Testing the effects of high-dimensional covariates via aggregating cumulative covariances

Runze Li, Kai Xu, Yeqing Zhou and Liping Zhu
The Pennsylvania State University, Anhui Normal University, Tongji University and Renmin University of China

Abstract

In this paper, we test for the effects of high-dimensional covariates on the response. In many applications, different components of covariates usually exhibit various levels of variation, which is ubiquitous in high-dimensional data. To simultaneously accommodate such heteroscedasticity and high dimensionality, we propose a novel test based on an aggregation of the marginal cumulative covariances, requiring no prior information on the specific form of regression models. Our proposed test statistic is scale-invariance, tuning-free and convenient to implement. The asymptotic normality of the proposed statistic is established under the null hypothesis. We further study the asymptotic relative efficiency of our proposed test with respect to the state-of-art universal tests in two different settings: one is designed for high-dimensional linear model and the other is introduced in a completely model-free setting. A remarkable finding reveals that, thanks to the scale-invariance property, even under the high-dimensional linear models, our proposed test is asymptotically much more powerful than existing competitors for the covariates with heterogeneous variances while maintaining high efficiency for the homoscedastic ones.

Key words: Conditional mean independence; cumulative covariance; high dimension; martingale difference divergence.

*Runze Li (Email: rzli@psu.edu) is the Eberly Family Chair in Statistics, The Pennsylvania State University, University Park, PA 16802, USA. Kai Xu (Email: tjxxukai@ahnu.edu.cn) is Associate Professor, School of Mathematics and Statistics, Anhui Normal University, Wuhu 241002, China. Yeqing Zhou (Email: zhouyeqing@tongji.edu.cn) is Assistant Professor, School of Mathematical Sciences, Tongji University, Shanghai 200092, China. Liping Zhu (Email: zhu.liping@ruc.edu.cn) is Professor and corresponding author, Center for Applied Statistics and Institute of Statistics and Big Data, Renmin University of China, Beijing 100872, China. The authors make equal contributions to this paper and their names are listed in the alphabetical order.

†We are grateful to the Editor, the Associate Editor and reviewers for their constructive comments, which lead to a significant improvement of this work. We are also very grateful to Professor Min Qian for providing us with the R codes of McKeague and Qian (2015). This work is supported by National Natural Science Foundation of China (12171477, 11901006, 12001405, 11731011, 11931014), Fundamental Research Funds for the Central Universities (22120210557), Natural Science Foundation of Anhui Province (1908085QA06) and National Science Foundation (DMS-1820702, 1953196 and 2015539).

1 Introduction

In regression analysis, it is of fundamental importance to infer whether a set of covariates $\mathbf{x} = (X_1, \ldots, X_p)^{\mathrm{T}} \in \mathbb{R}^p$ has any effect on a response $Y \in \mathbb{R}^1$. Consider the null hypothesis

$$H_0 : E(Y \mid \mathbf{x}) = E(Y), \quad \text{almost surely.} \tag{1.1}$$

If H_0 holds, it implies that $\mathbf{x}$ does not contribute to the conditional mean function of Y. Thus there is no need to build a regression model for the conditional mean function. In genomic studies, exploring whether a set of genes is predictive for certain clinical outcomes can be formulated as the hypothesis in (1.1). See, for example, the significant gene set selection in Subramanian et al. (2005), Efron and Tibshirani (2007) and Zhong and Chen (2011).

Testing for the covariates effects has received much attention, and many tests have been developed for low and fixed-dimensional covariates. By assuming $E(Y \mid \mathbf{x}) = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}$, testing (1.1) is equivalent to checking whether $\boldsymbol{\beta} = 0$. The classical F-test can be used to infer the overall significance of linear regression coefficients. Moreover, many model-specification tests are also designed to test (1.1), including the local (Hardle and Mammen, 1993; Zheng, 1996; Guo et al., 2016) and global smoothing tests (Stute, 1997; Stute et al., 1998; Escanciano, 2006). Without specifying a functional form of $E(Y \mid \mathbf{x})$, Wang and Akritas (2006) developed a test for covariate effects in a completely nonparametric fashion. Shao and Zhang (2014) proposed a martingale difference divergence in the setting of fixed dimension. However, these methods do not target the case of high-dimensional covariates and suffer from the curse of dimensionality when the covariate dimension p diverges. In particular, Zhong and Chen (2011) proved the power of F-test is adversely impacted by an increasing ratio p/n even when $p < n - 1$, where n is the sample size. For the martingale difference divergence that can capture any type of conditional mean dependence in the fixed dimension, Zhang et al. (2018) showed that it can only measure the component-wise linear dependence in high dimension.

In the case of high dimensional covariates, many tests have been proposed under parametric model settings. For example, in high-dimensional linear regression models, Goeman et al. (2006) considered a score test in an empirical Bayesian model. Zhong and Chen (2011) proposed a simultaneous test for coefficients based on a U-statistic. This work is further modified by Feng et al. (2013) with Wilcoxon scores, and by Cui et al. (2018) with refitted cross-validation. In high-dimensional generalized linear regression models, Goeman et al. (2011) extended the test of Goeman et al. (2006) and derived its asymptotic distribution. Guo and Chen (2016) introduced a test that is robust to a wide range of link functions. It is desirable to develop a test that can accommodate the high-dimensionality without any parametric model assumptions.

Directly testing (1.1) without specifying any model structure is very challenging in high dimensions. McKeague and Qian (2015) proposed an adaptive resampling test (ART) using the maximum-type statistic on the slopes of marginal linear regressions. The ART may effectively detect the presence of significant covariates under working model: $Y = \beta_0 + \mathbf{x}^\top \boldsymbol{\beta} + \varepsilon$, where ε and $\mathbf{x}$ are uncorrelated, and $\boldsymbol{\beta} = \arg\min_{\boldsymbol{\gamma}} E(Y - \mathbf{x}^\top \boldsymbol{\gamma})^2$. Thus, the ART procedure may be directly applied for the hypothesis (1.1) under condition that $\boldsymbol{\beta} = 0$ implies that $E(Y \mid \mathbf{x}) = E(Y)$. Luedtke and van der Laan (2018) established theoretical properties of the standardized ART procedure under high-dimensional setting and further introduced a more computationally tractable approach to the ART. Zhang et al. (2018) considered a relatively weaker null hypothesis:

$$H_0' : E(Y \mid X_s) = E(Y) \ \text{ almost surely, } \ \text{for all } 1 \leq s \leq p. \qquad (1.2)$$

It is clear that H_0 in (1.1) implies H_0' in (1.2), but not vice versa. In other words, the distance between $E(Y \mid \mathbf{x})$ and $E(Y)$ cannot be fully captured by pairwise distances between $E(Y \mid X_s)$ and $E(Y)$, for $s = 1, \ldots, p$. However, Zhang et al. (2018) pointed out that the difference between $E(Y \mid X_s)$ and $E(Y)$ can be regarded as the marginal effect of X_s contributing to Y. From this point of view, the pairwise distances between $E(Y \mid X_s)$ and $E(Y)$ are still informative in testing H_0 in (1.1). The pairwise comparison partly motivates us to develop a novel high-dimensional nonparametric test for H_0' in (1.2).

In real applications, different components of covariates usually exhibit different levels of variation. For example, in high-dimensional microarray data, the variation of expression differs substantially from gene to gene (Nettleton et al., 2008). The heterogeneous variances of covariates may affect the performances of the testing procedures, when the test statistics are not scale invariant. See, for example, McKeague and Qian (2015) and its discussions and rejoinder for extensive discussions on this issue. To deal with the issue, a widely used strategy is to standardize each covariate by its corresponding standard deviation prior to implementing the aforementioned tests. Examples include Zhong and Chen (2011), McKeague and Qian (2015) and Zhang et al. (2018). This strategy however brings difficulties in theoretical justifications for diverging p and requires implicitly the variances of all covariates be finite.

To accommodate the issues of heterogeneous covariates variances and high dimensionality, we introduce a novel test to detect the mean effects of high-dimensional covariates on the response. Without any model assumptions, our test statistic is built on cumulative covariance (Zhou et al., 2020), which utilizes the rank of covariates and hence is scale-invariance. To illustrate the appealing features of our proposed test, we compare it with two state-of-art tests: one is proposed by Zhong and Chen (2011) and the other by Zhang et al. (2018). The first test is designed for high-dimensional model and originates from the classical F-test that is known to be powerful when p is fixed as $n \to \infty$. The second test is developed in a high-dimensional model-free setting. The asymptotic properties of these two tests are universal (Paindaveine and Verdebout, 2016), in the sense that p may go to infinity in an arbitrary rate as n goes. This also applies to our proposed test. In what follows, we summarize our contributions as well as the desirable properties of the new test.

- A direct implementation of the proposed test statistic according to its definition has a computational complexity of order $O(n^5 p)$, which is computationally expensive. By sorting the covariates in an increasing order, we provide an efficient algorithm to reduce the computational cost of its numerator to $O\{np\log(n)\}$. This algorithm can also be adapted to Zhong and Chen (2011)'s and Zhang et al. (2018)'s test statistics to improve the computational efficiency of their numerators. The denominators of three statistics can be calculated in $O(n^2 p)$ operations. Therefore, the overall computational complexities of three tests are all of order $O(n^2 p)$.

- We derive the universally asymptotic normality of our proposed test statistic under the high-dimensional null hypothesis, without any assumption on relative growth rate between p and n. No bootstrap or random permutation is required to approximate the asymptotic null distributions. In this sense, our testing procedure is tuning-free and distribution-free.

- We derive the asymptotic power function of our proposed test and carefully study its asymptotic relative efficiency with respect to the tests proposed by Zhong and Chen (2011) and Zhang et al. (2018) in high-dimensional linear models. When each covariate has the same variance, we prove that the asymptotic relative efficiency of our proposal is near 0.872 with respect to Zhong and Chen (2011), and 0.979 with respect to Zhang et al. (2018). However, under the heterogeneous variances of covariates, the asymptotic relative efficiency can even go to infinity as p goes to infinity. See Section 3 for detailed discussions. This implies that our proposed test has little efficiency loss for homoscedastic covariates, but substantial efficiency gain for heteroscedastic cases.

The rest of the paper is organized as follows. In Section 2, we introduce a new conditional mean testing procedure based on the cumulative covariance, and derive its asymptotic distribution under the null hypothesis and alternatives. Section 3 carefully studies the asymptotic relative efficiency of the proposed test with respect to the tests in Zhong and Chen (2011) and Zhang et al. (2018). We assess the finite sample performance of the proposed test through Monte Carlo simulations and a real data application in Sections 4.1 and 4.2, respectively. A short discussion is given in Section 5. All technical proofs and extra simulations are relegated to the supplement.

2 A New Test Procedure

For the purpose of comparison, we first review the tests proposed by Zhong and Chen (2011) and Zhang et al. (2018). We will compare the asymptotic relative efficiency of these two tests with our proposed procedure in Section 3.

2.1 Two existing tests allowing for universal (n, p)-asymptotics

Let $\mathbf{x}_i = (X_{i1}, \ldots, X_{ip})^{\mathrm{T}}$. Suppose we have a random sample $\{(\mathbf{x}_i, Y_i), i = 1, \ldots, n\}$ drawn independently from the joint distribution of $(\mathbf{x}, Y)$. Throughout this paper, we denote $(n)_m = n(n-1)\ldots(n-m+1)$ and $C(n,m) = (n)_n/\{(n)_m(n-m)_{n-m}\}$ for $1 \le m \le n$, and

$$\sum_{(i,j)}^{n}, \sum_{(i,j,k)}^{n}, \sum_{(i,j,k,l)}^{n} \text{ and } \sum_{(i,j,k,l,r)}^{n}$$

denote summations that are taken over all possible permutations of distinctive indices.

The test in Zhong and Chen (2011) is constructed through a modified F-statistic under linear model assumption,

$$\mathrm{ZC}_{n,p} = \{4(n)_4\widehat{\sigma}_1\}^{-1} \sum_{s=1}^{p} \sum_{(i,j,k,l)}^{n} (Y_i - Y_j)(Y_k - Y_l)(X_{is} - X_{js})(X_{ks} - X_{ls}), \qquad (2.1)$$

where $\widehat{\sigma}_1^2 = [\{2(n)_2\}^{-1} \sum_{(i,j)}^{n} (Y_i - Y_j)^2][\{4(n)_4\}^{-1} \sum_{(i,j,k,l)}^{n} \{(\mathbf{x}_i - \mathbf{x}_j)^{\mathrm{T}}(\mathbf{x}_k - \mathbf{x}_l)\}^2]$. The test statistic in Zhang et al. (2018) is built upon the martingale difference divergence without model assumptions,

$$\begin{aligned} \mathrm{ZYS}_{n,p} &= \{4(n)_4\widehat{\sigma}_2\}^{-1}\sum_{s=1}^{p}\sum_{(i,j,k,l)}^{n}(Y_i-Y_j)(Y_k-Y_l)(|X_{is}-X_{ls}| \\ &+ |X_{js}-X_{ks}|-|X_{is}-X_{ks}|-|X_{js}-X_{ls}|), \end{aligned} \tag{2.2}$$

where $\widehat{\sigma}_2^2 = [(n-1)^3/\{n(n-3)^4\}]\sum_{(i,j)}^{n}\sum_{s,t=1}^{p}\widetilde{A}_{ij}(s)\widetilde{A}_{ij}(t)\widetilde{B}_{ij}$, and $\widetilde{A}_{ij}(s)$ and $\widetilde{B}_{ij}$ are the U-centered versions of $A_{ij}(s)=|X_{is}-X_{js}|$ and $\widetilde{B}_{ij}=(Y_i-Y_j)^2/2$, respectively. See more details in Zhang et al. (2018, Equation (7)). Zhong and Chen (2011, Theorem 3) and Zhang et al. (2018, Theorem 2.2) established the asymptotic normality properties for $\{n(n-1)/2\}^{1/2}\mathrm{ZC}_{n,p}$ and $\{n(n-1)/2\}^{1/2}\mathrm{ZYS}_{n,p}$ when both the dimension and the sample size go to infinity.

Both tests of Zhong and Chen (2011) and Zhang et al. (2018) require the existence of the second moments of covariates and are not invariant under scale transformations, which indicates their power performances heavily depend on the variance magnitudes of covariates. In Sections 3 and 4, we will show the advantages of scale-invariance property from both the asymptotic and the numerical perspectives.

In high-dimension setting, easy implementation is a desirable property for testing. Therefore, we are interested in the computational complexity of these two tests. Naively computing $\mathrm{ZC}_{n,p}$ and $\mathrm{ZYS}_{n,p}$ by (2.1) and (2.2) is very complicated. Our fast algorithm given in the following Section 2.4 can be adapted to calculate the numerators of $\mathrm{ZC}_{n,p}$ and $\mathrm{ZYS}_{n,p}$, which have the computational complexity of $O(np)$ and $O\{np\log(n)\}$, respectively. The details of the adapted algorithms can be found in Section S.7 of the supplement. The denominators of $\mathrm{ZC}_{n,p}$ and $\mathrm{ZYS}_{n,p}$, namely $\widehat{\sigma}_1$ and $\widehat{\sigma}_2$, can be computed in $O(n^2p)$ operations.

2.2 Cumulative covariance revisited

We next review the cumulative covariance that is introduced by Zhou et al. (2020). Let X and Y denote two random variables. Under the assumption that the second moment of Y is finite, the cumulative covariance $\mathrm{CCov}(Y \mid X)$ is defined as

$$\mathrm{CCov}(Y \mid X) = E[\mathrm{cov}^2\{Y, I(X < \widetilde{X}) \mid \widetilde{X}\}], \tag{2.3}$$

where $(\widetilde{X}, \widetilde{Y})$ is an independent copy of (X, Y) and $I(\cdot)$ is an indicator function. The cumulative covariance is non-negative and equals zero if and only if $E(Y \mid X) = E(Y)$. In this sense, the cumulative covariance can fully characterize the conditional mean dependence. An appealing property of the cumulative covariance is that it keeps invariant with respect to arbitrary strictly monotone transformation of X. This invariance property, however, is not shared by martingale difference divergence. Therefore, the test proposed by Zhang et al. (2018), which is built upon martingale difference divergence, requires the second moment of X be finite, while the CCov-based test allows it to be infinity. We shall elaborate this in detail in the sequel.

2.3 The CCov-based test statistic in high dimension

To test H_0' in (1.2), it is natural to use the summation of all marginal cumulative covariances,

$$\sum_{s=1}^{p} \mathrm{CCov}(Y \mid X_s), \tag{2.4}$$

which is non-negative and equals zero if and only if the pairwise differences between $E(Y \mid X_s)$ and $E(Y)$, for all $s = 1, \ldots, p$, are identically zero. For simplicity, we assume that all the components of $\mathbf{x}$ are continuous so that the probability of a tie occurring in the data is zero. A natural estimator of (2.4) can be defined as

$$W_{n,p} = n^{-3} \sum_{s=1}^{p} \sum_{j=1}^{n} \left[\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right) \{I(X_{is} < X_{js}) - F_{n,s}(X_{js})\} \right]^2,$$

where

$$\overline{Y} = n^{-1} \sum_{i=1}^{n} Y_i, \text{ and } F_{n,s}(X_{js}) = n^{-1} \sum_{i=1}^{n} I(X_{is} < X_{js}).$$

This sample version appears straightforward, however, $W_{n,p}$ involves several redundant terms that bring in asymptotically non-negligible bias-terms in high dimension, resulting in a fragile size performance for the test based on $W_{n,p}$. The details of bias-terms can be found in Appendix S.2. To formulate the CCov-based test in high dimension, we consider instead

$$T_{n,p} = \{4(n)_5\}^{-1} \sum_{s=1}^{p} \sum_{(i,j,k,l,r)} (Y_i - Y_j)(Y_k - Y_l)\psi(X_{is}, X_{js}, X_{rs})\psi(X_{ks}, X_{ls}, X_{rs}), \quad (2.5)$$

where $\psi(X_1, X_2, X_3) = I(X_1 < X_3) - I(X_2 < X_3)$. Proposition 1 in Appendix S.1 ensures that $T_{n,p}$ is an unbiased estimator of (2.4). Thus $T_{n,p}$ is basically all we need to test H_0' in (1.2). For arbitrary strictly monotone transformations M_s, we have $\psi(X_{is}, X_{js}, X_{rs})\ \psi(X_{ks}, X_{ls}, X_{rs}) = \psi\{M_s(X_{is}), M_s(X_{js}), M_s(X_{rs})\}\psi\{M_s(X_{ks}), M_s(X_{ls}), M_s(X_{rs})\}$ for $i, j, k, l, r = 1, \ldots, n$. Consequently, the proposed test statistic $T_{n,p}$ is automatically scale-invariance.

2.4 Computational algorithm

Directly computing $T_{n,p}$ through (2.5) has a computational complexity of order $O(n^5p)$. To reduce the computational burden, we provide a computationally efficient algorithm to implement the proposed statistic $T_{n,p}$. First, for any $s = 1, \ldots, p$, sort the n observation of $\{X_{is} : i = 1, \ldots, n\}$ to be $X_{(1)s} < \ldots < X_{(n)s}$. Second, find the corresponding response $Y_{(i)s}$ associated with $X_{(i)s}$. Denote $\dot{Y}_{(i)s} = Y_{(i)s} - \overline{Y}$. After these two preliminary steps, the following theorem shows how to compute $T_{n,p}$ efficiently.

Theorem 1. *For a random sample* $\{\mathbf{x}_i, Y_i\}_{i=1}^n$ *drawn from joint distribution of* $(\mathbf{x}, Y)$,

$$
\begin{aligned}
T_{n,p} \;=\; & \{(n)_5\}^{-1}\Big[(n-2)(n-3)\sum_{s=1}^{p}\sum_{j=1}^{n}\Big(\sum_{i=1}^{j-1}\dot{Y}_{(i)s}\Big)^2 + 2\sum_{s=1}^{p}\sum_{j=1}^{n}\Big\{\big(nj-2n-2j+2\big) \\
& \times\dot{Y}_{(j)s}\sum_{i=1}^{j-1}\dot{Y}_{(i)s}\Big\} - \sum_{s=1}^{p}\sum_{j=1}^{n}\Big\{\big(n^2-2nj-n+4j-4\big)\sum_{i=1}^{j-1}\dot{Y}_{(i)s}^2\Big\} \\
& -\{n(n^2-3n+8)/3\}\sum_{s=1}^{p}\sum_{i=1}^{n}\dot{Y}_{(i)s}^2 + 2\sum_{s=1}^{p}\sum_{i=1}^{n}(i-1)^2\dot{Y}_{(i)s}^2\Big].
\end{aligned}
$$

Theorem 1 guarantees that $T_{n,p}$ can be computed in only $O\{np\log(n)\}$ operations.

2.5 Asymptotic null distribution

To establish the asymptotic normality of our test statistic, we study Hoeffding decomposition (Serfling, 1980) for the variance of $T_{n,p}$, which is valid for diverging p. Define $\phi_s(i,j,k,l,r) = (Y_i - Y_j)(Y_k - Y_l)\psi(X_{is}, X_{js}, X_{rs})\psi(X_{ks}, X_{ls}, X_{rs})/4$ for

$i, j, k, l, r = 1, \ldots, n$ and $s = 1, \ldots, p$, and symmetrize it by

$$
\begin{aligned}
h_s(i,j,k,l,r) \;=\; & \big\{\phi_s(i,j,k,l,r) + \phi_s(i,k,j,l,r) + \phi_s(i,l,j,k,r) \\
& + \phi_s(r,j,k,l,i) + \phi_s(r,k,j,l,i) + \phi_s(r,l,j,k,i) + \phi_s(i,r,k,l,j) \\
& + \phi_s(i,k,r,l,j) + \phi_s(i,l,r,k,j) + \phi_s(i,j,r,l,k) + \phi_s(i,r,j,l,k) \\
& + \phi_s(i,l,j,r,k) + \phi_s(i,j,k,r,l) + \phi_s(i,k,j,r,l) + \phi_s(i,r,j,k,l)\big\}/15.
\end{aligned}
$$

Write $h(i,j,k,l,r) = \sum_{s=1}^{p} h_s(i,j,k,l,r)$, for $i,j,k,l,r = 1,\ldots,n$. Then, the statistic $T_{n,p}$ has the following expression,

$$
T_{n,p} = 1/\{n(n-1)(n-2)(n-3)(n-4)\} \sum_{(i,j,k,l,r)}^{n} h(i,j,k,l,r).
$$

It is clear that $T_{n,p}$ is actually a U-statistic of order five. This finding is very useful in subsequent derivations. For $c = 1,\ldots,5$, let $h^{(c)}(\mathbf{z}_1,\ldots,\mathbf{z}_c) = E\{h(1,2,3,4,5) \mid \mathbf{z}_1,\ldots,\mathbf{z}_c\}$ be projections of h to lower-dimensional sample spaces, where $\mathbf{z}_i = (\mathbf{x}_i, Y_i)$ is the i-th observation.

To determine the asymptotic form of $T_{n,p}$, we study its variance decomposition for high-dimensional data. Towards this end, we impose the following assumption.

Assumption 1. Assume that

$$
0 < c \le \operatorname{var}(Y \mid \mathbf{x}) \le E^{1/2}[\{Y - E(Y \mid \mathbf{x})\}^4 \mid \mathbf{x}] \le C < \infty, \tag{2.6}
$$

almost surely for some constants c and C.

This assumption is also considered by Patilea et al. (2016) and Zhang et al. (2018)

to derive the asymptotic properties of their test statistics. Assumption 1 holds true for any p when $Y = E(Y \mid \mathbf{x}) + \sigma(\mathbf{x})\varepsilon$, where $E(\varepsilon \mid \mathbf{x}) = 0$, $E(\varepsilon^2 \mid \mathbf{x}) \geq c_1 > 0$, $E(\varepsilon^4 \mid \mathbf{x}) \leq C_1 < \infty$ and $0 < c < \sigma(\mathbf{x}) < C < \infty$. If the model error $\{Y - E(Y \mid \mathbf{x})\}$ and $\mathbf{x}$ are independent, and the fourth moments of the model error are bounded, then (2.6) is trivially true.

Denote $\xi_c = \text{var}\{h^{(c)}(\mathbf{z}_1, \ldots, \mathbf{z}_c)\}$, for $c = 1, \ldots, 5$. The following lemma states the Hoeffding decomposition (Serfling, 1980) for the variance of $T_{n,p}$ as $p \to \infty$.

Lemma 1. *Suppose that Assumption 1 holds. Under H_0', we have $\xi_1 = 0$, $var(T_{n,p}) = \{C(n,5)\}^{-1} \sum_{c=2}^{5} C(5,c)C(n-5,5-c)\xi_c$ and the terms ξ_2, ξ_3, ξ_4 and ξ_5 are of the same order as $p \to \infty$. In particular, $var(T_{n,p}) = \{n(n-1)/2\}^{-1} S^2\{1 + o(1)\}$, where*

$$S^2 = 4^{-1} var\left[\sum_{s=1}^{p} K_0(Y_1, Y_2) K_1\{F_s(X_{1s}), F_s(X_{2s})\}\right],$$

$K_0(Y_1, Y_2) = \{Y_1 - E(Y)\}\{Y_2 - E(Y)\}$, $K_1\{F_s(X_{1s}), F_s(X_{2s})\} = F_s^2(X_{1s}) + F_s^2(X_{2s}) - 2\max\{F_s(X_{1s}), F_s(X_{2s})\} + 2/3$, and $F_s(\cdot)$ is the cumulative distribution function of X_s for $s = 1, \ldots, p$.

Define

$$\widetilde{T}_{n,p} = \{n(n-1)/2\}^{-1/2} \sum_{(i,j)}^{n} \sum_{s=1}^{p} K_0(Y_i, Y_j) K_1\{F_s(X_{is}), F_s(X_{js})\}/4.$$

Based on Lemma 1, we derive that $[\{n(n-1)/2\}^{1/2} T_{n,p} - \widetilde{T}_{n,p}]/S \to 0$ in probability under H_0' in the supplement. Thus, it suffices to derive the asymptotic distribution of $\widetilde{T}_{n,p}/S$ under the null.

Define $V(\mathbf{x}_1, \mathbf{x}_2) = \sum_{s=1}^{p} K_1\{F_s(X_{1s}), F_s(X_{2s})\}$. To establish the asymptotic normality of $\widetilde{T}_{n,p}/S$, we use the martingale central limit theorem (Hall and Heyde, 2014, CLT). The following assumption is imposed to facilitate the proof of martingale CLT and is closely related to the typical condition (2.1) of Hall (1984).

Assumption 2. As $p \to \infty$ and $n \to \infty$,

$$E\{V(\mathbf{x}_1,\mathbf{x}_2)V(\mathbf{x}_2,\mathbf{x}_3)V(\mathbf{x}_3,\mathbf{x}_4)V(\mathbf{x}_4,\mathbf{x}_1)\}/E^2\{V(\mathbf{x}_1,\mathbf{x}_2)^2\} \to 0,$$
$$E\{V(\mathbf{x}_1,\mathbf{x}_2)^4\}/[nE^2\{V(\mathbf{x}_1,\mathbf{x}_2)^2\}] \to 0.$$

Assumption 2 is presented in an abstract way and can be made more explicit under specific dependence structures. To illustrate this, we consider the commonly encountered banded dependence structure, where the random vector $\mathbf{x}$ is m-dependent. In Section S.9 of the supplement, it can be verified that

$$E\{V(\mathbf{x}_1,\mathbf{x}_2)^2\} \geq 8p/45 \to \infty,$$
$$0 \leq E\{V(\mathbf{x}_1,\mathbf{x}_2)V(\mathbf{x}_2,\mathbf{x}_3)V(\mathbf{x}_3,\mathbf{x}_4)V(\mathbf{x}_4,\mathbf{x}_1)\} = O(pm^3),$$
$$E\{V(\mathbf{x}_1,\mathbf{x}_2)^4\} = O\left[pm^3 + E^2\{V(\mathbf{x}_1,\mathbf{x}_2)^2\}\right].$$

Assumption 2 is trivially satisfied if $m = o(p^{1/3})$ for the divergent p. In particular, if m is a fixed constant, the above conditions are fairly mild when p is divergent. Moreover, there is no explicit relationship between p and n in Assumption 2. If the coordinates of $\mathbf{x}$ are independent but not necessarily identically distributed, p can grow to infinity freely as $n \to \infty$.

Theorem 2. *Suppose that Assumptions 1 and 2 hold. Under the null hypothesis H_0', as $n, p \to \infty$, $\{n(n-1)/2\}^{1/2}\, T_{n,p}/S$ converges in distribution to $N(0,1)$.*

To formulate a test procedure based on $T_{n,p}$, we need to provide a suitable variance estimator for S^2. We consider the following estimator,

$$S_{n,p}^2 = \{4c_n n(n-1)\}^{-1} \sum_{(i,j)}^{n} K_0(\dot{Y}_i, \dot{Y}_j)^2 \Big[\sum_{s=1}^{p} K_1\{F_{n,s}(X_{is}), F_{n,s}(X_{js})\} \Big]^2,$$

where $\dot{Y}_i = Y_i - \overline{Y}$, $F_{n,s}$ is the empirical distribution function of X_s and $c_n = \{(1 - n^{-1})^2 + n^{-2}\}^2$ is a finite sample adjustment factor to reduce the bias of $S_{n,p}^2$. See the proof of Theorem 3 for details. The variance estimator $S_{n,p}^2$ has a computational cost of order $O(n^2 p)$. In what follows, we derive the ratio-consistency of this variance estimator.

Theorem 3. *Suppose that Assumptions 1 and 2 hold. Then, we have the ratio consistency that as $n, p \to \infty$, $S_{n,p}^2/S^2$ converges in probability to 1. Consequently, $\{n(n-1)/2\}^{1/2}\, T_{n,p}/S_{n,p}$ converges in distribution to the standard normal distribution under the H_0'.*

Theorems 2 and 3 reveal that the asymptotic normality of our proposed statistic under the null hypothesis holds with no restriction on relative growth rate between p and n. Theorem 3 suggests that our proposed test rejects the null hypothesis H_0' in (1.2) at significant level α if $\{n(n-1)/2\}^{1/2}\, T_{n,p}/S_{n,p} > z_\alpha$, where z_α is the $1-\alpha$ quantile of standard normal.

2.6 Asymptotic distribution under alternatives

For the power analysis, we consider a class of alternatives H_1' satisfying

$$\mathrm{var}\left\{\sum_{s=1}^{p} L_s(Y_1, X_{1s})\right\} = o(n^{-1}S^2) \quad \text{and} \quad \mathrm{var}\left\{\sum_{s=1}^{p} L_s(Y_1, X_{2s})\right\} = o(S^2),$$

where $L_s(u,v) = E\Big[K_0(u, Y_2)K_1\{F_s(v), F_s(X_{2s})\}\Big]$. These two conditions are assumed to describe a small difference between H_0' and H_1' in intuitive way. Thus, the underlying alternatives may be viewed as 'local' alternatives. Rigorous definition of local alternatives perhaps may be arguably phrased in terms of contiguity, but this is beyond the scope of this paper. Under H_1', the variance of $T_{n,p}$ defined in Lemma 1 remains valid. In the following theorem, we derive the asymptotic distribution of our proposed test statistic under the alternatives, which allows for power evaluations.

Theorem 4. *Suppose that Assumptions 1 and 2 hold. Under H_1', as $n, p \to \infty$,*

$$\{n(n-1)/2\}^{1/2}\left\{T_{n,p} - \sum_{s=1}^{p} \mathrm{CCov}(Y \mid X_s)\right\}/S$$

converges in distribution to the standard normal.

Based on Theorems 3 and 4 as well as Slutsky's theorem, the power of the proposed test under H_1' is

$$\Psi_{n,p} = \{1 + o(1)\}\Phi\left[-z_\alpha + \{n(n-1)/2\}^{1/2}\sum_{s=1}^{p} \mathrm{CCov}(Y \mid X_s)/S\right],$$

where $\Phi(\cdot)$ is the cumulative distribution function of $N(0,1)$, and z_α denotes the $1-\alpha$

quantile of $N(0,1)$. The power of our test is in spirit controlled by

$$\mathrm{SNR}_{\mathrm{NEW}} = \{n(n-1)/2\}^{1/2} \sum_{s=1}^{p} \mathrm{CCov}(Y \mid X_s)/S,$$

which can be viewed as a signal to noise ratio.

3 Asymptotic Relative Efficiency

It is challenging to compare our test with the tests of Zhong and Chen (2011) and Zhang et al. (2018) in a completely model-free context. We study the asymptotic powers of these three tests under high-dimensional linear models, and anticipate that similar conclusions can be drawn from nonlinear models. Let us consider the model

$$Y = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon, \tag{3.1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$, $\mathbf{x} = (X_1, \ldots, X_p)^{\mathrm{T}} \sim N(0, \boldsymbol{\Sigma})$, and ε is independent of $\mathbf{x}$ with $E(\varepsilon) = 0$ and $\mathrm{var}(\varepsilon) = \sigma^2$. To illustrate the implication of $\mathrm{SNR}_{\mathrm{NEW}}$, we consider a diagonal matrix $\boldsymbol{\Sigma} = \mathrm{diag}(d_1, \ldots, d_p)$. Following Theorem 1(3) in Zhou et al. (2020), we can derive that

$$\mathrm{SNR}_{\mathrm{NEW}} \quad = \quad \left\{15n(n-1)/(4\pi^2\sigma^4 p)\right\}^{1/2} \sum_{s=1}^{p} d_s \beta_s^2,$$

By contrast, the asymptotic power of Zhong and Chen (2011)'s test depends on

$$\mathrm{SNR}_{\mathrm{ZC}} \quad = \quad \left\{n(n-1)/(2\sigma^4)\right\}^{1/2} \sum_{s=1}^{p} d_s^2 \beta_s^2 \Big(\sum_{s=1}^{p} d_s^2\Big)^{-1/2}.$$

From Shao and Zhang (2014, Theorem 1(3)) and Székely et al. (2007, Theorem 7(ii)), the asymptotic power of Zhang et al. (2018)'s test is related to

$$\mathrm{SNR}_{\mathrm{ZYS}} = \Big[n(n-1)/\{8\sigma^4(1-\sqrt{3}+\pi/3)\}\Big]^{1/2} \sum_{s=1}^{p} d_s^{3/2}\beta_s^2 \Big(\sum_{s=1}^{p} d_s\Big)^{-1/2}.$$

Based on the signal to noise ratios of three tests, the asymptotic relative efficiency of Zhong and Chen (2011)'s test with respect to our proposal is

$$\mathrm{ARE}(\mathrm{NEW},\mathrm{ZC}) = \big\{15/(2\pi^2 p)\big\}^{1/2}\Big(\sum_{s=1}^{p} d_s\beta_s^2\Big)\Big(\sum_{s=1}^{p} d_s^2\Big)^{1/2}\Big(\sum_{s=1}^{p} d_s^2\beta_s^2\Big)^{-1}.$$

The asymptotic relative efficiency of Zhang et al. (2018)'s test with respect to ours is

$$\mathrm{ARE}(\mathrm{NEW},\mathrm{ZYS}) = \Big\{30(1-\sqrt{3}+\pi/3)/(\pi^2 p)\Big\}^{1/2}\Big(\sum_{s=1}^{p} d_s\beta_s^2\Big)\Big(\sum_{s=1}^{p} d_s\Big)^{1/2}\Big(\sum_{s=1}^{p} d_s^{3/2}\beta_s^2\Big)^{-1}.$$

To view a rough picture of the asymptotic power comparison, we consider three scenarios in what follows.

3.1 Homoscedastic case: the marginal variance of each covariate is the same

In the homoscedastic case, $d_1 = \ldots = d_p$. Direct computation shows $\mathrm{ARE}(\mathrm{NEW},\mathrm{ZC}) \approx 0.872$ and $\mathrm{ARE}(\mathrm{NEW},\mathrm{ZYS}) \approx 0.979$. It follows that $\mathrm{ARE}(\mathrm{ZYS},\mathrm{ZC}) \approx 0.891$, which is in line with Remark 2.6 in Zhang et al. (2018). These results imply that our proposed test is asymptotically less powerful than Zhong and Chen (2011)'s test, which is specially designed for the linear models. Both our proposed test and Zhang et al.

(2018) 's test are model-free, and perform comparably under the homoscedastic case.

3.2 Heteroscedastic case when the number of non-zero effects is fixed

For simplicity, we assume that all non-zero coefficients β_s have the same magnitude, that is, $\beta_s = \kappa I(1 \leq s \leq q), s = 1, \ldots, p$, for $\kappa \neq 0$, $q \in \{1, \ldots, p\}$ is fixed. In this setting, if we assume the condition

$$p = o\left\{\min\left(\sum_{s=1}^{p} d_s^2, \sum_{s=1}^{p} d_s\right)\right\}, \tag{3.2}$$

we have $\mathrm{ARE}(\mathrm{NEW}, \mathrm{ZC}) \to \infty$ and $\mathrm{ARE}(\mathrm{NEW}, \mathrm{ZYS}) \to \infty$, as $p \to \infty$. The condition (3.2) is a sufficient, but not necessary condition for our test to be more powerful than Zhong and Chen (2011)'s test and Zhang et al. (2018) 's test. In Section S.8 of the supplement, we show that the proposed test may still have better power performances when $p = O\left\{\min\left(\sum_{s=1}^{p} d_s^2, \sum_{s=1}^{p} d_s\right)\right\}$. The condition (3.2) is trivially true in the case that different components of $\mathbf{x}$ have distinctive scales. Under this condition, our test is substantially more powerful than both Zhong and Chen (2011)'s and Zhang et al. (2018)'s tests. To further compare the asymptotic power performances of these two tests under the heteroscedastic case, we further impose the condition

$$\sum_{s=1}^{p} d_s = o\left(\sum_{s=1}^{p} d_s^2\right). \tag{3.3}$$

This condition is also mild if the variance of each covariate differs much. Under this assumption, we show the asymptotic power of Zhang et al. (2018)'s test is superior to that of Zhong and Chen (2011)'s test, which is also a totally new finding and have not been discovered by Zhang et al. (2018). Therefore, the asymptotic powers of three tests arranged in a descending order are those of our proposed test, Zhang et al. (2018) 's test and Zhong and Chen (2011)'s test.

Consider an explicit scenario satisfying (3.2) and (3.3): There is a parameter $\delta > 0$ not depending on the dimension p such that

$$d_s \asymp s^{\delta}, \text{for } s = 1, \ldots, p,$$

where for two sequences $\{a_s\}$ and $\{b_s\}$, we write $a_s \asymp b_s$ if there exist positive constants c and C such that $c \leq \liminf_s a_s/b_s \leq \limsup_s a_s/b_s \leq C$.

In the ultrahigh dimension setting $\log p \asymp n^{\theta}$, the signal to noise ratios of three tests are respectively

$$\mathrm{SNR}_{\mathrm{ZC}} \asymp (\log p)^{1/\theta} p^{-(1+2\delta)/2}, \mathrm{SNR}_{\mathrm{ZYS}} \asymp (\log p)^{1/\theta} p^{-(1+\delta)/2}, \mathrm{SNR}_{\mathrm{NEW}} \asymp (\log p)^{1/\theta} p^{-1/2}.$$

As the dimension $p \to \infty$, all the three tests will have trivial powers and cannot distinguish the alternatives from the null. The statistical intuition behind this phenomenon is that for fixed-dimensional signals, high dimensionality is a total curse and the signal to noise ratios of three tests converge to zero. Even in this case, the convergence rate of our proposed test is still slower than those of Zhong and Chen (2011)'s and Zhang et al. (2018)'s tests. We can also give the explicit order of asymptotic relative efficiency of these three tests in this case.

$$\mathrm{ARE}(\mathrm{NEW}, \mathrm{ZC}) \asymp p^{\delta}, \mathrm{ARE}(\mathrm{NEW}, \mathrm{ZYS}) \asymp p^{\delta/2}, \text{ and } \mathrm{ARE}(\mathrm{ZYS}, \mathrm{ZC}) \asymp p^{\delta/2}.$$

In terms of the asymptotic power, the performance of our proposed test is the best, followed by Zhang et al. (2018)'s test. Zhong and Chen (2011)'s test is unfortunately the worst among the three tests.

3.3 Heteroscedastic case when the number of non-zero effects is diverging

Assume that $\beta_s = \kappa I(1 \le s \le q), s = 1, \ldots, p$, for $\kappa \neq 0$, $q(\le p)$ is diverging. All other settings are remained exactly the same as those in Section 4.2.

Suppose that $q \asymp p^{\tau}$ with $0 < \tau < 1$. In the ultrahigh dimension setting of $\log p \asymp n^{\theta}$ for some $0 < \theta < 1$, the signal to noise ratios of three tests are respectively

$$\mathrm{SNR}_{\mathrm{ZC}} \asymp (\log p)^{1/\theta} p^{\tau(1+2\delta)-1/2-\delta}, \mathrm{SNR}_{\mathrm{ZYS}} \asymp (\log p)^{1/\theta} p^{\tau(1+3\delta/2)-1/2-\delta/2}, \text{ and}$$
$$\mathrm{SNR}_{\mathrm{NEW}} \asymp (\log p)^{1/\theta} p^{\tau(1+\delta)-1/2}.$$

(a) If the signals are dense, that is, in the order of $q = p^{\tau}$ with $1/2 \le \tau < 1$, all the signal to noise ratios $\mathrm{SNR}_{\mathrm{ZC}}$, $\mathrm{SNR}_{\mathrm{ZYS}}$ and $\mathrm{SNR}_{\mathrm{NEW}}$ go to infinity as $p \to \infty$. Therefore, these three tests have nontrivial power under H_1'.

(b) If the signals are sparse, that is, in the order of $q = p^{\tau}$ with $0 < \tau < 1/2$. Apparently, $\tau(1 + 2\delta) - 1/2 - \delta < 0$ in $\mathrm{SNR}_{\mathrm{ZC}}$, which implies that Zhong and Chen (2011)'s test may suffer from low power under the sparse alternatives even when the covariate is homoscedastic, which is consistent with the fact that Zhong and Chen (2011)'s test is designed to target dense alternatives. By contrast, Zhang et al. (2018)'s test may break down for $0 < \tau \le 1/3$, since $\tau(1 + 3\delta/2) - 1/2 - \delta/2 < 0$

in SNR_{ZYS} for all $\delta > 0$. Our proposed test has a nontrivial power as long as $\delta \geq (2\tau)^{-1} - 1$.

Moreover, no matter the signal is dense or sparse, the asymptotic relative efficiency of these three tests are

$$\text{ARE}(\text{NEW}, \text{ZC}) = p^{\delta(1-\tau)}, \text{ARE}(\text{NEW}, \text{ZYS}) = p^{\delta(1-\tau)/2}, \text{and } \text{ARE}(\text{ZYS}, \text{ZC}) \asymp p^{\delta(1-\tau)/2},$$

which implies that our proposed test is still asymptotically more powerful than Zhong and Chen (2011)'s and Zhang et al. (2018)'s tests.

In summary, the aforementioned power analysis suggests that compared to Zhong and Chen (2011)'s and Zhang et al. (2018)'s tests, our proposed test has a substantial efficiency gain in heteroscedastic case while maintaining high power efficiency in homoscedastic case. We shall verify this finding through numerical studies.

4 Numerical Studies

4.1 Simulation studies

We conduct simulations to evaluate the finite-sample performance of the proposed test and compare it with the two universal (n, p)-asymptotic tests proposed by Zhong and Chen (2011) and Zhang et al. (2018). In addition, we compare our proposal test with the ART (McKeague and Qian, 2015) and its related test proposed by Zhang and Laber (2015) in the supplement.

Let us consider the following three models:

$$Y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_1 + \varepsilon_i, \quad (4.1)$$

$$Y_i = 3\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_3 + \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_4/2) + \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_2 - 1)\varepsilon_i, \quad (4.2)$$

$$Y_i = (\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_5)\exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_2/\sqrt{2}) + \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_5/\sqrt{2q}) + \varepsilon_i, \quad (4.3)$$

where $\mathbf{x}_i = (X_{i1}, \ldots, X_{ip})^{\mathrm{T}}$ is generated from the following moving average model:

$$X_{is} = s^{\delta/2}\{\rho_1 Z_{is} + \rho_2 Z_{i(s+1)} + \cdots + \rho_T Z_{i(s+T-1)}\}, \quad (4.4)$$

for $\delta \geq 0$, $T = 8$ and $s = 1, \ldots, p$. Here, $\mathbf{z}_i = (Z_{i1}, \ldots, Z_{i(p+T-1)})^{\mathrm{T}}$ is $(p+T-1)$-dimensional standard normal. The coefficients $\{\rho_k\}_{k=1}^{T}$ are generated independently from the uniform distribution on $[0, 1]$ and are kept fixed once generated. The moving average model (4.4) implies that $\boldsymbol{\Sigma} = \mathrm{cov}(\mathbf{x}_i) = (\sigma_{st})_{p\times p}$, consists of

$$\sigma_{st} = (st)^{\delta/2}\sum_{k=1}^{T}\rho_k\rho_{k+|s-t|}I\{|s-t| < T\}, \text{ for } s, t = 1, \ldots, p.$$

Therefore,

$$\mathrm{var}(X_s) = s^{\delta}\sum_{k=1}^{T}\rho_k^2, \text{ for } s = 1, \ldots, p.$$

The parameter δ controls the degree of heteroscedasticity. We consider $\delta = 0, 0.25, 0.5,$ 0.75 and 1, where $\delta = 0$ indicates that all covariates are homogeneous. The error term ε_i follows two different distributions: $N(0, 1)$ and the centralized gamma distribution with shape parameter 1 and scale parameter 1, where the centralized gamma distribution is skewed to the right.

Under the null hypothesis H_0', the coefficients in all three models are all zero. Following Zhong and Chen (2011) and Zhang et al. (2018), we consider two configurations of alternative hypothesis. (a) Non-sparse case: the total number of active covariates $q = [p^{0.7}/2]$, where $[x]$ denotes the largest integer not greater than x. (b) Sparse case: the total number of active covariates $q = [3p^{0.3}/2]$. The coefficients are defined as follows.

$$
\begin{aligned}
\boldsymbol{\beta}_1 &: \ \beta_j = ||\boldsymbol{\beta}||/\sqrt{q}, \quad \text{for} \ \ j = 1, \ldots, q, \\
\boldsymbol{\beta}_2 &: \ \beta_j = 1, \qquad\qquad\ \ \text{for} \ \ j = 1, \\
\boldsymbol{\beta}_3 &: \ \beta_j = ||\boldsymbol{\beta}||/\sqrt{q}, \quad \text{for} \ \ j = 2, \ldots, [q/2] + 1, \\
\boldsymbol{\beta}_4 &: \ \beta_j = ||\boldsymbol{\beta}||/\sqrt{q}, \quad \text{for} \ \ j = [q/2] + 2, \ldots, q + 1, \\
\boldsymbol{\beta}_5 &: \ \beta_j = ||\boldsymbol{\beta}||/\sqrt{q}, \quad \text{for} \ \ j = 2, \ldots, q + 1,
\end{aligned}
$$

where $||\boldsymbol{\beta}||^2 = 0.04$. All other entries are identically 0. We choose n =80, 120 and $p = 550$, 1116, according to $p = [\exp(n^{0.4}) + 230]$. Model (4.1) is a linear regression model where the first q covariates have the same magnitude of signals. Model (4.2) is a partially linear model with heteroscedastic errors while Model (4.3) is a nonlinear model. The significance level α is fixed at 0.05 and all results are based on 1,000 Monte Carlo replications.

Table 1 reports the empirical sizes and powers of our proposed test as well as those of Zhong and Chen (2011)'s and Zhang et al. (2018)'s tests for linear model (4.1). The empirical sizes of all three tests are reasonably close to 5% under two different error distributions. We display the kernel density estimates for the standardized test statistic $T_{n,p}$ for linear model (4.1) in Figure 1, which can be well approximated by a standard normal distribution. This confirms the theoretical results in Theorem 3. Since our proposed test is scale-invariance, its empirical sizes stay the same under different values of δ. This property is not shared by Zhong and Chen (2011)'s or Zhang et al. (2018)'s test. In terms of power, when $\delta = 0$, Zhong and Chen (2011)'s test is an obvious winner, and Zhang et al. (2018)'s test is slightly better than our proposed test. The differences among three tests are not remarkable. This echoes the theoretical finding in Section 3 that our proposed test has little efficiency loss when each covariate has the same variance. However, when δ is larger than zero, the story becomes totally different. Our proposed test has the best performances, followed by Zhang et al. (2018)'s test, and then Zhong and Chen (2011)'s test. When $\delta = 1$, the empirical powers of our proposed test and the two competitors are 0.998, 0.133, 0.056 respectively under sparse H_1 with $(n, p) = (120, 1116)$ and normal errors. Under this scenario, our proposed test significantly outperforms the competitors while the empirical power of Zhong and Chen (2011)'s test is close to the significance level. This implies that even under the linear models, our proposed test has substantial efficiency gain compared with the two competitors for the heteroscedastic covariates, in accordance with asymptotic power comparison in Section 3. Compared to non-sparse alternatives, all three tests have power reductions under sparse ones.

Tables 2 and 3 summarize the results of all three tests for models (4.2) and (4.3). The sizes of all three tests are satisfactory regardless of δ, (n, p) and error distributions. Under alternatives, Zhong and Chen (2011)'s test and Zhang et al. (2018)'s test gradually break down as δ increases. By contrast, our proposed test remains valid for a wide range of δ. When $\delta = 1$, the empirical powers of our test arrive at 0.922 in model (4.2) and 0.911 in model (4.3) under non-sparse H_1 with $(n,p) = (120, 1116)$ and normal errors. Similar to the phenomenon in linear model, our scale-invariance test is still much more powerful for nonlinear models with heteroscedastic covariates.

Table 1: The empirical sizes and powers for linear model (4.1) at the significance level 5%, where δ controls the degree of heterogeneity in terms of the covariate variances. Throughout our numerical studies, we refer to our proposed test, and the tests proposed by Zhong and Chen (2011) and Zhang et al. (2018) as NEW, ZC and ZYS, respectively.

			Normal error			Gamma error		
(n,p)	Hypothesis	δ	ZC	ZYS	NEW	ZC	ZYS	NEW
$(80,550)$	H_0	0.00	0.054	0.050	0.045	0.052	0.054	0.059
		0.25	0.054	0.048	0.045	0.059	0.059	0.059
		0.50	0.053	0.052	0.045	0.064	0.065	0.059
		0.75	0.062	0.052	0.045	0.066	0.063	0.059
		1.00	0.067	0.053	0.045	0.060	0.066	0.059
	Non-sparse H_1	0.00	0.772	0.735	0.718	0.783	0.757	0.735
		0.25	0.665	0.830	0.934	0.688	0.834	0.933
		0.50	0.422	0.833	0.996	0.458	0.846	0.991
		0.75	0.240	0.771	1.000	0.260	0.783	1.000
		1.00	0.140	0.627	1.000	0.152	0.638	1.000
	Sparse H_1	0.00	0.522	0.488	0.470	0.523	0.515	0.497
		0.25	0.218	0.371	0.652	0.228	0.390	0.638
		0.50	0.115	0.271	0.795	0.099	0.270	0.779
		0.75	0.072	0.197	0.911	0.079	0.184	0.905
		1.00	0.065	0.138	0.965	0.068	0.134	0.972
$(120,1116)$	H_0	0.00	0.047	0.048	0.044	0.052	0.049	0.052
		0.25	0.043	0.041	0.044	0.057	0.052	0.052
		0.50	0.042	0.044	0.044	0.058	0.053	0.052
		0.75	0.039	0.046	0.044	0.064	0.057	0.052
		1.00	0.039	0.047	0.044	0.063	0.057	0.052
	Non-sparse H_1	0.00	0.849	0.814	0.797	0.842	0.811	0.794
		0.25	0.731	0.918	0.981	0.749	0.909	0.979
		0.50	0.466	0.912	1.000	0.471	0.918	0.999
		0.75	0.246	0.830	1.000	0.251	0.836	1.000
		1.00	0.139	0.655	1.000	0.150	0.661	1.000
	Sparse H_1	0.00	0.670	0.612	0.593	0.643	0.620	0.602
		0.25	0.231	0.452	0.796	0.242	0.452	0.796
		0.50	0.101	0.303	0.933	0.110	0.304	0.941
		0.75	0.063	0.190	0.982	0.079	0.201	0.989
		1.00	0.056	0.133	0.998	0.063	0.131	1.000

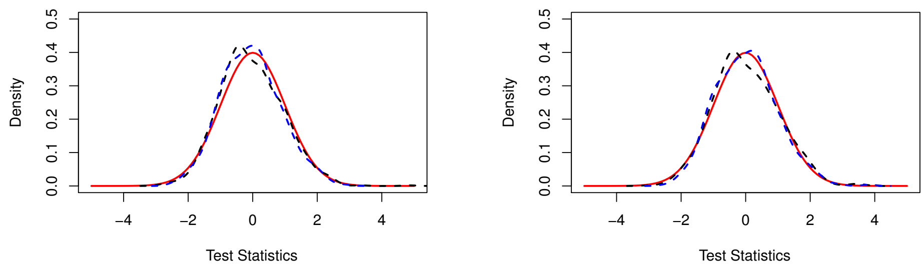

Figure 1: Density curves of the asymptotic null distribution of our test statistic under $(n, p) = (80, 550)$ (black dashed line) and $(n, p) = (120, 1116)$ (blue dashed line) compared with the standard normal distribution (red solid line).

4.2 An application

We apply our proposed test to a gene expression data set. The cardiomyopathy microarray data comes from a study by Redfern et al. (2000), and has been analyzed by several important researchers, including Segal et al. (2003), Hall and Miller (2009), Li et al. (2012) and Shao and Zhang (2014). It contains 6,319 gene expression values from 30 mice. Redfern et al. (2000) reported that the overexpression of G protein-coupled receptor Ro1 in hearts of adult mice would lead to a lethal dilated cardiomyopathy. This finding helps geneticists look into the etiology of human disease. We test whether the gene set contributes to expression level of Ro1.

Figure 2 shows the standard deviation of each gene expression level, which ranges from 17.34 to 18,437.96. This implies that the variation of gene differs substantially from each other. We divide the whole dataset into two subsets with $n_1 = 16$ and $n_2 = 14$. On the first subset, we follow Li et al. (2012) and Shao and Zhang (2014) to screen out unimportant genes by marginally testing the conditional mean independence between the expression levels of each gene and Ro1. The Benjamini-Hochberg

Table 2: The empirical sizes and powers for partially linear model (4.2) at the significance level 5%, where δ controls the degree of heterogeneity in terms of the covariate variances. Refer to the captions in Table 1 for abbreviations.

			Normal error			Gamma error		
(n,p)	Hypothesis	δ	ZC	ZYS	NEW	ZC	ZYS	NEW
$(80,550)$	H_0	0.00	0.054	0.050	0.045	0.052	0.054	0.059
		0.25	0.054	0.048	0.045	0.059	0.059	0.059
		0.50	0.053	0.052	0.045	0.064	0.065	0.059
		0.75	0.062	0.052	0.045	0.066	0.063	0.059
		1.00	0.067	0.053	0.045	0.060	0.066	0.059
	Non-sparse H_1	0.00	0.646	0.684	0.683	0.684	0.728	0.714
		0.25	0.487	0.690	0.814	0.479	0.715	0.837
		0.50	0.233	0.634	0.903	0.243	0.644	0.924
		0.75	0.128	0.504	0.951	0.143	0.506	0.964
		1.00	0.072	0.321	0.952	0.070	0.329	0.967
	Sparse H_1	0.00	0.343	0.391	0.380	0.354	0.396	0.402
		0.25	0.131	0.289	0.474	0.130	0.306	0.505
		0.50	0.064	0.195	0.561	0.058	0.193	0.608
		0.75	0.046	0.136	0.648	0.048	0.133	0.713
		1.00	0.044	0.107	0.749	0.037	0.104	0.804
$(120,1116)$	H_0	0.00	0.047	0.048	0.044	0.052	0.049	0.052
		0.25	0.043	0.041	0.044	0.057	0.052	0.052
		0.50	0.042	0.044	0.044	0.058	0.053	0.052
		0.75	0.039	0.046	0.044	0.064	0.057	0.052
		1.00	0.039	0.047	0.044	0.063	0.057	0.052
	Non-sparse H_1	0.00	0.669	0.700	0.691	0.687	0.747	0.733
		0.25	0.483	0.725	0.862	0.520	0.749	0.870
		0.50	0.261	0.672	0.939	0.276	0.720	0.946
		0.75	0.125	0.533	0.973	0.133	0.546	0.976
		1.00	0.064	0.275	0.922	0.047	0.266	0.906
	Sparse H_1	0.00	0.450	0.479	0.468	0.478	0.523	0.523
		0.25	0.159	0.339	0.600	0.162	0.381	0.639
		0.50	0.068	0.211	0.727	0.064	0.233	0.751
		0.75	0.055	0.148	0.830	0.047	0.151	0.846
		1.00	0.050	0.109	0.903	0.045	0.099	0.906

procedure is applied to control the false discovery rate at 0.001. After the screening procedure, the tests proposed by Zhong and Chen (2011) and Zhang et al. (2018) and our proposal retain 145, 79 and 163 genes respectively.

Table 3: The empirical sizes and powers for nonlinear model (4.3) at the significance level 5%, where δ controls the degree of heterogeneity in terms of the covariate variances. Refer to the captions in Table 1 for abbreviations.

			Normal error			Gamma error		
(n,p)	Hypothesis	δ	ZC	ZYS	NEW	ZC	ZYS	NEW
$(80,550)$	H_0	0.00	0.054	0.050	0.045	0.052	0.054	0.059
		0.25	0.054	0.048	0.045	0.059	0.059	0.059
		0.50	0.053	0.052	0.045	0.064	0.065	0.059
		0.75	0.062	0.052	0.045	0.066	0.063	0.059
		1.00	0.067	0.053	0.045	0.060	0.066	0.059
	Non-sparse H_1	0.00	0.735	0.762	0.753	0.740	0.766	0.743
		0.25	0.429	0.650	0.831	0.443	0.662	0.846
		0.50	0.215	0.512	0.862	0.206	0.530	0.890
		0.75	0.101	0.371	0.877	0.105	0.375	0.912
		1.00	0.056	0.272	0.884	0.080	0.266	0.922
	Sparse H_1	0.00	0.638	0.695	0.670	0.672	0.707	0.667
		0.25	0.226	0.464	0.741	0.212	0.448	0.751
		0.50	0.087	0.288	0.801	0.079	0.263	0.797
		0.75	0.051	0.170	0.827	0.055	0.160	0.845
		1.00	0.042	0.110	0.847	0.048	0.116	0.862
$(120,1116)$	H_0	0.00	0.047	0.048	0.044	0.052	0.049	0.052
		0.25	0.043	0.041	0.044	0.057	0.052	0.052
		0.50	0.042	0.044	0.044	0.058	0.053	0.052
		0.75	0.039	0.046	0.044	0.064	0.057	0.052
		1.00	0.039	0.047	0.044	0.063	0.057	0.052
	Non-sparse H_1	0.00	0.794	0.816	0.805	0.804	0.820	0.803
		0.25	0.434	0.689	0.871	0.430	0.696	0.881
		0.50	0.207	0.512	0.906	0.179	0.528	0.905
		0.75	0.107	0.378	0.911	0.090	0.348	0.914
		1.00	0.075	0.271	0.911	0.055	0.225	0.922
	Sparse H_1	0.00	0.732	0.767	0.744	0.733	0.764	0.740
		0.25	0.199	0.471	0.809	0.202	0.461	0.815
		0.50	0.088	0.246	0.847	0.077	0.263	0.851
		0.75	0.056	0.158	0.877	0.054	0.155	0.870
		1.00	0.050	0.116	0.882	0.043	0.107	0.881

We then compare the power performances of these three tests. We randomly pick 6, 7, 8, 9 and 10 samples from the second subset of data and test the overall effects of selected genes. Each testing procedure is repeated 1,000 times and the empirical power is reported in Table 4. The powers of all tests gradually approach to one

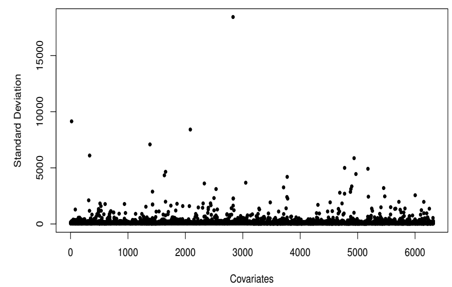

Figure 2: The standard deviation of each covariate.

as the sample size increases, and our proposed test outperforms both competitors significantly, particularly when the sample size is small.

Table 4: The empirical powers for ZC, ZYS tests and our proposed test.

random samples	ZC	ZYS	NEW
6	0.363	0.105	0.463
7	0.432	0.328	0.731
8	0.569	0.632	0.912
9	0.692	0.845	0.980
10	0.831	0.975	1.000

5 Discussion

In this paper, we develop a new test to examine the effects of high-dimensional covariates on the response without any model assumptions. Our test statistic is built on the cumulative covariance, which has an explicit form and is completely free of tuning parameters. The limiting distributions of our proposed test statistic are normal under both the null hypothesis and the alternatives. Our asymptotic power analysis and numerical studies show that even under the high-dimensional linear models, our proposed test has substantial power improvement compared to the tests of Zhong and

Chen (2011) and Zhang et al. (2018) in the heteroscedastic covariates setting, while maintaining high efficiency in the homoscedastic cases. It is also important to remark here that our proposed testing procedure can be easily generalized to multivariate response case. For the multivariate response $\mathbf{y} \in \mathbb{R}^q$ $(q > 1)$, where q is allowed to be large but finite, we can analogously define

$$E\Big[\mathrm{cov}\{\mathbf{y}^{\mathrm{T}}, I(X < \widetilde{X}) \mid \widetilde{X}\}\mathrm{cov}\{\mathbf{y}, I(X < \widetilde{X}) \mid \widetilde{X}\}\Big],$$

where $(\widetilde{X}, \widetilde{\mathbf{y}})$ is an independent copy of $(X, \mathbf{y})$. This metric reduces to (2.3) in the univariate case of $q = 1$. The corresponding test statistic is further defined by

$$\{4(n)_5\}^{-1} \sum_{s=1}^{p} \sum_{(i,j,k,l,r)}^{n} (\mathbf{y}_i - \mathbf{y}_j)^{\mathrm{T}}(\mathbf{y}_k - \mathbf{y}_l)\psi(X_{is}, X_{js}, X_{rs})\psi(X_{ks}, X_{ls}, X_{rs}).$$

The computationally efficient algorithm and theoretical analysis for our test statistic in (2.5) can be directly applied to the above statistic. It is also worth noting that our test procedure is built on a sum-of-squares-based statistic and targets dense alternatives. To enhance its power for sparse signals, we suggest to follow the ideas of McKeague and Qian (2015), Fan et al. (2015), Chen et al. (2019) and Zheng et al. (2019) to construct test statistics, which deserve further investigations.

REFERENCE

Chen, S.X., Li, J., and Zhong, P.S. (2019). "Two-sample and anova tests for high dimensional means." *The Annals of Statistics*, **3**, 1443–1474.

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

Efron, B. and Tibshirani, R. (2007). "On testing the significance of sets of genes." *The Annals of Applied Statistics*, **1(1)**, 107–129.

Escanciano, J.C. (2006). "A consistent diagnostic test for regression models using projections." *Econometric Theory*, **22**, 1030–1051.

Fan, J., Liao, Y., and Yao, J. (2015). "Power enhancement in high-dimensional cross-sectional tests." *Econometrica*, **83(4)**, 1497–1541.

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

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

Goeman, J.J., Van Houwelingen, H.C., and Finos, L. (2011). "Testing against a high-dimensional alternative in the generalized linear model: asymptotic type i error control." *Biometrika*, **98(2)**, 381–390.

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

Guo, X., Wang, T., and Zhu, L. (2016). "Model checking for parametric single-index models: a dimension reduction model-adaptive approach." *Journal of the Royal Statistical Society, Series B*, **78(5)**, 1013–1035.

Hall, P. (1984). "Central limit theorem for integrated square error of multivariate nonparametric density estimators." *Journal of multivariate analysis*, **14(1)**, 1–16.

Hall, P. and Heyde, C. (2014). *Martingale Limit Theory and Its Application*. New York: Academic Press.

Hall, P. and Miller, H. (2009). "Using generalized correlation to effect variable selection in very high dimensional problems." *Journal of Computational and Graphical Statistics*, **18(3)**, 533–550.

Hardle, W. and Mammen, E. (1993). "Comparing nonparametric versus parametric regression fits." *The Annals of Statistics*, **21(4)**, 1926–1947.

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

Luedtke, A.R. and van der Laan, M.J. (2018). "Parametric-rate inference for one-sided differentiable parameters." *Journal of the American Statistical Association*, **113(522)**, 780–788.

McKeague, I.W. and Qian, M. (2015). "An adaptive resampling test for detecting the presence of significant predictors (with discussions)." *Journal of the American Statistical Association*, **110(512)**, 1422–1462.

Nettleton, D., Recknor, J., and Reecy, J.M. (2008). "Identification of differentially expressed gene categories in microarray studies using nonparametric multivariate analysis." *Bioinformatics*, **24(2)**, 192–201.

Paindaveine, D. and Verdebout, T. (2016). "On high-dimensional sign tests." *Bernoulli*, **22**, 1745–1769.

Patilea, V., Sánchez-Sellero, C., and Saumard, M. (2016). "Testing the predictor effect on a functional response." *Journal of the American Statistical Association*, **111(516)**, 1684–1695.

Redfern, C.H., Degtyarev, M.Y., Kwa, A.T., Salomonis, N., Cotte, N., Nanevicz, T., Fidelman, N., Desai, K., Vranizan, K., Lee, E.K., Coward, P., Shah, N., Warrington, J.A., Fishman, G.I., Bernstein, D., Baker, A.J., and Conklin, B.R. (2000). "Conditional expression of a gi-coupled receptor causes ventricular conduction delay and a lethal cardiomyopathy." *Proceedings of the National Academy of Sciences*, **97(9)**, 4826–4831.

Segal, M.R., Dahlquist, K.D., and Conklin, B.R. (2003). "Regression approaches for microarray data analysis." *Journal of Computational Biology*, **10(6)**, 961–980.

Serfling, R.J. (1980). *Approximation theorems of mathematical statistics*. New York: Wiley.

Shao, X. and Zhang, J. (2014). "Martingale difference correlation and its use in high-dimensional variable screening." *Journal of the American Statistical Association*, **109(507)**, 1302–1318.

Stute, W. (1997). "Nonparametric model checks for regression." *The Annals of Statistics*, **25(2)**, 613–641.

Stute, W., Thies, S., and Zhu, L.X. (1998). "Model checks for regression: an innovation process approach." *The Annals of Statistics*, **26(5)**, 1916–1934.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles." *Proceedings of the National Academy of Sciences*, **102(43)**, 15545–15550.

Székely, G.J., Rizzo, M.L., and Bakirov, N.K. (2007). "Measuring and testing dependence by correlation of distances." *The Annals of Statistics*, **35(6)**, 2769–2794.

Wang, L. and Akritas, M.G. (2006). "Testing for covariate effects in the fully nonparametric analysis of covariance model." *Journal of the American Statistical Association*, **101(474)**, 722–736.

Zhang, X., Yao, S., and Shao, X. (2018). "Conditional mean and quantile dependence testing in high dimension." *The Annals of Statistics*, **46(1)**, 219–246.

Zhang, Y. and Laber, E.B. (2015). "Comment: an adaptive resampling test for detecting the presence of signifficant predictors." *Journal of the American Statistical Association*, **110(512)**, 1451–1454.

Zheng, J.X. (1996). "A consistent test of functional form via nonparametric estimation techniques." *Journal of Econometrics*, **75(2)**, 263–289.

Zheng, S., Cheng, G., Guo, J., and Zhu, H. (2019). "Test for high-dimensional correlation matrices." *The Annals of Statistics*, **47**, 2887–2921.

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

Zhou, T., Zhu, L., Xu, C., and Li, R. (2020). "Model-free forward screening via cumulative divergence." *Journal of the American Statistical Association*, **115(531)**, 1393–1405.

Supplement to "Testing the effects of high-dimensional covariates via aggregating cumulative covariances"

S.1. Proposition 1 and Its Proof.

To justify the unbiasedness of $T_{n,p}$ in high dimension, we provide a simple and equivalent expression for $\mathrm{CCov}(Y \mid X)$ in (2.3).

Proposition 1. *Let $\{(\widetilde{X}_i, \widetilde{Y}_i), i = 1, \dots 5\}$ be the five independent copies of (X, Y). Assume that $var(X) > 0$, $0 < var(Y) < \infty$ and X is continuous, then $\mathrm{CCov}(Y \mid X)$ can be expressed as*

$$E\big[(\widetilde{Y}_1 - \widetilde{Y}_3)(\widetilde{Y}_2 - \widetilde{Y}_4)\psi(\widetilde{X}_1, \widetilde{X}_3, \widetilde{X}_5)\psi(\widetilde{X}_2, \widetilde{X}_4, \widetilde{X}_5)\big]/4, \tag{S.1}$$

where the function $\psi(\cdot, \cdot, \cdot)$ is defined as

$$\psi(X_1, X_2, X_3) = I(X_1 < X_3) - I(X_2 < X_3).$$

In contrast to the standard version in (2.3), this expression in (S.1) plays an important role in constructing a scale-invariance test statistic for high-dimensional conditional mean independence without bias correction.

Proof: Using the equality $\mathrm{cov}\{Y, I(X < x)\} = E\big[(\widetilde{Y}_1 - \widetilde{Y}_2)\{I(\widetilde{X}_1 < x) - I(\widetilde{X}_2 < x)\}\big]/2$ for any $x \in \mathbb{R}^1$, we obtain

$$\begin{aligned} \mathrm{cov}^2\{Y, I(X < x)\} &= E\big[(\widetilde{Y}_1 - \widetilde{Y}_2)(\widetilde{Y}_3 - \widetilde{Y}_4)\{I(\widetilde{X}_1 < x) - I(\widetilde{X}_2 < x)\} \\ &\quad \times \{I(\widetilde{X}_3 < x) - I(\widetilde{X}_4 < x)\}\big]/4. \end{aligned}$$

By the law of iterated expectations, we complete the proof. $\square$

S.2. Proof of Theorem 1

It is noted that

$$
\begin{aligned}
&\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)\{I(X_{is}<X_{js})-F_{n,s}(X_{js})\}\\
=\ &(2n)^{-1}\sum_{i=1}^{n}\sum_{k=1}^{n}(Y_i-Y_k)\{I(X_{is}<X_{js})-I(X_{ks}<X_{js})\},
\end{aligned}
$$

for $j=1,\ldots,n$ and $s=1,\ldots,p$. Define

$$
\begin{aligned}
J_1 &= \sum_{s=1}^{p}\sum_{(i,j,k,l,r)}^{n}(Y_i-Y_j)(Y_k-Y_l)\psi(X_{is},X_{js},X_{rs})\psi(X_{ks},X_{ls},X_{rs}),\\
J_2 &= \sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}(Y_i-Y_j)(Y_k-Y_l)\psi(X_{is},X_{js},X_{is})\psi(X_{ks},X_{ls},X_{is}),\\
J_3 &= \sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}(Y_i-Y_j)(Y_i-Y_k)\psi(X_{is},X_{js},X_{ls})\psi(X_{is},X_{ks},X_{ls}),\\
J_4 &= \sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}(Y_i-Y_j)(Y_j-Y_k)\psi(X_{is},X_{js},X_{is})\psi(X_{js},X_{ks},X_{is}),\\
J_5 &= \sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}(Y_i-Y_j)(Y_j-Y_k)\psi(X_{is},X_{js},X_{js})\psi(X_{js},X_{ks},X_{js}),\\
J_6 &= \sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}(Y_i-Y_j)^2\psi^2(X_{is},X_{js},X_{ks}),\\
J_7 &= \sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}(Y_i-Y_j)^2\psi^2(X_{is},X_{js},X_{js}).
\end{aligned}
$$

Then, after straightforward but laborious computations, we have

$$
\begin{aligned}
W_{n,p} &= n^{-5}(J_1/4+J_2+J_3-2J_4-2J_5-J_6/2+2J_7),\\
T_{n,p} &= \{4(n)_5\}^{-1}J_1.
\end{aligned}
$$

To derive an efficient algorithm for $T_{n,p}$, it suffices to analyze $W_{n,p}$, J_2, J_3, J_4, J_5, J_6 and J_7, respectively. For any $s=1,\ldots,p$, sort the n observations of this covariate $\{X_{is}:i=1,\ldots,n\}$ to be $X_{(1)s}<\ldots<X_{(n)s}$. Next, find the corresponding response

$Y_{(i)s}$ associated with $X_{(i)s}$. Denote $\dot{Y}_{(i)s} = Y_{(i)s} - \overline{Y}$ with $\overline{Y} = n^{-1}\sum_{i=1}^{n} Y_i$. Following the computational algorithm in Section 3.2 of Zhu et al. (2010), we have

$$W_{n,p} = n^{-3}\sum_{s=1}^{p}\sum_{j=1}^{n}\big(\sum_{i<j}^{n}\dot{Y}_{(i)s}\big)^2.$$

Then, we turn to deal with the terms J_2, J_3, J_4, J_5, J_6 and J_7. Denote $\dot{Y}_i = Y_i - \overline{Y}_i$, it can be shown that

$$\begin{aligned} J_2 &= \sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}(\dot{Y}_i - \dot{Y}_j)(\dot{Y}_k - \dot{Y}_l)I(X_{js} < X_{is})I(X_{ls} < X_{is}) \\ &\quad - \sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}(\dot{Y}_i - \dot{Y}_j)(\dot{Y}_k - \dot{Y}_l)I(X_{js} < X_{is})I(X_{ks} < X_{is}) \\ &= 2n\sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\dot{Y}_j\dot{Y}_k I(X_{js} < X_{is})I(X_{ks} < X_{is}) \\ &\quad -2n\sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\dot{Y}_i\dot{Y}_k I(X_{js} < X_{is})I(X_{ks} < X_{is}) \\ &= 2n\sum_{s=1}^{p}\sum_{j=1}^{n}\left(\sum_{i=1}^{j-1}\dot{Y}_{(i)s}\right)^2 - 2n\sum_{s=1}^{p}\sum_{j=1}^{n}\left\{(j-1)\dot{Y}_{(j)s}\sum_{i=1}^{j-1}\dot{Y}_{(i)s}\right\}, \end{aligned}$$

where the second equality follows from the fact that $\sum_{i=1}^{n}\dot{Y}_i = 0$. Similarly, it follows that

$$\begin{aligned} J_3 &= n^2\sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\dot{Y}_i^2 I(X_{is} < X_{js}) - 2n\sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\dot{Y}_i^2 I(X_{is} < X_{ks})I(X_{js} < X_{ks}) \\ &\quad +3n\sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\dot{Y}_i\dot{Y}_j I(X_{is} < X_{ks})I(X_{js} < X_{ks}) \\ &\quad +\sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\dot{Y}_i^2 I(X_{js} < X_{ls})I(X_{ks} < X_{ls}) \\ &= \sum_{s=1}^{p}\sum_{j=1}^{n}\left[\left\{n^2 - 2n(j-1)\right\}\sum_{i=1}^{j-1}\dot{Y}_{(i)s}^2\right] + 3n\sum_{s=1}^{p}\sum_{j=1}^{n}\left(\sum_{i=1}^{j-1}\dot{Y}_{(i)s}\right)^2 \end{aligned}$$

$$+\{(n-1)n(2n-1)/6\}\sum_{s=1}^{p}\sum_{i=1}^{n}\dot{Y}_{(i)s}^2.$$

Moreover, we have

$$\begin{aligned}
J_4 &= n\sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\dot{Y}_i^2 I(X_{is}<X_{js}) + \sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\dot{Y}_i\dot{Y}_j I(X_{is}<X_{ks})I(X_{js}<X_{ks}) \\
&- n\sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\dot{Y}_i\dot{Y}_j I(X_{is}<X_{js}) - \sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\dot{Y}_i^2 I(X_{is}<X_{ks})I(X_{js}<X_{ks}) \\
&= \sum_{s=1}^{p}\sum_{j=1}^{n}\left(\sum_{i=1}^{j-1}\dot{Y}_{(i)s}\right)^2 + \sum_{s=1}^{p}\sum_{j=1}^{n}\left\{(n-j+1)\sum_{i=1}^{j-1}\dot{Y}_{(i)s}^2\right\} - n\sum_{s=1}^{p}\sum_{j=1}^{n}\left\{\dot{Y}_{(j)s}\sum_{i=1}^{j-1}\dot{Y}_{(i)s}\right\}.
\end{aligned}$$

Similar calculation shows that

$$\begin{aligned}
J_5 &= \sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\dot{Y}_k^2 I(X_{is}<X_{ks})I(X_{js}<X_{ks}) \\
&\quad -2\sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\dot{Y}_i\dot{Y}_k I(X_{is}<X_{ks})I(X_{js}<X_{ks}) \\
&\quad +\sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\dot{Y}_i\dot{Y}_j I(X_{is}<X_{ks})I(X_{js}<X_{ks}) \\
&= \sum_{s=1}^{p}\sum_{i=1}^{n}\left\{(i-1)^2\dot{Y}_{(i)s}^2\right\} - 2\sum_{s=1}^{p}\sum_{j=1}^{n}\left\{(j-1)\dot{Y}_{(j)s}\sum_{i=1}^{j-1}\dot{Y}_{(i)s}\right\} + \sum_{s=1}^{p}\sum_{j=1}^{n}\left(\sum_{i=1}^{j-1}\dot{Y}_{(i)s}\right)^2.
\end{aligned}$$

Furthermore, it can be shown that

$$\begin{aligned}
J_6 &= 2\sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\dot{Y}_i^2 I(X_{is}<X_{js}) + 2\sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\dot{Y}_j^2 I(X_{is}<X_{ks}) \\
&\quad -4\sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\dot{Y}_i^2 I(X_{is}<X_{ks})I(X_{js}<X_{ks}) \\
&\quad +4\sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\dot{Y}_i\dot{Y}_j I(X_{is}<X_{ks})I(X_{js}<X_{ks})
\end{aligned}$$

$$= \quad 2\sum_{s=1}^{p}\sum_{j=1}^{n}\left\{(n-2j+2)\sum_{i=1}^{j-1}\dot{Y}_{(i)s}^2\right\} + n(n-1)\sum_{s=1}^{p}\sum_{i=1}^{n}\dot{Y}_{(i)s}^2 + 4\sum_{s=1}^{p}\sum_{j=1}^{n}\left(\sum_{i=1}^{j-1}\dot{Y}_{(i)s}\right)^2,$$

and

$$J_7 = \sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\dot{Y}_i^2 I(X_{is} < X_{js}) + \sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\dot{Y}_j^2 I(X_{is} < X_{js}) = n\sum_{s=1}^{p}\sum_{i=1}^{n}\dot{Y}_{(i)s}^2.$$

Putting the above together, we complete the proof. □

S.3. Proof of Lemma 1

It is easy to see that

$$\begin{aligned} & E\{I(X_{1s} < X_{5s})I(X_{2s} < X_{5s}) \mid \mathbf{z}_1, \mathbf{z}_2\} \\ = \quad & 1 - |F_s(X_{1s}) - F_s(X_{2s})|/2 - F_s(X_{1s})/2 - F_s(X_{2s})/2, \end{aligned}$$

where $F_s(\cdot)$ is the cumulative distribution function of X_s for $s = 1, \ldots, p$. Based on this, we have

$$\begin{aligned} & E\{\psi(X_{1s}, X_{3s}, X_{5s})\psi(X_{2s}, X_{4s}, X_{5s}) \mid \mathbf{z}_1, \ldots, \mathbf{z}_4\} \\ = \quad & \{-|F_s(X_{1s}) - F_s(X_{2s})| - |F_s(X_{3s}) - F_s(X_{4s})| \\ & +|F_s(X_{1s}) - F_s(X_{4s})| + |F_s(X_{2s}) - F_s(X_{3s})|\}/2 \\ = \quad & \big[K_1\{F_s(X_{1s}), F_s(X_{2s})\} + K_1\{F_s(X_{3s}), F_s(X_{4s})\} \\ & -K_1\{F_s(X_{1s}), F_s(X_{4s})\} - K_1\{F_s(X_{2s}), F_s(X_{3s})\}\big]/2, \end{aligned} \tag{S.2}$$

where

$$\begin{aligned} & K_1\{F_s(X_{1s}), F_s(X_{2s})\} = 2E\{\psi(X_{1s}, X_{3s}, X_{5s})\psi(X_{2s}, X_{4s}, X_{5s}) \mid \mathbf{z}_1, \mathbf{z}_2\} \\ & = -|F_s(X_{1s}) - F_s(X_{2s})| - F_s(X_{1s}) - F_s(X_{2s}) + F_s^2(X_{1s}) + F_s^2(X_{2s}) + 2/3, \end{aligned}$$

is the double centered version of $-|F_s(X_{1s})-F_s(X_{2s})|$ and satisfies that $E[K_1\{F_s(X_{1s}),F_s(X_{2s})\} \mid \mathbf{z}_1] = E[K_1\{F_s(X_{1s}),F_s(X_{2s})\} \mid \mathbf{z}_2] = 0$. By the useful property of the double centered distance, it can be shown that the projections of $\widetilde{h} = 5h/3$ are,

$$
\begin{aligned}
\widetilde{h}^{(1)}(\mathbf{z}_1) &= \sum_{s=1}^{p}\Big\{L_s(Y_1,X_{1s}) + 2\mathrm{CCov}(Y \mid X_s)\Big\}/4,\\
\widetilde{h}^{(2)}(\mathbf{z}_1,\mathbf{z}_2) &= \sum_{s=1}^{p}\Big[K_0(Y_1,Y_2)K_1\{F_s(X_{1s}),F_s(X_{2s})\} + 2\mathrm{CCov}(Y \mid X_s)\\
&\quad +2L_s(Y_1,X_{1s}) + 2L_s(Y_2,X_{2s}) - L_s(Y_1,X_{2s}) - L_s(Y_2,X_{1s})\Big]/12,\\
\widetilde{h}^{(3)}(\mathbf{z}_1,\mathbf{z}_2,\mathbf{z}_3) &= \sum_{s=1}^{p}\sum_{(i,j,k)}^{3}\Big\{K_0(Y_i,Y_j)\Big[2K_1\{F_s(X_{is}),F_s(X_{js})\}\\
&\quad -K_1\{F_s(X_{is}),F_s(X_{ks})\} - K_1\{F_s(X_{js}),F_s(X_{ks})\}\Big]\\
&\quad +2L_s(Y_i,X_{is}) - L_s(Y_i,X_{js}) - L_s(Y_i,X_{ks})\Big\}/48,\\
\widetilde{h}^{(4)}(\mathbf{z}_1,\mathbf{z}_2,\mathbf{z}_3,\mathbf{z}_4) &= \sum_{s=1}^{p}\sum_{(i,j,k,l)}^{4}\Big\{K_0(Y_i,Y_j) + K_0(Y_k,Y_l)\Big\}\\
&\quad \times\Big[2K_1\{F_s(X_{is}),F_s(X_{js})\} + 2K_1\{F_s(X_{ks}),F_s(X_{ls})\}\\
&\quad -K_1\{F_s(X_{is}),F_s(X_{ks})\} - K_1\{F_s(X_{is}),F_s(X_{ls})\}\\
&\quad -K_1\{F_s(X_{js}),F_s(X_{ks})\} - K_1\{F_s(X_{js}),F_s(X_{ls})\}\Big]/192,
\end{aligned}
$$

where $K_0(Y_1,Y_2) = (Y_1-EY)(Y_2-EY)$ and $L_s(Y_1,X_{1s}) = E[K_0(Y_1,Y_2)\times K_1\{F_s(X_{1s}), F_s(X_{2s})\} \mid \mathbf{z}_1]$. Moreover, $\widetilde{h}^{(5)}(\mathbf{z}_1,\mathbf{z}_2,\mathbf{z}_3,\mathbf{z}_4,\mathbf{z}_5) = \sum_{s=1}^{p}\widetilde{h}_s(1,2,3,4,5)$.

Under either H_0' or H_1', it is straightforward to verify that

$$
\xi_1 = \mathrm{var}\{h^{(1)}(\mathbf{z}_1)\} = O\left[\mathrm{var}\left\{\sum_{s=1}^{p} L_s(Y_1,X_{1s})\right\}\right] = o(n^{-1}S^2).
$$

To obtain the variance of $h^{(2)}(\mathbf{z}_1,\mathbf{z}_2)$, we first show that

$$
\begin{aligned}
& \operatorname{var}\left[\sum_{s=1}^{p} K_1\{F_s(X_{1s}), F_s(X_{2s})\}\right] \\
= & \sum_{s=1}^{p}\sum_{t=1}^{p} \operatorname{cov}\left[K_1\{F_s(X_{1s}), F_s(X_{2s})\}, K_1\{F_t(X_{1t}), F_t(X_{2t})\}\right] \\
= & \frac{1}{4}\sum_{s=1}^{p}\sum_{t=1}^{p} E\Big(\big[K_1\{F_s(X_{1s}), F_s(X_{2s})\} + K_1\{F_s(X_{3s}), F_s(X_{4s})\} \\
& -K_1\{F_s(X_{1s}), F_s(X_{4s})\} - K_1\{F_s(X_{2s}), F_s(X_{3s})\}\big]\big[K_1\{F_t(X_{1t}), F_t(X_{2t})\} \\
& +K_1\{F_t(X_{3t}), F_t(X_{4t})\} - K_1\{F_t(X_{1t}), F_t(X_{4t})\} - K_1\{F_t(X_{2t}), F_t(X_{3t})\}\big]\Big) \\
= & \sum_{s=1}^{p}\sum_{t=1}^{p} E\left\{\psi(X_{1s}, X_{3s}, X_{5s})\psi(X_{2s}, X_{4s}, X_{5s})\psi(X_{1t}, X_{3t}, X_{6t})\psi(X_{2t}, X_{4t}, X_{6t})\right\},
\end{aligned}
$$

which further equals

$$
\begin{aligned}
& 4\sum_{s=1}^{p}\sum_{t=1}^{p} E[\operatorname{cov}^2\{I(X_{1s} < X_{2s}), I(X_{1t} < X_{3t}) \mid X_{2s}, X_{3t}\}] \\
= & 4\sum_{s=1}^{p}\sum_{t=1}^{p} \mathrm{BKR}(X_s, X_t) \geq 4\sum_{s=1}^{p} \mathrm{BKR}(X_s, X_s) = 2p/45 \to \infty, \qquad \text{(S.3)}
\end{aligned}
$$

as $p \to \infty$. Here $\mathrm{BKR}(X_s, X_t) = \int\int\{F_{st}(u,v) - F_s(u)F_t(v)\}^2 dF_s(u)dF_t(v)$ is Blum-Kiefer-Rosenblatt coefficient between X_s and X_t introduced by Blum et al. (1961), and $F_{st}(\cdot,\cdot)$, $F_s(\cdot)$ and $F_t(\cdot)$ denote the joint and marginal cumulative distributions of X_s and X_t, respectively. Under Assumption 1, it directly follows that $S^2 \to \infty$ and $S^2 \asymp \sum_{s=1}^{p}\sum_{t=1}^{p} \mathrm{BKR}(X_s, X_t) \geq p/90$, as $p \to \infty$. Recalling the expression of $h^{(2)}(\mathbf{z}_1, \mathbf{z}_2)$, we then calculate the difference between ξ_2 and $S^2/100$.

$$
\begin{aligned}
|\xi_2 - S^2/100| \leq & \operatorname{var}\Big[\sum_{s=1}^{p}\{L_s(Y_1, X_{1s}) + L_s(Y_2, X_{2s}) - L_s(Y_1, X_{2s})/2 \\
& -L_s(Y_2, X_{1s})/2\}\Big]/36 + \Big|\operatorname{cov}\Big[\sum_{s=1}^{p} K_0(Y_1, Y_2)K_1\{F_s(X_{1s}), F_s(X_{2s})\}, \\
& +\sum_{s=1}^{p}\{L_s(Y_1, X_{1s}) + L_s(Y_2, X_{2s}) - L_s(Y_1, X_{2s})/2 - L_s(Y_2, X_{1s})/2\}\Big]\Big|/6
\end{aligned}
$$

$$
\begin{aligned}
&\leq \left[\mathrm{var}\left\{\sum_{s=1}^{p} L_s(Y_1, X_{1s})\right\} + \mathrm{var}\left\{\sum_{s=1}^{p} L_s(Y_1, X_{2s})\right\}\right]/23 \\
&+ (2^{3/2}S/3)\left[\mathrm{var}\left\{\sum_{s=1}^{p} L_s(Y_1, X_{1s})\right\} + \mathrm{var}\left\{\sum_{s=1}^{p} L_s(Y_1, X_{2s})\right\}\right]^{1/2},
\end{aligned}
$$

where the last inequality holds due to Hölder's inequality. Consequently, under either H_0' or H_1', we have $|\xi_2 - S^2/100| = o(S^2)$. This, together with (Serfling, 1980, Lemma 5.2.1 A) and the fact that $S^2 \to \infty$ as $p \to \infty$, implies

$$
\xi_2 = 100^{-1}S^2\{1 + o(1)\}, \quad \text{as} \ \ p \to \infty.
$$

Again, by Hölder's inequality, we obtain

$$
\begin{aligned}
\xi_3 \ &\leq \ (6^5/48^2)\Bigg(16S^2 + 8\mathrm{var}\left[\sum_{s=1}^{p} K_0(Y_1, Y_2)K_1\{F_s(X_{1s}), F_s(X_{3s})\}\right] \\
&\quad +4\mathrm{var}\left\{\sum_{s=1}^{p} L_s(Y_1, X_{1s})\right\} + 2\mathrm{var}\left\{\sum_{s=1}^{p} L_s(Y_1, X_{2s})\right\}\Bigg) \\
&\leq \ (6^5/48^2)\left(16S^2\{1 + o(1)\} + 8\mathrm{var}\left[\sum_{s=1}^{p} K_0(Y_1, Y_2)K_1\{F_s(X_{1s}), F_s(X_{3s})\}\right]\right) \\
&\leq \ (6^5/48^2)\left[16S^2\{1 + o(1)\} + 8c\sum_{s=1}^{p}\sum_{t=1}^{p} \mathrm{BKR}(X_s, X_t)\right] \leq CS^2,
\end{aligned}
$$

where the second inequality follows from the condition (2.7) and the third inequality holds by Assumption 1. Here and in what follows, the notations C and c are generic constants, which may take different values at each appearance. Because $2\xi_3/3 \geq \xi_2 = 100^{-1}S^2\{1 + o(1)\}$, we have that ξ_2 and ξ_3 are of the same order.

Similarly, it can be shown that

$$
\begin{aligned}
\xi_4 &\leq (8^{23}/192^2)\Bigg(32S^2 + 8\mathrm{var}\left[\sum_{s=1}^{p} K_0(Y_1, Y_2)K_1\{F_s(X_{1s}), F_s(X_{3s})\}\right] \\
&\quad +8\mathrm{var}\{K_0(Y_1, Y_2)\}\mathrm{var}\left[\sum_{s=1}^{p} K_1\{F_s(X_{1s}), F_s(X_{3s})\}\right]\Bigg) \\
&\leq (8^{23}/192^2)\left\{32S^2 + 8c\sum_{s=1}^{p}\sum_{t=1}^{p}\mathrm{BKR}(X_s, X_t)\right\} \leq CS^2,
\end{aligned}
$$

where the second inequality follows from (S.3) and Assumption 1. Because $3\xi_4/4 \geq \xi_3 \asymp S^2$, we also have that ξ_3 and ξ_4 are of the same order.

Using similar arguments those in the derivation of (S.3), it follows by Assumption 1 and Minkowski's inequality that

$$
\begin{aligned}
\xi_5 &\leq C\mathrm{var}\{\sum_{s=1}^{p}\psi(X_{1s}, X_{3s}, X_{4s})\psi(X_{2s}, X_{5s}, X_{6s})\} \\
&= 4C\sum_{s=1}^{p}\sum_{t=1}^{p}\mathrm{BKR}(X_s, X_t).
\end{aligned}
$$

This, together with the inequality $4\xi_5/5 \geq \xi_4$, gives that ξ_4 and ξ_5 are of the same order. Hence, the third, fourth and fifth terms in the Hoeffding decomposition are all of smaller order.

By the definitions of $C(n, m)$ and $(m)_m$, we can easily see that

$$
\{C(n, 5)\}^{-1}C(5, c)C(n-5, 5-c) = n^{-c}\{C(5, c)\}^2\{(c)_c\}\{1 + o(1)\},
$$

for $5 \geq c \geq 1$. Under either H_0' or H_1', we apply the Hoeffding decomposition for $\mathrm{var}(T_{n,p})$ to obtain

$$
\mathrm{var}(T_{n,p}) = \{n(n-1)/2\}^{-1}S^2\{1 + o(1)\}.
$$

This completes the proof of Lemma 1. $\square$

S.4. Proof of Theorem 2

Under H_0', it is straightforward to show that $h^{(1)}(\mathbf{z}_1) = \sum_{s=1}^{p} \Big\{ L(Y_1, X_{1s}) + 2\text{CCov}(Y \mid X_s) \Big\}/4 = 0$. Then, the projection of h to two-dimensional sample spaces is

$$h^{(2)}(\mathbf{z}_1, \mathbf{z}_2) = \sum_{s=1}^{p} K_0(Y_1, Y_2) K_1\{F_s(X_{1s}), F_s(X_{2s})\}/20.$$

Let $\widehat{T}_{n,p}$ be the projection of $T_{n,p}$, where

$$\widehat{T}_{n,p} = 10\{n(n-1)/2\}^{-1} \sum_{1 \le i < j \le n}^{n} h^{(2)}(\mathbf{z}_i, \mathbf{z}_j).$$

We can decompose $T_{n,p} = \widehat{T}_{n,p} + T_{n,p} - \widehat{T}_{n,p}$, where $T_{n,p} - \widehat{T}_{n,p}$ can still be written as a U-statistics with kernel

$$g(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4, \mathbf{z}_5) = h(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4, \mathbf{z}_5) - \sum_{1 \le i_1 < i_2 \le 5} h^{(2)}(\mathbf{z}_{i_1}, \mathbf{z}_{i_2}).$$

By the useful property of the double centered distance, it can be shown that the projections of g are,

$$\begin{aligned} g^{(1)}(\mathbf{z}_1) &= 0, \\ g^{(2)}(\mathbf{z}_1, \mathbf{z}_2) &= 0, \\ g^{(3)}(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3) &= h^{(3)}(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3) - \sum_{1 \le i_1 < i_2 \le 3} h^{(2)}(\mathbf{z}_{i_1}, \mathbf{z}_{i_2}), \\ g^{(4)}(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4) &= h^{(4)}(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4) - \sum_{1 \le i_1 < i_2 \le 3} h^{(2)}(\mathbf{z}_{i_1}, \mathbf{z}_{i_2}) \\ &\quad - \sum_{1 \le i_1 < i_2 < i_3 \le 4} h^{(3)}(\mathbf{z}_{i_1}, \mathbf{z}_{i_2}, \mathbf{z}_{i_2}). \end{aligned}$$

Under the null hypothesis H_0', $\text{var}\{g^{(1)}(\mathbf{z}_1)\} = \text{var}\{g^{(2)}(\mathbf{z}_1, \mathbf{z}_2)\} = 0$. By the Hoeffding's variance formula in Lemma 1, $\text{var}(\widehat{T}_{n,p}) = O[n^{-2}\text{var}\{h^{(2)}(\mathbf{z}_1, \mathbf{z}_2)\}]$ and $\text{var}(T_{n,p} - \widehat{T}_{n,p}) = o[n^{-2}\text{var}\{h^{(2)}(\mathbf{z}_1, \mathbf{z}_2)\}]$, which follows the fact that $\text{var}\{h^{(2)}\}$, $\text{var}\{h^{(3)}\}$, $\text{var}\{h^{(4)}\}$

and $\text{var}\{h^{(5)}\}$ are of the same order. We further obtain that

$$T_{n,p}\text{var}^{-1/2}(\widehat{T}_{n,p}) = \widehat{T}_{n,p}\text{var}^{-1/2}(\widehat{T}_{n,p}) + o_P(1).$$

From (Serfling, 1980, Lemma 2.5.1A), we have

$$\text{var}(\widehat{T}_{n,p}) = \{n(n-1)/2\}^{-1}S^2.$$

Hence we only need to show that

$$\{n(n-1)/2\}^{1/2}\widehat{T}_{n,p}/S \xrightarrow{D} N(0,1),$$

where $\xrightarrow{D}$ denotes convergence in distribution. According to the definitions of $\widehat{T}_{n,p}$ and $h^{(2)}(\mathbf{z}_1, \mathbf{z}_2)$, $\widehat{T}_{n,p}/S$ can also be written as

$$\{n(n-1)/2\}^{-1}\sum_{i<j}^{n}\sum_{s=1}^{p} K_0(Y_i, Y_j)K_1\{F_s(X_{is}), F_s(X_{js})\}/(2S).$$

To prove the theorem, we only need to show

$$\{n(n-1)/2\}^{-1/2}\sum_{i<j}^{n}\sum_{s=1}^{p} K_0(Y_i, Y_j)K_1\{F_s(X_{is}), F_s(X_{js})\}/(2S) \xrightarrow{D} N(0,1).$$

Write $\widetilde{T}_{n,k} = \sum_{i=2}^{k} Z_{ni}$, where

$$Z_{ni} = \sum_{j=1}^{i-1}\sum_{s=1}^{p} K_0(Y_i, Y_j)K_1\{F_s(X_{is}), F_s(X_{js})\}/\left\{2n(n-1)\right\}^{1/2}.$$

Let $\mathcal{F}_i = \sigma\{(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_i, \mathbf{y}_i)\}$ be the σ-field generated by $\{(\mathbf{x}_j, \mathbf{y}_j), j \le i\}$. It is easy to see that $E(Z_{ni} \mid \mathcal{F}_{i-1}) = 0$ and it follows that $\{\widetilde{T}_{n,k}, \mathcal{F}_k : 2 \le k \le n\}$ is a zero mean martingale. Let $v_{ni} = E(Z_{ni}^2 \mid \mathcal{F}_{i-1})$, $2 \le i \le n$, and $V_n = \sum_{i=2}^{n} v_{ni}$. The central limit theorem will hold (Hall (1984)) if we can verify

$$V_n/\text{var}(\widetilde{T}_{n,p}) \xrightarrow{pr} 1, \tag{S.4}$$

and for any $\varepsilon > 0$

$$\sum_{i=1}^{n} S^{-2} E\{Z_{ni}^2 I(Z_{ni} > \varepsilon S) \mid \mathcal{F}_{i-1}\} \xrightarrow{pr} 0, \tag{S.5}$$

where $\xrightarrow{pr}$ denotes convergence in probability. It can be shown that $v_{ni} = v_{ni}^{(1)} + v_{ni}^{(2)}$, where

$$\begin{aligned} v_{ni}^{(1)} &= \{2n(n-1)\}^{-1} \sum_{j=1}^{i-1} \sum_{s=1}^{p} \sum_{t=1}^{p} E[K_0^2(Y_i, Y_j) K_1\{F_s(X_{is}), F_s(X_{js})\} \\ &\qquad K_1\{F_t(X_{it}), F_t(X_{jt})\} \mid \mathbf{z}_j], \end{aligned}$$

and

$$\begin{aligned} v_{ni}^{(2)} &= \{n(n-1)\}^{-1} \sum_{1 \le j < k \le i-1} \sum_{s=1}^{p} \sum_{t=1}^{p} E[K_0(Y_i, Y_j) K_1\{F_s(X_{is}), F_s(X_{js})\} \\ &\qquad K_0(Y_i, Y_k) K_1\{F_t(X_{it}), F_t(X_{kt})\} \mid \mathbf{z}_j, \mathbf{z}_k]. \end{aligned}$$

By Fubini's theorem,

$$\begin{aligned} &\mathrm{var}\left\{ \sum_{i=2}^{n} v_{ni}^{(1)} / \mathrm{var}(\widetilde{T}_{n,p}) \right\} \\ =\ & \left\{2n(n-1)S^2\right\}^{-2} \mathrm{var}\Bigg(\sum_{j=1}^{n} \sum_{s=1}^{p} \sum_{t=1}^{p} (n-j) E[K_0^2(Y, Y_j) \\ & K_1\{F_s(X_s), F_s(X_{js})\} K_1\{F_t(X_t), F_t(X_{jt})\} \mid \mathbf{z}_j] \Bigg) \\ =\ & \left\{2n(n-1)S^2\right\}^{-2} \sum_{j=1}^{n} (n-j)^2 \mathrm{var}\Bigg(\sum_{s=1}^{p} \sum_{t=1}^{p} E[K_0^2(Y, Y_j) \\ & K_1\{F_s(X_s), F_s(X_{js})\} K_1\{F_t(X_t), F_t(X_{jt})\} \mid \mathbf{z}_j] \Bigg) \\ \le\ & c \left\{n(n-1)S^2\right\}^{-2} \sum_{j=1}^{n} (n-j)^2 E\{V(\mathbf{x}_1, \mathbf{x}_2)^4\} \\ =\ & o(1) n^{-3} \sum_{j=1}^{n} (n-j)^2, \end{aligned}$$

where the first inequality follows from Assumption 1 and Hölder inequality, and the second inequality holds due to Assumption 2. Together the fact $\sum_{j=1}^{n}(n-j)^2 = O(n^3)$, we obtain that

$$\text{var}\left\{\sum_{i=2}^{n} v_{ni}^{(1)}/\text{var}(\widetilde{T}_{n,p})\right\} = o(1).$$

As $E(v_{ni}) = \{2n(n-1)\}^{-1}(i-1)\text{var}\left[\sum_{s=1}^{p} K_0(Y_1, Y_2)K_1\{F_s(X_{1s}), F_s(X_{2s})\}\right]$, it follows immediately that $E\left\{\sum_{i=2}^{n} v_{ni}^{(1)}/\text{var}(\widetilde{T}_{n,p})\right\} = 1$. By Markov's inequality,

$$\sum_{i=2}^{n} v_{ni}^{(1)}/\text{var}(\widetilde{T}_{n,p}) \xrightarrow{pr} 1.$$

Similar discussions can be performed on the term $\sum_{i=2}^{n} v_{ni}^{(2)}/\text{var}(\widetilde{T}_{n,p})$. It is essential to obtain $E\{\sum_{i=2}^{n} v_{ni}^{(2)}/\text{var}(\widetilde{T}_{n,p})\} = 0$ and

$$\begin{aligned}
& \text{var}\left\{\sum_{i=2}^{n} v_{ni}^{(2)}/\text{var}(\widetilde{T}_{n,p})\right\} \\
= & \left\{n(n-1)S^2\right\}^{-2}\text{var}\Bigg(\sum_{1\le j<k\le n}\sum_{s=1}^{p}\sum_{t=1}^{p}(n-k)E[K_0(Y, Y_j) \\
& K_1\{F_s(X_s), F_s(X_{js})\}K_0(Y, Y_k)K_1\{F_t(X_t), F_t(X_{kt})\} \mid \mathbf{z}_j, \mathbf{z}_k]\Bigg) \\
\le & c\left\{n(n-1)S^2\right\}^{-2}\sum_{j=1}^{n}(j-1)(n-j)^2 \\
& \times E\{V(\mathbf{x}_1, \mathbf{x}_2)V(\mathbf{x}_2, \mathbf{x}_3)V(\mathbf{x}_3, \mathbf{x}_4)V(\mathbf{x}_4, \mathbf{x}_1)\}.
\end{aligned}$$

Thus, Assumption 2 and the fact $\sum_{j=1}^{n}(j-1)(n-j)^2 = O(n^4)$ together yield

$$\sum_{i=2}^{n} v_{ni}^{(2)}/\text{var}(\widetilde{T}_{n,p}) \xrightarrow{pr} 0.$$

In summary, (S.4) holds.

It remains to show (S.5). Since

$$0 \le E\{Z_{ni}^2 I(Z_{ni} > \varepsilon S) \mid \mathcal{F}_{i-1}\} \le \varepsilon^{-2} S^{-2} E(Z_{ni}^4 \mid \mathcal{F}_{i-1}),$$

we only need to prove that

$$\sum_{i=1}^{n} E(Z_{ni}^4) = o(S^4). \tag{S.6}$$

It is noted that for every integer $m \ge 4$,

$$\left(\sum_{i=1}^{m} x_i\right)^4 = \sum_{i=1}^{m} x_i^4 + 4\sum_{i\neq j}^{m} x_i^3 x_j + 3\sum_{i\neq j}^{m} x_i^2 x_j^2 + 6\sum_{i\neq j\neq k\neq i}^{m} x_i^2 x_j x_k + \sum_{i\neq j\neq k\neq l\neq i}^{m} x_i x_j x_k x_l.$$

Through simple calculation, we have

$$\begin{aligned}\sum_{i=1}^{n} E(Z_{ni}^4) &\le cn^{-2} E\left[\sum_{s=1}^{p} K_1\{F_s(X_{is}), F_s(X_{js})\}\right]^4 \\ &\quad + Cn^{-1} E^2\left[\sum_{s=1}^{p} K_1\{F_s(X_{is}), F_s(X_{js})\}\right]^2 \\ &= cn^{-2} E\{V(\mathbf{x}_1, \mathbf{x}_2)^4\} + 4Cn^{-1} S^4.\end{aligned}$$

Under Assumption 2, (S.6) follows immediately. This completes the proof. □

S.5. Proof of Theorem 3

Denote the infeasible variance estimator for S^2 by

$$\check{S}_{n,p}^2 = \{4c_n n(n-1)\}^{-1} \sum_{(i,j)}^{n} K_0(\dot{Y}_i, \dot{Y}_j)^2 \left[\sum_{s=1}^{p} K_1\{F_s(X_{is}), F_s(X_{js})\}\right]^2,$$

where $\dot{Y}_i = Y_i - \overline{Y}$, and $c_n = \{(1-n^{-1})^2 + n^{-2}\}^2$. By Slutsky's theorem, we only need to prove

$$S_{n,p}^2/\check{S}_{n,p}^2 - 1 \overset{pr}{\longrightarrow} 0, \tag{S.7}$$

$$\check{S}_{n,p}^2/S^2 - 1 \overset{pr}{\longrightarrow} 0. \tag{S.8}$$

We divide the proof into the following two steps.

Step 1: We first aim at proving (S.7).

Applying the triangle inequality and the boundedness of empirical distribution function, we obtain

$$|S_{n,p}^2 - \check{S}_{n,p}^2|/S^2 \le c\widehat{\Delta}_1,$$

where

$$\begin{aligned}\widehat{\Delta}_1 &= \{n(n-1)\}^{-1}\sum_{i\neq j}^{n}\sum_{s=1}^{p}\Big\{|F_{n,s}(X_{is}) - F_s(X_{is})| \\ &\quad + |F_{n,s}(X_{js}) - F_s(X_{js})|\Big\}K_0(\dot{Y}_i, \dot{Y}_j)^2/S^2.\end{aligned}$$

It is noted that $\widehat{\Delta}_1 \ge 0$. To establish (S.7), it suffices to show $E(\widehat{\Delta}_1) = o(1)$.

Under Assumptions 1 and 2, it can be verified that

$$E(\widehat{\Delta}_1 \mid \mathbf{x}_1, \cdots, \mathbf{x}_p) \le Cn^{-1}\sum_{i=1}^{n}\sum_{s=1}^{p} \mid F_{n,s}(X_{is}) - F_s(X_{is}) \mid S^{-2}.$$

Due to Dvoretzky-Kiefer-Wolfowitz inequality (Kosorok, 2008, Theorem 11.6), we have

$$n^{-1}\sum_{i=1}^{n}\sum_{s=1}^{p}|F_{n,s}(X_{is}) - F_s(X_{is})| = O_P\left[p\{\log(n)/n\}^{1/2}\right],$$

which yields

$$E(\widehat{\Delta}_1) = O\left[\{\log(n)/n\}^{1/2} pS^{-2}\right].$$

It is noted that

$$Cp = C\sum_{s=1}^{p} \mathrm{BKR}(X_s, X_s) \le S^2.$$

Then $E(\widehat{\Delta}_1) = o(1)$ is immediate.

Step 2: We turn to prove (S.8) in this step.

Without loss of generality, we assume $E(Y) = 0$, because $\dot{Y}_i$ is invariant under location shifts. For any $i \ne j$, it follows by definition that $\dot{Y}_i = (1-n^{-1})Y_i - n^{-1}Y_j - n^{-1}\sum_{k\notin\{i,j\}}^{n} Y_k$. Direct calculation yields that

$$\begin{aligned} K_0(\dot{Y}_i, \dot{Y}_j) &= \{(1-n^{-1})^2 + n^{-2}\}Y_iY_j - n^{-1}(1-n^{-1})(Y_i^2 + Y_j^2) \\ &- n^{-1}(1-2n^{-1})(Y_i + Y_j)\sum_{k\notin\{i,j\}}^{n} Y_k + n^{-2}\sum_{k\notin\{i,j\}}^{n}\sum_{l\notin\{i,j\}}^{n} Y_kY_l. \end{aligned} \quad \text{(S.9)}$$

By the independence of Y_i and Y_j, these four terms on the right-hand sides of (S.9) are uncorrelated with each other. Recalling (S.3) and the notation $c_n = \{(1-n^{-1})^2 + n^{-2}\}^2$, we obtain

$$E(\check{S}_{n,p}^2)/S^2 = 1 + o(1). \quad \text{(S.10)}$$

Again, it follows from (S.9) that

$$\check{S}_{n,p}^2/S^2 \le c(\widehat{\Delta}_2 + \widehat{\Delta}_3 + \widehat{\Delta}_4),$$

where $\widehat{\Delta}_2$, $\widehat{\Delta}_3$ and $\widehat{\Delta}_4$ are defined as follows,

$$\begin{aligned}
\widehat{\Delta}_2 &= \{n(n-1)\}^{-1}\sum_{i\neq j}^{n}\left[\sum_{s=1}^{p}K_1\{F_s(X_{is}),F_s(X_{js})\}\right]^2 K_0(Y_i,Y_j)^2/S^2,\\
\widehat{\Delta}_3 &= \{n^2(n-1)\}^{-1}\sum_{i\neq j}^{n}\sum_{k\notin\{i,j\}}^{n}\left[\sum_{s=1}^{p}K_1\{F_s(X_{is}),F_s(X_{js})\}\right]^2 |K_0(Y_i,Y_j)K_0(Y_i,Y_k)|/S^2,\\
\widehat{\Delta}_4 &= \{n^3(n-1)\}^{-1}\sum_{i\neq j}^{n}\sum_{k\notin\{i,j\}}^{n}\sum_{l\notin\{i,j\}}^{n}\left[\sum_{s=1}^{p}K_1\{F_s(X_{is}),F_s(X_{js})\}\right]^2 |K_0(Y_i,Y_j)K_0(Y_k,Y_l)|/S^2.
\end{aligned}$$

Since $K_1\{F_s(X_{is}),F_s(X_{js})\}$ is double centered, we have $E[K_1\{F_s(X_{is}),F_s(X_{js})\} \mid X_{1s}] = E[K_1\{F_s(X_{is}),\ F_s(X_{js})\} \mid X_{2s}] = 0$. From (S.3), we obtain $E[K_1\{F_s(X_{1s}), F_s(X_{2s})\}K_1\{F_t(X_{1t}),F_t(X_{2t})\}] = \int\int\{F_{st}(u,v)-F_s(u)F_t(v)\}^2 dF_s(u)dF_t(v) = \mathrm{BKR}(X_s,X_t)$. These results, together with Assumptions 1 and 2, give

$$\begin{aligned}
\mathrm{var}(\check{S}^2_{n,p}/S^2) &\leq 4c^2\{\mathrm{var}(\widehat{\Delta}_2/S^2)+\mathrm{var}(\widehat{\Delta}_3/S^2)+\mathrm{var}(\widehat{\Delta}_3/S^2)\}\\
&\leq CE\{V(\mathbf{x}_1,\mathbf{x}_2)V(\mathbf{x}_2,\mathbf{x}_3)V(\mathbf{x}_3,\mathbf{x}_4)V(\mathbf{x}_4,\mathbf{x}_1)\}/E^2\{V(\mathbf{x}_1,\mathbf{x}_2)^2\}\\
&\quad +E\{V(\mathbf{x}_1,\mathbf{x}_2)^4\}/[nE^2\{V(\mathbf{x}_1,\mathbf{x}_2)^2\}] = o(1). \qquad (S.11)
\end{aligned}$$

The statement for (S.8) follows from (S.10), (S.11) and Slutsky's theorem.

Combining the results in Step 1 and 2, the proof of ratio consistency is completed. Moreover, based on the results in Theorem 2 and Slutsky's theorem, we have under the H_0',

$$\{n(n-1)/2\}^{1/2}\,T_{n,p}/S_{n,p} \xrightarrow{D} N(0,1).$$

□

S.6. Proof of Theorem 4

Since $T_{n,p}$ is a U-statistic, it follows by the proof of Lemma 1 that under H_1',

$$\begin{aligned}
& \{n(n-1)/2\}^{1/2}\, S^{-1}[T_{n,p} - \sum_{s=1}^{p} \mathrm{CCov}(Y \mid X_s)] \\
= & \{n(n-1)/2\}^{-1/2}\, (2S)^{-1} \sum_{i<j}^{n} \sum_{s=1}^{p} [K_0(Y_i, Y_j) K_1\{F_s(X_{is}), F_s(X_{js})\} \\
- & L_s(Y_i, X_{is}) - L_s(Y_j, X_{js}) + \mathrm{CCov}(Y \mid X_s)] + o_P(1).
\end{aligned}$$

Following the notation of Hall (1984), we denote

$$\begin{aligned}
H_n(\mathbf{z}_i, \mathbf{z}_j) &= \{n(n-1)/2\}^{-1/2}\, (2S)^{-1} \sum_{s=1}^{p} [K_0(Y_i, Y_j) K_1\{F_s(X_{is}), F_s(X_{js})\} \\
&- L_s(Y_i, X_{is}) - L_s(Y_j, X_{js}) + \mathrm{CCov}(Y \mid X_s)].
\end{aligned}$$

By the above definition, it is not hard to verify that

$$E\{H_n(\mathbf{z}_1, \mathbf{z}_2) \mid \mathbf{z}_1\} = 0.$$

Under H_1', we further have

$$(n^2/2) E\{H_n(\mathbf{z}_1, \mathbf{z}_2)^2\} = 1 + O\left[\mathrm{var}\left\{\sum_{s=1}^{p} L_s(Y_1, X_{1s})\right\} S^{-2}\right] = 1 + o(1).$$

To establish the asymptotic normality of $T_{n,p}$, it suffices to verify the condition (2.1) in Theorem 1 of Hall (1984), namely,

$$\frac{E\{G_n(\mathbf{z}_1, \mathbf{z}_2)^2\} + n^{-1} E\{H_n(\mathbf{z}_1, \mathbf{z}_2)^4\}}{E^2\{H_n(\mathbf{z}_1, \mathbf{z}_2)^2\}} \to 0,$$

as $n, p \to \infty$, where $G_n(\mathbf{z}_1, \mathbf{z}_2) = E\{H_n(\mathbf{z}_3, \mathbf{z}_1) H_n(\mathbf{z}_3, \mathbf{z}_2) \mid \mathbf{z}_1, \mathbf{z}_2\}$. Under Assumptions 1 and 2, and following the proof of Lemma 1, we can show that

$$\begin{aligned}
E\{G_n(\mathbf{z}_1,\mathbf{z}_2)^2\} &\leq Cn^{-4}E\{V(\mathbf{x}_1,\mathbf{x}_2)V(\mathbf{x}_2,\mathbf{x}_3)V(\mathbf{x}_3,\mathbf{x}_4)V(\mathbf{x}_4,\mathbf{x}_1)\}/S^4 \\
&\quad +Cn^{-4}\mathrm{var}^2\left\{\sum_{s=1}^{p} L_s(Y_1,X_{2s})\right\}/S^4, \\
E^2\{H_n(\mathbf{z}_1,\mathbf{z}_2)^2\} &= 4n^{-4}\{1+o(1)\}, \\
n^{-1}E\{H_n(\mathbf{z}_1,\mathbf{z}_2)^4\} &\leq Cn^{-5}E\{V(\mathbf{x}_1,\mathbf{x}_2)^4\}/S^4 \\
&\quad +Cn^{-4}\mathrm{var}^2\left\{\sum_{s=1}^{p} L_s(Y_1,X_{2s})\right\}/S^4.
\end{aligned}$$

By the boundness of distribution function and Assumption 2, it follows that

$$\begin{aligned}
E\{G_n(\mathbf{z}_1,\mathbf{z}_2)^2\}/E^2\{H_n(\mathbf{z}_1,\mathbf{z}_2)^2\} &\to 0, \quad \text{and} \\
n^{-1}E\{H_n(\mathbf{z}_1,\mathbf{z}_2)^4\}/E^2\{H_n(\mathbf{z}_1,\mathbf{z}_2)^2\} &\to 0,
\end{aligned}$$

as $n, p \to \infty$. Therefore, all assumptions in Theorem 1 in Hall (1984) are satisfied with the kernel $H_n(\mathbf{z}_1,\mathbf{z}_2)$ in his Equation (2.1). This completes the proof of this theorem. □

S.7. Fast Algorithms for $\mathrm{ZC}_{n,p}$ and $\mathrm{ZYS}_{n,p}$

By applying the idea of Theorem 1, we introduce fast algorithms to calculate the numerators of standardized ZC and ZYS statistics: $\widehat{\sigma}_1\mathrm{ZC}_{n,p}$ and $\widehat{\sigma}_2\mathrm{ZYS}_{n,p}$. For any $s = 1,\ldots,p$, sort the n observation of $\{X_{is} : i = 1,\ldots,n\}$ to be $X_{(1)s} < \ldots < X_{(n)s}$ and find the corresponding response $Y_{(i)s}$ associated with $X_{(i)s}$. Denote $\dot{Y}_i = Y_i - \overline{Y}$, $\dot{Y}_{(i)s} = Y_{(i)s} - \overline{Y}$, $\dot{X}_{is} = X_{is} - \overline{X}_s$ and $\dot{X}_{(i)s} = X_{(i)s} - \overline{X}_s$, where $\overline{Y} = n^{-1}\sum_{i=1}^{n} Y_i$ and

$\overline{X}_s = n^{-1}\sum_{i=1}^{n} X_{is}$. Observe that $\sum_{i=1}^{n}\dot{Y}_i = \sum_{i=1}^{n}\dot{X}_{is} = 0$. Then we can express $\widehat{\sigma}_1\mathrm{ZC}_{n,p}$ as

$$
\begin{aligned}
\widehat{\sigma}_1\mathrm{ZC}_{n,p} &= \{4(n)_4\}^{-1}\sum_{s=1}^{p}\sum_{(i,j,k,l)}^{n}(Y_i - Y_j)(Y_k - Y_l)(X_{is} - X_{js})(X_{ks} - X_{ls}) \\
&= (n)_2^{-1}\sum_{s=1}^{p}\sum_{(i,j)}^{n}\dot{Y}_i\dot{Y}_j\dot{X}_{is}\dot{X}_{js} + (n)_4^{-1}\sum_{s=1}^{p}\sum_{(i,j,k,l)}^{n}\dot{Y}_i\dot{Y}_j\dot{X}_{ks}\dot{X}_{ls} \\
&\quad -2(n)_3^{-1}\sum_{s=1}^{p}\sum_{(i,j,k)}^{n}\dot{Y}_i\dot{Y}_j\dot{X}_{is}\dot{X}_{ks} \\
&= \{(n)_2^{-1} + 2(n)_4^{-1} + 2(n)_3^{-1}\}\sum_{s=1}^{p}(\sum_{i=1}^{n}\dot{Y}_i\dot{X}_{is})^2 \\
&\quad -\{(n)_2^{-1} + 6(n)_4^{-1} + 4(n)_3^{-1}\}\sum_{s=1}^{p}\sum_{i=1}^{n}\dot{Y}_i^2\dot{X}_{is}^2 \\
&\quad +(n)_4^{-1}\sum_{s=1}^{p}(\sum_{i=1}^{n}\dot{Y}_i^2)(\sum_{i=1}^{n}\dot{X}_{is}^2),
\end{aligned}
$$

which only requires $O(np)$ operations. By contrast, $\widehat{\sigma}_2\mathrm{ZYS}_{n,p}$ can be expressed

$$
\begin{aligned}
\widehat{\sigma}_2\mathrm{ZYS}_{n,p} &= \{4(n)_4\}^{-1}\sum_{s=1}^{p}\sum_{(i,j,k,l)}^{n}(Y_i - Y_j)(Y_k - Y_l)(X_{is} - X_{js})(X_{ks} - X_{ls}) \\
&= -(n)_2^{-1}\sum_{s=1}^{p}\sum_{(i,j)}^{n}\dot{Y}_i\dot{Y}_j|\dot{X}_{is} - \dot{X}_{js}| - (n)_4^{-1}\sum_{s=1}^{p}\sum_{(i,j,k,l)}^{n}\dot{Y}_i\dot{Y}_j|\dot{X}_{ks} - \dot{X}_{ls}| \\
&\quad +2(n)_3^{-1}\sum_{s=1}^{p}\sum_{(i,j,k)}^{n}\dot{Y}_i\dot{Y}_j|\dot{X}_{is} - \dot{X}_{ks}| \\
&= -\{(n)_2^{-1} + 2(n)_4^{-1} + 2(n)_3^{-1}\}\sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\dot{Y}_i\dot{Y}_j|\dot{X}_{is} - \dot{X}_{js}| \\
&\quad -\{4(n)_4^{-1} + 2(n)_3^{-1}\}\sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}\dot{Y}_i^2|\dot{X}_{is} - \dot{X}_{js}| \\
&\quad +(n)_4^{-1}(\sum_{i=1}^{n}\dot{Y}_i^2)\sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n}|\dot{X}_{is} - \dot{X}_{js}|.
\end{aligned}
$$

Note that $|\dot{X}_{is}-\dot{X}_{js}|=2(\dot{X}_{is}-\dot{X}_{js})I(\dot{X}_{is}>\dot{X}_{js})-(\dot{X}_{is}-\dot{X}_{js})$. Employing the similar arguments to those for dealing with $T_{n,p}$, we obtain that

$$\begin{aligned}
\widehat{\sigma}_2\mathrm{ZYS}_{n,p} \quad &= -\{2(n)_2^{-1}+4(n)_4^{-1}+4(n)_3^{-1}\}\sum_{s=1}^{p}\{\sum_{i=1}^{n}\dot{Y}_{(i)s}\dot{X}_{(i)s}\sum_{j=1}^{i-1}\dot{Y}_{(j)s}\} \\
&+ \{2(n)_2^{-1}+4(n)_4^{-1}+4(n)_3^{-1}\}\sum_{s=1}^{p}\{\sum_{i=1}^{n}\dot{Y}_{(i)s}\sum_{j=1}^{i-1}\dot{Y}_{(j)s}\dot{X}_{(j)s}\} \\
&- \{8(n)_4^{-1}+4(n)_3^{-1}\}\sum_{s=1}^{p}\{\sum_{i=1}^{n}(i-1)\dot{Y}_{(i)s}^2\dot{X}_{(i)s}-\sum_{i=1}^{n}\dot{Y}_{(i)s}^2\sum_{j=1}^{i-1}\dot{X}_{(j)s}\} \\
&+ \{4n(n)_4^{-1}+2n(n)_3^{-1}\}\sum_{s=1}^{p}\sum_{i=1}^{n}\dot{Y}_i^2\dot{X}_{is}+2(n)_4^{-1}(\sum_{i=1}^{n}\dot{Y}_i^2)\sum_{s=1}^{p}\sum_{i=1}^{n}(i-1)\dot{X}_{(i)s} \\
&- 2(n)_4^{-1}(\sum_{i=1}^{n}\dot{Y}_i^2)\sum_{s=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{i-1}\dot{X}_{(j)s},
\end{aligned}$$

which can be computed at order $O\{np\log(n)\}$.

S.8. Further Discussion on Asymptotic Relative Efficiency

The condition $p=o(\min\{\sum_{s=1}^{p}d_s^2,\sum_{s=1}^{p}d_s\})$ or $\sum_{s=1}^{p}d_s=o(\sum_{s=1}^{p}d_s^2)$ is a sufficient, but not necessary condition for our proposed test to be more powerful than the ZC and ZYS tests. If the marginal variance of each covariate is at the same magnitude (not necessarily the same), our proposed test may still have better power performances than the ZC and ZYS tests. We consider the case where $0<c_1\le d_s\le c_2<\infty$, for bounded constants c_1 and c_2, and $s=1,\ldots,p$. Then we have the following inequalities for asymptotic relative efficiency:

$$\begin{aligned}
&\left\{15/(2\pi^2)\right\}^{1/2}c_1^2/c_2^2\le \mathrm{ARE}(\mathrm{NEW},\mathrm{ZC})\le\left\{15/(2\pi^2)\right\}^{1/2}c_2^2/c_1^2,\\
&\left\{30(1-\sqrt{3}+\pi/3)/\pi^2\right\}^{1/2}c_1^{3/2}/c_2^{3/2}\le \mathrm{ARE}(\mathrm{NEW},\mathrm{ZYS})\\
&\le\left\{30(1-\sqrt{3}+\pi/3)/\pi^2\right\}^{1/2}c_2^{3/2}/c_1^{3/2}.
\end{aligned}$$

When $c_1 = c_2$, we can draw the same conclusions as Section 3.1 of the manuscript. When $c_1 \neq c_2$, ARE(NEW, ZC) and ARE(NEW, ZYS) depend on different combinations of $(c_1, c_2, q, p, \beta_s, d_s)$, for $s = 1, \ldots, p$. It is worth noting that it is possible to make the proposed test more powerful, provided that $\{15/(2\pi^2)\}^{1/2} c_2^2/c_1^2 > 1$ and $\left\{30(1-\sqrt{3}+\pi/3)/\pi^2\right\}^{1/2} c_2^{3/2}/c_1^{3/2} > 1$. To appreciate this, we assume that $\beta_s = \kappa I(1 \leq s \leq q), s = 1, \ldots, p$, for $\kappa \neq 0$, and $q \in \{1, \ldots, p\}$ satisfies

$$q/p \to c_0 \in [0, 1).$$

Set $d_1 = \ldots = d_q = c_1$ and $d_{q+1} = \ldots = d_p = c_2$. As $p \to \infty$, we have

$$\begin{aligned}
&\mathrm{ARE(NEW, ZC)} \to \{15/(2\pi^2)\}^{1/2}\{c_0 + (c_2/c_1)^2(1-c_0)\}^{1/2}, \\
&\mathrm{ARE(NEW, ZYS)} \to \{30(1-\sqrt{3}+\pi/3)/\pi^2\}^{1/2}\{c_0 + (c_2/c_1)(1-c_0)\}^{1/2}.
\end{aligned}$$

This implies that the proposed test can have power gain than the ZC and ZYS tests if

$$\frac{c_2}{c_1} \geq \max\left(\left[\frac{\{15/(2\pi^2)\}^{-1} - c_0}{1-c_0}\right]^{1/2}, \frac{\{30(1-\sqrt{3}+\pi/3)/\pi^2\}^{-1} - c_0}{1-c_0}\right).$$

Furthermore, we conduct new simulations to verify this finding. The covariates $\mathbf{x}_i = (X_{i1}, \ldots, X_{ip})^{\mathrm{T}}$ are generated from the following moving average model:

$$\begin{cases} X_{is} = \sqrt{c_1}\{\rho_1 Z_{is} + \rho_2 Z_{i(s+1)} + \cdots + \rho_T Z_{i(s+T-1)}\}, \text{ for } s = 1, \ldots, q, \\ X_{is} = \sqrt{c_2}\{\rho_1 Z_{is} + \rho_2 Z_{i(s+1)} + \cdots + \rho_T Z_{i(s+T-1)}\}, \text{ for } s = q+1, \ldots, p, \end{cases} \tag{S.12}$$

for $T = 8$, where Z_{ij} are i.i.d from $N(0, 1)$. We set $c_0 = q/p$, $c_1 = 1$ and

$$c_2 = 2\max\left(\left[\frac{\{15/(2\pi^2)\}^{-1} - c_0}{1-c_0}\right]^{1/2}, \frac{\{30(1-\sqrt{3}+\pi/3)/\pi^2\}^{-1} - c_0}{1-c_0}\right).$$

The sample size $n = 80$, and dimension $p = 550$. Other model settings are remained the same as the Section 4 of manuscript. The empirical sizes and powers at the significance level 5% are summarized in Table S.1. It is clear that the empirical sizes of three tests are reasonably close to 5%. Our proposed test always has the highest power among the three, even when the covariates have bounded variances.

Table S.1: The empirical sizes and powers at the significance level 5%.

		Normal			Gamma		
Model	Hypothesis	ZC	ZYS	NEW	ZC	ZYS	NEW
Linear	H_0	0.054	0.050	0.045	0.055	0.059	0.059
	Non-sparse H_1	0.320	0.489	0.718	0.339	0.528	0.735
	Sparse H_1	0.214	0.319	0.470	0.233	0.328	0.497
Partially Linear	H_0	0.054	0.050	0.045	0.055	0.059	0.059
	Non-sparse H_1	0.300	0.519	0.686	0.324	0.533	0.716
	Sparse H_1	0.141	0.262	0.388	0.135	0.257	0.411
Nonlinear	H_0	0.054	0.050	0.045	0.055	0.059	0.059
	Non-sparse H_1	0.296	0.515	0.752	0.302	0.523	0.750
	Sparse H_1	0.312	0.503	0.687	0.288	0.495	0.697

S.9. Further Discussion on Conditions in Assumption 2

We study the conditions imposed in Assumption 2 when the random vector $\mathbf{x}$ is m-dependent. By the definition of $V(\mathbf{x}_1, \mathbf{x}_2)$, it is straightforward to show that

$$\begin{aligned} E\{V(\mathbf{x}_1, \mathbf{x}_2)^2\} &= 16 \sum_{s=1}^{p} \sum_{t=1}^{p} E[\mathrm{cov}^2\{I(X_{3s} < X_{1s}), I(X_{3t} < X_{2t}) \mid X_{1s}, X_{2t}\}] \\ &\geq 16 \sum_{s=1}^{p} E[\mathrm{cov}^2\{I(X_{3s} < X_{1s}), I(X_{3s} < X_{2s}) \mid X_{1s}\}] \\ &= 8p/45 \to \infty, \end{aligned}$$

$$\begin{aligned} &E\{V(\mathbf{x}_1, \mathbf{x}_2)V(\mathbf{x}_2, \mathbf{x}_3)V(\mathbf{x}_3, \mathbf{x}_4)V(\mathbf{x}_4, \mathbf{x}_1)\} \\ &\quad = E([E\{V(\mathbf{x}_1, \mathbf{x}_3)V(\mathbf{x}_2, \mathbf{x}_3) \mid \mathbf{x}_1, \mathbf{x}_2\}]^2) \geq 0. \end{aligned}$$

Recall that X_s and X_t are independent provided that $|s-t| > m$. By the boundedness of $K_1(\cdot, \cdot)$ and $F_s(\cdot)$, we obtain

$$\begin{aligned}
& E\{V(\mathbf{x}_1, \mathbf{x}_2)V(\mathbf{x}_2, \mathbf{x}_3)V(\mathbf{x}_3, \mathbf{x}_4)V(\mathbf{x}_4, \mathbf{x}_1)\} \\
= & \sum_{s_1=1}^{p} \sum_{s_2=s_1-m}^{s_1+m} \sum_{s_3=s_2-m}^{s_2+m} \sum_{s_4=s_3-m}^{s_3+m} E[K_1\{F_{s_1}(X_{1s_1}), F_{s_1}(X_{2s_1})\}K_1\{F_{s_2}(X_{2s_2}), F_{s_2}(X_{3s_2})\} \\
& \times K_1\{F_{s_3}(X_{3s_3}), F_{s_3}(X_{4s_3})\}K_1\{F_{s_4}(X_{4s_4}), F_{s_4}(X_{1s_4})\}] \le 14pm^3/3,
\end{aligned}$$

and

$$\begin{aligned}
E\{V(\mathbf{x}_1, \mathbf{x}_2)^4\} \quad & = \sum_{s_1=1}^{p} E[K_1\{F_{s_1}(X_{1s_1}), F_{s_1}(X_{2s_1})\}^4] \\
& + \ 4 \sum_{(s_1,s_2)}^{p} E[K_1\{F_{s_1}(X_{1s_1}), F_{s_1}(X_{2s_1})\}^3 K_1\{F_{s_2}(X_{1s_2}), F_{s_2}(X_{2s_2})\}] \\
& + \ 3 \sum_{(s_1,s_2)}^{p} E[K_1\{F_{s_1}(X_{1s_1}), F_{s_1}(X_{2s_1})\}^2 K_1\{F_{s_2}(X_{1s_2}), F_{s_2}(X_{2s_2})\}^2] \\
& + \ 6 \sum_{(s_1,s_2,s_3)}^{p} E[K_1\{F_{s_1}(X_{1s_1}), F_{s_1}(X_{2s_1})\}^2 K_1\{F_{s_2}(X_{1s_2}), F_{s_2}(X_{2s_2})\} \\
& \qquad \times K_1\{F_{s_3}(X_{1s_3}), F_{s_3}(X_{2s_3})\}] \\
& + \sum_{(s_1,s_2,s_3,s_4)}^{p} E[K_1\{F_{s_1}(X_{1s_1}), F_{s_1}(X_{2s_1})\}K_1\{F_{s_2}(X_{1s_2}), F_{s_2}(X_{2s_2})\} \\
& \qquad \times K_1\{F_{s_3}(X_{1s_3}), F_{s_3}(X_{2s_3})\}K_1\{F_{s_4}(X_{1s_4}), F_{s_4}(X_{2s_4})\}] \\
& = \ O\left[pm^3 + E^2\{V(\mathbf{x}_1, \mathbf{x}_2)^2\}\right].
\end{aligned}$$

Then it follows that

$$\begin{aligned}
E\{V(\mathbf{x}_1, \mathbf{x}_2)V(\mathbf{x}_2, \mathbf{x}_3)V(\mathbf{x}_3, \mathbf{x}_4)V(\mathbf{x}_4, \mathbf{x}_1)\}/E^2\{V(\mathbf{x}_1, \mathbf{x}_2)^2\} & \le 1575m^3/(32p), \\
E\{V(\mathbf{x}_1, \mathbf{x}_2)^4\}/[nE^2\{V(\mathbf{x}_1, \mathbf{x}_2)^2\}] & = O\{m^3/(np) + 1/n\},
\end{aligned}$$

which implies that Assumption 2 holds true for $m = o(p^{1/3})$.

S.10. Comparison with Maximum-type Tests

We make some numerical comparisons with the adaptive resampling test (ART) by McKeague and Qian (2015) and its standardized version, as well as the scale-invariance t test by Zhang and Laber (2015). To compare their power performances, we consider three models:

$$
\begin{aligned}
Y_i &= \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_1 + \varepsilon_i, & \text{(S.13)}\\
Y_i &= 3\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_3 + \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_4/2) + \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_2 - 1)\varepsilon_i, & \text{(S.14)}\\
Y_i &= (\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_5)\exp(4\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_2/5 - 1/2) + \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_5/\sqrt{2q}) + \varepsilon_i, & \text{(S.15)}
\end{aligned}
$$

where the generations of $\mathbf{x}_i$ and $\mathbf{z}_i$ are the same as Section 4. The definitions of the coefficients in each model are also identical with Section 4. The error term ε_i is generated from $N(0,1)$. We set $||\boldsymbol{\beta}||^2 = 0.03$ and consider two configurations of alternative hypothesis. (a) Non-sparse case: the total number of active covariates $q = [p^{0.7}]$, where $[x]$ denotes the largest integer not greater than x. (b) Sparse case: the total number of active covariates $q = [2p^{0.3}]$. We fix the significance level α at 0.05 and $(n,p) = (80, 550)$. The critical values of our proposed test are determined by the asymptotically normal distribution. McKeague and Qian (2015) used the double bootstrap to obtain the critical values. In our numerical study, we follow Section 3 of McKeague and Qian (2015) to choose the tuning parameter $\lambda_n = \max\{\sqrt{a \log n}, \text{the upper } \alpha/(2p) \text{ quantile of } N(0,1)\}$, where $a = 2$ for ART and $a = 4$ for standardized ART. To save computational cost due to the double bootstrap, Zhang and Laber (2015) proposed a parametric bootstrap procedure to mimic the limiting null distribution.

Table S.2 reports the empirical sizes and powers of three tests for linear model (S.13). Since the proposed test, standardized ART and Zhang and Laber (2015)'s test are all scale-invariance, their empirical sizes stay the same under different values of δ. This property is not shared by the original ART. When δ is small, the empirical type-I error rates of ART test are slightly inflated. A more careful selection of thresholding value λ_n may be needed to avoid this size distortion. In terms of power, our sum-of-

Table S.2: The empirical sizes and powers for linear model (S.13) at the significance level 5%, where δ controls the degree of heterogeneity in terms of the covariate variances. Throughout, we refer to our proposed test, and the tests proposed by McKeague and Qian (2015) and Zhang and Laber (2015) as NEW, ART and ZL, respectively. STD-ART denotes that the ART procedure is applied to the standardized covariates.

Error	Hypothesis	δ	NEW	ZL	ART	STD-ART
Normal	H_0	0.00	0.045	0.037	0.085	0.045
		0.25	0.045	0.037	0.080	0.045
		0.50	0.045	0.037	0.055	0.045
		0.75	0.045	0.037	0.045	0.045
		1.00	0.045	0.037	0.025	0.045
	Non-sparse H_1	0.00	0.613	0.266	0.505	0.285
		0.25	0.914	0.580	0.740	0.615
		0.50	0.996	0.859	0.780	0.880
		0.75	1.000	0.954	0.735	0.970
		1.00	1.000	0.986	0.690	0.990
	Sparse H_1	0.00	0.459	0.718	0.820	0.700
		0.25	0.672	0.920	0.940	0.910
		0.50	0.855	0.994	0.990	0.995
		0.75	0.956	1.000	0.985	1.000
		1.00	0.991	1.000	0.990	1.000

squares-type test outperforms the other two tests under the non-sparse H_1. Since ART variations and t test of Zhang and Laber (2015) are based on maximum-type statistics, all of them achieve excellent power performances under the sparse H_1, which are superior to our test. Compared with the original ART, the standardized ART really makes a difference, especially for dense alternatives with heteroscedastic covariates. Moreover, the scale-invariance Zhang and Laber (2015)'s test and standardized ART have similar power trends as δ increases. They are useful in sparse settings while ours is powerful in dense ones.

Tables S.3 and S.4 summarize the results of all tests for models (S.14) and (S.15). Similar to the phenomenon in the linear model, our scale-invariance test is still the most powerful to detect dense signals of heteroscedastic covariates. When $\delta = 1$, the empirical powers of our proposed test arrive at 0.932 in model (S.14) and 0.816 in model (S.15) under non-sparse H_1 with normal errors. McKeague and Qian (2015)'s ART tests and Zhang and Laber (2015)'s test show their advantages to deal with sparse alternatives.

Table S.3: The empirical sizes and powers for partially linear model (S.14) at the significance level 5%, where δ controls the degree of heterogeneity in terms of the covariate variances. Refer to the captions in Table S.2 for abbreviations.

Error	Hypothesis	δ	NEW	ZL	ART	STD-ART
Normal	H_0	0.00	0.045	0.037	0.085	0.045
		0.25	0.045	0.037	0.080	0.045
		0.50	0.045	0.037	0.055	0.045
		0.75	0.045	0.037	0.045	0.045
		1.00	0.045	0.037	0.025	0.045
	Non-sparse H_1	0.00	0.613	0.486	0.480	0.520
		0.25	0.810	0.688	0.670	0.710
		0.50	0.920	0.841	0.695	0.865
		0.75	0.961	0.887	0.755	0.875
		1.00	0.932	0.659	0.505	0.665
	Sparse H_1	0.00	0.406	0.621	0.590	0.630
		0.25	0.517	0.723	0.670	0.745
		0.50	0.619	0.812	0.645	0.815
		0.75	0.734	0.867	0.680	0.870
		1.00	0.843	0.922	0.705	0.940

Table S.4: The empirical sizes and powers for nonlinear model (S.15) at the significance level 5%, where δ controls the degree of heterogeneity in terms of the covariate variances. Refer to the captions in Table S.2 for abbreviations.

Error	Hypothesis	δ	NEW	ZL	ART	STD-ART
Normal	H_0	0.00	0.045	0.037	0.085	0.045
		0.25	0.045	0.037	0.080	0.045
		0.50	0.045	0.037	0.055	0.045
		0.75	0.045	0.037	0.045	0.045
		1.00	0.045	0.037	0.025	0.045
	Non-sparse H_1	0.00	0.468	0.221	0.320	0.210
		0.25	0.649	0.287	0.295	0.295
		0.50	0.750	0.370	0.295	0.365
		0.75	0.789	0.419	0.270	0.445
		1.00	0.816	0.447	0.280	0.470
	Sparse H_1	0.00	0.501	0.756	0.780	0.740
		0.25	0.600	0.849	0.810	0.815
		0.50	0.687	0.912	0.705	0.845
		0.75	0.753	0.934	0.605	0.885
		1.00	0.806	0.948	0.575	0.885

S.11. Comparison under Heavy-tailed Covariates

Our test does not require any moment conditions on the covariates, and hence is applicable in the scenarios of generally distributed covariates including the heavy-tailed ones. As suggested by a reviewer, we conduct some simulation studies under heavy-tailed covariates. We generate the $(p+T-1)$-dimensional $\mathbf{z}_i = (Z_{i1}, \ldots, Z_{i(p+T-1)})^{\mathrm{T}}$ from the following two distributions: (i) the first q components of $\mathbf{z}_i$ is independently drawn from $t(2)$ distribution, and others are from $N(0,1)$; (ii) the first q components of $\mathbf{z}_i$ is independently drawn from $t(3)$ distribution, and others are from $N(0,1)$. In these two scenarios, $t(2)$ distribution has an infinite variance while $t(3)$ has a finite one. The error terms follow $N(0,1)$. We set $\delta = 0$, $(n,p) = (80, 550)$ and keep other settings same as those in Section 4. The empirical sizes and powers of our proposed test as well as the ZC and ZYS tests at the significance level $\alpha = 5\%$ are reported in Table S.5. The sizes of all three tests are satisfactory regardless of models and covariate distributions. Under the alternatives, our scale-invariance test is the most powerful to handle heavy-tailed covariates, even in the linear models. For example, the empirical powers of our proposed test and the two competitors are 0.743, 0.470, 0.207 respectively under the non-sparse H_1 in linear model (4.1) with $t(2)$ covariates.

Table S.5: The empirical sizes and powers at the significance level 5%.

		$t(2)$			$t(3)$		
Model	Hypothesis	ZC	ZYS	NEW	ZC	ZYS	NEW
Linear	H_0	0.056	0.061	0.057	0.062	0.049	0.049
	Non-sparse H_1	0.207	0.470	0.743	0.335	0.548	0.734
	Sparse H_1	0.431	0.677	0.755	0.518	0.569	0.662
Partially Linear	H_0	0.056	0.061	0.057	0.062	0.049	0.049
	Non-sparse H_1	0.065	0.377	0.667	0.239	0.502	0.657
	Sparse H_1	0.187	0.410	0.455	0.304	0.414	0.487
Nonlinear	H_0	0.056	0.061	0.057	0.062	0.049	0.049
	Non-sparse H_1	0.181	0.496	0.722	0.297	0.559	0.753
	Sparse H_1	0.340	0.683	0.743	0.524	0.694	0.759

REFERENCE

Blum, J.R., Kiefer, J., and Rosenblatt, M. (1961). “Distribution free tests of independence based on the sample distribution function.” *The Annals of Mathematical Statistics*, **32(2)**, 485–498.

Hall, P. (1984). “Central limit theorem for integrated square error of multivariate nonparametric density estimators.” *Journal of multivariate analysis*, **14(1)**, 1–16.

Kosorok, M. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer.

McKeague, I.W. and Qian, M. (2015). “An adaptive resampling test for detecting the presence of significant predictors (with discussions).” *Journal of the American Statistical Association*, **110(512)**, 1422–1462.

Serfling, R.J. (1980). *Approximation theorems of mathematical statistics*. New York: Wiley.

Zhang, Y. and Laber, E.B. (2015). “Comment: an adaptive resampling test for detecting the presence of signifficant predictors.” *Journal of the American Statistical Association*, **110(512)**, 1451–1454.

Zhu, L.P., Zhu, L.X., and Feng, Z.H. (2010). “Dimension reduction in regressions through cumulative slicing estimation.” *Journal of the American Statistical Association*, **105(492)**, 1455–1466.