Asia Mathematika
Volume: 6 Issue: 3 , (2022) Pages: 55 – 75
Available online at **www.asiamath.org**

Model average feature screening for ultrahigh dimensional data with responses missing at random

Mingyue WANG[1], **Yufei GAO**[2], **Peng LAI**[3*]
School of Mathematics and Statistics, Nanjing University of Information Science & Technology, Nanjing, China. [1]ORCID: 0000-0003-1407-8848

Received: 01 Dec 2022 • **Accepted:** 07 Dec 2022 • **Published Online:** 30 Dec 2022

Abstract: This article focuses on the feature screening problem of ultrahigh dimensional data with responses missing at random. A model average feature screening method is proposed based on the inverse probability weighted method. The proposed screening procedure is robust to the heavy-tailed distribution data and potential outliers. It can deal with missing data when the response variable is missing at random. It is model-free and the sure screening properties are satisfied. We illustrate its screening performance through Monte Carlo simulations.

Key words: Ultrahigh dimensional data, feature screening, missing at random, sure screening property

1. Introduction

With the rapid development of science and technology and data acquisition capability, ultrahigh dimensional data has become more frequent in many fields, such as biomedicine, economics, social sciences, and so on. The dimensionality p of the collected data increases exponentially with the sample size n, that is $\log p = O(n^{\xi})$ for some $\xi > 0$. Ultrahigh dimensional features lead to the invalidation of traditional statistical methods, the inability to extract information effectively, and low computational efficiency. To handle this kind of ultrahigh dimensional problem, some fast and simple dimensionality reduction methods without information loss need to be studied.

In the case of complete data, many dimensionality reduction methods have been formed. [1] proposed the feature screening method SIS for the ultrahigh dimensional linear model based on the marginal Pearson correlation coefficients between covariates and response variable to construct screening indicators. Under some regularity conditions, the ranking consistency and screening consistency of predictors are satisfied. Thus, selecting features based on the sorted screening indicators can obtain the possible active predictors. On the basis of their research, many statisticians have carried out corresponding research works. [2] proposed a feature screening procedure under the assumption of a generalized linear model by ranking the maximum marginal likelihood estimates or the maximum marginal likelihood itself. [3] significantly improved SIS by using distance correlation coefficients to reduce the dimensionality of ultrahigh dimensional covariates without specifying any model assumptions. [4] proposed a conditional quantile-based feature screening method for the ultrahigh dimensional data without model assumptions, which to some extent, optimized the method proposed by [5]. [6] proposed a feature screening method based on the Blum-Kiefer-Rosenblatt correlation coefficient, which is effective in dealing with nonlinear effects. [7] proposed a model-free procedure called covariate information

*Correspondence: laipengnuist@163.com

number sure independence screening (CIS). CIS used a marginal utility connected to the notion of the traditional Fisher information and possessed the sure screening property. Many feature screening procedures can also be found in [8–10], and so on.

However, in many cases, due to some unavoidable factors, part of the information of some samples would not be observed, which is the so-called missing data. In many practical problems studied [11–13], there are often missing data. When the sample is missing, it cannot entirely reflect the actual characteristics of the interested problem. Failure to properly handle missing values may lead to bias in the final analysis results. In the missing mechanisms ([14]), missing at random (MAR) is a common phenomenon in practice, and many methods have been proposed to solve the missing data problems. The ultrahigh dimensional features coupled with missing data lead to many challenging problems. There are few studies on ultrahigh dimensional feature screening with responses missing at random. [15] proposed an inverse probability weighted Pearson correlation coefficient under the assumption of a linear model. Combined with the inverse probability weights and the SIRS feature screening method proposed by [8], [16] proposed a model-free ultrahigh dimensional feature screening method with responses missing at random. Furthermore, [17] studied the feature screening in ultrahigh dimensional partially linear models with responses missing at random and [18] studied the feature screening for ultrahigh dimensional categorical data with covariates missing at random.

In addition to the problem of missing data, observations are often accompanied by outliers, heavy tails, etc. Some robust methods which can handle the problems of outliers, heavy tails and responses missing at random are needed. Since quantiles are not sensitive to outliers and heavy-tailed distribution data, this paper combines the inverse probability weighted method, the idea of model average, and conditional quantiles to propose a feature screening procedure on the ultrahigh dimensional continuous data with responses missing at random. Considering that the missing mechanism of the response variable is missing at random, the feature screening method for a binary ultrahigh dimensional data based on the model averaging technique proposed by [19] is used to screen out the predictors related to the missing indicator at first. Then the inverse probability weights would be obtained on this basis.

The second part of this paper will introduce an inverse probability weighted conditional quantile feature screening method (MMACQ) based on the model average technique under the assumption that the response variable is missing at random and study its theoretical properties under some regularity conditions. The third part uses Monte Carlo numerical simulations to verify the finite sample properties of MMACQ and compare it with other screening methods. The relevant proofs of the theoretical properties will be presented in the supplementary material.

2. Model Average Feature Screening

Let Y be the response variable with support set $\boldsymbol{\Psi}_y$. $\boldsymbol{X} = (X_1, \ldots, X_p)^T$ is the covariate vector. The dimension p is much larger than the sample size n, which increases exponentially with the increase of the sample size n and satisfies the sparsity assumption. That is, only a small number of covariates affect the response variable.

2.1. Screening method

In order to identify those important covariates that have significant effects on the response variable, the active and inactive predictor sets are defined as $\mathcal{A} = \left\{k : F\left(y \middle| \boldsymbol{X}\right) \text{ depends on } X_k, k = 1, \ldots, p \text{ for some } y \in \boldsymbol{\Psi}_y\right\}$ and $\mathcal{L} = \left\{k : F(y|\boldsymbol{X}) \text{ does not depend on } X_k, k = 1, \ldots, p \text{ for any } y \in \boldsymbol{\Psi}_y\right\}$. If $k \in \mathcal{A}$, then X_k is an important predictor. If $k \in \mathcal{L}$, it means that X_k has no effect on the response variable and is an unimportant predictor variable.

Considering that quantile regression has obvious robustness when dealing with heavy-tailed data and outliers, [4] focused on the conditional quantile, $Q_\tau\left(Y \middle| X\right) = \inf\left(t : \operatorname{pr}\left(Y \le t \middle| X\right) \ge \tau\right)$, $\tau \in (0,1)$. Define the set of important predictors and the set of unimportant predictors under a given τ quantile as $\mathcal{A}_\tau = \left\{k : Q_\tau\left(Y \middle| \boldsymbol{X}\right) \text{ depends on } X_k, k = 1, \ldots, p\right\}$ and $\mathcal{L}_\tau = \left\{k : Q_\tau\left(Y \middle| \boldsymbol{X}\right) \text{ does not depend on } X_k, k = 1, \ldots, p\right\}$. Obviously, $\mathcal{A}_\tau$ is a subset of $\mathcal{A}$.

In the case where the response variable is missing at random, it is assumed that the observed sample data is $(Y_i, \delta_i, \boldsymbol{X}_i)$, $i = 1, \ldots, n$. If Y_i is missing, then $\delta_i = 0$, otherwise $\delta_i = 1$, and $P\left(\delta = 1 \middle| \boldsymbol{X}, Y\right) = P\left(\delta = 1 \middle| \boldsymbol{X}\right)$. Based on the MAR assumption and the sparsity property, the value of the missing indicator δ is only affected by some predictors. Thus we can define the active predictors set and the inactive predictors set for the missing indicator δ as $\mathcal{A}_\delta = \left\{k : P\left(\delta = 1 \middle| \boldsymbol{X}\right) \text{ depends on } X_k\right\}$ and $\mathcal{L}_\delta = \left\{k : P\left(\delta = 1 \middle| \boldsymbol{X}\right) \text{ does not depend on } X_k\right\}$.

When the ultrahigh dimensional data is completely observed, [4], inspired by the definition of conditional expectation, defined the feature screening index as

$$d_k(t) = E\left(\left[\tau - I\left\{Y < Q_\tau(Y)\right\}\right] I\left(X_k < t\right)\right).$$

However, this method is no longer applicable when the response variable is missing at random. An inverse probability weighted feature screening measurement $d_k^*(t)$ is considered here,

$$d_k^*(t) = E\left\{\frac{\delta}{P\left(\delta = 1 \middle| X_{\mathcal{A}_\delta}\right)}\left[\left[\tau - I\left\{Y < Q_\tau(Y)\right\}\right] I\left(X_k < t\right)\right]\right\} = d_k(t).$$

Given X_k, if the τ quantile conditional quantile of Y is independent of X_k, that is, $Q_\tau(Y|X_k) = Q_\tau(Y)$, then $d_k^*(t) = 0$. Given the τ quantile, the estimate of $d_k^*(t)$ is

$$\hat{d}_k^*(t) = n^{-1}\sum_{i=1}^{n}\frac{\delta_i}{\hat{P}\left(\delta_i = 1 \middle| X_{\hat{\mathcal{A}}_\delta}\right)}\left[\tau - I\left\{Y_i < \hat{Q}_\tau(Y)\right\}\right] I\left\{X_{ik} < t\right\}, k = 1, \ldots, p,$$

where $\hat{Q}_\tau(Y)$ is the quantile estimate of $Y_1, Y_2, \ldots, Y_n$ at a given τ quantile. Thus, the marginal utility of the kth predictor is

$$\left\|\hat{d}_k^*\right\| = n^{-1}\sum_{i=1}^{n}\hat{d}_k^*\left(X_{ik}\right)^2, k = 1, \ldots, p.$$

When the response variable Y is independent of the predictor variable X_k, $\left\|\hat{d}_k^*\right\|$ will fluctuate around zero and approach zero, it can be used to measure the importance of covariates. The estimated active predictor set would be

$$\hat{\mathcal{A}}_\tau = \left\{k : \left\|\hat{d}_k^*\right\| \ge cn^{-\alpha}, k = 1, \ldots, p\right\},$$

where c and α are predetermined thresholds.

In order to calculate $\left\|\hat{d}_k^*\right\|$, we have to first estimate $P\left(\delta_i = 1 \middle| X_{\mathcal{A}_\delta}\right)$ and $Q_\tau(Y)$. Note that the value of δ is 0 or 1. Therefore, we can select covariates using some binary classification feature screening method.

[19] proposed a conditional quantile feature screening method called MACQFS, based on the model average technique for the binary response variable, which is used here to get $\hat{\mathcal{A}}_\delta = \left\{k : \hat{\omega}_k^* \geq vn^{-\alpha}, k = 1, \ldots, p\right\}$, where $\hat{\omega}_k^*$ is the estimated index to measure the connection between missing indicator δ and covariate X_k, details can be seen in [19], v and n are the pre-set threshold values. Since v and n are difficult to be determined in advance in the actual operation process, [1] proposed that the estimates of the screening index $\hat{\omega}_k^*$ can be sorted in descending order, and the top $\left[n/\log(n)\right]$ are selected to constitute the estimated set of important predictors, where $[\cdot]$ is the rounding function, $d_\delta = \left[n/\log(n)\right]$. The estimated set of important predictors that affect the value of δ is $\hat{\mathcal{A}}_\delta = \left\{k : \text{among the largest } d_\delta \text{ of } \hat{\omega}_k^*\right\}$. According to the Theorem 2 of [19], under some regularity conditions, $P(A_\delta \subseteq \hat{\mathcal{A}}_\delta) \to 1$. It means using the MACQFS feature screening method can select the important predictors which are related to δ.

Based on the estimated set of important predictors for missing indicator δ, we can estimate $P\left(\delta = 1 \middle| X\right) = P\left(\delta = 1 \middle| X_{\hat{\mathcal{A}}_\delta}\right)$ by some nonparametric estimation methods. However, although using the MACQFS method to screen the original covariates can compress the original covariate dimension p to the dimension $d_\delta = \left[n/\log(n)\right] < n$, there may still be high dimension for the possible nonparametric estimation. Assuming that $P\left(\delta = 1 \middle| X_{\hat{\mathcal{A}}_\delta}\right) = \exp\left(X_{\hat{\mathcal{A}}_\delta}^T \beta\right) / \left(1 + \exp\left(X_{\hat{\mathcal{A}}_\delta}^T \beta\right)\right)$ is a logistic model, SCAD ([20]) can be used to perform another variable selection on the covariates and get the estimates set $\hat{\mathcal{A}}_\delta$. After re-selection of variables, new important predictor variable estimation sets $\hat{\mathcal{A}}_\delta^*$ for missing indicator δ and $\hat{\beta}_{SCAD}$ are obtained, and then $\hat{P}\left(\delta = 1 \middle| X\right) = \hat{P}\left(\delta = 1 \middle| X_{\hat{\mathcal{A}}_\delta}\right) = \hat{P}\left(\delta = 1 \middle| X_{\hat{\mathcal{A}}_\delta^*}\right)$ can be obtained. Another difficulty is that $\hat{Q}_\tau(Y)$ is hard to be obtained because of the missing data of Y. The method proposed by [21] will be used to estimate the quantile of the response variable. The core of this method is non-parametric verification.

Since the number of values for $\tau \in (0,1)$ is infinite, the impact of the predictor variable on the response variable Y at different quantiles may differ. If we only consider the method under a certain τ quantile alone, we will lose lots of information, so we use the idea of model average. Take the quantile sequence $0 < \tau_1 < \cdots < \tau_m < 1$. We regard the local important predictor sets under different quantiles as multiple candidate models and form an important global predictor set by weighted combination.

Define the set of locally significant predictors and the set of unimportant predictors as

$$\mathcal{A}_{\tau_s} = \left\{k : Q_{\tau_s}\left(\boldsymbol{y} \middle| \boldsymbol{X}\right) \text{ depends on } X_k, k = 1, \ldots, p \text{ for some } \boldsymbol{y} \in \boldsymbol{\Psi}_y\right\},$$

$$\mathcal{L}_{\tau_s} = \left\{k : Q_{\tau_s}\left(\boldsymbol{y} \middle| \boldsymbol{X}\right) \text{ does not depend on } X_k, k = 1, \ldots, p \text{ for any } \boldsymbol{y} \in \boldsymbol{\Psi}_y\right\}.$$

We can estimate $\mathcal{A}_{\tau_s}$ as

$$\hat{\mathcal{A}}_{\tau_s} = \left\{k : \left\|\hat{d}_{k,\tau_s}^*\right\| \geq cn^{-\alpha}, k = 1, \ldots, p\right\},$$

where $\left\|\hat{d}_{k,\tau_s}^*\right\| = \dfrac{1}{n}\sum\limits_{i=1}^{n} \hat{d}_{k,\tau_s}^{*2}\left(X_{ik}\right)$ and

$$\hat{d}_{k,\tau_s}^*(t) = n^{-1}\sum_{i=1}^{n} \frac{\delta_i}{\hat{P}\left(\delta_i = 1 \middle| X_{\hat{\mathcal{A}}_\delta^*}\right)}\left[\tau - I\left\{Y_i < \hat{Q}_{\tau_s}(Y)\right\}\right] I\left(X_{ik} < t\right).$$

To comprehensively consider the influence of predictors on the response variable at different τ quantiles, combined with the idea of model average, we weighted the local important predictor variable screening indicators to form the global important predictor variable screening indicators

$$R_k = \sum_{s=1}^{m} \omega_{\tau_s}^{*} I\Big(X_k \in \mathcal{A}_{\tau_s}\Big).$$

Regarding the model average weights $\omega_{\tau_s}^{*}$, $s = 1, \ldots, m$, such as [22], considering the local important predictor set $\mathcal{A}_{\tau_s}$ in the m local important variable sets, we get

$$\omega_{\tau_s}^{*} = \frac{\tilde{\omega}_{\tau_s}^{*}}{\sum_{s=1}^{m} \tilde{\omega}_{\tau_s}^{*}},\ \tilde{\omega}_{\tau_s}^{*} = \frac{1}{|\mathcal{A}_{\tau_s}|} \Sigma_{k \in \mathcal{A}_{\tau_s}} \left\| d_{k,\tau_s}^{*} \right\|,$$

where $\|d_{k,\tau_s}\| = E\{d_{k,\tau_s}(X_k)^2\}$. The estimate of the globally significant predictor screening metrics is

$$\hat{R}_k = \sum_{s=1}^{m} \hat{\omega}_{\tau_s}^{*} I\left(X_k \in \hat{\mathcal{A}}_{\tau_s}\right),$$

where

$$\hat{\omega}_{\tau_s}^{*} = \frac{\hat{\tilde{\omega}}_{\tau_s}^{*}}{\sum_{s=1}^{m} \hat{\tilde{\omega}}_{\tau_s}^{*}},\ \hat{\tilde{\omega}}_{\tau_s}^{*} = \frac{1}{|\hat{\mathcal{A}}_{\tau_s}|} \Sigma_{k \in \hat{\mathcal{A}}_{\tau_s}} \left\| \hat{d}_{k,\tau_s}^{*} \right\|.$$

Regarding the globally important predictor variable screening indicator R_k, there may be two situations when the estimated value is large: (1) Its corresponding predictor X_k exists in multiple sets of locally important predictors; (2) Even if X_k does not exist in multiple sets of locally important predictors, the set of important local predictors which is located in is more important according to the model average weight. The above two situations show that when the estimated value of R_k is large, the corresponding predictor variable X_k has a more significant impact on the response variable so that the predictor variables can be screened by $\hat{R}_k$. The set of globally significant predictor estimates is

$$\hat{\mathrm{A}} = \left\{k : \hat{R}_k \geq cn^{-\alpha}, k = 1, \ldots, p\right\},$$

where c and α are pre-set threshold values. Also, we can select the largest $d = [n/\log(n)]$ variables as

$$\hat{A} = \left\{k : \text{among the largest } d \text{ of } \hat{R}_k\right\}.$$

The above screening method based on the model average and the inverse probability weights is called MMACQ.

2.2. Sure screening properties

In order to study the theoretical properties of the MMACQ method, the following conditions are required.

- C1. There exists constants $c > 0$, $M > 0$, and $\alpha \in [0, 1/4)$, such that

$$\infty > M \geq \max_{k \in \mathcal{A}_\tau} \|d_k^{*}\| \geq \min_{k \in \mathcal{A}_\tau} \|d_k^{*}\| \geq 2cn^{-\alpha} \text{ and } \min_{k \in \mathrm{A}} R_k \geq 2cn^{-\alpha}.$$

- C2. In the neighborhood of $Q_\tau(Y)$, $F(y)$ is second-order differentiable. The density function $f(y)$ of Y is uniformly bounded away from 0 and infinity, and $f'(y)$ is also uniformly bounded.
- C3. There exists a constant $\ell_1 > 0$, which is the lower limit of P_i, that is $0 < \ell_1 \le P_i \le 1$.

Remark: The condition (C1) ensures that the marginal utility of carrying predictor information does not decay too quickly. Condition (C2) is a necessary condition for quantile regression. In condition (C3), ℓ_1 is a small positive number, and the purpose is to ensure that the denominator of δ_i/P_i is not 0, where $P_i = P(\delta_i = 1|X_{\mathcal{A}_\delta}, \theta)$, θ is the parameter.

Theorem 2.1. *(Sure screening property) Under conditions (C1)-(C3), there are positive constants* c_1 *and* c_2, *such that*

$$P\left(\max_{1\le k\le p}\left|\left\|\hat{d}_k^*\right\| - \|d_k^*\|\right| \ge cn^{-\alpha}\right) \le O\left\{p\exp\left(-c_1 n^{3-2\alpha}\right) + p\exp\left(-c_2 n^{1-2\alpha}\right)\right\}.$$

Because of $\min_{k\in\mathcal{A}_\tau}\|d_k^*\| \ge 2cn^{-\alpha}$, *we can further get*

$$P\left(\mathcal{A}_\tau \subseteq \hat{\mathcal{A}}_\tau\right) \ge 1 - O\left(a_n\exp(-c_1 n^{3-2\alpha}) + a_n\exp\left(-c_2 n^{1-2\alpha}\right)\right),$$

where $a_n = |\mathcal{A}_\tau|$ *represents the number of elements in the important variable set* $\mathcal{A}_\tau$.

Theorem 2.2. *(Minimum Model Size) Under the conditions (C1)-(C3), there are positive constants* c_3, c_4, c_5, c_6, *such that*

$$\begin{aligned}P\left(\max_{1\le k\le p}\left|\hat{R}_k - R_k\right| \ge cn^{-\alpha}\right) \le & O\left\{pnm\exp\left(-c_3 n^{3-2\alpha}\right) + pmn\exp\left(-c_4 n^{1-2\alpha}\right)\right\} \\ & + O\left(a_{n,m}mp\exp\left(-c_5 n^{3-2\alpha}\right) + a_{n,m}mp\exp\left(-c_6 n^{1-2\alpha}\right)\right).\end{aligned} \tag{1}$$

Because of $\min_{k\in \mathrm{A}} R_k \ge 2cn^{-\alpha}$, *we can further get*

$$\begin{aligned}P\left(\mathrm{A} \subseteq \hat{\mathrm{A}}\right) \ge & 1 - O\left\{A_n nm\exp\left(-c_3 n^{3-2\alpha}\right) + A_n mn\exp\left(-c_4 n^{1-2\alpha}\right)\right\} \\ & - O\left(A_n a_{n,m} m\exp\left(-c_5 n^{3-2\alpha}\right) + A_n a_{n,m} m\exp\left(-c_6 n^{1-2\alpha}\right)\right),\end{aligned} \tag{2}$$

where $A_n = \left|\mathrm{A}\right|$ *represents the number of elements in the important variable set* A, $a_{n,m}$ *is the maximum value of the set* $\{a_{n,\tau_s}, 1 \le s \le m\}$, $a_{n,\tau_s} = \left|\mathrm{A}_{\tau_s}\right|$ *represents the number of elements in the important variable set* A_{τ_s}.

Combined with the definition of ultrahigh dimensional data, it is considered that the dimension p of the predictor variable increases exponentially with the increase of the sample size n, that is, $p = o\left(\exp\left(n^\gamma\right)\right)$. Assuming the number of quantiles $m = O\left(n^\beta\right)$, where $\beta > 0$, when $0 < \gamma < \log\left(n^{1-2\alpha} - \log(n^{\beta+1})\right)$ and $0 < \beta < \left(n^{1-2\alpha} - \log(n) - 1\right)/\log(n)$, the right side of the formula (1) tends to 0, which is $P(\mathrm{A} \subseteq \hat{\mathrm{A}}) \to 1$. It shows that when the conditions (C1)-(C3) are satisfied, the important variable estimation set obtained by the MMACQ method contains the real important variable set with a probability tends to 1. The proofs of these theorems are detailed in Supplementary Material.

3. Numerical examples

3.1. Monte Carlo Simulations

In this section, Monte Carlo numerical simulation will be used to test the screening performance of the average feature screening method MMACQ for the ultrahigh dimensional model with responses missing at random proposed in Section 2 of this paper.

To compare the difference between the screening method that comprehensively considers multiple quantiles, that is, the MMACQ feature screening method, and the screening method that only considers a given quantile, this section uses the important local variable defined in Section 2 to screen the index $\left\|\hat{d}^*_{k,\tau_s}\right\|$, in the case of given quantiles $\tau = 0.5$ and $\tau = 0.35$, the set of predictors is screened respectively, and the screening processes under the above two quantiles are called MCQ(0.5) and MCQ(0.35). To make the screening results more convincing, this section proposes to compare the MMACQ method with the MCQ(0.5) and MCQ(0.35) screening processes mentioned above, as well as the CC screening process (feature screening after removing all samples with missing responses), F screening (feature screening under full data), and the MAR feature screening method proposed by [16] for comparison.

According to the definition of the global screening index in the MMACQ feature screening method, in the simulation process, we set the quantile sequence as $\{0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95\}$, which is $m = 10$. We use the logistic model $p(\delta = 1|X, \beta) = \exp(X^\top\beta)/(1 + \exp(X^\top\beta))$ to generate the missing index δ, where β is a $p-$ dimensional parameter vector, and by selecting the appropriate parameter β, the missing rate of the response variable can be controlled at around 50% and 70%. In order to reduce the randomness of data generated during the Monte Carlo numerical simulation, the results in this section are based on 500 repeated experiments. During all experiments, the size of the estimated set of important predictors is $d = [n/\log(n)]$. Set each example sample size n and predictor dimension p to be $(n, p) = (200, 1000)$ and $(n, p) = (200, 5000)$, respectively. In order to visually compare the screening performance of the above 6 screening processes, define P_a, the proportion of which all active predictors being selected into the submodel with size d over 500 replications.

Example 3.1 (Linear Model). *Consider the linear regression model*

$$Y_i = 3.5X_{1i} + 2X_{2i} + 2.5X_{3i} + 3X_{4i} + \varepsilon_i,$$

where the predictor variable $X_i = (X_{1i}, X_{2i}, \ldots, X_{pi})^T$, $(i = 1, \ldots, n)$ *is generated from a multivariate normal distribution with mean 0,* $\mathrm{cov}(X_{ji}, X_{ki}) = \rho^{|j-k|}$, $(j, k = 1, \ldots, p)$. *Let* ρ *be* 0.3, 0.6, 0.9, *respectively. The residuals* ε_i *follow a standard normal distribution. In order to test whether the above six screening processes can screen out predictors with heavy-tailed features, let* X_{1i} *obey the* t *distribution with* 3 *degrees of freedom, and let* X_{pi} *be the* $t(3) + 1$ *distribution. Selecting* $\beta = (2, 3, 0, \ldots, 0)$ *and* $\beta = (4, 6, 0, \ldots, 0, -5.5)$, *the random missing rates of the response variables are 50% and 70%, respectively. The screening results are shown in Tables 1-2.*

As shown in Tables 1-2, The MMACQ feature screening method is more effective in screening, and its screening performance is constantly improving with the strengthening of the correlation of the predictor variables. The MMACQ feature screening method is the most stable method when the response variable is missing at 50% and 70%, respectively, and its screening results are closest to those of the F method. The CC method cannot effectively screen out all important predictors because of the loss of information with missing

Table 1. The selecting rates P_a in Example 3.1 when $(n, p) = (200, 1000)$

ρ	method	50%					70%				
		X_1	X_2	X_3	X_4	P_a	X_1	X_2	X_3	X_4	P_a
0.3	MMACQ	1.000	0.940	1.000	1.000	0.940	1.000	0.940	0.940	0.935	0.850
	MCQ(0.5)	0.990	0.590	0.990	0.995	0.575	1.000	0.755	0.945	0.940	0.690
	MCQ(0.35)	1.000	0.930	1.000	1.000	0.930	1.000	0.925	0.920	0.900	0.805
	MAR	0.985	0.815	1.000	1.000	0.800	1.000	0.895	0.950	0.930	0.810
	CC	0.995	0.195	0.995	1.000	0.195	0.995	0.280	0.985	0.985	0.270
	F	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.6	MMACQ	0.990	0.995	0.995	0.995	0.985	0.990	0.995	1.000	0.990	0.975
	MCQ(0.5)	0.950	0.920	0.990	0.990	0.860	0.980	0.975	1.000	0.985	0.945
	MCQ(0.35)	0.985	0.995	0.995	0.995	0.980	0.990	0.995	1.000	0.975	0.960
	MAR	0.955	0.975	0.995	0.995	0.935	0.985	0.980	1.000	0.980	0.945
	CC	0.930	0.855	1.000	1.000	0.785	0.930	0.770	1.000	1.000	0.705
	F	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.9	MMACQ	0.985	1.000	1.000	1.000	0.985	1.000	1.000	1.000	1.000	1.000
	MCQ(0.5)	0.790	0.995	1.000	1.000	0.785	0.980	1.000	1.000	1.000	0.980
	MCQ(0.35)	0.930	1.000	1.000	1.000	0.930	0.995	1.000	1.000	1.000	0.995
	MAR	0.890	1.000	1.000	1.000	0.890	0.990	1.000	1.000	1.000	0.990
	CC	0.765	1.000	1.000	1.000	0.765	0.765	1.000	1.000	1.000	0.765
	F	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 2. The selecting rate P_a in Example 3.1 when $(n, p) = (200, 5000)$

ρ	method	50%					70%				
		X_1	X_2	X_3	X_4	P_a	X_1	X_2	X_3	X_4	P_a
0.3	MMACQ	0.995	0.905	0.990	0.975	0.885	0.995	0.925	0.900	0.860	0.745
	MCQ(0.5)	0.970	0.405	0.950	0.965	0.360	0.985	0.620	0.905	0.885	0.515
	MCQ(0.35)	1.000	0.850	0.980	0.985	0.835	0.990	0.890	0.890	0.785	0.640
	MAR	0.995	0.720	0.955	0.975	0.690	0.995	0.770	0.905	0.845	0.595
	CC	0.990	0.095	0.995	1.000	0.085	0.950	0.100	0.870	0.950	0.095
	F	1.000	1.000	1.000	1.000	1.000	1.000	0.995	1.000	1.000	0.995
0.6	MMACQ	0.995	0.995	1.000	1.000	0.990	1.000	0.985	0.980	0.985	0.970
	MCQ(0.5)	0.950	0.875	0.990	0.985	0.820	0.990	0.955	0.980	0.990	0.930
	MCQ(0.35)	0.985	0.985	0.995	1.000	0.965	0.995	0.980	0.980	0.975	0.950
	MAR	0.975	0.980	0.995	0.995	0.945	1.000	0.975	0.980	0.985	0.965
	CC	0.875	0.605	1.000	1.000	0.515	0.785	0.545	0.955	0.975	0.355
	F	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.9	MMACQ	1.000	1.000	1.000	1.000	1.000	0.985	1.000	1.000	1.000	0.985
	MCQ(0.5)	0.710	0.985	0.995	0.995	0.700	0.925	0.995	1.000	0.995	0.915
	MCQ(0.35)	0.940	1.000	1.000	1.000	0.940	0.980	1.000	1.000	1.000	0.980
	MAR	0.910	1.000	1.000	1.000	0.555	0.620	0.985	1.000	0.990	0.600
	CC	0.560	0.995	1.000	1.000	0.555	0.620	0.985	1.000	0.990	0.600
	F	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

data being deleted. When the dimension increases with a fixed sample size, the feature screening performance turns worse. But the proposed MMACQ shows the best performance.

Example 3.2 (Linear model with outliers)**.** *Use the same model settings as Example 3.1, and let 95% of the residuals follow a standard normal distribution, and the other 5% follow a normal distribution with a mean of 2 and a variance of 3. The screening results are shown in Tables 3-4. From Tables 3-4, we can see that under the outlier situation, the proposed MMACQ procedure shows the best performance which is similar to the results of Example 3.1.*

Example 3.3 (Nonlinear Model)**.** *Consider the nonlinear model*

$$Y_i = 3X_{1i}^{1/3} + 3.5I(X_{2i} < 0) + 2.5X_{3i} + 3X_{4i} + \varepsilon_i,$$

where the covariate $X_i = (X_{1i}, \ldots, X_{pi})^T, (i = 1, \ldots, n)$ *obeys a multivariate normal distribution with mean 0 and covariance* $\text{cov}(X_{ji}, X_{ki}) = \rho^{|j-k|}, (j, k = 1, \ldots, p)$. *Let* ρ *be 0.3, 0.6 and 0.9, respectively. The residual*

Table 3. The selecting rate P_a in Example 3.2 when $(n,p)=(200,1000)$

ρ	method	50%					70%				
		X_1	X_2	X_3	X_4	P_a	X_1	X_2	X_3	X_4	P_a
0.3	MMACQ	0.995	0.900	0.985	0.990	0.875	1.000	0.965	0.960	0.960	0.905
	MCQ(0.5)	0.995	0.560	0.960	0.970	0.540	0.995	0.790	0.960	0.945	0.730
	MCQ(0.35)	0.995	0.885	0.985	0.985	0.860	1.000	0.935	0.955	0.885	0.820
	MAR	1.000	0.790	0.975	0.985	0.765	1.000	0.920	0.965	0.940	0.865
	CC	1.000	0.225	1.000	1.000	0.225	0.975	0.280	0.970	0.970	0.235
	F	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.6	MMACQ	1.000	0.985	1.000	1.000	0.985	1.000	0.990	0.990	0.995	0.980
	MCQ(0.5)	0.935	0.935	1.000	1.000	0.875	0.995	0.955	0.985	0.990	0.945
	MCQ(0.35)	1.000	0.985	1.000	1.000	0.985	1.000	0.990	0.990	0.995	0.985
	MAR	0.990	0.980	1.000	1.000	0.970	1.000	0.975	0.990	0.990	0.960
	CC	0.985	0.850	1.000	1.000	0.835	0.945	0.775	0.990	1.000	0.725
	F	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.9	MMACQ	0.995	1.000	1.000	1.000	0.995	0.995	0.995	1.000	1.000	0.990
	MCQ(0.5)	0.795	1.000	1.000	1.000	0.795	0.950	1.000	1.000	1.000	0.950
	MCQ(0.35)	0.965	1.000	1.000	1.000	0.965	0.975	0.995	1.000	1.000	0.970
	MAR	0.935	1.000	1.000	1.000	0.935	0.975	1.000	1.000	1.000	0.975
	CC	0.735	1.000	1.000	1.000	0.735	0.805	0.990	1.000	1.000	0.795
	F	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 4. The selecting rate P_a in Example 3.2 when $(n,p)=(200,5000)$

ρ	method	50%					70%				
		X_1	X_2	X_3	X_4	P_a	X_1	X_2	X_3	X_4	P_a
0.3	MMACQ	1.000	0.940	0.985	0.990	0.915	0.990	0.940	0.870	0.860	0.715
	MCQ(0.5)	0.985	0.380	0.945	0.980	0.360	0.985	0.620	0.880	0.885	0.490
	MCQ(0.35)	0.995	0.920	0.980	0.985	0.895	0.990	0.875	0.855	0.785	0.625
	MAR	0.990	0.720	0.980	0.975	0.695	0.985	0.800	0.855	0.855	0.605
	CC	0.985	0.090	1.000	1.000	0.075	0.930	0.125	0.855	0.890	0.090
	F	1.000	1.000	1.000	1.000	1.000	1.000	0.990	1.000	1.000	0.990
0.6	MMACQ	1.000	1.000	1.000	1.000	1.000	0.985	0.985	0.980	0.975	0.960
	MCQ(0.5)	0.925	0.890	0.990	1.000	0.820	0.980	0.950	0.980	0.970	0.925
	MCQ(0.35)	0.990	0.990	1.000	1.000	0.980	0.990	0.975	0.980	0.965	0.950
	MAR	0.975	0.960	1.000	1.000	0.935	1.000	0.955	0.975	0.965	0.950
	CC	0.835	0.655	1.000	1.000	0.530	0.850	0.490	0.965	0.985	0.385
	F	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.9	MMACQ	0.995	1.000	1.000	1.000	0.995	0.995	0.990	0.990	0.990	0.985
	MCQ(0.5)	0.685	0.965	0.995	0.995	0.650	0.970	0.990	0.985	0.995	0.955
	MCQ(0.35)	0.940	1.000	1.000	1.000	0.940	0.985	0.985	0.985	0.990	0.970
	MAR	0.885	0.995	0.995	1.000	0.880	0.995	0.980	0.980	0.990	0.975
	CC	0.525	1.000	1.000	1.000	0.525	0.605	0.980	0.995	0.990	0.590
	F	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

ε_i follows a standard normal distribution. To test the sensitivity of the above six screening processes to heavy-tailed features, let X_{1i} be generated from a t distribution with 3 degrees of freedom, and X_{pi} follow a $t(3)+1$ distribution. The missing rates are the same to Example 1. The screening results are shown in Tables 5-6.

Tables 5-6 show that MMACQ consistently selects all significant variables with a probability higher than 0.85 under the nonlinear model assumptions, indicating that the performance of our proposed MMACQ feature screening method is the best compared with other methods.

4. Conclusion

This paper combines the inverse probability weighting technique, conditional quantile, and model averaging ideas to propose the MMACQ feature screening method in the context of ultrahigh dimensional data with random missing response variable. We confirm under certain conditions that the MMACQ feature screening method satisfies the sure screening property, implying that the method can make the set of significant predictor variables filtered with probability converging to 1 so that all true significant predictor variables are included.

Table 5. The selecting rate P_a in Example 3.3 when $(n, p) = (200, 1000)$

ρ	method	50%					70%				
		X_1	X_2	X_3	X_4	P_a	X_1	X_2	X_3	X_4	P_a
0.3	MMACQ	0.985	0.970	1.000	0.990	0.945	0.995	0.970	0.990	0.970	0.935
	MCQ(0.5)	0.940	0.840	0.995	0.970	0.770	0.970	0.865	0.980	0.970	0.825
	MCQ(0.35)	0.990	0.975	1.000	0.985	0.950	0.995	0.970	0.990	0.930	0.900
	MAR	0.985	0.980	1.000	0.980	0.945	1.000	0.965	0.980	0.915	0.875
	CC	0.915	0.445	1.000	1.000	0.375	0.825	0.490	0.995	0.990	0.370
	F	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.6	MMACQ	0.960	0.995	1.000	1.000	0.960	0.985	0.995	1.000	1.000	0.980
	MCQ(0.5)	0.675	0.980	0.995	1.000	0.650	0.935	0.995	1.000	0.995	0.935
	MCQ(0.35)	0.945	0.995	1.000	1.000	0.940	0.985	0.995	1.000	1.000	0.980
	MAR	0.945	1.000	1.000	1.000	0.945	0.995	0.995	1.000	1.000	0.990
	CC	0.470	0.965	1.000	1.000	0.440	0.615	0.910	1.000	0.995	0.535
	F	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.9	MMACQ	0.930	1.000	1.000	1.000	0.930	0.990	1.000	1.000	1.000	0.990
	MCQ(0.5)	0.270	1.000	1.000	1.000	0.270	0.930	1.000	1.000	1.000	0.930
	MCQ(0.35)	0.755	1.000	1.000	1.000	0.755	0.975	1.000	1.000	1.000	0.975
	MAR	0.815	1.000	1.000	1.000	0.815	0.975	1.000	1.000	1.000	0.975
	CC	0.150	1.000	1.000	1.000	0.150	0.330	1.000	1.000	1.000	0.330
	F	0.990	1.000	1.000	1.000	0.990	1.000	1.000	1.000	1.000	1.000

Table 6. The selecting rate P_a in Example 3.3 when $(n, p) = (200, 5000)$

ρ	method	50%					70%				
		X_1	X_2	X_3	X_4	P_a	X_1	X_2	X_3	X_4	P_a
0.3	MMACQ	0.995	0.975	1.000	0.980	0.950	0.990	0.955	0.990	0.905	0.855
	MCQ(0.5)	0.920	0.655	0.995	0.965	0.600	0.945	0.825	0.990	0.915	0.735
	MCQ(0.35)	0.995	0.970	1.000	0.970	0.935	0.985	0.960	0.980	0.785	0.755
	MAR	0.995	0.975	1.000	0.985	0.955	0.985	0.960	0.970	0.755	0.710
	CC	0.805	0.195	1.000	0.990	0.140	0.675	0.170	0.995	0.905	0.085
	F	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.6	MMACQ	0.975	1.000	1.000	0.995	0.970	0.995	0.985	1.000	1.000	0.980
	MCQ(0.5)	0.780	0.990	1.000	0.990	0.970	0.925	0.985	1.000	0.990	0.905
	MCQ(0.35)	0.960	1.000	1.000	0.995	0.955	0.995	0.985	1.000	0.980	0.965
	MAR	0.980	1.000	1.000	0.995	0.975	0.990	0.985	1.000	0.985	0.960
	CC	0.450	0.885	1.000	1.000	0.365	0.365	0.760	0.995	0.995	0.220
	F	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.9	MMACQ	0.960	1.000	1.000	1.000	0.960	0.975	1.000	1.000	1.000	0.975
	MCQ(0.5)	0.385	0.995	1.000	1.000	0.385	0.785	1.000	1.000	1.000	0.785
	MCQ(0.35)	0.875	1.000	1.000	1.000	0.875	0.960	1.000	1.000	1.000	0.960
	MAR	0.935	1.000	1.000	1.000	0.935	0.960	1.000	1.000	0.995	0.955
	CC	0.175	1.000	1.000	1.000	0.175	0.160	1.000	1.000	1.000	0.160
	F	0.995	1.000	1.000	1.000	0.995	0.995	1.000	1.000	1.000	0.995

Since the quantile is insensitive to data with outliers and data with heavy-tailed distribution characteristics, the screening results of the MMACQ method should not be affected by it. The Monte Carlo numerical simulation verifies this conjecture.

Acknowledgment

Peng Lai's research is supported by National Natural Science Foundation of China (11771215).

5. Supplementary material: Technical Proofs

Let $\pi(X, \theta) = P(\delta = 1 | X_{\mathcal{A}_\delta}, \theta)$, $\pi(X, \theta)$ represents a function with parameter θ that is used to estimate $P(\delta = 1 | X_{\mathcal{A}_\delta}, \theta)$. $P_i = \pi(X_{\mathcal{A}_\delta i}, \theta) = P(\delta_i = 1 | X_{\mathcal{A}_\delta i}, \theta)$ represents the probability that the ith sample response variable is not missing. $\hat{P}_i = \pi(X_{\hat{\mathcal{A}}_\delta i}, \hat{\theta})$ represents the sample estimate of $P_i = \pi(X_{\mathcal{A}_\delta i}, \theta)$. To facilitate the proof, we introduce the following lemmas.

Lemma 5.1. *There exists a constant* $v_1 > 0$ *such that* $\left\| P_i^2 - \hat{P}_i^2 \right\| \leq v_1 O_p(n^{-1/2})$.

Proof. From the theoretical properties of the SCAD variable selection method, we obtained that under certain regularity conditions, $P\left(\mathcal{A}_\delta = \hat{\mathcal{A}}_\delta^*\right) = 1$. According to the properties of logistic regression coefficient estimation, it can be obtained that there are constants $v > 0$ and $v_1 > 0$, such that

$$\begin{aligned}\left\| P_i^2 - \hat{P}_i^2 \right\| &= \left| \pi\left(X_{\mathcal{A}_\delta i}, \theta\right) - \pi\left(X_{\hat{\mathcal{A}}_\delta i}, \hat{\theta}\right) \right| \left| \pi\left(X_{\mathcal{A}_\delta i}, \theta\right) + \pi\left(X_{\hat{\mathcal{A}}_\delta i}, \hat{\theta}\right) \right| \\ &\leq v \left| \pi'\left(X_{\mathcal{A}_\delta i}, \xi\right)\left(\theta - \hat{\theta}\right) \right| \leq v_1 O_p\left(n^{-1/2}\right).\end{aligned}$$

□

Lemma 5.2. *There exists a constant* $v_2 > 0$ *such that* $\left| P_i - \hat{P}_i \right| \leq v_2 O_p\left(n^{-1/2}\right)$.

Proof. The proof is similar to the proof of Lemma 5.1. □

Proof of Theorem 1. Similar to the proof of Theorem 1 by [4], let

$$\tilde{d}_k^*(t) = n^{-1} \sum_{i=1}^n \frac{\delta_i}{P_i}\left[\tau - I\left\{Y_i < Q_\tau(Y)\right\}\right] I\left(X_{ik} < t\right).$$

Define $\left\| \tilde{d}_k^* \right\| = n^{-1} \sum_{i=1}^n \tilde{d}_k^*\left(X_{ik}\right)^2 = \dfrac{(n-1)(n-2)}{n^2}\left(\dfrac{1}{n-2}\tilde{D}_{k1}^* + \tilde{D}_{k2}^*\right)$, where

$$\begin{aligned}\tilde{D}_{k1}^* &= \frac{2}{n(n-1)} \sum_{i<j} \frac{1}{2}\left\{ \frac{\delta_i}{P_i^2}\left[\tau - I\left\{Y_i < Q_\tau(Y)\right\}\right]^2 I\left(X_{ik} < X_{jk}\right)\right. \\ &\quad \left. + \frac{\delta_j}{P_j^2}\left[\tau - I\left\{Y_j < Q_\tau(Y)\right\}\right]^2 I\left(X_{jk} < X_{ik}\right)\right\} \\ &= \frac{2}{n(n-1)} \sum_{i<j} \varphi_1\left(X_{ik}, Y_i, \delta_i, P_i; X_{jk}, Y_j, \delta_j, P_j; Q_\tau\right),\end{aligned}$$

$$\begin{aligned}\tilde{D}_{k2}^* &= \frac{6}{n(n-1)(n-2)} \sum_{i<j<l} \frac{1}{3}\left\{ \frac{\delta_i\delta_j}{P_iP_j}\left[\tau - I\{Y_i < Q_\tau(Y)\}\right]\left[\tau - I\{Y_j < Q_\tau(Y)\right]\right\} I\left(X_{ik} < X_{jk}\right) \\ &\quad \cdot I\left(X_{jk} < X_{ik}\right) + \frac{\delta_j\delta_l}{P_jP_l}\left[\tau - I\left\{Y_j < Q_\tau(Y)\right\}\right]\left[\tau - I\left\{Y_l < Q_\tau(Y)\right\}\right] I\left(X_{jk} < X_{lk}\right) I\left(X_{lk} < X_{jk}\right) \\ &\quad + \frac{\delta_l\delta_i}{P_lP_i}\left[\tau - I\left\{Y_l < Q_\tau(Y)\right\}\right]\left[\tau - I\left\{Y_i < Q_\tau(Y)\right\}\right] I\left(X_{lk} < X_{ik}\right) I\left(X_{ik} < X_{lk}\right)\} \\ &= \frac{6}{n(n-1)(n-2)} \sum_{i<j<l} \varphi_2\left(X_{ik}, Y_i, \delta_i, P_i; X_{jk}, Y_j, \delta_j, P_j; X_{lk}, Y_l, \delta_l, P_l; Q_\tau\right),\end{aligned}$$

where $\varphi_1(\cdot)$ and $\varphi_2(\cdot)$ are the kernel functions of the U-statistics $\tilde{D}_{k1}^*$ and $\tilde{D}_{k2}^*$, respectively. By definition

$$\left\| \hat{d}_k^* \right\| = \frac{(n-1)(n-2)}{n^2}\left(\frac{1}{n-2}\hat{D}_{k1}^* + \hat{D}_{k2}^*\right), \tag{3}$$

where the composition of $\hat{D}_{k1}^{*}$ and $\hat{D}_{k2}^{*}$ is the same as $\tilde{D}_{k1}^{*}$ and $\tilde{D}_{k2}^{*}$, but P_i, P_j, P_l, $Q_\tau(Y)$ need to be changed into $\hat{P}_i$, $\hat{P}_j$, $\hat{P}_l$, $\hat{Q}_\tau(Y)$.

First look at $\left|\hat{D}_{k1}^{*}-\tilde{D}_{k1}^{*}\right|$. Taking full advantage of the triangle inequality, there exist constants v_2^{*}, $v_3>0$ such that the following inequality holds

$$\begin{aligned}\left|\hat{D}_{k1}^{*}-\tilde{D}_{k1}^{*}\right| &\le \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\delta_i}{\hat{P}_i^2}\left[\tau^2+(1-2\tau)\hat{Q}_i\right]-\frac{\delta_i}{P_i^2}\left[\tau^2+(1-2\tau)Q_i\right]\right| \\ &\le \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\delta_i}{\hat{P}_i^2}(1-2\tau)\left(\hat{Q}_i-Q_i\right)\right|+\left|\frac{P_i^2-\hat{P}_i^2}{P_i^2\hat{P}_i^2}\delta_i\left[\tau^2+(1-2\tau)Q_i\right]\right| \\ &\le \frac{\left|1-2\tau\right|}{n}\sum_{i=1}^{n}\frac{\delta_i}{P_i^2}\left|\left(\hat{Q}_i-Q_i\right)\right|+v_3O_p\left(n^{-1/2}\right),\end{aligned}$$

where $Q_i\triangleq I\{Y_i<Q_\tau(Y)\}$, $\hat{Q}_i\triangleq I\left\{Y_i<\hat{Q}_\tau(Y)\right\}$.

Note that for $\forall\varepsilon>0$, there is

$$\sup_{t':|t'-t|\le\varepsilon}\Big|I\{Y<t'\}-I\{Y<t\}\Big|\le I\{t-\varepsilon\le Y<t+\varepsilon\}.$$

From Theorem 2.2 of [21], we can get $\left|\hat{Q}_\tau(Y)-Q_\tau(Y)\right|=O\left(n^{-1/2}\right)$. Let

$$v_4=1+4\left|1-2\tau\right|\int\frac{1}{P}f\left(Q_\tau(Y),X_{\mathcal{A}_\delta}\right)dX_{\mathcal{A}_\delta},$$

where $P=P\left(\delta=1\middle|X_{\mathcal{A}_\delta}\right)$. For any $\eta>0$, under event $\left\{\left|\hat{Q}_\tau(Y)-Q_\tau(Y)\right|\le\eta n^{-\alpha}v_4^{-1}\right\}$, we can get

$$\begin{aligned}\left|\hat{D}_{k1}^{*}-\tilde{D}_{k1}^{*}\right| &\le \frac{\left|1-2\tau\right|}{n}\sum_{i=1}^{n}\frac{\delta_i}{P_i^2}\left|I\left\{Y_i<\hat{Q}_\tau(Y)\right\}-I\{Y_i<Q_\tau(Y)\}\right|+v_3O\left(n^{-1/2}\right) \\ &\le \frac{\left|1-2\tau\right|}{n}\sum_{i=1}^{n}\frac{\delta_i}{P_i^2}I\left\{Q_\tau(Y)-\eta n^{-\alpha}v_4^{-1}\le Y_i<Q_\tau(Y)+\eta n^{-\alpha}v_4^{-1}\right\}+v_3O_p\left(n^{-1/2}\right).\end{aligned}$$

Let

$$\begin{aligned}\mu_1 &= E\left[\frac{\delta}{P^2}I\left\{Q_\tau(Y)-\eta n^{-\alpha}v_4^{-1}\le Y_i<Q_\tau(Y)+\eta n^{-\alpha}v_4^{-1}\right\}\right] \\ &= E\left\{\frac{1}{P}E\left[I\left(Q_\tau(Y)-\eta n^{-\alpha}v_4^{-1}\le Y_i<Q_\tau(Y)+\eta n^{-\alpha}v_4^{-1}\right)\middle|X_{\mathcal{A}_\delta}\right]\right\}.\end{aligned}$$

The Taylor expansion of μ_1 is

$$\mu_1=2\eta n^{-\alpha}v_4^{-1}\int\frac{f\left\{Q_\tau(Y),X_{\mathcal{A}_\delta}\right\}}{P}dX_{\mathcal{A}_\delta}+\frac{\eta n^{-2\alpha}v_4^{-2}}{2}\int\frac{f'_{Y|X_{\mathcal{A}_\delta}}\{Q_\tau^{*}(Y)\}-f'_{Y|X_{\mathcal{A}_\delta}}\left\{Q_\tau^{+}(Y)\right\}}{P}f\left(X_{\mathcal{A}_\delta}\right)dX_{\mathcal{A}_\delta},$$

where $Q_\tau^*(Y)$ and $Q_\tau^+(Y)$ are both in the $\eta n^{-\alpha} v_4^{-1}$ neighborhood of $Q_\tau(Y)$. There exists a constant $v_5 > 0$ that can make

$$\left| \int \frac{1}{P} \left[f'_{Y|X_{\mathcal{A}_\delta}} \left\{ Q_\tau^* (Y) \right\} - f'_{Y|X_{\mathcal{A}_\delta}} \left\{ Q_\tau^+ (Y) \right\} \right] f\left(X_{\mathcal{A}_\delta} \right) dX_{\mathcal{A}_\delta} \right| \leq v_5.$$

Choose appropriate n or η to make $\left|1 - 2\tau\right| \eta n^{-\alpha} v_5 \leq 2v_4$. When $\tau \neq 1/2$, we can get

$$\begin{aligned}
\frac{\eta n^{-\alpha}}{\left|1-2\tau\right|} - \mu_1 &= \frac{\eta n^{-\alpha}}{\left|1-2\tau\right|} - \frac{\eta n^{-2\alpha} v_4^{-2}}{2} \int \frac{f'_{Y|X_{\mathcal{A}_\delta}} \left\{ Q_\tau^* (Y) \right\} - f'_{Y|X_{\mathcal{A}_\delta}} \left\{ Q_\tau^+ (Y) \right\}}{P} \cdot f(X_{\mathcal{A}_\delta}) dX_{\mathcal{A}_\delta} \\
&\quad - 2\eta n^{-\alpha} v_4^{-1} \int \frac{1}{P} f\left\{ Q_\tau(Y), X_{\mathcal{A}_\delta} \right\} dX_{\mathcal{A}_\delta} \\
&\geq \frac{\eta n^{-\alpha}}{\left|1-2\tau\right|} \cdot \frac{v_4 - 1}{v_4} - 2\eta n^{-\alpha} v_4^{-1} \int \frac{1}{P} f\left\{ Q_\tau(Y), X_{\mathcal{A}_\delta} \right\} dX_{\mathcal{A}_\delta} \\
&= 2\eta n^{-\alpha} v_4^{-1} \int \frac{1}{P} f\left\{ Q_\tau(Y), X_{\mathcal{A}_\delta} \right\} dX_{\mathcal{A}_\delta}.
\end{aligned}$$

When n satisfies $n \geq (v_3/\eta)^{2/(1-2\alpha)}$, it can be deduced by combining the above formula

$$\begin{aligned}
&P\left(\left| \hat{D}_{k1}^* - \tilde{D}_{k1}^* \right| \geq 2\eta n^{-\alpha} \right) \\
&\leq P\left(\left| \frac{\left|1-2\tau\right|}{n} \sum_{i=1}^{n} \frac{\delta_i}{P_i^2} I\left\{ Q_\tau(Y) - \eta n^{-\alpha} v_4^{-1} \leq Y_i < Q_\tau(Y) + \eta n^{-\alpha} v_4^{-1} \right\} + v_3 O\left(n^{-1/2} \right) \right| \geq 2\eta n^{-\alpha} \right) \\
&\leq P\left(\left| n^{-1} \sum_{i=1}^{n} \frac{\delta_i}{P_i^2} I\left\{ Q_\tau(Y) - \eta n^{-\alpha} v_4^{-1} \leq Y_i < Q_\tau(Y) + \eta n^{-\alpha} v_4^{-1} \right\} \right| - \mu_1 \right. \\
&\qquad \left. \geq 2\eta n^{-\alpha} v_4^{-1} \int \frac{1}{P} f\left\{ Q_\tau(Y), X_{\mathcal{A}_\delta} \right\} dX_{\mathcal{A}_\delta} \right).
\end{aligned}$$

From Hoeffding's inequality, we can get

$$P\left(\left| \hat{D}_{k1}^* - \tilde{D}_{k1}^* \right| \geq 2\eta n^{-\alpha} \right) \leq \exp\left[-8\eta^2 \left(\int p^{-1} f\left(Q_\tau(Y), X_{\mathcal{A}_\delta} \right) dX_{\mathcal{A}_\delta} \right)^2 n^{1-2\alpha} v_4^{-2} \right].$$

When $\tau = 1/2$,

$$\left| \hat{D}_{k1}^* - \tilde{D}_{k1}^* \right| \leq \frac{\left|1-2\tau\right|}{n} \sum_{i=1}^{n} \frac{\delta_i}{P_i^2} \left| \left(\hat{Q}_i - Q_i \right) \right| + v_3 O_p\left(n^{-1/2} \right) = v_3 O_p\left(n^{-1/2} \right).$$

$P\left(\left| \hat{D}_{k1}^* - \tilde{D}_{k1}^* \right| \geq 2\eta n^{-\alpha} \right) = 0$ is always established because of $\alpha \in \left[0, 1/4\right)$, so no matter what value τ takes, we have

$$P\left(\left| \hat{D}_{k1}^* - \tilde{D}_{k1}^* \right| \geq 2\eta n^{-\alpha} \right) \leq \exp\left[-8\eta^2 \left(\int p^{-1} f\left(Q_\tau(Y), X_{A_\delta} \right) dX_{A_\delta} \right)^2 n^{1-2\alpha} v_4^{-2} \right]. \tag{4}$$

Similarly, for $\left| \hat{D}_{k2}^* - \tilde{D}_{k2}^* \right|$, there exists some constant $v_8 > 0$, we can obtain

$$P\left(\left| \hat{D}_{k2}^* - \tilde{D}_{k2}^* \right| \geq 2\eta n^{-\alpha} \right) \leq \exp\left[-8\eta^2 f^2\left(Q_\tau(Y) \right) v_8^{-2} n^{1-2\alpha} \right]. \tag{5}$$

Combining equations (3)-(5), we can get

$$
\begin{aligned}
P\left(\left|\left\|\hat{d}_k^*\right\|-\left\|\tilde{d}_k^*\right\|\right| \geq 4\eta n^{-\alpha}\right) &\leq P\left(\left|\hat{D}_{k1}^*-\tilde{D}_{k1}^*\right| \geq 2\eta n^{1-\alpha}\right)+P\left(\left|\hat{D}_{k2}^*-\tilde{D}_{k2}^*\right| \geq 2\eta n^{-\alpha}\right) \\
&\leq \exp\Bigg[-8\eta^2\left(\int P^{-1} f\left(Q_\tau\left(Y\right), X_{\mathcal{A}_\delta}\right) dX_{\mathcal{A}_\delta}\right)^2 n^{3-2\alpha} v_4^{-2}\Bigg] \\
&\quad +\exp\Bigg[-8\eta^2 f^2\left(Q_\tau\left(Y\right) v_8^{-2} n^{1-2\alpha}\right)\Bigg]. \qquad (6)
\end{aligned}
$$

It is noted that $\|d_k^*\| = E\left\{\varphi_2\left(X_{ik}, Y_i, \delta_i, P_i; X_{jk}, Y_j, \delta_j, P_j; X_{lk}, Y_l, \delta_l, P_l; Q_\tau\right)\right\} = E\left(\tilde{D}_{k2}^*\right)$. Using Marcov's inequality, it can be obtained that for any $\varepsilon > 0$ and $t > 0$ we have

$$
P\left\{\tilde{D}_{k2}^*-\|d_k^*\| \geq \varepsilon\right\} \leq \exp\left(-t\varepsilon\right)\exp\left(-t\left\|d_k^*\right\|\right) E\left\{\exp\left(t\tilde{D}_{k2}^*\right)\right\}.
$$

Following the proof process of [4], as described in Section 5.1.6 of [23], $\tilde{D}_{k2}^*$ can be rewritten as

$$
\tilde{D}_{k2}^* = \frac{1}{n!}\sum_{n!} D_2^*\left(X_{1k}, Y_1, \delta_1, P_1; \ldots; X_{nk}, Y_n, \delta_n, P_n; Q_\tau\right),
$$

where $\sum_{n!}$ represents the sum of all possible permutations in $(1, 2, \ldots, n)$, and each $D_2^*\left(X_{1k}, Y_1, \delta_1, P_1; \ldots;\right.$ $\left.X_{nk}, Y_n, \delta_n, P_n; Q_\tau\right)$ is an independent and identically distributed random variable with mean $\phi \equiv \left[n/3\right]$. Let

$$
\psi\left(t\right) = E\Big[\exp\left\{t\varphi_2\left(X_{ik}, Y_i, \delta_i, P_i; X_{jk}, Y_j, \delta_j, P_j, X_{lk}, Y_l, \delta_l, P_l; Q_\tau\right)\right\}\Big],
$$

then using Jensen's inequality we can get

$$
E\left\{\exp\left(t\tilde{D}_{k2}^*\right)\right\} = E\left\{\exp\left[t(n!)^{-1}\sum_{n!} D_2^*\Big(X_{1k}, Y_1, \delta_1, P_1; \ldots; X_{nk}, Y_n, \delta_n, P_n; Q_\tau\Big)\right]\right\} \leq \psi^{\phi}\Big(t/\phi\Big).
$$

And then we can get

$$
\begin{aligned}
&P\left\{\tilde{D}_{k2}^*-\|d_k^*\| \geq \varepsilon\right\} \leq \exp\left(-t\varepsilon\right)\exp\left(-t\left\|d_k^*\right\|\right)\psi^{\phi}\left(t/\phi\right) \\
&\leq \exp\left(-t\varepsilon\right) E^{\phi}\left(\exp\left\{\phi^{-1} t\Big[\varphi_2\left(X_{ik}, Y_i, \delta_i, P_i; X_{jk}, Y_j, \delta_j, P_j; X_{lk}, Y_l, \delta_l, P_l; Q_\tau\right)-\|d_k^*\|\Big]\right\}\right).
\end{aligned}
$$

According to Lemma 1 in [3], we can get

$$
\begin{aligned}
&E\left(\exp\left\{\phi^{-1} t\Big[\varphi_2\left(X_{ik}, Y_i, \delta_i, P_i; X_{jk}, Y_j, \delta_j, P_j; X_{lk}, Y_l, \delta_l, P_l; Q_\tau\right)-\|d_k\|\Big]\right\}\right) \\
&\leq \exp\left\{\left(\phi^{-1} t\right)^2\left(2/\ell_1\right)^2/8\right\} = \exp\left\{t^2/\left(2\phi^2\ell_1^2\right)\right\}.
\end{aligned}
$$

Therefore

$$
P\left\{\left|\tilde{D}_{k2}^*-\|d_k^*\|\right| \geq \varepsilon\right\} \leq \exp\left(\Big[-t\varepsilon+t^2/\left(2\phi\ell_1^2\right)\Big]\right).
$$

Choose $t = \ell_1^2\phi\varepsilon$, $\ell_2 = \ell_1^2/2$, we can get $P\Big\{\tilde{D}_{k2}^* - \|d_k^*\| \geq \varepsilon\Big\} \leq \exp\Big(-\varepsilon^2\phi\ell_2\Big)$. From the symmetry of the U statistic, we can further obtain

$$P\left\{\left|\tilde{D}_{k2}^* - \|d_k^*\|\right| \geq \varepsilon\right\} \leq 2\exp\left(-\varepsilon^2\phi\ell_2\right).$$

Similarly, it can be proved that:

$$P\left\{\left|\tilde{D}_{k1}^* - E\left(\tilde{D}_{k1}^*\right)\right| \geq \varepsilon\right\} \leq 2\exp\left(-\varepsilon^2\phi\ell_2\right).$$

Obviously, $0 \leq \|d_k^*\| = E\left(\tilde{D}_{k2}^*\right) \leq E\left(\left|\tilde{D}_{k2}^*\right|\right) \leq 1/\ell_1^2$ and $0 \leq E\left(\tilde{D}_{k1}^*\right) \leq E\left(\left|\tilde{D}_{k1}^*\right|\right) \leq 1/\ell_1^2$. Furthermore, we can get

$$0 \leq \max_{1\leq s\leq 2}\sup_p\max_{1\leq k\leq p} E\left(\tilde{D}_{ks}^*\right) \leq 1/\ell_1^2.$$

Let $\varepsilon = \eta n^{-\alpha}$, when n is large enough, there are $(3n-2)\,n^{-2}E\left(\tilde{D}_{k2}^*\right) < \eta n^{-\alpha}$ and $(n-1)\,n^{-2}E\left(\tilde{D}_{k1}^*\right) < \eta n^{-\alpha}$, and then we can get

$$P\left(\left|\left\|\tilde{d}_k^*\right\| - \|d_k^*\|\right| \geq 4\eta n^{-\alpha}\right) \leq 2\exp\left(-\ell_3\eta^2 n^{3-2\alpha}\right) + 2\exp\left(-\ell_4\eta^2 n^{1-2\alpha}\right). \tag{7}$$

Combining equations (6) and (7), it can be deduced that:

$$P\left(\left|\left\|\hat{d}_k^*\right\| - \|d_k^*\|\right| \geq 8\eta n^{-\alpha}\right) \leq 3\exp\left(-\ell_5 n^{3-2\alpha}\right) + 3\exp\left(-\ell_6 n^{1-2\alpha}\right), \tag{8}$$

where

$$\ell_5 = \min\left\{8\eta^2\left(\int P^{-1}f\left(Q_\tau\left(Y\right), X_{\mathcal{A}_\delta}\right)dX_{\mathcal{A}_\delta}\right)^2 v_4^{-2}, \ell_3\eta^2\right\}, \ell_6 = \min\left\{8\eta^2 f^2\left(Q_\tau\left(Y\right)\right)v_8^{-2}, \ell_4\eta^2\right\}.$$

It can be obtained from (8), let $c = 8\eta$, there exist positive constants c_1 and c_2, such that

$$P\left(\max_{1\leq k\leq p}\left|\left\|\hat{d}_k^*\right\| - \|d_k^*\|\right| \geq cn^{-\alpha}\right) \leq O\left\{p\exp\left(-c_1 n^{3-2\alpha}\right) + p\exp\left(-c_2 n^{1-2\alpha}\right)\right\}.$$

If $\mathcal{A}_\tau \not\subset \hat{\mathcal{A}}_\tau$, there exists $k^* \in \mathcal{A}_\tau$ and $k^* \notin \hat{\mathcal{A}}_\tau$. According to the conditions, $k^* \notin \hat{\mathcal{A}}_\tau$ indicates $\left\|\hat{d}_{k^*}^*\right\| < cn^{-\alpha}$, and then $\min_{k\in\mathcal{A}_\tau}\|d_k^*\| \geq 2cn^{-\alpha}$ can get $\left|\left\|\hat{d}_{k^*}^*\right\| - \|d_{k^*}^*\|\right| > cn^{-\alpha}$, so there is:

$$P\left(\mathcal{A}_\tau \not\subset \hat{\mathcal{A}}_\tau\right) = P\left(\left|\left\|\hat{d}_{k^*}^*\right\| - \|d_{k^*}^*\|\right| > cn^{-\alpha}\right) \leq P\left(\max_{k\in\mathcal{A}_\tau}\left|\left\|\hat{d}_{k^*}^*\right\| - \|d_{k^*}^*\|\right| \geq cn^{-\alpha}\right).$$

Therefore

$$\begin{aligned}P\left(\mathcal{A}_\tau \subseteq \hat{\mathcal{A}}_\tau\right) &= 1 - P\left(\mathcal{A}_\tau \not\subset \hat{\mathcal{A}}_\tau\right) \geq 1 - P\left(\max_{k\in\mathcal{A}_\tau}\left|\left\|\hat{d}_{k^*}^*\right\| - \|d_{k^*}^*\|\right| \geq cn^{-\alpha}\right)\\ &\geq 1 - O\left(a_n\exp\left(-c_1 n^{3-2\alpha}\right) + a_n\exp\left(-c_2 n^{1-2\alpha}\right)\right),\end{aligned}$$

where $a_n = \left|\mathcal{A}_\tau\right|$ represents the number of elements in set $\mathcal{A}_\tau$.

So far, Theorem 1 has been proved. $\square$

Proof of Theorem 2. Next, we prove Theorem 2, which is similar to [22].

$$
\begin{aligned}
P\left(\left|\hat{R}_k - R_k\right| \geq cn^{-\alpha}\right) &= P\left(\left|\frac{m^{-1}\sum_{s=1}^{m}\hat{\tilde{\omega}}_{\tau_s}^{*} I\left(X_k \in \hat{\mathcal{A}}_{\tau_s}\right)}{m^{-1}\sum_{s=1}^{m}\hat{\tilde{\omega}}_{\tau_s}^{*}} - \frac{m^{-1}\sum_{s=1}^{m}\tilde{\omega}_{\tau_s}^{*} I\left(X_k \in \mathcal{A}_{\tau_s}\right)}{m^{-1}\sum_{s=1}^{m}\tilde{\omega}_{\tau_s}^{*}}\right| \geq cn^{-\alpha}\right) \\
&\leq P\left(\left|m^{-1}\sum_{s=1}^{m}\hat{\tilde{\omega}}_{\tau_s}^{*}\right|^{-1}\left|m^{-1}\sum_{s=1}^{m}\hat{\tilde{\omega}}_{\tau_s}^{*} I\left(X_k \in \hat{\mathcal{A}}_{\tau_s}\right) - m^{-1}\sum_{s=1}^{m}\tilde{\omega}_{\tau_s}^{*} I\left(X_k \in \mathcal{A}_{\tau_s}\right)\right| \geq \frac{cn^{-\alpha}}{2}\right) \\
&+ P\left(\frac{\left|m^{-1}\sum_{s=1}^{m}\tilde{\omega}_{\tau_s}^{*} I\left(X_k \in \mathcal{A}_{\tau_s}\right)\right|}{\left|m^{-1}\sum_{s=1}^{m}\tilde{\omega}_{\tau_s}^{*}\right|\left|m^{-1}\sum_{s=1}^{m}\hat{\tilde{\omega}}_{\tau_s}^{*}\right|}\left|m^{-1}\sum_{s=1}^{m}\hat{\tilde{\omega}}_{\tau_s}^{*} - m^{-1}\sum_{s=1}^{m}\tilde{\omega}_{\tau_s}^{*}\right| \geq \frac{cn^{-\alpha}}{2}\right) := S_3 + S_4.
\end{aligned}
$$

According to the definition $\tilde{\omega}_{\tau_s}^{*} = \left|\mathcal{A}_{\tau_s}\right|^{-1}\sum_{k\in\mathcal{A}_{\tau_s}}\left\|d_{k,\tau_s}^{*}\right\|$, combined with the condition (C1), we can get:

$$
\infty > M \geq \sup_s \frac{1}{\left|\mathcal{A}_{\tau_s}\right|}\sum_{k\in\mathcal{A}_{\tau_s}}\left\|d_{k,\tau_s}\right\| = \sup_s \tilde{\omega}_{\tau_s} \geq \inf_s \tilde{\omega}_{\tau_s} = \inf_s \frac{1}{\left|\mathcal{A}_{\tau_s}\right|}\sum_{k\in\mathcal{A}_{\tau_s}}\left\|d_{k,\tau_s}\right\| \geq 2cn^{-\alpha} = M_1.
$$

Let $M_2 = M_1 - \delta$, δ is a constant and satisfy $\delta \in (0,1)$. After calculation, we can get

$$
P\left(\left|m^{-1}\sum_{s=1}^{m}\hat{\tilde{\omega}}_{\tau_s}^{*}\right| \leq M_2\right) \leq P\left(\left|m^{-1}\sum_{s=1}^{m}\hat{\tilde{\omega}}_{\tau_s}^{*} - m^{-1}\sum_{s=1}^{m}\tilde{\omega}_{\tau_s}^{*}\right| \geq \delta\right) \leq \sum_{s=1}^{m} P\left(\left|\hat{\tilde{\omega}}_{\tau_s}^{*} - \tilde{\omega}_{\tau_s}^{*}\right| \geq \delta\right).
$$

First, we calculate S_3.

$$
\begin{aligned}
S_3 &= P\left(\left|m^{-1}\sum_{s=1}^{m}\hat{\tilde{\omega}}_{\tau_s}^{*}\right|^{-1}\left|m^{-1}\sum_{s=1}^{m}\hat{\tilde{\omega}}_{\tau_s}^{*} I\left(X_k \in \hat{\mathcal{A}}_{\tau_s}\right) - m^{-1}\sum_{s=1}^{m}\tilde{\omega}_{\tau_s}^{*} I\left(X_k \in \mathcal{A}_{\tau_s}\right)\right| \geq \frac{cn^{-\alpha}}{2}\right) \\
&\leq P\left(\left|\frac{1}{m}\sum_{s=1}^{m}\hat{\tilde{\omega}}_{\tau_s}^{*}\right| \leq M_2\right) + P\left(\left|\frac{1}{m}\sum_{s=1}^{m}\hat{\tilde{\omega}}_{\tau_s}^{*} I\left(X_k \in \hat{\mathcal{A}}_{\tau_s}\right) - \frac{1}{m}\sum_{s=1}^{m}\tilde{\omega}_{\tau_s}^{*} I\left(X_k \in \mathcal{A}_{\tau_s}\right)\right| \geq \frac{M_2 cn^{-\alpha}}{2}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
& P\left(\left|\frac{1}{m}\sum_{s=1}^{m}\hat{\tilde{\omega}}_{\tau_s}^{*} I\Big(X_k \in \hat{\mathcal{A}}_{\tau_s}\Big) - \frac{1}{m}\sum_{s=1}^{m}\tilde{\omega}_{\tau_s}^{*} I\Big(X_k \in \mathcal{A}_{\tau_s}\Big)\right| \geq \frac{M_2 c n^{-\alpha}}{2}\right) \\
& \leq P\Bigg(\left|\frac{1}{m}\sum_{s=1}^{m}\hat{\tilde{\omega}}_{\tau_s}^{*} I\Big(X_k \in \hat{\mathcal{A}}_{\tau_s}\Big) - \hat{\tilde{\omega}}_{\tau_s}^{*} I\Big(X_k \in \hat{\mathcal{A}}_{\tau_s}\Big)\right| \\
& \quad + \left|\frac{1}{m}\sum_{s=1}^{m}\hat{\tilde{\omega}}_{\tau_s}^{*} I\Big(X_k \in \mathcal{A}_{\tau_s}\Big) - \tilde{\omega}_{\tau_s}^{*} I\Big(X_k \in \mathcal{A}_{\tau_s}\Big)\right| \geq \frac{M_2 c n^{-\alpha}}{2}\Bigg) \\
& \leq \sum_{s=1}^{m} P\left(\left|\hat{\tilde{\omega}}_{\tau_s}^{*} I\Big(X_k \in \hat{\mathcal{A}}_{\tau_s}\Big) - \hat{\tilde{\omega}}_{\tau_s}^{*} I\Big(X_k \in \hat{\mathcal{A}}_{\tau_s}\Big)\right| \geq \frac{M_2 c n^{-\alpha}}{4}, \mathcal{A}_{\tau_s} = \hat{\mathcal{A}}_{\tau_s}\right) \\
& \quad + \sum_{s=1}^{m} P\left(\left|\hat{\tilde{\omega}}_{\tau_s}^{*} I\Big(X_k \in \hat{\mathcal{A}}_{\tau_s}\Big) - \hat{\tilde{\omega}}_{\tau_s}^{*} I\Big(X_k \in \hat{\mathcal{A}}_{\tau_s}\Big)\right| \geq \frac{M_2 c n^{-\alpha}}{4}, \mathcal{A}_{\tau_s} \neq \hat{\mathcal{A}}_{\tau_s}\right) \\
& \quad + \sum_{s=1}^{m} P\left(\left|\hat{\tilde{\omega}}_{\tau_s}^{*} - \tilde{\omega}_{\tau_s}^{*}\right| \geq \frac{M_2 c n^{-\alpha}}{4}\right) \\
& \leq 0 + \sum_{s=1}^{m} P\left(\mathcal{A}_{\tau_s} \neq \hat{\mathcal{A}}_{\tau_s}\right) + \sum_{s=1}^{m} P\left(\left|\hat{\tilde{\omega}}_{\tau_s}^{*} - \tilde{\omega}_{\tau_s}^{*}\right| \geq \frac{M_2 c n^{-\alpha}}{4}\right).
\end{aligned}
$$

From Theorem 1, it can be obtained that there exist positive constants ℓ_7 and ℓ_8 such that:

$$
\begin{aligned}
& P\left(\left|\left\|\hat{d}_{k,\tau_s}^{*}\right\| - \left\|d_{k,\tau_s}^{*}\right\|\right| \geq c n^{-\alpha}\right) \leq 3\exp\left(-\ell_7 n^{3-2\alpha}\right) + 3\exp\left(-\ell_8 n^{1-2\alpha}\right), \\
& P\left(\mathcal{A}_{\tau_s} \subseteq \hat{\mathcal{A}}_{\tau_s}\right) \geq 1 - O\left(a_{n,\tau_s}\exp\left(-c_1 n^{3-2\alpha}\right) + a_{n,\tau_s}\exp\left(-c_2 n^{1-2\alpha}\right)\right),
\end{aligned}
$$

where $a_{n,\tau_s} = \left|\mathcal{A}_{\tau_s}\right|$, represents the number of elements in the local important variable set $\mathcal{A}_{\tau_s}$.

According to the definition of the local important variable screening index and $\left\{\hat{\mathcal{A}}_{\tau_s} \not\subset \mathcal{A}_{\tau_s}\right\} = \Big\{\exists j^* \in \hat{\mathcal{A}}_{\tau_s}, \text{s.t.} j^* \notin \mathcal{A}_{\tau_s}\Big\}$, we can get

$$
j^* \notin \mathcal{A}_{\tau_s} \Rightarrow \left\|d_{j^*,\tau_s}^{*}\right\| = 0, j^* \in \hat{\mathcal{A}}_{\tau_s} \Rightarrow \left\|\hat{d}_{j^*,\tau_s}^{*}\right\| \geq c n^{-\alpha}.
$$

Then there is $\left|\left\|\hat{d}_{j^*,\tau_s}^{*}\right\| - \left\|d_{j^*,\tau_s}^{*}\right\|\right| \geq c n^{-\alpha}$, so that we can get

$$
\begin{aligned}
& P\left(\hat{\mathcal{A}}_{\tau_s} \not\subseteq \mathcal{A}_{\tau_s}\right) = P\left(\left|\left\|\hat{d}_{j^*,\tau_s}^{*}\right\| - \left\|d_{j^*,\tau_s}^{*}\right\|\right| \geq c n^{-\alpha}\right) \\
& \leq P\left(\max_{k}\left|\left\|\hat{d}_{k,\tau_s}^{*}\right\| - \left\|d_{k,\tau_s}^{*}\right\|\right| \geq c n^{-\alpha}\right) \\
& \leq O\left(a_{n,\tau_s}\exp\left(-\ell_7 n^{3-2\alpha}\right) + a_{n,\tau_s}\exp\left(-\ell_8 n^{1-2\alpha}\right)\right).
\end{aligned}
$$

By further derivation, we can get:

$$P\left(\hat{\mathcal{A}}_{\tau_s} \subseteq \mathcal{A}_{\tau_s}\right) = 1 - P\left(\hat{\mathcal{A}}_{\tau_s} \not\subset \mathcal{A}_{\tau_s}\right) \geq 1 - O\left(a_{n,\tau_s} \exp\left(-\ell_7 n^{3-2\alpha}\right) + a_{n,\tau_s} \exp\left(-\ell_8 n^{1-2\alpha}\right)\right).$$

There exist positive constants c_3 and c_4:

$$\begin{aligned}
&P\left(\mathcal{A}_{\tau_s} = \hat{\mathcal{A}}_{\tau_s}\right) = P\left(\left(\mathcal{A}_{\tau_s} \subseteq \hat{\mathcal{A}}_{\tau_s}\right) \cap \left(\hat{\mathcal{A}}_{\tau_s} \subseteq \mathcal{A}_{\tau_s}\right)\right) \\
&= P\left(\mathcal{A}_{\tau_s} \subseteq \hat{\mathcal{A}}_{\tau_s}\right) + P\left(\hat{\mathcal{A}}_{\tau_s} \subseteq \mathcal{A}_{\tau_s}\right) - P\left(\left(\mathcal{A}_{\tau_s} \subseteq \hat{\mathcal{A}}_{\tau_s}\right) \cup \left(\hat{\mathcal{A}}_{\tau_s} \subseteq \mathcal{A}_{\tau_s}\right)\right) \\
&\geq 1 - O\left(a_{n,\tau_s} \exp\left(-c_1 n^{3-2\alpha}\right) + a_{n,\tau_s} \exp\left(-c_2 n^{1-2\alpha}\right)\right) \\
&\quad + 1 - O\left(a_{n,\tau_s} \exp\left(-\ell_7 n^{3-2\alpha}\right) + a_{n,\tau_s} \exp\left(-\ell_8 n^{1-2\alpha}\right)\right) - 1 \\
&= 1 - O\left(a_{n,\tau_s} \exp\left(-c_3 n^{3-2\alpha}\right) + a_{n,\tau_s} \exp\left(-c_4 n^{1-2\alpha}\right)\right).
\end{aligned}$$

Therefore we can get

$$\begin{aligned}
S_3 &\leq \textstyle\sum_{s=1}^m P\left(\left|\hat{\tilde{\omega}}^*_{\tau_s} - \tilde{\omega}^*_{\tau_s}\right| \geq \delta\right) + \textstyle\sum_{s=1}^m P\left(\left|\hat{\tilde{\omega}}^*_{\tau_s} - \tilde{\omega}^*_{\tau_s}\right| \geq \dfrac{M_2 c n^{-\alpha}}{4}\right) \\
&\quad + O\left(a_{n,m} m \exp\left(-c_3 n^{3-2\alpha}\right) + a_{n,m} m \exp\left(-c_4 n^{1-2\alpha}\right)\right),
\end{aligned}$$

where $a_{n,m}$ is the maximum value of the set $\{a_{n,\tau_s}, 1 \leq s \leq m\}$.

Next, we consider S_4

$$\begin{aligned}
S_4 &= P\left(\left|\frac{\left|m^{-1}\sum_{s=1}^m \tilde{\omega}^*_{\tau_s} I\left(X_k \in \mathcal{A}_{\tau_s}\right)\right|}{\left|m^{-1}\sum_{s=1}^m \tilde{\omega}^*_{\tau_s}\right|\left|m^{-1}\sum_{s=1}^m \hat{\tilde{\omega}}^*_{\tau_s}\right|}\right|\left|m^{-1}\sum_{s=1}^m \hat{\tilde{\omega}}^*_{\tau_s} - m^{-1}\sum_{s=1}^m \tilde{\omega}^*_{\tau_s}\right| \geq \frac{c n^{-\alpha}}{2}\right) \\
&= P\left(\left|\frac{\left|m^{-1}\sum_{s=1}^m \tilde{\omega}^*_{\tau_s} I\left(X_k \in \mathcal{A}_{\tau_s}\right)\right|}{\left|m^{-1}\sum_{s=1}^m \tilde{\omega}^*_{\tau_s}\right|\left|m^{-1}\sum_{s=1}^m \hat{\tilde{\omega}}^*_{\tau_s}\right|}\right|\left|m^{-1}\sum_{s=1}^m \hat{\tilde{\omega}}^*_{\tau_s} - m^{-1}\sum_{s=1}^m \tilde{\omega}^*_{\tau_s}\right| \geq \frac{c n^{-\alpha}}{2}, \left|m^{-1}\sum_{s=1}^m \hat{\tilde{\omega}}^*_{\tau_s}\right| \leq M_2\right) \\
&+ P\left(\left|\frac{\left|m^{-1}\sum_{s=1}^m \tilde{\omega}^*_{\tau_s} I\left(X_k \in \mathcal{A}_{\tau_s}\right)\right|}{\left|m^{-1}\sum_{s=1}^m \tilde{\omega}^*_{\tau_s}\right|\left|m^{-1}\sum_{s=1}^m \hat{\tilde{\omega}}^*_{\tau_s}\right|}\right|\left|m^{-1}\sum_{s=1}^m \hat{\tilde{\omega}}^*_{\tau_s} - m^{-1}\sum_{s=1}^m \tilde{\omega}^*_{\tau_s}\right| \geq \frac{c n^{-\alpha}}{2}, \left|m^{-1}\sum_{s=1}^m \hat{\tilde{\omega}}^*_{\tau_s}\right| > M_2\right) \\
&\leq \textstyle\sum_{s=1}^m P\left(\left|\hat{\tilde{\omega}}^*_{\tau_s} - \tilde{\omega}^*_{\tau_s}\right| \geq \delta\right) + \textstyle\sum_{s=1}^m P\left(\left|\hat{\tilde{\omega}}^*_{\tau_s} - \tilde{\omega}^*_{\tau_s}\right| \geq \dfrac{M_2 c n^{-\alpha}}{2}\right).
\end{aligned}$$

Observe that S_3 and S_4 are mainly related to $P\left(\left|\hat{\tilde{\omega}}_{\tau_s}^{*}-\tilde{\omega}_{\tau_s}^{*}\right|\geq\delta\right)$, so next calculate $P\left(\left|\hat{\tilde{\omega}}_{\tau_s}^{*}-\tilde{\omega}_{\tau_s}^{*}\right|\geq cn^{-\alpha}\right)$.

$$
\begin{aligned}
&P\left(\left|\hat{\tilde{\omega}}_{\tau_s}^{*}-\tilde{\omega}_{\tau_s}^{*}\right|\geq cn^{-\alpha}\right)=P\left(\left|\left|\hat{\mathcal{A}}_{\tau_s}\right|^{-1}\textstyle\sum_{k\in\hat{\mathcal{A}}_{\tau_s}}\left\|\hat{d}_{k,\tau_s}^{*}\right\|-\left|\mathcal{A}_{\tau_s}\right|^{-1}\textstyle\sum_{k\in\mathcal{A}_{\tau_s}}\left\|d_{k,\tau_s}^{*}\right\|\right|\geq cn^{-\alpha}\right)\\
&=P\left(\left|\left|\hat{\mathcal{A}}_{\tau_s}\right|^{-1}\textstyle\sum_{k\in\hat{\mathcal{A}}_{\tau_s}}\left\|\hat{d}_{k,\tau_s}^{*}\right\|-\left|\mathcal{A}_{\tau_s}\right|^{-1}\textstyle\sum_{k\in\mathcal{A}_{\tau_s}}\left\|d_{k,\tau_s}^{*}\right\|\right|\geq cn^{-\alpha},\mathcal{A}_{\tau_s}\subset\hat{\mathcal{A}}_{\tau_s}\right)\\
&\quad+P\left(\left|\left|\hat{\mathcal{A}}_{\tau_s}\right|^{-1}\textstyle\sum_{k\in\hat{\mathcal{A}}_{\tau_s}}\left\|\hat{d}_{k,\tau_s}^{*}\right\|-\left|\mathcal{A}_{\tau_s}\right|^{-1}\textstyle\sum_{k\in\mathcal{A}_{\tau_s}}\left\|d_{k,\tau_s}^{*}\right\|\right|\geq cn^{-\alpha},\mathcal{A}_{\tau_s}\not\subset\hat{\mathcal{A}}_{\tau_s}\right)\\
&\leq P\left(\left|\mathcal{A}_{\tau_s}\right|^{-1}\textstyle\sum_{k\in\mathcal{A}_{\tau_s}}\left|\left\|\hat{d}_{k,\tau_s}^{*}\right\|-\left\|d_{k,\tau_s}^{*}\right\|\right|\geq\dfrac{cn^{-\alpha}}{2}\right)\\
&\quad+P\left(\left|\hat{\mathcal{A}}_{\tau_s}\right|^{-1}\textstyle\sum_{k\in\hat{\mathcal{A}}_{\tau_s}\setminus\mathcal{A}_{\tau_s}}\left|\left\|\hat{d}_{k,\tau_s}^{*}\right\|-\left\|d_{k,\tau_s}^{*}\right\|\right|\geq\dfrac{cn^{-\alpha}}{2}\right)+P\left(\mathcal{A}_{\tau_s}\not\subset\hat{\mathcal{A}}_{\tau_s}\right)\\
&\leq\textstyle\sum_{k\in\mathcal{A}_{\tau_s}}P\left(\left|\left\|\hat{d}_{k,\tau_s}^{*}\right\|-\left\|d_{k,\tau_s}^{*}\right\|\right|\geq\dfrac{cn^{-\alpha}}{2}\right)+\textstyle\sum_{k\in\hat{\mathcal{A}}_{\tau_s}\setminus\mathcal{A}_{\tau_s}}P\left(\left|\left\|\hat{d}_{k,\tau_s}^{*}\right\|-\left\|d_{k,\tau_s}^{*}\right\|\right|\geq\dfrac{cn^{-\alpha}}{2}\right)\\
&\quad+P\left(\mathcal{A}_{\tau_s}\not\subset\hat{\mathcal{A}}_{\tau_s}\right).
\end{aligned}
$$

Because $\max_{1\leq s\leq m}\left|\hat{\mathcal{A}}_{\tau_s}\right|\leq n$, $\delta\in(0,1)$ and $\delta<M_1$, choose the appropriate c and n, we can make $cn^{-\alpha}\leq\delta$, then we can deduce

$$
\begin{aligned}
&P\left(\left|\hat{R}_k-R_k\right|\geq cn^{-\alpha}\right)\\
&\leq 4n\textstyle\sum_{s=1}^{m}P\left(\left|\left\|\hat{d}_{k,\tau_s}^{*}\right\|-\left\|d_{k,\tau_s}^{*}\right\|\right|\geq\dfrac{\delta}{2}\right)+4\displaystyle\sum_{s=1}^{m}P\left(\mathcal{A}_{\tau_s}\not\subset\hat{\mathcal{A}}_{\tau_s}\right)\\
&\quad+O\left(a_{n,m}m\exp\left(-c_3n^{3-2\alpha}\right)+a_{n,m}m\exp\left(-c_4n^{1-2\alpha}\right)\right)\\
&\quad+4n\sum_{s=1}^{m}P\left(\left|\left\|\hat{d}_{k,\tau_s}^{*}\right\|-\left\|d_{k,\tau_s}^{*}\right\|\right|\geq\frac{M_2cn^{-\alpha}}{8}\right)\\
&\leq 8n\sum_{s=1}^{m}P\left(\left|\left\|\hat{d}_{k,\tau_s}^{*}\right\|-\left\|d_{k,\tau_s}^{*}\right\|\right|\geq\frac{M_2cn^{-\alpha}}{8}\right)+4\sum_{s=1}^{m}P\left(\mathcal{A}_{\tau_s}\not\subset\hat{\mathcal{A}}_{\tau_s}\right)\\
&\quad+O\left(a_{n,m}m\exp\left(-c_3n^{3-2\alpha}\right)+a_{n,m}m\exp\left(-c_4n^{1-2\alpha}\right)\right).
\end{aligned}
$$

Further we can get that there exist positive constants c_5 and c_6, such that

$$
\begin{aligned}
P\left(\left|\hat{R}_k-R_k\right|\geq cn^{-\alpha}\right)\leq 8nm&\left\{3\exp\left(-c_5n^{3-2\alpha}\right)+3\exp\left(-c_6n^{1-2\alpha}\right)\right\}\\
&+4mO\left(a_{n,m}\exp\left(-c_1n^{3-2\alpha}\right)+a_{n,m}\exp\left(-c_2n^{1-2\alpha}\right)\right)\\
&+O\left(a_{n,m}m\exp\left(-c_3n^{3-2\alpha}\right)+a_{n,m}m\exp\left(-c_4n^{1-2\alpha}\right)\right).
\end{aligned}
$$

Therefore, we have

$$P\left(\max_{1\le k\le p}\left|\hat{R}_k - R_k\right| \ge cn^{-\alpha}\right) \le O\left\{pmn\exp\left(-c_7 n^{3-2\alpha}\right) + pmn\exp\left(-c_8 n^{1-2\alpha}\right)\right\}$$
$$+ O\left(a_{n,m}mp\exp\left(-c_9 n^{3-2\alpha}\right) + a_{n,m}mp\exp\left(-c_{10} n^{1-2\alpha}\right)\right)$$
$$+ O\left(a_{n,m}mp\exp\left(-c_{11} n^{3-2\alpha}\right) + a_{n,m}mp\exp\left(-c_{12} n^{1-2\alpha}\right)\right).$$

If $\mathrm{A} \not\subset \hat{\mathrm{A}}$, there exist $k^* \in \mathrm{A}$ and $k^* \notin \hat{\mathrm{A}}$. According to the conditions, $k^* \notin \hat{\mathrm{A}}$ indicates $\hat{R}_{k^*} < cn^{-\alpha}$, and then $\min_{k\in\mathrm{A}} R_k \ge 2cn^{-\alpha}$ can get $\left|\hat{R}_{k^*} - R_{k^*}\right| > cn^{-\alpha}$, so that

$$P\left(\mathrm{A} \not\subset \hat{\mathrm{A}}\right) = P\left(\left|\hat{R}_{k^*} - R_{k^*}\right| > cn^{-\alpha}\right) \le P\left(\max_{k\in\mathrm{A}}\left|\hat{R}_k - R_k\right| > cn^{-\alpha}\right).$$

So

$$P\left(\mathrm{A} \subseteq \hat{\mathrm{A}}\right) = 1 - P\left(\mathrm{A} \not\subset \hat{\mathrm{A}}\right) \ge 1 - P\left(\max_{k\in\mathrm{A}}\left|\hat{R}_k - R_k\right| > cn^{-\alpha}\right)$$
$$\ge 1 - O\left\{A_n nm\exp\left(-c_7 n^{3-2\alpha}\right) + A_n nm\exp\left(-c_8 n^{1-2\alpha}\right)\right\}$$
$$- O\left(A_n a_{n,m}m\exp\left(-c_9 n^{3-2\alpha}\right) + A_n a_{n,m}m\exp\left(-c_{10} n^{1-2\alpha}\right)\right)$$
$$- O\left(A_n a_{n,m}m\exp\left(-c_{11} n^{3-2\alpha}\right) + A_n a_{n,m}m\exp\left(-c_{12} n^{1-2\alpha}\right)\right),$$

where $A_n = \left|\mathrm{A}\right|$ represents the number of elements in the important variable set A.

So far, Theorem 2 has been proved. □

References

[1] Fan J, Lv J. Sure independence screening for ultrahigh dimensional feature space. J R Stat Soc B 2008; 70(5): 849-911.

[2] Fan J, Song R. Sure independence screening in generalized linear models with NP-dimensionality. Ann Stat 2010; 38(6): 3567-3604.

[3] Li R, Zhong W, Zhu L. Feature screening via distance correlation learning. J Am Stat Assoc 2012; 107(499): 1129-1139.

[4] Wu Y, Yin G. Conditional quantile screening in ultrahigh-dimensional heterogeneous data. Biometrika 2015; 102(1): 65-76.

[5] He X, Wang L, Hong HG. Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. Ann Stat 2013; 41(1): 342-369.

[6] Zhou Y, Zhu L. Model-free feature screening for ultrahigh dimensional datathrough a modified blum-kiefer-rosenblatt correlation. Stat Sin 2018; 28(3): 1351-1370.

[7] Nandy D, Chiaromonte F, Li R. Covariate information number for feature screening in ultrahigh-dimensional supervised problems. J Am Stat Assoc 2022; 117(539): 1516-1529.

[8] Zhu LP, Li L, Li R, Zhu LX. Model-free feature screening for ultrahigh-dimensional data. J Am Stat Assoc 2011; 106(496): 1464-1475.

[9] Liu Y, Chen X. Quantile screening for ultra-high-dimensional heterogeneous data conditional on some variables. J Stat Comput. Simul 2018; 88(2): 329-342.

[10] Peng M, Xiang L. Correlation-based joint feature screening for semi-competing risks outcomes with application to breast cancer data. Stat Methods Med Res 2021; 30(11): 2428-2446.

[11] Manigandan P, Alagirisamy K, Lokesh K, Pachiyappan D, Hajira Banu G. Forecasting the prevalence of COVID-19 outbreak in world using the ARIMA and exponential smoothing model. Asia Mathematika 2020; 4(2): 150-160.

[12] Mohammed MK. Modeling treatments prediction of coronary artery disease based on fuzzy soft expert system. Asia Mathematika 2020; 4(2): 1-108.

[13] Ibrahim MI, Ibrahim LK. Non-immunogenic Microenvironmental Factors Effect on Tumor Growth With Immune Response. Asia Mathematika 2020; 4(3): 14-23.

[14] Rubin DB. Inference and missing data. Biometrika 1976; 63(3): 581-592.

[15] Lai P, Ji J, Liu Y. Feature selection method for super high dimensional models with random missing response variables. Stat Decis 2017; 20-23.

[16] Lai P, Liu Y, Liu Z, Wan Y. Model free feature screening for ultrahigh dimensional data with responses missing at random. Comput Stat Data Anal 2017; 105: 201-216.

[17] Tang N, Xia L, Yan X. Feature screening in ultrahigh-dimensional partially linear models with missing responses at random. Comput Stat Data Anal 2019; 133: 208-227.

[18] Ni L, Fang F, Shao J. Feature screening for ultrahigh dimensional categorical data with covariates missing at random. Comput Stat Data Anal 2020; 142: 106824.

[19] Gao Y, Peng L, He M, Xia W. Feature filtering method for ultra high dimensional data based on model average. J Yangzhou Univ (Nat Sci Ed) 2020; 23, 7-14.

[20] Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. J Am Stat Assoc 2001; 96(456): 1348-1360.

[21] Shu X, Zhang L, Zhou Y. Quantile estimation of samples with random missing data. Acta Math Sinica 2017; 60, 155–172.

[22] Lai P, Song F, Gao Y. Stable feature screening for ultrahigh dimensional data. J Korean Stat Soc 2019; 48(2): 221-232.

[23] Serfling RJ. Approximation theorems of mathematical statistics. John Wiley & Sons 1980.