeta_j^*(k)) + (\zeta_j^*(k) - \zeta_j(k)) \right| \geq 2c_3^* n^{-\kappa} \right\} \\ & \leq p \left(1 - \Pr \left\{ \left| \hat{\zeta}_j(k) - \zeta_j^*(k) \right| < c_3^* n^{-\kappa} \right\} \cdot \Pr \left\{ \left| \zeta_j^*(k) - \zeta_j(k) \right| < c_3^* n^{-\kappa} \right\} \right) \\ & \leq p(1 - p_{I_1} \cdot p_{I_2}), \end{aligned}$$

where $p_{I_1} \equiv \Pr \left\{ \left| \hat{\zeta}_j(k) - \zeta_j^*(k) \right| < c_3^* n^{-\kappa} \right\}$ and $p_{I_2} \equiv \Pr \left\{ \left| \zeta_j^*(k) - \zeta_j(k) \right| < c_3^* n^{-\kappa} \right\}$. For p_{I_1} , we have that

$$\begin{aligned} p_{I_1} &= 1 - \Pr \left\{ \left| \hat{\zeta}_j(k) - \zeta_j^*(k) \right| \geq c_3^* n^{-\kappa} \right\} \\ & \geq 1 - \Pr \left\{ \left| \tilde{\zeta}_j(k) - \zeta_j^*(k) \right| \geq \frac{(n + 1)(c_4^* n^{-\kappa} - |\Delta|)}{n} \right\} \\ & \geq 1 - 4(n + 2) \exp(-c_{10} n^{1-2\kappa-\xi}), \end{aligned}$$

where c_{10} is a positive constant. For p_{I_2} , we have that

$$p_{I_2} = 1 - \Pr \left\{ \left| \zeta_j^*(k) - \zeta_j(k) \right| \geq c_3^* n^{-\kappa} \right\} \geq 1 - 4 \exp(-c_{11} n^{1-2\kappa-2\xi})$$

where $c_{11} > 0$ is a constant. The last inequality above holds due to the fact that $\Delta = O(1/n)$. Thus,

$$\begin{aligned} & \Pr\left(\max_{1 \leq j \leq p} |\hat{\tau}_{jk} - \tau_{j,k}| \geq 2c_3^* n^{-\kappa}\right) \\ & \leq p \left[1 - \left(1 - 4(n+2) \exp(-c_{10} n^{1-2\kappa-\xi})\right) \cdot \left(1 - 4 \exp(-c_{11} n^{1-2\kappa-2\xi})\right) \right] \\ & \leq 8p \exp(-c_4 n^{1-2\kappa-2\xi}) \end{aligned}$$

Let $c_3 = 4\hat{p}_k / (1 - \hat{p}_k)(c_3^*)^2$; by the definition of \hat{v}_{jk} , we have

$$\Pr\left(\max_{1 \leq j \leq p} |\hat{v}_{jk} - v_{j,k}| \geq c_4 n^{-\kappa}\right) \leq 8p \exp(-c_4 n^{1-2\kappa-\xi}).$$

□

Appendix A.6. Proof of Theorem 2

Proof. It follows from Condition (C2) and Theorem 1 that

$$\begin{aligned} & \Pr\left(\min_{j_1 \in \mathcal{A}_{k_1}} |\hat{v}_{jk}| - \max_{j_2 \notin \mathcal{A}_{k_2}} |\hat{v}_{jk}| < \frac{\rho_0}{2}\right) \\ & \leq \Pr\left\{\left(\min_{j_1 \in \mathcal{A}_{k_1}} |\hat{v}_{jk}| - \max_{j_2 \notin \mathcal{A}_{k_2}} |\hat{v}_{jk}|\right) - \left(\min_{j_1 \in \mathcal{A}_{k_1}} |v_{j,k}| - \max_{j_2 \notin \mathcal{A}_{k_2}} |v_{j,k}|\right) < -\frac{\rho_0}{2}\right\} \\ & \leq \Pr\left\{\left|\left(\min_{j_1 \in \mathcal{A}_{k_1}} |\hat{v}_{jk}| - \max_{j_2 \notin \mathcal{A}_{k_2}} |\hat{v}_{jk}|\right) - \left(\min_{j_1 \in \mathcal{A}_{k_1}} |v_{j,k}| - \max_{j_2 \notin \mathcal{A}_{k_2}} |v_{j,k}|\right)\right| > \frac{\rho_0}{2}\right\} \\ & \leq \Pr\left(2 \max_{1 \leq j \leq p} |\hat{v}_{jk} - v_{j,k}| > \frac{\rho_0}{2}\right) \\ & \leq 8p \exp(-c_7 n \rho_0^2 / K), \end{aligned}$$

where $c_7 > 0$ is a constant. $K \log(p) = o(n\rho_0^2)$ ensures that there exists some $n_0 > 0$ for $n > n_0$, $p \leq \exp(c_7 n \rho_0^2 / 2K)$. Consequently, we have that $\liminf_{n \rightarrow \infty} \left\{ \min_{j_1 \in \mathcal{A}_{k_1}} \hat{v}_{jk} - \max_{j_2 \notin \mathcal{A}_{k_2}} \hat{v}_{jk} \right\} > 0$ almost surely.

□

Appendix A.7. Proof of Theorem 3

Proof. Due to the equivalence of $12 \cdot (n+1) \cdot \hat{v}_{jk} \geq \rho$ and $|\hat{\tau}_{jk}| \geq \sqrt{\rho \frac{1-\hat{p}_k}{12(n+1)\hat{p}_k}}$, proof of Theorem 3 will be obtained by $|\hat{\tau}_{jk}|$. Note that

$$\begin{aligned} F_{|\hat{\tau}_{jk}|}(u) & := \Pr\left\{|\hat{\tau}_{jk}| < u \sqrt{\frac{1-\hat{p}_k}{12(n+1)\hat{p}_k}}\right\}, \quad (j \in \{j : j \notin \mathcal{A}_k\}; k = 1, \dots, K); \\ F_{|\hat{\tau}_k|}(u) & := \Pr\left\{|\hat{\tau}_{jk}| < u \sqrt{\frac{1-\hat{p}_k}{12(n+1)\hat{p}_k}}, \forall j \in \{j : j \notin \mathcal{A}_k\}\right\}, \quad (k = 1, \dots, K); \\ F_{|\hat{\tau}_k|^c}^c(u) & := \Pr\left\{|\hat{\tau}_{jk}| \geq u \sqrt{\frac{1-\hat{p}_k}{12(n+1)\hat{p}_k}}, \exists j \in \{j : j \notin \mathcal{A}_k\}\right\}, \quad (k = 1, \dots, K). \end{aligned}$$

Then, denote $q_k = p - s_k$, and we have

$$F_{|\hat{\tau}_{jk}|}(u) = F_{\hat{\tau}_{jk}}(u) - F_{\hat{\tau}_{jk}}(-u) = 1 - 2F_{\hat{\tau}_{jk}}(-u),$$

$$F_{|\hat{\tau}_k|}(u) = [F_{|\hat{\tau}_{jk}|}(u)]^{q_k} = [1 - 2\Pr_{\hat{\tau}_{jk}}(-u)]^{q_k},$$

$$F_{|\hat{\tau}_k|}^c(u) = 1 - F_{|\hat{\tau}_k|}(u) = 1 - [1 - 2\Pr_{\hat{\tau}_{jk}}(-u)]^{q_k}.$$

When $\beta \in (2, +\infty)$, we have $\rho = o(n^{1/4})$ as a constant. Hence, we obtain

$$\frac{1 - F_{\hat{\tau}_{jk}}(\sqrt{\rho})}{1/2p} = \frac{F_{\hat{\tau}_{jk}}(-\sqrt{\rho})}{1/2p} = 1 + O(n^{-1/2}). \tag{A10}$$

By the definition of $F_{|\hat{\tau}_k|}^c(x)$, it implies that

$$F_{|\hat{\tau}_k|}^c(k) = 1 - \left\{ 1 - p^{-1} \left[1 + O(n^{-1/2}) \right] \right\}^{q_k}. \tag{A11}$$

From the definition of ultra-high dimensional data, we have $p = o(\exp\{n^\alpha\})$, $\alpha > 0$ and $s_k = o(n)$. If $\alpha > 1/2$, according to Equations (A10) and (A11), we have

$$\begin{aligned} \left\{ 1 - p^{-1} \left[1 + O(n^{-1/2}) \right] \right\}^{q_k} &= \exp \left\{ q_k \cdot \log \left\{ 1 - p^{-1} \left[1 + O(n^{-1/2}) \right] \right\} \right\} \\ &= \exp \left\{ -q_k \cdot \left(p^{-1} \left[1 + O(n^{-1/2}) \right] + o(1/p) \right) \right\} \\ &= \exp \left\{ -1 + O(n^{-1/2}) \right\} \exp \{ o(1/p) \} \rightarrow e^{-1}. \end{aligned}$$

To conclude, $F_{|\hat{\tau}_k|}^c(k) \rightarrow 1 - e^{-1}$ a.s. $n \rightarrow \infty$; in other words,

$$\lim_{p \rightarrow \infty} \Pr\{\text{FD}_{k,\hat{\rho}_0} > 0\} = \lim_{p \rightarrow \infty} \Pr\left\{ \sum_{j \in \mathcal{A}_k^c} I(\hat{v}_{jk} \geq \rho) > 0 \right\} = 1 - e^{-1} \quad (k = 1, \dots, K).$$

The number of variables screened into the adaptive FD set is subjecting to the p times Bernoulli test; then, the expectation and variance of the number of the false discovery are

$$\begin{aligned} \text{EFD} &= q_k \cdot 1/p = (1 + O(n^{-1/2}))(1 - o(p^{-1})) = 1 + O(n^{-1/2}), \\ \text{VFD} &= q_k \cdot 1/p \cdot (1 - 1/p) = (1 + O(n^{-1/2}))(1 - o(p^{-1}))^2 = 1 + O(n^{-1/2}). \end{aligned}$$

□

Appendix A.8. Proof of Corollary 2

Proof. According to the Condition (C2), the definition of $\hat{\rho}_0$ in Section 3.1 and $\max_{j \in \mathcal{A}_k} |\hat{v}_{jk} - v_{jk}| \leq cn^{-\kappa}$ in Theorem 1, we have

$$\min_{j \in \mathcal{A}_k} |\hat{v}_{jk}| \geq \min_{j \in \mathcal{A}_k} (|v_{jk}| - |\hat{v}_{jk} - v_{jk}|) \geq \min_{j \in \mathcal{A}_k} |v_{jk}| - \max_{j \in \mathcal{A}_k} |v_{jk} - \hat{v}_{jk}| \geq cn^{-\kappa}.$$

Therefore, we obtain that

$$\Pr(\mathcal{A}_k \subset \hat{\mathcal{A}}_{k,\hat{\rho}_0}) \geq \Pr\left(\max_{j \in \mathcal{A}_k} |\hat{v}_{jk} - v_{jk}| \leq cn^{-\kappa} \right) \geq 1 - 8ps_r \exp(-c_6 n^{1-2\kappa-\xi})$$

holds for some constant $c_6 > 0$.

□

Appendix A.9. Proof of Theorem 4

Proof. Similar to Proof of Corollary 2, it is clear that

$$\Pr(\mathcal{A}_k \subset \hat{\mathcal{A}}_{k,\alpha}) \geq 1 - 8ps_k \exp(-c_7 n^{1-2\kappa-\xi})$$

holds for some constant $c_7 > 0$.

In order to prove $\widehat{\text{FDR}}_{\hat{\rho}_k} \rightarrow \alpha$ in probability, under the assumption that $q_k/p \rightarrow 1$ as $p \rightarrow \infty$ and for any $\rho > 0$, by Corollary 1 and the Hoeffding’s inequality, it suffices to show that

$$\sup_{0 \leq \rho \leq \epsilon} \left[\frac{\sum_{j \in \mathcal{A}_k^c} I(\hat{v}_{jk} \geq \rho)}{pS_{\chi_1^2}(\rho)} \right] \rightarrow 1 \tag{A12}$$

in probability as $n \rightarrow \infty$, where $\epsilon > 0$ is a constant $S_{\chi_1^2}(\rho) = 1 - F_{\chi_1^2}(\rho)$, the survival function of the distribution of χ_1^2 . Thus,

$$\begin{aligned} \text{FDP}_{k,\rho} &= \frac{\text{FD}_{k,\rho}}{\max\{\sum_{j \in \mathcal{I}} I(\hat{v}_{jk} \geq \rho), 1\}} = \frac{\sum_{j \in \mathcal{A}_k^c} I(\hat{v}_{jk} \geq \rho)/q_k}{\max\{\sum_{j \in \mathcal{I}} I(\hat{v}_{jk} \geq \rho), 1\}/p} \\ &= \frac{pS_{\chi_1^2}(\rho)}{\max\{\sum_{j \in \mathcal{I}} I(\hat{v}_{jk} \geq \rho), 1\}} = \frac{pS_{\chi_1^2}(\rho)}{\max\{pS_{\chi_1^2}(\rho) + \sum_{j \in \mathcal{A}_k} I(\hat{v}_{jk} \geq \rho), 1\}}. \end{aligned}$$

Notice that $\sum_{j \in \mathcal{A}_k} I(\hat{v}_{jk} \geq \rho)$ is monotone in ρ and asymptotically converges to s_k , and $S_{\chi_1^2}(\rho)$ is continuous and monotone. Then, there exists a unique constant $0 < \hat{\rho}_k \leq Cn^{-\beta}$ such that

$$\frac{pS_{\chi_1^2}(\hat{\rho}_k)}{\max\{\sum_{j \in \mathcal{I}} I(\hat{v}_{jk} \geq \hat{\rho}_k), 1\}} = \alpha \tag{A13}$$

in probability as $n \rightarrow \infty$. Therefore, according to the Equations (A12) and (A13), we obtain that

$$\lim_{n \rightarrow \infty} \frac{\widehat{\text{FDR}}_{\hat{\rho}_k}}{\alpha} = 1.$$

□

Appendix B

Table A1. The quantiles of minimum model size in Scenario 1.3 of Section 4.1.

Method	$\rho = 0$					$\rho = 0.5$				
	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%
	$p = 1000$									
QA-SVS-A(4)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	5.0
QA-SVS-A(5)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	5.0
QA-SVS-A(6)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	5.0
QA-SVS-A(7)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	5.0
QA-SVS-A(8)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	5.0
QA-SVS-A(9)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	5.0
QA-SVS-A(10)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	5.0
QCS(4)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.5	5.0
QCS(5)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	7.0
QCS(6)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	5.5
QCS(7)	3.0	3.0	3.0	3.0	3.0	3.5	3.0	3.0	4.0	11.5
QCS(8)	3.0	3.0	3.0	3.0	3.0	4.0	3.0	3.0	4.0	12.0
QCS(9)	3.0	3.0	3.0	3.0	3.0	4.5	3.0	3.0	4.0	7.5
QCS(10)	3.0	3.0	3.0	3.0	3.0	3.5	3.0	3.0	4.0	10.0
SIS	3.0	3.0	3.0	3.0	3.0	6.5	3.0	3.0	5.0	6.0
DC-SIS	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	4.5
QA-SIS(0.1)	3.0	3.0	3.0	3.0	3.0	4.5	3.0	3.0	3.5	16.5
QA-SIS(0.3)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QA-SIS(0.5)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	5.0
QA-SIS(0.7)	3.0	3.0	3.0	3.0	3.0	4.0	3.0	3.0	4.0	10.5
QA-SIS(0.9)	4.0	15.0	32.5	74.5	217.0	4.5	14.5	36.5	81.5	299.0

Table A1. *Cont.*

Method	$\rho = 0$					$\rho = 0.5$				
	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%
	$p = 5000$									
QA-SVS-A(4)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	5.0
QA-SVS-A(5)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	5.0
QA-SVS-A(6)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	6.0
QA-SVS-A(7)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	8.0
QA-SVS-A(8)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	6.0
QA-SVS-A(9)	3.0	3.0	3.0	3.0	3.5	3.0	3.0	3.0	4.0	7.0
QA-SVS-A(10)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	4.0	6.5
QCS(4)	3.0	3.0	3.0	3.0	4.0	3.0	3.0	3.0	4.0	6.0
QCS(5)	3.0	3.0	3.0	3.0	7.0	3.0	3.0	3.0	4.0	6.0
QCS(6)	3.0	3.0	3.0	3.0	5.0	3.0	3.0	3.0	4.0	54.0
QCS(7)	3.0	3.0	3.0	3.0	5.0	3.0	3.0	3.0	4.0	32.0
QCS(8)	3.0	3.0	3.0	3.0	8.0	3.0	3.0	3.0	4.0	60.5
QCS(9)	3.0	3.0	3.0	3.0	6.5	3.0	3.0	3.0	4.0	19.5
QCS(10)	3.0	3.0	3.0	3.0	7.5	3.0	3.0	4.0	6.0	92.0
SIS	3.0	3.0	3.0	3.0	18.5	3.0	3.0	4.0	5.0	11.5
DC-SIS	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	5.0
QA-SIS(0.1)	3.0	3.0	3.0	3.0	31.0	3.0	3.0	3.0	4.0	28.0
QA-SIS(0.3)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.5
QA-SIS(0.5)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	7.5
QA-SIS(0.7)	3.0	3.0	3.0	4.0	18.5	3.0	3.0	3.0	7.0	27.5
QA-SIS(0.9)	16.0	58.5	130.5	323.5	1382.0	6.0	31.0	124.5	421.5	1844.0

Notes: QA-SVS-A(4), QA-SVS-A(5), . . . , and QA-SVS-A(10), our proposed method defined in Remark 3 with different quantile grid points ($K = 4, \dots, 10$); QCS-A(4), QCS-A(5), . . . , and QCS-A(10), the quantile correlation-based screening method (Tang et al., 2013) [19] with different quantile grid points ($K = 4, \dots, 10$); SIS, the sure independence screening (Fan and Lv, 2008) [1]; DC-SIS, the distance correlation-based screening (Li et al., 2012) [8]; QA-SIS(0.1), QA-SIS(0.3), . . . , QA-SIS(0.9), the quantile-adaptive model-free sure independence screening (He et al., 2013) [9] at different quantile levels.

Table A2. The quantiles of minimum model size in Scenario 1.4 of Section 4.1.

Method	$\rho = 0$					$\rho = 0.5$				
	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%
	$p = 1000$									
QA-SVS-A(4)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(5)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(6)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(7)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(8)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(9)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(10)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QCS(4)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0
QCS(5)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	7.0
QCS(6)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	5.0
QCS(7)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	5.0
QCS(8)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	8.0
QCS(9)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	6.5
QCS(10)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	7.5
SIS	414.0	686.0	781.5	894.0	979.5	558.5	706.5	824.5	932.0	988.0
DC-SIS	441.5	619.0	742.5	840.0	962.0	341.5	601.0	747.5	895.0	971.0
QA-SIS(0.1)	135.5	223.0	321.5	428.5	626.5	82.5	134.5	201.5	303.0	543.0
QA-SIS(0.3)	14.0	28.0	49.5	93.5	237.5	10.0	24.5	65.0	116.0	593.5
QA-SIS(0.5)	8.0	20.0	32.0	54.5	162.5	3.0	5.0	6.0	8.0	15.0
QA-SIS(0.7)	37.5	73.0	146.5	223.0	394.5	9.0	15.5	20.0	32.5	56.5
QA-SIS(0.9)	152.5	291.0	418.0	562.0	816.0	60.5	145.0	215.5	307.5	548.5

Table A2. *Cont.*

Method	$\rho = 0$					$\rho = 0.5$				
	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%
	$p = 5000$									
QA-SVS-A(4)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(5)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(6)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(7)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(8)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(9)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(10)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QCS(4)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QCS(5)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QCS(6)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QCS(7)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QCS(8)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QCS(9)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
QCS(10)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
SIS	2126.0	3353.0	3956.5	4469.5	4874.5	3038.5	3633.5	4071.0	4528.5	4954.5
DC-SIS	1949.0	3324.0	4045.5	4387.5	4832.0	1485.5	3097.5	3959.0	4368.0	4702.0
QA-SIS(0.1)	507.5	1137.5	1470.0	1981.5	3183.0	433.0	747.5	1177.5	1573.5	2852.5
QA-SIS(0.3)	54.5	97.0	177.0	348.5	1122.5	39.0	126.0	288.0	595.5	2742.0
QA-SIS(0.5)	33.5	95.0	184.5	351.5	711.0	6.0	11.0	15.0	29.0	62.5
QA-SIS(0.7)	143.0	419.5	690.0	1263.5	2217.5	34.5	60.5	100.0	175.0	402.0
QA-SIS(0.9)	670.5	1140.0	1775.5	2586.5	3707.5	439.5	717.0	1022.0	1547.0	2662.5

Notes: QA-SVS-A(4), QA-SVS-A(5), . . . , and QA-SVS-A(10), our proposed method defined in Remark 3 with different quantile grid points ($K = 4, \dots, 10$); QCS-A(4), QCS-A(5), . . . , and QCS-A(10), the quantile correlation-based screening method (Tang et al., 2013) [19] with different quantile grid points ($K = 4, \dots, 10$); SIS, the sure independence screening (Fan and Lv, 2008) [1]; DC-SIS, the distance correlation-based screening (Li et al., 2012) [8]; QA-SIS(0.1), QA-SIS(0.3), . . . , QA-SIS(0.9), the quantile-adaptive model-free sure independence screening (He et al., 2013) [9] at different quantile levels.

Table A3. The quantiles of minimum model size in Scenario 1.5 of Section 4.1.

Method	$\rho = 0$					$\rho = 0.5$				
	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%
	$p = 1000$									
QA-SVS-A(4)	3.0	4.0	9.0	28.5	317.0	3.0	3.0	6.0	24.0	156.5
QA-SVS-A(5)	3.0	3.0	8.0	40.0	179.5	3.0	3.0	4.0	8.0	42.5
QA-SVS-A(6)	3.0	3.0	5.0	12.0	93.5	3.0	3.0	4.0	5.0	23.0
QA-SVS-A(7)	3.0	3.0	5.0	13.0	82.0	3.0	3.0	4.0	9.5	59.5
QA-SVS-A(8)	3.0	3.0	4.0	11.5	52.0	3.0	3.0	3.0	7.5	71.0
QA-SVS-A(9)	3.0	3.0	4.0	11.0	47.5	3.0	3.0	3.0	6.0	53.5
QA-SVS-A(10)	3.0	3.0	4.0	10.5	89.5	3.0	3.0	4.0	6.5	59.0
QCS(4)	3.0	3.0	4.0	10.0	46.0	3.0	3.0	3.0	6.0	60.5
QCS(5)	3.0	3.0	4.0	8.0	58.0	3.0	3.0	4.0	6.5	66.5
QCS(6)	3.0	3.0	4.0	11.0	251.0	3.0	3.0	3.0	6.0	44.0
QCS(7)	3.0	3.0	4.0	11.5	76.5	3.0	3.0	4.0	6.0	58.5
QCS(8)	3.0	3.0	6.0	17.5	118.0	3.0	3.0	4.0	9.5	76.5
QCS(9)	3.0	4.0	6.0	22.5	118.0	3.0	3.0	4.0	15.0	66.5
QCS(10)	3.0	4.0	7.0	20.0	156.5	3.0	3.0	5.0	9.5	72.5
SIS	260.0	559.0	802.5	922.5	981.5	227.5	574.5	769.5	907.0	988.0
DC-SIS	3.0	3.0	6.0	28.0	450.0	3.0	3.0	4.0	19.5	350.0
QA-SIS(0.1)	82.0	168.5	300.0	521.5	849.5	70.0	132.0	241.5	395.0	737.5
QA-SIS(0.3)	10.5	21.0	42.5	98.0	195.0	4.5	12.0	21.0	53.0	306.0
QA-SIS(0.5)	5.5	20.5	56.0	166.5	517.5	3.0	9.5	24.5	90.5	439.5
QA-SIS(0.7)	4.0	11.0	30.5	85.5	309.0	7.5	15.5	35.5	114.5	355.0
QA-SIS(0.9)	93.5	191.0	382.0	587.0	798.5	111.5	259.5	467.5	660.5	852.5

Table A3. *Cont.*

Method	$\rho = 0$					$\rho = 0.5$				
	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%
	$p = 5000$									
QA-SVS-A(4)	3.0	12.0	58.5.0	200.0	831.0	3.0	5.0	13.0	81.0	615.5
QA-SVS-A(5)	3.0	6.0	18.0	98.0	791.0	3.0	4.0	8.0	33.0	242.5
QA-SVS-A(6)	3.0	4.0	9.0	45.0	322.0	3.0	4.0	8.0	31.5	241.0
QA-SVS-A(7)	3.0	4.0	6.5	42.0	549.5	3.0	4.0	7.0	29.0	165.0
QA-SVS-A(8)	3.0	4.0	7.0	26.0	152.5	3.0	3.0	6.0	19.5	183.5
QA-SVS-A(9)	3.0	4.0	12.5	70.0	552.0	3.0	3.0	5.5	19.0	423.5
QA-SVS-A(10)	3.0	5.0	12.5	44.0	539.0	3.0	3.0	6.0	28.5	334.5
QCS(4)	3.0	3.0	6.0	14.5	166.0	3.0	3.0	4.0	14.5	163.5
QCS(5)	3.0	4.0	6.5	30.0	209.5	3.0	3.0	4.0	8.5	73.5
QCS(6)	3.0	4.0	9.5	40.0	594.5	3.0	3.0	5.0	17.0	160.5
QCS(7)	3.0	4.0	12.0	50.0	417.0	3.0	3.0	7.0	25.0	125.5
QCS(8)	3.0	5.0	20.0	64.5	619.0	3.0	3.0	7.0	25.0	94.0
QCS(9)	3.0	6.0	22.0	97.0	642.5	3.0	4.0	9.0	39.5	324.0
QCS(10)	3.0	6.0	21.5	163.0	1131.5	3.0	4.0	8.0	37.5	431.5
SIS	1341.0	3294.5	4114.0	4586.0	4968.0	1252.5	3027.5	3972.0	4524.0	4937.5
DC-SIS	3.0	5.0	12.0	110.0	1952.0	3.0	4.0	10.0	57.5	1297.5
QA-SIS(0.1)	423.0	1111.5	1827.0	3080.0	4736.5	350.0	709.5	1175.5	1655.0	3101.0
QA-SIS(0.3)	19.0	77.0	198.0	556.0	1755.5	14.0	45.5	128.0	402.0	1319.0
QA-SIS(0.5)	13.5	82.5	377.5	787.0	2140.0	4.0	32.0	177.5	469.5	2936.0
QA-SIS(0.7)	16.5	48.0	177.5	472.0	1475.0	22.5	96.5	336.0	777.5	2261.5
QA-SIS(0.9)	549.0	1025.5	1727.0	2413.0	3997.5	434.5	1132.5	2093.5	3309.0	4599.5

Notes: QA-SVS-A(4), QA-SVS-A(5), . . . , and QA-SVS-A(10), our proposed method defined in Remark 3 with different quantile grid points ($K = 4, \dots, 10$); QCS-A(4), QCS-A(5), . . . , and QCS-A(10), the quantile correlation-based screening method (Tang et al., 2013) [19] with different quantile grid points ($K = 4, \dots, 10$); SIS, the sure independence screening (Fan and Lv, 2008) [1]; DC-SIS, the distance correlation-based screening (Li et al., 2012) [8]; QA-SIS(0.1), QA-SIS(0.3), . . . , QA-SIS(0.9), the quantile-adaptive model-free sure independence screening (He et al., 2013) [9] at different quantile levels.

Table A4. The quantiles of minimum model size in Scenario 1.6 of Section 4.1.

Method	$\rho = 0$					$\rho = 0.5$				
	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%
	$p = 1000$									
QA-SVS-A(4)	3.0	11.0	41.5	211.0	717.0	3.0	3.0	3.0	4.0	5.0
QA-SVS-A(5)	3.0	4.5	26.5	126.5	380.5	3.0	3.0	3.0	3.0	4.0
QA-SVS-A(6)	3.0	3.0	13.0	69.5	319.0	3.0	3.0	3.0	3.0	4.0
QA-SVS-A(7)	3.0	3.0	6.0	18.0	309.0	3.0	3.0	3.0	3.0	3.5
QA-SVS-A(8)	3.0	3.0	6.0	16.5	107.0	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(9)	3.0	3.0	4.5	10.0	94.0	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(10)	3.0	3.0	4.0	13.0	112.5	3.0	3.0	3.0	3.0	3.0
QCS(4)	4.0	31.5	95.3	329.5	730.5	3.0	3.0	3.0	5.5	71.0
QCS(5)	8.0	39.0	148.5	437.0	742.0	3.0	3.0	3.0	4.0	10.5
QCS(6)	4.0	28.5	88.0	287.0	664.5	3.0	3.0	3.0	4.0	46.0
QCS(7)	3.5	22.5	95.5	256.0	563.5	3.0	3.0	3.0	3.0	6.0
QCS(8)	3.0	19.0	67.0	201.0	535.0	3.0	3.0	3.0	3.0	11.0
QCS(9)	3.0	8.0	50.0	165.5	494.5	3.0	3.0	3.0	3.0	11.0
QCS(10)	3.0	9.5	38.0	124.0	539.0	3.0	3.0	3.0	3.0	7.5
SIS	3.0	3.0	3.0	3.0	7.0	3.0	3.0	3.0	3.0	3.0
DC-SIS	3.0	3.0	3.0	3.0	4.0	3.0	3.0	3.0	3.0	3.0
QA-SIS(0.1)	4.0	7.5	14.5	28.0	152.0	3.0	4.0	5.0	7.5	15.0
QA-SIS(0.3)	4.0	35.0	110.0	247.0	840.5	3.0	3.0	3.0	5.0	17.0
QA-SIS(0.5)	27.5	141.0	262.0	516.5	880.5	3.0	4.0	7.0	35.5	258.0
QA-SIS(0.7)	49.5	185.0	425.0	736.5	929.0	8.5	32.0	160.5	385.5	865.0
QA-SIS(0.9)	241.5	456.5	648.5	878.0	985.0	67.0	275.0	541.5	754.5	972.0

Table A4. Cont.

Method	$\rho = 0$					$\rho = 0.5$				
	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%
	$p = 5000$									
QA-SVS-A(4)	6.0	52.0	347.0	1262.5	3451.5	3.0	3.0	3.0	3.0	5.5
QA-SVS-A(5)	3.0	17.5	149.0	472.0	2285.5	3.0	3.0	3.0	3.5	6.0
QA-SVS-A(6)	3.0	11.0	47.5	293.0	1288.0	3.0	3.0	3.0	3.0	4.0
QA-SVS-A(7)	3.0	5.0	19.5	89.5	1288.0	3.0	3.0	3.0	3.0	4.0
QA-SVS-A(8)	3.0	3.0	10.0	95.0	923.0	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(9)	3.0	4.0	8.0	33.5	783.5	3.0	3.0	3.0	3.0	3.0
QA-SVS-A(10)	3.0	4.0	11.0	58.0	725.5	3.0	3.0	3.0	3.0	3.0
QCS(4)	40.5	329.0	853.5	1965.0	3801.0	3.0	3.0	4.0	32.0	277.0
QCS(5)	12.5	164.5	515.0	1677.5	3445.5	3.0	3.0	3.0	3.0	112.0
QCS(6)	6.5	62.5	262.0	917.0	2845.5	3.0	3.0	3.0	5.5	35.0
QCS(7)	18.5	105.0	404.5	937.0	2855.0	3.0	3.0	3.0	4.0	31.5
QCS(8)	5.5	84.0	333.0	788.5	3004.0	3.0	3.0	3.0	4.0	16.0
QCS(9)	3.0	46.0	173.5	596.5	1803.5	3.0	3.0	3.0	4.0	28.5
QCS(10)	6.5	40.5	213.5	939.0	2590.5	3.0	3.0	3.0	5.0	13.0
SIS	3.0	3.0	3.0	3.0	43.5	3.0	3.0	3.0	3.0	7.5
DC-SIS	3.0	3.0	3.0	3.0	4.0	3.0	3.0	3.0	3.0	3.0
QA-SIS(0.1)	8.0	23.0	52.5	149.0	919.0	5.0	11.0	17.5	27.0	55.5
QA-SIS(0.3)	3.0	117.5	562.0	1466.0	3196.5	3.0	3.0	4.0	6.5	109.0
QA-SIS(0.5)	69.5	611.5	1754.5	3333.5	4554.0	3.5	8.0	59.0	552.5	1786.5
QA-SIS(0.7)	184.5	1131.0	2307.0	3419.0	4573.0	17.0	287.5	886.5	1798.0	3692.5
QA-SIS(0.9)	867.0	1881.5	3007.0	3863.5	4763.5	841.5	1862.5	2873.0	3834.0	4603.0

Notes: QA-SVS-A(4), QA-SVS-A(5), . . . , and QA-SVS-A(10), our proposed method defined in Remark 3 with different quantile grid points ($K = 4, \dots, 10$); QCS-A(4), QCS-A(5), . . . , and QCS-A(10), the quantile correlation-based screening method (Tang et al., 2013) [19] with different quantile grid points ($K = 4, \dots, 10$); SIS, the sure independence screening (Fan and Lv, 2008) [1]; DC-SIS, the distance correlation-based screening (Li et al., 2012) [8]; QA-SIS(0.1), QA-SIS(0.3), . . . , QA-SIS(0.9), the quantile-adaptive model-free sure independence screening (He et al., 2013) [9] at different quantile levels.

Table A5. The result of criteria in all scenarios under $p = 5000$ with $\alpha = 0.05$ of Section 4.2.

Method	QA-SVS-AFD(K)					QA-SVS-FDR(K)					QCS-FDR(K)				
	2	3	4	5	6	2	3	4	5	6	2	3	4	5	6
	Scenario 2.1														
$ \hat{\mathcal{A}} $	12.08	10.69	10.52	10.23	10.04	11.68	11.69	11.92	11.88	11.53	11.17	11.22	11.25	11.29	11.10
FDR	0.16	0.06	0.05	0.02	0.00	0.14	0.14	0.16	0.15	0.13	0.10	0.10	0.10	0.11	0.09
F ₁ -score	0.91	0.97	0.98	0.99	1.00	0.92	0.92	0.91	0.92	0.93	0.95	0.94	0.94	0.94	0.95
	Scenario 2.2														
$ \hat{\mathcal{A}} $	50.33	48.23	45.02	38.21	27.37	52.35	51.98	52.22	52.19	52.05	50.22	50.67	49.77	49.41	48.43
FDR	0.02	0.00	0.00	0.00	0.00	0.05	0.04	0.05	0.05	0.05	0.05	0.05	0.04	0.05	0.04
F ₁ -score	0.98	0.98	0.95	0.86	0.70	0.97	0.98	0.97	0.97	0.97	0.96	0.96	0.95	0.95	0.94
	Scenario 2.3														
$ \hat{\mathcal{A}} $	12.09	10.67	10.40	10.10	10.02	11.74	11.58	11.52	11.60	11.57	11.08	10.90	11.16	10.90	10.90
FDR	0.17	0.06	0.04	0.01	0.00	0.14	0.13	0.13	0.13	0.13	0.09	0.08	0.10	0.08	0.08
F ₁ -score	0.91	0.97	0.98	1.00	1.00	0.92	0.93	0.93	0.93	0.93	0.95	0.96	0.95	0.96	0.96
	Scenario 2.4														
$ \hat{\mathcal{A}} $	50.05	45.91	39.33	25.62	15.27	52.23	51.96	51.91	52.07	51.56	49.53	48.03	47.22	44.47	43.53
FDR	0.02	0.00	0.00	0.00	0.00	0.05	0.05	0.05	0.05	0.04	0.05	0.05	0.05	0.04	0.05
F ₁ -score	0.98	0.96	0.88	0.67	0.46	0.97	0.97	0.97	0.97	0.97	0.95	0.93	0.92	0.90	0.89

Table A5. Cont.

Method	QA-SVS-AFD(K)					QA-SVS-FDR(K)					QCS-FDR(K)					
	K	2	3	4	5	6	2	3	4	5	6	2	3	4	5	6
Scenario 2.5																
$ \hat{\mathcal{A}} $	10.89	9.88	9.30	7.95	5.90	10.47	10.43	10.54	10.47	10.40	10.00	10.09	9.88	9.62	9.48	
FDR	0.08	0.00	0.00	0.00	0.00	0.04	0.04	0.05	0.04	0.04	0.05	0.05	0.05	0.04	0.05	
F ₁ -score	0.96	0.99	0.96	0.88	0.73	0.98	0.98	0.97	0.98	0.98	0.94	0.95	0.94	0.93	0.92	
Scenario 2.6																
$ \hat{\mathcal{A}} $	14.81	3.65	0.49	0.08	0.00	12.65	13.93	12.88	10.88	9.95	1.83	1.30	0.94	0.72	0.45	
FDR	0.05	NaN	NaN	NaN	NaN	0.04	0.04	0.05	0.04	0.04	NaN	NaN	NaN	NaN	NaN	
F ₁ -score	0.43	0.13	0.02	0.00	0.00	0.38	0.41	0.38	0.33	0.31	0.06	0.05	0.03	0.02	0.02	
Scenario 2.7																
$ \hat{\mathcal{A}} $	10.92	10.02	9.99	9.99	9.90	10.39	10.51	10.46	10.52	10.60	10.66	10.72	10.63	10.53	10.50	
FDR	0.08	0.00	0.00	0.00	0.00	0.03	0.04	0.04	0.05	0.05	0.06	0.06	0.05	0.05	0.04	
F ₁ -score	0.96	1.00	1.00	1.00	0.99	0.98	0.98	0.98	0.98	0.97	0.97	0.97	0.97	0.98	0.98	
Scenario 2.8																
$ \hat{\mathcal{A}} $	38.60	18.94	6.36	1.24	0.19	43.52	41.46	40.05	38.34	36.21	22.75	17.21	11.87	8.08	6.15	
FDR	0.02	0.00	0.00	NaN	NaN	0.05	0.04	0.05	0.05	0.04	0.04	0.05	NaN	NaN	NaN	
F ₁ -score	0.85	0.55	0.22	0.05	0.01	0.89	0.86	0.84	0.83	0.80	0.59	0.48	0.35	0.26	0.20	

Notes: QA-SVS-AFD(K), our proposed method by controlling FD adaptively defined in Remark 4 with different quantile grid points ($K = 2, \dots, 6$); QA-SVS-FDR(K), our proposed method by controlling FDR defined in Remark 5 with different quantile grid points ($K = 2, \dots, 6$); QCS-FDR(K), the quantile correlation-based screening method (Tang et al., 2013) [19] with different quantile grid points ($K = 2, \dots, 6$). $|\hat{\mathcal{A}}|$: the average number of selected predictors; FDR: the average of empirical false discovery proportion, where 'NaN' indicates the method loss validity; F₁-score: the average of F1-score.

Appendix C

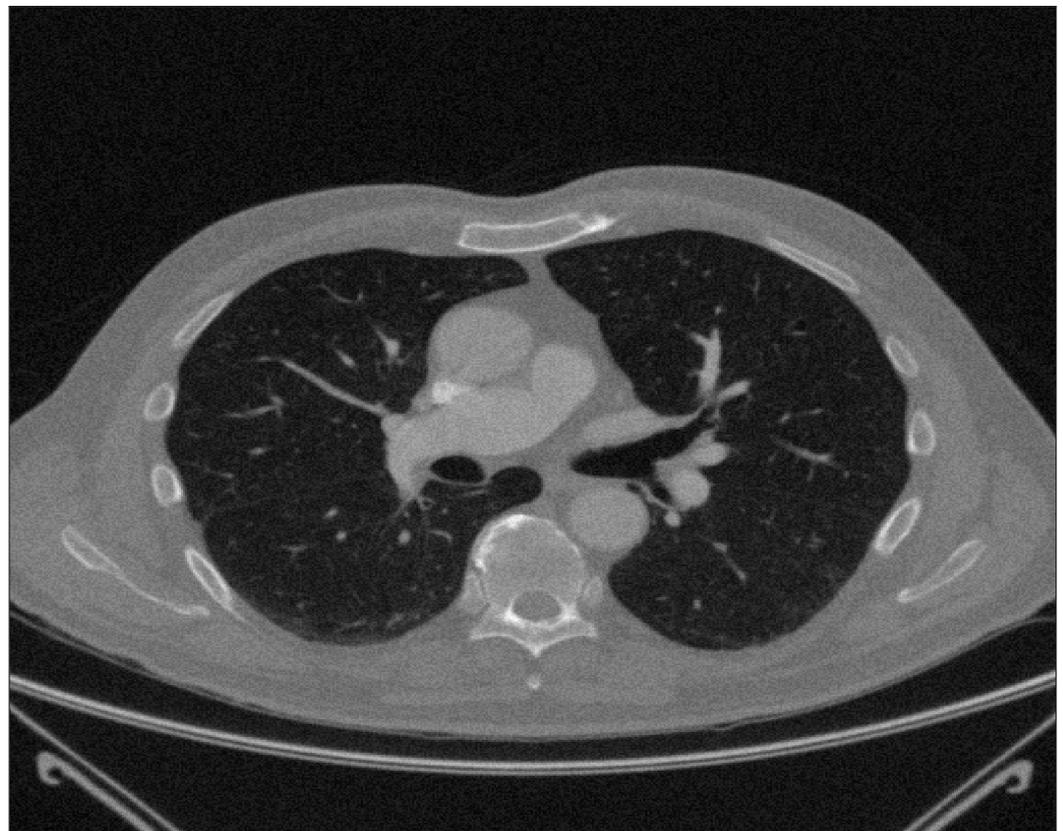


Figure A1. A lung CT figure of a subject in the dataset.

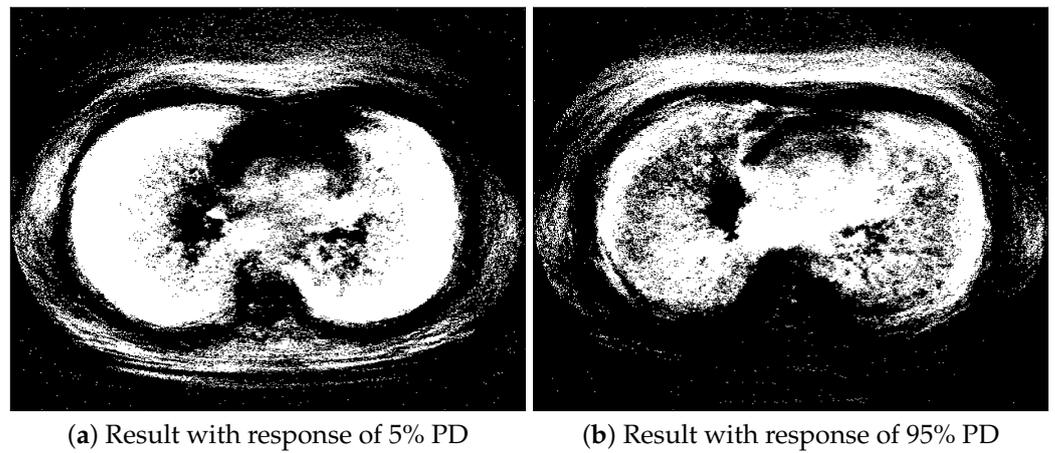


Figure A2. Screened important pixels of lung CT by QA-SVS-FDR under $K = 2$. The black pixels indicate the useless pixels, and the white pixels represent the important pixels.

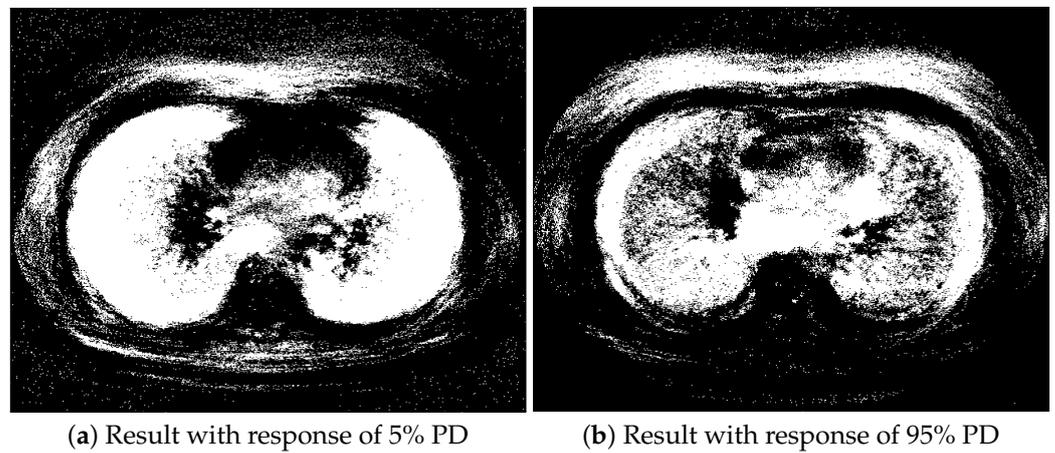


Figure A3. Screened important pixels of lung CT by QA-SVS-FDR under $K = 3$.

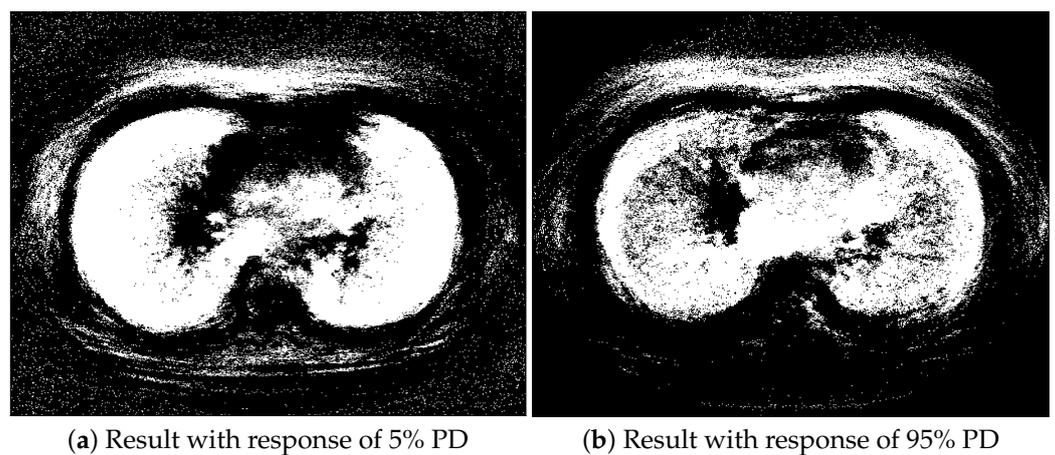


Figure A4. Screened important pixels of lung CT by QA-SVS-FDR under $K = 4$.

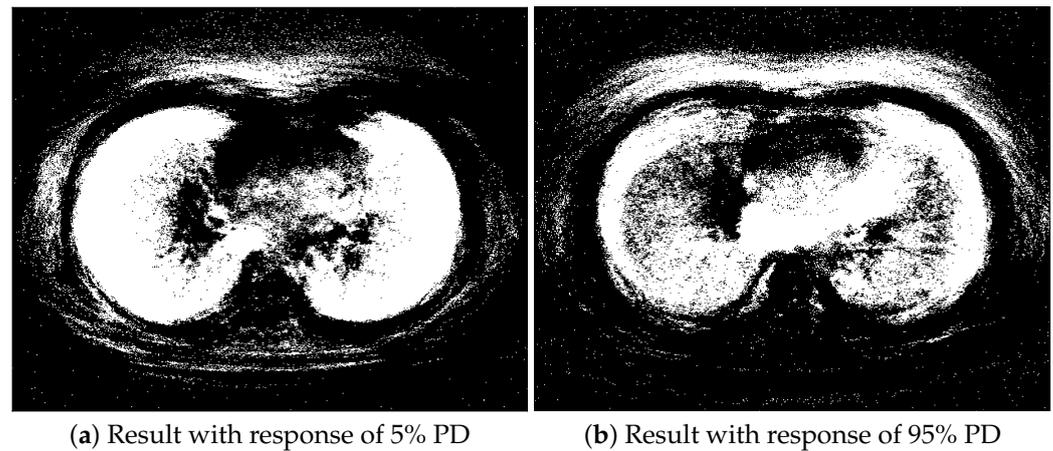


Figure A5. Screened important pixels of lung CT by QA-SVS-FDR under $K = 5$.

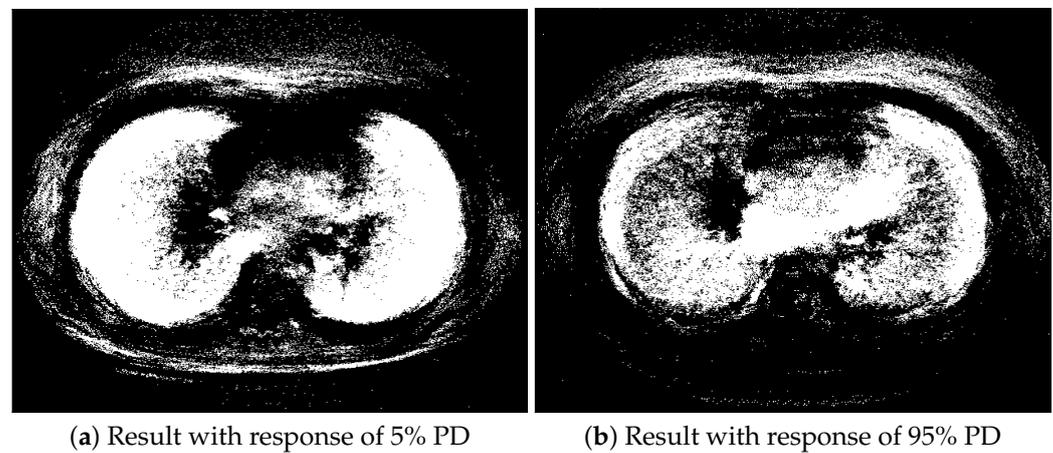


Figure A6. Screened important pixels of lung CT by QA-SVS-FDR under $K = 6$.

References

1. Fan, J.; Lv, J. Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* **2008**, *70*, 849–883.
2. Liu, W.; Li, R., Variable Selection and Feature Screening. In *Macroeconomic Forecasting in the Era of Big Data: Theory and Practice*; Fuleky, P., Ed.; Springer: Cham, Switzerland, 2020; pp. 293–326.
3. Fan, J.; Song, R. Sure Independence Screening in Generalized Linear Models with N_p -Dimensionality. *Ann. Stat.* **2010**, *38*, 3567–3604. [[CrossRef](#)]
4. Fan, J.; Feng, Y.; Song, R. Nonparametric Independence Screening in Sparse Ultra-High-Dimensional Additive Models. *J. Am. Stat. Assoc.* **2011**, *106*, 544–557. [[CrossRef](#)] [[PubMed](#)]
5. Li, G.; Peng, H.; Zhang, J.; Zhu, L. Robust Rank Correlation Based Screening. *Ann. Stat.* **2012**, *40*, 1846–1877. [[CrossRef](#)]
6. Chang, J.; Tang, C.Y.; Wu, Y. Marginal Empirical Likelihood In addition, Sure Independence Feature Screening. *Ann. Stat.* **2013**, *41*, 2123–2148. [[CrossRef](#)] [[PubMed](#)]
7. Zhu, L.P.; Li, L.; Li, R.; Zhu, L.X. Model-Free Feature Screening for Ultrahigh-Dimensional Data. *J. Am. Stat. Assoc.* **2011**, *106*, 1464–1475. [[CrossRef](#)] [[PubMed](#)]
8. Li, R.; Zhong, W.; Zhu, L. Feature Screening via Distance Correlation Learning. *J. Am. Stat. Assoc.* **2012**, *107*, 1129–1139. [[CrossRef](#)] [[PubMed](#)]
9. He, X.; Wang, L.; Hong, H.G. Quantile-Adaptive Model-Free Variable Screening for High-Dimensional Heterogeneous Data. *Ann. Stat.* **2013**, *41*, 342–369. [[CrossRef](#)]
10. Lin, L.; Sun, J.; Zhu, L. Nonparametric feature screening. *Comput. Stat. Data Anal.* **2013**, *67*, 162–174. [[CrossRef](#)]
11. Lu, J.; Lin, L. Model-free conditional screening via conditional distance correlation. *Stat. Pap.* **2020**, *61*, 225–244. [[CrossRef](#)]
12. Mai, Q.; Zou, H. The Kolmogorov filter for variable screening in high-dimensional binary classification. *BIOMETRIKA* **2013**, *100*, 229–234. [[CrossRef](#)]
13. Huang, D.; Li, R.; Wang, H. Feature Screening for Ultrahigh Dimensional Categorical Data with Applications. *J. Bus. Econ. Stat.* **2014**, *32*, 237–244. [[CrossRef](#)]

14. Cui, H.; Li, R.; Zhong, W. Model-Free Feature Screening for Ultrahigh Dimensional Discriminant Analysis. *J. Am. Stat. Assoc.* **2015**, *110*, 630–641. [[CrossRef](#)]
15. Han, X. Nonparametric screening under conditional strictly convex loss for ultrahigh dimensional sparse data. *Ann. Stat.* **2019**, *47*, 1995–2022.
16. Zhou, T.; Zhu, L.; Xu, C.; Li, R. Model-free forward screening via cumulative divergence. *J. Am. Stat. Assoc.* **2020**, *115*, 1393–1405. [[CrossRef](#)]
17. Xie, J.; Lin, Y.; Yan, X.; Tang, N. Category-Adaptive Variable Screening for Ultra-High Dimensional Heterogeneous Categorical Data. *J. Am. Stat. Assoc.* **2020**, *115*, 747–760. [[CrossRef](#)]
18. Hao, N.; Zhang, H.H. A note on high-dimensional linear regression with interactions. *Am. Stat.* **2017**, *71*, 291–297. [[CrossRef](#)]
19. Tang, W.; Xie, J.; Lin, Y.; Tang, N. Quantile Correlation Based Variable Selection. *J. Bus. Econ. Stat.* **2021**, *40*, 1801–1903. [[CrossRef](#)]
20. Liu, W.; Ke, Y.; Liu, J.; Li, R. Model-free feature screening and fdr control with knockoff features. *J. Am. Stat. Assoc.* **2022**, *117*, 428–443. [[CrossRef](#)]
21. Guo, X.; Ren, H.; Zou, C.; Li, R. Threshold selection in feature screening for error rate control. *J. Am. Stat. Assoc.* **2022**, 1–13. [[CrossRef](#)]
22. Cook, R.D. Testing predictor contributions in sufficient dimension reduction. *Ann. Stat.* **2004**, *32*, 1062–1092. [[CrossRef](#)]
23. Yin, X.; Hilafu, H. Sequential Sufficient Dimension Reduction for Large p , Small n Problems. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2015**, *77*, 879–892. [[CrossRef](#)]
24. Yuan, Q.; Chen, X.; Ke, C.; Yin, X. Independence index sufficient variable screening for categorical responses. *Comput. Stat. Data Anal.* **2022**, *174*, 107530. [[CrossRef](#)]
25. Hyndman, R.J.; Fan, Y. Sample Quantiles in Statistical Packages. *Am. Stat.* **1996**, *50*, 361–365.
26. Mohamed, I.B.; Mirakhmedov, S.M. Approximation by Normal Distribution for A Sample Sum in Sampling Without Replacement from a Finite Population. *Sankhya A* **2016**, *78*, 188–220. [[CrossRef](#)]
27. Benjamini, Y.; Hochberg, Y. Controlling The False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **1995**, *57*, 289–300. [[CrossRef](#)]
28. Shalmon, T.; Salazar, P.; Horie, M.; Hanneman, K.; Pakkal, M.; Anwari, V.; Fratesi, J. Predefined and data driven CT densitometric features predict critical illness and hospital length of stay in COVID-19 patients. *Sci. Rep.* **2022**, *12*, 8143. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.