Electronic Journal of Statistics
Vol. 17 (2023) 429–463
ISSN: 1935-7524
https://doi.org/10.1214/23-EJS2109

Tests for high-dimensional single-index models*

Leheng Cai,

Center for Statistical Science and Department of Industrial Engineering, Tsinghua University
e-mail: cailh22@mails.tsinghua.edu.cn

Xu Guo†, Gaorong Li

School of Statistics, Beijing Normal University
e-mail: xustat12@bnu.edu.cn; ligaorong@bnu.edu.cn

and

Falong Tan

Department of Statistics, Hunan University
e-mail: falongtan@hnu.edu.cn

Abstract: In this paper, we aim to test the overall significance of regression coefficients in high-dimensional single-index models. We first reformulate the hypothesis testing problem under elliptical distributions for predictors. Applying distribution-based transformation, we introduce a high-dimensional score-type test statistic. Notably, no moment condition for the error term is required. Our introduced procedures are thus robust with respect to outliers in response. Moreover our procedure is free of variance estimation of the error term. We establish the test statistic's asymptotic normality under null hypothesis. Power analysis is also investigated. To further improve computational efficiency and enhance empirical powers, we also introduce a two-stage test procedure under ultrahigh-dimensional settings based on random data splitting. To eliminate the additional randomness induced by data splitting, we further develop a powerful ensemble algorithm based on multiple data splitting. We show that the ensemble algorithm can control the type I error rate at a given significance level. Extension to partial significance testing problem is also investigated. Lastly, numerical studies and real data analysis are conducted to compare with existing approaches and to illustrate the robustness and validity of our proposed test procedures.

Keywords and phrases: Data splitting, hypothesis testing, score test, single-index model.

Received April 2022.

*Xu Guo's research was supported by Beijing Natural Science Foundation (1212004) and the National Natural Science Foundation of China (12071038); Gaorong Li's research was supported by the National Natural Science Foundation of China (11871001, 12131006 and 11971001) and the Fundamental Research Funds for the Central Universities of China (2019NTSS18); Falong Tan's research was supported by the National Natural Science Foundation of China (12071119), the Natural Science Foundation of Hunan Province (2020JJ4162), and the Fundamental Research Funds for the Central Universities.

†Corresponding author: Xu Guo.

Contents

1. Introduction

With the rapid development of information technology, high-dimensional data are now frequently collected in various areas, including biomedical engineering, microarray analysis, and finance. In these areas, a critical question is whether predictors $\boldsymbol{X} = (X_1, \cdots, X_p)^T \in \mathbb{R}^p$ as a whole contribute to an interested response $Y \in \mathbb{R}$ or not. The main challenge is that the dimension p is usually very large, even much larger than the sample size n. The high-dimensional nature leads classical procedures, such as F-test, fail.

Recently, many efforts have been devoted to solve this problem for high-dimensional data. [13] proposed an empirical Bayes test. [30] illustrated the failure of F-test for high-dimensional data, introduced a novel test statistic based on a U-statistic, and established its asymptotic distribution. [11] proposed a rank-based score test for high-dimensional linear model. [4] employed a U-statistic of order two and enhanced the power using refitted cross-validation variance estimation. For other recent developments, see also [21] and [22].

However, these test procedures are limited to linear models, which are often restrictive for high-dimensional data modeling. In this paper, we consider high-dimensional single-index model, that is,

$$Y = g(\boldsymbol{\beta}^T \boldsymbol{X}, \epsilon) \quad \text{with } \epsilon \perp\!\!\!\perp \boldsymbol{X}. \tag{1.1}$$

Here $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^T$ is a p-dimensional unknown parametric vector, the link function $g(\cdot, \cdot)$ is unknown and $\perp\!\!\!\perp$ means independence. Compared with linear model, model (1.1) is very general. Actually, it includes linear models, generalized linear models and transformed models [12]. In model (1.1), $Y \perp\!\!\!\perp \boldsymbol{X} | \boldsymbol{\beta}^T \boldsymbol{X}$, which means that the response Y is independent of the predictors $\boldsymbol{X}$ given the index variable $\boldsymbol{\beta}^T \boldsymbol{X}$.

We mainly concern about whether the predictors $\boldsymbol{X}$ are significant for the response. This leads to a global hypothesis testing problem for the index parameter $\boldsymbol{\beta}$

$$H_0 : \boldsymbol{\beta} = \mathbf{0}, \quad \text{versus} \quad H_1 : \boldsymbol{\beta} \neq \mathbf{0}. \tag{1.2}$$

Recently, [6] developed test procedures for an individual regression coefficient of interest in high-dimensional single-index model, while we aim to test the overall significance of predictors in this paper. Besides we want to introduce robust procedures. The existence of the unknown link function $g(\cdot, \cdot)$ makes the above inference problem difficult. A key observation is that when the predictors follow elliptical distributions, the general single-index model can be recast into pseudo-linear models with transformed response [31, 28]. Thus estimation of the unknown link function $g(\cdot, \cdot)$ can be avoided. For robustness consideration, we adopt the distribution-based transformation. Based on this transformation, we construct a high-dimensional score-type test procedure. Notably, moment condition for the error term ϵ is totally avoided for our procedure. Our introduced procedures are thus robust with respect to outliers in response. As noted by [4], empirical performances of many existing procedures are adversely affected by the overestimation of the variance. While our procedure is free of error variance estimation.

We establish the asymptotic normality under the null hypothesis. Power analysis is also investigated under local alternative hypotheses. Different from [1], [30], and [11], the asymptotic distributions in our paper are not derived under the pseudo-independence assumption under which the predictors $\boldsymbol{X}$ are generated by larger dimensional factors. Instead we assume that the predictors follow the elliptical distribution, which is very general and includes multivariate normal distribution and multivariate t distribution. We note that [4] and [22] also made elliptical distribution assumption in high-dimensional inference for linear model.

When the dimension p is much larger than the sample size n in ultrahigh-dimensional data, the aforementioned score test might perform unsatisfactorily. Indeed even though the response Y depends on the predictors $\boldsymbol{X}$, usually only a small subset of predictors are significant to the response. This then motivates us to introduce a two-stage test procedure based on random data splitting. The idea of data splitting has been used successfully in various statistical problems. In fact, [29] and [24] adopted the data splitting strategy to conduct variable selection with error rate control. [8] used the data splitting idea to estimate the error variance in an ultrahigh-dimensional linear model. The basic idea here is to use one part of the data to detect potentially significant variables and reduce dimensionality. Then we apply the proposed high-dimensional score test procedure to the other part of the data and the reduced predictors to further determine their overall significance. We should emphasize that the potentially significant variables detected in the first step may include many noise predictors. Actually, all variables are inactive under the null hypothesis. Thus their significance should be checked. By exploiting the sparsity information, the two-stage procedure can enhance the detection power. To reduce the additional randomness induced by data splitting, we further develop a powerful ensemble algorithm based on a multiple data splitting strategy. We show that the ensemble algorithm can control the type I error rate.

Besides global testing, testing the significance of a part $\boldsymbol{\beta}$ is also of great importance. Many authors have investigated this important problem for linear or generalized linear regression models. See for instance [18], [14], and more recently [15]. Inspired by these works, we then further investigate partial significance testing problem in high-dimensional single-index model. Suitable test statistic is introduced and its asymptotic normality is also established.

The paper is organized as follows. In Section 2, we develop the high-dimensional score test and study its asymptotic distributions. We introduce the two-stage procedure in Section 3. Section 4 discusses the extension to partial significance testing problem. Section 5 examines the finite-sample performance of the proposed procedures using Monte Carlo simulations and a real-data example. Section 6 concludes the paper and all the technical proofs are provided in Appendix.

2. Test statistic construction and its asymptotic distributions

2.1. Reformulation of the hypothesis

Without loss of generality, we assume that $\mathrm{E}(\boldsymbol{X}) = \mathbf{0}$ and $\Sigma = \mathrm{E}(\boldsymbol{X}\boldsymbol{X}^T) > 0$ throughout this paper. In high-dimensional inference literature, the pseudo-independence assumption is commonly imposed. The pseudo-independence assumption resembles a factor model structure where the p-dimensional random vector $\boldsymbol{X}$ is generated linearly by a larger dimensional factor vector $\boldsymbol{Z}$. Though it is assumed in [1], [30], and [11], it may be difficult to validate in practice and excludes some multivariate distributions such as multivariate t-distribution.

In this paper, we instead consider that the random vector $\boldsymbol{X}$ follows a p-dimensional elliptical distribution. The elliptical distribution is often assumed in multivariate analysis [10]. It contains a large family of multivariate distributions, such as multivariate normal distribution, multivariate t-distribution, and multivariate logistic distribution.

Now we present the definition of elliptical distributions.

Definition 1. (The elliptical distribution assumption) A random vector $\boldsymbol{X}$ follows an elliptical distribution if and only if $\boldsymbol{X}$ has the following explicit expression

$$\boldsymbol{X} = R \times \Gamma \boldsymbol{U}.$$

Here Γ is a $p \times p$ matrix; $\boldsymbol{U}$ is a random vector uniformly distributed on the unit sphere in $\mathbb{R}^p$, and the generating variate R is a nonnegative random variable, satisfying $\mathrm{E}(R^2) = p$, $\mathrm{Var}(R^2) = O(p)$, and also independent with $\boldsymbol{U}$. Note that the calculation rule between Γ and $\boldsymbol{U}$ is matrix multiplication, and $\times$ denotes scalar multiplication.

In high-dimensional inference for linear model, [4] and [22] also made the elliptical distribution assumption. In practice, we may adopt existing tests for elliptical distributions (see for instance [17] and [5]) based on several principal components of $\boldsymbol{X}$. Alternatively, we may apply coordinatewise gaussianization to transform predictors into normal distributions, i.e. $\tilde{X}_{ij} = \Phi^{-1}\left(\frac{n}{n+1}\hat{F}_j(X_{ij})\right)$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable and $\hat{F}_j$ is the empirical cumulative distribution function of the j-th component of $\boldsymbol{X}$. See [23] for more details of coordinatewise gaussianization. In this paper, the elliptical distribution assumption is imposed for theoretical developments.

With the elliptical distribution assumption, we can recast the single-index model into pseudo-linear models with transformed response. We state this result in the following lemma.

Lemma 1. *Assume that $\boldsymbol{X}$ follows an elliptical distribution. Then for any given transformation function $h(\cdot)$ of the response Y, under model (1.1), there exists some constant κ depending on function $h(\cdot)$ such that*

$$\boldsymbol{\beta}_h =: \Sigma^{-1} Cov(\boldsymbol{X}, h(Y)) = \kappa \times \boldsymbol{\beta}.$$

The above lemma follows directly from Proposition 1 in [12]. See also [31] and [28]. Throughout the paper, we assume that $\kappa \neq 0$. When $h(\cdot)$ is monotone and $g(\cdot,\cdot)$ is monotone with respect to the first argument, this assumption is satisfied. See [12] for more discussions.

Lemma 1 tells us that we can use linear regression structure to recover the unknown index parameter $\boldsymbol{\beta}$ in model (1.1). To be more specific, applying a given transformation function $h(Y)$, we obtain a transformed linear model

$$h(Y) = \boldsymbol{\beta}_h^T \boldsymbol{X} + e. \tag{2.1}$$

By the definition of $\boldsymbol{\beta}_h$, it is clear that the predictors $\boldsymbol{X}$ and the error term $e = h(Y) - \boldsymbol{\beta}_h^T \boldsymbol{X}$ are uncorrelated. That is, $\mathrm{E}(\boldsymbol{X}e) = \mathrm{E}[\boldsymbol{X}(h(Y) - \boldsymbol{X}^T\boldsymbol{\beta}_h)] = 0$. Since $\boldsymbol{\beta}_h$ is proportional to $\boldsymbol{\beta}$, Lemma 1 provides us an opportunity to convert the hypothesis (1.2) to the following equivalent one

$$H_0 : \boldsymbol{\beta}_h = \mathbf{0}, \quad \text{versus} \quad H_1 : \boldsymbol{\beta}_h \neq \mathbf{0}. \tag{2.2}$$

For the choice of the transformation function, in this paper we specially focus on the function $h(Y) = F(Y) - 1/2$, where $F(Y)$ is the cumulative distribution function of Y. This choice makes our procedure be robust with respect to outliers in response. A brief discussion with another simple choice $h(Y) = Y$ is made in Remark 1 later.

Different form the classical linear model, the predictors $\boldsymbol{X}$ and the error term e now may not be independent in the transformed linear model. While the existing works in linear models greatly rely on the independence between the predictors and the random error term. Further the transformed response $F(Y)$ is unknown and has to be estimated. These points bring many difficulties in methodological and theoretical developments in our model setting.

2.2. High-dimensional score test

In this subsection, we propose a score-type test statistic for hypothesis (2.2) when the dimension p diverges to infinity. Suppose that $(\boldsymbol{X}_i, Y_i)_{i=1}^n$ is a random sample from the population $(\boldsymbol{X}, Y)$. Note that $\boldsymbol{X}$ is uncorrelated with $F(Y) - 1/2$ under H_0. Thus a natural measurement of the closeness between $\boldsymbol{\beta}_h$ and $\mathbf{0}$ is the L_2 norm $\mathrm{E}\left\|\boldsymbol{X}(F(Y)-1/2)\right\|^2$. One may consider to use the sample version to estimate $\mathrm{E}\left\|\boldsymbol{X}(F(Y)-1/2)\right\|^2$ as follows:

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{X}_i^T\boldsymbol{X}_j\left(F_n(Y_i)-\frac{1}{2}\right)\left(F_n(Y_j)-\frac{1}{2}\right).$$

Here, $F_n(Y_i)$ is the empirical distribution function

$$F_n(Y_i) = \frac{1}{n}\sum_{j=1}^{n} I(Y_j \le Y_i).$$

However, the diagonal elements add some technical difficulty. To this end, we remove the unwanted diagonal elements and consider the following U-statistic with kernel $\boldsymbol{X}_i^T\boldsymbol{X}_j(F_n(Y_i)-1/2)(F_n(Y_j)-1/2)$

$$S_n = \frac{12}{n(n-1)}\sum_{i\neq j}\boldsymbol{X}_i^T\boldsymbol{X}_j\left(F_n(Y_i)-\frac{1}{2}\right)\left(F_n(Y_j)-\frac{1}{2}\right). \tag{2.3}$$

The number 12 in S_n is to normalize the variance of the term $F_n(Y_i) - 1/2$. Note that $\sum_{j=1}^n I(Y_j \le Y_i)$ is the rank of Y_i. Since statistics with ranks are well-known to be robust, this then intuitively explains why our procedures are robust with respect to outliers in response.

To derive the asymptotic properties of S_n, we need to claim some mild conditions. We first present the condition to regulate the diverging dimension p and the sample size n.

$$\Sigma > 0 \text{ and } tr(\Sigma^4) = o(tr^2(\Sigma^2)) \quad p \to \infty \ as \ n \to \infty. \tag{C1}$$

This mild condition frequently appeared in the literature [30, 4]. The positive definiteness of the covariance matrix Σ ensures the identification of the parameter $\boldsymbol{\beta}_h$ in model (2.1). We also allow the eigenvalues of Σ diverge to infinity when $p \to \infty$. Note that if all the eigenvalues are bounded, then $tr(\Sigma^4) = o(tr^2(\Sigma^2))$ holds trivially.

Next, we define the following local alternative hypotheses for the local power analysis.

$$\boldsymbol{\beta}_h^T\Sigma\boldsymbol{\beta}_h = o(1), \quad \boldsymbol{\beta}_h^T\Sigma^2\boldsymbol{\beta}_h = o\left(\sqrt{\frac{tr(\Sigma^2)}{n}}\right), \quad \boldsymbol{\beta}_h^T\Sigma^3\boldsymbol{\beta}_h = o\left(\frac{tr(\Sigma^2)}{n}\right).$$

Similar conditions were also considered in [30] and [11]. The above local alternative hypotheses H_{1n} obviously describe a small discrepancy between $\boldsymbol{\beta}_h$ and $\mathbf{0}$. The following theorem shows the asymptotic behavior of the test statistic S_n under H_0.

Theorem 1. *Assume the condition (C1) holds and assume that $\boldsymbol{X}$ follows an elliptical distribution, under H_0, as $n \to \infty$ and $p \to \infty$, then*

$$\frac{n}{\sqrt{2tr(\Sigma^2)}} S_n \xrightarrow{d} N(0,1).$$

Remark 1. From Lemma 1, we may also consider another simple and natural choice of $h(Y) = Y$. With this choice, the corresponding test statistic is $S_n' = \frac{1}{n(n-1)} \sum_{i \neq j} \boldsymbol{X}_i^T \boldsymbol{X}_j Y_i Y_j$ and under null hypothesis, we would have $\frac{n}{\sqrt{2tr(\Sigma^2)\sigma^2}} S_n' \xrightarrow{d} N(0,1)$. Here $\sigma^2 = \mathrm{Var}(e)$ denotes the variance of the error term. Hence, it would generally fail when the stochastic error is heavy-tailed or there exist outliers. Notably, our procedure with $h(Y) = F(Y) - 1/2$ does not need to estimate the variance of the error term, which brings large convenience and simplifies computation. Further no moment condition for the error term is required. Thus our procedures are robust with respect to outliers in response.

To establish the score test based on Theorem 1, we now need to estimate $tr(\Sigma^2)$. Following [30], we use the ratio consistent estimator of $tr(\Sigma^2)$ below.

$$\widehat{tr(\Sigma^2)} = \frac{1}{2\binom{n}{4}} \sum_{i_1 < i_2 < i_3 < i_4} (\boldsymbol{X}_{i_1} - \boldsymbol{X}_{i_2})^T (\boldsymbol{X}_{i_3} - \boldsymbol{X}_{i_4})(\boldsymbol{X}_{i_2} - \boldsymbol{X}_{i_3})^T (\boldsymbol{X}_{i_4} - \boldsymbol{X}_{i_1}).$$

Combining Theorem 1 and the Slutsky Theorem, our proposed score test rejects H_0 at a significance level α if

$$nS_n \geq \sqrt{2\widehat{tr(\Sigma^2)}} z_\alpha.$$

Here z_α is the upper-α quantile of standard normal distributions.

Due to the nonlinear dependence between the predictors $\boldsymbol{X}$ and the error term e, it is hard to calculate the theoretical distribution under local alternatives. Provided that S_n diverges to infinity under local alternatives, our proposed test procedure is consistent, which means the power tends to 1 as $n \to \infty$. We present the following theorem for power analysis.

Theorem 2. *Suppose that there exist a sufficiently small positive number δ and a positive constant C such that $\boldsymbol{\beta}_h^T \Sigma^2 \boldsymbol{\beta}_h \geq C n^{-1+\delta} \sqrt{tr(\Sigma^2)}$. Under the conditions in Theorem 1 and the local alternatives, as $n \to \infty$ and $p \to \infty$, then*

$$\frac{n}{\sqrt{2tr(\Sigma^2)}} S_n \to \infty.$$

Remark 2. Regard $\boldsymbol{\beta}_h^T \Sigma^2 \boldsymbol{\beta}_h$ as the signal strength. The condition $\boldsymbol{\beta}_h^T \Sigma^2 \boldsymbol{\beta}_h \geq C n^{-1+\delta} \sqrt{tr(\Sigma^2)}$ ensures that the signal can be well detected and then the proposed test is consistent.

3. Ultrahigh-dimensional test with screening

Although the above score test procedure performs well for large dimensional data, it does not give a satisfying solution for ultrahigh-dimensional data due to power loss and computation limit. Actually even though the response Y depends on the predictors $\boldsymbol{X}$, usually only a small subset of predictors contribute to the response. This motivates us to propose a two-stage test procedure based on random data splitting.

We randomly split the data into two subsets $\mathcal{D}_1$ and $\mathcal{D}_2$. Let n_1 and n_2 be the sample sizes of $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively. Here, the balanced data splitting is adopted, namely $n_1 = [n/2]$. The sparsity assumption often holds in ultrahigh-dimensional setting, which means only quite a small subset of the predictors are significant to the response variable even under alternative hypotheses. We denote the small subset of predictors as $\mathcal{M} = \{k : \beta_{h_k} \neq 0\}$. If we can first reduce the dimensionality from ultra-high level to a moderate-high level, we can then increase the detection power. This goal is achieved by data splitting. In the first step, we adopt some feature screening procedure to the subset $\mathcal{D}_1$ to screen out noise predictors and retain potentially significant variables. In the second step, we then apply the high-dimensional score test in subsection 2.2 on the other data subset $\mathcal{D}_2$ and reduced predictors to further check their significance. Data splitting is crucial to eliminate the effect of spurious correlations due to ultrahigh dimensionality, and to avoid an inflation of the type-I error. This point is illustrated in Section 4.

Since the seminal work of [9], there are many proposals for feature screening. Considering the robustness of screening procedures, here we adopt the robust rank correlation based screening in [19]. To be more specific, we consider the Kendall's τ rank correlation coefficient of each predictor X_k with the response Y. Denote

$$\widehat{\omega}_k = \frac{1}{n_1(n_1-1)} \sum_{i \neq j}^{n_1} \operatorname{sgn}(X_{ki} - X_{kj}) \operatorname{sgn}(Y_i - Y_j),\ k = 1, \cdots, p.$$

We select a set of important predictors with large $\hat{\omega}_k$. That is, we define

$$\widehat{\mathcal{M}} = \{1 \leq k \leq p : |\widehat{\omega}_k| \text{ is among the first } d_n \text{ largest ones}\},$$

where d_n is a pre-specified threshold value. Under some regularity conditions, it has been demonstrated in [19] that the following sure screening property holds.

$$\Pr(\mathcal{M} \subseteq \widehat{\mathcal{M}}) \xrightarrow{p} 1.$$

With the sure screening property, the hypothesis (2.2) is asymptotically equivalent to the following hypothesis:

$$H_0 : \boldsymbol{\beta}_{h\widehat{\mathcal{M}}} = \mathbf{0}, \quad \text{versus} \quad H_1 : \boldsymbol{\beta}_{h\widehat{\mathcal{M}}} \neq \mathbf{0},$$

where $\boldsymbol{\beta}_{h\widehat{\mathcal{M}}} = \{\beta_{h_k} : k \in \widehat{\mathcal{M}}\}$. As a result, the screening procedure in the first step helps us to transform the ultrahigh-dimensional testing problem to a moderate high-dimensional testing problem. The benefit of screening is further illustrated in Section 4.

In the second step, we adopt the high-dimensional score test in subsection 2.2 to test $\boldsymbol{\beta}_{h\widehat{\mathcal{M}}}$ based on the other subset $\mathcal{D}_2$. While data splitting plays an essential role in dimension reduction and power enhancement, it only uses half of the data to make inference. As suggested by an anonymous reviewer, we can adopt cross-fitting method to achieve further improvement. See for instance [3] and [6]. Actually we can reverse the data set for screening and the data set for testing, obtain two corresponding screening sets $\widehat{\mathcal{M}}_1$ and $\widehat{\mathcal{M}}_2$, and construct two test statistics S_{n1} and S_{n2}, respectively. From Theorem 1, we have

$$S_n^* =: \frac{1}{\sqrt{2}} \times \left[\frac{nS_{n1}}{2\sqrt{2\widehat{tr(\Sigma^2_{\widehat{\mathcal{M}}_1})}}} + \frac{nS_{n2}}{2\sqrt{2\widehat{tr(\Sigma^2_{\widehat{\mathcal{M}}_2})}}} \right] \xrightarrow{d} N(0,1). \tag{3.1}$$

Here n is the total sample size.

We present the whole test procedure in **Algorithm 1**.

Algorithm 1 Test with screening based on single data splitting

Step 1. Randomly split the data into two subsets, denoted by $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively. Let $n_1 = |\mathcal{D}_1|$ and $n_2 = |\mathcal{D}_2|$. We use $n_1 = [n/2]$.

Step 2. For the subset $\mathcal{D}_1 = (\mathbb{X}^{(1)}, \mathbf{Y}^{(1)})$, where $\mathbb{X}^{(1)}$ is the $n_1 \times p$ design matrix and $\mathbf{Y}^{(1)}$ is the $n_1 \times 1$ response vector, calculate

$$\widehat{\omega}_k = \frac{1}{n_1(n_1-1)} \sum_{i \neq j}^{n_1} \text{sgn}(X_{ki} - X_{kj}) \text{sgn}(Y_i - Y_j), \quad k = 1, 2, \cdots, p$$

and define a submodel

$$\widehat{\mathcal{M}} = \{1 \le k \le p : |\omega_k| \text{ is among the first } d_n \text{ largest ones}\}$$

Following [9], we recommend to set $d_n = [n_1/\log(n_1)]$.

Step 3. For the subset $\mathcal{D}_2$, calculate the test statistic S_{n1} according to (2.3).

Step 4. Exchange the role of two subsets. Repeat Step 2 and Step 3 to obtain the test statistic S_n^* and output the p-value based on (3.1).

Although the test with screening based on single data splitting can enhance the empirical power, its performance in practice might be affected by additional randomness induced by data splitting. In fact, some significant predictors could be missed at the sample level due to randomness and the sample reduction. This would inflate the empirical size of the test or reduce the empirical testing power. Therefore, following [24], we introduce an ensemble algorithm based on multiple data splitting.

Now we repeat **Algorithm 1** B times and get the p-values denoted as $\{p_b,\ b = 1, 2, \cdots, B\}$. For any $\gamma \in (\gamma_{\min}, 1)$, where $\gamma_{\min}$ is a predetermined value, such as 0.05 or $1/B$, define

$$Q(\gamma) = \min\left\{1, q_\gamma\left(\frac{p_b}{\gamma}; b = 1, 2, \cdots, B\right)\right\}.$$

Here $q_\gamma(A)$ represents the lower-γ quantile of the number set A. Further let

$$Q^* = \min\left\{1, (1 - \log\gamma_{\min}) \inf_{\gamma \in (\gamma_{\min}, 1)} Q(\gamma)\right\}.$$

The above Q^* is the final adjusted p-value for the testing procedure. The ensemble test algorithm is presented in **Algorithm 2**.

The following theorem shows that the type I error rate can be controlled at α level.

Theorem 3. *Given any significance level α and any $\gamma_{\min} > 0$, the type I error rate is asymptotically controlled at α. That is, under H_0,*

$$\limsup_{n\to\infty} \Pr(Q^* \le \alpha) \le \alpha.$$

Actually, for any fixed $\gamma \in (0,1)$, $Q(\gamma)$ is an asymptotically correct p-value. However, a proper choice of γ may be difficult. Hence, we use Q^* to give a final version of the adjusted p-value, which is an adaptive method to select a suitable value of γ. Here, $\gamma_{\min}$ is a lower bound for γ. Then, the correction factor $1 - \log \gamma_{\min}$ is upper bounded. For more discussions about $Q(\gamma)$ and Q^*, kindly see [24].

Algorithm 2 Test with screening based on multiple data splitting

Step 1. Apply **Algorithm 1** to the original data and get the output p-value, denoted as p_1.
Step 2. Repeat Step 1 B times. Obtain the set of p-values $\{p_1, p_2, \cdots, p_B\}$. Recommend $B = 10$ or 20.
Step 3. Calculate the adjusted p-value

$$Q^* = \min\left\{1, (1 - \log \gamma_{\min}) \inf_{\gamma \in (\gamma_{\min}, 1)} Q(\gamma)\right\},$$

where $\gamma_{\min}$ is a predetermined value, such as 0.05 or $1/B$,

$$Q(\gamma) = \min\left\{1, q_\gamma\left(\frac{p_b}{\gamma}; b = 1, 2, \cdots, B\right)\right\},$$

and $q_\gamma(A)$ represents the lower-γ quantile of the number set A.
Step 4. Reject H_0 at a significance level α when $Q^* < \alpha$.

4. Extension to partial significance test

In this section, we further investigate how to construct partial significance tests for single-index models. Suppose that $\boldsymbol{X}_i = (\boldsymbol{X}_{ia}^T, \boldsymbol{X}_{ib}^T)^T$, where $\boldsymbol{X}_{ia} \in \mathbb{R}^q$ and $\boldsymbol{X}_{ib} \in \mathbb{R}^{p-q}$. Then, we rewrite model (1.1) as follows,

$$Y_i = g(\boldsymbol{X}_{ia}^T \boldsymbol{\beta}_a + \boldsymbol{X}_{ib}^T \boldsymbol{\beta}_b, \epsilon_i), \quad i = 1, 2, \cdots, n. \tag{4.1}$$

The dimension of the interested parameter $\boldsymbol{\beta}_b$ could be higher than the sample size n, while the dimension of the nuisance parameter $\boldsymbol{\beta}_a$ could also diverge to infinity. Consider the following null hypothesis

$$\tilde{H}_0 : \boldsymbol{\beta}_b = 0, \qquad \tilde{H}_1 : \boldsymbol{\beta}_b \neq 0. \tag{4.2}$$

This problem is very vital. It aims to assert that given $\boldsymbol{X}_a \in \mathbb{R}^q$, whether $\boldsymbol{X}_b \in \mathbb{R}^{p-q}$ brings additional information for the response Y. Since the dimension

p and q diverge, classical F-test also fails. In practice, $\boldsymbol{X}_a$ usually comes from some preliminary results and often includes a few but quite important variables. Instead, $\boldsymbol{X}_b$ usually includes much more predictors that remain to be explored.

Applying the distribution transformation we discussed in subsection 2.1, we could recast model (4.1) as follows,

$$F(Y_i) - \frac{1}{2} = \boldsymbol{X}_i^T \boldsymbol{\beta}_h + e_i = \boldsymbol{X}_{ia}^T \boldsymbol{\beta}_a + \boldsymbol{X}_{ib}^T \boldsymbol{\beta}_b + e_i.$$

Then, we convert the hypothesis (4.2) to the following equivalent one

$$\tilde{H}_0' : \boldsymbol{\beta}_{hb} = 0, \qquad \tilde{H}_1' : \boldsymbol{\beta}_{hb} \neq 0.$$

Let $\mathbb{X} = (\mathbb{X}_a, \mathbb{X}_b)$ denote the matrix form of sample data. Define the conditional predictor $\boldsymbol{X}_{ib}^* := \boldsymbol{X}_{ib} - \mathrm{Cov}(\boldsymbol{X}_{ib}, \boldsymbol{X}_{ia}) \mathrm{Var}(\boldsymbol{X}_{ia})^{-1} \boldsymbol{X}_{ia}$ and the projection matrix $\mathbb{P}_a := \mathbb{X}_a (\mathbb{X}_a^T \mathbb{X}_a)^{-1} \mathbb{X}_a^T$. Since $\boldsymbol{X}_i$ follows elliptical distributions, the linear combination $\boldsymbol{X}_{ib}^*$ also follows elliptical distributions. Let $\mathbf{Y}$ and $\boldsymbol{e}$ be the vector forms of sample data.

We first give some mild conditions.

(C2) The covariance matrix $\Sigma > 0$. The dimension $p \to \infty$ and $q \to \infty$, when $n \to \infty$. Assume that $p^{\frac{1}{2}} q n^{-1} \to 0$, $q n^{-\frac{1}{4}} \to 0$ and $q p^{-\frac{1}{4}} \to 0$.

(C3) Suppose that $\mathrm{E}(e_i | \boldsymbol{X}_i) = 0$, $\mathrm{E}(e_i^2 | \boldsymbol{X}_i) = \sigma^2$, and $\mathrm{E}(e_i^4 | \boldsymbol{X}_i) < \infty$.

(C4) Assume that $\boldsymbol{X}_i$ follows the elliptical distribution and the corresponding generating variate R_i^* of $\boldsymbol{X}_{ib}^*$ satisfies $\mathrm{E}(R_i^{*8}) = (1 + o(1)) p^4$.

(C5) There exist positive constants τ_1 and τ_2 such that

$$\tau_1 < \lambda_{\min}(\Sigma_{b|a}) \leq \lambda_{\max}(\Sigma_{b|a}) < \tau_2.$$

where $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ represent the smallest and largest eigenvalues of a semipositive definite matrix A, respectively. Further $\Sigma_{b|a}$ is the conditional covariance matrix as follows,

$$\Sigma_{b|a} = \mathrm{E}(\mathrm{Cov}(\boldsymbol{X}_{ib} | \boldsymbol{X}_{ia})) = (\sigma_{j_1 j_2}^*).$$

The above conditions are mild and widely imposed. Condition (C2) controls the diverging number of predictors. Condition (C3) is a conditional moment assumption for e, which is common in the literature, such as [14] and [30]. However, we should recognize that the assumption $\mathrm{E}(e_i | \boldsymbol{X}_i) = 0$ and $\mathrm{E}(e_i^2 | \boldsymbol{X}_i) = \sigma^2$ may not hold in general, since in the transformed linear model e is only uncorrelated with $\boldsymbol{X}$. This assumption is imposed to establish the asymptotic normality. Under Condition (C4), $\boldsymbol{X}_{ib}^*$ also follows the elliptical distribution. Besides, $\mathrm{E}(R_i^{*8}) = (1 + o(1)) p^4$ is imposed for theoretical derivation and similar conditions are assumed in [18]. Condition (C5) has been widely used in the literature; see for instance [7], [18], [27] and many others.

Notice that $\mathrm{E}[(F(\mathbf{Y}) - 1/2 - \mathbb{X}_a \boldsymbol{\beta}_a)^T \mathbb{X}_b] = \mathrm{E}(\mathbb{X}_b^T \boldsymbol{e}) = 0$ holds under $\tilde{H}_0'$. Hence, it naturally leads to the following plug-in test statistic L_n,

$$L_n = \frac{1}{n} \hat{\boldsymbol{e}}^T \mathbb{X}_b \mathbb{X}_b^T \hat{\boldsymbol{e}},$$

where $\hat{\boldsymbol{e}}$ is the residual, that is

$$\hat{\boldsymbol{e}} = (\mathbb{I} - \mathbb{P}_a)\left(F_n(\mathbf{Y}) - \frac{1}{2}\right)$$
$$= (\mathbb{I} - \mathbb{P}_a)\boldsymbol{e} + (\mathbb{I} - \mathbb{P}_a)(F_n(\mathbf{Y}) - F(\mathbf{Y})).$$

In order to eliminate bias, we consider the following test statistic T_n,

$$T_n = \frac{1}{n}\hat{\boldsymbol{e}}^T \mathbb{X}_b \mathbb{X}_b^T \hat{\boldsymbol{e}} - \frac{\hat{\sigma}^2}{n} tr((\mathbb{I} - \mathbb{P}_a)\mathbb{X}_b \mathbb{X}_b^T),$$

where $\hat{\sigma}^2 := (n-q)^{-1}\hat{\boldsymbol{e}}^T\hat{\boldsymbol{e}}$.

Theorem 4. *Suppose Conditions (C2)-(C5) hold. Under the null $\tilde{H}_0'$, when $n \to \infty$ and $p \to \infty$, we have*

$$\frac{T_n}{\sqrt{2\sigma^4 tr(\Sigma_{b|a}^2)}} \xrightarrow{d} N(0,1).$$

In order to construct score tests based on Theorem 4, we need to obtain consistent estimators of σ^2 and $tr(\Sigma_{b|a}^2)$. For σ^2, we use $\hat{\sigma}^2 = (n-q)^{-1}\hat{\boldsymbol{e}}^T\hat{\boldsymbol{e}}$. Following [26], we employ the following consistent estimator of $tr(\Sigma_{b|a}^2)$,

$$\widehat{tr(\Sigma_{b|a}^2)} = \frac{n^2}{(n+1-q)(n-q)}\Big(tr(\hat{\Sigma}_{b|a}^2) - (n-q)^{-1} tr^2(\hat{\Sigma}_{b|a})\Big),$$

where $\hat{\Sigma}_{b|a} = n^{-1}\mathbb{X}_b^T(\mathbb{I} - \mathbb{P}_a)\mathbb{X}_b$. Based on Theorem 4 and the Slusky Theorem, our proposed test procedure rejects $\tilde{H}_0'$ or equivalently $\tilde{H}_0$ at a significant level α if

$$T_n \geq \sqrt{2\hat{\sigma}^4 \widehat{tr(\Sigma_{b|a}^2)}}\, z_\alpha.$$

Here z_α is the upper-α quantile of standard normal distributions.

5. Numerical studies

5.1. Simulation studies

In this section, we present three simulation examples to evaluate the finite sample performance of our proposed tests.

Study 1. The first set of numerical simulations are carried out to evaluate the performance of the score test for high-dimensional data. Considering the following two models:

$$Y_i = a \times \boldsymbol{X}_i^T \boldsymbol{\beta} + \epsilon_i, \qquad \text{(Model 1)}$$

$$Y_i = 5\sqrt{|a \times \boldsymbol{X}_i^T \boldsymbol{\beta} + 10|} + 4 + \epsilon_i. \qquad \text{(Model 2)}$$

The predictors $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \cdots, X_{ip})^T$ are generated from a multivariate normal distribution $N(0, \Sigma_1)$ with $\Sigma_1 = \{0.2^{|i-j|}\}_{p\times p}$ or a multivariate t-distribution $\sqrt{3/5}t_5(0, \Sigma_1)$ with 5 degrees of freedom, respectively. For the high-dimensional setting, a nonsparse case and a sparse case are taken into account. The so-called nonsprase case means that the first half elements of the parametric vector $\boldsymbol{\beta}$ are nonzeros with equal magnitude, while for the sparse case, only the first five elements of $\boldsymbol{\beta}$ are nonzeros with equal magnitude. The two cases above both satisfy that $\|\boldsymbol{\beta}\| = 1$. The tuning parameter a is selected to be $0, 0.3, \cdots, 1.5$, where $a = 0$ represents the null hypothesis H_0. There are two cases of error distributions *Case*(1): standard normal distribution $N(0, 1)$; and *Case*(2): Cauchy distribution or equivalently Student's t-distribution with 1 degree of freedom $t(1)$.

We also compare the proposed score test with the test in [30] (denoted by ZC) and the test in [4] (denoted by CGZ). The corresponding numerical results of *Study 1* are respectively reported in Tables 1-2 for nonsparse case and Tables 3-4 for sparse case under two different dimensions $p = 75, 150$ and fixed sample size $n = 50$ at the significance level $\alpha = 0.05$. For each setting, 500 realizations are conducted to get the empirical powers or sizes.

From Tables 1-4, we have the following findings. Firstly, the empirical sizes of all three tests under each setting are closely around the predetermined significance level 0.05. Secondly, the simulation results in *Case*(2) provide a strong evidence that our proposed test procedure is sufficiently robust and hardly affected by the heavy-tailed error terms in both linear and nonlinear models, while the other two existing methods lose much effectiveness. In fact, their empirical powers can be as low as the nominal level. Thirdly, it is also noted that the empirical powers of the proposed test increase quickly as the level of signal strength a increases. However, when the dimension rises from 75 to 150, the empirical powers decrease slightly in different cases.

Table 1

Comparison of empirical sizes and powers for nonsparse case in Study 1 at the significance level $\alpha = 0.05$. $\boldsymbol{X} \sim N(0, \Sigma_1)$. S_n: the proposed test; ZC: test in [30]; CGZ: test in [4].

Non-sparse		Error terms											
		Normal			$t(1)$			Normal			$t(1)$		
Model	a	S_n	ZC	CGZ	S_n	ZC	CGZ	S_n	ZC	CGZ	S_n	ZC	CGZ
		$(n,p) = (50, 75)$						$(n,p) = (50, 150)$					
Model 1	0.0	0.062	0.076	0.070	0.044	0.012	0.006	0.060	0.066	0.054	0.060	0.024	0.022
	0.3	0.198	0.208	0.192	0.096	0.040	0.028	0.108	0.130	0.110	0.078	0.020	0.020
	0.6	0.542	0.546	0.560	0.178	0.030	0.028	0.348	0.356	0.346	0.148	0.056	0.048
	0.9	0.834	0.846	0.854	0.316	0.048	0.044	0.588	0.594	0.632	0.232	0.068	0.058
	1.2	0.926	0.940	0.944	0.452	0.066	0.056	0.760	0.768	0.784	0.304	0.066	0.062
	1.5	0.954	0.960	0.970	0.520	0.110	0.100	0.814	0.818	0.834	0.368	0.106	0.098
Model 2	0.0	0.064	0.068	0.056	0.062	0.024	0.026	0.046	0.056	0.044	0.040	0.022	0.020
	0.3	0.340	0.336	0.310	0.124	0.026	0.018	0.230	0.232	0.226	0.094	0.046	0.032
	0.6	0.540	0.544	0.528	0.196	0.020	0.024	0.390	0.388	0.398	0.130	0.030	0.030
	0.9	0.690	0.706	0.698	0.242	0.034	0.030	0.524	0.522	0.526	0.144	0.058	0.052
	1.2	0.790	0.796	0.800	0.334	0.032	0.040	0.626	0.602	0.610	0.196	0.056	0.044
	1.5	0.864	0.860	0.874	0.330	0.090	0.090	0.650	0.634	0.644	0.214	0.046	0.038

Study 2. In the second set of numerical simulations, we investigate the performance of the two-stage test procedures for ultrahigh dimension, based on single

TABLE 2
Comparison of empirical sizes and powers for nonsparse case in Study 1 at the significance level $\alpha = 0.05$. $\boldsymbol{X} \sim \sqrt{3/5}t_5(0, \Sigma_1)$. S_n: the proposed test; ZC: test in [30]; CGZ: test in [4].

Non-sparse		Error terms											
		Normal			$t(1)$			Normal			$t(1)$		
Model	a	S_n	ZC	CGZ	S_n	ZC	CGZ	S_n	ZC	CGZ	S_n	ZC	CGZ
		$(n,p) = (50, 75)$						$(n,p) = (50, 150)$					
Model 1	0.0	0.060	0.064	0.060	0.044	0.028	0.026	0.038	0.058	0.040	0.058	0.018	0.018
	0.3	0.208	0.218	0.200	0.092	0.026	0.022	0.148	0.158	0.140	0.110	0.030	0.026
	0.6	0.578	0.598	0.582	0.170	0.026	0.024	0.398	0.420	0.408	0.124	0.036	0.032
	0.9	0.842	0.842	0.860	0.338	0.042	0.036	0.590	0.588	0.612	0.226	0.040	0.034
	1.2	0.894	0.906	0.902	0.414	0.096	0.084	0.762	0.774	0.788	0.340	0.088	0.068
	1.5	0.960	0.964	0.968	0.546	0.128	0.138	0.856	0.854	0.864	0.356	0.096	0.092
Model 2	0.0	0.058	0.064	0.060	0.054	0.028	0.026	0.038	0.058	0.040	0.058	0.018	0.018
	0.3	0.406	0.410	0.402	0.156	0.028	0.020	0.238	0.270	0.244	0.114	0.032	0.028
	0.6	0.610	0.610	0.600	0.206	0.026	0.024	0.424	0.446	0.426	0.132	0.038	0.032
	0.9	0.762	0.766	0.770	0.284	0.032	0.026	0.476	0.476	0.500	0.176	0.034	0.028
	1.2	0.786	0.788	0.800	0.294	0.064	0.060	0.602	0.606	0.612	0.216	0.062	0.048
	1.5	0.834	0.862	0.864	0.376	0.076	0.078	0.680	0.710	0.716	0.256	0.062	0.058

TABLE 3
Comparison of empirical sizes and powers for sparse case in Study 1 at the significance level $\alpha = 0.05$. $\boldsymbol{X} \sim N(0, \Sigma_1)$. S_n: the proposed test; ZC: test in [30]; CGZ: test in [4].

Sparse		Error terms											
		Normal			$t(1)$			Normal			$t(1)$		
Model	a	S_n	ZC	CGZ	S_n	ZC	CGZ	S_n	ZC	CGZ	S_n	ZC	CGZ
		$(n,p) = (50, 75)$						$(n,p) = (50, 150)$					
Model 1	0.0	0.062	0.076	0.070	0.046	0.032	0.024	0.066	0.068	0.062	0.054	0.030	0.014
	0.3	0.192	0.212	0.186	0.104	0.022	0.026	0.090	0.126	0.110	0.052	0.018	0.018
	0.6	0.478	0.530	0.522	0.170	0.028	0.020	0.358	0.374	0.372	0.120	0.046	0.030
	0.9	0.804	0.818	0.826	0.278	0.066	0.056	0.522	0.558	0.556	0.210	0.036	0.036
	1.2	0.886	0.896	0.908	0.420	0.092	0.076	0.698	0.732	0.726	0.296	0.050	0.038
	1.5	0.952	0.962	0.972	0.502	0.116	0.124	0.786	0.808	0.814	0.352	0.080	0.082
Model 2	0.0	0.064	0.068	0.056	0.062	0.024	0.026	0.048	0.056	0.044	0.064	0.022	0.020
	0.3	0.322	0.324	0.288	0.130	0.022	0.018	0.238	0.236	0.228	0.124	0.046	0.040
	0.6	0.534	0.518	0.532	0.190	0.018	0.018	0.354	0.338	0.324	0.160	0.030	0.034
	0.9	0.688	0.678	0.690	0.232	0.042	0.038	0.496	0.480	0.476	0.210	0.058	0.052
	1.2	0.782	0.768	0.780	0.300	0.048	0.046	0.600	0.586	0.596	0.224	0.038	0.032
	1.5	0.832	0.822	0.832	0.314	0.080	0.074	0.652	0.628	0.642	0.246	0.038	0.040

TABLE 4
Comparison of empirical sizes and powers for sparse case in Study 1 at the significance level $\alpha = 0.05$. $\boldsymbol{X} \sim \sqrt{3/5}t_5(0, \Sigma_1)$. S_n: the proposed test; ZC: test in [30]; CGZ: test in [4].

Sparse		Error terms											
		Normal			$t(1)$			Normal			$t(1)$		
Model	a	S_n	ZC	CGZ	S_n	ZC	CGZ	S_n	ZC	CGZ	S_n	ZC	CGZ
		$(n,p) = (50, 75)$						$(n,p) = (50, 150)$					
Model 1	0.0	0.060	0.064	0.060	0.046	0.028	0.030	0.052	0.058	0.050	0.038	0.030	0.026
	0.3	0.144	0.154	0.166	0.102	0.032	0.030	0.126	0.148	0.122	0.102	0.026	0.018
	0.6	0.480	0.496	0.506	0.154	0.038	0.032	0.330	0.350	0.328	0.126	0.032	0.024
	0.9	0.712	0.732	0.742	0.264	0.054	0.050	0.532	0.548	0.546	0.216	0.054	0.046
	1.2	0.850	0.878	0.882	0.344	0.074	0.070	0.706	0.720	0.712	0.258	0.062	0.060
	1.5	0.934	0.936	0.936	0.456	0.104	0.090	0.754	0.762	0.776	0.308	0.066	0.052
Model 2	0.0	0.058	0.064	0.060	0.054	0.028	0.026	0.050	0.058	0.040	0.076	0.018	0.018
	0.3	0.340	0.352	0.328	0.136	0.036	0.024	0.218	0.228	0.212	0.108	0.032	0.026
	0.6	0.540	0.558	0.550	0.146	0.026	0.022	0.402	0.374	0.366	0.128	0.036	0.034
	0.9	0.692	0.674	0.688	0.230	0.026	0.028	0.468	0.456	0.466	0.140	0.040	0.028
	1.2	0.772	0.758	0.770	0.262	0.054	0.054	0.554	0.564	0.566	0.222	0.052	0.044
	1.5	0.786	0.790	0.786	0.356	0.072	0.080	0.616	0.594	0.590	0.252	0.064	0.056

data splitting and multiple data splitting. We call them SST (Single Splitting Test) and MST (Multiple Splitting Test) for short respectively. We introduce the following models for illustration.

$$Y_i = a \times \boldsymbol{X}_i^T \boldsymbol{\beta} + \epsilon_i, \qquad \text{(Model 1)}$$

$$Y_i = \frac{30}{1 + 30\exp(-a \times \boldsymbol{X}_i^T \boldsymbol{\beta})} + 8 + \epsilon_i. \qquad \text{(Model 3)}$$

We set $\beta_j = 1/\sqrt{5}$ for $j = 1, \cdots, 5$ while the other components are zeros. The tuning parameter a is set to be $0, 0.5, \cdots, 2.5$, in which $a = 0$ represents the null H_0. The predictors $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \cdots, X_{ip})^T$ are generated from a multivariate normal distribution $N(0, \Sigma_2)$ with $\Sigma_2 = \{0.2^{|i-j|}\}_{p\times p}$. The two error distributions in Study 1 are also considered here. For MST, we set $B = 10$ and $\gamma_{\min} = 0.05$. We set the sample size $n = 100$ and the dimension $p = 5000$. And 500 realizations are repeated for each setting.

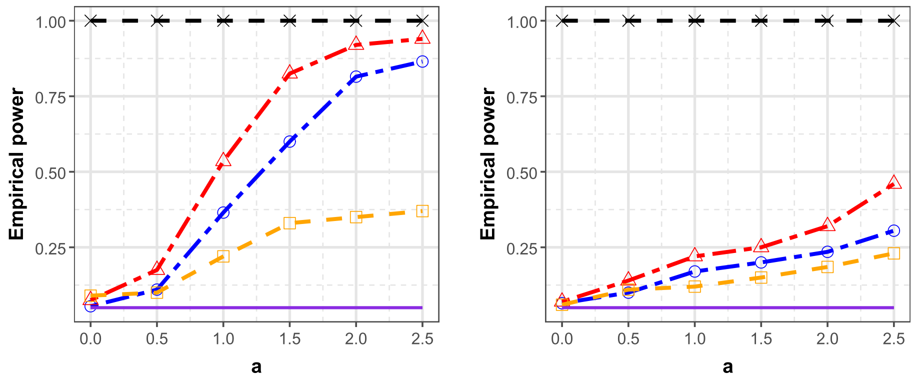
By displaying the empirical size and power curves with different values of a, the corresponding numerical results of *Study 2* are illustrated in Figure 1. We also consider two naive procedures, that is, the high-dimensional test procedure without screening, and a procedure which screens and tests without data splitting. From Figure 1, we can see that the proposed two-stage test procedures control the empirical sizes well and the empirical powers increase as the signal strength a increases. Besides, it is observed that the powers of the MST are apparently higher than SST. It also indicates that the two-stage test procedures with screening are much more efficient and sensitive to the signal than the high-dimensional test procedure without screening when the dimension p is rather large. Moreover, as shown in Figure 1, the procedure which screens and tests without data splitting cannot control the type-I error. In fact, screening and testing based on the same data would be misled by the spurious correlations in such ultrahigh-dimensional data.

Study 3. In the last set of numerical simulations, we assess the performance of the high-dimensional partial test procedure. The following models are introduced for illustration

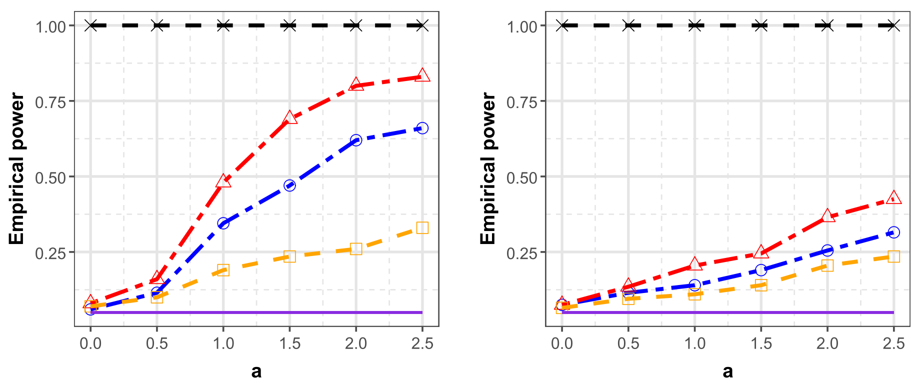
$$Y_i = c \times \boldsymbol{X}_i^T \boldsymbol{\beta} + \epsilon_i, \qquad \text{(Model 1)}$$

$$Y_i = 0.3c \times (\boldsymbol{X}_i^T \boldsymbol{\beta})^3 + 0.3c \times \boldsymbol{X}_i^T \boldsymbol{\beta} + \epsilon_i. \qquad \text{(Model 4)}$$

The predictors $\boldsymbol{X}_i = (\boldsymbol{X}_{ia}^T, \boldsymbol{X}_{ib}^T)^T = (X_{i1}, X_{i2}, \cdots, X_{ip})^T$ are generated from a multivariate normal distribution $N(0, \Sigma_3)$ with $\Sigma_3 = \{0.2^{|i-j|}\}_{p\times p}$. A nonsparse case and a sparse case are also taken into account. The nonsprase case: the first half elements of the parametric vector $\boldsymbol{\beta}_b$ are nonzeros with equal magnitude, and the sparse case: only the first five elements of $\boldsymbol{\beta}_b$ are nonzeros with equal magnitude. The two cases above both satisfy that $\|\boldsymbol{\beta}_b\| = 1$. We set $\boldsymbol{\beta}_a = (1, 1, \cdots, 1)^T$. The value of c is selected to be $0, 0.4, \cdots, 2.0$, where $c = 0$ represents the null hypothesis $\tilde{H}_0$. Two cases of error distributions are the same with those in Study 1.



(a) Model 1; left: Case (1), right: Case (2)



(b) Model 3; left: Case (1), right: Case (2)

FIG 1. *The empirical size and power curves with* $n = 100$ *and* $p = 5000$ *at the significance level* $\alpha = 0.05$*, which is highlighted by the solid lines. In four pictures above, the twodash lines with circles: SST; the twodash lines with triangles: MST; the dash lines with squares: test without screening; the dash lines with crosses: test and screen without data splitting.*

We also compare the proposed partial test with the test in [18] (denoted by LWT). The corresponding numerical results of *Study 3* are respectively reported in Table 5 for nonsparse case and Table 6 for sparse case under different dimensions and sample sizes $n = 50, p = 100, q = 5$ and $n = 100, p = 200, q = 10$ at the significance level $\alpha = 0.05$. For each setting, 500 realizations are conducted to get the empirical powers or sizes.

From Tables 5-6, we find that the empirical sizes of both two tests under each setting are closely around the predetermined significance level 0.05. Secondly, the simulation results corresponding to $t(1)$ error show that our proposed test procedure is relatively robust in both linear and nonlinear models, while the procedure in [18] lose much power. Furthermore, it could be observed that the

Table 5
Comparison of empirical sizes and powers for nonsparse case in Study 3 at the significance level $\alpha = 0.05$. $\boldsymbol{X} \sim N(0, \Sigma_3)$. T_n: the proposed partial test; LWT: test in [18].

Nonsparse		Error terms							
		Normal		$t(1)$		Normal		$t(1)$	
Model	c	T_n	LWT	T_n	LWT	T_n	LWT	T_n	LWT
		$(n,p,q) = (50,100,5)$				$(n,p,q) = (100,200,10)$			
Model 1	0.0	0.044	0.038	0.044	0.080	0.036	0.058	0.052	0.084
	0.4	0.184	0.204	0.082	0.086	0.266	0.278	0.082	0.082
	0.8	0.500	0.666	0.114	0.100	0.718	0.844	0.124	0.078
	1.2	0.684	0.868	0.228	0.108	0.834	0.986	0.200	0.092
	1.6	0.762	0.910	0.276	0.136	0.914	0.994	0.266	0.138
	2.0	0.814	0.924	0.352	0.180	0.938	0.998	0.376	0.176
Model 4	0.0	0.046	0.050	0.048	0.056	0.048	0.044	0.044	0.050
	0.4	0.512	0.332	0.196	0.138	0.788	0.222	0.384	0.148
	0.8	0.708	0.340	0.348	0.208	0.870	0.292	0.534	0.190
	1.2	0.804	0.352	0.420	0.234	0.934	0.264	0.660	0.214
	1.6	0.814	0.360	0.526	0.268	0.950	0.318	0.718	0.210
	2.0	0.820	0.362	0.546	0.270	0.954	0.266	0.758	0.194

Table 6
Comparison of empirical sizes and powers for sparse case in Study 3 at the significance level $\alpha = 0.05$. $\boldsymbol{X} \sim N(0, \Sigma_3)$. T_n: the proposed partial test; LWT: test in [18].

Sparse		Error terms							
		Normal		$t(1)$		Normal		$t(1)$	
Model	a	T_n	LWT	T_n	LWT	T_n	LWT	T_n	LWT
		$(n,p,q) = (50,100,5)$				$(n,p,q) = (100,200,10)$			
Model 1	0.0	0.052	0.044	0.052	0.078	0.056	0.032	0.052	0.086
	0.4	0.164	0.178	0.080	0.074	0.218	0.258	0.070	0.066
	0.8	0.424	0.526	0.122	0.082	0.572	0.790	0.098	0.092
	1.2	0.624	0.752	0.176	0.114	0.768	0.940	0.176	0.100
	1.6	0.726	0.836	0.216	0.136	0.846	0.970	0.236	0.118
	2.0	0.748	0.900	0.286	0.150	0.890	0.990	0.290	0.158
Model 4	0.0	0.050	0.050	0.048	0.060	0.040	0.044	0.046	0.050
	0.4	0.448	0.344	0.198	0.132	0.728	0.226	0.316	0.114
	0.8	0.638	0.332	0.296	0.194	0.858	0.232	0.476	0.152
	1.2	0.748	0.318	0.386	0.252	0.898	0.234	0.544	0.214
	1.6	0.758	0.324	0.472	0.192	0.916	0.218	0.622	0.174
	2.0	0.810	0.336	0.518	0.226	0.924	0.196	0.700	0.176

empirical powers of the proposed test is higher than LWT in nonlinear models.

5.2. Real data analysis

In this subsection, we apply our methodology to a Cardiomyopathy microarray data [25], which was once analyzed by [20], [11] and many others. This dataset consists of a $n \times p$ matrix of gene expression values $\mathbb{X} = (X_{ij})$, where X_{ij} is the expression level of the j-th gene for the i-th mouse. The response Y is the G protein-coupled receptor Ro1 expression level, and the predictors $\boldsymbol{X}$ are other potentially related gene expression levels. Here, the dimension of predictors $p = 6319$ is much larger than the sample size $n = 30$. The goal is to identify the influential genes for overexpression of the Ro1 in mice. A high-dimensional single-index model is now considered for this data set. Before modeling, both the response and predictors have been standardized to be zero mean and unit variance.

Noticing the dimension p is very large compared with the sample size n, it is then important to firstly check whether there exists any significant predictor to affect the response. A global testing problem can be considered to tackle this problem. If the null hypothesis (1.2) is not rejected, we do not need to pursue further. We adopt the score test, SST and MST, respectively. Additionally, in order to assess the robustness of our proposed methods, we add some random outliers to the response variable. To be more specific, we randomly choose a quarter of the sample and add the random noise which follows the cauchy distribution to the response. The results are presented in Table 7. From this table, we can see clearly that there truly exists some significant predictors to affect the response. Besides, since the results of the noised data are similar to those of the original data, we are confident that our proposed procedures are robust to the outlier in response.

We then consider to adopt the Kendall's τ correlation based screening approach in [19] to select the influential genes. Eight important variables are selected. It is then of interest to assert that given those eight chosen variables, the other eliminated predictors by the screening procedure are indeed irrelevant or not. This can be formulated as a partial significance testing problem. The proposed partial test statistic T_n is then applied. From Table 7, we assert that the eliminated predictors are conditionally uncorrelated with the response.

TABLE 7
The p-values of the test statistics for the real data analysis

		Score test	SST	MST
Original	full sample	4e-9(5.77)	9e-11(6.37)	6e-13
	eliminated	0.43(0.18)	—	—
Noised	full sample	6e-6(4.37)	7e-7(4.82)	6e-8
	eliminated	0.65(-0.38)	—	—

The values of test statistics are given in parentheses.

6. Conclusions and discussions

In this paper, we investigate the overall significance testing problem of regression coefficients in high-dimensional single-index models. We tackle this problem by exploiting the property of elliptical distributions. By adopting a transformed linear model framework, we propose the high-dimensional score test. With the distribution function of the response, our procedure is robust with respect to outliers in response. We show that the introduced score test statistic S_n asymptotically follows a normal distribution under null hypothesis and diverges under local alternative hypotheses with some mild conditions.

To improve computational efficiency and enhance empirical powers, we also discuss two test procedures, SST and MST, for ultrahigh-dimensional settings based on random data splitting and feature screening. We control the type I error rate by modifying the obtained p-values in MST. Additionally, we further investigate how to construct partial significance tests in high-dimensional single-index models. The asymptotic normality of the proposed test statistic T_n is derived. And the validity of the procedures is also illustrated through simulations and real data analysis.

Our proposed test statistic T_n does not allow the dimensions p and q be too large for the partial significance testing problem. It would be very interesting and challenging to test partial significance when both p and q are ultrahigh-dimensional. It is beyond the scope of this paper. We aim to investigate this issue in near future.

Appendix: technical proofs

Proofs for Theorem 1

To demonstrate the Theorems, we introduce some following lemmas.

Lemma 2. *Suppose that $\boldsymbol{U} = (U_1, U_2, \cdots, U_p)^T$ is a random vector uniformly distributed on the unit sphere in $\mathbb{R}^p$. Then $Var(\boldsymbol{U}) = p^{-1} I_p$, $E(U_j^4) = 3p^{-1}(p+2)^{-1}$, $E(U_i^2 U_j^2) = p^{-1}(p+2)^{-1}$ $i \neq j$.*

See Theorem 2.8 in [10] for the proof of Lemma 2.

Lemma 3. *Assume $\boldsymbol{X} = R\Gamma\boldsymbol{U}$ follows an elliptical distribution. Let $\boldsymbol{Z}$ denote the random vector $R\boldsymbol{U}$. For any symmetric matrix $A = (a_{ij})$ and $B = (b_{ij})$, $E(\boldsymbol{Z}^T A\boldsymbol{Z})(\boldsymbol{Z}^T B\boldsymbol{Z}) = (1 + o(1))[tr(A)tr(B) + 2tr(AB)]$ holds.*

Proof of Lemma 3. From the definition of elliptical distributions, we have $\mathrm{E}(R^4) = p^2 + O(p)$. Note that

$$(\boldsymbol{Z}^T A\boldsymbol{Z})(\boldsymbol{Z}^T B\boldsymbol{Z}) = R^4 \sum_{i,j,k,l} a_{ij} b_{kl} U_i U_j U_k U_l$$

$$= R^4 \left[\sum_{i\neq k} a_{ii} b_{kk} U_i^2 U_k^2 + \sum_{i\neq j} a_{ij} b_{ij} U_i^2 U_j^2 + \sum_{i\neq j} a_{ij} b_{ji} U_i^2 U_j^2 + \sum_i a_{ii} b_{ii} U_i^4 \right].$$

It then follows that

$$\mathrm{E}(\boldsymbol{Z}^T A\boldsymbol{Z})(\boldsymbol{Z}^T B\boldsymbol{Z}) = \mathrm{E}(R^4)\mathrm{E}(U_1^2 U_2^2) \left[tr(A)tr(B) + 2tr(AB) - 3\sum_i a_{ii} b_{ii} \right]$$

$$+ \mathrm{E}(R^4)\mathrm{E}(U_1^4)\sum_i a_{ii}b_{ii}.$$

Then from Lemma 2, we conclude that

$$\begin{aligned}\mathrm{E}(\boldsymbol{Z}^T A\boldsymbol{Z})(\boldsymbol{Z}^T B\boldsymbol{Z}) &= \mathrm{E}(R^4)\mathrm{E}(U_1^2U_2^2)\left[tr(A)tr(B) + 2tr(AB)\right] \\ &= (1+o(1))[tr(A)tr(B) + 2tr(AB)].\end{aligned}$$

Proof of Theorem 1. Note that S_n has the following decomposition

$$\begin{aligned}\frac{1}{12}S_n &= \frac{1}{n(n-1)}\sum_{i\neq j}\boldsymbol{X}_i^T\boldsymbol{X}_j\left(F_n(Y_i)-\frac{1}{2}\right)\left((F_n(Y_j)-\frac{1}{2}\right) \\ &= \frac{1}{n(n-1)}\sum_{i\neq j}\boldsymbol{X}_i^T\boldsymbol{X}_j\left(F(Y_i)-\frac{1}{2}\right)\left(F(Y_j)-\frac{1}{2}\right) \\ &\quad + \frac{2}{n(n-1)}\sum_{i\neq j}\boldsymbol{X}_i^T\boldsymbol{X}_j\left(F_n(Y_i)-F(Y_i)\right)\left(F(Y_j)-\frac{1}{2}\right) \\ &\quad + \frac{1}{n(n-1)}\sum_{i\neq j}\boldsymbol{X}_i^T\boldsymbol{X}_j\left(F_n(Y_i)-F(Y_i)\right)\left(F_n(Y_j)-F(Y_j)\right) \\ &\coloneqq A_{n1} + A_{n2} + A_{n3}.\end{aligned}$$

For the first term A_{n1}, we rewrite it as follows

$$\begin{aligned}A_{n1} &= \frac{1}{n(n-1)}\sum_{i\neq j}\boldsymbol{X}_i^T\boldsymbol{X}_j\boldsymbol{X}_i^T\boldsymbol{\beta}_h\boldsymbol{X}_j^T\boldsymbol{\beta}_h + \frac{2}{n(n-1)}\sum_{i\neq j}\boldsymbol{X}_i^T\boldsymbol{X}_j\boldsymbol{X}_i^T\boldsymbol{\beta}_h e_j \\ &\quad + \frac{1}{n(n-1)}\sum_{i\neq j}\boldsymbol{X}_i^T\boldsymbol{X}_j e_i e_j \\ &\coloneqq A_{n11} + A_{n12} + A_{n13}.\end{aligned}$$

Step 1. We compute the expectations and variances of the terms A_{n11} and A_{n12}. Under H_0 or the local alternatives, we have

$$\begin{aligned}\mathrm{E}(e_i^2) &= \mathrm{E}\left[\left(F(Y_i)-\frac{1}{2}\right)^2\right] + \mathrm{E}[(\boldsymbol{X}_i^T\boldsymbol{\beta}_h)^2] - 2\mathrm{E}\left[(\boldsymbol{X}_i^T\boldsymbol{\beta}_h)\left(F(Y_i)-\frac{1}{2}\right)\right] \\ &= \frac{1}{12} + o(1).\end{aligned}$$

For the expectations of A_{n12} and A_{n13}, we have $\mathrm{E}(A_{n12}) = 0$ and $\mathrm{E}(A_{n13}) = 0$. It is easy to show that $\mathrm{E}(A_{n11}) = \boldsymbol{\beta}_h^T\Sigma^2\boldsymbol{\beta}_h$. Next, we consider the variances of A_{n11} and A_{n12}. Applying Lemma 3, we have

$$\mathrm{Var}(A_{n11}) = O(n^{-2})\mathrm{E}\left[(\boldsymbol{X}_i^T\boldsymbol{X}_j)^2(\boldsymbol{X}_i^T\boldsymbol{\beta}_h)^2(\boldsymbol{X}_j^T\boldsymbol{\beta}_h)^2\right] + O(n^{-1})\left(\mathrm{E}(A_{n11})\right)^2$$

$$
\begin{aligned}
&+ O(n^{-1})\mathrm{E}\left[(\boldsymbol{X}_i^T\boldsymbol{X}_j)(\boldsymbol{X}_i^T\boldsymbol{X}_s)(\boldsymbol{X}_i^T\boldsymbol{\beta}_h)^2(\boldsymbol{X}_j^T\boldsymbol{\beta}_h)(\boldsymbol{X}_s^T\boldsymbol{\beta}_h)\right]\\
&= o(n^{-2}tr(\Sigma^2)).
\end{aligned}
$$

Actually, the following facts hold:

$$
\begin{aligned}
&\mathrm{E}\left((\boldsymbol{X}_i^T\boldsymbol{X}_j)^2(\boldsymbol{X}_i^T\boldsymbol{\beta}_h)^2(\boldsymbol{X}_j^T\boldsymbol{\beta}_h)^2\right)\\
&= \mathrm{E}\left((\boldsymbol{X}_i^T\boldsymbol{\beta}_h)^2\mathrm{E}((\boldsymbol{X}_i^T\boldsymbol{X}_j)^2(\boldsymbol{X}_j^T\boldsymbol{\beta}_h)^2|\boldsymbol{X}_i)\right)\\
&\le C\mathrm{E}\left((\boldsymbol{X}_i^T\boldsymbol{\beta}_h)^2\left(\boldsymbol{\beta}_h^T\Sigma\boldsymbol{\beta}_h\boldsymbol{X}_i^T\Sigma\boldsymbol{X}_i + \boldsymbol{\beta}_h^T\Sigma\boldsymbol{X}_i\boldsymbol{X}_i^T\Sigma\boldsymbol{\beta}_h\right)\right)\\
&= O\left((\boldsymbol{\beta}_h^T\Sigma^3\boldsymbol{\beta}_h)(\boldsymbol{\beta}_h^T\Sigma\boldsymbol{\beta}_h) + tr(\Sigma^2)(\boldsymbol{\beta}_h^T\Sigma\boldsymbol{\beta}_h)^2 + (\boldsymbol{\beta}_h^T\Sigma^2\boldsymbol{\beta}_h)^2\right)\\
&= o(tr(\Sigma^2)),
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathrm{E}\left((\boldsymbol{X}_i^T\boldsymbol{X}_j)(\boldsymbol{X}_i^T\boldsymbol{X}_s)(\boldsymbol{X}_i^T\boldsymbol{\beta}_h)^2(\boldsymbol{X}_j^T\boldsymbol{\beta}_h)(\boldsymbol{X}_s^T\boldsymbol{\beta}_h)\right)\\
&= \mathrm{E}\left((\boldsymbol{X}_i^T\boldsymbol{\beta}_h)^2\mathrm{E}\left((\boldsymbol{X}_j^T\boldsymbol{X}_i)(\boldsymbol{X}_j^T\boldsymbol{\beta}_h)(\boldsymbol{X}_s^T\boldsymbol{X}_i)(\boldsymbol{\beta}_h^T\boldsymbol{X}_s)|\boldsymbol{X}_i\right)\right)\\
&= \mathrm{E}\left(\boldsymbol{X}_i^T\boldsymbol{\beta}_h\boldsymbol{\beta}_h^T\Sigma\boldsymbol{X}_i\right)^2\\
&= O\left((\boldsymbol{\beta}_h^T\Sigma^2\boldsymbol{\beta}_h)^2 + (\boldsymbol{\beta}_h^T\Sigma^3\boldsymbol{\beta}_h)(\boldsymbol{\beta}_h^T\Sigma\boldsymbol{\beta}_h)\right)\\
&= o\left(n^{-1}tr(\Sigma^2)\right).
\end{aligned}
$$

Further observe that

$$
\begin{aligned}
\mathrm{Var}(A_{n12}) &= O(n^{-1})\mathrm{E}[(\boldsymbol{X}_i^T\boldsymbol{X}_j)(\boldsymbol{X}_s^T\boldsymbol{X}_j)(\boldsymbol{X}_i^T\boldsymbol{\beta}_h)(\boldsymbol{X}_s^T\boldsymbol{\beta}_h)e_j^2]\\
&\quad + O(n^{-2})\mathrm{E}[(\boldsymbol{X}_i^T\boldsymbol{X}_j)^2(\boldsymbol{X}_i^T\boldsymbol{\beta}_h)^2e_j^2]\\
&= o(n^{-2}tr(\Sigma^2)).
\end{aligned}
$$

Therefore, A_{n1} has the following form

$$
A_{n1} = \frac{1}{n(n-1)}\sum_{i\neq j}\boldsymbol{X}_i^T\boldsymbol{X}_je_ie_j + o_p(\sqrt{n^{-2}tr(\Sigma^2)}) + \boldsymbol{\beta}_h^T\Sigma^2\boldsymbol{\beta}_h.
$$

Step 2. We establish the asymptotic distribution of A_{n1} by martingale central limit theorem [16].

In this step, our goal is to verify the following limit distribution under H_0.

$$
\frac{12}{n(n-1)}\sum_{j\neq i}\boldsymbol{X}_i^T\boldsymbol{X}_je_ie_j \xrightarrow{d} N\left(0, \frac{2tr(\Sigma^2)}{n^2}\right).
$$

Define $\boldsymbol{\eta}_{in} = 12\sqrt{\frac{2}{n(n-1)}}\sum_{j=1}^{i-1}\boldsymbol{X}_i^T\boldsymbol{X}_je_ie_j$. Let $S_{kn} = \sum_{i=2}^{k}\boldsymbol{\eta}_{in}$, $\mathscr{F}_i = \sigma\{(\boldsymbol{X}_i, e_i)\ i = 1, 2, \cdots, n\}$. Obviously, we have $\mathrm{E}(\boldsymbol{\eta}_{in}|\mathscr{F}_{i-1}) = 0$, which follows that $(S_{kn}, \mathscr{F}_k)$ is a zero-mean martingale sequence. Define $v_{in} = \mathrm{Var}(\boldsymbol{\eta}_{in}|\mathscr{F}_{i-1})$ and $V_n = \sum_{i=2}^n v_{in}$. By martingale central limit theorem, it is sufficient to show that the following two conditions hold under H_0.

$$\frac{V_n}{\mathrm{Var}(S_{nn})} \xrightarrow{P} 1, \tag{A.1}$$

and

$$\sum_{i=2}^{n} (tr(\Sigma^2))^{-1} \mathrm{E}(\boldsymbol{\eta}_{in}^2 I_{\{|\boldsymbol{\eta}_{in}|>\varepsilon\sqrt{tr(\Sigma^2)}\}}|\mathscr{F}_{i-1}) \xrightarrow{P} 0. \tag{A.2}$$

Simple calculation leads to the following results

$$v_{in} = \frac{24}{n(n-1)} \sum_{j=1}^{i-1} e_j^2 \boldsymbol{X}_j^T \Sigma \boldsymbol{X}_j + \frac{24}{n(n-1)} \sum_{1\le j\ne k\le i-1} e_j e_k \boldsymbol{X}_j^T \Sigma \boldsymbol{X}_k.$$

and

$$V_n = \frac{24}{n(n-1)} \sum_{i=2}^{n}\sum_{j=1}^{i-1} e_j^2 \boldsymbol{X}_j^T \Sigma \boldsymbol{X}_j + \frac{24}{n(n-1)} \sum_{i=2}^{n}\sum_{1\le j\ne k\le i-1} e_j e_k \boldsymbol{X}_j^T \Sigma \boldsymbol{X}_k.$$

Since $\mathrm{Var}(S_{nn}) = tr(\Sigma^2)$, then we write

$$\begin{aligned}\frac{V_n}{\mathrm{Var}(S_{nn})} &= \frac{24}{n(n-1)tr(\Sigma^2)} \left(\sum_{i=2}^{n}\sum_{j=1}^{i-1} e_j^2 \boldsymbol{X}_j^T \Sigma \boldsymbol{X}_j + \sum_{i=2}^{n}\sum_{1\le j\ne k\le i-1} e_j e_k \boldsymbol{X}_j^T \Sigma \boldsymbol{X}_k \right) \\ &=: G_{n1} + G_{n2}.\end{aligned}$$

It can be shown that $\mathrm{E}(G_{n1}) = 1$ and $\mathrm{E}(G_{n2}) = 0$ under H_0. Observe the fact that

$$\mathrm{Var}(G_{n1}) = O(n^{-4}) tr^{-2}(\Sigma^2) \sum_{j=1}^{n-1} j^2 \mathrm{E}\left[(\boldsymbol{X}_j^T \Sigma \boldsymbol{X}_j)^2 - tr^2(\Sigma^2) \right].$$

Condition (C1), Lemma 3 and algebra calculation imply that $\mathrm{Var}(G_{n1}) = o(n^{-1})$. Similarly, we can also obtain that

$$\begin{aligned}\mathrm{Var}(G_{n2}) &= O(n^{-4}) tr^{-2}(\Sigma^2) \sum_{j_1<k_1}\sum_{j_2<k_2} (n-k_1)(n-k_2) \\ &\quad \times \mathrm{E}\left[e_{j_1} e_{j_2} e_{k_1} e_{k_2} \boldsymbol{X}_{j_1}^T \Sigma \boldsymbol{X}_{k_1} \boldsymbol{X}_{j_2}^T \Sigma \boldsymbol{X}_{k_2} \right] \\ &= O(n^{-4}) \sum_{k=1}^{n} (n-k)^2 (k-1) \frac{tr(\Sigma^4)}{tr^2(\Sigma^2)}.\end{aligned}$$

Thus $\mathrm{Var}(G_{n2}) = o(1)$. Markov inequality yields that $G_{n1} \xrightarrow{P} 1$ and $G_{n2} \xrightarrow{P} 0$. Up to now, (A.1) has been verified.

Observe that

$$\sum_{i=2}^{n} \mathrm{E}(\boldsymbol{\eta}_{in}^4) = O(n^{-4}) \sum_{i=2}^{n} \mathrm{E} \left(\sum_{j=1}^{i-1} e_i e_j \boldsymbol{X}_i^T \boldsymbol{X}_j \right)^4$$
$$= O(n^{-4}) \sum_{i=2}^{n} \sum_{s \neq t} \mathrm{E}(\boldsymbol{X}_i^T \boldsymbol{X}_s \boldsymbol{X}_i^T \boldsymbol{X}_s \boldsymbol{X}_i^T \boldsymbol{X}_t \boldsymbol{X}_i^T \boldsymbol{X}_t) \qquad \text{(A.3)}$$
$$+ O(n^{-4}) \sum_{i=2}^{n} \sum_{j=1}^{i-1} \mathrm{E}(\boldsymbol{X}_i^T \boldsymbol{X}_j)^4.$$

For the first term in (A.3), algebra calculation based on Lemma 3 shows that

$$O(n^{-4}) \sum_{i=2}^{n} \sum_{s \neq t} \mathrm{E}(\boldsymbol{X}_i^T \boldsymbol{X}_s \boldsymbol{X}_i^T \boldsymbol{X}_s \boldsymbol{X}_i^T \boldsymbol{X}_t \boldsymbol{X}_i^T \boldsymbol{X}_t) = O(n^{-1} tr^2(\Sigma^2)).$$

Define $\Gamma^T \Gamma = (v_{ij})$. For the second term in (A.3), we have the following facts

$$O(n^{-4}) \sum_{i=2}^{n} \sum_{j=1}^{i-1} \mathrm{E}(\boldsymbol{X}_i^T \boldsymbol{X}_j)^4 = O(n^{-4}) \sum_{i \neq j} \mathrm{E} \left(\sum_{k,l=1}^{p} v_{kl} Z_{ik} Zjl \right)^4$$
$$= O(n^{-2}) \left(\sum_{k,l=1}^{p} v_{kl}^4 \mathrm{E}(Z_{ik}^4) \mathrm{E}(Z_{jl}^4) + \sum_{k \neq l} \sum_{s \neq t} v_{kl}^2 v_{st}^2 \mathrm{E}(Z_{ik}^2) \mathrm{E}(Z_{is}^2) \mathrm{E}(Z_{jl}^2) \mathrm{E}(Z_{jt}^2) \right.$$
$$+ \sum_{k \neq l} \sum_{s \neq t} v_{kl} v_{kt} v_{sl} v_{st} \mathrm{E}(Z_{ik}^2) \mathrm{E}(Z_{is}^2) \mathrm{E}(Z_{jl}^2) \mathrm{E}(Z_{jt}^2)$$
$$\left. + \sum_{k=l} \sum_{s \neq t} v_{ks}^2 v_{kl}^2 \mathrm{E}(Z_{ik}^4) \mathrm{E}(Z_{js}^2) \mathrm{E}(Z_{jt}^2) \right).$$

Note that the following inequalities holds

$$\sum_{k,l=1}^{p} v_{kl}^4 \leq \left(\sum_{k,l}^{p} v_{kl}^2 \right)^2 = (tr(\Sigma^2))^2,$$
$$\sum_{k=l} \sum_{s \neq t} v_{ks}^2 v_{kl}^2 \leq \left(\sum_{k,l}^{p} v_{kl}^2 \right)^2 = (tr(\Sigma^2))^2,$$
$$\sum_{k \neq l} \sum_{s \neq t} v_{kl} v_{kt} v_{sl} v_{st} \leq tr(\Sigma^4).$$

Combining condition (C1), we obtain $\sum_{i=2}^{n} \mathrm{E}(\boldsymbol{\eta}_{in}^4) = o((tr(\Sigma^2))^2)$. In other words, for any $\varepsilon > 0$, we have

$$E \left\{ \sum_{i=2}^{n} \frac{\mathrm{E}(\boldsymbol{\eta}_{in}^4 | \mathscr{F}_{i-1})}{\varepsilon^2 (tr(\Sigma^2))^2} \right\} = o(1).$$

Moreover, considering the following fact

$$\mathrm{E}\left(\boldsymbol{\eta}_{in}^2 I_{\{|\boldsymbol{\eta}_{in}|>\varepsilon\sqrt{tr(\Sigma^2)}\}}|\mathscr{F}_{i-1}\right) \le \varepsilon^{-2} tr(\Sigma^2)^{-1}\mathrm{E}(\boldsymbol{\eta}_{in}^4|\mathscr{F}_{i-1}),$$

then we can conclude (A.2) holds.

Step 3. It suffices to show that the variances of A_{n3} and A_{n2} are $o(n^{-2}tr(\Sigma^2))$ under H_0 to complete the proof.

We note that $\boldsymbol{X}$ is uncorrelated with Y under H_0. It follows that $\mathrm{E}(A_{n2}) = \mathrm{E}(A_{n3}) = 0$ obviously. Besides,

$$\begin{aligned}
\mathrm{E}(A_{n2}^2) &= O(n^{-4})\sum_{i\neq j}\sum_{k\neq l}\mathrm{E}(\boldsymbol{X}_i^T\boldsymbol{X}_j\boldsymbol{X}_k^T\boldsymbol{X}_l)\mathrm{E}\Big[(F_n(Y_i)-F(Y_i))(F_n(Y_k)-F(Y_k)) \\
&\quad \times \Big(F(Y_j)-\frac{1}{2}\Big)\Big(F(Y_l)-\frac{1}{2}\Big)\Big] \\
&= O(n^{-4})\sum_{i\neq j}\mathrm{E}(\boldsymbol{X}_i^T\boldsymbol{X}_j)^2\mathrm{E}\Big[(F_n(Y_i)-F(Y_i))^2\Big(F(Y_j)-\frac{1}{2}\Big)^2\Big] \\
&= o(n^{-2}tr(\Sigma^2)),
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}(A_{n3}^2) &= O(n^{-4})\sum_{i\neq j}\sum_{k\neq l}\mathrm{E}(\boldsymbol{X}_i^T\boldsymbol{X}_j\boldsymbol{X}_k^T\boldsymbol{X}_l)\mathrm{E}\Big[(F_n(Y_i)-F(Y_i))(F_n(Y_j)-F(Y_j)) \\
&\quad \times (F_n(Y_k)-F(Y_k))(F_n(Y_l)-F(Y_l))\Big] \\
&= O(n^{-4})\sum_{i\neq j}\mathrm{E}(\boldsymbol{X}_i^T\boldsymbol{X}_j)^2\mathrm{E}\Big[(F_n(Y_i)-F(Y_i))^2(F_n(Y_j)-F(Y_j))^2\Big] \\
&= o(n^{-2}tr(\Sigma^2)).
\end{aligned}$$

Summarizing the above results, we conclude that Theorem 1 holds.

Proofs for Theorem 2

According to the proof of Theorem 1, we have already demonstrated the fact that

$$\begin{aligned}
\frac{nS_n}{12\sqrt{tr(\Sigma^2)}} &= \frac{n}{\sqrt{tr(\Sigma^2)}}\Big(\frac{1}{n(n-1)}\sum_{i\neq j}\boldsymbol{X}_i^T\boldsymbol{X}_j e_i e_j + o_p(\sqrt{n^{-2}tr(\Sigma^2)}) \\
&\quad + \boldsymbol{\beta}_h^T\Sigma^2\boldsymbol{\beta}_h + A_{n2} + A_{n3}\Big),
\end{aligned}$$

under H_0 or the local alternatives.

To verify S_n tend to infinite as $n\to\infty$, it suffices to show the orders of A_{n2} and A_{n3} are controlled by the term $\boldsymbol{\beta}_h^T\Sigma^2\boldsymbol{\beta}_h$ under the conditions of Theorem 2.

Step 1 For the term A_{n3}, we have

$$|A_{n3}| \leq \frac{1}{n(n-1)} \sum_{i \neq j} \left| \boldsymbol{X}_i^T \boldsymbol{X}_j \right| \max_{1 \leq i \leq n} |F_n(Y_i) - F(Y_i)|^2 .$$

Since $\mathrm{E}(\boldsymbol{X}_i^T \boldsymbol{X}_j)^2 = tr(\Sigma^2)$, we show that A_{n3} is of order $O_p(n^{-1} \log n \sqrt{tr(\Sigma^2)})$.

Step 2 For the term A_{n2}, we rewrite it as below

$$\begin{aligned}
\frac{1}{2} A_{n2} = & \frac{1}{n(n-1)} \sum_{i \neq j} \boldsymbol{X}_i^T \boldsymbol{X}_j (F_n(Y_i) - F(Y_i)) \boldsymbol{X}_j^T \beta_h \\
& + \frac{1}{n(n-1)} \sum_{i \neq j} \boldsymbol{X}_i^T \boldsymbol{X}_j (F_n(Y_i) - F(Y_i)) e_j \\
=: & D_{n1} + D_{n2}.
\end{aligned}$$

Note that

$$|D_{n1}| \leq \max_{1 \leq i \leq n} |F_n(Y_i) - F(Y_i)| \frac{1}{n(n-1)} \sum_{i \neq j} \left| \boldsymbol{X}_i^T \boldsymbol{X}_j \boldsymbol{X}_j^T \beta_h \right| ,$$

and

$$\begin{aligned}
& \mathrm{E}\left((\boldsymbol{X}_i \boldsymbol{X}_j)^2 (\boldsymbol{X}_j^T \boldsymbol{\beta}_h)^2 \right) = \mathrm{E}\left(\boldsymbol{Z}_j^T \Gamma^T \boldsymbol{\beta}_h \boldsymbol{Z}_i^T \Gamma^T \Gamma \boldsymbol{Z}_j \right)^2 \\
& = \mathrm{E}\left(\mathrm{E}\left(\boldsymbol{Z}_j^T \Gamma^T \boldsymbol{\beta}_h \boldsymbol{Z}_i^T \Gamma^T \Gamma \boldsymbol{Z}_j \right)^2 | \boldsymbol{Z}_i \right) \\
& = O(1) \mathrm{E}\left(tr(\Gamma^T \boldsymbol{\beta}_h \boldsymbol{Z}_i^T \Gamma^T \Gamma \Gamma^T \boldsymbol{\beta}_h \boldsymbol{Z}_i^T \Gamma^T \Gamma) \right) \\
& \quad + O(1) \mathrm{E}\left(tr^2(\Gamma^T \boldsymbol{\beta}_h \boldsymbol{Z}_i^T \Gamma^T \Gamma) \right) \\
& = O(\boldsymbol{\beta}_h^T \Sigma^3 \boldsymbol{\beta}_h).
\end{aligned}$$

Therefore, we readily obtain D_{n1} is $o_p(n^{-1} (\log n)^{1/2} \sqrt{tr(\Sigma^2)})$.

For D_{n2}, we write

$$\begin{aligned}
D_{n2} = & \frac{1}{n^2(n-1)} \sum_{i \neq j} \sum_{k=1}^{n} \boldsymbol{X}_i^T \boldsymbol{X}_j e_j \left(I_{\{Y_k \leq Y_i\}} - F(Y_i) \right) \\
= & \frac{1}{n^2(n-1)} \sum_{i \neq j} \sum_{k=i} \boldsymbol{X}_i^T \boldsymbol{X}_j e_j \left(I_{\{Y_k \leq Y_i\}} - F(Y_i) \right) \\
& + \frac{1}{n^2(n-1)} \sum_{i \neq j} \sum_{k=j} \boldsymbol{X}_i^T \boldsymbol{X}_j e_j \left(I_{\{Y_k \leq Y_i\}} - F(Y_i) \right) \\
& + \frac{1}{n^2(n-1)} \sum_{i \neq j \neq k} \boldsymbol{X}_i^T \boldsymbol{X}_j e_j \left(I_{\{Y_k \leq Y_i\}} - F(Y_i) \right) \\
= & D_{n21} + D_{n22} + D_{n23}.
\end{aligned}$$

By Cauchy-Schwarz inequality, we have

$$\begin{aligned} \mathrm{E}\left|D_{n21}\right| &\leq Cn^{-1}\mathrm{E}\left|\boldsymbol{X}_i^T\boldsymbol{X}_j e_j\right| \\ &\leq Cn^{-1}\mathrm{E}^{1/2}\left(\boldsymbol{X}_i^T\boldsymbol{X}_j\boldsymbol{X}_j^T\boldsymbol{X}_i\right)\mathrm{E}^{1/2}\left(e_j^2\right) \\ &= Cn^{-1}\sqrt{tr(\Sigma^2)}, \end{aligned}$$

which implies that D_{n21} is of order $O_p(n^{-1}\sqrt{tr(\Sigma^2)})$. Analogously, D_{n22} is also of order $O_p(n^{-1}\sqrt{tr(\Sigma^2)})$. Since e_j is bounded in probability under the local alternative, we derive that

$$\begin{aligned} \mathrm{E}(D_{n23}^2) &= \frac{1}{n^4(n-1)^2}\sum_{i\neq j\neq k}\sum_{s\neq t\neq l}\mathrm{E}\left(\boldsymbol{X}_i^T\boldsymbol{X}_j\boldsymbol{X}_s^T\boldsymbol{X}_t\left(I_{\{Y_k\leq Y_i\}}-F(Y_i)\right)\right. \\ &\quad\left.\times\left(I_{\{Y_l\leq Y_s\}}-F(Y_s)\right)e_j e_t\right) \\ &= O(n^{-2})\mathrm{E}\left(\boldsymbol{X}_i^T\boldsymbol{X}_j\boldsymbol{X}_s^T\boldsymbol{X}_j e_j^2+(\boldsymbol{X}_i^T\boldsymbol{X}_j)^2 e_j^2\right) \\ &= O(n^{-2}tr(\Sigma^2)), \end{aligned}$$

by noticing the fact that

$$\begin{aligned} &\mathrm{E}\left(\boldsymbol{X}_i^T\boldsymbol{X}_j\left(I_{\{Y_k\leq Y_i\}}-F(Y_i)\right)e_j\right)=0, \\ &\mathrm{E}\left(\boldsymbol{X}_i^T\boldsymbol{X}_j\left(I_{\{Y_k\leq Y_i\}}-F(Y_i)\right)e_j|\boldsymbol{X}_i,Y_i,\boldsymbol{X}_j,e_j\right)=0. \end{aligned}$$

Hence, we verify that D_{n2} is of order $O_p(n^{-1}tr(\Sigma^2))$. Then we obtain $A_{n2}=o_p(n^{-1}(\log n)^{1/2}\sqrt{tr(\Sigma^2)})$.

So far, we conclude that $|A_{ni}|^{-1}\boldsymbol{\beta}_h^T\Sigma^2\boldsymbol{\beta}_h\to\infty$ $i=2,3$ holds under the conditions that $\boldsymbol{\beta}_h^T\Sigma^2\boldsymbol{\beta}_h\geq Cn^{-1+\delta}\sqrt{tr(\Sigma^2)}$ for any small positive number δ. Lastly, applying similar techniques to those used in the above proofs, we are able to show $n^{-1}(n-1)^{-1}\sum_{i\neq j}\boldsymbol{X}_i^T\boldsymbol{X}_j e_i e_j$ is $O_p(n^{-1}tr(\Sigma^2))$, which completes the proof.

Proofs for Theorem 3

According to [24], $\min\{1,\cdot\}$ can be ignored in the proof. Thus, it is sufficient to show that

$$\Pr((1-\log\gamma_{\min})\inf_{\gamma\in(\gamma_{\min},1)}Q(\gamma)\leq\alpha)\leq\alpha.$$

Define $\pi(u)=B^{-1}\sum_{b=1}^{B}I_{\{p_b\leq u\}}$. Then the two events $\{Q(\gamma)\leq\alpha\}$ and $\{\pi(\alpha\gamma)\geq\gamma\}$ are equivalent. In fact, p_b follows a uniform distribution $U(0,1)$

under H_0. Applying Markov inequality, we obtain that

$$\begin{aligned}
&\Pr(Q(\gamma) \leq \alpha | H_0) = \Pr(\pi(\alpha\gamma) \geq \gamma | H_0) \\
&= \Pr\left(B^{-1} \sum_{b=1}^{B} I_{\{p_b \leq \alpha\gamma\}} \geq \gamma | H_0\right) \\
&\leq (\gamma B)^{-1} \sum_{b=1}^{B} \Pr(p_b \leq \alpha\gamma | H_0) \\
&= \alpha.
\end{aligned}$$

Using Markov inequality again, we derive

$$\begin{aligned}
&\Pr\left(\inf_{\gamma \in (\gamma_{\min}, 1)} Q(\gamma) \leq \alpha\right) = \mathrm{E}\left(\inf_{\gamma \in (\gamma_{\min}, 1)} I_{\{Q(\gamma) \leq \alpha\}}\right) \\
&= \mathrm{E}\left(\sup_{\gamma \in (\gamma_{\min}, 1)} I_{\{\pi(\alpha\gamma) \geq \gamma\}}\right) = \mathrm{E}\left(\sup_{\gamma \in (\gamma_{\min}, 1)} I_{\{B^{-1} \sum_{b=1}^{B} I\{p_b \leq \alpha\gamma\} \geq \gamma\}}\right) \\
&\leq \mathrm{E}\left(\sup_{\gamma \in (\gamma_{\min}, 1)} \frac{1}{\gamma} B^{-1} \sum_{b=1}^{B} I_{\{p_b \leq \alpha\gamma\}}\right) \\
&= \mathrm{E}\left(\sup_{\gamma \in (\gamma_{\min}, 1)} \frac{1}{\gamma} I_{\{p_b \leq \alpha\gamma\}}\right).
\end{aligned}$$

Consider the following fact,

$$\sup_{\gamma \in (\gamma_{\min}, 1)} \frac{1}{\gamma} I_{\{p_b \leq \alpha\gamma\}} = \begin{cases} 0 & p_b \geq \alpha, \\ \alpha p_b^{-1} & \alpha\gamma_{\min} \leq p_b < \alpha, \\ \gamma_{\min}^{-1} & p_b < \alpha\gamma_{\min}. \end{cases}$$

Therefore, we can easily calculate the expectation under H_0

$$\begin{aligned}
&\mathrm{E}\left(\sup_{\gamma \in (\gamma_{\min}, 1)} \frac{1}{\gamma} I_{\{p_b \leq \alpha\gamma\}} \Big| H_0\right) = \int_0^{\alpha\gamma_{\min}} \gamma_{\min}^{-1} dx + \int_{\alpha\gamma_{\min}}^{\alpha} \frac{\alpha}{x} dx \\
&= \alpha(1 - \log(\gamma_{\min})).
\end{aligned}$$

Summarizing the above results, $\Pr\left(\inf_{\gamma \in (\gamma_{\min}, 1)} Q(\gamma) \leq \alpha | H_0\right) \leq \alpha(1 - \log(\gamma_{\min}))$ holds to complete the proof.

Proofs for Theorem 4

Without loss of generality, we assume that $\sigma_{jj}^* = 1$ for any $j \in S := \{q + 1, \cdots, p\}$. To prove Theorem 4, we fist introduce the following lemma.

Lemma 4. *Suppose $\boldsymbol{X}_{ib}^*$ follows the elliptical distribution and Condition (C4) holds, we have the following facts*

$$E\left(\sum_{j\in S}(X_{ij}^{*2}-1)\right)^4 = O(tr^2(\Sigma_{b|a}^2)), \tag{A.4}$$

$$E\left(\sum_{j\in S}X_{i_1j}^*X_{i_2j}^*\right)^4 = O(tr^2(\Sigma_{b|a}^2)). \tag{A.5}$$

Proof of Lemma 4. Under Condition (C4), one directly proves (A.4) by some algebras in [2] and thus omit here. Denote $\Sigma_{b|a} := \left(\sigma_{ij}^*\right)$. By the definition of elliptical distributions, we have

$$\mathrm{E}(X_{ij_1}^*X_{ij_2}^*X_{ij_3}^*X_{ij_4}^*) = \sigma_{j_1j_2}^*\sigma_{j_3j_4}^* + \sigma_{j_1j_3}^*\sigma_{j_2j_4}^* + \sigma_{j_1j_4}^*\sigma_{j_2j_3}^* + C\sum_{k\in S}\sigma_{j_1k}^*\sigma_{j_2k}^*\sigma_{j_3k}^*\sigma_{j_4k}^*.$$

Thus, by Cauchy-Schwarz inequality,

$$\begin{aligned}
&\sum_{j_1j_2j_3j_4\in S}\left(\mathrm{E}(X_{ij_1}^*X_{ij_2}^*X_{ij_3}^*X_{ij_4}^*)\right)^2 \\
&\le 2\sum_{j_1j_2j_3j_4\in S}\left(\sigma_{j_1j_2}^*\sigma_{j_3j_4}^* + \sigma_{j_1j_3}^*\sigma_{j_2j_4}^* + \sigma_{j_1j_4}^*\sigma_{j_2j_3}^*\right)^2 \\
&\quad + 2C^2\sum_{j_1j_2j_3j_4\in S}\left(\sum_{k\in S}\sigma_{j_1k}^*\sigma_{j_2k}^*\sigma_{j_3k}^*\sigma_{j_4k}^*\right)^2 \\
&= 2\left(3tr^2(\Sigma_{b|a}^2) + 4tr(\Sigma_{b|a}^4)\right) + 2C^2\sum_{k_1,k_2}\left(\sum_j \sigma_{jk_1}^*\sigma_{jk_2}^*\right)^4 \\
&\le 2\left(3tr^2(\Sigma_{b|a}^2) + 4tr(\Sigma_{b|a}^4)\right) + 2C^2\left(\sum_{k_1,k_2}\left(\sum_j \sigma_{jk_1}^*\sigma_{jk_2}^*\right)^2\right)^2 \\
&\le 2\left(3tr^2(\Sigma_{b|a}^2) + 4tr(\Sigma_{b|a}^4)\right) + 2C^2tr^2(\Sigma_{b|a}^2) \\
&= O(tr^2(\Sigma_{b|a}^2)).
\end{aligned}$$

By noticing the following facts to complete the proof for (A.5),

$$\mathrm{E}\left(\sum_{j\in S}X_{i_1j}^*X_{i_2j}^*\right)^4 = \sum_{j_1,j_2,j_3,j_4}\mathrm{E}^2(X_{i_1j_1}^*X_{i_1j_2}^*X_{i_1j_3}^*X_{i_1j_4}^*). \tag{A.6}$$

Now we present the proof of Theorem 4.

Step 1 We first define $\tilde{T}_n$ as follows,

$$\tilde{T}_n = \frac{1}{n}(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y}))^T \mathbb{X}_b^* \mathbb{X}_b^{*T} (\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y})) - \frac{p-q}{n}(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y}))^T (\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y})).$$

Next, we would verify that $p^{-\frac{1}{2}}(T_n - \tilde{T}_n)$ is $o_p(1)$. Notice that

$$\begin{aligned} & T_n - \tilde{T}_n \\ &= \left\{ \frac{1}{n}\hat{e}^T \mathbb{X}_b \mathbb{X}_b^T \hat{e} - \frac{1}{n}(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y}))^T \mathbb{X}_b^* \mathbb{X}_b^{*T} (\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y})) \right. \\ &\quad \left. + \frac{p-q}{n}(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y}))^T \mathbb{P}_a (\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y})) \right\} \\ &\quad - \left\{ \frac{\hat{\sigma}^2}{n} tr(\mathbb{X}_b^* \mathbb{X}_b^{*T} (\mathbb{I} - \mathbb{P}_a)) \right. \\ &\quad \left. - \frac{p-q}{n}(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y}))^T (\mathbb{I} - \mathbb{P}_a)(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y})) \right\} \\ &= J_{n1} + J_{n2}. \end{aligned} \tag{A.7}$$

We now calculate the order of the maximum eigenvalue of matrix $\mathbb{H} := \mathbb{X}_b^* \mathbb{X}_b^{*T} - (p-q)\mathbb{I} = (h_{ij})$.

Note that

$$\mathrm{E}(tr(\mathbb{H}^4)) = \mathrm{E}\left(\sum_{i_1, i_2, i_3, i_4} h_{i_1 i_2} h_{i_2 i_3} h_{i_3 i_4} h_{i_4 i_1} \right).$$

Further note that

$$\begin{aligned} & \mathrm{E}\left(\sum_{i_1 \neq i_2 \neq i_3 \neq i_4} h_{i_1 i_2} h_{i_2 i_3} h_{i_3 i_4} h_{i_4 i_1} \right) \\ &= \mathrm{E}\left(\sum_{i_1 \neq i_2 \neq i_3 \neq i_4} \sum_{j_1 \in S} X^*_{i_1 j_1} X^*_{i_2 j_1} \sum_{j_2 \in S} X^*_{i_2 j_2} X^*_{i_3 j_2} \sum_{j_3 \in S} X^*_{i_3 j_3} X^*_{i_4 j_3} \sum_{j_4 \in S} X^*_{i_4 j_4} X^*_{i_1 j_4} \right) \\ &= \sum_{i_1 \neq i_2 \neq i_3 \neq i_4} \sum_{j_1 j_2 j_3 j_4 \in S} \sigma^*_{j_1 j_2} \sigma^*_{j_2 j_3} \sigma^*_{j_3 j_4} \sigma^*_{j_4 j_1} \\ &= O(n^4 tr(\Sigma^4_{b|a})). \end{aligned}$$

Denote Λ as the complement set of $\{i_1 \neq i_2 \neq i_3 \neq i_4\}$. Using Cauchy-Schwarz inequality, we get

$$\begin{aligned} & \mathrm{E}\left(\sum_{\Lambda} h_{i_1 i_2} h_{i_2 i_3} h_{i_3 i_4} h_{i_4 i_1} \right) \\ &\leq \frac{1}{4} n^3 \left[\mathrm{E}(h^4_{i_1 i_2}) + \mathrm{E}(h^4_{i_2 i_3}) + \mathrm{E}(h^4_{i_3 i_4}) + \mathrm{E}(h^4_{i_4 i_1}) \right]. \end{aligned}$$

According to (A.6), when $i_1 \neq i_2$, we have

$$\begin{aligned} \mathrm{E}(h_{i_1 i_2}^4) &= \sum_{j_1 j_2 j_3 j_4 \in S} \mathrm{E}(X_{i_1 j_1}^* X_{i_1 j_2}^* X_{i_1 j_3}^* X_{i_1 j_4}^* X_{i_2 j_1}^* X_{i_2 j_2}^* X_{i_2 j_3}^* X_{i_2 j_4}^*) \\ &= \sum_{j_1 j_2 j_3 j_4 \in S} \mathrm{E}^2(X_{i_1 j_1}^* X_{i_1 j_2}^* X_{i_1 j_3}^* X_{i_1 j_4}^*) \\ &= O(tr^2(\Sigma_{b|a}^2)). \end{aligned}$$

According to (A.4), when $i_1 = i_2$, we have

$$\mathrm{E}(h_{i_1 i_1}^4) = \mathrm{E}\left(\sum_{j_1 \in S} X_{i_1 j_1}^{*2} - (p-q)\right)^4 = O(tr^2(\Sigma_{b|a}^2)).$$

Therefore, $\lambda_{\max}(\mathbb{H}) = O_p(n^{\frac{3}{4}} p^{\frac{1}{2}} + np^{\frac{1}{4}})$, which is implied by $\lambda_{\max}^4(\mathbb{H}) \leq tr(\mathbb{H}^4)$.

For the first term on the right hand side of (A.7), we have

$$\begin{aligned} J_{n1} &= n^{-1}(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y}))^T(\mathbb{I} - \mathbb{P}_a)\mathbb{X}_b^* \mathbb{X}_b^{*T}(\mathbb{I} - \mathbb{P}_a)(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y})) \\ &\quad - n^{-1}(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y}))^T \mathbb{X}_b^* \mathbb{X}_b^{*T}(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y})) \\ &\quad + n^{-1}(p-q)(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y}))^T \mathbb{P}_a(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y})) \\ &= n^{-1}(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y}))^T \mathbb{P}_a \mathbb{H} \mathbb{P}_a(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y})) \\ &\quad - 2n^{-1}(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y}))^T \mathbb{P}_a \mathbb{H}(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y})). \end{aligned}$$

Denote $\mathbb{P}_a =: (p_{ij})$. Under Condition (C3), we have

$$\begin{aligned} \mathrm{E}\left(\boldsymbol{e}^T \mathbb{P}_a \boldsymbol{e}\right) &= \mathrm{E}\left(\sum_{i,j} p_{ij} e_i e_j\right) = \mathrm{E}\left(\mathrm{E}\left(\sum_{i,j} p_{ij} e_i e_j | \mathbb{X}\right)\right) \\ &= \sigma^2 \mathrm{E}\left(tr(\mathbb{P}_a)\right) = O(q). \end{aligned}$$

Considering the fact that $\boldsymbol{e}^T \mathbb{P}_a \mathbb{H} \mathbb{P}_a \boldsymbol{e} \leq \lambda_{\max}(\mathbb{H}) \boldsymbol{e}^T \mathbb{P}_a \boldsymbol{e}$, then

$$\boldsymbol{e}^T \mathbb{P}_a \mathbb{H} \mathbb{P}_a \boldsymbol{e} = O_p\left(n^{\frac{3}{4}} q p^{\frac{1}{2}} + nqp^{\frac{1}{4}}\right).$$

Similarly, we could also obtain $\boldsymbol{e}^T \mathbb{P}_a \mathbb{H} \boldsymbol{e} = O_p\left(n^{\frac{3}{4}} q p^{\frac{1}{2}} + nqp^{\frac{1}{4}}\right)$.

We write

$$\begin{aligned} &(F_n(\mathbf{Y}) - F(\mathbf{Y}))^T(F_n(\mathbf{Y}) - F(\mathbf{Y})) = \sum_{i=1}^n (F_n(Y_i) - F(Y_i))^2 \\ &= \frac{1}{n^2} \sum_{i,j,k} \left(I_{\{Y_j \leq Y_i\}} - F(Y_i)\right)\left(I_{\{Y_k \leq Y_i\}} - F(Y_i)\right). \end{aligned}$$

Simple algebra leads to $\mathrm{E}(F_n(\mathbf{Y}) - F(\mathbf{Y}))^T(F_n(\mathbf{Y}) - F(\mathbf{Y})) = O(1)$, which is followed by $(F_n(\mathbf{Y}) - F(\mathbf{Y}))^T(F_n(\mathbf{Y}) - F(\mathbf{Y})) = O_p(1)$. Further observe that

$$\begin{aligned}
&(F_n(\mathbf{Y}) - F(\mathbf{Y}))^T \mathbb{P}_a \mathbb{H} \mathbb{P}_a (F_n(\mathbf{Y}) - F(\mathbf{Y})) \\
&\le (F_n(\mathbf{Y}) - F(\mathbf{Y}))^T (F_n(\mathbf{Y}) - F(\mathbf{Y})) \lambda_{\max}(\mathbb{H}) \\
&= O_p\left(n^{\frac{3}{4}} p^{\frac{1}{2}} + n p^{\frac{1}{4}}\right).
\end{aligned}$$

Analogously, we derive the order of $(F_n(\mathbf{Y}) - F(\mathbf{Y}))^T \mathbb{P}_a \mathbb{H} (F_n(\mathbf{Y}) - F(\mathbf{Y}))$ is also $O_p\left(n^{\frac{3}{4}} p^{\frac{1}{2}} + n p^{\frac{1}{4}}\right)$. By Cauchy-Schwarz inequality,

$$\begin{aligned}
&\left|\boldsymbol{e}^T \mathbb{P}_a \mathbb{H} \mathbb{P}_a (F_n(\mathbf{Y}) - F(\mathbf{Y}))\right| \\
&\le \left(\boldsymbol{e}^T \mathbb{P}_a \mathbb{H} \mathbb{P}_a \boldsymbol{e}\right)^{1/2} \left((F_n(\mathbf{Y}) - F(\mathbf{Y}))^T \mathbb{P}_a \mathbb{H} \mathbb{P}_a (F_n(\mathbf{Y}) - F(\mathbf{Y}))\right)^{1/2} \\
&= O_p\left(n^{\frac{3}{4}} q^{\frac{1}{2}} p^{\frac{1}{2}} + n q^{\frac{1}{2}} p^{\frac{1}{4}}\right).
\end{aligned}$$

Other cross terms can be similarly controlled. Thus, when Condition (C2) holds, the order of J_{n1} is $o_p(p^{1/2})$.

For the second term on the right hand side of (A.7), we have

$$\begin{aligned}
J_{n2} &= \frac{(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y}))^T (\mathbb{I} - \mathbb{P}_a)(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y}))}{n(n-q)} tr(\mathbb{X}_b^* \mathbb{X}_b^{*^T} (\mathbb{I} - \mathbb{P}_a)) \\
&\quad - \frac{p-q}{n} (\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y}))^T (\mathbb{I} - \mathbb{P}_a)(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y})) \\
&= \frac{(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y}))^T (\mathbb{I} - \mathbb{P}_a)(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y}))}{n-q} (p-q) \\
&\quad \times \left(\frac{1}{n(p-q)} tr(\mathbb{X}_b^* \mathbb{X}_b^{*T}) - \frac{n-q}{n}\right) \\
&\quad - \frac{(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y}))^T (\mathbb{I} - \mathbb{P}_a)(\boldsymbol{e} + F_n(\mathbf{Y}) - F(\mathbf{Y}))}{n(n-q)} tr(\mathbb{X}_b^* \mathbb{X}_b^{*T} \mathbb{P}_a) \\
&= J_{n21} + J_{n22}.
\end{aligned}$$

According to Lemma 3, one easily shows that

$$\begin{aligned}
&\frac{1}{n(p-q)} tr(\mathbb{X}_b^* \mathbb{X}_b^{*T}) - 1 = \frac{1}{n(p-q)} \sum_{i=1}^n \left(\boldsymbol{X}_{ib}^{*T} \boldsymbol{X}_{ib}^* - (p-q)\right) \\
&= O_p(n^{-1/2} p^{-1/2}).
\end{aligned}$$

Notice that

$$\begin{aligned}
&\frac{(F_n(\mathbf{Y}) - F(\mathbf{Y}))^T (\mathbb{I} - \mathbb{P}_a)(F_n(\mathbf{Y}) - F(\mathbf{Y}))}{n-q} \\
&\le \frac{(F_n(\mathbf{Y}) - F(\mathbf{Y}))^T (F_n(\mathbf{Y}) - F(\mathbf{Y}))}{n-q} \lambda_{\max}(\mathbb{I} - \mathbb{P}_a)
\end{aligned}$$

$$= O_p(n^{-1});$$
$$\frac{\boldsymbol{e}^T(\mathbb{I}-\mathbb{P}_a)\boldsymbol{e}}{n-q} \leq \frac{\boldsymbol{e}^T\boldsymbol{e}}{n-q}\lambda_{\max}(\mathbb{I}-\mathbb{P}_a) = O_p(1).$$

Then, Cauchy-Schwarz inequality imples that $n^{-1}\boldsymbol{e}^T(\mathbb{I}-\mathbb{P}_a)(F_n(\mathbf{Y})-F(\mathbf{Y}))$ is of order $O_p(n^{-1/2})$. Hence, we verify that $J_{n21} = o_p(p^{1/2})$. Further observe that $tr(\mathbb{X}_b^*\mathbb{X}_b^{*T}\mathbb{P}_a) \leq \lambda_{\max}(\mathbb{X}_b^*\mathbb{X}_b^{*T})tr(\mathbb{P}_a)$, then we could derive that $J_{n22} = o_p(p^{1/2})$, since $p^{1/2}qn^{-1} \to 0$.

Up to now, we have verified that

$$\frac{T_n - \tilde{T}_n}{(2\sigma^4 tr(\Sigma_{b|a}^2))^{\frac{1}{2}}} \xrightarrow{P} 0.$$

Step 2 Then, it suffices to show

$$\frac{\tilde{T}_n}{(2\sigma^4 tr(\Sigma_{b|a}^2))^{\frac{1}{2}}} \xrightarrow{d} N(0,1). \tag{A.8}$$

For $\tilde{T}_n$, we have

$$\begin{aligned}\tilde{T}_n &= \frac{1}{n}\boldsymbol{e}^T\mathbb{H}\boldsymbol{e}+\frac{1}{n}(F_n(\mathbf{Y})-F(\mathbf{Y}))^T\mathbb{H}(F_n(\mathbf{Y})-F(\mathbf{Y}))+\frac{2}{n}\boldsymbol{e}^T\mathbb{H}(F_n(\mathbf{Y})-F(\mathbf{Y}))\\ &= K_{n1}+K_{n2}+K_{n3}.\end{aligned}$$

where

$$\begin{aligned}K_{n1} &= \frac{1}{n}\boldsymbol{e}^T\mathbb{X}_b^*\mathbb{X}_b^{*T}\boldsymbol{e} - \frac{p-q}{n}\boldsymbol{e}^T\boldsymbol{e}\\ &= \frac{1}{n}\sum_{i_1\neq i_2}\sum_{j\in S} e_{i_1}e_{i_2}X_{i_1j}^*X_{i_2j}^* + \frac{1}{n}\sum_{i=1}^{n}\sum_{j\in S} e_i^2(X_{ij}^{*2}-1)\\ &= K_{n11}+K_{n12}.\end{aligned}$$

It is clear that $\mathrm{E}(K_{n12}) = 0$. Recall that $\mathrm{E}(e_i^2|\boldsymbol{X}_i) = \sigma^2$ and $\mathrm{E}(e_i^4|\boldsymbol{X}_i) < \infty$. Applying Cauchy-Schwarz inequality and Lemma 3,

$$\begin{aligned}n^2\mathrm{Var}(K_{n12}) &= \sum_{i_1,i_2}\sum_{j_1,j_2\in S}\mathrm{E}\left(e_{i_1}^2e_{i_2}^2(X_{i_1j_1}^{*2}-1)(X_{i_2j_2}^{*2}-1)\right)\\ &= \sum_{i_1\neq i_2}\sum_{j_1,j_2\in S}\sigma^4\mathrm{E}(X_{i_1j_1}^{*2}-1)\mathrm{E}(X_{i_2j_2}^{*2}-1))\\ &\quad+\sum_{i=1}^{n}\sum_{j_1,j_2\in S}\mathrm{E}(e_i^4(X_{ij_1}^{*2}-1)(X_{ij_2}^{*2}-1))\\ &\leq C\sum_{i=1}^{n}\mathrm{E}\left(X_{ib}^{*T}X_{ib}-\mathrm{E}(X_{ib}^{*T}X_{ib})\right)^2\\ &= O(np).\end{aligned}$$

Thus, $\mathrm{Var}(K_{n12}) = O(pn^{-1}) = o(p)$, which is followed by

$$\frac{K_{n12}}{(2\sigma^4 tr(\Sigma_{b|a}^2))^{\frac{1}{2}}} \xrightarrow{P} 0.$$

For K_{n11}, it is easy to see that $\mathrm{E}(K_{n11}) = 0$. Furthermore,

$$\begin{aligned} n^2\mathrm{Var}(K_{n11}) &= \sum_{i_1\neq i_2}\sum_{i_3\neq i_4}\sum_{j_1,j_2\in S} \mathrm{E}(e_{i_1}e_{i_2}e_{i_3}e_{i_4}X^*_{i_1j_1}X^*_{i_2j_1}X^*_{i_3j_2}X^*_{i_4j_2}) \\ &= 2\sum_{i_1\neq i_2}\sum_{j_1,j_2\in S} \mathrm{E}(e_{i_1}^2 e_{i_2}^2 X^*_{i_1j_1}X^*_{i_2j_1}X^*_{i_1j_2}X^*_{i_2j_2}) \\ &= 2n(n-1)\sigma^4 tr(\Sigma_{b|a}^2), \end{aligned}$$

which shows $\mathrm{Var}(K_{n11}) = 2\sigma^4 tr(\Sigma_{b|a}^2)(1 + o(1))$.

In order to verify (A.8), it suffices to show

$$\frac{K_{n11}}{\mathrm{Var}(K_{n11})^{\frac{1}{2}}} \xrightarrow{d} N(0, 1).$$

Following the techniques used in the proof of Theorem 1, we could obtain the asymptotic normality above by the martingale central limit theorem.

Finally, note that

$$(F_n(\mathbf{Y})-F(\mathbf{Y}))^T\mathbb{H}(F_n(\mathbf{Y})-F(\mathbf{Y})) \leq \lambda_{\max}(\mathbb{H})(F_n(\mathbf{Y})-F(\mathbf{Y}))^T(F_n(\mathbf{Y})-F(\mathbf{Y})),$$

thus we obtain the order of K_{n2} is $o_p(p^{1/2})$. Cauchy-Schwarz inequality shows that K_{n3} is also $o_p(p^{1/2})$, which completes the whole proof.

References

[1] Bai, Z. and H. Saranadasa (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica 6*(2), 311–329. MR1399305

[2] Bao, Y. and A. Ullah (2010). Expectation of quadratic forms in normal and nonnormal variables with applications. *Journal of Statistical Planning and Inference 140*(5), 1193–1205. MR2581122

[3] Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal 21*(1), C1–C68. MR3769544

[4] Cui, H., W. Guo, and W. Zhong (2018). Test for high-dimensional regression coefficients using refitted cross-validation variance estimation. *The Annals of Statistics 46*(3), 958–988. MR3797993

[5] Ducharme, G. R. and P. L. de Micheaux (2020). A goodness-of-fit test for elliptical distributions with diagnostic capabilities. *Journal of Multivariate Analysis 178*, 104602. MR4076341

[6] Eftekhari, H., M. Banerjee, and Y. Ritov (2021). Inference in high-dimensional single-index models under symmetric designs. *Journal of Machine Learning Research 22*(1), 1247–1309. MR4253720

[7] Fan, J., Y. Fan, and J. Lv (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics 147*(1), 186–197. MR2472991

[8] Fan, J., S. Guo, and N. Hao (2012). Variance estimation using refitted cross-validation in ultrahigh dimensional regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 74*(1), 37–65. MR2885839

[9] Fan, J. and J. Lv (2008). Sure independence screening for ultra-high dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70*(5), 849–911. MR2530322

[10] Fang, K.-T., S. Kotz, and K. W. Ng (2018). *Symmetric Multivariate and Related Distributions*. Chapman and Hall/CRC. MR1071174

[11] Feng, L., C. Zou, Z. Wang, and B. Chen (2013). Rank-based score tests for high-dimensional regression coefficients. *Electronic Journal of Statistics 7*, 2131–2149. MR3104951

[12] Feng, Z., T. Wang, and L. Zhu (2014). Transformation-based estimation. *Computational Statistics & Data Analysis 78*, 186–205. MR3212166

[13] Goeman, J. J., S. A. Van De Geer, and H. C. Van Houwelingen (2006). Testing against a high dimensional alternative. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(3), 477–493. MR2278336

[14] Guo, B. and S. X. Chen (2016). Tests for high dimensional generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78*(5), 1079–1102. MR3557190

[15] Guo, Z., C. Renaux, P. Bühlmann, and T. Cai (2021). Group inference in high dimensions with applications to hierarchical testing. *Electronic Journal of Statistics 15*(2), 6633–6676. MR4357274

[16] Hall, P. and C. C. Heyde (2014). *Martingale Limit Theory and Its Application*. Academic press. MR0624435

[17] Huffer, F. W. and C. Park (2007). A test for elliptical symmetry. *Journal of Multivariate Analysis 98*(2), 256–281. MR2301752

[18] Lan, W., H. Wang, and C.-L. Tsai (2014). Testing covariates in high-dimensional regression. *Annals of the Institute of Statistical Mathematics 66*(2), 279–301. MR3171406

[19] Li, G., H. Peng, J. Zhang, and L. Zhu (2012). Robust rank correlation based screening. *The Annals of Statistics 40*(3), 1846–1877. MR3015046

[20] Li, R., W. Zhong, and L. Zhu (2012). Feature screening via distance correlation learning. *Journal of the American Statistical Association 107*(499), 1129–1139. MR3010900

[21] Li, Y., I. Kim, and Y. Wei (2020). Randomized tests for high-dimensional regression: A more efficient and powerful solution. *Advances in Neural Information Processing Systems 33*.

[22] Liu, C., X. Zhao, and J. Huang (2021). New tests for high-dimensional linear regression based on random projection. *Statistica Sinica*, to apear. MR4527780

[23] Mai, Q., D. He, and H. Zou (2022). Coordinatewise gaussianization: Theories and applications. *Journal of the American Statistical Association 0*(0), 1–15.

[24] Meinshausen, N., L. Meier, and P. Buehlmann (2009). P-values for high-dimensional regression. *Journal of the American Statistical Association 104*(488), 1671–1681. MR2750584

[25] Redfern, C. H., P. Coward, M. Y. Degtyarev, E. K. Lee, A. T. Kwa, L. Hennighausen, H. Bujard, G. I. Fishman, and B. R. Conklin (1999). Conditional expression and signaling of a specifically designed gi-coupled receptor in transgenic mice. *Nature biotechnology 17*(2), 165–169.

[26] Srivastava, M. S. (2005). Some tests concerning the covariance matrix in high dimensional data. *Journal of the Japan Statistical Society 35*(2), 251–272. MR2328427

[27] Tony Cai, T., W. Liu, and Y. Xia (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76*(2), 349–372. MR3164870

[28] Wang, T., P.-R. Xu, and L.-X. Zhu (2012). Non-convex penalized estimation in high-dimensional models with single-index structure. *Journal of Multivariate Analysis 109*, 221–235. MR2922865

[29] Wasserman, L. and K. Roeder (2009). High dimensional variable selection. *Annals of Statistics 37*(5A), 2178–2201. MR2543689

[30] Zhong, P.-S. and S. X. Chen (2011). Tests for high-dimensional regression coefficients with factorial designs. *Journal of the American Statistical Association 106*(493), 260–274. MR2816719

[31] Zhu, L. and L. Zhu (2010). Stable direction recovery in single-index models with a diverging number of predictors. *Science China Mathematics 53*(7), 1817–1826. MR2665516